Carbon Pricing and Innovation: The Impact of the European Carbon Trading System

Markus Trunschke*

November 2023

- Preliminary draft -

Abstract

Pricing carbon emissions increases firms' incentives to develop innovations aimed at reducing their productions' carbon emission intensities. This paper incorporates this mechanism in a dynamic discrete choice model of firms' innovation decisions while differentiating between emission-reducing and non-emission-reducing innovations. I apply the model to the European Union's Emission Trading System and estimate its parameters using administrative carbon emission data and patent information for a large set of German manufacturing firms between 2008-2017. I find that emission-reducing innovations decrease a firm's carbon emission intensity on average by about 13.7% while simultaneously decreasing its productivity by 1.5%. In contrast, non-emission-reducing innovations increase productivity by 2.2%. Furthermore, startup costs of emission-reducing innovations are lower than those of non-emission-reducing innovations. However, the costs of maintaining emission-reducing innovation activities are substantially higher than maintaining the development of non-emission-reducing innovation. Simulating counterfactual emission price changes substantially impacts emission-reducing innovation activity while non-emission-reducing innovations stays stable.

Keywords: Environmental innovations, Porter hypothesis, dynamic structural model, patents, carbon emissions; carbon pricing

*ZEW–Leibniz Centre for European Economic Research Mannheim, MaCCI, KU Leuven. markus.trunschke@zew.de

JEL Classification: Q55, Q52, O3, L50

1 Introduction

Environmental regulations combating global climate change are a top priority of policymakers and will likely remain so for the next decades. The European Union implemented with the Emission Trading System (EU-ETS) one of the most prominent environmental regulations in recent years. It represents the first and one of the largest carbon market programs worldwide, with a coverage of about 1.3 billion tons of carbon emissions[1]. Introduced in 2005, it aims at decreasing the EU's carbon emissions by at least 55% compared to 1990 by 2030 as an intermediate goal and achieving net-zero carbon emissions by 2050. Two competing arguments are most prevalent when implementing policies and regulations to protect the environment. Critiques argue that, at least in the short term, economic agents bear a new cost through the regulation by being penalized with, e.g., fines for behavior negatively impacting the environment and being forced to costly adjust their processes to comply with the regulations. This might negatively affect the regulated economy by increasing consumer prices and reducing firm competitiveness and economic growth, especially in an international context. However, Porter first argued in 1991 that such regulations increase incentives for firms to invest in solutions reducing the penalized behavior, such as innovations that decrease the environmental impact of firms' production processes and products. Porter and Linde (1995) continue to argue in the strong version of their hypothesis that these innovations do not just have the potential to decrease environmental harm while lowering a firm's compliance costs but increase the productivity of innovating firms, at least compensating for the induced short-term cost.

Despite the EU-ETS's importance and the fact that Porter's hypothesis played a strong role in its implementation (see European Commission (2007, 2011)), comprehensive empirical evidence of its innovation-inducing and productivity-enhancing effects remains scarce. Existing studies in the context of the EU-ETS either only address parts of the Porter Hypothesis, solely focusing on its impact on innovations or productivity, or leave out essential mechanisms such as the emission price itself. However, when evaluating and designing efficient and cost-effective tools for carbon emission reductions, it is essential for both economists and policymakers to understand and quantify all key mechanisms through which emission trading systems like the EU-ETS impact firm performance and innovation activity.

This paper aims to fill this gap by analyzing how the EU-ETS influences the innovation activities of regulated firms and how these innovation activities affect their carbon emission intensity and productivity. To achieve this, I develop a structural dynamic programming model that describes how the EU-ETS impacts the innovation choices of regulated firms. I explicitly model a firm's innovation decisions while differentiating between technological innovations reducing carbon emissions and non-carbon emission-reducing technological innovations to be able to analyze not just an impact on emission-reducing technology development but also a potential substitution between both technology types. I estimate the

[1] Source: European Transfer Log, Own Calculations.

short-run impact of both innovation types on developing firms' carbon emission intensity and productivity using detailed information on German firms'. Furthermore, I estimate each technology type's development costs and long-run benefits. Using these estimates, the model allows me to evaluate how different carbon prices affect innovation activities of both types and how they impact the total carbon emissions of regulated firms by simulating counterfactual situations.

I estimate the model's parameters using a combination of three data sources. First, I obtain yearly carbon emissions and allowance allocations for each firm in the EU-ETS from a publicly available administrative data set. Second, I merge this data financial and balance sheet information of German firms from Bureau van Dijk's Orbis database. Finally, I retrieve and merge information on firms' innovation activities from patent application data provided by PATSTAT, a worldwide patent database. In total, 1,336 German firms regulated under the EU-ETS are present in all three data sets, which represents coverage of over 93%.

The results show that emission-reducing innovations decrease carbon emissions on average by about 13.7% while simultaneously decreasing the firm's productivity by 1.5%. Therefore, firms would not develop these innovations if their carbon emissions were not priced in through the EU-ETS, as the entire benefits of the innovations stem from reducing their emission intensity while reducing their productivity. This contrasts the strong version of Porter's hypothesis. Non-emission-reducing innovations increase productivity by 2.2% on average. Development costs for both types of innovations differ substantially. Innovation cost distributions for firms starting to innovate have substantially higher averages than those for experienced firms. Moreover, the innovation cost distribution for firms starting to develop non-emission-reducing innovations has a substantially lower average than for emission-reducing innovations. However, the innovation cost distribution for continuing to develop emission-reducing innovations has a lower average than its counterpart for non-emission-reducing innovations. Simulating an emission price of 101€increases average emission-reducing innovation activity in the sample by about 8.78% while leaving non-emission-reducing innovation decisions stable.

This paper contributes to at least three strands of literature. First, it adds to the literature examining the impact of environmental regulation on firm behavior (Becker and Henderson 2000; Martin et al. 2014; Fowlie et al. 2016) by not just focusing on the impact of introducing a regulation on the target quantities such as emission levels but instead building a comprehensive framework explicitly incorporating the impact on firms' innovation behavior which dynamically affects target quantities. Second, I contribute to the literature evaluating the impact of the EU-ETS on the targeted emission levels and other firm outcomes (Martin et al. 2016; Colmer et al. 2022; Calel and Dechezleprêtre 2016). Most of the studies in this literature solely focus on the immediate, direct impact of the EU-ETS disregarding any intertemporal dynamics caused by firms' innovation activities, or do not incorporate the impact of the carbon emission price itself. In contrast, I explicitly incorporate the carbon price in my model through which the EU-ETS can incentivize firms to develop emission-reducing innovations, which in turn dynamically affect the firm's emission intensity and productivity. This methodology allows me to analyze multiple outcomes in one framework and to simulate the impact of emission price changes on firms' direction of innovation, affecting the effectiveness of the EU-ETS. Third, I extend the literature on structural modeling of firm innovation behavior on a microeconomic level, drawing on recent work from Aw et al. (2011); Peters et al. (2017); Maican et al. (2022); Peters et al. (2022), and Peters and Trunschke (2022) by building a model explicitly tailored to incorporate an environmental policy focusing on fostering emission reductions through technological change. This model allows firms to make two independent innovation decisions with separate benefits and development costs while incorporating an environmental regulation that incentivizes the development of one innovation type.

Section 2 of this paper gives an overview of the EU-ETS. Section 3 outlines the model, and 4 explains my approach to estimate the model's parameters. Section 5 presents the data used in the estimation and 6 presents the results. I conduct a counterfactual simulation of an emission price increase in 7 and the 8^{th} and final section concludes.

2 European Union's Emission Trading System and the Economy

In an effort to comply with the Kyoto Protocol, the European Union introduced the European Union Emission Trading System (EU-ETS) in 2005. It implements an EU-wide market-based mechanism aimed at reducing carbon emissions in the European Economy. The EU-ETS covered about 11,000 industrial installations from 7,804 companies in 2020. The number of regulated companies increased over time as new sectors were added to the EU-ETS in each new phase. This represents about 40% of the EU's total carbon emissions ($\approx$ 1.3 billion tons in 2020)[2]. It was the first major carbon market worldwide and is currently the second-largest worldwide in terms of regulated yearly carbon emissions[3].

The EU has chosen a cap-and-trade approach, in which a total amount of carbon emission allowances is allocated to the participating countries each year. One allowance gives the holder the right to emit one ton of CO_2 equivalents. The cap is reduced each year by a predefined reduction factor which was at 1.74% in phase 3 and 2.2% in the fourth phase. Allowances are distributed to all regulated installations either for free or in country-wide auctions and 1,571,583,007 total allowances were initially issued in 2021. EU-ETS-regulated Firms emitting less than their initially allocated allowances can then sell their surplus via carbon allowance exchanges to firms emitting more than their initial allowances permit. At the end of each year, firms need to provide at least as many allowances as tons of carbon emissions emitted. For each missing allowance, a firm had to pay a penalty of

[2]Source: European Transfer Log, Own Calculations. The EU-ETS does not cover relatively small industrial installations and only incorporates specified sectors.

[3]The only larger carbon trading market is currently the carbon trading system in China introduced in 2021.

Figure 1: Carbon Price over Time

Notes: The figure shows the daily average of the continuous settlement price of emission allowance futures (EUA) traded over the Intercontinental Exchange (ICE).

40€ in phase one and 100€ since the second phase while having to submit the missing allowances in the following year (Council of European Union 2003).

First implemented in 2005 with a three-year trial phase, the EU-ETS entered its fourth phase at the beginning of 2021. All emission allowances were allocated free of charge to the regulated installations during the first period. At the beginning of the second phase in 2008, the EU revised the EU-ETS, corrected an oversupply of emission allowances during the first phase, and began to gradually phase out free allowance allocation, replacing it with allowance auctions. The aviation sector was included as well, and Norway, Iceland, and Lichtenstein joined the EU-ETS. In the third phase, which started in 2013, the EU replaced national carbon caps with an EU-wide one and further restricted the free allocation of allowances.

Figure 1 below plots the price of a carbon emission allowance since the introduction of the EU-ETS. The price of allowances traded over allowance exchanges started in 2005 at about 5€ but quickly rose to 20-30€ over the next months, where it remained until the beginning of 2006. When it was announced that verified emissions in 2005 were lower than the total number of allowances, the price suddenly dropped to about 15€ before slowly declining towards zero. After correcting the allowance allocation in 2008, the initial price of 20€ steadily declined until 2013 to about 5€. Since then, the price has increased first slowly before strongly increasing from 2017 onwards to its current level of about 65€.

The way the emission trading system is designed it exhibits incentives for firms to reduce their emissions no matter their current emission levels. This is because the marginal benefit of avoiding emitting a ton of CO_2 does not depend on the the firm's emission level but on the allowance price as it can sell every unused allowance on the secondary market. Even if a firm emits less tons of CO_2 than allowances it received in the beginning of the year it can still profit from reducing its emissions further and sell the additional allowances. In line with this more than half of observations in the sample who innovate in emission-reducing technologies had excess allowances in the in the years leading to their patent application and the correlation between having excess allocations and environmental innovation activity is significantly positive.

3 Model

This paper aims to answer the question of how the EU-ETS affects firms' incentive to innovate, their emission intensity development, and productivity evolution. To answer this question, I develop a dynamic discrete choice model that includes all relevant mechanisms through which the EU-ETS affects firms' innovation activities. The primary idea is that pricing a firm's carbon emissions internalizes negative externalities to the extent that firms must consider the cost of their carbon emissions in their profit maximization decision while developing emission-reducing innovations allows them to improve their emission intensity in subsequent periods. The model consists mainly of two parts - the static profit maximization decision and the future-oriented innovation development decisions.

In the first part, firms make their production decisions given their current level of revenue productivity, carbon emission intensity, and carbon emission price. I build this part on a production function framework with carbon emissions as a by-product of production caused by a subset of production inputs. If the emitted carbon is not priced, firms would not include it in their profit maximization (or cost minimization) decision. However, if emitting carbon is costly, firms include these additional costs in their production decision. Fernández et al. (2002), Førsund (2009), Kumbhakar and Tsionas (2016), and Murty and Russell (2020) propose this multi-equation setup because it possesses acceptable theoretical properties as opposed to the commonly applied approach of modeling pollution as an input or output of a single production function[4].

In the second part, I follow previous work from Aw et al. (2011), Peters et al. (2017), and Doraszelski and Jaumandreu (2013) and model both revenue productivity and emission intensity to develop endogenously, affected by the firms' innovation development decisions. However, I extend these models by allowing firms to choose to develop either emission-reducing innovations or non-emission-reducing innovations, which differ in the way they influence the firm's future state. Both types of innovations can impact the developing firm's future revenue productivity. Emission-reducing innovations additionally influence the firm's subsequent period's emission intensity. Furthermore, both of these effects can

[4] Murty and Russell (2020) argue that treating emissions as an additional regular input in a production function for only the desired output, such as in the early work of Baumol (1988), leads to unrealistic implications. For example, holding output fixed, increasing any (emission-generating) input leads to decreased emissions. Modeling emissions as an output in a production function would lead to similarly problematic implications. E.g., under standard free disposability assumptions, firms would be able to decrease the level of their emission output without decreasing the desired output or inputs, as Murty and Russell (2020) argue.

carry over to some extent into future periods because the development processes of both variables are allowed to depend on previous periods' values. The data does not include innovation costs of either technology. I, therefore, model them as random variables similar to Aw et al. (2011) or Peters et al. (2017). Combining all model parts, I can further define simple decision rules for a firm's innovation development decisions, which allow me to (i) calculate the long-run benefits of each innovation type and (ii) simulate firm behavior in counterfactual situations such as increases in the carbon-emission price.

Static Part

This part describes the static profit maximization decision of the firm under a technology that generates carbon emissions as a by-product. It forms the grounds for the dynamic innovation decisions in the following section. The basis of this model is the production function with which firm i produces an output q_{it} in period t

$$q_{it} = F(X_{it}^N, X_{it}^E, \psi_{it}, \beta). \tag{1}$$

Within their technological constraints, firms can choose to use non-emission-producing inputs X_{it}^N such as capital or labor, or inputs that generate carbon emissions X_{it}^E such as material or fuel. The parameter vector β contains both output and substitution elasticities of their production technology. Production efficiency ψ_{it} describes how efficiently the firm uses its production inputs and is only observed by the firm before making its production decision but not by the econometrician. The amount of Carbon emitted as a by-product of the production process is a function of Carbon emission-generating inputs X_{it}^E and the firm's emission intensity α_{it}

$$E_{it} = G(X_{it}^E, \alpha_{it}). \tag{2}$$

This directly links the amount of the carbon-emitting inputs chosen by the firm and emissions generated as a by-product of production. The higher a firm's emission intensity, the more Carbon it emits during production when using emission-generating production inputs. Relating the amount of emissions to the production technology's emission-generating inputs allows to solely capture the direct link between the inputs emitting Carbon in the production process. If instead emissions would be a function of production output, the firm would be able to affect its emission intensity by increasing non-emission-generating inputs which, ceteris paribus, increase output without increasing the amount of emissions. Then estimated effects on the firm's emission intensity would not only capture technological advancements but also substitution effects within the same production technology. Furthermore, unlike the commonly employed approach of including emissions as an input in the production function, the approach in this paper does not allow firms to choose an emission level similar to a conventional input. As explained in Førsund (2009), the input approach generates a negative connection between the amount of output produced and emissions and allows firms to produce output without emitting any carbon. Modeling Carbon emissions in a separate function depending on emission-generating inputs, however, implies that firms emit carbon as long as the emission-generating input is used in production. It also creates a positive relation between the amount of Carbon emitted and the level of output as Murty and Russell (2020) explain.

I assume firms to be in a monopolistically competitive environment à la Dixit and Stiglitz (1977) as in Aw et al. (2011) or Peters et al. (2017). This setting provides the most flexible demand specification without the need to model firms' strategic interactions. Consumers' utility maximization leads to demand for a firm's output q_{it} to be given by

$$q_{it} = \left(\frac{p_{it}}{P_{jt}}\right)^{\eta_{jt}} \frac{I_{jt}}{P_{jt}} e^{\phi_{it}} = \Phi_{jt} p_{it}^{\eta_{jt}} e^{\phi_{it}}, \tag{3}$$

where p_{it} is the price that firm i asks for its output. P_{jt} represents a price index of all product variants in firm i's market j while I_{jt} is the size of the market. The price elasticity of demand in the market η_j, which I assume to be constant over time, explains how strongly demand reacts to price changes. The demand shock ϕ_{it} shifts the demand for firm i's product in period t and can be interpreted as the desirability or quality of the firm's product. Similar to the production efficiency ψ_{it}, it is known to the firm when making the current period's decisions while being unobserved to the econometrician. Assuming output markets to be in equilibrium, I can express a firm's revenue as

$$\begin{aligned} R_{it} = p_{it} \cdot q_{it} &= \left(\frac{P_{jt}^{1+\eta_{jt}}}{e^{\phi_{it}} I_{jt}}\right)^{\frac{1}{\eta_{jt}}} q_{it}^{\frac{1}{\eta_{jt}}} q_{it} \\ &= \left(\frac{P_{jt}^{1+\eta_{jt}}}{e^{\phi_{it}} I_{jt}}\right)^{\frac{1}{\eta_{jt}}} \left[F(X_{it}^N, X_{it}^E, \psi_{it}, \beta)\right]^{\frac{1+\eta_{jt}}{\eta_{jt}}} \\ &= \left(\frac{P_{jt}^{1+\eta_{jt}}}{I_{jt}}\right)^{\frac{1}{\eta_{jt}}} \left[F(X_{it}^N, X_{it}^E, \omega_{it}, \beta)\right]^{\frac{1+\eta_{jt}}{\eta_{jt}}} . \end{aligned} \tag{4}$$

I follow Peters et al. (2017) and combine both unobserved production efficiency ψ_{it} and the unobserved demand shock ϕ_{it} to revenue productivity ω_{it} because I cannot disentangle these two quantities with the data at hand in the empirical approach. It also resembles most closely the commonly estimated quantities in empirical applications of production function estimations in the literature using revenue data.[5]

Using all parts from above, I can express a firms profit π as

$$\begin{aligned} \pi_{it} &= p_{it} q_{it} - c(X_{it}^N, X_{it}^E, w_{it}^N, w_{it}^E, p_{it}^E, \alpha_{it}) \\ &= R_{it} - w_{it}^N X_{it}^N - w_{it}^E X_{it}^E - p_{it}^E E_{it}, \end{aligned} \tag{5}$$

with w_{it}^N and w_{it}^E representing prices for non-emission-generating inputs and emission generating inputs, respectively. Regulated firms have a strictly positive Carbon emission price p_{it}^E. If firms' emissions are not regulated, the Carbon emission price would essentially equal

[5] See De Ridder et al. (2022) for a discussion about estimating production functions with revenue data.

zero, and the amount of Carbon emissions they generate as a by-product of production would drop out of the profit function. However, the larger the Carbon emission price is, the larger the trade-off between producing the desired output with emission-generating production inputs and paying the cost of emitted Carbon.

Dynamic Part

This part focuses on the development processes of both revenue productivity ω_{it} and emission intensity α_{it} and the dynamic innovation decisions using the above-developed static profit maximization of the firm. Firm productivity evolves dynamically with some amount of persistency. However, differently than in a substantial part of the empirical production function literature, it can be influenced by the firm's decisions to develop emission-reducing innovations i^E_{it-1} or non-emission-reducing innovations i^N_{it-1}. Therefore,

$$\omega_{it} = \Upsilon^{\omega}(\omega_{it-1}, i^E_{it-1}, i^N_{it-1}). \tag{6}$$

Even though firms primarily develop emission-reducing innovations to reduce their carbon emission intensity, they are also likely to impact a firm's general productivity. However, the sign of this impact is not clear ex-ante. Porter (1991) argues innovations incentivized by environmental regulations can also increase productivity because they represent a new technology that might make the firm's production more efficient or increase its products' quality. Even in the case of a positive productivity impact of emission-reducing innovations, firms might not necessarily develop these innovations without environmental regulations since they either face budget or capacity constraints in dimensions relevant to innovation decisions or do not have perfect information about every possible technological direction. In contrast, when environmental regulations are introduced, firms will develop emissions-reducing innovations even if they have a negative effect on productivity, as long as the increase in profits from the reduction in emission costs is larger than the profit loss due to reduced productivity.

A firm's emission intensity α_{it} develops in a similar fashion as productivity. It evolves over time as a persistent process. The firm can influence this evolution by developing emission-reducing innovations, however, not by developing non-emission-reducing innovations, i.e.

$$\alpha_{it} = \Upsilon^{\alpha}(\alpha_{it-1}, i^E_{it-1}). \tag{7}$$

Any reduction of emission levels at constant output when non-emission-reducing innovations were developed, therefore, does not come from a reduction of the emission intensity but only stems indirectly from increased productivity.

Developing innovations are costly, and a firm's decision to engage in any possible innovation development activity not only depends on the expected potential benefits of the innovation but also on their respective development costs. These costs are likely to differ substantially between different technologies. For example, Peters and Trunschke (2022) show this for different technology types. I follow this idea and allow development costs in this model to differ between emission-reducing and non-emission-reducing innovations. Since I do not observe the innovation costs of any of the two innovation types (C_{it}^E, C_{it}^N) in my data, I model them as random variables drawn from distributions Λ^E and Λ^N with parameters Θ^E and Θ^N, respectively, i.e.

$$\begin{aligned} C_{it}^E &\sim \Lambda^E(\Theta^E; i_{it-1}^E), \\ C_{it}^N &\sim \Lambda^N(\Theta^N; i_{it-1}^N). \end{aligned} \tag{8}$$

Second, innovation development costs do not just differ between technology types but also depend on the experience of the developer. If a firm is a first-time developer in the respective technology, its development costs are likely to be substantially higher than for experienced developers. This difference arises not just because of the fixed costs of building up an innovation department but also because of the inexperience and inefficiencies of the firm in developing the technology. The model, therefore, allows the moments of the innovation development cost distributions to vary for firms starting or continuing developing innovations.

Combining all pieces of the model, the firm's intertemporal maximization problem can then be expressed in the following Bellman equation

$$\begin{aligned} V_{it} &= \pi^*(s_{it}) + \delta \max_{i^E, i^N \in \{0,1\}} \mathbb{E}\left[V(s_{it+1}, i_{it}^E, i_{it}^N) - C_{it}^E i_{it}^E - C_{it}^N i_{it}^N\right] \\ &= \pi^*(s_{it}) + \delta \max_{i^E, i^N \in \{0,1\}} \mathbb{E}\left[\int\limits_{C^E}\int\limits_{C^N} \Big(\left[V(s_{it+1}|\omega_{it}, \alpha_{it}; i_{it}^E, i_{it}^N)\right] - C_{it}^E i_{it}^E - C_{it}^N i_{it}^N\Big) dC^N dC^E\right]. \end{aligned} \tag{9}$$

The first part is the contemporary profit after the firm made its production decision to maximize its profit given its current state variables $s_{it} = (\omega_{it}, \alpha_{it}, k_{it}, i_{it-1}^E, i_{it-1}^N)$. The second part contains the maximum of the firm's expected future value, which depends on its current innovation choices (i_{it}^E, i_{it}^N) and is discounted by the discount factor δ, net the associated development costs if the firm decides to develop any of the innovation types. Because the innovation costs are random variables, their expectation can be expressed as the integral over all their possible values. However, the expected value of the firm itself depends on the future values of revenue productivity ω_{it} and emission intensity α_{it} which are influenced by the firm's innovation decisions. I assume the firm to make innovation decisions for both types sequentially. First, firms make the emission-reducing innovation decision i_{it}^E and afterwards the non-emission-reducing innovation decision i_{it}^N in the same period. Equation (9) can then be expressed in terms of the firm's emission-reducing innovation decision at the beginning of the period after observing its emission-reducing innovation development costs as

$$\mathbb{E}\left[V(s_{it+1}|\omega_{it}, \alpha_{it}; i_{it}^E, i_{it}^N)\right] = \int\limits_{\omega}\int\limits_{\alpha} \begin{Bmatrix} V(s_{it+1}|\omega_{it}, \alpha_{it}; i_{it}^E = 1) - C_{it}^E, \\ V(s_{it+1}|\omega_{it}, \alpha_{it}; i_{it}^E = 0) \end{Bmatrix} dF(\alpha_{it+1}|\alpha_{it}, i_{it}^E) dG(\omega_{it+1}|\omega_{it}, i_{it}^N, i_{it}^E) \tag{10}$$

However, the future value in the interim value function in (10) still depends on the second choice the firm makes in the same period and can, therefore, be rewritten in terms of the non-emission-reducing innovation choice after the firm observed C_{it}^N as

$$V(s_{it+1}|\omega_{it},\alpha_{it};i_{it}^E) = \left\{ \begin{array}{l} V(s_{it+1}|\omega_{it},\alpha_{it};i_{it}^N=1) - C_{it}^N, \\ V(s_{it+1}|\omega_{it},\alpha_{it};i_{it}^N=0) \end{array} \right\} \tag{11}$$

Firms choose to invest in either innovation if the benefits of innovating in the specific innvoation type is larger than the corresponding innovation cost. The marginal benefit of each technology is the discounted difference between the expected future value of the firm if it decides to innovate in the respective technology and if it does not

$$\begin{aligned} \Delta_E \delta E[V(s_{it+1})] &= E[V(s_{it+1}|\omega_{it},\alpha_{it};i_{it}^E=1)] - E[V(s_{it+1}|\omega_{it},\alpha_{it};i_{it}^E=0)], \\ \Delta_N \delta E[V(s_{it+1})] &= E[V(s_{it+1}|\omega_{it},\alpha_{it};i_{it}^N=1)] - E[V(s_{it+1}|\omega_{it},\alpha_{it};i_{it}^N=0)]. \end{aligned} \tag{12}$$

Without knowing the firm's contemporary innovation costs, the ex-ante conditional innovation choice probabilities for emission-reducing and non-emission-reducing innovations can, therefore, be expressed as the probability that the marginal benefit of developing the innovation is larger than its innovation costs.

$$\begin{aligned} &P(i_{it}^E=1|s_{it}) = P\left(\Delta_E \delta E[V(s_{it+1})] \geq C_{it}^E\right) \quad \text{and} \\ &P(i_{it}^N=1|s_{it}) = P\left(\Delta_N \delta E[V(s_{it+1})] \geq C_{it}^N\right). \end{aligned} \tag{13}$$

4 Empirical Approach

The approach to estimate all primitives of the model consists mainly of two parts. First, I estimate all parameters influencing the firm's short-run profit maximization decision and the parameters governing the development processes of both revenue productivity and emission intensity. I then use these estimates to estimate all (expected) value functions and the parameters of the development cost distributions in the second part.

Static Part

The estimation of all short-run parameters has three steps. Step one identifies the parameters of the emission generation function and the emission intensity development process. Step two then estimates the demand elasticities for each industry, followed by the simultaneous estimation of the revenue function and the productivity development process in the third step. A key challenge in this last step is identifying unobserved productivity and its development process.

I begin with calculating the firm's emission intensity α_{it} and estimating the parameters of the emission generation function (2) and the parameters of the emission intensity development process (7). Assuming material to be the only emission-generating input and its relationship with emissions to be linear. Therefore, the emission generating function (2) takes the form

$$E_{it} = M_{it} e^{\alpha_{it}}. \tag{14}$$

Rearranging terms leads to a simple expression for emission intensity

$$\alpha_{it} = ln \left(\frac{E_{it}}{M_{it}} \right). \tag{15}$$

Assuming a firm's emission intensity to develop following a controlled Markov process that is linear in the previous period's emission intensity and the emission-reducing innovation decision. Adding an i.i.d. zero-mean error term κ_{it} leads to the estimation equation, which I can estimate using OLS

$$\alpha_{it} = \gamma_0 + \gamma_1 \alpha_{it-1} + \gamma_2 i^E_{it-1} + \kappa_{it}. \tag{16}$$

Following insights from Peters et al. (2017), who show that profit-maximizing firms in a monopolistic competition environment set their output price as $p_{it} = \left(\frac{\eta_{jt}}{1+\eta_{jt}}\right) \cdot MC_{it}$, with MC_{it} representing marginal costs, short-run profits can be expressed as

$$\pi_{it} = R_{it} - MC_{it} q_{it} = -\frac{1}{\eta_{jt}} R_{it}. \tag{17}$$

Rearranging terms leads to

$$\frac{MC_{it} q_{it}}{R_{it}} = \frac{w^N_{it} X^N_{it} - w^E_{it} X^E_{it} - p^E_{it} E_{it}}{R_{it}} = 1 + \frac{1}{\eta_{jt}}. \tag{18}$$

I can, therefore, regress the variable cost-to-revenue ratio onto a constant for each industry separately using OLS, which allows me to back out demand elasticity estimates $\hat{\eta}_{jt}$.

The remaining step is the estimation of the revenue function- and the productivity development process parameters. For simplicity, I assume the firm's production technology to be of Cobb-Douglas type that is Leontief in material input. However, the setup can easily allow for other, more general, production functions

$$F(K_{it}, L_{it}, \omega_{it}; \beta) = q_{it} = K^{\beta_k}_{it} L^{\beta_l}_{it} e^{\omega_{it} + \epsilon_{it}}, \tag{19}$$

Including (19) in the revenue equation (4) and taking the natural logarithm leads to the basic form of the estimation equation

$$r_{it} = \left(\frac{1}{\eta_{jt}}\right) \lambda_{jt} + \left(\frac{1+\eta_{jt}}{\eta_{jt}}\right) (\beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \epsilon_{it}), \tag{20}$$

where lower-case variables represent the natural logarithms of their respective capital-letter counterparts. I include the estimated demand elasticities as data and a set of industry- and time dummies λ_{jt} which subsumes all industry-level variables and variation over time.

As in most applications, I do not observe revenue productivity in the data. Trying to estimate the parameters of the revenue function (20) without accounting for ω_{it} leads to a substantial, well-known simultaneity bias as explained in Olley and Pakes (1996). Following Ackerberg et al. (2015), I employ a two-step control function approach, proxying for ω_{it} in the first step, which allows me to find an expression that includes ω_{it}. The second step then identifies all output elasticities. I can express material demand as a function of all other production inputs, revenue productivity, and emission intensity[6]

$$m_{it} = h_m(k_{it}, l_{it}, \omega_{it}, \alpha_{it}). \tag{21}$$

Assuming that ω_{it} is the only unobserved factor in the material demand equation, I can invert (21) such that productivity becomes a function of only observed variables

$$\omega_{it} = h_m^{-1}(k_{it}, l_{it}, m_{it}, \alpha_{it}), \tag{22}$$

which I substitute in equation (20) for ω_{it}

$$r_{it} = \left(\frac{1}{\eta_{jt}}\right)\lambda_{jt} + \left(\frac{1+\eta_{jt}}{\eta_{jt}}\right)\left(\beta_k k_{it} + \beta_l l_{it} + h^{-1}(k_{it}, l_{it}, \omega_{it}, \alpha_{it}) + \epsilon_{it}\right). \tag{23}$$

Instead of using my model's structure to find a closed form of this equation as in Peters et al. (2017), I follow Ackerberg et al. (2015) and approximate (23) using a 4^{th} order Taylor approximation and estimate its parameters using OLS. As explained above, this step does not identify any of the structural parameters but allows me to retrieve an estimate $\hat{q}_{it}$ that includes the unobserved productivity term ω_{it} but not the i.i.d error ϵ. I can use this approximation to identify all output elasticities and the parameters of the productivity development process in the second stage. Assuming the productivity development process to be cubic in past productivity and linear in the innovation decisions, as common in the literature (Aw et al. 2011)

$$\omega_{it} = \rho_1\omega_{it-1} + \rho_2\omega_{it-1}^2 + \rho_3\omega_{it-1}^3 + \rho_4 i_{it-1}^E + \rho_5 i_{it-1}^N + \xi_{it}. \tag{24}$$

ξ_{it} represents a contemporary i.i.d. zero mean productivity shock. Based on this, I can formulate all necessary moments to estimate all parameters using an efficient two-step GMM estimator as proposed in Hayashi (2000)

$$\left[\begin{pmatrix} k_{it} \\ l_{it-1} \\ \omega_{it-1} \\ \omega_{it-1}^2 \\ \omega_{it-1}^3 \\ i_{it-1}^E \\ i_{it-1}^N \end{pmatrix} \otimes \xi_{it}\right] = 0. \tag{25}$$

[6]For unregulated firms, emission intensity would drop out of the equation because firms would not account for it in their production decisions.

This formulation assumes that all investments into contemporary capital k_{it} were already made in the previous period, ruling out any dependence of the capital stock in the contemporary productivity shock. Contrary to this, I assume that firms can at least partially decide on their labor input in the contemporary period and, therefore, use lagged labor input to define the moment.

Dynamic Part

The only primitives of the model that are left to estimate are the innovation development cost distribution parameters θ^E, θ^N, and the (expected) value functions. I estimate these quantities using a nested fixed-point algorithm (NFXP) proposed by Rust (1987), which is based on a Likelihood function consisting of the conditional choice probabilities for each innovation type. The assumed sequential innovation decision order allows me to express the joint conditional choice probability of both types of innovations as the multiplication of each separate probability. However, note that the conditional choice probability of emission-reducing innovations is conditional on the contemporary non-emission-reducing innovation decision, while its counterpart for non-emission-reduing innovations is not.

$$\begin{aligned}\mathcal{L}(\theta|i_{it}^E, i_{it}^N, s_{it}) &= \prod_i \prod_t P(i_{it}^E, i_{it}^N|s_{it}, \theta) \\ &= \prod_i \prod_t P(i_{it}^E|i_{it}^N, s_{it}, \theta) P(i_{it}^N|s_{it}, \theta)\end{aligned} \tag{26}$$

These choice probabilities of both innovation types represent the probability that the discounted expected marginal benefit of choosing to innovate exceeds the associated innovation costs as shown in (13). These probabilities have analytical expressions when assuming both innovation costs to be drawn from exponential distributions,

$$\begin{aligned}P\left(i_{it}^E|s_{it}\right) &= 1 - exp\left(\frac{\delta V\left(i_{it}^E = 1|s_{it}\right) - \delta V\left(i_{it}^E = 0|s_{it}\right)}{\theta^E}\right), \\ P\left(i_{it}^N|s_{it}\right) &= 1 - exp\left(\frac{\delta V\left(i_{it}^N = 1|s_{it}\right) - \delta V\left(i_{it}^N = 0|s_{it}\right)}{\theta^N}\right),\end{aligned} \tag{27}$$

with θ^E and θ^N describing the means of the respective exponential distributions. Appendix A.2 derives the likelihood function in detail and provides some technical details of the computational procedure. Calculating these probabilities rely on solving the system of equations defined by the (expected) value functions (9), (10), and (11), over a state space grid of 100 equally spaced points between the minimum and maximum values for revenue productivity ω_{it}, 100 equally spaced points for emission intensity α_{it}, and the observed combinations of material, employees, capital stock, and industry classification. I match the solutions computed on the state space grid to observations in my dataset by interpolating value functions and expected value functions between the grid points with cubic B-splines.

5 Data

The analysis in the empirical part of the paper uses information from three different data sets. I take yearly information on German firms from the Orbis database, which I merge with data on carbon emissions from the European Carbon Transfer Log (EUTL). Patent data from the worldwide patent database - PATSTAT provides information on each firm's innovation behavior. The analysis concentrates on firms in manufacturing sectors mainly because (i) the EUTL focuses on these sectors, and therefore more than 95% of emissions in my sample come from firms in manufacturing, and (ii) it focuses on technical innovations using patenting information, which is predominantly done in manufacturing, as well.

Firm Data

The Orbis database from Bureau van Dijk provides financial indicators and balance sheet information from firms worldwide. It is, apart from administrative data, one of the most renowned and complete source of firm information in Germany. I take information on production inputs such as material, number and cost of employees, and fixed capital, turnover, and industry classifications from firms in manufacturing industries. This gives me a total of 9,145,934 observations from 1,270,101 firms from all industries until 2020. However, the sample restrictions described below and missing information in important variables reduce the total number of observations substantially.

Carbon Emission Data

I obtain firms' carbon emission data from the European Transfer Log (EUTL), which constitutes a publically available dataset containing each industrial installation regulated under the EU-ETS [7]. This administrative data provides names and addresses of all installations and the companies who own them. It also contains records on yearly emissions, initially allocated allowances, and all allowance transactions between installations for each installation since the second EU-ETS phase in 2007. In total, the dataset contains emission information for 25,205 observations from 2,794 industrial installations in Germany. I aggregate the data on the firm-year level and obtain yearly emission data for 1,336 firms. I am able to find above 93% of installations from the emission dataset to firms in the Orbis database using a fuzzy name and address matching algorithm on the installation owners' addresses and names[8]. Appendix A.1 provides further details on the matching quality and process. Table 1 shows that total emissions in the sample are highly concentrated in

[7] The dataset is freely available under https://climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets/union-registry_en

[8] The algorithm I use compares the similarity of strings from each candidate, weighting words by their frequency. This essentially gives low weights to fill words and legal forms while increasing the relative importance of informative words. After matching these weighted strings, I manually validate each potential match. For all firms in the emission data that were not properly matched, I conduct a manual search of the Orbis online database. See https://github.com/ThorstenDoherr/searchengine and Doherr (2023) for further information on the fuzzy name matching algorithm.

a small number of sectors. The vast majority of emissions originate from manufacturing industries, and almost 70% come solely from mining, oil processing, and energy-supplying sectors.

Table 1: Emissions per Industry

Industry	Share	Description
1	.007	Food, beverages, tobacco
2	.000	Textiles, clothing, leather
3	.013	Wood, paper
4	.052	Chemicals, pharmaceuticals
5	.001	Rubber, plastics
6	.054	Glass, ceramics, concrete
7	.120	Metals, metal products
8	.000	Electronics, med. instruments
9	.008	Machinery
10	.008	Automotive, other vehicles
11	.000	Furniture, other consumer products
12	.690	Mining, oil processing, energy supply
Total	.953	

Notes: The table presents each industry's share on total emissions over all periods in the sample. Service industries are omitted.

Innovation Data

I use information on firms' patenting activities as an innovation indicator. Though it is well known that not all innovation activities are patented, it represents the most often used indicator for innovation. Especially non-technical, less valuable, and easily hidable innovations are less likely to be patented. However, the analysis focuses on technical innovations that are likely to be covered reasonably well by patent information. An advantage of using patent data is that information on the universe of all patents is accessible in its entirety and that the novelty of the invention is externally validated by patent examiners and does not solely rely on self-reported information. Another advantage is that patents are accurately located in the technology space via technology classifications. I follow Calel and Dechezleprêtre (2016) and use this property to classify patents into carbon-reducing innovations and non-carbon-reducing innovations using the "YO2" CPC class (see Angelucci et al. (2018)) of patents for climate change mitigation. This class was created by the European Patent Office to identify patents aimed at combating climate change which is largely driven by firms' carbon emissions. I match firm-year observations from the Orbis dataset to patent applications using the same fuzzy string-matching approach of names and addresses from firms in the Orbis sample and patent applications.[9]

[9] It is impossible for me to validate all potential matches manually because of the data size. Instead, only a subset of the results is manually validated and then fed into a machine-learning algorithm for the validation of the full sample. See Appendix A.1 for more information on the matching procedure.

Analyzing the transition rates displayed in Table 2 for emission-reducing patent applications and in Table 3 for non-emission-reducing patent applications between the number of a firm's patent applications from one year to the next reveals that most firms transition between no patent applications and some patent applications. Most firms transition from having no patent application at all to one or more applications in the next year or the other way around. Only relatively few firms transition from one patent application to more than one application in the next year. Therefore, the firm's discrete decision between not conducting any innovation project and conducting one seems to be more relevant the intensive margin. I reflect this in my model by modeling the firm's innovation decision as binary.

Table 2: Emission-reducing Patenting Transition Rates

		Emission-reducing patent applications$_t$							
		0	1	2	3	4	5	>5	Total
Emission-reducing patent applications$_{t-1}$	0	207,782	703	140	57	14	6	16	208,718
	1	776	167	61	29	17	5	10	1,065
	2	165	71	39	16	10	8	11	320
	3	64	30	29	20	9	5	12	169
	4	21	12	9	14	11	7	17	91
	5	8	7	5	9	5	2	16	52
	>5	25	8	12	16	14	15	210	300
	Total	208,841	998	295	161	80	48	292	210,715

Notes: Transition rates are based on the full merged sample. The estimation sample is smaller due to its focus on manufacturing industries, regulated firms, a shorter time frame, and missing values in included variables.

Table 3: Non-mission-reducing Patenting Transition Rates

Non-emission-reducing patent applications$_{t-1}$		Non-emission-reducing patent applications$_t$							
		0	1	2	3	4	5	>5	Total
	0	196,033	2,429	778	250	113	54	50	199,707
	1	2,969	866	370	191	107	51	70	4,624
	2	915	415	238	149	80	49	120	1,966
	3	328	231	136	91	58	48	116	1,008
	4	152	122	93	80	68	39	116	670
	5	81	57	63	45	48	38	81	413
	>5	141	120	139	121	113	116	1577	2,327
	Total	200,619	4,240	1,817	927	587	395	2,130	210,715

Notes: Transition rates are based on the full merged sample. The estimation sample is smaller due to its focus on manufacturing industries, regulated firms, a shorter time frame, and missing values in included variables.

The summary statistics in table 4 show that the sample focuses mostly on larger firms. The average firm has about 1,500 employees and receives more than 1.2bn € in annual revenues. Carbon emission costs vary greatly between firms making a profit from selling their allowances to firms that pay almost half a billion Euro. On average firms pay 6.4m € for their carbon emissions, which is on average 6.4% of their annual material costs. About 9% of firms in the sample have non-emission-reducing patent applications, whereas about 5% have emission-reducing patent applications. These shares are substantially lower than survey-based measures of innovation (Rammer et al. 2022). However, patent applications only represent a subset of valuable, mostly technical innovations, which is the focus of this paper.

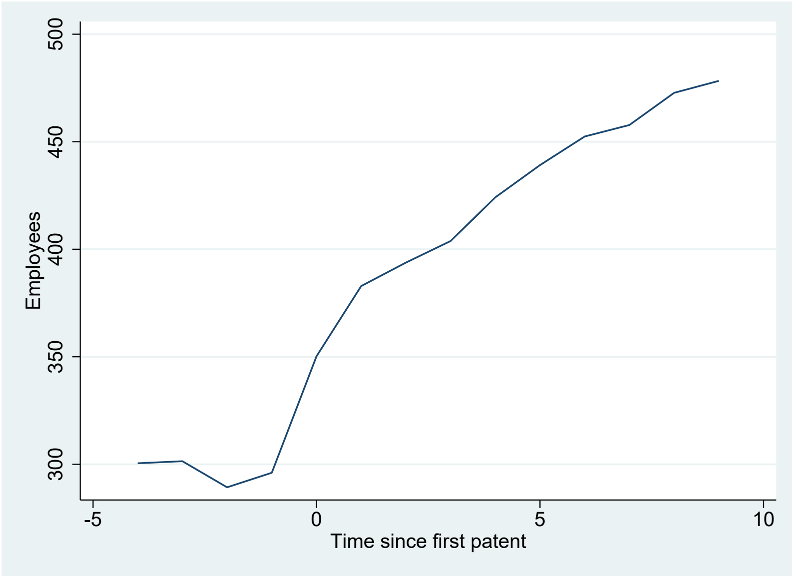
Table 4: Summary Statistics

Variable	Model	Unit	mean	med	sd	min	max
Revenues	R	mio €	1275.48	169.38	6,161.59	0.610	81,782.92
Fixed assets	K	mio €	758.93	76.56	4,990.26	0.060	110,551.19
Material cost	M	mio €	787.01	76.81	4,263.57	0.019	54,139.25
Labor cost	$p_l L$	mio €	136.38	20.90	193.10	0.006	10,500.63
Emission costs	$p_E E$	mio €	6.4	0.005	19.71	-49.198	484.63
Non-emission red. inno	i^N	0/1	0.09	0	0.28	0	1
Emission red. inno	i^E	0/1	0.05	0	0.22	0	1

Notes: Emission costs can be negative because they represent net emission costs and firms can sell their emission allowances if they do not need them.

The grating of patents generally takes several years. I, therefore, use the earliest filing date of the patent application in the analysis to get as close as possible to the true invention

Figure 2: Patent Application and Employment



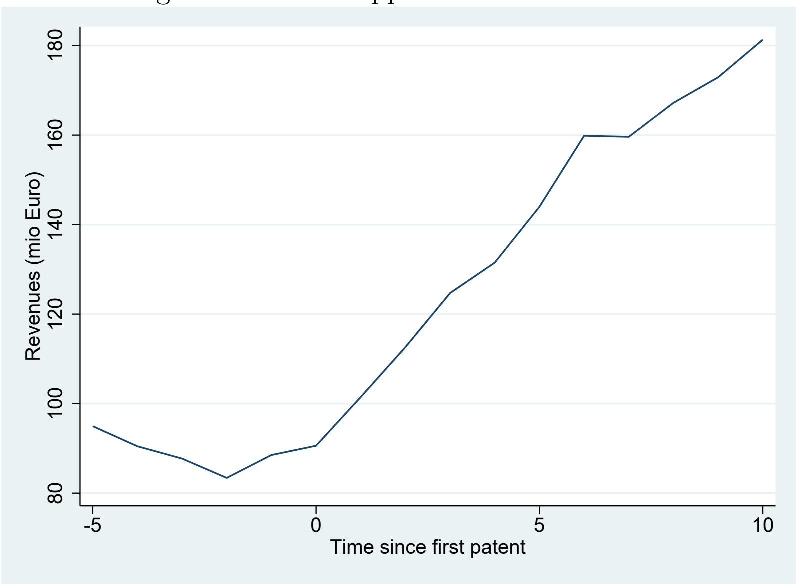
Notes: Average annual number of employees of a patenting firm up to 5 years before and 10 years after its first patent application.

date when matching patent applications with firm information. Looking at Figures 2-4 depicting the development of employment, revenues, and emission intensity[10] relative to the timing of the first (emission-reducing) patent application. The positive relationship of the first application with employment and revenue are already present in the year of the patent application while steadily increasing afterwards.[11] Similarly, the negative relationship between the first emission-reducing patent application and a firm's emission intensity seems to be directly present in initial year of the application. I additionally regress firms' employment, revenues, and emission intensity on a (emission-reducing) patent application dummy and three of its lags in separate models. The results in Tables A.2-A.4 in Appendix A.3 confirm the visual findings from Figures 2-4 as already the contemporary patent application dummy is highly significant. The relationship again seems to carry over to subsequent periods because the coefficients from lagged patent application often continue to be significantly different from zero. This evidence indicates towards an immediate relationship between patenting and firms' input and output choices without a strong time-lag. I, therefore, assume only a one year lag between the firm's innovation decisions and the related patent application and a direct impact of the patent application on emission-intensity and productivity in the estimation of the model's parameters.

[10] Emission intensity is defined as in the later part of the paper as $\frac{total\ emissions}{material\ input}$.

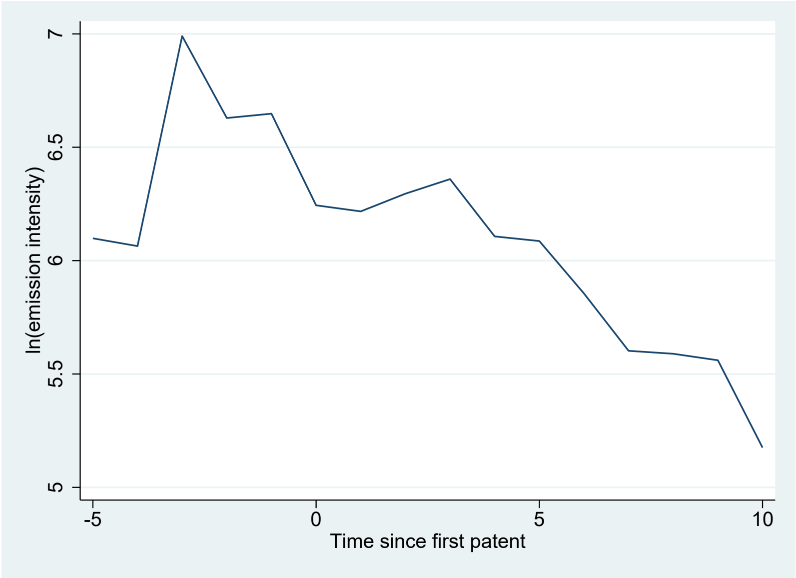
[11] Part of the continuously growth might come from subsequent patent applications after the initial one.

Figure 3: Patent Application and Revenue



Notes: Average annual revenue of a patenting firm up to 5 years before and 10 years after its first patent application.

Figure 4: Emission-reducing Patent Application and Emission Intensity



Notes: Average annual emission intensity of a firm patenting in emission-reducing technologies up to 5 years before and 10 years after its first patent application.

6 Results

I estimate all model parameters in two parts. First, I retrieve all static parameter estimates that do not involve solving the dynamic programming problem, namely the emission generation function's parameters, demand elasticities, revenue function parameters, and the parameters governing the laws of motion for emission intensities and productivity. Estimating these parameters before solving the dynamic programming problem substantially reduces the dimensionality of the subsequent estimation and all problems associated with it. Plugging the resulting estimates into the corresponding functions leaves estimating the innovation development cost distribution parameters and (expected-) value functions for the second step.

6.1 Static Part

I begin with estimating the parameters of the emission generation function. The results in Table 5 confirm that firms can improve their emission intensity through innovation. By developing emission-reducing innovations, the emission intensity of firms significantly decreases by 7.9% on average. The large and significant parameters for the lagged emission intensity show a high degree of persistency of firms' emission intensity levels over time. I include year- and industry fixed effects to test the robustness of the results in the second column. However, including firm fixed effects in the model makes it necessary to use the GMM estimator proposed by Arellano and Bond (1991). The negative impact of emission-reducing innovations persists, however, doubling in size, while the degree of persistency of the emission intensity decreases. The differences of the results in these two models might, to some degree stem from the reduced sample size in the second column due to missing values in the lagged variables that the estimator uses as instruments.

Table 5: Emission Generation Function Estimates

Variable	OLS	Areallano-Bond
α_{it-1}	0.954***	0.349***
	(0.011)	(0.117)
i^E_{it-1}	-0.079**	-0.137**
	(0.051)	(0.064)
cons	0.292***	4.270 ***
	(0.071)	(0.772)
SE($\hat{\zeta}$)	0.711	1.517
Observations	3,117	2,555

Notes: Dependent variable is α_{it}. Standard Errors are in parentheses below the point estimates. Significance at the * 10% level, ** 5% level, *** 1% level. Time- and firm dummy variables are included in the second model but not reported.

Before jointly estimating the parameters of the revenue function and the productivity development process, I estimate demand elasticities. As described in the empirical approach in section 4, rearranging the firm's profit equation (17) leads to equation (18), which I can estimate for each industry separately, allowing me to back out demand elasticities. The results in Table 6 show the expected negative sign and are comparable in size to estimates of similar models with different datasets (Peters et al. 2017; Peters and Trunschke 2022). The size of the demand elasticity affects how high firms' markups over production costs are. This, in turn, affects the marginal benefit of each innovation decision. The higher the demand elasticity, the smaller the marginal profit a firm receives per unit of revenue. This especially influences the benefits of increasing its productivity or decreasing its emission intensity. Using the results in Table 6, equation (5) shows that in my model, a marginal euro of revenues translates, e.g., for mining and oil processing into 38.3 cents additional profit while firms in food, beverage, and tobacco production only receive a 26.1 cents additional profit.

Table 6: Demand Estimates

	Industry	Demand Elasticity (η_j)
1.	Food, beverages, tobacco	-3.825
2.	Textiles, clothing, leather	-3.197
3.	Wood, paper	-3.512
4.	Chemicals, pharmaceuticals	-3.124
5.	Rubber, plastics	-3.524
6.	Glass, ceramics, concrete	-2.903
7.	metals, metal products	-3.185
8.	Electronics, instruments, electrical equipment	-3.452
9.	Machinery	-3.398
10.	Automotive, other vehicles	-3.618
11.	Furniture, other consumer products	-3.128
12.	Mining, oil processing, energy supply	-2.611

Notes: Industry demand elasticity estimates are based on a larger sample as I only require total variables costs and revenues to be non-missing and also include firms that are observed only once or with gaps.

Plugging the demand elasticity estimates into the revenue equation (20), I estimate the revenue elasticities and the productivity development process parameters using material inputs as a proxy for unobserved productivity ω_{it} in the first stage of the GMM estimator. The results in Table 7 show that the revenue elasticity of capital is with about 0.53 substantially lower than it's equivalent for labor (0.83). The parameter estimates for the lagged productivity terms in the productivity development process show a high degree of persistency. This means that contemporary effects on productivity are carried over to subsequent periods to a substantial amount. Developing an emission-reducing innovation decreases productivity in the next period by 1.5% while non-emission-reducing innovations increase productivity by 2.2%, however, non-significantly.

Table 7: Revenue Function Estimates

Variable	Coef	SE
k_{it}	0.525***	0.099
l_{it}	0.826***	0.077
ω_{it-1}	0.919***	0.026
ω_{it-1}^2	0.071***	0.021
ω_{it-1}^3	-0.020***	0.006
i_{it-1}^E	-0.015	0.026
i_{it-1}^N	0.022	0.029
SE($\hat{\xi}$)	0.786	
Observations	3104	

Notes: Significance at the * 5% level, ** 1% level, *** 0.1% level. Time- and industry dummy variables are included in the first stage of the estimator but not reported.

6.2 Dynamic Part

In the second step, I estimate the parameter of the dynamic model. I use the estimated static parameters to calculate profits for each firm type on the grid as explained in section 4. I then estimate the averages of the development cost distributions of each firm type while differentiating between unexperienced and experienced innovators respectively. Similar to previous results in the literature startup costs averages are substantially higher than continuation cost averages for both technology types. However, they are lower for emission-reducing innovations than for non-emission-reducing innovations. However, costs of maintaining innovation activities for emission-reducing innovations are, on average, double the costs for non-emission-reducing innovations.

Table 8: Innovation Cost Distribution Averages

Parameter		Point Estimate
Startup cost emis. red. inno.	(θ^{ES})	4,884.543
Maintenance cost emis. red. inno.	(θ^{EM})	672.302
Startup cost non-emis. red. inno.	(θ^{NS})	5,656.351
Maintenance cost non-emis. red. inno.	(θ^{NM})	317.057
Observations	3,104	

Notes: I calculate the fixed point of the Bellman equation using a grid of 100 points for ω, 100 points for α, 12 industries, 10 points for capital, 10 points for employees, and 10 points for material input.

7 Simulating the Impact of the Carbon Price

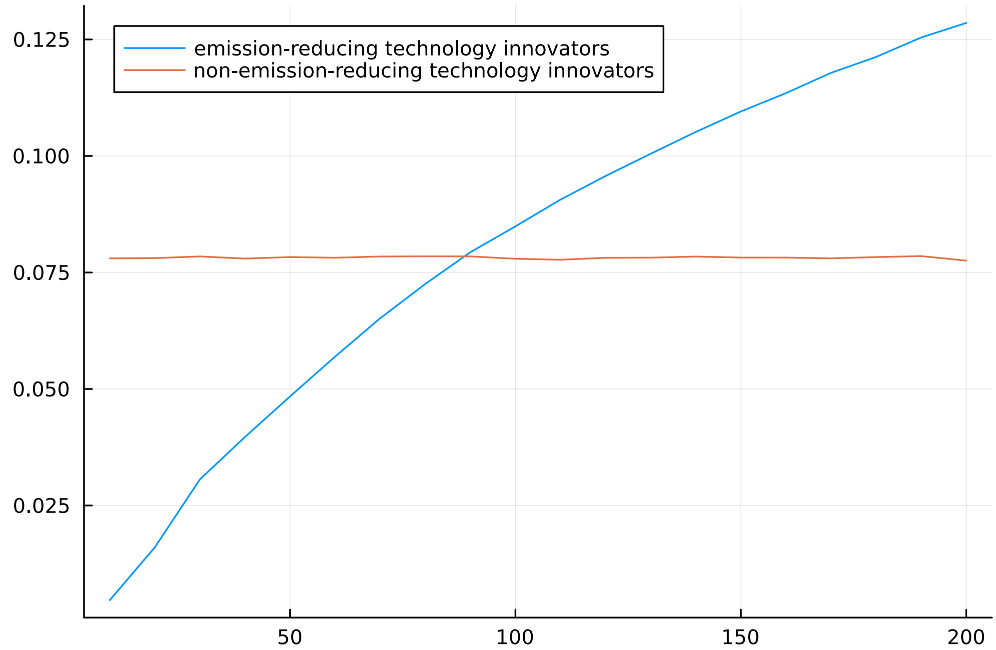
The main quantity affecting through which the EU-ETS affects firms' Carbon emission behavior is the level of the emission price. The higher the price, the more costly it is for firms to emit Carbon in their production processes. To reduce the burden of the policy firms can invent emission-reducing technologies and reduce their emission-intensity. This allows firms, ceteris paribus, to increase their production while keeping emission levels constant. In the model in Chapter 3, an increase in the price for carbon emissions influences firm profits by increasing the marginal cost of production. The higher the total cost of a firm's emissions, the larger its marginal benefits from decreasing its emission intensity. Policymakers repeatedly made this argument when arguing in favor of implementing the EU-ETS; however, up to now, there is little empirical evidence supporting this channel. A low emission price over most of the implementation period of the EU-ETS also made it difficult to detect any large-scale impacts. This study circumvents this problem by modeling firms' innovation decisions allowing for simulating counterfactual situations with varying emission prices instead of trying to identify the effect from the data directly.

I, therefore, utilize the estimated model parameters from the previous section and simulate variations in the emission price between 10 and 200€. Forecasting the Carbon price development is notoriously difficult, however the simulated range is well within expected price ranges of the near future of 80-160€ per allowance.[12] For each price level, I then simulate firms' innovation decisions for the next year and present the average emission-reducing or non-emission-reducing innovator rates in the sample from 100 simulation runs in Figure 5.

The results show that the emission price affects innovation activity in the sample substantially. Generally, the higher the emission price, the higher is the rate of emission-reducing innovators. At a price of 10€ only 0.4% of firms would decide to innovate in emission-reducing technologies. However, at a price of 200€ per carbon allowance, almost 13% of firms in the sample would decide to invest in emission-reducing technologies. This positive relationship between the emission-price and innovation decisions is highly nonlinear. While the emission-reducing innovator rate increases steeply with low emission prices, the rate of increase declines steadily with higher emission prices. However, the rate of non-emission-reducing innovators seems to be unaffected by the emission-price. In general, the model allows for a connection between the emission price and non-emission-reducing innovations through productivity increases. Non-emission-reducing innovations affect productivity positively, decreasing the amount of any input needed to, ceteris paribus, produce the same amount of output. The fewer emission generating inputs a firm needs in production the lower are its emission costs. Therefore, firms should have a higher incentive to increase productivity through non-emission-reducing innovations when facing higher emission costs. However, the simulations do not reveal such a connection. This might be caused by marginal benefits of non-emission-reducing innovations still not exceeding

[12]This range is the result from a survey of seven Carbon market expert organizations conducted by the Ariadne Project in 2022. See Pahle et al. (2022) for more details.

Figure 5: Simulation Emission Price Change - Results



Notes: The figure displays the average innovator rates in the sample for varying emission price levels. The simulation for each emission price level is repeated 100 times and the average innovator rate over all simulations is taken.

associated development costs, even with high carbon prices for those firms that did not innovate in the technology before.

These results confirm the weak version of the Porter Hypothesis stating that the environmental policy increases environmentally oriented innovation activity in the case of the EU-ETS. The simulations show that if the emission price is high enough it builds substantial incentives for firms to consider taking on the costs of innovating for the benefit of reducing the cost of the environmental regulation. However, I do not find support for the strong Porter Hypothesis stating that the environmental regulation additionally increases firms' general productivity. This is because I find that mission-reducing innovations at the same time reduce the developing firm's productivity while non-emission-reducing innovations are not affected by the policy.

8 Conclusion

Pricing undesired production by-products such as carbon emissions is one of the most prevalent type of environmental regulation. Most of the political and academic discussions of its impact focus on the direct additional cost for firms, which tends to reduce their competitiveness and increase consumer prices. This neglects important mechanisms through which the policy has a dynamic impact on the economy. Carbon prices have the additional effect of incentivizing firms to develop innovations aiming at reducing their carbon emissions. Existing studies, however, either only focus on parts of this dynamic mechanisms or on the introduction of the policy instead of the carbon price itself. I address this gap in the literature by developing a dynamic structural model of the EU-ETS's impact on firms' innovation activities. In the model, firms generate carbon emissions as a by-product of their production process. Pricing these emissions affects firms' profit, incentivizing them to develop innovations that reduce their future emission intensity. The model differentiates between emission-reducing and non-emission-reducing innovations while accounting for the dynamic nature of those decisions. This general model allows me to evaluate the impact of carbon price changes on innovation decisions and their impact on emissions and productivity, providing a substantial contribution to both the academic discussion of the Porter Hypothesis and the contemporary political discussion on the impact of environmental policies such as the EU-ETS.

I estimate the model's parameters for a large sample of German manufacturing firms using administrative carbon emission data combined with patent application information and firm's financial data. The combined data set represents over 93% of German firms that are regulated by the EU-ETS. My results show that innovations aimed at reducing emissions have a substantial impact on the emission intensity of developing firms. I confirm that innovations aimed at reducing carbon emissions indeed significantly lower the carbon emission intensity. This effect is substantial with 13.7% on average. This effect is carried over to a large extent to subsequent periods by a highly persistent emission intensity development process. At the same time, these innovations reduce the developing firm's productivity by about 1.5%. In contrast, non-emission-reducing innovations increase productivity by on average about 2.2%. These results imply that on average firms would not engage in developing emission-reducing innovations if their emissions were not priced through the EU-ETS and only develop non-emission-reducing innovations. Thus, carbon pricing due to the EU-ETS has stimulated emission-reducing innovation, as argued by Porter. These results emphasize that the EU-ETS has a substantial impact on the innovation behavior of regulated firms. Without pricing firms' carbon emissions, they would not develop innovations that are explicitly reducing their emission intensity because of their average negative impact on productivity. However, to quantify the impact the carbon price has on innovation activities, I simulate a counterfactual carbon prices between 10-200€. The results show that emission-reducing innovation activity reacts strongly to changes in the carbon price. Initially, emission-reducing innovation activity rises steeply with the emission price. However, the higher the price rises the less steep is the increase in innovation activity. In contrast non-emission-reducing innovation activity seems to be unaffected by the emission price. This implies that a high enough price significantly increases firms' emission-reducing innovation activity without reducing other non-emission-reducing innovations.

The results further show that innovation costs for inexperienced developers are substantially higher than for developers already experienced in innovating in the respective technology. Development costs for inexperienced developers of emission-reducing innovations are substantially lower than for non-emission-reducing innovations, while maintaining the development of emission-reducing innovations is more expensive than maintaining non-emission-reducing innovations.

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015): "Identification Properties of Recent Production Function Estimators," *Econometrica*, 83, 2411–2451.

Angelucci, S., F. J. Hurtado-Albir, and A. Volpe (2018): "Supporting global initiatives on climate change: The EPO's "Y02-Y04S" tagging scheme," *World Patent Information*, 54, S85–S92.

Arellano, M. and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297.

Aw, B. Y., M. J. Roberts, and D. Y. Xu (2011): "R&D Investment, Exporting, and Productivity Dynamics," *American Economic Review*, 101, 1312–44.

Baumol, W. J. (1988): *The Theory of Environmental Policy*, Cambridge: Cambridge University Press, 2 ed ed.

Becker, R. and V. Henderson (2000): "Effects of Air Quality Regulations on Polluting Industries," *Journal of Political Economy*, 108, 379–421.

Calel, R. and A. Dechezleprêtre (2016): "Environmental Policy and Directed Technological Change: Evidence from the European Carbon Market," *The Review of Economics and Statistics*, 98, 173–191.

Colmer, J., R. Martin, M. Muûls, and U. J. Wagner (2022): "Does Pricing Carbon Mitigate Climate Change? Firm-Level Evidence from the European Union Emissions Trading Scheme," Discussion Paper DP16982, CEPR.

Council of European Union (2003): "Directive 2003/87/EC on establishing a scheme for greenhouse gas emission allowance trading within the Community and amending Council Directive 96/61/EC," Official Journal of the European Union.

De Ridder, M., B. Grassi, and G. Morzenti (2022): "The Hitchhiker's Guide to Markup Estimation," Discussion Paper 17532, CEPR.

Dixit, A. K. and J. E. Stiglitz (1977): "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67, 297–308.

Doherr, T. (2023): "The SearchEngine: A Holistic Approach to Matching," Discussion Paper 23-001, ZEW.

Doraszelski, U. and J. Jaumandreu (2013): "R&D and Productivity: Estimating Endogenous Productivity," *The Review of Economic Studies*, 80, 1338–1383.

European Commission (2007): "EU action against climate change - EU emissions trading: an open system promoting global innovation:," Brochure, available at https://op.europa.eu/s/y0OJ, European Commission.

——— (2011): "A Roadmap for moving to a competitive low carbon economy in 2050," EU technical report COM 112, European Commission.

Fernández, C., G. Koop, and M. F. J. Steel (2002): "Multiple-Output Production With Undesirable Outputs," *Journal of the American Statistical Association*, 97, 432–442.

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

Førsund, F. R. (2009): "Good Modelling of Bad Outputs: Pollution and Multiple-Output Production," *International Review of Environmental and Resource Economics*, 3, 1–38.

Hayashi, F. (2000): *Econometrics*, Princeton: Princeton University.

Kumbhakar, S. C. and E. G. Tsionas (2016): "The Good, the Bad and the Technology: Endogeneity in Environmental Production Models," *Journal of Econometrics*, 190, 315–327.

Maican, F. G., M. Orth, M. J. Roberts, and V. A. Vuong (2022): "The Dynamic Impact of Exporting on Firm R&D Investment," *Journal of the European Economic Association*, jvac065.

Martin, R., L. B. De Preux, and U. J. Wagner (2014): "The impact of a carbon tax on manufacturing: Evidence from microdata," *Journal of Public Economics*, 117, 1–14.

Martin, R., M. Muûls, and U. J. Wagner (2016): "The Impact of the European Union Emissions Trading Scheme on Regulated Firms: What Is the Evidence after Ten Years?" *Review of Environmental Economics and Policy*.

Murty, S. and R. R. Russell (2020): *Bad Outputs*, Springer, chap. 12, 1–53.

Olley, G. S. and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

Pahle, M., J. Sitarz, and S. Osori (2022): "The EU-ETS price through 2030 and beyond: A closer look at drivers, models and assumptions. Input material and takeaways from a workshop in Brussels," The Ariadne Project.

Peters, B., M. J. Roberts, and V. A. Vuong (2022): "Firm R&D Investment and Export Market Exposure," *Research Policy*, 51, 104601.

Peters, B., M. J. Roberts, V. A. Vuong, and H. Fryges (2017): "Estimating dynamic R&D choice: an analysis of costs and long-run benefits," *RAND Journal of Economics*, 48, 409–437.

Peters, B. and M. Trunschke (2022): "Choosing Technologies: Benefits of Developing Fourth Industrial Revolution Technologies," Unpublished.

Porter, M. (1991): "America's Green Strategy," *Scientific American*, 264, 168.

Porter, M. E. and C. v. d. Linde (1995): "Toward a New Conception of the Environment-Competitiveness Relationship," *Journal of Economic Perspectives*, 9, 97–118.

Rammer, C., T. Doherr, B. Krieger, H. Marks, H. Niggemann, B. Peters, T. Schubert, M. Trunschke, J. von der Burg, and S. Eibelshäuser (2022): "Innovationen in der deutschen Wirtschaft: Indikatorenbericht zur Innovationserhebung 2021, ZEW-Innovationserhebung - Mannheimer Innovationspanel (MIP)," Technical Report 251141, ZEW - Leibniz Centre for European Economic Research.

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55, 999–1033.

A Appendix

A.1 Note on Data Matching Procedure and Quality

I merge the three data sources (EUTL, Orbis, PATSTAT) using firms names and addresses present in each data set. The European Union Transfer Log has 2663 industrial installations from 1465 German firms (account holders). Using Thorsten Doherrs fuzzy name matching program *Search Engine* and a manual Creditreform firm identifiers (crefos) search in the Orbis data for non-matched EUTL account holders, I can match 1331 crefos to 1336 account holders (2663) installations.[13] The Orbis data uses the crefo as a firm identifier which allows me to merge the two data sets based on crefo-year observations. Table A.1 provides some merging statistics. It shows that a high share of firms from the EUTL data are successfully merged to the Orbis data. However, the high number of missing values in the Orbis data reduces the number of observations in the analysis substantially.

Table A.1: Merge Statistics EUTL-Orbis

		Observations directly matched	Firms	Observations from matched firms
Merged		8,978	1,240	21,176
Not merged	EUTL	8,443	151	1,960
	Orbis	9,137,167	1,268,947	9,131,452

I then merge the combined Orbis-EUTL data to patent information from PATSTAT. In the first step, applicants of patents need to be matched to firms using names and addresses in both the Creditreform database and PATSTAT. This match was already completed in a previous project by Thorsten Doherr and Vanessa Behrens using the program *Search Engine*. Because of the large number of observations, they employed a neural net approach to selecting false positives using absolute meta-information instead of a manual cleaning approach. With patent applications matched to crefos, I can, in a second step, merge all available patent applications to the Orbis-EUTL data on a crefo-year basis.

A.2 Note on the Computationally Efficient Implementation of the Likelihood Function

The most computationally expensive part of the empirical approach is the estimation of the development cost distribution parameters θ. The estimation utilizes a numerical maximum likelihood procedure to retrieve the point estimates. Similar to Peters et al. (2017), the procedure uses a Nested Fixed Point Algorithm, which is based on a likelihood function of

[13] 60 firms are only matched to pseudo crefos given to airlines. They are dropped from the sample as it is impossible to find, even with a manual search, any valid firm/crefo in the Orbis data.

the conditional choice probabilities, i.e.,

$$\mathcal{L}(\theta|i^E_{it}, i^N_{it}, s_{it}) = \prod_i \prod_t P(i^E_{it}|i^N_{it}, s_{it}, \theta) P(i^N_{it}|i^N_{it-1}, s_{it}, \theta). \tag{28}$$

I can rewrite the likelihood function as in equation 29 to specifically show the contribution of each observation to the likelihood function's value, where θ^E and θ^N are the parameter vectors specific to each technology type, respectively. θ^{Es} and θ^{Ns} represent the parameters for inexperienced and θ^{Em} θ^{Nm} for experienced firms.

$$\begin{aligned}
\mathcal{L}(\theta|i^E_{it}, i^N_{it}, s_{it}) =& \prod_i \prod_t P(i^E_{it}|i^N_{it}, s_{it}, \theta) P(i^N_{it}|i^N_{it-1}, s_{it}, \theta) \\
=& \prod_i \prod_t P(i^E_{it} = 1|i^N_{it}, s_{it}, \theta^E)^{i^E_{it}} P(i^E_{it} = 0|i^N_{it}, s_{it}, \theta^N)^{1-i^E_{it}} \\
& P(i^N_{it} = 1|i^N_{it-1}, s_{it}, \theta^N)^{i^N_{it}} P(i^N_{it} = 0|i^N_{it-1}, s_{it}, \theta^N)^{1-i^N_{it}} \\
=& \prod_i \prod_t P(i^E_{it} = 1|i^N_{it} = 1, i^E_{it-1} = 1, s'_{it}, \theta^{Em})^{i^E_{it} i^N_{it} i^E_{it-1}} \\
& P(i^E_{it} = 1|i^N_{it} = 1, i^E_{it-1} = 0, s'_{it}, \theta^{Es})^{i^E_{it} i^N_{it}(1-i^E_{it-1})} \ldots
\end{aligned} \tag{29}$$

The conditional choice probabilities represent the probabilities of a firm choosing to develop a specific technology in a given year. This is equal to the probability that the discounted marginal expected long-run value of choosing to develop the respective technology is greater than the associated development cost, i.e.,

$$\begin{aligned}
P\left(i^E_{it}|s_{it}\right) &= P\left(\delta\,\mathbb{E}\left[V\left(i^E_{it} = 1|s_{it}\right) - V\left(i^E_{it} = 0|s_{it}\right)\right] \geq \mathbb{E}[C^E(\theta^E)]\right), \\
P\left(i^N_{it}|s_{it}\right) &= P\left(\delta\,\mathbb{E}\left[V\left(i^N_{it} = 1|s_{it}\right) - V\left(i^N_{it} = 0|s_{it}\right)\right] \geq \mathbb{E}[C^N(\theta^N)]\right).
\end{aligned} \tag{30}$$

Assuming development costs $C^E(\theta^E), C^N(\theta^N)$ to be exponentially distributed, the empirical representations of these probabilities are given by

$$\begin{aligned}
P\left(i^E_{it}|s_{it}\right) &= 1 - exp\left(\frac{\delta V\left(i^E_{it} = 1|s_{it}\right) - \delta V\left(i^E_{it} = 0|s_{it}\right)}{\theta^E}\right), \\
P\left(i^N_{it}|s_{it}\right) &= 1 - exp\left(\frac{\delta V\left(i^N_{it} = 1|s_{it}\right) - \delta V\left(i^N_{it} = 0|s_{it}\right)}{\theta^N}\right).
\end{aligned} \tag{31}$$

Calculating these probabilities is the computationally most expensive part of the estimation procedure because retrieving the probabilities relies on calculating the fixed point of the value function. I approximate the fixed point using value function interactions, starting from an arbitrary initial value function value to calculate expected value functions and a new value for the value function. This new value is then used in the next iteration instead of the starting value. This procedure is repeated until the value function value used at the beginning of the iteration does not differ more than a critical value from the newly calculated value function value. This procedure needs to be repeated for each grid point. Using a highly efficient programming language in combination with well-optimized code is essential for the feasibility of the estimation. Interpreted languages like Matlab, Stata, Python, or R are relatively slow, as each command in the script needs to be interpreted by the respective program at run-time. Compiled languages such as C++, Fortran, or Julia translate the code into compiled programs ahead of executing it. Unlike most other compiled languages, Julia uses a just-in-time compiler which compiles the user-written code into machine language during the first execution of the code. It, therefore, combines a lot of the flexibility of interpreted languages with the speed of compiled ones. In order to achieve the necessary efficiency, I pre-allocate memory for objects used in the performance-critical parts in caches and only replace their values of the objects' elements during the calculation of the likelihood function values. This avoids slow memory allocations during execution. Instead of running each iteration of the value function iterations in one function for all firm types, I calculate each iteration separately for each firm type. Using Julia's multi-threading capabilities on a server with 64 CPU cores allows this approach to reduce run time substantially. I additionally employ Julia's multi-threading options and highly optimized linear algebra libraries where they promise performance increases during the likelihood function calculation.[14]

[14] Julia supports Basic Linear Algebra Subprograms (BLAS) and LAPACK routines. When using these routines inside a multi-threading framework (such as a parallelized loop) it is important to carefully consider parallel computing options from these packages because a nested multi-threading can lead to performance bottlenecks.

A.3 Tables

Table A.2: Patent Application Effect Lag: Employeemnt

	(1) Employees	(2) Employees	(3) Employees	(4) employees
Patent_t	34.38^{***}			
	(8.692)			
Patent_{t-1}		29.87^{***}		
		(8.668)		
Patent_{t-2}			16.47^{**}	
			(7.767)	
Patent_{t-3}				8.008
				(8.421)
N	88739	64986	51581	43178

Notes: Method: OLS; all models include firm-year fixed effects; heteroscedasticity robust standard errors in parentheses below point estimates; * p<0.1, ** p<0.05, *** p< 0.01.

Table A.3: Patent Application Effect Lag: Employeemnt

	(1) ln(revenues)	(2) ln(revenues)	(3) ln(revenues)	(4) ln(revenues)
Patent_t	0.0190^{**}			
	(0.00793)			
Patent_{t-1}		0.0145^{**}		
		(0.00659)		
Patent_{t-2}			0.0133^{*}	
			(0.00715)	
Patent_{t-3}				0.0222^{**}
				(0.00755)
N	72013	56659	46985	40160

Notes: Method: OLS; all models include firm-year fixed effects; heteroscedasticity robust standard errors in parentheses below point estimates; * p<0.1, ** p<0.05, *** p< 0.01.

Table A.4: Patent Application Effect Lag: Employemnt

	(1) ln(emission int)	(2) ln(emission int)	(3) ln(emission int)	(4) ln(emission int)
Em. red. pat_t	-0.197*			
	(0.104)			
Em. red. pat_{t-1}		-0.143		
		(0.106)		
Em. red. pat_{t-2}			0.0220	
			(0.0663)	
Em. red. pat_{t-3}				0.0599
				(0.158)
N	3354	3107	2933	2769

Notes: Method: OLS; all models include firm-year fixed effects; heteroscedasticity robust standard errors in parentheses below point estimates; * p<0.1, ** p<0.05, *** p< 0.01.