LMU
LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN
VOLKSWIRTSCHAFTLICHE FAKULTÄT

Gregor Zöttl:

Emission Trading Systems and the Optimal Technology Mix

Munich Discussion Paper No. 2011-1

Department of Economics
University of Munich

Volkswirtschaftliche Fakultät
Ludwig-Maximilians-Universität München

Emission Trading Systems and the Optimal Technology Mix *

Gregor Zoettl†

University of Munich

February 16, 2011

Abstract

Cap and trade mechanisms enjoy increasing importance in environmental legislation worldwide. The most prominent example is probably given by the European Union Emission Trading System (EU ETS) designed to limit emissions of greenhouse gases, several other countries already have or are planning the introduction of such systems.[2] One of the important aspects of designing cap and trade mechanisms is the possibility of competition authorities to grant emission permits for free. Free allocation of permits which is based on past output or past emissions can lead to inefficient production decisions of firms' (compare for example Böhringer and Lange (2005), Rosendahl (2007), Mackenzie et al. (2008), Harstad and Eskeland (2010)). Current cap and trade systems grant free allocations based on installed production facilities, which lead to a distortion of firms' investment incentives, however.[1] It is the purpose of the present article to study the impact of a cap and trade mechanism on firms' investment and production decisions and to analyze the optimal design of emission trading systems in such an environment.

Keywords: Emissions Trading, Free Allocation, Investment Incentives, Technology Mix.
JEL classification: H21, H23, Q55, Q58, 033

*We want to thank Veronika Grimm, Karsten Neuhoff and Yves Smeers for many valuable comments.
†University of Munich, Dep. of Economics, Ludwigstrasse 28, Munich, Germany, zoettl@lmu.de.
[1]See the present phase II of the EU ETS, compare for example German Parliament (2007) for the case of Germany.

1 Introduction

In the present article we analyze the impact of a cap and trade mechanism on the technology mix of production facilities and on firms' final output decisions. We determine the optimal design of such a mechanism for ideal market conditions but also for non-ideal situations where competition authorities' decisions are partially constrained by requirements of the political or legislative process.

Cap and trade mechanisms designed to internalize social cost of pollution enjoy increasing importance in environmental legislation worldwide. A prominent example is probably given by the European Union Emission Trading System (EU ETS), several other countries already have or are planning the introduction of such systems.[2] An important aspect when introducing cap and trade mechanisms is the possibility of competition authorities to grant emission permits for free. This apparently allowed to crucially facilitate the political processes which finally lead to the introduction of currently adopted cap and trade systems. As Convery(2009) in a recent survey on the origins and the development of the EU ETS observes: "The key quid pro quos to secure industry support in Germany and across the EU were agreements that allocation would take place at Member State level [...], and that the allowances would be free". Very similar observations can also be found in many other contributions to that issue.[3]

Clearly, a one and for all lump sum allocation of permits, which is entirely independent of firms' actions has a purely distributive impact as the seminal contributions to the design of cap and trade mechanisms have already illustrated (See Coase (1960), Dales (1968) and Montgomery (1972)).[4] However, the design of free allocations in currently active cap and trade systems is not very likely to have such lump sum property but will include explicit or implicit features of updating, as several authors argue. Compare for example Neuhoff et al. (2006) who observe for the case of the EU ETS: "For the phase I trading period, incumbent firms received allowances based on their historic emissions.[...] For future trading periods, the Member States have to again define NAPs for the ETS. [...] It is likely that the base

[2]Those include New Zealand, Australia, Canada, United States, for an overview See IEA (2010).

[3]As for example Tietenberg (2006) observes: "free distribution of permits (as opposed to auctioning them off) seems to be a key ingredient in the successful implementation of emissions trading programs".

Bovenberg et al. (2008) state: "The compensation issue has come to the fore in recent policy discussions. For example, several climate change policy bills recently introduced in the U.S. Congress (for example, one by Senator Jeff Bingaman of New Mexico and another by Senator Dianne Feinstein of California) contain very specific language stating that affected energy companies should receive just enough compensation to prevent their equity values from falling."

[4]For a recent discussion of the "conditions under which the independence property is likely to hold both in theory and in practice", see Hahn and Stavis (2010).

period will be adjusted over time to reflect changes in the distribution of plants over time. It is, for example, difficult to envisage that in phase II a government will decide to allocate allowances to a power plant that closed down in phase I. This suggests that some element of 'updating' of allocation plans cannot be avoided if such plans are made sequentially."

Updating of free allocation schemes designed to consistently adapt to an industry's dynamic development has an impact on firms' behavior, however. First, it leads to a distortion of the operation of existing production facilities if firms believe that current output or emissions do have an impact on allocations granted to those facilities in the future.[5] Second, it has an impact on firms' incentives to modify their production facilities through upgrading, retiring and building of new facilities if free (technology specific) allocations are granted for all installed facilities.[6] Most contributions to the literature which analyze the impact of free allocations and the optimal design of emission trading systems have focused on the first effect and abstract from the latter. That is, they provide very rich insight on the impact of updating on firms' production and emission decisions, but abstract from an explicit analysis of firms' incentives to modify their production facilities. The most prominent contributions include Moledina et al. (2003), Böhringer and Lange (2005), Rosendahl(2007), Mackenzie et al. (2008), or Harstadt and Eskeland (2010).[7]

In the present article we want to explicitly analyze the impact of updating on firms' investment incentives which determine their technology mix in the long run. Since the long run implications to a large extent are responsible for the final success of an environmental legislation this seems to provide an important aspect for the ongoing debate on the optimal design of emission trading systems. In order to do so we provide an analytical framework with an endogenous emission permit market where (strategic) firms chose to invest in two different production technologies (with different emission intensities) which allow for production during a longer horizon of time. We then analyze the impact of a cap and trade

[5]Compare for example Böhringer and Lange (2005):"As a case in point, one major policy concern is that [...] the allocation should account for (major) changes in the activity level of firms. Free allocation schemes must then abstain from lump-sum transfers and revert to output- or emission-based allocation."

[6]A recent review of current emission trading schemes by the International Energy Agency (IEA 2010) reveals that most legislations which provide free emission permits do update their allocation schemes:" An important detail of systems using grandfathered allocation is the treatment of companies that establish new facilities or close down. Current or proposed schemes generally provide new entrants with the same support as existing facilities. The rationale for this is to avoid investment moving to jurisdictions without carbon pricing." A prominent example in this respect is given by the legislation currently observed in phase II of the EU ETS, in the case of Germany for example new production facilities receive technology specific free allocations when they start operations, retiring facilities loose their allocations, compare German Parliament (2007).

[7]In a recent empirical study on phase I of the EU ETS Anderson and Di Maria (2011) indeed find evidence that firms' output decisions have been inflated, "possibly due to future policy design features".

mechanism on firms' technology choices and their production decisions. As a benchmark we determine the first best solution. Analogous to the previous literature, if distributional concerns do not matter, in an ideal market with perfectly competitive firms it is optimal to grant no free allocation to any technology and to set the total emission cap such that the permit price equals to marginal social cost of pollution.

In the main part of the paper we then analyze the optimal design of a cap and trade system if the market is not ideal. First, we consider the case that firms behave imperfectly competitive when making their investment and their production decisions. It is then optimal to grant free allocations in order to stimulate inefficiently low investment incentives. As we show, however, in a closed system with endogenous permit market it is not optimal to implement total investment at first best levels since this would imply an inefficiently high permit price. It can be optimal, furthermore, to set free allocations such as to induce firms to choose a technology mix which is even cleaner than in the first best scenario in order to depress the endogenous permit price.[8]

Second, we analyze the case where the design of the cap and trade mechanism is subject to political constraints (as extensively discussed above) and the competition authority has to determine the optimal market design given those constraints.[9] We first analyze how the optimal target on total emissions should be set in case free allocations in all technologies are exogenously fixed. As we find, for moderate levels of free allocations the target on total emissions should be set such that the equilibrium permit price is above marginal social cost of pollution. For high levels of free allocation (as for example for the case of full allocation where all permits used by a certain technology during a compliance period are freely allocated, compare for example German Parliament (2007))[6] the total cap on emissions should be set such that the equilibrium permit price should be below marginal social cost of pollution.

We then analyze the case that free allocation only for a specific technology is exogenously fixed and determine the optimal level of free allocation for the remaining technology. In order to avoid excessive distortions of the resulting technology mix it is typically optimal to grant free allocation for the remaining technology. That is, the insights obtained from

[8]Those results have a direct implication also for other measures designed to stimulate investment incentives of firms, as for example capacity mechanisms introduced in electricity markets, compare for example Cramton and Stoft (2008).

[9]Observe that to some extend this parallels the fundamental approach found in the previous literature: Böhringer and Lange(2005) provide second best rules if (for political reasons) updating has to be based on past output, Harstadt and Eskeland (2010) analyze market design in case governments cannot commit to full auctioning of permits and Bovenberg et al. (2005, 2008) consider the constraint that firms have to be fully compensated for the regulatory burden.

the first best benchmark that free allocations are never optimal are no longer true in case allocation to one of the technologies is exogenously fixed. Moreover, if this technology is relatively dirty (as compared to the technology with exogenously fixed allocation) the level of free allocation should remain below the exogenously fixed allocation. If on the contrary the remaining technology is relatively clean, the level of free allocation should even be above the exogenously fixed allocation. Observe that the current practice of full allocation (as currently granted in phase II of the EU ETS, compare German Parliament (2007) for the case of Germany) induces a pattern of free allocation which is completely opposed to those findings.

Let us finally mention that from a modeling perspective the present paper also contributes to the literature of peak load pricing which analyzes optimal investment decisions in several technologies. For a survey on this literature see Crew and Kleindorfer (1995). More recent contributions include Ehrenmann and Smeers (2011), Zöttl (2010), or Zöttl (2011).[10] Our framework introduces an endogenous emission permit market with the purpose to internalizes social cost of emissions. This setup allows us to analyze the optimal design of a cap and trade mechanism by taking into account firms' investment and production decisions.

The remainder of the article is structured as follows: Section 2 states the model analyzed throughout this article, section 3 derives the market equilibrium for a given cap and trade mechanism. In section 4 we determine the optimal markets design, section 5 concludes.

2 The Model

We consider n firms which first have to choose production facilities from two different technologies prior to competing on many consecutive spot markets with fluctuating demand. Inverse Demand is given by the function $P(Q,\theta)$, which depends on $Q \in \mathbb{R}^+$, and the variable $\theta \in \mathbb{R}$ that represents the demand scenario. The parameter θ takes on values in the interval $[\underline{\theta},\overline{\theta}]$ with frequencies $f(\theta)$. The corresponding distribution is denoted $F(\theta) = \int_{\underline{\theta}}^{\theta} f(\theta)d\theta$.[11] We denote by $q(\theta) = (q_1(\theta),\ldots,q_n(\theta))$ the vector of spot market outputs of the n firms in demand scenario θ, and by $Q(\theta) = \sum_{i=1}^{n} q_i$ total quantity produced in scenario

[10] Based on those analytical frameworks a number of numerical studies tries to quantify the impact of a cap and trade mechanism on firms' investment decisions for different levels of an exogenously fixed permit price (compare for example Neuhoff et al. (2006), Matthes (2006) or recently Pahle, Fan and Schill (2011)).

[11] Mathematically we treat the frequencies associated to the realizations of θ by making use of a density and a distribution-function. Notice, however, that there is no uncertainty in the framework presented — all realizations of $\theta \in [\underline{\theta},\overline{\theta}]$ indeed realize, with the corresponding frequency $f(\theta)$.

θ. Demand in each scenario satisfies standard regularity assumptions, i.e.[12]

Assumption 1 (Demand) *Inverse demand satisfies* $P_q(Q,\theta) < 0$, $P_\theta(Q,\theta) \geq 0$, $P_{q\theta}(Q,\theta) \geq 0$ *and* $P_q(Q,\theta) + P_{qq}(Q,\theta)\frac{Q}{n} < 0$ *for all* $Q, \theta \in \mathbb{R}$.

Technologies differ with respect to investment and production cost and emission factors.

Assumption 2 (Technologies) *Firms can choose between two different technologies, t=1,2. Each technology t has constant marginal cost of investment* k_t*, constant marginal cost of production* c_t*, and an emission factor* w_t *which measures the amount of the pollutant emitted per unit of output.*

We denote *total* investment of firm i in *both* technologies by x_{1i} and investment of firm i in technology 2 by x_{2i}, aggregate total investment is denoted by X_1 and aggregate investment in technology 2 by X_2.[13] We denote aggregate output produced in scenario θ by $Q(\theta)$. Each unit of output produced with technology $t = 1, 2$ causes emissions w_t. We denote total emissions (for example of a greenhouse gas) produced at all markets $\theta \in [\underline{\theta}, \bar{\theta}]$ by $\mathcal{T}$. The social cost associated to emissions is denoted by $D(\mathcal{T})$. The competition authority designs a cap and trade mechanism to internalize this social cost.

Assumption 3 (Cap and Trade Mechanism and Social Cost of Pollution) *Total Pollution* $\mathcal{T}$ *causes a social damage* $D(\mathcal{T})$*, which satisfies* $D_T(\mathcal{T}) \geq 0$ *and* $D_{TT}(\mathcal{T}) \geq 0$*. A cap and trade mechanism limits total emissions such that* $\mathcal{T} \leq T$*. Each unit invested in technology* $t = 1, 2$ *is assigned the amount* A_t *of permits for free.*

Permits are tradeable, we make the following assumptions regarding the permit market.

Assumption 4 (Permit Markets) *(i) Emission permit trading is arbitrage–free and storage of permits is costless.*

(ii) Firms are price takers at the permit market.[14]

We denote the market price for emission permits by e. For given investment decisions of a firm (x_{1i}, x_{2i}) we can now write down marginal production cost of firm i as follows:

$$C(q_i, x_{1i}, x_{2i}) = \begin{cases} c_2 + w_2 e & \text{for} \quad 0 < q_i \leq x_{2i}, \\ c_1 + w_1 e & \text{for} \quad x_{2i} < q_i \leq x_{1i}, \\ \infty & \text{for} \quad x_{1i} < q_i. \end{cases}$$

[12]We denote the derivative of a function $g(x,y)$ with respect to the argument x, by $g_x(x,y)$, the second derivative with respect to that argument by $g_{xx}(x,y)$, and the cross derivative by $g_{xy}(x,y)$.

[13]Thus, aggregate investment in technology 1 is given by $X_1 - X_2$.

[14]Since emission trading systems typically encompass large regions (several countries in the case of the EU ETS) this seems to be a quite natural assumption.

To sum up, at the first stage, firms simultaneously invest in the two different technologies at marginal cost of investment k_1, k_2. Investment choices are observed by all firms. Then, given their investment choices, firms compete at a sequence of spot markets with fluctuating demand in the presence of a cap and trade mechanism. At each spot market θ, firms simultaneously choose output $q_i(\theta)$ which causes emissions. Each firm i has to cover its total emissions by permits. Depending on the allocation rule (A_1, A_2) firms obtain permits for free, contingent on their investment decision. Firms have to purchase permits needed in excess of the free allocation at the permit market at price e, which is the price at which the permit market clears given the target T.

3 The Market Equilibrium

In this section we derive the market equilibrium with cap and trade mechanism, for the case of perfect and imperfect competition. Observe that in the framework analyzed, where demand fluctuates over time it is optimal for firms to invest into a mix of both technologies. We will consider the case that technology 2 allows cheaper production but exhibits higher investment cost. Those units have to run most of the time in order to recover their high investment cost(this is typically denoted "baseload–technology"). Technology 1 has relatively low investment cost but produces at high marginal cost. Those units are built in order to serve during periods of high demand (this is typically denoted "peakload–technology") but run idle if demand is low. In order to be able to characterize the market equilibrium for a given cap and trade mechanism (T, A_1, A_2), we first determine firms' profits, given investments x_1, x_2 and given spot market output $q(\theta)$.

$$\begin{aligned}\pi_i(x_{1i}, x_{2i}) \;=\; & \int_{\underline{\theta}}^{\theta_{\overline{B}}} \left(P(Q,\theta) - c_2 - w_2 e\right) q_i(\theta, x) dF(\theta) + \int_{\theta_{\overline{B}}}^{\theta_{\underline{P}}} \left(P(X_2,\theta) - c_2 - w_2 e\right) x_{2i} dF(\theta) \qquad (1)\\ & + \int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} \left(P(Q,\theta) - c_1 - w_1 e\right) q_i(\theta, x) dF(\theta) + \int_{\theta_{\overline{P}}}^{\overline{\theta}} \left(P(X_1,\theta) - c_1 - w_1 e\right) x_{1i} dF(\theta)\\ & - \int_{\theta_{\underline{P}}}^{\overline{\theta}} \left((c_1 + w_1 e) - (c_2 + w_2 e)\right) x_{2i} dF(\theta) - (k_2 - A_2 e) x_{2i} - (k_1 - A_1 e)(x_{1i} - x_{2i}).\end{aligned}$$

Note that the permit market affects both, the firms' marginal production cost as well as their investment cost. No matter whether permits have been allocated for free or have to be bought at the permit market, firms face opportunity cost of $w_t e$ when deciding to produce one unit of output with technology $t = 1, 2$. This opportunity cost increases their marginal production cost to $c_t + w_t e$, $t = 1, 2$. Investment cost is affected by the firms' anticipation of a free allocation of permits. A free allocation is equivalent to a subsidy paid upon investment: If each unit of capacity invested is assigned A_t permits, investment cost k_t is reduced by their value, that is by $A_t e$ for $t = 1, 2$.

The critical spot market scenarios[15] $\theta_{\overline{B}}, \theta_{\underline{P}}, \theta_{\overline{P}}$ indicate wether firms produce either at the capacity bounds x_2, x_1 (that is, at the vertical pieces of their marginal cost curves), or on the flat (i.e. unconstrained) parts of their marginal cost curves. They depend on the intensity of competition at the spot market and are illustrated in Figure 1 both for the case of perfect and imperfect competition. For $\theta \in [\underline{\theta}, \theta_{\overline{B}}]$ firms produce the output at

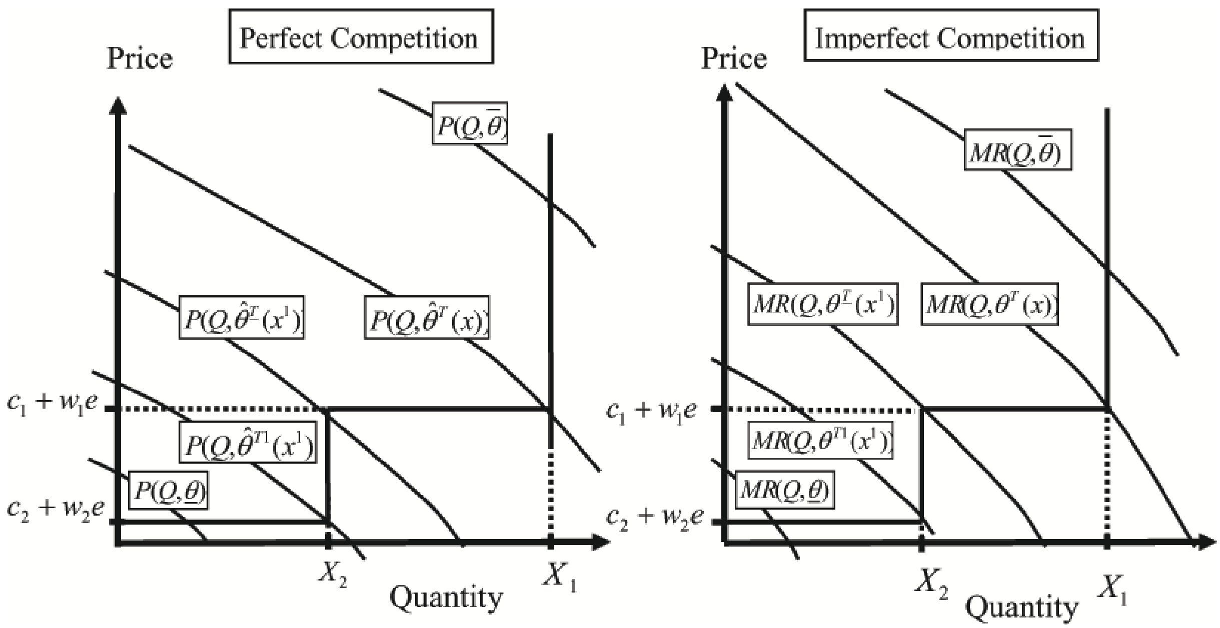

Figure 1: Illustration of the critical spot market scenarios. Left: The case of a perfectly competitive market, right: the general case with imperfect competition. In the figure we denote marginal revenue by $MR(Q, \theta) := P(Q, \theta) + P_q(Q, \theta)\frac{Q}{n}$.

marginal cost c_2. For $\theta \in [\theta_{\overline{B}}, \theta_{\underline{P}}]$ firms are constrained by their investment in the base load technology and produce X_2, still at marginal cost c_2, and prices are driven by the demand function. At those demand levels, using the peak load technology 1 is not yet profitable. Observe that $F(\theta_{\underline{P}}) - F(\theta_{\overline{B}})$ measures the fraction of time where investment in the base load technology is binding, which we will refer to as *constrained base duration*. For $\theta \in [\theta_{\underline{P}}, \theta_{\overline{P}}]$ firms produce output at marginal cost c_1, we denote $1 - F(\theta_{\underline{P}})$ as *peak duration*.[16] Finally, for all realizations above $\theta_{\overline{P}}$, firms are constrained by their total capacity choice X_1, and

[15]For the precise definition of those critical spot market scenarios, see appendix A.

[16]The equivalent *base duration* would be given by $1 - F(\underline{\theta}) = 1$, it is not explicitly introduced, however.

prices are driven exclusively by the demand function, we denote $1 - F(\theta_{\bar{P}})$ as *constrained peak duration*. In the subsequent lemma we characterize the market equilibrium when firms invest in the base load and in the peak load technology.

LEMMA 1 *For a given cap and trade mechanism (T, A_1, A_2), define the toal investment condition Ψ_I, the base investment condition Ψ_{II} and the permit pricing condition*[17] *Ψ_E as follows:*

$$\Psi_I := \int_{\theta_{\overline{P}}}^{\bar{\theta}} \left[P(X_1^*, \theta) + P_q(X_1^*, \theta) \frac{X_1^*}{n} - (c_1 + w_1 e^*) \right] dF(\theta) - (k_1 - A_1 e^*) \tag{2}$$

$$\Psi_{II} := \int_{\theta_{\overline{B}}}^{\theta_{\underline{P}}} \left[P(X_2^*, \theta) + P_q(X_2^*, \theta) \frac{X_2^*}{n} - (c_2 + w_2 e^*) \right] dF(\theta) + \tag{3}$$
$$\int_{\theta_{\underline{P}}}^{\bar{\theta}} (c_1 - c_2) + (w_1 - w_2) e^* dF(\theta) - (k_2 - k_1) + (A_2 - A_1) e^*$$

$$\Psi_E := \int_{\underline{\theta}}^{\theta_{\overline{B}}} w_2 Q(e^*, \theta) dF(\theta) + \int_{\theta_{\overline{B}}}^{\theta_{\underline{P}}} w_2 X_2^* dF(\theta) + \int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} w_1 Q(e^*, \theta) dF(\theta) \tag{4}$$
$$+ \int_{\theta_{\overline{P}}}^{\bar{\theta}} w_1 X_1^* dF(\theta) - \int_{\theta_{\underline{P}}}^{\bar{\theta}} (w_1 - w_2) X_2^* dF(\theta) - T$$

Equilibrium investment X^{1*}, X^{2*} and the equilibrium permit price e^* simultaneously solve $\Psi_I = \Psi_{II} = \Psi_E = 0$.

PROOF See appendix A. □

In the lemma, (2) is the first order condition that determines total investment. Firms choose their total investment $\frac{X_1^*}{n}$ as to equal marginal profits generated by their last running unit (running at total marginal cost $c_1 + w_1 e^*$) to the investment cost of that unit (given by $k_1 - A_1 e^*$). As already mentioned above, under a cap and trade mechanism, the value of the permits reuired for production at the spot market is part of the firms' marginal production cost, the value of free allocations is of firms' marginal cost of investment.

Now let us provide some intuition on the determinants of the optimal base load investment. Since total investment X_1 has already been fixed (it is determined by (2)), the firms' decision when choosing X_2 has to be interpreted as a decision of virtually replacing units of technology 1 by technology 2. The cost of such virtual replacement of the marginal unit (given by $k_2 - k_1 - (A_2 - A_1)e^*$) has to equal the extra profits generated by that unit

[17]For a positive permit price, we might also obtain the situation, where production for very low demand realizations is suppressed and positive output is produced only for demand realizations which satisfy θ : $P(0, \theta) - C(0, \theta) - e > 0$. For ease of notation we disregard this corner solution, which could be easily included in the entire analysis.

due to lower marginal production cost. Lower production cost of one additional unit has two effects: First, for all demand realizations $\theta \in [\theta_{\overline{B}}, \theta_{\underline{P}}]$ one more unit is produced (that would not have been produced without the replacement); for $\theta \in [\theta_{\underline{P}}, \overline{\theta}]$ one more unit can be produced at lower marginal cost $c_2 + w_2 e^*$ (instead of $c_1 + w_1 e^*$) due to the replacement. (Compare also figure 1).

The market price for permits, e^*, depends on the emission target T set by the market designer as well as the technology mix installed by the firms. At the equilibrium permit price the market exactly clears, allowing for total emission of T units of the pollutant. Notice that the left hand side of expression (4) is just total production at all spot markets multiplied by the emission factors of the respective Technologies (w_1, w_2), total emissions are obtained by integrating over emissions at all spot markets $\theta \in [\underline{\theta}, \overline{\theta}]$.

Finally observe that lemma 1 characterizes the market solution when firms decide to invest in both technologies, that is, when indeed $0 < X_2^* < X_1^*$ obtains. First, whenever the base load technology (k_2, c_2) is very unattractive,[18] then only the peak load technology (k_1, c_1) is active. Second, if the base load technology (k_2, c_2) is always more attractive[19] than the peak load technology (k_1, c_1), then only technology (k_2, c_2) is active in the market equilibrium. Notice that in principle the case of investment in a single technology is covered by out framework, it obtains by eliminating the possibility to invest in technology 2, expression (2)) then determines investment in the single technology. To keep the notational burden limited, however, we do not explicitly include those corner solution in the exposition of the paper, but opted to focus on all those cases when firms indeed choose to investment in both technologies.

To conclude the discussion of lemma 1 let us already at this point mention the relevance of endogenously modeling the emission permit market as compared to the case which assumes an exogenously fixed price for pollution. Observe that, for a constant emission price equilibrium investment under imperfect competition differs from that obtained under perfect competition by the terms $\int_{\theta_{\overline{P}}}^{\overline{\theta}} P_q(X_1, \theta)\frac{X_1}{n}$ and $\int_{\theta_{\overline{B}}}^{\theta_{\underline{P}}} P_q(X_2, \theta)\frac{X_2}{n} dF(\theta)$ respectively, which corresponds to the difference between scarcity prices and marginal scarcity profits. Since those terms are negative (and profits concave given our assumptions) investment incentives under imperfect competition are lower than under perfect competition. That is, in the absence of an explicit market for emission permits (when pollution is for example taxed at some fixed level e^0) subsidies for investment (for example by granting free tax vouchers $A_1 > 0$ and $A_2 > A_1$ respectively) which exactly compensate for those differences would induce optimal investment incentives. Since the emission price is endogenous in our

[18]That is expression (3) yields $X_2^* \leq 0$.

[19]That is expressions (3) and (2) yield $X_2^* \geq X_1^*$.

framework, however, we will obtain a different result (compare theorem 2).

Before we now discuss existence of the market equilibrium we introduce the following definitions which will simplify the subsequent analysis and allow for a more intuitive discussion of our results:

DEFINITION 1 *(i) We denote the impact of increased total investment on total emissions (for fixed e) by* $A_1^E := \frac{\partial \Psi_E}{\partial X_1^*} = (1 - F(\theta_{\bar{P}}))\, w_1$*, observe* $A_1^E > 0$*. This allows to state the impact of changed emission price* e^* *on the equilibrium condition* Ψ_I *as follows* $\frac{\partial \Psi_I}{\partial e^*} = A_1 - A_1^E$*.*

(ii) We denote the impact of increased base load investment on total emissions (for fixed e) by $A_2^E := \frac{\partial \Psi_E}{\partial X_2^*} = (1 - F(\theta_{\bar{B}}))\, w_2 - (1 - F(\theta_{\underline{P}}))\, w_1$*. This allows to state the impact of changed emission price* e^* *on the equilibrium condition* Ψ_{II} *as follows* $\frac{\partial \Psi_{II}}{\partial e^*} = A_2 - A_1 - A_2^E$*. We furthermore denote* $w_2^E := \frac{1-F(\theta_{\underline{P}})}{1-F(\theta_{\bar{B}})} w_1$ *(which implies* $A_2^E > 0$ *if and only if* $w_2 > w_2^E$*) and* $w_2^L := \frac{F(\theta_{\overline{P}})-F(\theta_{\underline{P}})}{1-F(\theta_{\bar{B}})} w_1$ *(which implies* $A_1^E + A_2^E > 0$ *if and only if* $w_2 > w_2^L$*).*

(iii) We denote the impact of changed X_1 *on the equilibrium condition* Ψ_I *by* $\Psi_{I1} := \frac{\partial \Psi_I}{\partial X_1^*}$*, the impact of changed* X_2 *on the equilibrium condition* Ψ_{II} *by* $\Psi_{II2} := \frac{\partial \Psi_{II}}{\partial X_2^*}$ *and the impact of changed e on the equilibrium condition* Ψ_E *by* $\Psi_{Ee} := \frac{\partial \Psi_E}{\partial e^*}$*. Observe that those three expressions are negative.*

Observe[20] that $A_1^E = (1 - F(\theta_{\bar{P}}))\, w_1$ determines the total amount of additionally necessary permits resulting from an additionally invested unit of total capacity (formally given by the partial derivative of total emission with respect to X_1, i.e. $\frac{\partial \Psi_E}{\partial X_1}$). An increase of the permit price e now has two opposing effects on total investment incentives: on the one hand investment incentives are reduced by the amount A_1^E, on the other hand they increase by A_1 due to the increased value of free allocations.

A similar reasoning obtains for investment incentives in the base load technology. A_2^E determines the total amount of additionally necessary permits resulting from the replacement of one unit of the peak technology with one unit of the base technology (formally given by the partial derivative of total emission with respect to X_2, i.e. $\frac{\partial \Psi_E}{\partial X_2}$). An increase of the permit price e has two opposing effects on total investment incentives: on the one hand they are reduced by the amount A_2^E, on the other hand they increase by $(A_2 - A_1)$ due to the increased value of free allocations.

[20]Notice that the statements of definition 1 and the subsequent discussion exclusively refer to partial derivatives. In equilibrium total emissions do not change since they are capped at T.

Notice that $A_1^E \geq 0$ whereas A_2^E can also become negative. That is, an increased level of total investment X_1^* always implies additionally necessary emission permits. An increased level of base investment X_2^* does only imply additionally necessary emission permits if the base technology is "dirtier" than the peak technology. Interestingly the cut–off point obtains for $w_2 = w_2^E < w_1$, since an increased level of X_2^* leads to increased emissions for $\theta \in [\theta_{\underline{P}}, \overline{\theta}]$ if $w_2 > w_2$ but also leads to one unit of additional output for the demand levels $\theta \in [\theta_{\overline{B}}, \theta_{\underline{P}}]$.

As already argued, lemma 1 only characterizes the market equilibrium by establishing necessary conditions. In the subsequent lemma ne now want to establish conditions second order conditions for the existence of the market equilibrium.

Lemma 2 (Second Order Conditions) *(i) Lemma 1 characterizes the market equilibrium if*

(a) $\left(A_1 - A_1^E\right) A_1^E - \Psi_{I1}\Psi_{Ee} < 0,$ (b) $\left(A_2 - A_1 - A_2^E\right) A_2^E - \Psi_{II2}\Psi_{Ee} < 0$

(c) $\left(\left(A_1 - A_1^E\right) A_1^E - \Psi_{I1}\Psi_{Ee}\right)\left(\left(A_2 - A_1 - A_2^E\right) A_2^E - \Psi_{II2}\Psi_{Ee}\right) > \left(A_1 - A_1^E\right) A_1^E \left(A_2 - A_1 - A_2^E\right) A_2^E$

(ii) If the levels of free allocation satisfy $\left(A_1 - A_1^E\right) A_1^E \leq 0$ *and* $\left(A_2 - A_1 - A_2^E\right) A_2^E \leq 0$, *then condition (i) is satisfied.*

(iii) Define by A_1^{lim} *the highest* A_1 *yielding* $\left(A_1 - A_1^E\right) A_1^E - \Psi_{I1}\Psi_{Ee} \leq 0$, *define by* A_2^{lim} *the highest* A_2 *yielding* $\left(A_2 - A_1^E - A_2^E\right)\left(A_1^E + A_2^E\right) - \left(\Psi_{I1} + \Psi_{II2}\right)\Psi_{Ee} \leq 0$. *The second order conditions (i) cannot be satisfied if either* $A_1 \geq A_1^{lim}$, *or* $A_2 \geq A_1^{lim}$.

Proof See appendix B. □

Part (i) of the lemma establishes the standard second order conditions which establishes negative semi–definiteness of the Hessian matrix of firms' optimization problem. It allows the usual application of the implicit function theorem in order to conduct an analysis of comparative statics for the equilibrium characterized in lemma 1. In part (ii) we establish conditions when those second order conditions are satisfied and part (iii) provides an upper bound on the levels of free allocation such that higher allocations always violate those second order conditions.

Let us explicitly mention at this point that our analysis throughout this article focuses on symmetric investment decisions, the second order conditions established in lemma 2(i) guarantee that lemma 1 characterizes a unique symmetric solution. Since for the case of a monopolistic or a perfectly competitive market asymmetric investment levels are irrelevant[21] lemma 2(i) guarantees a existence and uniqueness of the market equilibrium in those cases.

[21]For perfect competition observe that both marginal cost of investment and marginal cost of production are constant, for monopoly observe that asymmetries cannot arise by definition.

For the case of oligopoly, when firms behave strategically, asymmetric investment levels might be relevant, however. Indeed, as we show in a companion paper (Zoettl(2010)) for investment decisions in a discrete number of technologies symmetric equilibria can only exist if technologies are sufficiently different, for sufficiently similar technologies a symmetric equilibrium of the investment game always fails to exist and asymmetric equilibria might arise.[22]

After having established the market equilibrium, we now determine the impact of changing the parameters of the cap and trade mechanism (A_1, A_2, T) in an analysis of comparative statics. If the second order conditions specified in lemma 2 (i) are satisfied we obtain the following results:

LEMMA 3 (COMPARATIVE STATICS OF THE MARKET EQUILIBRIUM) *(i) Higher free allocation for the base load technology A_2 always yields higher investment in the base load technology (i.e. $\frac{dX_2}{dA_2} > 0$). We furthermore obtain $\frac{dX_1}{dA_2} < 0$ if and only if $\left(A_1 - A_1^E\right) A_2^E < 0$. Define A_1^{cross} as the highest A_1 yielding $\left(A_1 - A_1^E\right) A_2^E \leq \Psi_{Ee}\Psi_{I1} - \left(A_1 - A_1^E\right) A_1^E$, we obtain $\frac{dX_2^*}{dA_2} < \frac{dX_1^*}{dA_2}$ if and only if $\left(w_2 > w_2^E\right)$ and $A_1 \in (A_1^{cross}, A_1^{lim})$.*

(ii) Higher free allocation for the peak load technology A_1 always yields higher investment in the peak load technology (i.e. $\frac{dX_1}{dA_1} > \frac{dX_2}{dA_1}$). Define by A_2^{total} the highest A_2 which yields $\left(A_2 - A_1^E - A_2^E\right) A_2^E - \Psi_{II2}\Psi_{Ee} \leq 0$ and by A_2^{cross} the highest A_2 which yields $\left(A_2 - A_1^E - A_2^E\right) A_1^E - \Psi_{I1}\Psi_{Ee} \leq 0$. There exists a unique w_2^S with $w_2^E < w_2^S \leq w_1$ such that $\frac{dX_1^*}{dA_1} < 0$ if and only if $w_2 > w_2^S$ and $A_2 \in (A_2^{total}, A_2^{lim})$. Furthermore, we obtain $\frac{dX_2^*}{dA_1} > 0$ if and only if $w_2 < w_2^S$ and $A_2 \in (A_2^{cross}, A_2^{lim})$.

(iii) For a change of the total emission cap T we obtain $\frac{dX_1^*}{dT} > 0$ if and only if $\left(A_1 < A_1^E\right)$, we furthermore obtain $\frac{dX_2^*}{dT} > 0$ if and only if $(A_2 - A_1 < A_2^E)$.

PROOF See appendix C. □

As we establish in the theorem, an increase of the free allocation A_2 in the base load technology always leads to increased base load investment (i.e. $\frac{dX_2^*}{dA_2} > 0$, see point (i)), an increase of the free allocation A_1 in the peak load technology always leads to increased investment in the peak load technology (i.e. $\frac{dX_1^*}{dA_1} > \frac{dX_2^*}{dA_1}$, see point (ii)). The impact of

[22]As analyzed in Zoettl(2010) those problems can be overcome when firms are allowed to choose from a continuum of technologies, when existence and uniqueness of the symmetric equilibrium can be reestablished. Those findings to some extend seem to be parallel the discussion on supply function equilibria, where Klemperer and Meyer (1989) show existence and uniqueness of the market equilibrium when firms can bid smooth supply functions, whereas von der Fehr and Harbord (1993) show than a symmetric equilibrium in discrete step functions fails to exist.

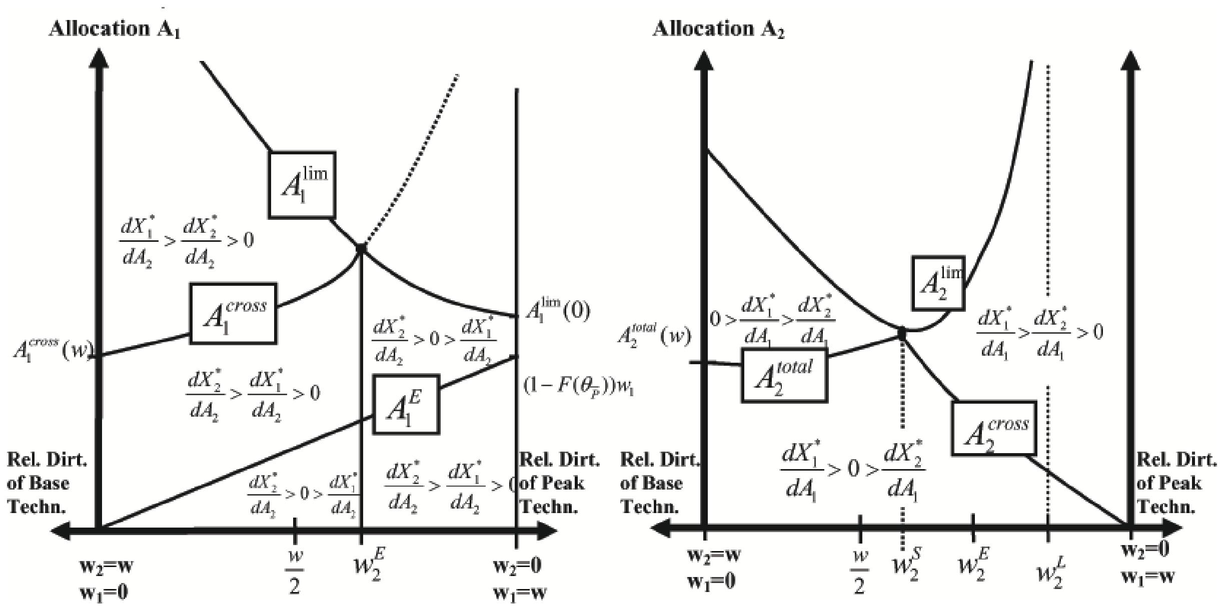

Figure 2: Results of comparative statics in the degree of free allocation. Left: For the degree of free allocation to the base load technology A_2, Right: For the degree of free allocation to the peak load technology A_1. For the case of linear demand we obtain,
left: $A_1^{cross}(w) = \frac{1-F(\theta_{\overline{P}})}{1-F(\theta_{\overline{B}})} F(\theta_{\overline{B}}) w_2$, $A_1^{cross}(w_2^E) = w_2^E$, $A_1^{lim}(0) = (1 - F(\theta_{\underline{P}}))\, w_1$ and
right: $A_2^{total}(w) = \left(1 - \frac{1-F(\theta_{\overline{P}})}{1-F(\theta_{\overline{B}})} F(\theta_{\overline{B}})\right) w_2$, $A_2^{cross}(w_2^S) = w_2^E$.

such changes on the remaining investment decisions is more ambiguous. In the subsequent paragraphs we briefly sketch the central trade-offs, a complete proof is only provided in the appendix, however. First consider a variation of the free allocation A_2 and determine its impact on the system of equilibrium conditions established in lemma 1. The total differential yields:[23]

$$\frac{d\Psi_I}{dA_2} = \Psi_{I1}\frac{dX_1^*}{dA_2} + \Psi_{Ie}\frac{de^*}{dA_2} = 0 \tag{5}$$

$$\frac{d\Psi_{II}}{dA_2} = \Psi_{II2}\frac{dX_2^*}{dA_2} + \Psi_{IIe}\frac{de^*}{dA_2} + \frac{\partial\Psi_{II}}{\partial A_2} = 0 \tag{6}$$

$$\frac{d\Psi_E}{dA_2} = \Psi_{E1}\frac{dX_1^*}{dA_2} + \Psi_{E2}\frac{dX_2^*}{dA_2} + \Psi_{Ee}\frac{de^*}{dA_2} = 0 \tag{7}$$

In order to directly evaluate the impact of the changed emission price $\frac{de^*}{dA_2}$ on the equilibrium conditions for total investment and investment in the base load technology, we solve

[23] For a better traceability of our computations we denote the partial derivatives $\frac{\partial\Psi_I}{\partial e^*} = \Psi_{Ie}$, $\frac{\partial\Psi_{II}}{\partial e^*} = \Psi_{IIe}$, $\frac{\partial\Psi_E}{\partial X_1^*} = \Psi_{E1}$ and $\frac{\partial\Psi_E}{\partial X_2^*} = \Psi_{E2}$, in a second step we make use of A_1^E and A_2^E introduced in definition 1.

expression (7) for $\frac{de^*}{dA_2} = \frac{\Psi_{E1}}{-\Psi_{Ee}}\frac{dX_1^*}{dA_2} + \frac{\Psi_{E2}}{-\Psi_{Ee}}\frac{dX_2^*}{dA_2}$ and plug into expression (5), which yields:

$$\begin{aligned} \frac{d\Psi_I}{dA_2} &= \left(\Psi_{I1} + \Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right)\frac{dX_1^*}{dA_2} + \left(\Psi_{Ie}\frac{\Psi_{E2}}{-\Psi_{Ee}}\right)\frac{dX_2^*}{dA_2} = 0 \\ \Leftrightarrow & \quad \left(-\Psi_{Ee}\Psi_{I1} + \left(A_1 - A_1^E\right)A_1^E\right)\frac{dX_1^*}{dA_2} + \left(\left(A_1 - A_1^E\right)A_2^E\right)\frac{dX_2^*}{dA_2} = 0 \end{aligned} \tag{8}$$

Observe that the coefficient on the expression $\frac{dX_1^*}{dA_1}$ determines the total impact of changed X_1^* on the equilibrium condition Ψ_I. This is given by the direct impact (i.e. Ψ_{I1}) and the indirect impact which takes into account the impact of changed X_1^* on the emission price and its feed back on the equilibrium condition Ψ_I (i.e. $\Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}} = \frac{1}{-\Psi_{Ee}}\left(A_1 - A_1^E\right)A_1^E$). Observe that the total impact of changed X_1^* on the equilibrium condition Ψ_I is negative if the second order conditions established in lemma 2(i) are to be satisfied. This directly illustrates why $\frac{dX_2^*}{dA_2}$ cannot drop to zero.

Furthermore, observe that the total impact of changed X_2^* on the equilibrium condition Ψ_I is only indirect, since Ψ_I does not directly depend on X_2. That is, we only have to take into account the impact of increased X_2^* on the emission price e^* and its feedback on the equilibrium condition Ψ_I. According to definition 1 an increase of X_2 leads to an increased equilibrium emission price if $A_2^E > 0$ (i.e. for $w_2 > w_2^E$, we obtain a decreased equilibrium emission price if $A_2^E < 0$, i.e. for $w_2 < w_2^E$). The impact of an increased emission price on the equilibrium condition Ψ_I depends on the the degree of free allocation A_1. Whenever $A_1 < A_1^E$ (i.e. $\frac{\partial \Psi_I}{\partial e^*} < 0$, compare definition 1) an increased emission price leads to a decrease of firms' total investment activity X_1^*. In this case the reduction of scarcity rents (obtained when total capacity is binding) caused by the increased emission price dominates the increased value associated to the permits granted for free. The reverse holds true for a high level of free allocation, i.e. $A_1 > A_1^E$ where an increased emission price leads to increased total investment X_1^*. Whenever the impact of increased investment X_2^* yields a decreased emission price, which obtains for cleaner base load technologies (for $A_2^E < 0$, i.e. $w_2 < w_2^E$), we obtain the opposite results. In sum, $\frac{dX_1^*}{dA_2} > 0$ if and only if $\left(A_1 - A_1^E\right)A_2^E > 0$, as stated in the theorem.

Finally expression (8) also provides the intuition under which conditions we obtain $\frac{dX_1^*}{dA_2} \geq \frac{dX_2^*}{dA_2}$ (i.e. also investment in the peak load technology increases). To this end observe that $\frac{dX_1^*}{dA_2} = \frac{dX_2^*}{dA_2}$ if and only if in expression (8) the total impact of changed X_1^* is precisely of the same size as the total impact of changed X_2^*, but of opposite sign. As shown in the theorem this only obtains in case the increase of investment in the base technology leads to an increase of the emission price (for $A_2^E > 0$, i.e. $w_2 > w_2^E$) and if this increase has a sufficiently positive impact on the equilibrium condition Ψ_I, i.e. for allocation A_1 sufficiently big (for $A_1 > A_1^{cross} > A_1^E$). All those results are illustrated in the left graph of figure 2.

Likewise we can analyze the impact of changing A_1, as established in theorem 3(ii). Analogous to expressions (19) (20) and (21) we can determine the total derivative and solve for $\frac{de^*}{dA_1}$. After plugging in, we obtain for $\frac{d\Psi_I}{dA_1} + \frac{d\Psi_{II}}{dA_1}$ (observe $\frac{\partial \Psi_I}{\partial A_1} = -\frac{\partial \Psi_{II}}{\partial A_1} = e$):

$$\frac{d\Psi_I}{dA_1} + \frac{d\Psi_{II}}{dA_1} = \left(\Psi_{I1} + \Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}} + \Psi_{IIe}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right)\frac{dX_1^*}{dA_1} + \left(\Psi_{II2} + \Psi_{IIe}\frac{\Psi_{E2}}{-\Psi_{Ee}} + \Psi_{Ie}\frac{\Psi_{E2}}{-\Psi_{Ee}}\right)\frac{dX_2^*}{dA_1} = 0 \quad (9)$$

$$\Leftrightarrow \left(-\Psi_{Ee}\Psi_{I1} + \left(A_2 - A_1^E - A_2^E\right)A_1^E\right)\frac{dX_1^*}{dA_1} + \left(-\Psi_{Ee}\Psi_{II2} + \left(A_2 - A_1^E - A_2^E\right)A_2^E\right)\frac{dX_2^*}{dA_1} = 0$$

Analogous to above the coefficients on the expressions $\frac{dX_2^*}{dA_1}$ and $\frac{dX_1^*}{dA_1}$ determine the impact of changed investment X_1^* or X_2^* on both equilibrium conditions. The sum of both coefficients is strictly negative if second order conditions are not to be violated (compare lemma 2(iii)). This directly illustrates why $\frac{dX_2}{dA_1}$ cannot reach the level of $\frac{dX_1}{dA_1}$ (in other words, increased free allocation A_1 cannot leave investment in the peak load technology unchanged).

Furthermore, as we show, for small A_2 both coefficients are negative (thus $\frac{dX_1^*}{dA_1}$ and $\frac{dX_2^*}{dA_1}$ have opposite signs), since $\frac{dX_1^*}{dA_1} > \frac{dX_2^*}{dA_1}$ this implies $\frac{dX_2^*}{dA_1} < 0$. Observe that the coefficient of the expression $\frac{dX_1^*}{dA_1}$ is increasing in A_2, the coefficient of expression $\frac{dX_1^*}{dA_1}$ is increasing in A_2 if $A_2^E > 0$ (i.e. $w_2 > w_2^E$). That is, for A_2 high enough the coefficients become non–negative, leading to altered monotonicity behavior. As we show in the theorem we can establish a relative level of dirtiness w_2^S (with $w_2^S \geq w_2^E$ and $w_2^S = w_2^E$ in the case of linear demand), which separates the cases when either of the coefficients becomes zero for higher levels of A_2 (Remember the sum of both coefficients has to be negative in order to satisfy the second order conditions, see above). Whenever the coefficient of $\frac{dX_1^*}{dA_1}$ equals to zero, expression (9) directly implies $\frac{dX_2^*}{dA_1} = 0$ and vice versa, as stated in the theorem.

Finally, in theorem 3(iii) we provide the results of comparative statics with respect to the parameter T. For an intuition of those results observe first of all that an increase of the total emission cap T leads to a reduction of the equilibrium permit price. This in turn induces increased total investment X_1^* if (similar to the intuition for part (i)) the increase of scarcity rents (which obtains due to lower emission price) dominates the decreased value of the emission permits granted for free, i.e. $A_1 < A_1^E$. The opposite result obtains for $A_1 > A_1^E$. Similarly, the reduced emission price induces increased investment in the base load technology X_2^* if the total impact of reduced emission price on the base load investment condition is negative, i.e. if and only if $A_2 < A_1 + A_2^E$ (i.e. $\Psi_{IIe} < 0$). If we denote total emissions which obtain in the absence of any environmental policy by $\bar{T}$. Lowering the cap on total emissions T below $\bar{T}$ corresponds to the introduction of a cap and trade mechanism.

To provide the direct connection of our framework with current practice in competition policy, let us conclude this section by briefly discussing the impact of introducing a cap and trade mechanism as observed for example during the current phase of the EU ETS. In this phase, the free allocations granted for free to a unit of each class of technologies were such as to cover the total needs necessary on average to operate that unit.[24] In our framework that would correspond to levels of free allocation $A_1^{full} \in [1-(F(\theta_{\overline{P}}))w_1, 1-(F(\theta_{\underline{P}}))w_1]$ and $A_2^{full} \in [1-(F(\theta_{\overline{B}}))w_2, w_2]$. [25] With an allocation scheme (A_1^{full}, A_2^{full}) which aims at covering the average total needs of a unit of specific technology we obtain increased total investment and increased base load investment when introducing the trading system (i.e. the emission cap is lowered below $\bar{T}$ in our framework). To see this, first observe that $A_1^{full} > A_1^E$ which according to lemma 3 (iii) leads to an increase of X_1^*. Second observe that $A_2^{full} - A_1^{full} > A_2^E$ (since $A_2^{full} > (1-F(\theta_{\overline{B}}))w_2$ and $A_1^{full} < (1-F(\theta_{\underline{P}}))w_1$), which according to lemma 3 (iii) leads to increased investment in the base load technology.

In order to apply our findings of lemma 3(i) consider again our example of an electricity markets with lignite or coal fired plants as a representative base load technology and open cycle gas turbines as a representative peak load technology. Since open cycle gas turbines have lower emission factors, we obtain $w_2 > w_1$, which directly implies $w_2 > w_2^E$ (compare definition 1). Since furthermore $A_1^{full} > A_1^E$ as established above, we can directly conclude that an increase (decrease) of the free allocation A_2 not only would yield increased base load investment but also an increased (decreased) emission price and increased (decreased) total investment.

After having analyzed the market equilibrium which obtains in the presence of an emission trading system and derived its properties of comparative statics we now proceed to the main part of this article and analyze the optimal design of a cap and trade mechanism.

4 The Optimal Cap and Trade Mechanism

In this section we determine the optimal cap and trade mechanism. We first determine the first best solution as a benchmark, which obtains for the case of a perfectly competitive market when a regulator can freely choose all parameters (A_1, A_2, T) of the cap and trade mechanism (see theorem 1). We then analyze several market imperfections and solve for

[24]To give a specific example: In the German electricity market free allocation is determined by a technology specific emission factor which measures average emissions per unit of electricity produced (0.365 $tCO2/MWh$ for gaseous fuels and 0.750 $tCO2/MWh$ for solid and liquid fuels) multiplied by a preestablished technology specific average usage. For open cycle gas turbines in Germany the average usage is established at 0.11 (i.e. 1000 hours per year), for coal and combined cycle gas turbines it is given by 0.86 (i.e. 7500 hours per year) and for lignite plants it is given by 0.94 (i.e. 8250 hours per year), See appendices 3 and 4 of German Parliament (2007).

[25]To be precise, in our framework average usage of the base technology is $\frac{1}{X_2^*}\int_{\underline{\theta}}^{\theta_{\overline{B}}} Q^*(\theta)dF(\theta) + (1 - F(\theta_{\overline{B}}))$, the average usage of the peak technology is $\frac{1}{X_1^*-X_2^*}\int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} (Q^*(\theta) - X_2^*)\, dF(\theta) + (1 - F(\theta_{\overline{P}}))$. The corresponding emission factors are given by w_1 and w_2 respectively.

the corresponding second best solutions. We first determine the optimal cap and trade mechanism which should be chosen for an imperfectly competitive market (see section 4.1). We then analyze the case when competition authorities cannot freely choose all parameters (A_1, A_2, T) of the cap and trade mechanism but only a subset of them (see section 4.2). In order to answer all those questions we first determine total welfare generated in a market with some cap and trade mechanism (A_1, A_2, T):

$$\begin{aligned} W(A_1, A_2, T) &= \int_{\underline{\theta}}^{\theta_{\overline{B}}} \left[\int_0^{Q^*} (P(Y,\theta) - c_2)\, Y dY \right] dF(\theta) + \int_{\theta_{\overline{B}}}^{\theta_{\underline{P}}} \left[\int_0^{X_2^*} (P(X_2^*,\theta) - c_2)\, Y dY \right] dF(\theta) \\ & \int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} \left[\int_0^{Q^*} (P(Y,\theta) - c_1)\, Y dY \right] dF(\theta) + \int_{\theta_{\overline{P}}}^{\overline{\theta}} \left[\int_0^{X_1^*} (P(X_1^*,\theta) - c_1)\, Y dY \right] dF(\theta) \\ & - \int_{\theta_{\underline{P}}}^{\overline{\theta}} (c_1 - c_2) X_2^* dF(\theta) - k_2 X_2^* - k_1 (X_1^* - X_2^*) - D(T). \end{aligned} \tag{10}$$

Observe, that welfare does not directly depend on the parameters (A_1, A_2, T) chosen for the cap and trade mechanism but only indirectly through the implied investment and production decisions X_1^*, X_2^* and Q^*. In order to maintain presentability of the results, we relegated all computations to the appendix and directly characterize the optimal cap and trade mechanism in the subsequent lemma.

Lemma 4 *The optimal cap and trade mechanism solves the following conditions:*

$$\begin{aligned} (i) \qquad & W_{A_1} := \frac{dX_1^*}{dA_1}\Omega_I + \frac{dX_2^*}{dA_1}\Omega_{II} = 0 \\ (ii) \qquad & W_{A_2} := \frac{dX_1^*}{dA_2}\Omega_I + \frac{dX_2^*}{dA_2}\Omega_{II} = 0 \\ (iii) \qquad & W_T := \frac{dX_1^*}{dT}\Omega_I + \frac{dX_2^*}{dT}\Omega_{II} - D_T(T) + e^* + \frac{\Delta}{n} = 0. \end{aligned}$$

The expressions Ω_I and Ω_{II} determine the total impact of changed X_1^* and changed X_2^* respectively on total Welfare. They are defined as follows:

$$\Omega_I := \int_{\theta_{\overline{P}}}^{\overline{\theta}} \frac{-P_q X_1^*}{n} dF(\theta) - A_1 e^* - \frac{\Delta}{n} A_1^E \qquad \Omega_{II} := \int_{\theta_{\overline{B}}}^{\theta_{\underline{P}}} \frac{-P_q X_2^*}{n} dF(\theta) - (A_2 - A_1) e^* - \frac{\Delta}{n} A_2^E.$$

The term $\frac{\Delta}{n} := \frac{\int_{\underline{\theta}}^{\theta_{\overline{B}}} \frac{dQ^*}{de^*}\left(-P_q \frac{Q^*}{n}\right) dF(\theta) + \int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} \frac{dQ^*}{de^*}\left(-P_q \frac{Q^*}{n}\right) dF(\theta)}{\int_{\underline{\theta}}^{\theta_{\overline{B}}} \left(\frac{dQ^*}{de^*} w_2\right) dF(\theta) + \int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} \left(\frac{dQ^*}{de^*} w_1\right) dF(\theta)} > 0$ *determines the impact of changed emissions on welfare for those spot markets where investment is not binding.*

Proof See appendix D. □

We now provide some intuition for the conditions which characterize an optimal cap and trade mechanism. We first consider the optimal choice of the free allocation to the peak load technology given by A_1. Observe that the optimality conditions (i) and (ii) express the impact of changed free allocation on total welfare exclusively through the channel of changed investment in the base load technology X_2^* and changed total investment X_1^*. The total impact of changed investment on total welfare is denoted by Ω_I and Ω_{II}, this total impact can be broken down into three components corresponding to the three summands of Ω_I and Ω_{II} respectively.

First, observe that at all those spot markets where total investment is binding (i.e. for $\theta \in [\theta_{\overline{B}}, \theta_{\underline{P}}]$ and $\theta \in [\theta_{\overline{P}}, \overline{\theta}]$ respectively) imperfectly competitive investment behavior induces too low investment incentives, an increase of investment X_2^* or X_1^* leads to increased welfare given by the markup $-P_q \frac{X}{n}$. Second, free allocation $A_1 > 0$ or $(A_2 - A_1) > 0$ induces too high investment incentives, thus an increase of investment would lead to a reduction of welfare given by the monetary value of the free allocation (i.e. $A_1 e^*$ and $(A_2 - A_1)e^*$). Notice that in a world with exogenously fixed emission price e^* the optimal level of free allocation should be chosen such as to balance those two effects.[26] Since the emission price is endogenous in our analysis, an additional term obtains. An increase of investment dX_1^* or dX_2^* leads to increased emissions of $dX_1^* A_1^E$ and $dX_2^* A_2^E$ at those spot markets where investment is binding. Since total emissions are capped by T, however, this necessarily has to imply an equivalent reduction of emissions at those spot markets where investment is not binding (i.e. for $\theta \in [\underline{\theta}, \theta_{\overline{B}}]$ or $\theta \in [\theta_{\underline{P}}, \theta_{\overline{P}}]$). Since production decisions are also imperfectly competitive, a reduction of output leads to reduced welfare generated at those spot markets. This impact is quantified by the term $\frac{\Delta}{n}$ defined in the lemma. That is, taking into account the endogenous nature of the emission price leads to a lower degree of optimal free allocation A_1 than suggested by an analysis with exogenously fixed emission price.

The impact of a changed emission cap T on total welfare has a similar structure than the impact of changed free allocations. Analogous to above, a changed emission cap leads to changed investment incentives, the impact of changed investment incentives on welfare is given by the terms Ω_I and Ω_{II}, which have already been discussed above. As we will see later on in theorems 1 and 2, if the levels of free allocation are chosen optimally such as to obtain $\Omega_I = \Omega_{II} = 0$ those terms will not be relevant for the optimal choice of the emission cap. If the levels of free allocation are not chosen optimally, however, they have to be considered when determining the optimal level of the emission cap T (compare theorems 3, 4, 5 and 6).

[26] That is, the monetary subsidy $A_1 e^*$ for example should then equate to the integral of the markups over all relevant spot markets. The intuition for this result in some sense parallels the quite well known insight obtained for a simple static model where a monopolist can be induced to produce first best output if he obtains a subsidy corresponding to his markup.

Apart from having an impact on investment incentives, a changed emission cap T leads to changed welfare also through several other channels. First, most apparently an increased emission cap leads to increased emissions which reduce welfare by the marginal social cost of pollution D_T. Second, observe that on the other hand an increased emission cap leads to a welfare increase since it implies a reduced emission price which allows for increased output. The welfare increase at each spot market is given by the changed output multiplied by the difference between marginal cost as perceived by the firms and true marginal cost, i.e. $dQ(w_i e^*)$, for $i = 1, 2$. Put differently however, this corresponds to the changed pollution at each spot market multiplied by the emission price e^*, the change in welfare at all spot markets then is simply given by the total change of emissions multiplied by the emission price i.e. dTe^*. As we will see in the subsequent theorem 1, for a perfectly competitive market the optimal cap and trade mechanism only balances those two effects and equates the marginal social cost of pollution to the emission price (i.e. $e^* = D_T$).[27] Third, observe that an increased emission cap T leads to a reduced emission price. This allows to reduce the welfare loss obtained due to imperfect competition at those spot markets where investment is not binding and output too low. Notice that the impact of changed emissions on welfare at those spot markets where investment is not binding has already been discussed above, it is given by $\frac{\Delta}{n}$.

Based on the findings of lemma 4 as the first best benchmark we can now directly establish the optimal cap and trade mechanism which obtains for a perfectly competitive market

Theorem 1 (Optimal Market Design, First Best Benchmark) *Under perfect competition the optimal market design satisfies*

$$(i)\ A_1^* = 0 \qquad (ii)\ A_2^* = 0 \qquad (iii)\ T^*: \quad e^* = D_T(T).$$

Proof See appendix E. □

The theorem demonstrates that in a competitive market (i.e. $n \to \infty$), full auctioning is unambiguously optimal (i.e. no free allocations should be granted). A brief glance to lemma 4 and the intuition provided reveals that investment incentives of firms under perfect competition are optimal, positive free allocation would lead to reduced welfare. Moreover, as condition (iii) shows, the emission target T should be set such that the equilibrium permit price equals marginal social cost of environmental damage. That is, as already discussed above, the optimal cap and trade mechanism balances welfare losses due to foregone

[27] This parallels the fundamental tradeoff obtained in a simple static model where a Pigou tax should just equal to the marginal social damage of pollution.

production at all spot markets given by e^* with the marginal social cost of pollution given by D_T.

in the subsequent two sections we now consider market imperfections which make an attainment of the first best outcome impossible. First, we determine the design of an optimal cap and trade mechanism for an imperfectly competitive market (see section 4.1). Apart from imperfect competition, another source of market imperfection arises when the competition authorities cannot freely choose all parameters (A_1, A_2, T) of the cap and trade mechanism, but only a subset. Such situations arise for example when the level of free allocation for (some of) the different technologies or the total emission cap is exogenously fixed due to political arrangements or lobbing of firms and the competition authority can only determine the remaining parameters (see section 4.2).

4.1 Optimal Market Design under Imperfect Competition

After having determined the first best benchmark (theorem 1) we now determine the optimal cap and trade mechanism for an imperfectly competitive market.

Theorem 2 (Optimal Market Design under Imperfect Competition) *Under imperfect competition the optimal market design satisfies*

$$(i) \qquad A_1^* = \frac{1}{e^*}\left(\int_{\theta_{\overline{P}}}^{\overline{\theta}} \left(\frac{-P_q X_1^*}{n}\right) dF(\theta) - \frac{\Delta}{n} A_1^E\right)$$

$$(ii) \qquad A_2^* = \frac{1}{e^*}\left(\int_{\theta_{\overline{B}}}^{\theta_{\underline{P}}} \left(\frac{-P_q X_2^*}{n}\right) dF(\theta) + \int_{\theta_{\overline{P}}}^{\overline{\theta}} \left(\frac{-P_q X_1^*}{n}\right) dF(\theta) - \frac{\Delta}{n}\left(A_1^E + A_2^E\right)\right)$$

$$(iii) \qquad T^*: \quad e^* = D_T(T) - \frac{\Delta}{n}.$$

Now assume that $P_{q\theta} = 0$. We then obtain $A_1^* > 0$. For $w_2 \leq w_2^E$ we obtain $A_2^* > A_1^*$, for $w_2 > w_2^E$ we can obtain $A_2^* = 0$.

Proof See appendix E. □

The optimal levels of free allocations (A_1^*, A_2^*) under imperfect competition are thus typically different from zero, a striking difference to the result obtained under perfect competition (see theorem 1). The fundamental reason why this is the case follows directly from the insights provided by lemma 1 and the subsequent discussion of the results: Imperfectly competitive firms not only exercise market power at the spot markets, but also choose their capacity such that they optimally benefit from scarcity prices, implying reduced production and investment incentives.

As already discussed in the text following lemma 1 (compare the last paragraph which discusses lemma 1), for an exogenously fixed price for pollution (e.g. a pigouvian tax at some fixed level e^*) optimal investment incentives are obtained by subsidizing investment such as to precisely compensate for the difference between scarcity rents and marginal scarcity profits. To stick as close as possible to our notation such subsidy could be made by assigning the amounts A_1 and A_2 of free tax vouchers to each unit invested in either of the technologies. The optimal level of tax vouchers is then given by expressions (i) and (ii) of theorem 2 (notice that for exogenously fixed permit price we have $\Delta = 0$).

Remember that in our framework the expression $\frac{\Delta}{n}$ allowed to quantify the impact of changed emissions at those spot markets where investment is not binding. Positive free allocation leads to increased investment incentives, which (through an increased emission price) can lead to reduced output (and thus pollution) at those spot market where investment is not binding. The terms including the expression $\frac{\Delta}{n}$ take this welfare loss into account. This leads to a reduced level of the optimal degree of free allocation. As we show in the theorem, under imperfect competition the degree of free allocation for the peak load technology is always positive. For the optimal allocation for the base load technology ambiguous results obtain. If the base load technology is less emission intensive than the peak load technology (i.e. $w_2 \leq w_1$) increased investment in the base load technology leads to reduced emissions and thus allows for more output at spot markets where investment is not binding. As we show this always implies $A_2^* > A_1^*$. On the other hand, if the base load technology is more emission intensive than the peak load technology ($w_2 > w_1$, i.e. an increase of base load investment leads to increased emission price), then it might be optimal to set $A_2^* < A_1^*$ or even $A_2^* < 0$ as we show.

Finally consider the optimal choice of the total emission cap T for the case of imperfect competition. A brief look at the optimality condition (iii) established in lemma 4 reveals, that the impact of a changed emission cap on investment decisions can be neglected since the levels of free allocation are determined optimally (such as to obtain $\Omega_I = \Omega_{II} = 0$). What matters, however, is the fact that an increased emission cap leads to a reduced emission price which in turn allows to reduce the welfare loss induced by imperfectly competitive production decisions at those spot markets where investment is not binding (given by $\frac{\Delta}{n}$). As a result the optimal cap on total emissions is chosen such as to yield an emission price below the marginal social cost of pollution.

In sum, the main intuition why an optimal cap and trade mechanism (with endogenous emission price) implies levels of free allocation which are different from zero is similar to the intuition obtained for the case of an exogenously fixed price for pollution (e.g. a pigouvian tax): Under imperfect competition firms' investment and production incentives are too low, leading to decreased welfare. Free allocations can provide adequate incentives which lead to an increase of welfare. However, for the case of an endogenous emission price, as modeled in the present paper, increased investment incentives also lead to an increased emission price which in turn aggravates welfare losses at those spot market where investment is not binding.

Let us finally discuss those results in the light of recently proposed measures thought to increase firms' investment incentives, as for example observed in liberalized electricity markets. In the perception of many economists and policy makers investment incentives in those markets are too low, one of the reasons potentially being market power as modeled in the present paper. To resolve those problems of too low investment incentives, several measures have been proposed, among them capacity mechanisms.[28] For the present discussion we clearly have to abstract form the specific problems encountered when designing real capacity markets, and just consider some subsidy s_t paid to the firms per unit of investment made in technology $t = 1, 2$. Notice that in the present framework it is equivalent if a monetary payment s_t or free allocations with value $A_t e^*$ for $t = 1, 2$ are granted to a firm per unit of investment in a technology t. What exclusively matters for firms' investment incentives is the total value $s_t + A_t e^*$ granted to firms per unit of investment, this total value should be set at an optimal level.[29]

This in turn implies, however, that, once a cap and trade mechanism is put in place in specific given market, the implications of this cap and trade have to be taken into account when designing the capacity market. More specifically the design of a capacity market which disregards the endogenous nature of emission prices, will lead to too high investment incentives. However, we establish furthermore that also when taking into account the endogenous nature of the emission price, the subsidy granted to the peak load technology should be positive, if the base load technology is less emission intensive it should receive a subsidy which exceeds that of the peak load technology. However, a relatively dirty base load technology should not receive any free allocations.

4.2 Optimal Design of a Partially Constrained Cap and Trade Mechanism

In theorems 1 and 2 we determined the optimal design of a cap and trade mechanism when all its parameters (A_1, A_2, T) can be freely chosen by the competition authority. We first

[28]In most restructured electricity markets in the United States so called "capacity markets" are installed in order to increase inefficiently low investment incentives, also in Europe policy makers consider their introduction (see e.g. Cramton and Stoft (2008)).

[29]Consequently, optimality just requires that the sum of both parameters satisfies the above optimality conditions. An immediate and interesting implication is that possible inefficiencies due to grandfathering could be healed by capacity payments that compensate for the distorting effect *without any efficiency losses* (as long as the subsidies resulting from free allocations are not higher than the sum of both parameters should be).

analyzed the case of a perfectly competitive market, which yields the first best benchmark (theorem 1) and then the case of imperfect competition (theorem 2). Another source of market imperfection, apart from imperfect competition, arises when the competition authorities cannot freely choose all parameters (A_1, A_2, T) of the cap and trade mechanism. Such rigidities might be due to political constraints and arrangements or due to lobbing of firms. As already discussed extensively in the introduction of this article free allocations have been key to guarantee the political support necessary to introduce cap and trade systems, compare Convery(2009), Tietenberg(2006), Bovenberg (2008), or for example Grubb and Neuhoff (2006)[30]. It is the purpose of the present section to analyze how a competition authority should optimally design a cap and trade mechanism if it can determine only a subset of the parameters of the cap and trade mechanism, whereas the remaining parameters are exogenously fixed due to the above discussed problems.

Theorem 3 determines the optimal degree of free allocations for the case of exogenously fixed level of the total emission cap T. In theorems 4, 5 and 6 we determine the optimal degree of free allocation to the remaining technologies and the corresponding level of the optimal total emission cap T. Observe that our results obtained in lemma 4 in principle would allow for a detailed analysis of those questions both for the cases of perfect and imperfect competition. In order to limit the notational burden in the present paper we restrict ourselves to the case of perfect competition, however. In this case the optimality conditions determined in lemma 4 read as follows

$$W_{A_1} := \frac{dX_1^*}{dA_1}(-A_1)\,e^* + \frac{dX_2^*}{dA_1}(A_1 - A_2)\,e^* = 0 \tag{11}$$

$$W_{A_2} := \frac{dX_1^*}{dA_2}(-A_1)\,e^* + \frac{dX_2^*}{dA_2}(A_1 - A_2)\,e^* = 0 \tag{12}$$

$$W_T := \frac{dX_1^*}{dT}(-A_1)\,e^* + \frac{dX_2^*}{dT}(A_1 - A_2)\,e^* - D_T(T) + e^* = 0. \tag{13}$$

We first analyze the case of an exogenously fixed level of the cap on total emissions T, an example might be a situation where politicians are willing to introduce a cap and trade mechanism but are reluctant to induce too severe (even though optimal from an overall welfare point of view) distortions on the economy. The above optimality conditions directly reveal that in a perfectly competitive market no free allocations should be granted to firms, independently of the level of the emission cap.[31] This is summarized in theorem 3.

[30] "Due in part to the sheer scale of the EU ETS, governments are subject to intense lobbying relating to the distributional impact of the scheme, and are constrained by this and by concerns about the impact of the system on industrial competitiveness. Few academics understand the real difficulties that policy-makers face when confronted with economically important industries claiming that government policy risks putting them at a disadvantage relative to competitors."

[31] The results of theorem 3 for the case of imperfect competition obtain analogously, the optimal levels of free allocation are given by conditions (i) and (ii) established in theorem 2.

Theorem 3 (Optimal Design for fixed emission cap T) *For any exogenously fixed total emission cap T it is optimal to choose the levels of free allocation $A_1^* = A_2^* = 0$.*

That is, the result obtained in the first best benchmark (theorem 1), where no free allocation has been found to be optimal also obtains if the total emission cap is not set at an optimal level. Observe that the reverse does not hold as we show in the subsequent theorem, however.

Theorem 4 (Optimal Design for fixed allocations A_1 and A_2) *Suppose the initial allocations A_1 and A_2 are fixed exogenously. Define*

$$\Gamma_0(A_1, A_2) := (A_1 - A_1^E)A_1\Psi_{I1} + (A_2 - A_1 - A_2^E)(A_2 - A_1)\Psi_{II2}. \tag{14}$$

The optimal emission cap T^* has to be set such as to satisfy $e^* = D_T(T^*)$ for $\Gamma_0(A_1, A_2) = 0$, $e^* > D_T(T^*)$ for $\Gamma_0(A_1, A_2) > 0$, and $e^* < D_T(T^*)$ for $\Gamma_0(A_1, A_2) < 0$.

Proof See appendix F. □

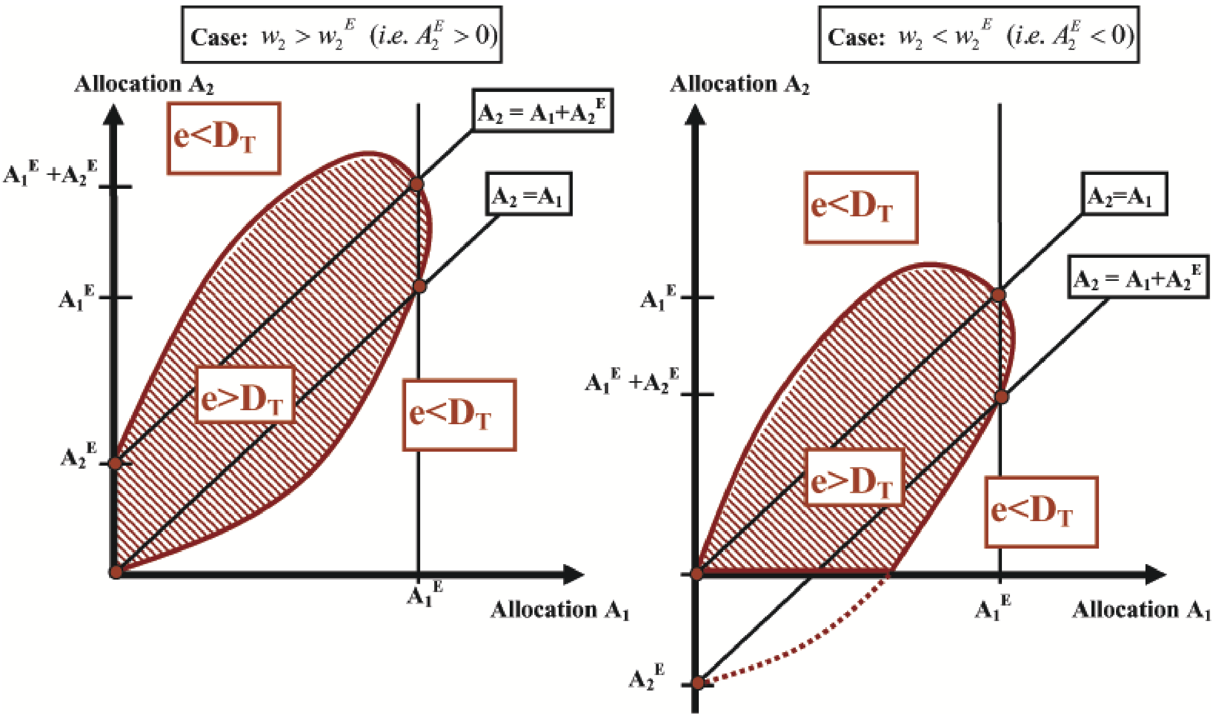

Figure 3: Choosing the optimal T^* for exogenously fixed initial allocations A_1 and A_2. Left: for relatively dirty base technology, i.e. $w_2 > w_2^E$, Right: for relatively clean base technology, i.e. $w_2 < w_2^E$.

That is, for levels of free allocation A_1, A_2 which are not set optimally the optimal cap on emissions T typically does not implement an emission price e^* equal to the social cost of pollution D_T. To get an intuition for the result, note first that the cap T on total emissions governs the price for emission certificates e^*, which in turn influences both investment decisions and unconstrained production decisions at those spot markets where investment is not binding. Optimal production decisions are induced by an emission price equal to the social cost of pollution. This is only overall optimal in case of optimal investment incentives.

Now first observe that in case of positive free allocations (as considered in the theorem) investment incentives are distorted, however. That is, for $A_1 > 0$ investment incentives in the peak load technology are too high, for $A_2 > A_1$ ($A_2 < A_1$) investment incentives in the base load technology are too high (low). A distortion of the emission price can then be suited to at least partially adjust investment incentives.

Second observe that the impact of a changed emission cap T on investment incentives already has been derived in lemma 3 (iii) and discussed in the subsequent text. As established there a higher emission cap T (implying a lower emission price e^*) leads to increased investment in the peak load technology X_1^* if and only if $A_1 < A_1^E$, it leads to increased investment in the base load technology X_2^* if and only if $A_2 < A_1 + A_2^E$.

Intuitively theorem 4 formally joins those two effects, that is, whenever the levels of free allocation A_1, A_2 are such as to induce over investment, the total cap on emissions should be set such as to induce an emission price which leads to a reduction of investment incentives and vice versa. All those findings are illustrated graphically in figure 3.

Consider the case $A_2 = A_1 > 0$, where all technologies get the same amount of free allocations (the 45-degree line of figure 3). In the light of the above discussion this implies first of all that investment incentives in the base load technology are undistorted (since $A_2 = A_1$) and investment incentives in the peak load technology are too high. For $A_1 < A_1^E$ investment incentives are reduced for a higher emission price, for $A_1 > A_1^E$ they are reduced for a lower emission price. Next consider the case $A_2 = A_1 + A_2^E$. In this case a changed emission price e^* has no impact on investment in the base load technology, analogous to above the optimal cap T is thus designed exclusively such as to reduce the too high investment incentives in the peak load technology (i.e. for $A_1 < A_1^E$ we have $e^* > D_T$ and vice versa).[32]

We conclude the discussion of theorem 4 by applying our findings to the current policy of full allocation (A_1^{full}, A_2^{full}) as observed during the current phase of the EU ETS and already introduced at the end of section 3. Remember that we derived the following properties for the levels of full allocation: $A_1^{full} > A_1^E$ and $A_2^{full} > A_1^{full} + A_2^E$. As already discussed,

[32]Observe that an analogous reasoning obtains for the case $A_1 = 0$, when only investment incentives in the base load technology are distorted and the case $A_1 = A_1^E$, when a changed emission price has no impact on investment in the peak load technology and only distortions of base load investment are to be adjusted by the total emission cap.

if we consider either lignite or coal fired plants as the representative base load technology and open cycle gas turbines as the representative peak load technology we also obtain $A_2^{full} > A_1^{full}$. For our framework we thus obtain that the optimal cap on total emissions has to be set such that the equilibrium permit price is lower than the social cost of pollution, i.e. $e^* < D_T$.

In the subsequent theorem 5 we consider the case that only allocation for the peak load technology A_1 is exogenously fixed, allocation for the base load technology A_2 and the total emission cap T can be determined optimally, however.

Theorem 5 (Optimal Design for fixed allocation A_1) *Suppose the allocation for the peak technology A_1 is exogenously fixed. The optimal allocation for the base technology then solves $A_2^* = \frac{dX_2^*/dA_2 - dX_1^*/dA_2}{dX_2^*/dA_2} A_1$. More specifically we obtain (see left graph of figure 4)*

$$\begin{cases} A_2^* = 0 & if \quad (A_1^{cross} \leq A_1 < A_1^{lim}) \\ 0 < A_2^* < A_1 & if \quad \big((0 < A_1 < A_1^E)\ \&\ (w_2 < w_2^E)\big)\ OR\ \big((A_1^E < A_1 < A_1^{cross})\ \&\ (w_2 > w_2^E)\big) \\ A_1 < A_2^* & if \quad \big((A_1^E < A_1 < A_1^{cross})\ \&\ (w_2 < w_2^E)\big)\ OR\ \big((0 < A_1 < A_1^E)\ \&\ (w_2 > w_2^E)\big). \end{cases}$$

The optimal cap T^* is such that $e^* > D_T$ if $A_1 < A_1^E$ and $e^* < D_T$ if $A_1 > A_1^E$.

Proof See appendix G. □

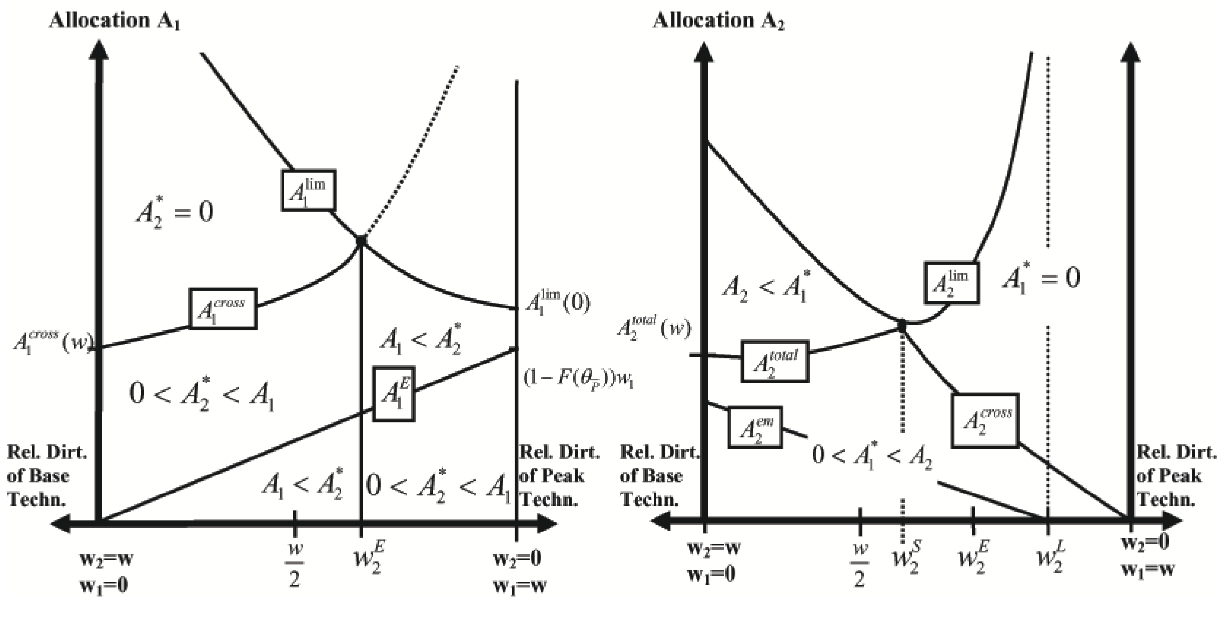

Figure 4: Left: Choosing the optimal A_2^* for exogenously fixed initial allocation A_1. Right: Choosing the optimal A_1^* for exogenously fixed initial allocation A_2.

Observe that the optimality condition for A_2 as stated in the theorem obtains directly by rearranging expression (12). To derive the properties of the optimal degree of allocation for the base load technology A_2^* as stated in the theorem we can now make use of the properties of comparative statics derived in lemma 3(i). Most importantly, as established there, we always obtain $\frac{dX_2^*}{dA_2} > 0$. We thus obtain $A_2^* > A_1$ if and only if $\frac{dX_1^*}{dA_2} < 0$. Since we only consider non–negative levels of free allocation we furthermore obtain $A_2^* = 0$ whenever $0 < \frac{dX_2^*}{dA_2} < \frac{dX_1^*}{dA_2}$. All those results of comparative statics have been derived in lemma 3 and have been discussed subsequently in section 3. Figures 2 and 4 do thus in principle look identical, observe however, that figure 2 exclusively states results of comparative statics, whereas figure 4 illustrates the properties of the optimal allocation A_2^* by making use of the previously obtained findings.

The Intuition for the optimal level of free allocation T^* in principle goes along the same lines as the one provided for the findings of theorem F. As compared to the first best benchmark, for positive allocation A_1 investment incentives in the peak load technology are too high. Whenever $A_1 < A_1^E$ we obtain $\Psi_{Ie} < 0$ which implies that a higher emission price e^* allows to reduce investment incentives in the peak load technology. Observe furthermore that investment incentives in the base load technology as induced by A_2^* are either too high or too low (i.e. $A_2^* > A_1$ for $w_2 > w_2^E$ and vice versa). As we show in the theorem it is optimal to set the total cap such as to obtain an emission price $e^* > D_T$ which induce reduced investment incentives in the peak load technology whenever $A_1 < A_1^C$. Observe that for $A_1 = A_1^C$ we obtain $A_2^* = A_1 = A_1^C$, which implies undistorted investment incentives in the base load technology, in this case we thus obtain $e^* = D_T$. The reverse hold true for the case $A_1 > A_1^C$ where the optimal emission cap T^* has to be set to obtain $e^* < D_T$ which induces reduced investment incentives in the peak load technology.

In the subsequent theorem 6 we consider the case that allocation for the peak load technology A_2 is exogenously fixed, allocation for the base load technology A_2 and the total emission cap T are determined optimally.

THEOREM 6 (OPTIMAL DESIGN FOR FIXED ALLOCATION A_2) *Suppose the allocation for the base technology A_2 is exogenously fixed. The optimal allocation for the peak technology then solves $A_1^* = \frac{dX_2^*/dA_1}{dX_2^*/dA_1 - dX_1^*/dA_1} A_2$. More specifically (see right graph of figure 4)*

$$\begin{cases} A_1^* = 0 & \textit{if} \quad (A_2^{cross} \leq A_2 < A_2^{lim}) \\ 0 < A_1^* < A_2 & \textit{if} \quad \big((0 < A_2 < A_2^{cross}) \;\&\; (w_2 < w_2^S)\big) \; OR \; \big((0 < A_2 < A_2^{total}) \;\&\; (w_2 > w_2^S)\big) \\ A_2 < A_1^* & \textit{if} \quad (A_2^{total} < A_2 < A_2^{lim}). \end{cases}$$

Define $A_2^{em} := A_1^E + A_2^E$. We obtain $A_2^{em} < A_2^{total}$ and $A_2^{em} < A_2^{cross}$. The optimal cap T^* is such that $e^* < D_T$ if $A_2 < A_2^{em}$ and $e^* > D_T$ if $A_2 > A_2^{em}$.

PROOF See appendix H. □

Observe that the optimality condition for A_1 as stated in the theorem obtains directly by rearranging expression (11). To derive the properties of the optimal degree of allocation for the peak load technology A_1^* as stated in the theorem we can now make use of the properties of comparative statics derived in lemma 3(ii). Most importantly, as established there, we always obtain $\frac{dX_1^*}{dA_1} > \frac{dX_2^*}{dA_1}$. Since we only consider non–negative levels of free allocation we obtain $A_1^* > 0$ whenever $\frac{dX_2^*}{dA_1} < 0$. Furthermore we obtain $A_1^* > A_2$ whenever $\frac{dX_2^*}{dA_1} < \frac{dX_1^*}{dA_1} < 0$.

Let us finally provide some intuition for the optimal cap on total emissions T. First of all observe that for $A_2 < A_2^{total}$ and $A_2 < A_2^{cross}$ we always obtain $0 < A_1^* < A_2$ which implies that investment incentives both in the base load and the peak load technology are too high as compared to the first best benchmark. For low levels of allocation to the base load technology (i.e. $A_2 < A_1^*(A_2) + A_2^E$) we obtain $\Psi_{IIe} < 0$ which makes it optimal to induce an emission price $e^* > D_T$ to lower investment incentives for both technologies.[33] Observe that for $A_2 = A_1^*(A_2) + A_2^E$ we obtain $\Psi_{IIe} = 0$, a distortion of the emission price above (or below) social cost of pollution has no impact on investment incentives in the base load technology. However, investment incentives in the peak load technology are too high (since $A_1^* > 0$). Since $\Psi_{Ie} < 0$ the distortion of the emission price above social cost of pollution is thus still suited to reduce investment incentives in the peak load technology. In total, the theorem thus balances increased investment incentives in the base load technology with reduced investment incentives in the peak load technology. The cut–off is reached where $\Psi_{IIe} + \Psi_{Ie} = 0$ which implies $A_2 = A_1^E + A_2^E = A_2^{em}$. That is for $A_2 < A_2^{em}$ it is optimal to set an emission cap T^* which induces $e^* > D_T$ and for $A_2 > A_2^{em}$ the optimal cap T^* induces $e^* < D_T$.[34] All those results are illustrated in figure 4.

We conclude the discussion of theorems 5 and 6 by applying our findings to the current policy of full allocation (A_1^{full}, A_2^{full}) as observed during the current phase of the EU ETS which served as the main illustrating example throughout this article:

First consider the case of theorem 5, where the allocation A_1 for the peak technology is exogenously fixed. As already shown, under full allocation we obtain $A_1^{full} \in [A_1^E, (1 - F(\theta_{\underline{P}}))w_1]$. For the case of a completely clean base load technology (that is $w_2 = 0$, in the context of electricity markets this would be the case for nuclear power plants for example) under the current rules such technology would not obtain any permits. As our result directly show, however, such technology should be granted more free allocations than the peak technology, i.e. $A_2^* > A_1^{full}$. Moreover the total emission cap T^* should be chosen such

[33] Observe that for $A_2^E < 0$ this range is degenerated at zero.

[34] For $A_2 > A_2^{cross}$ only investment incentives in the base load technology are distorted, since $A_2^{cross} > A_2^{em}$ the optimal cap then clearly has to implement $e^* > D_T$. For $A_2 > A_2^{total}$ the optimal $A_1^*(A_2) > A_2$ is so large that the optimal cap also has to implement $e^* > D_T$, as we show.

as to implement $e^* < D_T$ in order to dampen excessive investment incentives induced by those levels of free allocation.

Next consider the case of theorem 6, where the allocation A_2 for the base technology is exogenously fixed. The optimal level of free allocation for the peak technology has to be strictly positive if the peak technology is less emission intense than the base technology. In particular, if the peak technology is completely clean (that is $w_1 = 0$, in the context of electricity markets this would be the case for small bio-gas fired engines or turbines for example), under full free allocation this technology would not receive any free permits. As our result directly show, however, such technology should be granted a positive amount of free permits. Unlike in the case discussed in the preceding paragraph, however, the level of free allocation for the peak load technology should remain below the exogenously fixed level of free allocation for the base load technology. The reason for this difference lies in the fact that free allocations A_2 for the base load technology only have an impact on the resulting technology mix, free allocations A_1 to the peak load technology have an impact on the resulting technology mix and on firms' total investment activity. For exogenously fixed A_2 the optimal level of A_1^* is thus more moderate since it also leads to distorted total investment decisions. Finally notice for the optimal cap on total emissions: Since $A_2^{em} < (1 - F(\theta_{\overline{B}}))w_2 < A_2^{full}$ (compare definition 1 and the last paragraph of section 3) we can directly conclude that the total cap on emission has to be chosen such as to implement $e^* < D_T$ in order to dampen excessive investment incentives induced by those levels of free allocation.

In sum, if one of the technologies is granted an exogenously fixed level of free allocation (e.g. due to lobbying) then the optimal pattern of allocations to the remaining technology is completely different from the one that obtains under full allocation (for example as given by current legislation in the EU ETS, compare German Parliament (2007)). Furthermore for high levels of free allocation, the cap on total emissions should be chosen such as to induce an emission permit price which is below marginal social cost of pollution in order to reduce the distortions on the resulting technology mix.

5 Conclusion

Tradeable pollution permits are an increasingly important policy tool in environmental legislation worldwide. The possibility of freely allocate permits provides an important possibility to share the regulatory burden. This seems to significantly enhance the political support for recently introduced legislations (see for example Tietenberg(2006), Bovenberg(2008), or Convery(2009)). Since free allocations typically are subject to implicit or explicit updating the allocation of permits has an impact on firms' decisions: First, updating of free allocations on the one hand has an impact on firms' operation of existing production facilities if they believe that current output or emissions do have an impact on allowances granted in future periods. Second, updating also has an impact on firms' incentives to modify their production facilities if free allocations are granted based on all current installations of a firm.[35] The first phenomenon, where updating has an impact on current output and emissions, has already been intensively analyzed in the literature (compare for example Böhringer and Lange (2005), Rosendahl (2007), Mackenzie et al. (2008), or Harstadt and Eskeland (2010)). All those contributions abstract however from an explicit analysis of the impact of updating on firms' investment incentives which determine their technology mix in the long run. We thus want to contribute to this literature by explicitly analyzing the impact of a cap and trade scheme on firms' investment incentives. Since to a large extent the long run implications are responsible for the final success of a given environmental legislation we have the impression that this provides an important contribution to the ongoing debate on the optimal design of emission trading systems.

In the present article we have thus analyzed an analytical framework with tradeable permits and a cap on total emissions. Potentially strategically acting firms have been able to invest into production facilities (with different emission intensities), which allow for production for a longer horizon of time. After establishing the market equilibrium and the resulting technology mix, we have analyzed the optimal design of the cap and trade mechanism.

As a benchmark we established the first best solution which obtains for an ideal market. We then have derived the optimal design of the cap and trade system for a series of market imperfections. First we have analyzed the case of strategic investment and production decisions in an imperfectly competitive market. This allowed to highlight the close interdependency of mechanism to overcome low investment incentives (such as for example capacity markets in electricity markets) and cap and trade mechanisms: If the endogenous nature of emission prices in the presence of a cap and trade system is disregarded too high investment incentives are induced.

We then have analyzed the case that the competition authority cannot freely choose all parameters of the cap and trade system due to restrictions imposed by the political processes (as intensively discussed in the literature). The optimal choice of the remaining parameters differs substantially from that observed for the first best benchmark. Our result showed for example that if a certain technology receives free allocations it is typically optimal to grant free allocations also to the remaining technology. However, those free allocations should be the higher the less emission intensive the technology is. Interestingly those findings are

[35]Compare Neuhoff et al. (2006), IEA (2010), or the current legislation of the EU ETS, for example German Parliament (2007).

entirely opposed to the pattern of free allocations granted in a system of full free allocation as currently observed in EU ETS for example.

References

Anderson, B. and C. Di Maria (2011). Abatement and Allocation in the Pilot Phase of the EU ETS. *Environmental Resource Economics* , 48, 83-103

Böhringer, C. and A. Lange (2005). On the Design of Optimal Grandfathering Schemes for Emission Allowances. *European Economic Review* 49, 2041-2055.

Bovenberg, A.L., L.H. Goulder, and D.R. Gurney (2005). Efficiency Costs of Meeting Industry Distributional Constraints under Environmental Permits and Taxes. *Rand Journal of Economics* 36, 951-971.

Bovenberg, A.L., L.H. Goulder, and M.R. Jacobsen (2008). Costs of Alternative Environmental Policy Instruments in the Presence of Industry Compensation Requirements. *Journal of Public Economics*, 92, 1236-1253.

Coase, R. H. (1969). The Problem of Social Cost. *Journal of Law and Economics*, 3, 1-44.

Convery, F. (2009). Origins and Development of the EU ETS. *Environmental Resource Economics* , 43, 391-412.

Cramton, P. and S. Stoft (2008). Forward Reliability Markets: Less Risk, Less Market Power, More Efficiency. *Utilities Policy*, 16, 194–201.

Crew, M., C. Fernando, and P. Kleindorfer (1995). The Theory of Peak-Load Pricing: A Survey. *Journal of Regulatory Economics*, 8, 215–248.

Dales, J. (1968). Pollution, Property and Prices. University Press, Toronto.

Ehrenmann, A. and Y. Smeers (2011). Energy Only, Capacity Market and Security of Supply. A Stochastic Equilibrium Analysis. *Operations Research, forthcoming.*

European Council (2003). Directive 2003/87/EC , establishing a scheme for greenhouse gas emission allowance trading within the Community and amending Council Directive 96/61/EC, 13 October.

von der Fehr, N.-H. and D. Harbord (1993). Spot Market Competition in the UK Electricity Industry. *Economic Journal* 103 , 531 – 46.

German Parliament (2007). Gesetz zur Änderung der Rechtsgrundlagen zum Emissionshandel im Hinblick auf die Zuteilungsperiode 2008 bis 2012, Bundesgesetzblatt Jahrgang 2007 Teil I Nr. 38.

Grubb, M. and K. Neuhoff (2006). Allocation and competitiveness in the EU emission trading scheme: policy overview. *Climate Policy* 6, 7-30.

Hahn, R.W. and R.N. Stavins (2010). The Effect of Allowance Allocations on Cap-and-Trade System Performance. NBER Working Paper 15854.

Harstad, B. and G. Eskeland (2010). Trading for the Future: Signaling in Permit Markes. *Journal of Public Economics*, 94, 749–760.

IEA (2010). Reviewing Existing and Proposed Emission Trading Systems. International Energy Agency. Paris, November 2010.

Klemperer, P. and M. Meyer (1989). Supply Function Equilibria in Oligopoly under Uncertainty. *Econometrica*, 57, 1243–1277.

Mackenzie, I.A., N. Hanley, and T. Kornienko (2008). The Optimal Initial Allocation of Pollution Permits: a Relative Performance Approach. *Environmental Resource Economics* , 39, 265–282.

Matthes, F. (2006). Analyse der Wirtschaftlichkeit von Kraftwerksinvestitionen im Rahmen des EU-Emissionshandels. Institute for Applied Ecology Working Paper.

Moledina, A.A., J.S. Coggins, S. Polasky, and C. Costello (2003). Dynamic environmental policy with strategic firms: prices versus quantities. *Journal of Environmental Economics and Management* 45, 356376.

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

Neuhoff, K., K. Keats Martinez, and M. Sato (2006). Allocation, Incentives, and Distortions: The Impact of EU ETS Emissions Allowance Allocations to the Energy Sector. *Climate Policy* 6, 73–91.

Pahle, M., L. Fan, and W.–P. Schill (2011). How Emission Certificate Allocations Distort Fossil Investments, the German Example. *DIW-Discussion Paper 1097*

Rosendahl, K.E. (2008). Incentives and Prices in an Emission Trading Scheme with Updating. *Journal of Environmental Economics and Management*, 56, 69-82.

Tietenberg, T. (2006). Emissions Trading: Principles and Practice, second edition. RFF Press.

Zöttl, G. (2010). A Framework of Peak Load Pricing with Strategic Firms. *Operations Research*, 58, 1637–1649.

Zöttl, G. (2011). On Optimal Scarcity Prices. *International Journal of Industrial Organization*, forthcoming.

A Proof lemma 1

Note that given our assumptions on demand and cost, existence of spot market equilibrium at each demand scenario θ is ensured for the case of perfect and imperfect competition. We denote by $q_i^*(\theta, x)$ spot market output of firm i in scenario θ, given investments $x = (x_{11}, \ldots, x_{1n}, x_{21}, \ldots, x_{2n})$. Remember, X_1 and X_2 denote industry investment in either technology and $Q^*(\theta)$ industry output at each spot market θ, it is given as follows:

$$Q^* = \begin{cases} Q : P(Q,\theta) + P_q(Q,\theta)\frac{Q}{n} - c_2 - w_2 e^* = 0 & if\ \theta\ \in [\underline{\theta}, \theta_{\overline{B}}] \\ X_2 & if\ \theta\ \in [\theta_{\overline{B}}, \theta_{\underline{P}}] \\ Q : P(Q,\theta) + P_q(Q,\theta)\frac{Q}{n} - c_1 - w_1 e^* = 0 & if\ \theta\ \in [\theta_{\underline{P}}, \theta_{\overline{P}}] \\ X_1 & if\ \theta\ \in [\theta_{\overline{P}}, \overline{\theta}] \end{cases} \tag{15}$$

The critical spot market scenarios are defined as follows:

$$\begin{aligned} \theta_{\overline{B}} : \quad & P(X_2, \theta_{\overline{B}}) + P_q(X_2, \theta_{\overline{B}})\frac{1}{n} - c_2 - w_2 e^* = 0 \\ \theta_{\underline{P}} : \quad & P(X_2, \theta_{\underline{P}}) + P_q(X_2, \theta_{\underline{P}})\frac{1}{n} - c_1 - w_1 e^* = 0 \\ \theta_{\overline{P}} : \quad & P(X_1, \theta_{\overline{P}}) + P_q(X_1, \theta_{\overline{P}})\frac{1}{n} - c_1 - w_1 e^* = 0 \end{aligned}$$

That is, at spot market $\theta_{\overline{B}}$ investment X_2 in the base–load technology (c_2, k_2) starts to be binding, at $\theta_{\underline{P}}$ firms start to produce with the peak–load technology (c_1, k_1) and at $\theta_{\overline{P}}$ the total capacity bound X_1 is met. The first order conditions stated in lemma (1) obtain when equating expressions (16), (17), and (18) to zero. Notice that the case solution for the case perfect competition obtains as the special case where $n \to \infty$.

We first derive the first order conditions for optimal investment decisions. Note that, although in equilibrium at different demand realizations θ firm i might sometimes produce an unconstrained equilibrium quantity and sometimes is constrained by its choice x_{1i} or x_{2i}, equilibrium profit of firm i is continuous in θ. Thus, by Leibniz' rule, the first derivatives of the profit function are given as follows:

$$\frac{d\pi_i}{dx_{1i}} = \int_{\theta_{\overline{P}}}^{\bar{\theta}} \left[P(X_1,\theta) + P_q(X_1,\theta)\frac{X_1}{n} - (c_1 + w_1 e) \right] dF(\theta) - (k_1 - A_1 e) \tag{16}$$

$$\frac{d\pi_i}{dx_{2i}} = \int_{\theta_{\overline{B}}}^{\theta_{\underline{P}}} \left[P(X_2,\theta) + P_q(X_2,\theta)\frac{X_2}{n} - (c_2 + w_2 e) \right] dF(\theta) + \int_{\theta_{\underline{P}}}^{\bar{\theta}} (c_1 - c_2) + (w_1 - w_2) e dF(\theta) - (k_2 - A_2 e) + (k_1 - A_1 e) \tag{17}$$

In the market solution the emission price e has to be such as to equate the following expression to zero:

$$\int_{\underline{\theta}}^{\theta_{\overline{B}}} w_2 Q^* dF(\theta) + \int_{\theta_{\overline{B}}}^{\theta_{\underline{P}}} w_2 X_2 dF(\theta) + \int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} w_1 Q^* dF(\theta) + \int_{\theta_{\overline{P}}}^{\bar{\theta}} w_1 X_1 dF(\theta) - \int_{\theta_{\underline{P}}}^{\bar{\theta}} (w_1 - w_2) X_2 dF(\theta) - T \tag{18}$$

Which are the conditions Ψ_I, Ψ_{II} and Ψ_E as given in the lemma.

Let us directly at this point determine all partial derivatives of the equilibrium system characterized in the lemma. The partial derivatives of Ψ_I (expression (16)), Ψ_{II} (expression (17)) and Ψ_E (expression (18)) read as follows:

$$\frac{\partial \Psi_I}{\partial X_1^*} = \Psi_{I1} = \int_{\theta_{\bar{P}}}^{\bar{\theta}} P_q(X_1^*,\theta)\frac{n+1}{n} + P_{qq}(X_1^*,\theta)\frac{X_1^*}{n} dF(\theta) < 0$$

$$\frac{\partial \Psi_{II}}{\partial X_2^*} = \Psi_{II2} = \int_{\theta_{\bar{B}}}^{\theta_{\underline{P}}} P_q(X_2^*,\theta)\frac{n+1}{n} + P_{qq}(X_2^*,\theta)\frac{X_2^*}{n} dF(\theta) < 0$$

$$\frac{\partial \Psi_E}{\partial e} = \Psi_{Ee} = \int_{\underline{\theta}}^{\theta_{\bar{B}}} \frac{w_2^2}{P_q(Q^*,\theta)\frac{n+1}{n} + P_{qq}(Q^*,\theta)\frac{Q^*}{n}} dF(\theta) + \int_{\theta_{\underline{P}}}^{\theta_{\bar{P}}} \frac{w_1^2}{P_q(Q^*,\theta)\frac{n+1}{n} + P_{qq}(Q^*,\theta)\frac{Q^*}{n}} dF(\theta) < 0$$

$$\frac{\partial \Psi_E}{\partial X_1^*} = \Psi_{E1} = (1 - F(\theta_{\overline{P}}))\, w_1 = A_1^E > 0 \qquad \frac{\partial \Psi_I}{\partial e} = \Psi_{Ie} = A_1 - (1 - F(\theta_{\bar{P}}))w_1 = A_1 - A_1^E$$

$$\frac{\partial \Psi_E}{\partial X_2^*} = \Psi_{E2} = (1 - F(\theta_{\bar{B}}))w_2 - (1 - F(\theta_{\underline{P}}))w_1 = A_2^E$$

$$\frac{\partial \Psi_{II}}{\partial e} = \Psi_{IIe} = A_2 - A_1 - (1 - F(\theta_{\bar{B}}))w_2 + (1 - F(\theta_{\underline{P}}))w_1 = A_2 - A_1 - A_2^E$$

$$\frac{\partial \Psi_I}{\partial A_1} = e \qquad \frac{\partial \Psi_{II}}{\partial A_1} = -e \qquad \frac{\partial \Psi_{II}}{\partial A_2} = e \qquad \frac{\partial \Psi_E}{\partial T} = -1.$$

B Proof lemma 2

Part (i): To derive the second order conditions established in lemma 2, first observe that differentiation of the permit pricing condition Ψ_E with respect to X_1 and slight rearranging yields $\frac{de^*}{dX_1} = \frac{\Psi_{E1}}{-\Psi_{Ee}}$. Plugging into the derivatives of Ψ_I and Ψ_{II} and replacing for A_1^E and

A_2^E as introduced in definition 1 yields:

$$\frac{d\Psi_I}{dX_1} = \Psi_{I1} + \Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}} = \Psi_{I1} + (A_1 - A_1^E)\frac{A_1^E}{-\Psi_{Ee}} \qquad \frac{d\Psi_{II}}{dX_1} = \Psi_{IIe}\frac{\Psi_{E1}}{-\Psi_{Ee}} = (A_2 - A_1 - A_2^E)\frac{A_1^E}{-\Psi_{Ee}}$$

Likewise we obtain

$$\frac{d\Psi_I}{dX_2} = \Psi_{Ie}\frac{\Psi_{E2}}{-\Psi_{Ee}} = (A_1 - A_1^E)\frac{A_2^E}{-\Psi_{Ee}} \qquad \frac{d\Psi_{II}}{dX_1} = \Psi_{II2} + \Psi_{IIe}\frac{\Psi_{E2}}{-\Psi_{Ee}} = \Psi_{II2} + (A_2 - A_1 - A_2^E)\frac{A_2^E}{-\Psi_{Ee}}$$

The matrix $H = \begin{pmatrix} d\Psi_I/dX_1 & d\Psi_I/dX_2 \\ d\Psi_{II}/dX_1 & d\Psi_{II}/dX_2 \end{pmatrix}$ is negative definite if and only if conditions (a), (b) and (c) established in lemma 2 (i) are satisfied. To save on notation we introduce $C := det(H)$, observe that $C \geq 0$ if H is negative definite (compare (c) in lemma 2 (i)).

Part (ii): Since $\Psi_{I1} < 0$, $\Psi_{II2} < 0$ and $\Psi_{Ee} < 0$ (see appendix A) the conditions provided in lemma 2 (ii) are sufficient to guarantee negative definiteness of the matrix H.

Part (iii): To see why this is true, just observe that for $A_1 > A_1^{lim}$ condition (a) established in lemma 2(i) will be violated. The condition defining A_2^{lim} is given by the sum of conditions (a) and (b) of lemma 2(i), for $A_1 > A_1^{lim}$ at least one of those two conditions will be violated.

C Proof lemma 3

C.1 Preliminaries: Comparative Statics

The differentials for $\frac{dX_1^*}{dA_1}$, $\frac{dX_1^*}{dA_2}$, $\frac{dX_1^*}{d\mathcal{T}}$ and $\frac{dX_2^*}{dA_1}$, $\frac{dX_2^*}{dA_2}$, $\frac{dX_2^*}{d\mathcal{T}}$ are obtained by applying the implicity function theorem to the equilibrium conditions established in lemma 1. The total derivative of this equations system with respect to the parameter A_1 yields:

$$\Psi_I: \qquad \Psi_{I1}\frac{dX_1^*}{dA_1} + \Psi_{Ie}\frac{de^*}{dA_1} + \frac{\partial\Psi_I}{\partial A_1} \equiv 0 \tag{19}$$

$$\Psi_{II}: \qquad \Psi_{II2}\frac{dX_2^*}{dA_1} + \Psi_{IIe}\frac{de^*}{dA_1} + \frac{\partial\Psi_{II}}{\partial A_1} \equiv 0 \tag{20}$$

$$\Psi_E: \qquad \Psi_{E1}\frac{dX_1^*}{dA_1} + \Psi_{E2}\frac{dX_2^*}{dA_1} + \Psi_{Ee}\frac{de^*}{dA_1} + \frac{\partial\Psi_E}{\partial A_1} \equiv 0 \tag{21}$$

In order to derive an explicit formulation for $\frac{dX_1^*}{dA_1}$, we solve expression (21) for $\frac{de^*}{dA_1}$ and expression (20) for $\frac{dX_2^*}{dA_1}$. Plugging into expression (19) yields:

$$\left(\left(\Psi_{I1} + \Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right) - \left(\Psi_{Ie}\frac{\Psi_{E2}}{-\Psi_{Ee}}\right)\frac{\left(\Psi_{IIe}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right)}{\left(\Psi_{II2} + \Psi_{IIe}\frac{\Psi_{E2}}{-\Psi_{Ee}}\right)}\right)\frac{dX_1^*}{dA_1} + \frac{\partial\Psi_I}{\partial A_1} - \frac{\left(\Psi_{Ie}\frac{\Psi_{E2}}{-\Psi_{Ee}}\right)\frac{\partial\Psi_{II}}{\partial A_1}}{\left(\Psi_{II2} + \Psi_{IIe}\frac{\Psi_{E2}}{-\Psi_{Ee}}\right)} = 0$$

By making use of the definition of the variable C (compare appendix B) we can rearrange this expression and obtain:

$$\frac{dX_1^*}{dA_1} = (-1)\frac{\frac{\partial \Psi_I}{\partial A_1}\left(\Psi_{II2} + \Psi_{IIe}\frac{\Psi_{E2}}{-\Psi_{Ee}}\right) - \frac{\partial \Psi_{II}}{\partial A_1}\left(\Psi_{Ie}\frac{\Psi_{E2}}{-\Psi_{Ee}}\right)}{C} = \left((A_2 - A_1^E - A_2^E)A_2^E - \Psi_{II2}\Psi_{Ee}\right)\frac{-e}{-\Psi_{Ee}C} \quad (22)$$

Likewise, to obtain an explicit formulation for $\frac{dX_2^*}{dA_1}$, we analogously solve expression (21) for $\frac{de^*}{dA_1}$ and expression (19) for $\frac{dX_2^*}{dA_1}$. Plugging into expression (20) and solving for $\frac{dX_2^*}{dA_1}$ yields:

$$\frac{dX_2^*}{dA_1} = (-1)\frac{\frac{\partial \Psi_{II}}{\partial A_1}\left(\Psi_{I1} + \Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right) - \frac{\partial \Psi_I}{\partial A_1}\left(\Psi_{IIe}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right)}{C} = \left((A_2 - A_1^E - A_2^E)A_1^E - \Psi_{I1}\Psi_{Ee}\right)\frac{e}{-\Psi_{Ee}C} \quad (23)$$

Analogously we obtain:

$$\begin{aligned} \frac{dX_1^*}{dA_2} &= \left((A_1 - A_1^E)A_2^E\right)\frac{e}{-\Psi_{Ee}C} & \frac{dX_2^*}{dA_2} &= \left((A_1 - A_1^E)A_1^E - \Psi_{I1}\Psi_{Ee}\right)\frac{-e}{-\Psi_{Ee}C} \\ \frac{dX_1^*}{dT} &= \left(A_1 - A_1^E\right)\frac{\Psi_{II2}}{-\Psi_{Ee}C} & \frac{dX_2^*}{dT} &= \left(A_2 - A_1 - A_2^E\right)\frac{\Psi_{I1}}{-\Psi_{Ee}C} \end{aligned} \quad (24)$$

C.2 Proof lemma 3(i)

First, we define

$$\Gamma_1^{lim}(A_1) := \Psi_{I1} + \Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}} = \int_{\theta_{\overline{P}}}^{\overline{\theta}} P_q(X_1^*, \theta)dF(\theta) + \frac{(A_1 - A_1^E)A_1^E}{\int_{\underline{\theta}}^{\theta_{\overline{B}}} \frac{w_2^2}{-P_q(Q^*,\theta)}dF(\theta) + \int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} \frac{w_1^2}{-P_q(Q^*,\theta)}dF(\theta)}. \quad (25)$$

Observe, that the second order sufficient conditions for existence of the market equilibrium specified in 2(i) require $\Gamma_1^{lim}(A_1) < 0$. Since $\Gamma_1^{lim}(A_1)$ is increasing in A_1 we can define a unique A_1^{lim} which solves $\Gamma_1^{lim}(A_1^{lim}) = 0$ and conclude that $\frac{dX_2^*}{dA_2} > 0$ for all $A_1 < A_1^{lim}$.

Second, we define

$$\Gamma_1^{total}(A_1) := \Psi_{Ie}\frac{\Psi_{E2}}{-\Psi_{Ee}} = \frac{\left(A_1 - A_1^E\right)A_2^E}{\int_{\underline{\theta}}^{\theta_{\overline{B}}} \frac{w_2^2}{-P_q(Q^*,\theta)}dF(\theta) + \int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} \frac{w_1^2}{-P_q(Q^*,\theta)}dF(\theta)} \quad (26)$$

This allows us to rewrite $\frac{dX_1^*}{dA_2}$ as established in expression (24) as follows:

$$\frac{dX_1^*}{dA_2} = \Gamma_1^{total}(A_1)\frac{e^*}{C} \quad (27)$$

We finally show that $A_1^E < A_1^{lim}$. To see this, observe that $\Gamma_1^{lim}(A_1^E) = \Psi_{I1} < 0$. Since $\Gamma_1^{lim}(A_1)$ is increasing in A_1, we necessarily obtain $A_1^{lim} > A_1^E$.

Third, we define

$$\Gamma_1^{cross}(A_1) := \Psi_{I1} + \Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}} + \Psi_{Ie}\frac{\Psi_{E2}}{-\Psi_{Ee}} = \int_{\theta_{\overline{P}}}^{\overline{\theta}} P_q(X_1^*,\theta)dF(\theta) + \frac{\left(A_1 - A_1^E\right)\left(A_1^E + A_2^E\right)}{\int_{\underline{\theta}}^{\theta_{\overline{B}}} \frac{w_2^2}{-P_q(Q^*,\theta)}dF(\theta) + \int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} \frac{w_1^2}{-P_q(Q^*,\theta)}dF(\theta)}$$

Observe that $\frac{dX_2^*}{dA_2} < \frac{dX_1^*}{dA_2}$ if and only if $\Gamma_1^{cross}(A_1) > 0$ (compare expression (24)). We define the locus where $\Gamma_1^{cross}(A_1) = 0$ by A_1^{cross}. We now compare the critical allocation A_1^{cross} relative to the critical values A_1^{lim} and A_1^E:

- For $w_2 > w_2^E$, we can establish the following ranking: $A_1^E < A_1^{cross} < A_1^{lim}$.

  To show the first inequality observe that for all $A_1 \leq A_1^E$ we obtain $\Gamma_1^{total}(A_1) \leq 0$. As shown above we also obtain $\Gamma_1^{lim}(A_1) \leq 0$, which implies $\Gamma_1^{cross}(A_1) = \Gamma_1^{lim}(A_1) + \Gamma_1^{total}(A_1) \leq 0$. This directly implies, however, that A_1^{cross} cannot be in the interval $[0, A_1^E]$.

  To show the second inequality observe that for $A_1 > A_1^E$ we have $\Gamma_1^{total}(A_1) > 0$. Whenever $\Gamma_1^{cross}(A_1) = \Gamma_1^{lim}(A_1) + \Gamma_1^{total}(A_1) = 0$ we must thus have $\Gamma_1^{lim}(A_1) < 0$. Since $\Gamma_1^{lim}(A_1)$ is strictly increasing in A_1 this implies $A_1^{lim} > A_1^{cross}$.

- For $w_2 < w_2^E$ we establish that $\Gamma_1^{cross}(A_1) < 0$ (i.e. $A_1^* > 0$) for all $A_1 \in (0, A_1^{lim}]$.

  First, observe that for $A_1 > A_1^E$ we obtain $\Gamma_1^{total}(A_1) < 0$, which implies that $\Gamma_1^{cross}(A_1) = \Gamma_1^{lim}(A_1) + \Gamma_1^{total}(A_1) < 0$ for $A_1 \in [A_1^E, A_1^{lim}]$.

  Second observe that for $A_1 \leq A_1^E$ and $((1 - F(\theta_{\overline{B}}))w_2 - (F(\theta_{\overline{P}}) - F(\theta_{\underline{P}}))w_1) > 0$ we obtain $\Gamma_1^{cross}(A_1) < 0$.

  Third observe that for $A_1 \leq A_1^E$ and $((1 - F(\theta_{\overline{B}}))w_2 - (F(\theta_{\overline{P}}) - F(\theta_{\underline{P}}))w_1) < 0$, $\Gamma_1^{cross}(A_1)$ is maximized for $A_1 = 0$. Expression (28) then reads as follows:

$$\Gamma_1^{cross}(0) = \int_{\theta_{\overline{P}}}^{\overline{\theta}} P_q(X_1^*,\theta)dF(\theta) + \frac{-\left(1 - F(\theta_{\overline{P}})\right) w_1 \left((1 - F(\theta_{\overline{B}}))w_2 - (F(\theta_{\overline{P}}) - F(\theta_{\underline{P}}))w_1\right)}{\int_{\underline{\theta}}^{\theta_{\overline{B}}} \frac{w_2^2}{-P_q(Q^*,\theta)}dF(\theta) + \int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} \frac{w_1^2}{-P_q(Q^*,\theta)}dF(\theta)}$$

  Which can be guaranteed to be negative if $P_{qq} \leq 0$ and $P_{q\theta} \geq 0$. Without those additional assumptions it might happen that $A_1^* = 0$ in the region where $A_1 \leq A_1^E$ and $((1 - F(\theta_{\overline{B}}))w_2 - (F(\theta_{\overline{P}}) - F(\theta_{\underline{P}}))w_1) < 0$.

C.3 Proof lemma 3(ii)

First, we define

$$\Gamma_2^{lim}(A_2) := \left(\Psi_{II2} + \Psi_{I1} + (\Psi_{IIe} + \Psi_{Ie})\,\frac{\Psi_{E2} + \Psi_{E1}}{-\Psi_{Ee}}\right). \tag{28}$$

In order to precisely define A_2^{lim} first observe that $\Psi_{E2}+\Psi_{E1} = A_1^E + A_2^E$ (compare definition 1), we thus have to consider the following two cases:

- For the case $w_2 > w_2^L$ (i.e. $A_1^E + A_2^E > 0$) Γ_2^{lim} is strictly increasing in A_2, we can define a unique A_2^{lim} which solves $\Gamma_2^{lim}(A_2^{lim}) = 0$
- For the case $w_2 \leq w_2^L$ (i.e. $A_1^E + A_2^E \leq 0$) $\Gamma_2^{lim}(A_2)$ is non-increasing in A_2, it is thus minimized for $A_2 = 0$, which yields

$$\Gamma_2^{lim} = \Psi_{II2} + \Psi_{I1} + \left(A_2 - (A_1^E + A_2^E)\right) \frac{A_1^E + A_2^E}{-\Psi_{Ee}} < 0.$$

  That is, $\Gamma_2^{lim}(A_1) < 0$ for all $A_2 \geq 0$. For ease of notation we thus define $A_2^{lim} = \infty$ in this case.

We now determine $\frac{dX_2^*}{dA_1} - \frac{dX_2^*}{dA_1}$ as given by expressions (22) and (22). After plugging in for $\frac{\partial \Psi_I}{\partial A_1} = e^*$ and $\frac{\partial \Psi_{II}}{\partial A_1} = -e^*$ (compare appendix A) we obtain for $\frac{dX_1^*}{dA_1} - \frac{dX_2^*}{dA_1}$:

$$\left(\left(\Psi_{II2} + \Psi_{IIe}\frac{\Psi_{E2}}{-\Psi_{Ee}}\right) + \left(\Psi_{Ie}\frac{\Psi_{E2}}{-\Psi_{Ee}}\right) + \left(\Psi_{I1} + \Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right) + \left(\Psi_{IIe}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right)\right)\frac{-e^*}{C} = \frac{-e^*}{C}\Gamma_2^{lim}(A_2)$$

Since $\Gamma_2^{lim}(A_2) < 0$ as established above, we conclude that $\frac{dX_1^*}{dA_1} > \frac{dX_2^*}{dA_1}$ for all $A_2 \in [0, A_2^{lim}]$. Second, we define

$$\Gamma_2^{total}(A_2) := \left(\left(\Psi_{II2} + \Psi_{IIe}\frac{\Psi_{E2}}{-\Psi_{Ee}}\right) + \left(\Psi_{Ie}\frac{\Psi_{E2}}{-\Psi_{Ee}}\right)\right) = \left(\Psi_{II2} + \left(A_2 - A_1^E - A_2^E\right)\frac{A_2^E}{-\Psi_{Ee}}\right) \tag{29}$$

- For $w_2 > w_2^E$ (i.e. $A_2^E > 0$, see definition 1) $\Gamma_2^{total}(A_2)$ is strictly increasing in A_2. Since $\Gamma_2^{total}(0) < 0$, we can thus define a unique $A_2^{total} > 0$ which satisfies $\Gamma_2^{total}(A_2^{total}) = 0$.
- For the case $w_2 \leq w_2^E$ (i.e. $A_2^E \leq 0$, see definition 1) $\Gamma_2^{total}(A_2)$ is non-increasing in A_2. Observe that in this case Γ_2^{total} is maximized for $(A_2 = 0, w_2 = 0)$ which yields:

$$\Gamma_2^{total}(0) = \left(\Psi_{II2} + \left(A_2 - A_1^E - A_2^E\right)\frac{A_2^E}{-\Psi_{Ee}}\right) < 0. \tag{30}$$

  That is, for $A_2 \geq 0$ we obtain $\Gamma_2^{total}(A_2) < 0$. For ease of notation we thus define $A_2^{total} := \infty$ whenever $w_2 \leq w_2^E$.

Observe now that we can rewrite $\frac{dX_1^*}{dA_1}$ (established in expression (23)) in terms of Γ_2^{total}, which yields:

$$\frac{dX_1^*}{dA_1} = \Gamma_2^{total}(A_2)\frac{-e^*}{C} \tag{31}$$

Third, we define

$$\Gamma_2^{cross}(A_2) := \left(\Psi_{I1} + \Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right) + \left(\Psi_{IIe}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right) = \Psi_{I1} + \left(A_2 - A_1^E - A_2^E\right)\frac{A_1^E}{-\Psi_{Ee}} \tag{32}$$

We now rewrite $\frac{dX_2^*}{dA_1}$ as established in expression (23) in terms of Γ_2^{cross}, which yields:

$$\frac{dX_2^*}{dA_1} = \Gamma_2^{cross}(A_2)\frac{e^*}{C} \tag{33}$$

Observe that Γ_2^{cross} is strictly increasing in A_2 (see appendix A). Since $\Gamma_2^{cross}(0) < 0$, we can thus define a unique $A_2^{cross} > 0$ which satisfies $\Gamma_2^{cross}(A_2^{cross}) = 0$:

$$A_2^{cross} = A_1^E + A_2^E + \frac{\Psi_{I1}\Psi_{Ee}}{A_1^E} \tag{34}$$

Finally we compare the different critical values: A_2^{lim}, A_2^{total}, and A_2^{cross}. We have to consider the following three cases:

- For $w_2 \leq w_2^L$: In this case we obtain $\Gamma_2^{lim}(A_2) < 0$ and $\Gamma_2^{total}(A_2) < 0$ for all $A_2 \geq 0$. Thus A_2^{cross} provides the only critical level of initial allocation (remember, we defined $A_2^{lim} = \infty$ and $A_2^{total} = \infty$), we thus obtain $A_2^{cross} < A_2^{lim}$.
- For $w_2^L < w_2 \leq w_2^E$: In this case we obtain $\Gamma_2^{total}(A_2) < 0$ for all $A_2 \geq 0$ (remember, we defined $A_2^{total} = \infty$). Observe, furthermore that $\Gamma_2^{lim}(A_2) = \Gamma_2^{total}(A_2) + \Gamma_2^{cross}(A_2)$ for all A_2. This directly implies, however, that $\Gamma_2^{cross}(A_2^{lim}) > 0$ and thus $A_2^{cross} < A_2^{lim}$.
- For $w_2 \geq w_2^E$ we evaluate $\Gamma_2^{total}(A_2^{cross})$, which yields (compare expressions (29) and(34)):

$$\Gamma_2^{total}(A_2^{cross}) = \Psi_{II2} - \Psi_{I1}\frac{A_2^E}{A_1^E} = \overline{P_q^2}\left(F(\theta_{\underline{P}}) - F(\theta_{\overline{B}})\right) - \overline{P_q^1}\frac{(1-F(\theta_{\overline{B}}))w_2 - (1-F(\theta_{\underline{P}}))w_1}{w_1} \tag{35}$$

Observe that we denote by $\overline{P_q^1}$ the average slope of demand for those demand levels where total investment is binding and by $\overline{P_q^2}$ the average slope of demand for those demand levels where base load investment is binding, i.e.:

$$(i) \quad \overline{P_q^1} := \frac{\int_{\theta_{\overline{P}}}^{\overline{\theta}} P_q(X_1^*,\theta)dF(\theta)}{1-F(\theta_{\overline{P}})} \qquad\qquad (ii) \quad \overline{P_q^2} := \frac{\int_{\theta_{\overline{B}}}^{\theta_{\underline{P}}} P_q(X_2^*,\theta)dF(\theta)}{F(\theta_{\underline{P}}) - F(\theta_{\overline{B}})} \tag{36}$$

Rearranged this yields:

$$\Gamma_2^{total}(A_2^{cross}) = \frac{-\overline{P_q^1}(1-F(\theta_{\overline{B}}))}{w_1}\left(w_2 - \left(\frac{(1-F(\theta_{\underline{P}})) + (F(\theta_{\underline{P}}) - F(\theta_{\overline{B}}))\frac{\overline{P_q^2}}{\overline{P_q^1}}}{1-F(\theta_{\overline{B}})}\right)w_1\right) \tag{37}$$

Now define

$$w_2^S := \frac{(1-F(\theta_{\underline{P}})) + (F(\theta_{\underline{P}}) - F(\theta_{\overline{B}}))\frac{\overline{P_q^2}}{\overline{P_q^1}}}{1-F(\theta_{\overline{B}})}w_1 \tag{38}$$

Observe, that $w_2^E < w_2^S \leq w_1$ since $F(\theta_{\underline{P}}) - F(\theta_{\overline{B}}) > 0$ and $0 < \frac{\overline{P_q^2}}{\overline{P_q^1}} \leq 1$, notice that for $\overline{P_q^2} = \overline{P_q^1}$ (e.g. for $P_{qq} = P_{q\theta} = 0$) we obtain $w_2^S = w_1$. Furthermore, for

$w_2 > w_2^S$ we obtain $\Gamma_2^{total}(A_2^{cross}) > 0$ and for $w_2 < w_2^S$ we obtain $\Gamma_2^{total}(A_2^{cross}) < 0$. Since $\Gamma_2^{lim}(A_2) = \Gamma_2^{total}(A_2) + \Gamma_2^{cross}(A_2)$ for all A_2 we obtain:

$$\begin{array}{lll} 0 < A_2^{total} < A_2^{lim} < A_2^{cross} & \text{if} & w_2 > w_2^S \\ 0 < A_2^{total} = A_2^{lim} = A_2^{cross} & \text{if} & w_2 = w_2^S \\ 0 < A_2^{cross} < A_2^{lim} < A_2^{total} & \text{if} & w_2^E < w_2 < w_2^S \end{array} \tag{39}$$

C.4 Proof lemma 3(iii)

Observe that we have derived $\frac{dX_1^*}{dT}$ and $\frac{dX_2^*}{dT}$ in expression (24). The statements of lemma 3(iii) follow directly since $\Psi_{I1} < 0$, $\Psi_{II2} < 0$ and $\Psi_{Ee} < 0$ (see definition 1).

D Proof lemma 4

To derive the optimal design of the cap and trade mechanism (A_1, A_2, T) we first differentiate Welfare as given by expression (10) with respect to each of those parameters. We obtain for $\frac{dW}{dA_1}$:

$$\begin{aligned} \frac{dW}{dA_1} = & \int_{\underline{\theta}}^{\theta_{\overline{B}}} \frac{dQ^*}{dA_1} \left[P(Q^*, \theta) - c_2\right] dF(\theta) + \int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} \frac{dQ^*}{dA_1} \left[P(Q^*, \theta) - c_1\right] dF(\theta) + \\ & \frac{dX_2^*}{dA_1} \left[\int_{\theta_{\overline{B}}}^{\theta_{\overline{P}}} \left[P(X_2^*, \theta) - c_2\right] dF(\theta) + \int_{\theta_{\underline{P}}}^{\overline{\theta}} (c_1 - c_2) dF(\theta) - (k_2 - k_1) \right] + \\ & \frac{dX_1^*}{dA_1} \left[\int_{\theta_{\overline{P}}}^{\overline{\theta}} \left[P(X_1^*, \theta) - c_1\right] dF(\theta) - k_1 \right] \end{aligned}$$

We can now plug in the equilibrium conditions for firms' investment choices given by expressions (2) and (3) and we can plug in the optimality conditions for the unconstrained spot markets, whenever investment is not binding. [36] This yields:

$$\begin{aligned} \frac{dW}{dA_1} = & \int_{\underline{\theta}}^{\theta_{\overline{B}}} \frac{dQ^*}{dA_1} \left[-P_q \frac{Q^*}{n} + w_2 e\right] dF(\theta) + \int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} \frac{dQ^*}{dA_1} \left[-P_q \frac{Q^*}{n} + w_1 e\right] dF(\theta) \\ & \frac{dX_2^*}{dA_1} \left[\int_{\theta_{\overline{B}}}^{\theta_{\underline{P}}} \left[-P_q \frac{X_2^*}{n} + w_2 e\right] dF(\theta) + \int_{\theta_{\underline{P}}}^{\overline{\theta}} (w_2 - w_1) e dF(\theta) - (A_2 - A_1) e \right] + \\ & \frac{dX_1^*}{dA_1} \left[\int_{\theta_{\overline{P}}}^{\overline{\theta}} \left[-P_q + w_1 e\right] dF(\theta) - A_1 e \right] \end{aligned}$$

This can be further simplified by making use of the derivative of the permit pricing given by expression 4 with respect to A_1 (i.e. $\frac{\Psi_E}{A_1}$. This allows to eliminate all terms containing the emission factors w_1 and w_2 from the above expression (shown explicitly in expression

[36] That is for spot markets $\theta \in [\underline{\theta}, \theta_{\overline{B}}] \cup [\theta_{\underline{P}}, \theta_{\overline{P}}]$, in those cases the optimality conditions are simply given by $P(Q^*, \theta) + P_q \frac{Q^*}{n} - c_i - w_i e^* = 0$, for $i = 1, 2$.

(41) just below). We thus obtain

$$\frac{dW}{dA_1} = \int_{\underline{\theta}}^{\theta_{\overline{B}}} \frac{dQ^*}{dA_1}\left[-P_q\frac{Q^*}{n}\right]dF(\theta) + \int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} \frac{dQ^*}{dA_1}\left[-P_q\frac{Q^*}{n}\right]dF(\theta)$$
$$\frac{dX_2^*}{dA_1}\left[\int_{\theta_{\overline{B}}}^{\theta_{\underline{P}}}\left[-P_q\frac{X_2^*}{n}\right]dF(\theta) - (A_2 - A_1)e\right] + \frac{dX_1^*}{dA_1}\left[\int_{\theta_{\overline{P}}}^{\overline{\theta}}\left[-P_q\frac{X_1^*}{n}\right]dF(\theta) - A_1 e\right] \quad (40)$$

In order to show why indeed expression (40) is obtained, we now differentiate the permit pricing condition Ψ_E (compare lemma 1) with respect to A_1, this yields

$$-\int_{\underline{\theta}}^{\theta_{\overline{B}}} w_2\frac{dQ^*}{dA_1}dF(\theta) - \int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} w_1\frac{dQ^*}{dA_1}dF(\theta) =$$
$$\frac{dX_2^*}{dA_1}\left[\int_{\theta_{\overline{B}}}^{\theta_{\underline{P}}} w_2 dF(\theta) + \int_{\theta_{\underline{P}}}^{\overline{\theta}}(w_2 - w_1)dF(\theta)\right] + \frac{dX_1^*}{dA_1}\left[\int_{\theta_{\overline{P}}}^{\overline{\theta}} w_1 dF(\theta)\right]. \quad (41)$$

Now observe that $\frac{dQ^*}{dA_1} = \frac{dQ^*}{de^*}\frac{de^*}{dA_1^*}$, since unconstrained spot market output does not directly depend on the degree of free allocation A_1. By multiplying expression (41) with Δ (as defined in the lemma) we obtain:

$$\left(\int_{\underline{\theta}}^{\theta_{\overline{B}}} \frac{dQ^*}{de^*}\left[-P_q\frac{Q^*}{n}\right]dF(\theta) + \int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} \frac{dQ^*}{de^*}\left[-P_q\frac{Q^*}{n}\right]dF(\theta)\right)\frac{de^*}{dA_1} = \frac{-\Delta}{n}\left(\frac{dX_2^*}{dA_1}A_2^E + \frac{dX_1^*}{dA_1}A_1^E\right) \quad (42)$$

We can now plug expression (42) into expression (40), which yields

$$\frac{dW}{dA_1} = \frac{-\Delta}{n}\left(\frac{dX_2^*}{dA_1}A_2^E + \frac{dX_1^*}{dA_1}A_1^E\right) +$$
$$\frac{dX_2^*}{dA_1}\left[\int_{\theta_{\overline{B}}}^{\theta_{\underline{P}}}\left(-P_q\frac{X_2^*}{n}\right)dF(\theta) - (A_2 - A_1)e\right] + \frac{dX_1^*}{dA_1}\left[\int_{\theta_{\overline{P}}}^{\overline{\theta}}\left(-P_q\frac{X_1^*}{n}\right)dF(\theta) - A_1 e\right]$$

Rearranging finally yields

$$\frac{dW}{dA_1} = \frac{dX_2^*}{dA_1}\left[\int_{\theta_{\overline{B}}}^{\theta_{\underline{P}}}\left(-P_q\frac{X_2^*}{n}\right)dF(\theta) - (A_2 - A_1)e - \frac{\Delta}{n}A_2^E\right] +$$
$$\frac{dX_1^*}{dA_1}\left[\int_{\theta_{\overline{P}}}^{\overline{\theta}}\left(-P_q\frac{X_1^*}{n}\right)dF(\theta) - A_1 e - \frac{\Delta}{n}A_1^E\right]$$

Which corresponds exactly to the expression for $\frac{dW}{dA_1}$ stated in the lemma. The very same steps yield $\frac{dW}{dA_2} = \frac{dX_1^*}{dA_2}\Omega_I + \frac{dX_2^*}{dA_2}\Omega_{II}$ as stated in the lemma.

We finally determine $\frac{dW}{dT}$. Analogous to expression (40) we obtain:

$$\frac{dW}{dT} = \frac{dX_2^*}{dT}\left[\int_{\theta_{\overline{B}}}^{\theta_{\underline{P}}}\left[-P_q\frac{X_2^*}{n}\right]dF(\theta) - (A_2 - A_1)e\right] + \frac{dX_1^*}{dT}\left[\int_{\theta_{\overline{P}}}^{\overline{\theta}}\left[-P_q\frac{X_1^*}{n}\right]dF(\theta) - A_1 e\right] \quad (43)$$
$$\int_{\underline{\theta}}^{\theta_{\overline{B}}} \frac{dQ^*}{dT}\left[-P_q\frac{Q^*}{n}\right]dF(\theta) + \int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} \frac{dQ^*}{dT}\left[-P_q\frac{Q^*}{n}\right]dF(\theta) + e^* - D_T(T)$$

Which obtains since all terms containing the emission factors w_1 and w_2 integrate to 1. Why this is the case becomes clear when differentiating the permit pricing condition Ψ_E with respect to T:

$$-\int_{\underline{\theta}}^{\theta_{\overline{B}}} w_2 \frac{dQ^*}{dT} dF(\theta) - \int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} w_1 \frac{dQ^*}{dT} dF(\theta) = \frac{dX_2^*}{dT} A_2^E + \frac{dX_1^*}{dT} A_1^E - 1.$$

Now observe that $\frac{dQ^*}{dT} = \frac{dQ^*}{de^*}\frac{de^*}{dT}$, since unconstrained spot market output does not directly depend on the total emission cap T. By multiplying expression (44) with $\frac{\Delta}{n}$ we obtain:

$$\left(\int_{\underline{\theta}}^{\theta_{\overline{B}}} \frac{dQ^*}{de^*}\left[-P_q \frac{Q^*}{n}\right] dF(\theta) + \int_{\theta_{\underline{P}}}^{\theta_{\overline{P}}} \frac{dQ^*}{de^*}\left[-P_q \frac{Q^*}{n}\right] dF(\theta)\right) \frac{de^*}{dT} = \frac{-\Delta}{n}\left(\frac{dX_2^*}{dT} A_2^E + \frac{dX_1^*}{dT} A_1^E - 1\right) \quad (44)$$

We can now plug expression (44) into expression (43), which yields after rearranging:

$$\begin{aligned}\frac{dW}{dT} \quad = \quad & \frac{dX_2^*}{dT}\left[\int_{\theta_{\overline{B}}}^{\theta_{\underline{P}}} \left(-P_q \frac{X_2^*}{n}\right) dF(\theta) - (A_2 - A_1)e - \frac{\Delta}{n} A_2^E\right] + \\ & \frac{dX_1^*}{dT}\left[\int_{\theta_{\overline{P}}}^{\overline{\theta}} \left(-P_q \frac{X_1^*}{n}\right) dF(\theta) - A_1 e - \frac{\Delta}{n} A_1^E\right] + \frac{\Delta}{n} + e^* - D_T(T)\end{aligned}$$

Which corresponds exactly to the expression for $\frac{dW}{dT}$ stated in the lemma.

E Proof theorems 1 and 2

The optimality conditions established in lemma 4 are satisfied if the following conditions hold:

$$\begin{aligned}\Omega_I &= \int_{\theta_{\overline{P}}}^{\overline{\theta}} \frac{-P_q X_1^*}{n} dF(\theta) - A_1 e^* - \frac{\Delta}{n} A_1^E = 0 \\ \Omega_{II} &= \int_{\theta_{\overline{B}}}^{\theta_{\underline{P}}} \frac{-P_q X_2^*}{n} dF(\theta) - (A_2 - A_1)e^* - \frac{\Delta}{n} A_2^E = 0 \\ e^* &= D_T(T) - \frac{\Delta}{n}.\end{aligned}$$

The case of perfect competition as analyzed in theorem 1 obtains for $n->infty$. Observe that elimination of all terms involving the number of firms n in the denominator in the above conditions yields the characterization of the first best solution stated in theorem 1. In order to obtain the solution obtained for the case of imperfect competition as established in theorem 2 we solve the first two conditions for the levels of free allocation A_1 and A_2.

F Proof theorem 4

The optimality condition for $\mathcal{T}$ has been derived in lemma 4 (iii). After plugging in the results of comparative statics for $\frac{dX_1}{d\mathcal{T}}$ and $\frac{dX_2}{d\mathcal{T}}$ derived in expression (24) we obtain:

$$e^* - \mathcal{D}_{\mathcal{T}} = \left((A_2 - A_1)e\Psi_{IIe}\frac{\Psi_{I1}}{-\Psi_{Ee}\, C} + A_1 e \Psi_{Ie}\frac{\Psi_{II2}}{-\Psi_{Ee}\, C}\right) \quad (45)$$

We now make use of the notation introduced in definition 1, which allows us to rewrite expression (45) as follows:

$$e^* - \mathcal{D}_\mathcal{T} = \left((A_2 - A_1 - A_2^E)(A_2 - A_1)\Psi_{II2} + (A_1 - A_1^E)A_1\Psi_{I1}\right)\frac{e^*}{-\Psi_{Ee}\,C} \tag{46}$$

Notice that $\frac{e^*}{-\Psi_{Ee}\,C} > 0$ as established in appendix A. The remainder of the right hand side of expression (46) states $\Gamma_0(A_1, A_2)$ as defined in expression (14). The expression $e^* - D_T(T)$ and Γ_0 do thus exhibit the same sign, which proofs the theorem.

G Proof theorem 5

As a **first step** we determine the properties of the optimal allocation A_2^*. Observe that the optimality condition $A_2^* = \frac{dX_2^*/dA_2 - dX_1^*/dA_2}{dX_2^*/dA_2}A_1$ stated in the theorem directly obtains by rearranging expression (12). In lemma 3 (i) we have established $\frac{dX_2^*}{dA_2} > 0$ for all $A_1 < A_1^{lim}$. We thus obtain $A_2^* > A_1$ if and only if $\frac{dX_1^*}{dA_2} < 0$. Furthermore, we obtain $A_2^* = 0$ if $\frac{dX_2^*}{dA_2} < \frac{dX_1^*}{dA_2}$ since we only consider non–negative levels of free allocation. By making use of the properties of comparative statics established in lemma 3 (i) we directly obtain the properties of A_2^* as stated in the theorem.

As a **second step** we determine the optimal emission cap T^*. The optimality condition for T^* has been derived in expression (13) and yields after substituting for A_2^*:

$$e^* - \mathcal{D}_\mathcal{T} = \frac{dX_1^*}{d\mathcal{T}}A_1 e - \frac{dX_2}{d\mathcal{T}}A_1 e\left[\frac{dX_1/dA_2}{dX_2/dA_2}\right].$$

After substituting for $\frac{dX_1}{d\mathcal{T}}$, $\frac{dX_2}{d\mathcal{T}}$, $\frac{dX_1}{dA_2}$, and $\frac{dX_2}{dA_2}$ (expression (24)) this reads as follows:

$$e^* - \mathcal{D}_\mathcal{T} = \left((\Psi_{II2}\Psi_{Ie}) - (\Psi_{I1}\Psi_{IIe})\left[\frac{\left(-\Psi_{Ie}\frac{\Psi_{E2}}{\Psi_{Ee}}\right)}{-\left(\Psi_{I1} - \Psi_{Ie}\frac{\Psi_{E1}}{\Psi_{Ee}}\right)}\right]\right)\frac{A_1 e}{-\Psi_{Ee}\,C}$$

Rearranging and plugging in for Ψ_{Ie} (compare appendix A) we obtain:

$$e^* - \mathcal{D}_\mathcal{T} = \left(A_1^E - A_1\right)\;\frac{\left(\Psi_{I1}\left(\Psi_{IIe}\frac{\Psi_{E2}}{-\Psi_{Ee}}\right) + \Psi_{II2}\left(\Psi_{I1} + \Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right)\right)A_1 e}{\left(\Psi_{I1} + \Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right)\Psi_{Ee}\,C} \tag{47}$$

Observe that for $A_1 < A_1^{lim}$ the sign of the right hand side of expression (47) is entirely determined by the expression $(A_1^E - A_1)$, the remainder of expression (47) is strictly positive since $\left(\Psi_{I1} + \Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right) < 0$ and $\left(\Psi_{IIe}\frac{\Psi_{E2}}{-\Psi_{Ee}}\right) < 0$ (as shown further below in step three).

Finally notice that expression (47) has been derived without non–negativity constraint on A_2^*. As shown above, however, for $w_2 > w_2^E$ and $A_1 \in [A_1^E, A_1^{lim}]$ we obtain $A_2^* = 0$

(instead of a negative value as resulting in the computations leading to expression (47)). In this case the optimality condition given by expression (13) simplifies as follows:

$$e^* - \mathcal{D}_{\mathcal{T}} = \left(\frac{dX_1^*}{d\mathcal{T}} - \frac{dX_2}{d\mathcal{T}}\right) A_1 e = \left(\Psi_{II2}\Psi_{Ie} - \Psi_{I1}\Psi_{IIe}\right) \frac{A_1 e}{-\Psi_{Ee}\, C} < 0 \tag{48}$$

The inequality obtains since $\Psi_{Ie} > 0$ and $\Psi_{IIe} < 0$ for $w_2 > w_2^E$ and $A_1 \in [A_1^E, A_1^{lim}]$ (compare appendix A). We thus summarize the results obtained in expressions (47) and (54) as follows:

$$\begin{cases} e^* > D_T & \text{if} \quad A_1 < A_1^E \\ e^* = D_T & \text{if} \quad A_1 = A_1^E \\ e^* < D_T & \text{if} \quad A_1 > A_1^E \end{cases} \tag{49}$$

As a **third step** We finally show that the second order conditions established in lemma 2 (i) are indeed satisfied for all $A_1 \in [0, A_1^{lim}]$, A_2^*. Since $\left(\Psi_{I1} + \Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right) < 0$ for $A_1 < A_1^{lim}$ we just need to show that $\left(\Psi_{IIe}\frac{\Psi_{E2}}{-\Psi_{Ee}}\right) < 0$ for $A_1 < A_1^{lim}$. Notice first that we can rewrite Ψ_{IIe} (compare appendixA) and thus obtain:

$$\Psi_{IIe}\frac{\Psi_{E2}}{-\Psi_{Ee}} = \left(A_2^* - A_1 - \Psi_{E2}\right)\frac{\Psi_{E2}}{-\Psi_{Ee}} \tag{50}$$

Furthermore, we obtain for $(A_2^* - A_1)$ (compare expressions (12) and (24)):

$$(A_2^* - A_1) = -\frac{\frac{dX_1}{dA_2}}{\frac{dX_2}{dA_2}} A_1 = \frac{\left(\Psi_{Ie}\frac{\Psi_{E2}}{-\Psi_{Ee}}\right) A_1}{\left(\Psi_{I1} + \Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right)}$$

Plugging in allows us to rewrite expression (50) as follows:

$$\Psi_{IIe}\frac{\Psi_{E2}}{-\Psi_{Ee}} = \frac{\left(\Psi_{E2}\right)^2}{\left(\Psi_{I1} + \Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right)\left(-\Psi_{Ee}\right)}\left(\Psi_{Ie}\frac{A_1}{-\Psi_{Ee}} - \left(\Psi_{I1} + \Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right)\right)$$

Substituting for $\Psi_{Ie} = A_1 - A_1^E$ and $\Psi_{E1} = A_1^E$ (see appendix A) then yields:

$$\Psi_{IIe}\frac{\Psi_{E2}}{-\Psi_{Ee}} = \frac{\left(\Psi_{E2}\right)^2}{\left(\Psi_{I1} + \Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right)\left(\Psi_{Ee}\right)}\left(\Psi_{I1} - \frac{(A_1 - A_1^E)^2}{-\Psi_{Ee}}\right) < 0$$

We can thus conclude that the second order conditions established in in lemma 2 (i) are satisfied if and only if $A_1 < A_1^{lim}$ (Notice that the "only if" part follows directly from lemma 2 (iii)).

H Proof theorem 6

As a **first step** we determine the properties of the optimal allocation A_1^*. Observe that the optimality condition $A_1^* = \frac{dX_2^*/dA_1}{dX_2^*/dA_1 - dX_1^*/dA_1} A_2$ stated in the theorem directly obtains by rearranging expression (11). In lemma 3 (ii) we have established $\frac{dX_1^*}{dA_1} > \frac{dX_2^*}{dA_1}$ for all $A_2 < A_2^{lim}$. We thus obtain $A_1^* > 0$ if and only if $\frac{dX_2^*/dA_1}{<}0$. Furthermore, we obtain $A_1^* > A_2$ if and only if $\frac{dX_2^*}{dA_1} < \frac{dX_1^*}{dA_1} < 0$. By making use of the properties of comparative statics established in lemma 3 (ii) we directly obtain the properties of A_1^* as stated in the theorem.

As a **second step** we determine the optimal emission cap T^*. The optimality condition for T^* has been derived in expression (13) and yields after substituting for A_1^* (compare the first step above):

$$e^* - \mathcal{D}_{\mathcal{T}} = \frac{dX_2}{d\mathcal{T}}\left(A_2 - A_1^*\right)e^* + \frac{dX_1^*}{d\mathcal{T}}A_1^*e^* = \frac{dX_2}{d\mathcal{T}}\frac{\frac{dX_1}{dA_1}A_2e^*}{\frac{dX_1}{dA_1} - \frac{dX_2}{dA_1}} + \frac{dX_1^*}{d\mathcal{T}}\frac{-\frac{dX_2}{dA_1}A_2e^*}{\frac{dX_1}{dA_1} - \frac{dX_2}{dA_1}}$$

We can now plug in for $\frac{dX_1}{dA_1}$ and $\frac{dX_2}{dA_1}$ as derived in expressions (22) and (23) and for $\frac{dX_1}{d\mathcal{T}}$ and $\frac{dX_2}{d\mathcal{T}}$ as derived in expressions (24), which yields:

$$e^* - \mathcal{D}_{\mathcal{T}} = \frac{(-1)\,A_2\,(e^*)^2}{\left(\frac{dX_1}{dA_1} - \frac{dX_2}{dA_1}\right)(-\Psi_{Ee})\,C^2}\left((\Psi_{Ie}\Psi_{II2})\left[\left(\Psi_{I1} + \Psi_{Ie}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right) + \left(\Psi_{IIe}\frac{\Psi_{E1}}{-\Psi_{Ee}}\right)\right] + \right.$$
$$\left.(\Psi_{IIe}\Psi_{I1})\left[\left(\Psi_{II2} + \Psi_{IIe}\frac{\Psi_{E2}}{-\Psi_{Ee}}\right) + \left(\Psi_{Ie}\frac{\Psi_{E2}}{-\Psi_{Ee}}\right)\right]\right) = (\Psi_{Ie} + \Psi_{IIe})\,\frac{(-1)\,A_2\,(e^*)^2}{\left(\frac{dX_1}{dA_1} - \frac{dX_2}{dA_1}\right)(-\Psi_{Ee})\,C} \tag{51}$$

Now define:

$$\Gamma_2^{em}(A_2) := \Psi_{Ie} + \Psi_{IIe} = A_2 - (A_1^E + A_2^E). \tag{52}$$

Observe that $\Gamma_2^{em}(A_2^{em}) = 0$. Furthermore notice that $A_2^{em} < 0$ for $w_2 < w_2^L$, that is, $\Gamma_2^{em}(A_2) > 0$ for all $A_2 \geq 0$ whenever $w_2 < w_2^L$. In order to compare A_2^{em} to the previously established critical levels of initial allocation we make the following two observations:

$$\Gamma_2^{total}(A_2^{em}) = \Psi_{II2} < 0 \qquad \text{and} \qquad \Gamma_2^{cross}(A_2^{em}) = \Psi_{I1} < 0$$

This allows to directly conclude that $A_2^{em} < A_2^{total}$ and $A_2^{em} < A_2^{cross}$.

By making use of the newly introduced Γ_2^{em} we can rewrite expression (51) as follows:

$$e^* - \mathcal{D}_{\mathcal{T}} = -\Gamma_2^{em}(A_2)\frac{A_2e^*}{\left(\frac{dX_1}{dA_1} - \frac{dX_2}{dA_1}\right)(-\Psi_{Ee})\,C} \tag{53}$$

Finally notice that expression (53) has been derived without non–negativity constraint on A_1^*. As shown in step one of the present proof, however, for $w_2 < w_2^S$ and $A_2 \in [A_2^{cross}, A_2^{lim}]$ we obtain $A_1^* = 0$ (instead of a negative value as resulting in the computations leading to expression (53)). In this case the optimality condition given by expression (13) simplifies as follows:

$$e^* - \mathcal{D}_{\mathcal{T}} = \frac{dX_2^*}{d\mathcal{T}}A_2e^* = \frac{\Psi_{I1}\Psi_{IIe}}{-\Psi_{Ee}\,C}A_2e^* = \frac{dX_2^*}{d\mathcal{T}}A_2e^* = \left(A_2 - A_1 - A_2^E\right)\frac{\Psi_{I1}\,A_2e^*}{-\Psi_{Ee}\,C} < 0 \tag{54}$$

Observe that the above inequality is satisfied, since $A_2 \geq A_1 + A_2^E$ whenever $A_2^{cross} \leq A_2 \leq A_2^{lim}$ (compare expression (34), remember that $A_1 = A_1^* = 0$ in the case considered).

We can thus establish the following results for the optimal cap on total emissions:

$$\begin{cases} e^* > D_T & \text{if} \quad A_2 < A_2^{em} \\ e^* = D_T & \text{if} \quad A_2 = A_2^{em} \\ e^* < D_T & \text{if} \quad A_2 > A_2^{em} \end{cases} \tag{55}$$

As a **third step** We finally show that the second order conditions established in lemma 2(i) are indeed satisfied for all $A_2 \in [0, A_2^{lim}]$, A_1^*. Remember we obtained for A_1^*:

$$A_1^* = \frac{\Gamma_2^{cross}}{\Gamma_2^{lim}} A_2 = \frac{\Gamma_2^{cross}}{\Gamma_2^{cross} + \Gamma_2^{total}} A_2$$

In order to verify the second order conditions established in lemma 2(i) (a), (b), and (c) we now separately analyze the following cases:

- First, observe that

$$\Psi_{IIe} \frac{\Psi_{E2}}{-\Psi_{Ee}} = (A_2 - A_1 - \Psi_{E2}) \frac{\Psi_{E2}}{-\Psi_{Ee}} = \left(\frac{\Gamma_2^{total}}{\Gamma_2^{lim}} A_2 - \Psi_{E2} \right) \frac{\Psi_{E2}}{-\Psi_{Ee}} \tag{56}$$

  - For $w_2 < w_2^E$ (i.e. $\Psi_{E2} < 0$) expression (56) is negative, since $\Gamma_2^{total} < 0$ and $\Gamma_2^{lim} < 0$ if $A_2 < A_2^{lim}$ and $w_2 < w_2^E$ (compare appendix C).
  - For $w_2 > w_2^E$ (i.e. $\Psi_{E2} > 0$) and $A_2^{total} \leq A_2 \leq A_2^{lim}$ expression (56) is negative, since $\Gamma_2^{total} \geq 0$ and $\Gamma_2^{lim} < 0$.

  Whenever expression(56) is negative this directly implies that condition (b) is satisfied. Since furthermore $A_2 < A_2^{lim}$ also conditions (a) and (c) are satisfied.

- Second, for $A_2^{cross} \leq A_2 \leq A_2^{lim}$ we obtain $A_1^* = 0$ (compare step one of the present proof). We thus directly obtain:

$$\Psi_{Ie} \frac{\Psi_{E1}}{-\Psi_{Ee}} = -A_1^E \frac{\Psi_{E1}}{-\Psi_{Ee}} < 0 \tag{57}$$

  This directly implies that condition (a) is satisfied. Since furthermore $A_2 < A_2^{lim}$ also conditions (b) and (c) are satisfied.

- Third, for $w_2 > w_2^E$ (i.e. $\Psi_{E2} > 0$) and $0 \leq A_2 \leq min(A_2^{total}, A_2^{cross})$

  - Whenever $A_2 < A_1 + A_2^E$ (i.e. $\Psi_{IIe} < 0$), we directly obtain $\Psi_{IIe} \frac{\Psi_{E2}}{-\Psi_{Ee}} < 0$. This directly implies that condition (b) is satisfied. Since furthermore $A_2 < A_2^{lim}$, also conditions (a) and (c) are satisfied.
  - Whenever $A_2 \geq A_1 + A_2^E$ (i.e. $\Psi_{IIe} \geq 0$), then $\Psi_{IIe} \frac{\Psi_{E1}}{-\Psi_{Ee}} > 0$. Since $\Gamma_2^{cross} < 0$ in the region considered this directly implies that condition (a) is satisfied. Since furthermore $A_2 < A_2^{lim}$ also condition (b) is satisfied. Finally, since conditions (a) and (b) are satisfied and $\Gamma_2^{cross} < 0$ and $\Gamma_2^{total} < 0$, also condition (c) is satisfied.

We can thus conclude that the second order conditions as established in lemma 2(i) are satisfied for all $0 \leq A_2 \leq A_2^{lim}$.