Why Is Trading So Important in Cap and Trade? The Role of Economies of Scale and Productivity

Phuong Ho*

October 22, 2022

(Job Market Paper)

(Download the most recent version here)

Abstract

Economists have long established the cost effectiveness of cap and trade (CAT) owing to the cost heterogeneity between firms. I offer a new value of CAT: cost reduction *within* firms owing to productivity improvement and economies of scale. Overcoming the unobserved-cost challenge, I extend the literature on production function and introduce a method to estimate economies of scale using data on output and input quantity. I combine this method with the difference-in-difference strategy to exploit the policy transition from non-tradable cap to cap and trade in Norwegian cod fishery to identify the causal impacts of trading fishing quotas. I find trading increased vessels' productivity and facilitated the realization of existing economies of scale: vessels acquired quotas, expanded their operation, and moved toward the minimum average cost levels. Vessels realized economies of scale by upgrading their sizes and going fishing more often. I decompose the output-based value of traded quotas and find economies of scale played a main role in the first few years after a big vessel acquired quotas. These results highlight (i) flexibility through tradability in environmental regulations can reduce production costs of a firm, (ii) cost reduction is gained by both boosting existing production factors and advancements beyond scale economies, and (iii) consolidation can be a sign of cost efficiency owing to economies of scale rather than market power abuse.

Keywords: cap and trade, economies of scale, productivity, production function

JEL Classification: D22, D24, L11, Q22, Q28

*SNF-Centre for Applied Research at Norwegian School of Economics. Email: phuongho@arizona.edu. I thank Dan Ackerberg, Sarah Johnston, Ashley Langer, Derek Lemoine, Shanjun Li, Gautam Gowrisankaran, Dominic Parker, Mauricio Varela, and seminar participants at University of Arizona, Maastricht University, Cornell University (Dyson School), IIOC 2022, CREEA 2022, NBER Summer Institute IO 2022, and University of Wisconsin-Madison (AAE) for thoughtful discussions and comments. I thank Norwegian Directorate of Fisheries for providing access to the data as well as Tove Aasheim, Anita Kjeilen, Linda Nøstbakken, and Mads Wold for many clarifications on the dataset structure and Norwegian fishery regulations. I gratefully acknowledge funding from Research Council of Norway (grant 295197). Part of this work was completed while I stayed at Dyson School of Applied Economics and Management, Cornell University.

1 Introduction

Over the past few decades, the market-based cap-and-trade (CAT) approach has become popular to ration access to the commons in many contexts, notably air pollution, fisheries, and water management, and in many countries (Tietenberg, 2003; Stavins, 2011). Economists have long established that under specific conditions, CAT can minimize the total cost of reaching a predefined environmental target (Dales, 1968; Montgomery, 1972).[1] The reasoning is trading allows the permits to flow from those that value the permits less to those that value more (due to higher abatement cost of reducing emissions or lower cost of exploiting resources). However, in fishery management in which a cap defines the maximum amount of fish a vessel may catch (fishing quotas), many territories oppose to the trading part of the system. They concern trading will lead to consolidation of permits and outputs in certain big firms that will abuse market power (Grainger and Parker, 2013; Kroetz, Sanchirico and Lew, 2015; Helgesen, 2022).

This paper studies the value of trading, cap and trade as compared to non-tradable permits, using the case of fishing permits (quotas). Whereas economists have shown the cost effectiveness of CAT that is attributed to the cost heterogeneity between firms, I offer a new perspective of CAT: Trading reduces production costs *within* a firm. Why? I show that, first, trading shifts *down* average cost of the firm by improving quantity-based total factor productivity (TFPQ). Second, trading induces shifts *along* average cost when the firm has economies of scale. That is, if the firm's average cost is decreasing, trading allows the firm to adjust its inputs, expand its operation, and produce at lower per-unit cost. This paper presents a method to *empirically* examine this new cost efficiency hypothesis and demonstrates it using the case of cap and trade in Norwegian cod fishery. Furthermore, I show the value of trading, in terms of output unit, comprises those two components and how to *decompose* it.

A challenge to examine the new cost efficiency hypothesis is costs are rarely observed. I overcome this challenge by offering an approach to measure and estimate economies of scale using data on output and input quantity. I measure economies of scale using the cost elasticity of output, as the ratio of marginal cost to average cost, and show it equals the reciprocal of the sum of all output elasticities of inputs. The method relies on the standard cost minimization of the firms' input decisions and does *not* restrict the production function to certain specifications. I then use the contemporary proxy variable approach in the industrial organization (IO) literature (Ackerberg, Caves and Frazer, 2015) to estimate TFPQ and output elasticities of inputs (and hence, economies of scale).

I apply the method to study CAT in fishery for three important and interesting features of this context. First, environmental protection is often believed to increase production costs by restricting production or constraining the actions of firms. This issue would be starkly evident in fishery management where regulations aim to directly constrain output of a main product

[1]Conditions include zero transaction costs, complete information, perfectly competitive markets, and cost minimization behavior.

rather than a by-product like emissions. Stavins (2011) notes that conventional command and control (CAC) approaches in fishery through annual catch limits, restrictions on fishing tools, or closure of particular areas, in several case examples, triggered fishermen to employ excessively expensive fishing methods and even led to worse outcomes than open-access fisheries. However, besides the static cost effectiveness, CAT has been believed to reduce costs over time by providing stronger incentives for innovation and more likely improving productivity than CAC (Schmalensee and Stavins, 2013). Such belief relies on a "narrow" version of Porter hypothesis that argues flexible policies (such as CAT) give firms greater incentives to innovate than prescriptive regulations. Empirical findings of Porter hypothesis have been mixed and little focused on fishery (Ambec et al., 2013; Dechezleprêtre and Sato, 2017; Cohen and Tubb, 2018). My finding that CAT in fishery, relative to CAC (non-tradable permits), increases productivity agrees with the "narrow" Porter hypothesis. I provide a finding in fishery management where regulations directly control output of the main product. I also provide the first result that identifies trading as the causal factor that explains the outperformance of CAT over CAC on productivity change.

Second, although CAT is now mostly discussed in air emissions control, it was first implemented in natural resource management such as water and fisheries to avoid over-exploitation of the commons informally in the 1930s and formally in the 1970s (OECD, 2001). Despite such a long history, many territories have opposed to the trading part of the system for concerns about consolidation of fishing quotas and catches in a few big firms. For example, some territories used to adopt or currently adopt non-tradable quotas such as British Columbia, Denmark, Norway, and Peru (Fox et al., 2003; Asche et al., 2008; Natividad, 2016). Almost places that have adopted CAT have carefully monitored the degree of consolidation and imposed some restrictions on trade such as trade within a vessel class and cap ceilings (Grainger and Parker, 2013; Kroetz, Sanchirico and Lew, 2015). This paper documents the new value of trading—owing to productivity improvement and economies of scale—to provide additional benefits of tradability (or transferability as a more familiar term in fishery) for a regulator to consider when designing appropriate management measures. Furthermore, I highlight new insights on consolidation: Consolidation is a way for a firm to exploit economies of scale and achieve the necessary cost efficiency. Of course, verifying the realization of economies of scale is an empirical question. This paper provides an empirical method to answer it.

Third, the fishery context provides an ideal setting to credibly estimate productivity and production function, a key component to estimate economies of scale without observing cost data. Estimating a production function is challenging due to simultaneity and selection bias of productivity. Contemporary studies in the IO literature have provided a method, so called proxy variable approaches, to deal with these problems (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg, Caves and Frazer, 2015). However, whereas the method assumes output quantity is observed, typical datasets contain only revenue and many applications in practice replace output with revenue deflated by an industry-level price deflator. Such treatment can significantly bias both productivity and output elasticities (Foster, Haltiwanger and Syverson, 2008; De Loecker, 2011; Bond et al., 2021). The fishery context where output quantity is observed and the product is almost homogeneous allows me to address the output-price bias by estimating a quantity-based production function.[2]

Besides the challenges in inferring costs without observing costs and credibly estimating the production function, another empirical challenge is to identify the causal effects of trading component in a CAT regime. To do that, I exploit a novel feature of Norwegian cod fishery. I exploit the fact that (i) Norway switched from non-tradable quotas to tradable quotas for vessels within the same cod coastal fleet (fishing the same specie, cod, and having similar access to an ocean region and fishing gears), (ii) only a subset of the fleet (i.e. except vessels below 11m) is allowed to trade quotas, and (iii) the trade qualification is defined on a historical vessel length, that is, the length three years *before* the trade program was implemented. I use a difference-in-difference (DID) strategy to compare the changes in the outcomes of interest from before to after the trade program between vessels in the trade qualified group and vessels in the unqualified group. This strategy allows me not only to focus on the role of tradability rather than being confounded by the role of quotas, but also to identify the causal effects of trading.

The empirical analysis in this paper is divided into five parts. First, I estimate productivity and economies of scale in a flexible form of a production function, the translog function, to obtain yearly and vessel-specific productivity and economies of scale. I consider the production function with four input factors: vessel length, crew size, fishing distance (distance from a major catch location to a vessel's home municipality), and the number of fishing trips in a year. The focus is on vessel by year because a quota limits yearly maximum catch for a vessel.

The distributions of estimated productivity and economies of scale by trade-qualified status and over years present three important and stylized facts. First, vessels in the tradable groups with lowest productivity levels have exited fishery and stayers have seen improvement in their own productivity. Second, all vessels had had economies of scale before trading was allowed. Third, vessels in the tradable groups with little savings from exploiting economies of scale (due to very low cost elasticities of output) have exited fishery and stayers have seen increases in their own cost elasticities toward 1 (where average cost is minimized).

In the second part of the analysis, I examine the degree of consolidation and the causal impacts of trading on productivity and economies of scale using the above DID identification strategy. I find evidence of consolidation: The number of vessels in the trade-qualified groups significantly decreased and quota trading increased harvest of a vessel by 50–86%. Trading also increased productivity of a vessel by 25–46%. More importantly, trading is the causal factor that induces the increase in cost elasticity (or economies of scale). Together with the above

[2]A few studies also address the bias in this way, using restricted datasets in certain countries and industries to study the impacts of international trade reforms on productivity and markups (Lu and Yu, 2015; De Loecker et al., 2016), or the impacts of mergers and acquisitions on productivity (Braguinsky et al., 2015).

findings on the distribution of cost elasticity and output expansion, the results suggest trading helps a vessel exploit its existing economies of scale to expand harvest and move toward the minimum average cost level. A back-of-envelope calculation finds a vessel saves about 1.3–5.5% of its production cost owing to economies of scale.

Third, I decompose the value of trading, in terms of output unit, into two components: cost shifting due to productivity improvement and cost savings owing to economies of scale. This is a counterfactual experiment exercise using the estimate of the production function: How much the output would grow given a change in either productivity or economies of scale, ceteris paribus. In the studied Norwegian fishery context, their relative contributions to the value of trading vary substantially with the historical length (the length that defines tradability). Whereas productivity played a main role in expanding output for vessels in the two tradable middle-size groups (11–14.9m and 15–20.9m), I find economies of scale crucially contribute to output expansion in the first three years since the first time a vessel in the biggest group (21–28m) acquired quotas.

Fourth, I investigate where economies of scale come from. Specifically, I use the DID to examine the changes in input factors. I find trading significantly increases the number of fishing trips in a year of a vessel in all tradable groups, the vessel length of a vessel in the upper tradable group (21–28m), while having no impact on crew size and fishing distance. This result suggests fishermen have exploited economies of scale by investing in bigger boats and going fishing more often. Furthermore, the fact that the number of vessels has reduced while the number of fishing trips in a year of a vessel has increased suggests CAT can be more effective in prolonging fishing seasons than CAC, which underlies the realization of economies of scale.

Fifth, I discuss the implications of my findings on consolidation and market power matters, especially with an extra analysis on the change in observed fish sales price. As mentioned, my findings of cost reduction within a firm provide additional benefits of CAT for a policy maker to weigh the benefits against the costs. The benefit of realizing economies of scale formally establishes the tradeoff between consolidation concern and efficiency. However, I emphasize that reaching a final conclusion requires additional information and criteria.

Two following statements give examples of incomplete judgement. The first statement claims the evidence of realizing economies of scale dismisses the presence of market power. This statement is incomplete because economies of scale and market power may co-exist. For example, using the DID analysis and data on fish sales price, I find trading has no impact on price. One may state that this result implies fishermen have abused market power, because they could have lowered price given their cost reduction owing to economies of scale but they did not. Nevertheless, this second statement is also incomplete, because exploiting economies of scale induces reduction in average cost rather than marginal cost. If market power is defined as the wedge between marginal price and marginal cost, the second statement will be incomplete. In sum, my findings and methodology provide essential insights on the additional benefits of CAT vis à vis CAC, a necessary caveat on consolidation concern, and the tool to verify them with the limitation of data to request further information and criteria for final conclusion.

This paper makes two primary contributions on both empirical and methodological sides. On the empirical side, I provide the new perspective of CAT and its additional value for policy implications as discussed above: cost reduction *within* firms owing to productivity improvement and economies of scale. In environmental economics, studies on benefits of CAT relative to CAC have focused on the value of trading owing to heterogeneity in abatement costs between firms (Carlson et al., 2000; Keohane, 2006; Fowlie, Holland and Mansur, 2012). In fishery context, large sample analyses with attempts to establish causality and credible comparison have focused on the relative performance of CAT to open access (Costello, Gaines and Lynham, 2008; Costello et al., 2010), the impacts of trading restrictions on resource rent (Kroetz, Sanchirico and Lew, 2015), and the impacts of tradability on stock biomass (Isaksen and Richter, 2019).[3]

By focusing on the cost efficiency within firms, I fill in the gap of the CAT literature and connect it to the rich literature on the impacts of environmental regulation on firm performance. Given the long-standing perception of conflict of interests between business and regulators, a long literature has studied the impacts of environmental regulation on firm performance and competitiveness (Gray, 1987; Becker and Henderson, 2000; Ryan, 2012; Fabra and Reguant, 2014). A classical and important theory in this literature—Porter hypothesis—has stimulated empirical studies to test whether environmental regulations, especially the flexible approaches, give firms stronger incentives to innovate and more likely offer productivity improvement than prescriptive regulations; see recent reviews by (Dechezleprêtre and Sato, 2017) and Cohen and Tubb (2018).

My finding supports such Porter hypothesis, adding to the current mixed empirical evidence of the effects of CAT. However, I provide the first result that identifies the causal effect of trading, explaining the outperformance of CAT over CAC on productivity change. I also provide the first study on the effect of environmental regulation on firms' production cost owing to economies of scale. These results provide two channels to explain the technology responses to CAT: upgrade and else (including innovate). Specifically, my findings suggest first, flexibility through trading induces upgrade on current capital (vessel size) to help firms realize economies of scale. Second, flexibility through trading induces advancements beyond the existing input factors, returns to scale, and scale economies, such as development in management and operating skills and innovation of new inputs, which are captured through the increase in TFPQ.

On the methodological side, this paper is the first study that brings in the modern fast-growing production-based approach in IO to designing environmental regulations. The paper further extends the frontier by providing a method to estimate economies of scale that is applicable beyond fishery CAT context.

[3]Isaksen and Richter (2019) use a global database of fishery institutions and biomass statistics to study the effects of private property rights on the stock biomass status, comparing transferable quota regime to non-transferable quota regime and quota regime to open access. They find transferable quotas are more effective in preventing stock collapsing.

Estimating a production function has been useful and important for two reasons. First, one can obtain the estimate of productivity and then investigate the dynamics of productivity over time and across firms. Recent advancements in the IO literature introduce a proxy variable method to control for simultaneity and selection bias of productivity; see Olley and Pakes (1996); Levinsohn and Petrin (2003); Ackerberg, Caves and Frazer (2015); Gandhi, Navarro and Rivers (2020). A few studies have recently combined the production function estimation with a difference-in-difference framework to study the impact of international trade and industrial events such as exporting status and mergers and acquisitions on productivity and markups (De Loecker, 2013; Braguinsky et al., 2015; Stiebale and Vencappa, 2018; Rubens, 2021).

The second advantage of the production function estimation is one can recover the markup (the price-cost ratio) from the estimates of a production function. This approach to recover markup has been so-called the production approach to distinguish it from the demand approach that estimates the demand and relies on assumptions on how firms compete in the market.[4] In contrast to the demand approach, the production approach relies on the classic cost-minimizing behavior in firms' input allocation and the observed input's expenditure share in revenue, following the work by De Loecker and Warzynski (2012). Its applications are rapidly growing, spanning into studies in international trade and labor markets; see De Loecker et al. (2016); De Loecker, Eeckhout and Unger (2020); Autor et al. (2020); Rubens (2022).[5]

This paper could have estimated the markups to examine how CAT affects firms' markups. Such investigation will test whether CAT, especially in fishery, leads to serious consolidation and market power abuse, because this is ultimately an empirical question. Unfortunately, I do not observe the share of input cost in revenue, a key information to apply the production-based approach to recover markups.

However, and most importantly, I offer a method to measure economies of scale and infer changes in average costs. This method introduces the third advantage of the production function estimation and opens new directions for applications where consolidation and economies of scale matter.

The rest of the paper is outlined as follows. Section 2 provides a theoretical foundation of economies of scale: the application in fishery CAT, measuring, and estimating theory. Section 3 summarizes the Norwegian cod fisheries regulations. Section 4 presents the difference-in-difference (DID) strategy to estimate the intent to treat and average treatment on the treated of trading fishing quotas. The section also discusses the proxy variable approach to empirically estimate productivity and economies of scale, and identification. Section 5 describes data. Section 6 shows the impacts of trading on productivity and economies of scale. Section 7 discusses the decomposition of the output-based value of trading, where economies of scale

[4]Examples of the demand approach are huge, following advancements in demand estimation by Berry (1994) and Berry, Levinsohn and Pakes (1995).

[5]Raval (2020) recently points out a problem with estimating markups in a certain functional form of production functions. Doraszelski and Jaumandreu (2018); Demirer (2020) suggest solutions by showing how to estimate a production function with factor-augmenting productivity.

come from, and the implications of economies of scale on consolidation concerns. Section 8 concludes.

2 Economies of Scale: Theoretical Framework

This section describes the framework to estimate economies of scale using production data. Section 2.1 illustrates how economies of scale provide value in fishery CAT as a motivating example. Section 2.2 shows how to measure economies of scale. Section 2.3 presents the theoretical framework and its assumptions to estimate economies of scale. Ultimately, estimating economies of scale requires estimates of production function coefficients. This paper uses a proxy variable approach to estimate the production function (see Ackerberg, Caves and Frazer (2015)), along with OLS-with-fixed-effects and dynamic panel approach as robustness checks. Because I combine this method with the DID strategy to identify the causal impacts of trading, I delay the presentation of the DID regressions and the proxy variable approach to Section 4, after discussing the policy background. However, Section 2.3 highlights that the framework to measure and recover economies of scale using estimates of production function does not impose any restrictions on the production function. Section 2.4 goes back to the motivating example and formally shows how to decompose the value of trading into the component due to productivity improvement and the component due to economies of scale. Section 2.5 discusses caveats and extensions when the assumptions in the theoretical framework are relaxed.

2.1 Motivation: The Case of Cap and Trade in Fishery

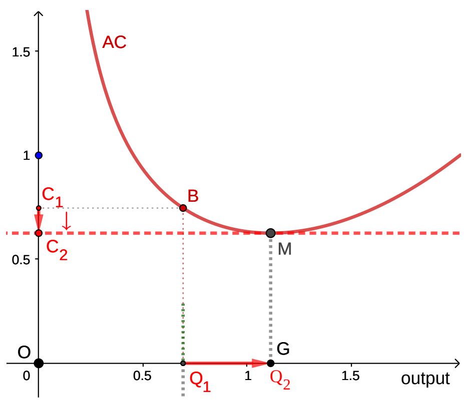


Figure 1: Economies of scale refer to decreasing average cost: movement from B to M

To remind, economies of scale refer to the cost advantages that a firm obtains for its scale of operation: decreasing average cost. Consider the average cost in Figure 1. The movement from B to M exhibits economies of scale: As quantity of production increases from Q_1 to Q_2, the average cost of each unit decreases from C_1 to C_2. On the other hand, diseconomies of scale are the output segment in which the average cost is increasing.

(a) A simple case: no change in productivity (b) When trading increases productivity of quota buyers

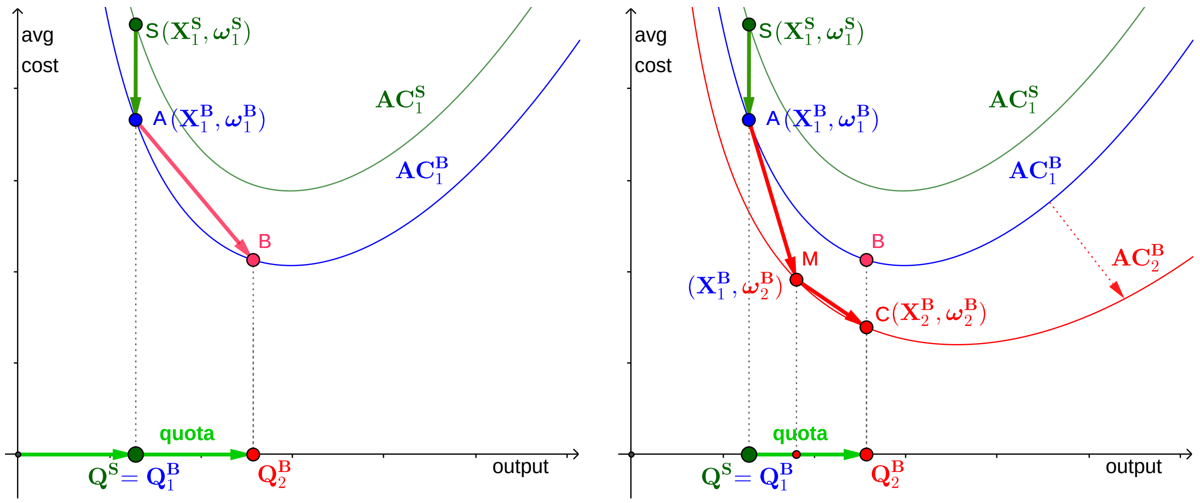


Figure 2: The Value of Trading

Note: Consider two average cost curves of two fishermen S and B. Any change in input factors X moves the fishing activities along the cost curve, whereas changes in productivity ω shift the cost curve. Figure 2a illustrates the value of trading if trading does not affect productivity (of quota buyer B). The value of traded quota in this case is owing to economies of scale AB. Figure 2b shows the cost reduction if trading affects productivity. The value of traded quota will consist of value due to productivity improvement and value due to economies of scale. Productivity improvement will shift the average cost from point A to M, whereas economies of scale will move the average cost from point M to C.

Economies of scale play an important role in explaining the cost savings in several contexts. I now discuss the case of cap and trade in fishery. Consider an example of two fishermen, S and B, with different average cost levels, AC_1^S and AC_1^B, respectively; see Figure 2a. Although they catch the same amount of fish ($Q^S = Q_1^B$), fisherman B catches at lower average cost: point A with the input use X_1^B and productivity ω_1^B instead of fisherman S's point S with the input use X_1^S and productivity ω_1^S. Thus, it will be more cost effective if fisherman S sells his quota (Q^S) to fisherman B. In that case, fisherman B will be able to double his catch to Q_2^B and operate at point B on his average cost curve. Studies on cap and trade in both resource and environmental economics have discussed the cost savings due to cost heterogeneity between agents: the cost change from quota seller's S to quota buyer's A. I provide another perspective of cost savings: the cost efficiency from operating at point A to point B *within* the quota buyer's fishing operation owing to economies of scale. This paper shows how to measure, estimate, and verify this new perspective of cost advantage.

Nevertheless, the goal of this paper does not stop there. What if trading affects productivity? A number of robust studies in the international trade literature have shown that trade reforms affect industry and firms' productivity; see De Loecker (2011, 2013); De Loecker and Goldberg (2014); Halpern, Koren and Szeidl (2015); De Loecker et al. (2016). Reasons include learning by exporting effects, intensified import competition that may lead to reduction in input-inefficiencies and adoption of better management practices, and the import of new intermediate products. Although trade between countries may differ from trading caps between firms in cap and trade, considering the possibility of an effect on productivity is useful. Indeed, I find this is the case in cap and trade in Norwegian cod fishery, which I show and discuss in Section 6.

Suppose that after acquiring the quota, fisherman B's productivity will increase, making his average cost curve shift down to the curve AC_2^B; see Figure 2b. So, with the double fishing quota, he will go fishing at the new cost level C on the cost curve AC_2^B instead of point B on AC_1^B. More importantly, note that the utilization of the double quota will require him adjust his input factors. For example, he may need to travel more often, stay longer on the sea, or build a bigger vessel. If he does not adjust inputs, he will be able to utilize only part of the double quota and operate fishing activities at point M rather than C. Thus, the value of traded quota, in terms of output value, includes the value of productivity improvement (that shifts the cost from point A to M) and the value of input adjustment owing to economies of scale (that slides the cost from M to C). This paper shows how to decompose the value of traded quotas into these two terms. Section 2.4 formally characterizes the decomposition formula and Section 7.1 presents the results in the Norwegian fishery example.

2.2 Measuring Economies of Scale

To measure economies of scale, I use the output elasticity of average cost that measures the percentage change in average cost when output increases by one percent: $\psi \equiv \frac{dAC}{dq} \cdot \frac{q}{AC}$. Negative elasticities are equivalent to economies of scale, whereas positive elasticities mean diseconomies of scale.

Another useful measure is the output elasticity of total cost that gives the percentage change in total cost when output increases by one percent: $\phi \equiv \frac{dC}{dq} \cdot \frac{q}{C}$. This measure is effectively the ratio of marginal cost to average cost. For every differentiable cost function, the two elasticities differ just by one unit:

$$\psi \equiv \frac{dAC}{dq} \cdot \frac{q}{AC} = \frac{MC - AC}{AC} = \phi - 1,$$

where AC denotes average cost and MC denotes marginal cost. Thus, an economy of scale (decreasing average cost) is equivalent to $\psi < 0$ or $\phi < 1$. A diseconomy of scale is equivalent to $\psi > 0$ or $\phi > 1$. A constant economy of scale means $\psi = 0$ or $\phi = 1$. Note that the results apply for every first differentiable cost function, regardless whether the cost is (locally) convex.

Figure 3 graphically illustrates the relations between average cost, marginal cost, and output elasticities of average cost and of total cost. The curve of output elasticity of average cost (ψ) always cuts the x-axis at constant-economy-of-scale level. If the cost is strictly convex, that constant-economy-of-scale is the minimum level of average cost, e.g. point M. Any point on the average cost to the left of M exhibits an economy of scale: decreasing average cost. These

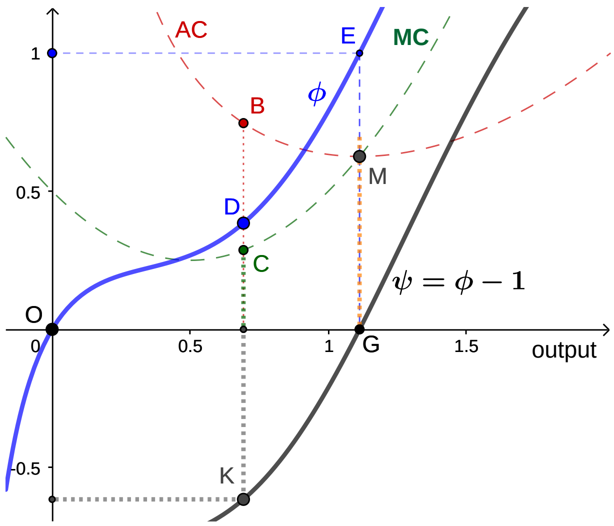


Figure 3: Measuring economies of scale using output elasticities of costs

Note: Two measures can be used: output elasticity of total cost ϕ (solid blue) and output elasticity of average cost ψ (solid black). They are plotted together with average cost curve AC (dash red) and marginal cost curve MC (dash green) on the graph. Important properties: (i) $\psi = \phi - 1$, (ii) $\phi = \frac{\text{marginal cost}}{\text{average cost}}$, and (iii) economies of scale $\Leftrightarrow \phi < 1 \Leftrightarrow \psi < 0$.

points are also where average cost is above marginal cost, negative ψ, or inelastic ϕ ($\phi < 1$). On the other hand, points to the right of M on the average cost curve has a diseconomy of scale, below marginal cost, positive ψ, or elastic ϕ.

2.3 Estimating Using Production Data

As shown, the two measures ψ and ϕ have one-to-one relation and one just needs to estimate either of them. Ideally, to estimate either ψ or ϕ, one would need data on cost and estimate the cost function. However, information on production cost is rarely observed. One even may not observe the input prices. I now discuss the method to estimate ϕ using data on output and input quantity only. The method relies on the standard definition of a long run cost with the following assumptions.

Assumption 1. Input prices are exogenous.

Assumption 2. All inputs are variable.

Assumption 3. Firms allocate inputs to minimize each own total cost.

Assumption 4. Production function $Q_{it}(\boldsymbol{X}_{it})$, where $\boldsymbol{X}_{it}$ is an input vector of a firm i at time t, is continuous and twice differentiable.

Proposition 1 (Cost elasticity and input elasticity)**.** *Under Assumptions 1–4, output elasticity of total cost is the reciprocal of the sum of all input elasticities of output. That is, let X be an* *input in the set of all variable inputs $\mathbb{X}$, then*

$$\phi_{it} = \left(\sum_{X \in \mathbb{X}} \theta_{it}^X \right)^{-1}, \text{ where } \theta_{it}^X = \frac{dQ_{it}}{dX_{it}} \cdot \frac{X_{it}}{Q_{it}}.$$

Proposition 1 is an important result that helps connect with the current literature on production function estimation and implies that we can use data on output and input quantity to estimate ϕ (the cost elasticity of total cost and a measure of economies of scale). Specifically, one can use various methods to estimate the production function $Q_{it}(\boldsymbol{X}_{it})$, calculate every input elasticities of output θ_{it}^X, and thus economy of scale ϕ. Note that the Proposition holds for *any* production function $Q_{it}(\boldsymbol{X}_{it})$. It can be in any form and have Hicksian neutral productivity and/or factor-augmenting productivity.

In the empirical part of this paper, Section 4.2, I discuss the use of the proxy variable approach (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg, Caves and Frazer, 2015)) with extra assumptions on the productivity evolution to identify a translog production function and then recover the economy of scale ϕ. I also discuss alternative identification assumptions and report the results if I estimate the production function using other methods: OLS with fixed effects and the dynamic panel approach by Blundell and Bond (1998) and Blundell and Bond (2000).

I now discuss the role of Assumptions 1–4 and present the proof of the proposition. Assumption 1 implies a firm's input price does not depend on input quantity. So, variation in input prices comes from exogenous factors rather than firms' input usage. This rules out market power in input markets. If the input market has market power, the relation between output elasticity of cost and input elasticities of output involves the price elasticity of input demand. Section 2.5.1 derives this relation. In that case, one would need additional information on the price elasticity of input demand to estimate ϕ.

Assumption 2 implies the cost function is the long run cost, because all inputs are variable. If there is a fixed input, the sum of output elasticities of inputs are still the main interest and there are two ways to interpret it. One way is to exclude the output elasticity of the fixed input in the formula in Proposition 1 and interpret ϕ as the output elasticity of total *variable* cost. phi in this case is the ratio of marginal cost to average *variable* cost. In the second way, assuming the fixed input has a dynamic implication on total cost, Section 2.5.2 shows the sum of all output elasticities of input (including the fixed-input elasticity) is the ratio of total expected average variable cost *and average adjustment cost* to marginal cost. Hence, ϕ can be still interpreted as the ratio of marginal cost to average total cost.

Assumption 3 is in fact the standard definition of a long run cost. Together with Assumptions 1 and 2, it assumes each firm solves the following cost minimization problem to design an input mix for a targeted output q_{it}:

$$C_{it}(q_{it}) = \min_{\boldsymbol{X}_{it}} \boldsymbol{W}_{it}^{\top}\boldsymbol{X}_{it} \text{ subject to } q_{it} \leq Q_{it}(\boldsymbol{X}_{it}),$$

where $\boldsymbol{X}_{it}$ is the input vector, $\boldsymbol{W}_{it}$ is the input-price vector, and $Q_{it}(\boldsymbol{X})$ is the production technology function.

Proof. With Assumption 4, the Lagrangian function of the cost-minimization problem is

$$\mathcal{L}_{it} = \boldsymbol{W}_{it}^{\top}\boldsymbol{X}_{it} + \lambda_{it}(q - Q_{it}(\boldsymbol{X}_{it})).$$

The first-order condition for any input $X \in \mathbb{X}$ is

$$\frac{\partial \mathcal{L}_{it}}{\partial X_{it}} = W_{it} - \lambda_{it} \cdot \frac{\partial Q_{it}}{\partial X_{it}} = 0.$$

Rearranging terms and multiplying by $\frac{X_{it}}{Q_{it}}$, we get

$$\frac{W_{it}X_{it}}{Q_{it}} \cdot \frac{1}{\lambda_{it}} = \frac{\partial Q_{it}}{\partial X_{it}} \cdot \frac{X_{it}}{Q_{it}}.$$

Because this relation applies for every input $X \in \mathbb{X}$, I sum all these relations side by side to get

$$\frac{\sum_{X\in\mathbb{X}}(W_{it}X_{it})}{Q_{it}} \cdot \frac{1}{\lambda_{it}} = \sum_{X\in\mathbb{X}} \left(\frac{\partial Q_{it}}{\partial X_{it}} \cdot \frac{X_{it}}{Q_{it}} \right).$$

In this equation, $\sum_{X\in\mathbb{X}}(W_{it}X_{it})$ is total cost and λ is marginal cost, because $\frac{dC_{it}}{dq} = \frac{\partial \mathcal{L}^*_{it}}{\partial q} = \lambda_{it}$. Hence, the left hand side is the ratio of average cost to marginal cost, or reciprocal ϕ. The right hand side is the sum of all input elasticities of output, which implies the Proposition 1. □

Although the Proposition 1 relies on the cost minimization assumption, note that it does not preclude the profit maximizing goal of the firms. Section 2.5.3 shows the compatibility and co-existence of the profit maximization and cost minimization. The key is in several market structure environments, designing an input mix to maximize profit is equivalent to designing an output to maximize profit and optimizing an input mix that miminizes total cost subject to the targeted output. Indeed, showing the compatibility between profit maximization and cost minimization in several market structure environments is also one of the methodological contributions of this paper. Literature on production function estimation has assumed firms maximize profits in a perfectly competitive output market; see Olley and Pakes (1996); Levinsohn and Petrin (2003); Ackerberg, Caves and Frazer (2015); Doraszelski and Jaumandreu (2018); Gandhi, Navarro and Rivers (2020). Section 2.5.3 shows such assumption is just a special case and stronger than Assumption 3.

2.4 Application: Decomposing the Value of Trading

Assumption 5. Production function contains only Hicksian productivity that is log-additive. That is, the production function is $Q_{it} = \mathbb{F}(\mathbf{X}_{it}, \omega_{it}) = F(\mathbf{X}_{it}) \exp(\omega_{it})$, where ω_{it} is the scalar logged productivity.

Proposition 2. *Under Assumptions 1–5, the output-based value of traded quotas can be decomposed into the utilization of quota due to economies of scale and the utilization due to productivity improvement. That is,*

$$\textit{quota} \equiv \Delta Q_{it} = \underbrace{\mathbb{F}(\mathbf{X}_{it}, \omega_{it}) - \mathbb{F}(\mathbf{X}_{i,t-1}, \omega_{it})}_{\textit{economies of scale } \Delta Q|_{\Delta \mathbf{X}}} + \underbrace{\mathbb{F}(\mathbf{X}_{i,t-1}, \omega_{it}) - \mathbb{F}(\mathbf{X}_{i,t-1}, \omega_{i,t-1})}_{\textit{productivity change } \Delta Q|_{\Delta \omega}}.$$

This result is obtained by decomposing the change in output between $t-1$ and t and notice this change is the traded quota. If the production function contains only Hicksian productivity, then the output elasticity of input $\theta^X = \frac{\partial \ln Q}{\partial \ln X}$ does not depend on productivity ω. Following Proposition 1, the cost elasticity ϕ will not depend on ω either. Hence, the same input choices, regardless of productivity levels, generate the same cost elasticity or economy of scale. The term of output change due to input adjustment captures the change along the average cost curve due to economies of scale, whereas the term of output change due to productivity improvement is the change due to the shift in the average cost curve.

I apply this formula to decompose the value of trading in the case of cap and trade in Norwegian cod fishery. Section 7.1 shows the results.

2.5 Notes and Extensions

The method to estimate economies of scale using an estimated production function (Proposition 1) relies on Assumptions 1–4. This section concludes the theoretical foundation of economies of scale by discussing caveats and extensions if these assumptions are relaxed.

2.5.1 Elasticities If Input Markets Have Market Power

Assumption 1 implies a perfectly competitive input market. I now discuss the relation between output elasticity of cost and input elasticity of output if input price depends on the input usage of the firms. In this case, the cost minimization problem in the production stage will be:

$$C_{it}(q_{it}) = \min_{\boldsymbol{X}_{it}} \sum_{X \in \mathbb{X}} W_{it}^X(X_{it}) X_{it} \text{ subject to } q_{it} \leq Q_{it}(\boldsymbol{X}_{it}), \tag{1}$$

where $W_{it}^X(\cdot)$ is the input price function of the X-input use. The Lagrangian function of the cost-minimization problem is

$$\mathcal{L}_{it} = \sum_{X \in \mathbb{X}} W_{it}^X(X_{it})X_{it} + \lambda_{it}(q - Q_{it}(\boldsymbol{X}_{it})). \tag{2}$$

The first-order condition for any input $X \in \mathbb{X}$ is

$$\frac{\partial \mathcal{L}_{it}}{\partial X_{it}} = W_{it}^{X'} X_{it} + W_{it}^X - \lambda_{it} \cdot \frac{\partial Q_{it}}{\partial X_{it}} = 0. \tag{3}$$

Summing this relation for all inputs and making a few algebra transformation, we arrive in:

$$\phi_{it} = \left(1 + \sum_{X \in \mathbb{X}} \eta_X \cdot \frac{W_{it}^X X_{it}}{C_{it}}\right) \left(\sum_{X \in \mathbb{X}} \theta_{it}^X\right)^{-1}, \tag{4}$$

where η_X is the price elasticity of demand for input X, i.e. $\eta_X \equiv \frac{dW^X}{dX} \cdot \frac{X}{W^X}$.

So, the output elasticity of cost is the ratio of average price elasticity of input demand, weighted by the share of input cost in total cost, to total input elasticities of output.

2.5.2 The Case of Dynamic Inputs

This extension relaxes Assumption 2 and quantifies the cost elasticity when an input has adjustment costs and dynamic implications on future cost values. Consider the classical cost minimization in a dynamic context in which capital K_{it} is dynamic and adjusted by endogenous investment level $I_{i,t-1}$ whereas labor L_{it} is variable. So, the capital evolves as $K_{it} = \delta K_{i,t-1} + I_{i,t-1}$ and the adjustment cost depends on both investment level and the capital state, $A(I_{i,t-1}, K_{i,t-1})$. For simplicity, I drop the notation i in this section. The dynamic cost minimization problem is

$$V(K_{t-1}, \Omega_t) = \min_{I_{t-1}, L_t} rI_{t-1} + wL_t + A(I_{t-1}, K_{t-1}) + \beta E[V(K_t, \Omega_{t+1})|\Omega_t], \tag{5}$$

$$\text{subject to } \; Q(K_t, L_t) \geq q_t, \tag{6}$$

$$K_t = \delta K_{t-1} + I_{t-1}. \tag{7}$$

Note that we can rewrite this problem into

$$V(K_{t-1}, \Omega_t) = \min_{I_{t-1}, L_t} r \cdot (K_t - \delta K_{t-1}) + w \cdot L_t + A(K_t - \delta K_{t-1}, K_{t-1}) + \beta E[V(K_t, \Omega_{t+1})|\Omega_t].$$

So, we can consider an equivalent problem with endogenous choices of capital and labor:

$$V(K_{t-1}, \Omega_t) = \min_{K_t, L_t} rK_t + wL_t + \mathcal{A}(K_t, K_{t-1}) + \beta E[V(K_t, \Omega_{t+1})|\Omega_t],$$
$$\text{subject to } \; Q(K_t, L_t) \geq q_t. \tag{8}$$

Lemma 1 (The dynamic version of the summed output elasticities of inputs)**.** *In a dynamic cost minimization with adjustment costs such that $\mathcal{A}(K^*_{t+1}, K^*_t) = K^*_t \cdot \frac{\partial \mathcal{A}}{\partial K_{t+1}} + K^*_t \cdot \frac{\partial \mathcal{A}}{\partial K_t}$, we have*

$$\frac{AVC_t + E[AAC_{t+1}|\Omega_t]}{MC_t} = \theta_{L_t} + \theta_{K_t},$$
$$\textit{where } \; AVC_t = \frac{rK_t + wL_t}{Q_t},$$
$$AAC_{t+1} = \frac{\mathcal{A}(K_{t+1}, K_t)}{Q_t}.$$

Note that $AVC_t = E[AVC_t|\Omega_t]$ and $MC_t = E[MC_t|\Omega_t]$. Intuitively, we have a dynamic equivalent version for the Proposition 1: The sum of all input elasticities of output is the ratio of expected average cost to marginal cost, as defined by the ratio of total variable cost and expected adjustment cost to marginal cost.

2.5.3 Compatibility of Profit Maximization and Cost Minimization

I now show that the cost minimization behavior of firms—Assumption 3—does not preclude the profit maximization behavior, which is often the assumed goal of the firms. Indeed, the two behaviors are compatible in a variety of market structures: perfect competition, Cournot competition, Cournot competition in the presence of bargaining power stemming from output size, price differentiation due to output-independent quality adjustment, and co-influence of output and input in a generalized cost function. To see why, let me first describe the two decision making processes of a firm.

Definition 1. The input choice problem of a firm to maximize its profit is

$$[\text{Problem 1:}] \quad \max_{\boldsymbol{X}_{it}} \mathcal{P}_{it}(Q_{it}(\boldsymbol{X}_{it})) \cdot Q_{it}(\boldsymbol{X}_{it}) - \mathcal{G}(\boldsymbol{X}_{it}),$$

where $\mathcal{P}_{it}(\cdot)$ is the firm individual output price function, $Q_{it}(\boldsymbol{X}_{it})$ is the production function, $\mathcal{G}(\cdot)$ is the generalized cost function.

Definition 2. The two-step decision problem where the firm decides output level to maximize profits in the first stage and decides inputs to minimize production cost of producing the targeted output in the second stage is

$$\text{[Problem 2:]} \quad \max_{q_{it}} \ \mathcal{P}_{it}(q_{it}) \cdot q_{it} - C(q_{it}) \text{ in stage 1, and}$$
$$C(q_{it}) = \min_{\boldsymbol{X}_{it}} \ \mathcal{G}(\boldsymbol{X}_{it}) \text{ subject to } Q_{it}(\boldsymbol{X}_{it}) \geq q_{it} \text{ in stage 2.}$$

Proposition 3 (Compatibility of profit maximization and cost minimization)**.** *Assume differentiability, concavity of profit function, and convexity of cost function. Problem 1 and Problem 2 are equivalent for the following market environments:*

i) Perfect competition in the output market.

ii) Cournot competition in the output market.

iii) Bargaining power stemming from output size. That is, $\mathcal{P}_{it}(q_{it}) = P(\mathcal{Q}(q_{it}), q_{it})$*.*

iv) Price differentiation stemming from an endogenous effort that affect quality but not quantity. That is, $\mathcal{P}_{it} = P(\mathcal{Q}(q_{it}), q_{it}, H(e_{it}))$ *and* $\mathcal{G}(\cdot) = \mathcal{G}(\boldsymbol{X}_{it}, e_{it})$*, where* e_{it} *is the endogenous efforts.*

v) Co-influence of output and input in cost function. That is, $\mathcal{G}(\cdot) = G(Q_{it}(\boldsymbol{X}_{it}), \boldsymbol{X}_{it})$*.*

If firms can differentiate their prices by allocating inputs to directly adjust product quality $H(\boldsymbol{X}_{it})$*, i.e.* $\mathcal{P}_{it} = P(\mathcal{Q}(q_{it}), q_{it}, H(\boldsymbol{X}_{it}))$*, then the two problems are not equivalent in general.*

Appendix B shows the proof. The main intuition is that the cost minimization problem lies in the production stage rather than being a whole single goal of the firm. The cost minimization problem aims to design inputs to produce the targeted output rather than to design the output that minimizes cost.

Proposition 3 also contributes to the production function literature that has traditionally assumed perfect competition in the output market to use the proxy variable approach to estimate a production function Olley and Pakes (1996); Levinsohn and Petrin (2003); Ackerberg, Caves and Frazer (2015); Doraszelski and Jaumandreu (2018); Gandhi, Navarro and Rivers (2020). I show that such assumption is only a special case and stronger than assuming cost minimization behavior in the input choice stage and profit maximization in the output choice stage.

2.5.4 Relation to Markups

Although I am the first that derives elasticities of costs to measure economies of scale and shows how to estimate it, I am not the first that exploits the cost-minimization condition of the input allocation problem. Indeed, an emerging literature on IO and macroeconomics has used this condition to estimate markups. This approach has been called production approach to distinguish it from the demand approach. In the demand approach championed by Bresnahan (1989) and Berry, Levinsohn and Pakes (1995), the markup estimation relies on assumptions on utility maximizing behavior of consumers and on how firms compete (for example, Bertrand-Nash price competition or Cournot quantity competition). This demand approach requires data on (at least) product market shares and product characteristics. In contrast, the production approach, established by De Loecker and Warzynski (2012), is posited on the cost minimization by producers and requires data on individual firm output, input, and a variable input's expenditure share in revenue. Examples of applications of this production approach include Braguinsky et al. (2015); De Loecker et al. (2016); De Loecker, Eeckhout and Unger (2020). I now discuss the relation between my output elasticity of cost and the markup in this production approach literature.

Let me begin with the review of the production approach. The production approach to estimate markups also relies on the cost minimizing problem. However, De Loecker and Warzynski (2012) rewrite the first order condition into

$$\frac{W_{it}X_{it}}{P_{it}Q_{it}} \cdot \frac{P_{it}}{\lambda_{it}} = \frac{\partial Q_{it}}{\partial X_{it}} \cdot \frac{X_{it}}{Q_{it}}, \tag{9}$$

where P_{it} is the output price. Hence, the markup ratio $\mu \equiv \frac{P_{it}}{\lambda_{it}}$ can be calculated through:

$$\mu_{it} = \frac{\theta_{it}^X}{\alpha_{it}^X}, \tag{10}$$

where α_{it}^X is the share of expenditure on input X in total sales, i.e. $\alpha_{it}^X \equiv \frac{W_{it}X_{it}}{P_{it}Q_{it}}$. Using this relation, De Loecker and Warzynski (2012) show firm-level markups can be inferred using production data. Specifically, one would need (i) data on output and input to estimate the production function and the output elasticity of one (or more) variable input(s) θ_{it}^X and (ii) data on expenditure share α_{it}^X, which is often available in the financial statement of the firms.

To think about the relation between output elasticity of cost ϕ and markup μ, we now can use Proposition 1 and equation (10) to get

$$\sum_{X\in\mathbb{X}} \theta_{it}^X = \sum_{X\in\mathbb{X}} \mu_{it}\alpha_{it}^X = \mu_{it} \sum_{X\in\mathbb{X}} \alpha_{it}^X = \mu_{it} \cdot \frac{C_{it}(Q_{it})}{P_{it}Q_{it}} \tag{11}$$

$$\implies \phi_{it} \cdot \mu_{it} = \frac{PQ}{C} \tag{12}$$

However, I want to emphasize that this relation in fact exists *without* the cost minimization assumption. Indeed, for every differentiable cost function, we have $\phi \cdot \mu = \frac{dC}{dQ} \cdot \frac{Q}{C} \cdot \frac{P}{dC/dQ} = \frac{PQ}{C} \equiv \frac{revenue}{cost}$, where $\mu \equiv \frac{P}{MC}$ is the markup. So, it is the relation between ϕ and θ or between μ and θ that requires the cost minimizing behavior in the input choice decision.

In my application of examining economies of scale to study cap and trade in fishery, I could have estimated the markup using the above approach, but I unfortunately do not observe the revenue cost ratio nor the revenue share of expenditure on input.

3 Background on Norwegian Cod Fishery Regulations

This paper explores the value of cap and trade (CAT) owing to productivity improvement and economies of scale. To do that, I exploit the policy transition from non-tradable cap to CAT in a subset of vessels in Norwegian cod fishery. This transition allows me to (i) focus on the impacts of trading rather than being confounded by the role of caps, and (ii) identify the causal impacts using the difference-in-difference (DID) strategy. Before discussing the empirical strategy as well as the proxy variable approach to estimate a production function, I provide a background on fishery regulations and Norwegian cod fishery below.

3.1 An Overview of Regulations in Fisheries

Before going through the detail context of the Norwegian cod fishery, understanding the big picture of regulations in fisheries in general is useful. Similar to other common goods and public goods such as water, forests, oil, atmosphere (e.g. air pollution as a bad public good), fishery in oceans suffers the tragedy of the commons in an open-access and unregulated environment. That is, individuals in an open access system pursue their own self-interest and neglect the well-being of society, leading to overconsumption and ultimately causing depletion of the resource (Tietenberg, 2003; Costello et al., 2010; Stavins, 2011).

As a result, hundreds of fisheries have followed the lead of other natural resources and have transitioned from open access systems to property right management. The first reforming property right management is limited entry, under which only fishermen that own permits are entitled to participate in the fishery. Together with limited entry, regulators also employ other command-and-control tools—fishing season limitations, gear restrictions, area closures—to limit fishing activities. However, resource rents may be dissipated by excessive capital investment, redundant effort, or inefficient timing of harvest; see Costello et al. (2010); Stavins (2011). Note that in contrast to public goods such as air, common goods are rivaled in consumption. Fishermen may race to catch as much as possible during a limited fishing season. In several cases, such CAC approaches led to worse outcomes than open-access fisheries.

Hence, regulators have stepped towards the next reform: catch share management. Under catch share management, the regulator defines the total allowable catch (TAC) of the whole fishery annually and each vessel owns a proportion of the TAC (catch share or quota) that entitles the vessel to catch up to the tonnage values of the quota. The catch share management has two types of arrangements: individual vessel quotas (IVQ) and individual transferable quotas (ITQ). In the IVQ system, a quota is attached to a boat and are not separately transferable.[6] If one wants to buy a quota, he has to buy the license and the boat. This approach is simply

[6]Specifically, a vessel has its own license that entitles the fishery entry and the regulator defines a quota on the license. The license can be transferred and the unique quota attached with the license will thus follow the license. However, the rule is "one vessel one license (thus one quota)." Hence, "transferability" of quotas in the IVQ system is effectively transferability of licenses and boats, which is not the trading property of the cap in a cap-and-trade program.

a command-and-control regulation in which a non-transferable cap of output or emissions is set on a facility. Although the cap can move from an owner to another owner by buying out the facility, the new cap is not allowed to combine with the existent cap to run on one facility. In contrast, the ITQ system mimics the cap-and-trade programs for air pollution. In the ITQ management, the regulator allocates shares of the harvest to individuals (individual vessels with active licenses) in the first time using a grandfathering rule and allows fishermen to trade those shares after the initial allocation.

Over the past three decades, many countries have employed catch share management. As of 2008, more than 140 fisheries in the world are managed by ITQs. Country examples include the Netherlands, Canada, Iceland, New Zealand, the U.S., Australia, Argentina, Chile, and so on (Costello, Gaines and Lynham, 2008; Chu, 2009; Costello et al., 2010). For example, New Zealand was the first country to adopt ITQs as a national policy in 1986.[7] Their ITQ system is very flexible in which quotas can be divisibly traded, sold or leased, and hold in perpetuity, establishing a well-functioning market of quotas (Newell, Sanchirico and Kerr, 2005). On the other hand, many territories have restricted trading for concerns about consolidation of quotas and catches. Several countries used to adopt or currently adopt non-tradable quotas (IVQs) such as Norway, Denmark, Sweden, the U.K., Peru (Asche et al., 2008; OECD, 2013). In the U.S., most of fisheries have several different trading restrictions such as consolidation caps, sunset provisions, and restrictions on leases or permanent trades (Grainger and Parker, 2013).

Recently, Norway has gradually switched from IVQ to ITQ by allowing quota trading in certain groups of vessels in the cod fishery, the most valuable capture species in Norwegian fishery. This paper exploits this policy transition to evaluate the impacts of ITQ relative to IVQ, or cap-and-trade relative to non-tradable cap in the fishery context. I now closely discuss the regulation transition in Norway.

3.2 Norwegian Cod Fishery

The regulation transition focuses on the coastal fleet in the Norwegian cod fishery. Cod is the most valuable catch in Norwegian fishing industry. As of 2019, the primary value of cod fishing was 7.2 billion Norwegian dollars (850 million US dollars), or 34%, followed by mackerel (12%) and herring (12%).[8] Cod in Norwegian sea is Atlantic cod (*Gadus morhua*). They can live for 25 years, attain reproductive maturity between ages two and six, grow to 2m long and 40kg (88lbs).[9] Appendix A shows how they look like and the distribution along Norwegian coast.

[7] Although the Netherlands, Canada, and Iceland were the first countries to adopt ITQs in the late 1970s, New Zealand became the world's largest ITQ system in 1986 by employing the system nationally.

[8] SSB `https://www.ssb.no/jord-skog-jakt-og-fiskeri/artikler-og-publikasjoner/fisket-verdt-21-milliardar-kroner`.

[9] See Animal Diversity Web, University of Michigan, `http://animaldiversity.org/accounts/Gadus_morhua/` and Norwegian Institute of Marine Research, `https://www.hi.no/en/hi/temasider/species/costal-cod--north-of-the-62-latitude`

Vessels that are allowed to fish cod in Norway are divided into two fleets: deep-sea fleet and coastal fleet. The deep-sea fleet typically consists of big commercial vessels that use active gears such as trawls and purse seines to find out the school of fish before putting the gear in the sea. The coastal fleet includes smaller vessels (less than 28 meters or have a cargo volume of less than 500m^3) that use passive gears such as yarns, long lines, hand lines, teine, net etc. that stand still in the sea and wait for the fish to reach the gear.[10] The coastal fleet has been closely monitored by the regulators, because this fleet consists of more than 2,000 vessels, contributing to main income and earnings in many communities along Norwegian coast (Nærings- og fiskeridepartementet, 2006). From 2016, the coastal fleet accounts for 97% of cod vessels and 70% of national cod quota; see Nærings- og fiskeridepartementet (2016, 2019). This paper focuses on this coastal fleet.

Table 1: Changes in the Norwegian management of the coastal fleet in the fishery of cod in the north of 62°N

Year	Event
1980s	Open fishery.
1990	Limited entry (closed fishery) with individual vessel quotas (IVQs).
2001	Length is recorded to legal length that is fixed, regardless of actual size upgrades.
	Fleet is divided into 4 legal length groups: 0–10.9, 11–14.9, 15–20.9, 21–27.9.
	Note that actual length has been less than 28m.
2003	Decommissioning scheme for coastal fleet up to 14.9m, from 1 Jul 2003 to 1 Jul 2009.
2004	Individual transferable quotas (ITQs) are introduced in certain license groups of the fishery.
	Quota can be transfered between vessels in legal length groups of 15–20.9m and 21–27.9m.
2005	Additional purchased quota has its life extended from 13/18 years to 20/25 years.
	Quota ceiling increases for groups 15m+.
2008	Quota trading is allowed for legal length group of 11–14.9m.
2008	Change from max length of 28m to max cargo of 300m^3.
2010	Change from max cargo of 300m^3 to max cargo of 500m^3.

Sources: Nærings- og fiskeridepartementet (2003, 2006, 2007, 2016, 2019); Armstrong and Clark (1997); Armstrong et al. (2014); Standal and Aarset (2008); Standal and Asche (2018).

Table 1 summarizes the regulation changes in the cod coastal fleet. Up to 1980s, the coastal fleet in the cod fishery was able to fish freely, without restriction of access and with spacious maximum national quotas. The reason was the agreed total quota in North-East Arctic cod fishery between Norway and the Soviet Union were set significantly higher than what would be considered sustainable. After many years of over-fishing, the demand for sustainability brought up a reduction in the total quotas agreed between Norway and the Soviet Union. In 1989, the coastal fleet quickly fished up the total quota, resulting in being halted on 18 April by the Directorate of Fisheries. Such decision was a bombshell because April was in the peak of the season and many communities even had not started to go fishing. Hence, a system

[10]Before 2008, vessels in the coastal fleet had a maximum length limit of 28 meters. From 2008, the length limit in the coastal fleet was removed and replaced by the maximum cargo limit of 300m^3 (and 500^3 in 2010).

for individual quotas became an immediate demand and quickly supported by the Norwegian Fisher's Association.

In 1990, the cod coastal fishery moved from free fishery (with a national cap) to a closed system with individual vessel quotas (IVQ). The aim was to avoid the fishing race that had happened in 1989. Based on a minimum catch requirement in historical years, a closed group was established. Vessels that did not satisfy the criterion could participate in an open group.[11] Vessels in the closed group were assigned individual catch shares (quotas).

Between 2001 and 2002, a length division of the coastal fleet (*Finnmarksmodellen*) was introduced to provide a fairer competition between vessels. The division categorized the coastal cod closed group into four length groups: 0–10.9m, 11–14.9m, 15–20.9m, 21–27.9m.[12] The closed group quota was divided into these four length groups before further distributed to individual vessels within a group. The intention was that vessels only competed with others within the same length group for their similar quotas. Furthermore, the physical lengths of vessels at this time were recorded into legal lengths that became fixed regardless of the size upgrades in future. The legal lengths became an attribute of the license and defined an individual quota share of a vessel in the license group.

Since 2004, quota trading has been allowed in order to reduce over-capacity and to increase profitability in the cod coastal fleet. The policy is known as a structural quota scheme (*strukturkvoteordningen*) in Norwegian regulation. The scheme allows vessels in certain license groups to trade their quotas. Due to the concern that quota trading would result in consolidation of quotas and catch into a few hands of fishermen, the trading scheme were implemented in vessels in only the two upper groups, 15–20.9 and 21–27.9, beginning on 1 January 2004. In 2007, after a review of the legislation, the scheme was expanded to cover vessels in the license group of 11–14.9, starting from 1 January 2008.

The scheme also imposes several restrictions on quota trading. First, vessels are only allowed to trade quotas within the same license group. Second, the vessel that sells the quota must exit the fisheries permanently (by being scrapped or sold). Third, a portion of the transferred quota (20%) must be deducted and given to the other vessels in the group. Fourth, the transferred quota is only valid for limited time, 13 years if being sold (15 years if the vessel is scrapped).[13] The final rule is a geographical restriction: Vessels in the south are not allowed to buy quotas from the north, although the North can buy quotas from the South.

Overall, the quota trading scheme in Norwegian cod coastal fishery has switched IVQs to ITQs in certain licensed length groups, 15–20.9 and 21–27.9 (and 11–14.9m), since 2004 (and 2008). As discussed in a pubic report by the Ministry of Trade and Industry (Nærings- og fiskeridepartementet, 2006), the goal of the transition is to reduce overcapacity and increase

[11]The open group is regulated by a total group quota rather than individual quotas and this group quota is substantially low, only about 5% to 10% of the cod fishery quota; see Nærings- og fiskeridepartementet (2019).

[12]Until 2008, vessels in the coastal fleet must be shorter than 28 meters.

[13]This valid duration was implemented for the scheme from 2004. After a legislation review in 2007, the valid durations became 20 years if being sold and 25 years if the vessel is scrapped.

profitability in the cod fishery. They justify that IVQs were able to prevent the expansion of further overcapacity, but IVQs would not give the industry incentives to remove existing overcapacity. Under an ITQ system, fishermen with a quota would like to have quasi-property rights to a certain proportion of the total quota. If overcapacity led to reduced profitability in the industry, quota could be bought and sold until costly overcapacity is gone. The least efficient vessels will be taken out of fishing in exchange for the more efficient ones increasing their catch. This vision of reducing overcapacity of the regulators explains the motivation for the scrapping condition that requires vessel that sells the quota give up its whole quota and exit the cod fishery permanently.[14]

This paper exploits the policy transition from IVQs to ITQs in only certain groups in the coastal cod fleet to study the effects of ITQs, relative to IVQs, on the vessel-level performance in the fishing market. The comparison between ITQs and IVQs is analogous to comparing cap-and-trade to command-and-control. In theory, cap-and-trade performs better to achieve economic efficiency goal by minimizing costs of compliance in the case of emissions reduction, or by maximizing the value of resources in the case of natural resources. In fishery, as revealed by the above stated vision of the Norwegian Ministry of Trade and Industry, ITQs are expected to increase efficiency and profitability. I will test this hypothesis by exploring the impacts of ITQs on productivity, fish sale prices, and production costs.

It is worth mentioning that the committee in 2003 proposed two measures to reduce the number of vessels in the coastal fleet: the ITQs system and the decommissioning scheme. Whereas the ITQs were implemented for vessels with legal lengths from 15m, the vessels below 15m (in terms of legal length) were subject to the decommissioning scheme in which regulator paid out the fisherman to buy back the vessel and license. The decommissioning scheme carried out from 1 July 2003 to 1 July 2009. This scheme led to a substantial reduction in the number of vessels in the control group, license of 0–14.9, which would raise a caveat for the interpretation of the empirical results using a difference-in-difference approach. The next section discusses the empirical methodology and identification in detail.

4 Empirical Methodology

The goal of this paper is to show trading in cap and trade offers cost reduction owing to productivity improvement and economies of scale. Because the quota is assigned at vessel level, I study the impact of trading at vessel level. I use difference-in-difference (DID) approach to compare the change in productivity and economies of scale between the treatment group and control group and between before and after the policy. Section 4.1 describes the DID strategy. Section 4.2 describes the empirical method to estimate time-varying and vessel-specific pro-

[14]The scrapping condition may imply that quotas are not traded divisibly. However, given that the trade happens between vessels within the same license group and vessels in a license group may have different quota due to different legal lengths, the whole policy scheme has certain degree of divisibility.

ductivity and economies of scale that are used as dependent variables in the DID regressions.

4.1 Estimating the Causal Effects of Cap and Trade

I estimate the effect of trading by exploiting the fact that only certain vessels are allowed to trade quotas. Vessels that hold licensed lengths below 11 meters are never allowed to trade and hence considered in a control group. The other vessels are in treatment groups with staggered adoption.[15] Licensed length group 2 (11–14.9m) can trade quotas from 2007 and licensed length groups 3 and 4 (15–20.9m and 21–27.9m) can trade quotas from 2004. My identification strategy is to estimate a DID specification comparing the difference between treatment and control groups and the pre-treatment and treatment-periods. This gives us an estimate of the intent-to-treat of the trading program. Because not every vessel in the trade-qualified group chooses to buy quotas, I also estimate the average treatment on the treated by looking at the difference between vessels that do acquire quotas and the others.

Intent-to-treat (ITT)

I estimate the following DID specification to identify ITT:

$$Y_{it} = \beta_{ITT} \textit{Trade Qualified}_{it} + \eta_i + \tau_t + \epsilon_{it}, \tag{13}$$

where Y_{it} is an outcome of interest, such as logged catch quantity, sale prices, productivity, economy of scale, and production factors of vessel i at time t. The variable $\textit{Trade Qualified}_{it}$ equals 1 if vessel i is in the trade-qualified group at time t, and zero otherwise. Equation (13) includes vessel fixed effects η_i to account for permanent differences in the operating skills of time-invariant ownership during the period 2001–2017. The model also includes time fixed effects τ_t to adjust for the average effects of time-varying factors (e.g. weather, sea temperature, stock levels, seasonality) that generate variation in the outcome of interest across all vessels.

The parameter of interest is β_{ITT}. It is a DID estimator that compares the change in outcome Y of trade-qualified vessels after trade qualified status to before, relative to the vessels that are not qualified for quota trading (licensed lengths below 11 meters). This is the intent-to-treat effect that measures the impact of the trade-qualification program.

The identification assumption of this DID approach is that, conditional on the fixed effects, differences between trade-qualified vessels and non-qualified vessels are on average similar for pre-trade-legalization and post-trade periods had the vessels not traded quotas. This assumption is untestable. However, the parallel trend assumption in the counterfactual condition can be

[15] The staggered DID estimate may be biased because treated individuals get treated at different time, making their presence in the sample unbalanced. Callaway and Sant'Anna (2020) and Goodman-Bacon (2021) suggest some modifications. In this paper, for the intent-to-treat of the trading program, we only have three treatment groups. So, we can illustrate the effects of each group separately relative to a clearly defined control group and time.

plausibly to believe if the two groups have parallel trend in the pre-trade-qualification period. I, hence, use the event study with lagged DID coefficients to test whether there is no difference in the outcome of interest between the two groups in pre-trade-qualification period, relative to the year the program is implemented. This event study also helps inform any anticipatory effects of the trade program. If fishermen expected the trade program would be enacted, they would adjust their production factors before the program officially enacted, showing a deviation in fishing outcomes from the parallel trend for years just right before the policy enforcement. Besides the lagged DID coefficients, I also include the leads coefficients in the event study to explore the dynamic effects of the policy. Results in Sections 6 and 7.1 confirm the parallel trend in the pre-trade period and show the effects of trading on catch quantity and revenue happened immediately after trading was allowed. Fishermen immediately utilize additional quotas by going fishing more often whereas take time to invest in capital (build bigger vessels).

Despite the pre-trade-period parallel trend verification, a threat to the identification of the causal effect is the spillovers of the trading impacts on vessels in the control group. Because the trading requires one of the two vessels in the transaction leave the fishery, vessels in the control group may benefit by facing fewer competitors both in the fishing ground and in the fish sales market. Specifically, vessels in the control group may catch more given lower congestion costs. Figure 4 in the next section shows the license group 0–10.9m, as a control group, goes fishing within 150km from the coast, which are distinctly distant from the fishing locations of vessels in the other groups, thereby alleviating the spillovers due to congestion costs. However, a small threat of spillovers in harvest may happen if fishing activities of big vessels in faraway locations in the ocean interfere the migration of cod. In that case, vessels in the control group may benefit from the fact that there would be fewer vessels, making the DID estimator underestimate the policy impact on trade-qualified vessels' harvest. Similarly, in the landing market, fewer competitors may help vessels in the control group sell their fish at higher prices than before, thereby causing the DID estimator underestimate the market-power-abuse impact of consolidation on fish prices.

Average treatment for the treated (ATT)

Because the trade-qualification program is not mandatory for all vessels in the trade-qualified group, the DID approach in equation (13) estimates the effect of the qualification program rather than the treatment effect of acquiring quotas through the program. I use DID to identify the average treatment for the vessels that do acquire quotas. Specifically, I estimate the following equation:

$$Y_{it} = \beta_{ATT} \mathit{Quota\ Acquisition}_{it} + \eta_i + \tau_t + \epsilon_{it}, \tag{14}$$

where $\mathit{Quota\ Acquisition}_{it}$ equals one for all periods after vessel i buys all quotas from another vessel in the trade-qualified group and zero otherwise.

The coefficient β_{ATT} measures the average treatment for the treated. It measures that the average changes in Y_{it} after a quota acquisition of the vessels that choose to acquire quotas, relative to the other vessels (including the unqualified trade vessels). The causal interpretation of this DID estimate for the ATT requires the vessels that acquire quotas have similar trends in fishing performances to the other vessels had the acquisition not happened. While this assumption cannot be completely tested, as in the case of intent-to-treat, I provide an event study that carefully looks at the changes in the vessel performance right before and after the quota acquisition to test the parallel trend patterns in the pre-acquisition period.

One can also instrument *Quota Acquisition*$_{it}$ using the trade-qualififed status *Trade Qualified*$_{it}$. The coefficient of such a DID instrumental variable approach measures the local average treatment effect (LATE), i.e. the average changes in Y_{it} from quota acquisition on "complier" vessels that will acquire quotas whenever they are qualified for trading. In this context with one-sided noncompliance, we only have either never-takers or compliers. Hence, the LATE would be another estimate of the average treatment effect on the treated. However, in contrast to β_{ATT}, the key identification assumption to interpret β_{LATE} as a causal effect is the exclusion restriction that requires the trade-qualification affect the outcome of interest only indirectly via an effect on trading execution. However, the Norwegian quota trading program by design sets a rule in which the quota acquirer may only take 80% of the quota and leave 20% equally shared to the other vessels in the group. Hence, there are vessels that would never trade but be able to gain higher quotas due to the program, causing a potential violation of the exclusion restriction condition. In this case, the DID estimation of 14 without IV, β_{ATT}, may give more convincible estimates of the quota acquisition impacts.

4.2 Estimating Productivity and Economies of Scale

I now discuss the estimation of a production function to obtain productivity and economies of scale. I consider the yearly production function of a vessel. Production factors include vessel size (length) K_{it}, crew size (labor) L_{it}, distance from the fishermen's municipality to major catch location D_{it}, and the number of trips in a year M_{it}. The reason is that the exact quota tonnage is set annually. Every year, the regulator decides the total allowable catch for the whole fishery and defines the conversion factor that converts a vessel's quota share to his quota tonnage. Quotas are not bankable. Hence, it is reasonable to assume the vessel's owner decides input factors that are critical for a production year instead of a trip.

I consider the standard production function, as in Assumption 5, with four factors and Hicks-neutral productivity ($\exp(\omega_{it})$):

$$Q_{it} = F(K_{it}, L_{it}, D_{it}, M_{it}; \beta) \exp(\omega_{it}). \tag{15}$$

In an empirical framework, we observe logged output y_{it} and assume $y_{it} = \ln Q_{it} + \epsilon_{it}$, where

ϵ_{it} is an exogenous unexpected shock to production. I estimate the following equation:

$$y_{it} = f(k_{it}, l_{it}, d_{it}, m_{it}; \beta) + \omega_{it} + \epsilon_{it}, \tag{16}$$

where $k_{it}, l_{it}, d_{it}, m_{it}, \omega_{it}$ are logged inputs and logged neutral productivity.

Because econometricians do not observe logged productivity ω_{it}, estimation has two challenges. First, an owner of a vessel chooses his inputs based on the realization of ω_{it}, causing simultaneity bias. Second, vessels that exit over time are those that have low productivity, causing selection bias. To address these challenges, literature has suggested four solutions: using input prices as instrument variables, using OLS with fixed effects, using the proxy variable (control function) approaches (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg, Caves and Frazer, 2015; Gandhi, Navarro and Rivers, 2020), and using dynamic panel approaches (Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998, 2000). In this paper, I use the proxy variable approach for the main results, and OLS with fixed effects and the dynamic panel approach as sensitivity checks.

The proxy variable approach in this paper relies on Assumptions 1–5 and the following assumptions:

Assumption 6. The conditional demand for a proxy variable input, $m_{it} = \mathcal{M}_{it}(k_{it}, l_{it}, d_{it}, \omega_{it})$, is strictly monotone in a single unobservable ω_{it}.

Assumption 7. The productivity ω_{it} evolves in a Markovian process:

$$\omega_{it} = g(\omega_{i,t-1}, \textit{Trade Qualified}_{i,t-1}) + \xi_{it},$$

where ξ_{it} is an exogenous shock in productivity that is uncorrelated with information at $t-1$, that is $\boldsymbol{E}[\xi_{it}|\mathbb{I}_{t-1}] = 0$.

All these assumptions are standard in the literature on the proxy variable approach, except two exceptions. First, Assumption 3 is weaker than assuming perfect competition in the literature, as discussed in Section 2.5.3. Second, Assumption 2 assumes all inputs are variable, which requires different moment conditions for the estimation. I discuss it in Section 4.3 (Identification).

Under the Assumption 6, the conditional demand for a proxy input is inverted and substituted into the production function to get

$$y_{it} = \underbrace{f(k_{it}, l_{it}, d_{it}, m_{it}; \beta) + \overbrace{\mathcal{M}_{it}^{-1}(k_{it}, l_{it}, d_{it}, m_{it})}^{\omega_{it}}}_{\equiv \chi_{it}} + \epsilon_{it}. \tag{17}$$

Hence, in the first stage, I run the following regression

$$y_{it} = \chi_{it}(k_{it}, l_{it}, d_{it}, m_{it}) + \epsilon_{it}, \tag{18}$$

to obtain estimates of expected output $\widehat{\chi}_{it}$. Although the coefficients in the first stage are not the coefficients for the production function, the goal is to separate productivity from shock ϵ_{it}.

In the second stage, I estimate the production function coefficients β using Assumption 7. With this assumption, I can compute productivity for any value of β, using $\omega_{it}(\beta) = \widehat{\chi}_{it} - f(k_{it}, l_{it}, m_{it}; \beta)$. By nonparametrically regressing $\omega_{it}(\beta)$ on its lag and event variables affecting productivity, I recover the exogenous shock $\xi_{it}(\beta)$. With the timing of the firm's decisions on k, l, d, m and the uncorrelation between exogenous shock ξ_{it} and past information, I can use the following moments to estimate β:

$$\mathbf{E}[\xi_{it}(\beta)x_{i,t-1}] = 0, \tag{19}$$

where x is the input vector (k, l, d, m).

For the specification of the production function, I use the translog function to obtain individual-specific and time-varying input elasticities of output. Given GMM estimates of β in the second stage, the output elasticity for capital, for example, is given by

$$\widehat{\theta}_{it}^{K} = \hat{\beta}_k + 2\hat{\beta}_{kk}k_{it} + \hat{\beta}_{kl}l_{it} + \hat{\beta}_{kd}d_{it} + \hat{\beta}_{km}m_{it}. \tag{20}$$

After getting all input elasticities of output, I calculate economies of scale $\widehat{\phi}_{it} = (\sum_{X\in\mathbb{X}} \widehat{\theta}_{it}^{X})^{-1}$.

4.3 Identification

Identifying the causal impacts of the trading policy on productivity and production cost relies on two main identifying strategies. The first is to identify productivity and economies of scale using either the proxy variable approach or the dynamic panel approach. The second is to identify the causal impacts of trading using the difference-in-difference identifying conditions.

4.3.1 Identifying the Production Function and Economies of Scale

Under the proxy variable approach in this paper, the gross output production function and economies of scale are identified upon the Assumptions 1–7. Compared to a variant of these assumptions under which Ackerberg, Caves and Frazer (2015) and Gandhi, Navarro and Rivers (2020) show the production function is identified, I make three differences. First, I show the cost minimization behavior holds in market environments other than perfect competition in the output market; see Section 2.5.3. So, we do not need to assume price takers in the output market to identify the production function.

Second, I assume capital is variable, whereas the literature has assumed capital is predetermined and dynamic (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg, Caves and Frazer, 2015; Gandhi, Navarro and Rivers, 2020). With this difference, instead of using the moment condition $E[k_{it}\xi_{it} = 0]$, I use the condition $E[k_{i,t-1}\xi_{it} = 0]$. As discussed in Ackerberg, Caves and Frazer (2015), if capital k_{it} is predetermined and depends on the investment adjustment in the previous period $t-1$, the current capital will be uncorrelated with exogenous shock ξ_{it}. For a variable input x_{it} that is chosen today and correlated with today's productivity, the past value of the variable input can be used instead in the moment condition. The coefficients of input variables are not identified in the first stage, but they are identified in the second stage of GMM upon the uncorrelation between the exogenous shock of innovation ξ_{it} and past information.

Third, I estimate the gross output production function, whereas Ackerberg, Caves and Frazer (2015) estimate the value added production function.[16] Whereas Ackerberg, Caves and Frazer (2015) exclude the proxy input m_{it} in the production function, including it in the production function is important in my case, because I need to obtain all elasticities of inputs with respect to output. The exclusion of the proxy input in the production function was due to the concern about the nonidentification of the gross output production function using the proxy variable approach as being raised by Gandhi, Navarro and Rivers (2020). However, Gandhi, Navarro and Rivers (2020) formally show the nonidentification of the gross output production function in the absence of time-series variation in relative prices (of input and/or output). As a result, the cross-sectional variation in the proxy variable is not enough to help identify the gross output production function. As shown in the below descriptive statistics (Figure 4), this is not the case in this paper: I consider the number of fishing trips in a year of a vessel as the proxy variable and this variable has variation both across vessels and time (year).

Ultimately, the key assumption that leads to the nonidentification of the gross output production function shown by Gandhi, Navarro and Rivers (2020) is the Assumption 6 that assumes productivity is the only scalar unobservable and the conditional demand for the proxy input is strictly monotone in productivity. The dynamic panel approach can avoid this assumption and can be used to estimate a gross output production function. This approach also allows other unobservables in the forms of fixed effects, besides productivity, affect the production function. Hence, I also consider the results using this approach. The limitation of this approach is to assume the linearity of the serial correlation in productivity.

Section 6 reports the main results using the proxy variable approach to estimate the production function. Appendices report results using OLS with fixed effects and the proxy variable approach for the production function estimation. Results show the productivity and economies of scale by the three approaches exhibit some differences in magnitudes but follow similar distribution shapes by vessel groups and years.

[16]Olley and Pakes (1996) estimate the value added production function. Levinsohn and Petrin (2003) estimate the gross output production function but they estimate the coefficient of labor in the first stage, which is shown to be biased in Ackerberg, Caves and Frazer (2015).

4.3.2 Identifying the Causal Impacts of Trading

The causal impacts of trading are identified upon the principle of difference-in-difference: assuming that the group of trade qualified vessels and the group of non-qualified vessels follow similar (parallel) trends in productivity and economies of scale, and that the impacts of trading do not spill over to non-qualified group, as noted in Section 4.1.

When combining with the estimation of a production function, separating the impact of trading on productivity from the impact on input factors right in the production function estimation step, before feeding into the difference-in-difference step, is important. Otherwise, the impact of the policy (trading) on productivity would be confounded with the impact of policy on input choices (and cost elasticity). To do such separation, including both *Trade Qualified*$_{i,t-1}$ and $\omega_{i,t-1}$ in the evolution of productivity is the key and solves two issues.[17] First, ignoring *Trade Qualified*$_{i,t-1}$ in the evolution process would let ξ_{it} absorb the trading impact. If trading also affects input choices ($k_{i,t-1}, l_{i,t-1}, d_{i,t-1}, m_{i,t-1}$), then ξ_{it} would correlate with the input choices. Second, more-productive firms tends to self select to the trading program. Including lagged productivity $\omega_{i,t-1}$ (together with *Trade Qualified*$_{i,t-1}$) helps control the potential self-selection of trading.

In summary, the estimated productivity in the estimation stage of a production function includes its own impact of trading without being confounded by the impacts of trading on production inputs. The estimated production coefficients β also keep their own impact of trading and, together with the variation in production inputs, contain the impact of trading on economies of scale (cost elasticity). In other words, whereas variation in each production input identifies the impact of trading on the input itself (if any), the total variation in *all* inputs weighted by the production coefficient (after being separated from the variation in productivity) identifies the impact of trading on economies of scale.

5 Data and Summary Statistics

The data are given by the Norwegian Directorate of Fisheries under a data confidentiality agreement. The data consist of four sets for the Norwegian cod fishery from January 2001 to December 2017. First, the vessel registry records the yearly registration status of a vessel and

[17] This productivity process has been noted by De Loecker (2013); Braguinsky et al. (2015) to study the effects of a firm's export status and plant acquisition on productivity, respectively. The original proxy variable approach by Olley and Pakes (1996); Levinsohn and Petrin (2003); Ackerberg, Caves and Frazer (2015) assumes the Markov productivity evolution without controls: $\omega_{it} = g_1(\omega_{i,t-1}) + \xi_{it}$. De Loecker (2013) also notes that such a productivity process as in (17) can directly estimate the impact of the event, trading in my context. A flexible specification of g can also offer the distribution of the heterogeneous impact of trading. However, I, similar to Braguinsky et al. (2015), take a two-stage approach where the second stage is the difference-in-difference regression. Although the two-stage approach delivers the average effect of trading instead of vessel-specific effect, this approach offers two advantages. First, I want to look at the changes from before to after the trading event not just for the trade-qualified vessels only, but also in comparison to a control group. Second, I want to use a consistent framework to investigate multiple outcomes.

its physical characteristics including length, engine power, tonnage, and built year. Second, the ownership registry describes the identity of a vessel's owner and their name, address, and organization type as of 31 December every year. Missing ownership is filled in using complement files of vessel events on changes in owner and vessel identification for continuous years. Third, the license registry records the license information and its valid duration a vessel holds. Fourth, the landing data record transactions of first hand sales of fish between a fisherman and a buyer (typically processing firms). The recorded transaction includes information on catch quantity, unit price, fishing vessel, landing date, the latest catch date, major catch location, fishing gear, crew size, and landing municipality. Unfortunately, trip duration is not available. When addressing fishing frequencies, I suppose each latest catch date represents for a trip in the corresponding week and month.

Table 2: Summary statistics

	count	mean	sd	min	max
Panel A: Sample of trip-level observations					
catch quantity (tonne)	1,158,557	1.61	4.04	0.00	224.62
revenue (thousand NOK)	1,158,557	24.24	65.24	0.00	4345.58
price (NOK/kg)	1,158,557	15.79	6.47	0.20	4955.00
crew (person)	1,158,557	2.11	1.49	1	99
distance (km)	1,158,557	133.78	261.43	0.05	2318.62
Panel B: Sample of yearly observations					
catch quantity (tonne)	30,776	60.83	94.98	0.00	1874.40
revenue (thousand NOK)	30,776	915.26	1424.22	0.01	28250.42
average value (NOK/kg)	30,776	16.10	5.66	4.79	128.83
length (m)	30,776	12.86	4.93	4.25	55
crew (person)	30,776	2.24	1.62	1	21.25
distance (km)	30,776	180.01	286.38	0.05	1642.74
# trips	30,776	37.68	23.80	1	213

I merge four datasets to compile two main samples for the analysis: trip-level sample and yearly sample. Table 2 shows the summary statistics of variables in the two samples.[18] Panel A summarizes the trip-level sample. Key variables include catch quantity, transacted unit price, revenue (as a product of quantity and unit price), crew size, and distance from the major catch location to the home municipality of the fishermen (also the vessel's register). These variables are recorded for every landing transaction. On average, a vessel catches 1.6 tonne (1.7 US ton), but there is a big gap across vessels. Some vessels catch only several kilograms of cod and some catch up to 224 tonnes (247 tons). The unit price has the mean value of 15.79 NOK/kg (82 cents/lb) and also varies a lot by transactions. Crew size on average is 2. There are 14 observations of 99 people on board, whereas the second highest value of the crew size is 61. Although 99 is definitely not the code of missing values, it may be a misreport by the

[18]One useful note is 10 nok $\approx$ 1 euro $\approx$ 1.15 usd. 1 kg $\approx$ 2.2 lbs. 1 tonne $\approx$ 1.1 US tons.

fishermen. Because there are only such 14 observations and it is important to keep the records of catch quantity, I decide to keep these observations in the sample for analysis.

Panel B summarizes the yearly observations. The yearly catch quantity and revenue are the sums of trip-level values for each vessel. The crew size and distance are the averages of trip-level values, weighted by the trip quantity.

For the main analysis, I use the yearly sample, except the investigation of transacted prices. Main fishing outcomes are catch quantity and revenue. Production inputs include vessel length, crew size, fishing distance (from fishermen's municipality to major catch location), and the number of trips in a year.

Figure 4 plots the average values of key variables, namely the number of vessels, catch quantity, revenue, vessel length, crew size, fishing distance, and the number of trips by licensed length group over years. Panel 4a shows the number of vessels over years. In all license groups, the numbers of vessels decrease over time. For vessels with licensed lengths from and above 15 meters, the number of vessels in these two groups has decreased since 2004, after the trading program started and allowed a vessel to buy out quota of another vessel that had to exit after selling the quota. Similarly, the number of vessel with licensed length of 11–14.9m has decreased after 2008. Between 2003 and 2008, a condemnation program in which the state bought back a number of vessels with licensed lengths below 15m played a key role to dramatically reduce the number of vessel in this group. The fall in the number of vessel in the license length 0–10.9m after 2010 is attributed to voluntary exits rather than the condemnation scheme or the trading policy. Note that although the number of vessel in the non-tradable group (0–10.9m) reduces significantly over years, the percent number in fact increases, see panel 4b Hence, the falls in the numbers of vessels in the tradable groups after the trading policy enacted are more substantial relative to the non-tradable group.

Panels 4c and 4d show vessels in a bigger licensed length group catch and earn more systematically. The tradable license groups catch significantly higher after the trading program applied.

Panels 4e–4h show the trend in yearly production factors: vessel actual length, crew size, distance from fishermen's municipality to major catch location, and the number of trips. Vessels in a bigger licensed length group systematically have certain longer vessel and bigger crew size. Since 2010, vessels have been allowed to be longer than 28 meters as long as their cargo sizes are below 500m^3. We see vessels in the license group 21–27.9m have taken this advantage and expanded their sizes beyond 28m. Given the big gap in actual vessel size among license groups, one may concern vessels in the big license group would crowd out small vessels in the fishing ground. Panel 4g alleviates this concern by showing that vessels in different license groups fish in areas distant from each other. Although the two license groups, 11–14.9m and 15–20.9m, fish in overlapped areas, the group 0–10.9m as a control group fishes within 150km from the vessel home municipality's coast, distinctly not overlapping with other groups. The separation in fishing ground between tradable group and control group is important, because it helps mitigate

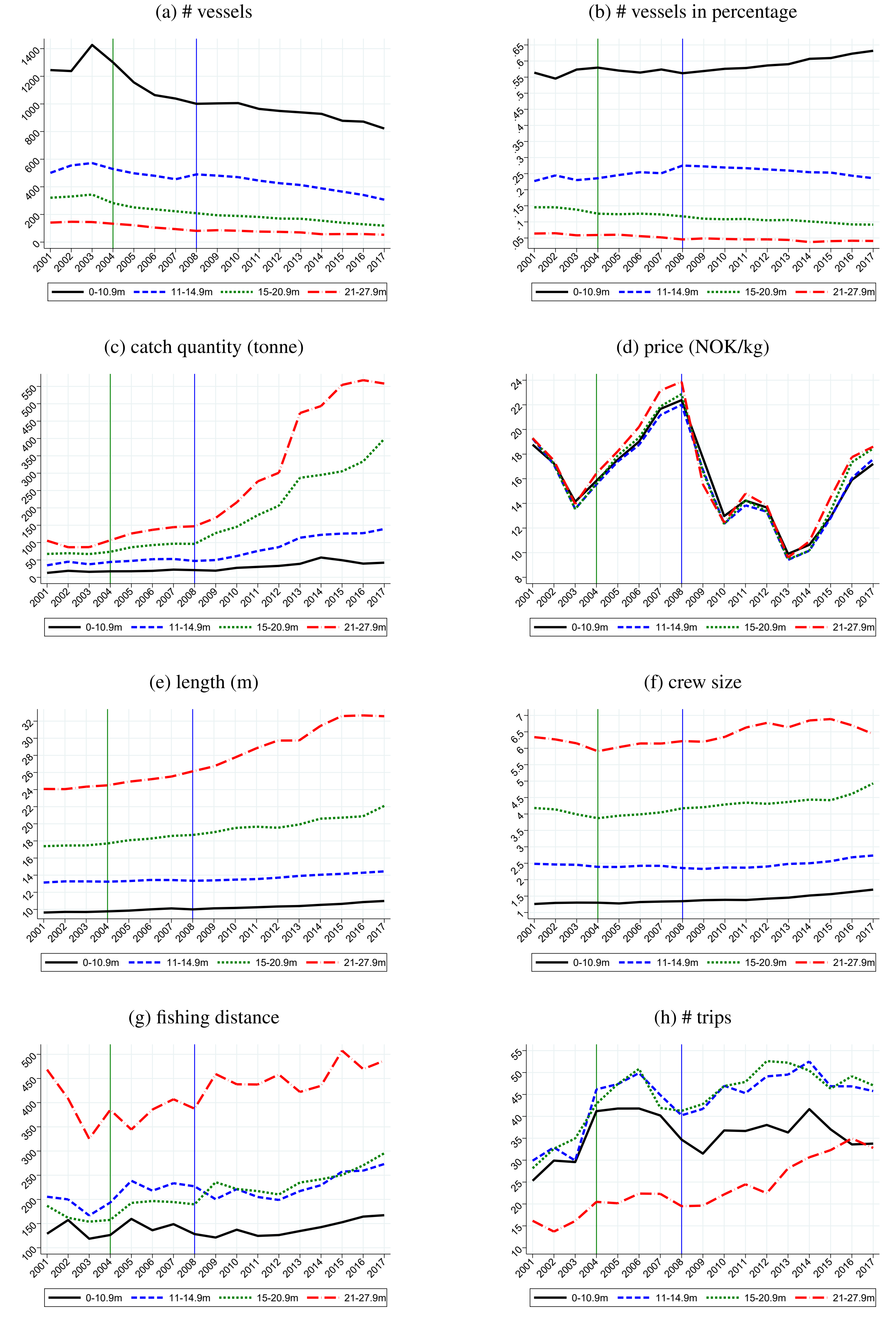


Figure 4: Description of key variables by licensed length group in a yearly-observation sample

the spillovers that threat the identification of the difference-in-difference estimation design in my context. A small threat of spillovers in harvest may still exist if the fishing activities of big vessels in faraway locations interfere the migration of the cod school. Figure A2 in Appendix A shows the distribution area and spawning area of cod in the north of the latitude $62°N$, the whole fishery that is studied in this paper. The spawning areas are near to fjords and the coastal areas, where small vessels fish. Big vessels fishing in the ocean may interfere with the migration, but there exists several management measures (closure areas and timing restrictions) during a year to limit fishing activities during the spawning period.

Figure 4h shows vessels in the big license groups go fishing less often than the ones in the smaller license groups. There may be two reasons. One is vessels in the big license groups travel on much bigger vessels. Another reason is the figure shows the number of trip in a year. Big vessels may prolong the trip thanks to higher safety and better equipped. Unfortunately, the trip duration is not reported. Hence, instead of investigating the trip production, I focus on the yearly production with four yearly inputs: length, crew size, fishing distance, and the number of trips. In fact, focusing on yearly production is also plausible for the reason that these four inputs are potentially all main factors a fisherman (or vessel's owner) would consider to design a plan to utilize the *yearly* quota.

6 Results

6.1 Consolidation of Catch Quantity

Table 3 report the effects of CAT program and quota acquisition on yearly catch quantity. All specifications include year fixed effects, vessel fixed effects, and owner fixed effects. Panel A shows the results for the difference-in-difference (DID) specification of equation (13) that estimates an intent to treat (ITT) of the trade program. Across the columns, the estimates suggest that the cap-and-trade program, compared to the previous nontradable cap, increases the harvest by 8.5%. Note that vessels are divided into four groups depending on licensed lengths: below 11m, 11–14.9m, 15–20.9m, and 21–27.9m. The latter two groups, 15–20.9m and 21–27.9m, are allowed to trade quotas from 2004. The licensed length group 11–14.9m may trade quotas from 2008. Columns (2)–(4) show the three biggest groups improve harvest by roughly 25%, 13%, and 4%, respectively.

Panel B estimates the average treatment on the treated (ATT) of quota acquisition. Results show vessels that acquire quotas catch at least 50% more than before. The ones in the biggest licensed length group, thanks to higher quota ceilings, even catch 85% more.

Figure 5 shows event-study graphs for the policy effects on catch quantity. Panel A show the relative changes between vessels in the trade-qualified groups and vessels in the never-treated group (licensed length below 11m) before and after 2004. Year 2004 is the first time when the trading program was introduced. We see that vessels in the trade-qualified groups improve their

Table 3: Effects of trading policy on catch quantity

coefficients	(1) staggered estimate	(2)	(3) estimate by group	(4)
		11-14.9m	15-20.9m	21-27.9m
Panel A: ITT				
Trade qualified	0.085***	0.044	0.127***	0.226***
	(0.027)	(0.035)	(0.033)	(0.047)
Panel B: ATT				
Quota acquisition	0.464***	0.408***	0.493***	0.623***
	(0.023)	(0.028)	(0.058)	(0.056)

Note: Panels represent specifications. Columns reports coefficients. All specifications use yearly observations and include year fixed effects, vessel fixed effects, and vessels' owner fixed effects. Panel A estimates ITT of the trading policy: $Y_{it} = \beta_{ITT} \textit{Trade Qualified}_{it} + \eta_i + \tau_t + \epsilon_{it}$. Panel B estimates ATT of quota acquisition: $Y_{it} = \beta_{ATT} \textit{Quota Acquisition}_{it} + \eta_i + \tau_t + \epsilon_{it}$. Standard errors in parentheses are clustered by vessel's municipality. Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

(a) logged quantity (ITT) (b) logged quantity (ATT)

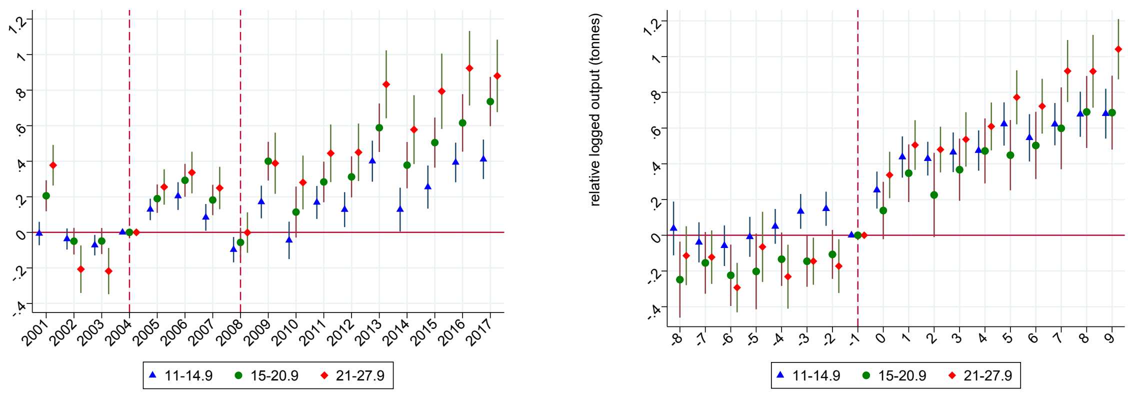


Figure 5: Event studies the impacts of trading on catch quantity growth

Note: Panel A plots the coefficients of the ITTs, except the indicator *Trading Qualified* is interacted with dummies for years before and after the program started. The base group is the licensed length below 11m and year 2004. Panel B plots the coefficients of the ATTs, except the indicator $\textit{Quota Acquisition}_{it}$ is interacted with dummies for years before and after the acquisition. The year prior to the quota acquisition year is normalized.

harvest modestly between 2004 and 2008. Since 2008, the trade-qualified groups dramatically increase their harvest. Since 2013, some vessels in the licensed length 21–27.9m even double their harvest.

Panel B illustrates the event studies for the effect of quota acquisition. They plot the coefficients of the regression (14), except the indicator $\textit{Quota Acquisition}_{it}$ is interacted with dummies for years before and after the acquisition. The one year prior to the quota-acquisition year is normalized. Panel B shows coefficients of interactions for years before acquisition are very close to zero, implying parallel trend assumptions satisfied. This suggests the trends in catch quantity of vessels that acquired quotas are very similar to the trends of vessels that did not. For years after quota acquisition time, we see an immediate effect of quota acquisition. Vessels acquiring quota immediately catch 20% higher than the preceding year.

These increases in catch quantity are not surprising, because one will expect vessels in the trade qualified group purchase fishing quotas from others to boost the harvest.

6.2 The Impacts of Trading on Productivity

Before examining the causal impact of trading on productivity, I explore the overall distribution of productivity by licensed length group and year. Figure 6 plots the productivity index estimated using the proxy variable approach.[19] Over years, productivity in all of the four groups has shifted to the right side. The shifts to the right are firstly attributed to the common improvement in productivity over time. One instant reason is technology in harvest may advance over time. Another reason is fisherman may adapt better to weather conditions and obtain productivity gains over time. Furthermore, although only the three upper license group has the trading program, spillovers of the trading impact on the whole fishery biomass may increase common fish stocks, allowing the regulator to increase total allowable catch in all four groups.

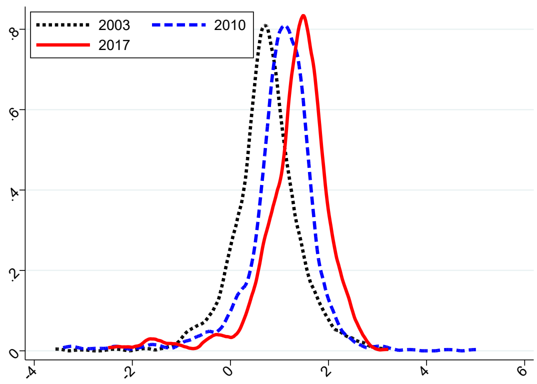
Despite the same direction in the shifts of productivity distribution in all four groups, the evolution of the distribution shows remarkably different patterns between the non-tradable group and the tradable groups. The distribution of productivity in the non-tradable group (0–10.9m) just simply shifts itself over years. Its peak level, variance, and tail length nearly remain the same as before. This implies the change in productivity of vessels in this group is purely induced by the systematic shocks over time. On the other hand, the distributions of productivity in the other tradable groups shift and change their shapes. The variance has decreased and more importantly, the lower tail has considerably shortened over time. In 2010, the distributions of productivity in the two middle groups (11–14.9m and 15–20.9m) had a long left tail, suggesting these two groups had a few vessels with extremely low productivity relative to the other vessels in the group. In 2017, these left tails were cut, and the new distributions even did not appear skewed. The shortening in the tails implies that the lowest productive vessels has exited from the fishery. For vessels in the license group 21–27.9m, exits of low productive vessels has happened sooner than the two middle groups.

Finally, I run difference-in-difference regressions on vessel productivity to examine the effects of trading policy on within-vessel productivity. Table 4 in reports the ITT and ATT estimates of the policy impacts on productivity.[20] Figure 7 plots the event-study coefficients. In general, we do not see the program has a significant effect on all individuals in the trading group. Instead, the trading program has substantial impact only on vessels that do acquire quotas. Acquiring additional quotas help vessel increase its productivity by 25–46% on average.

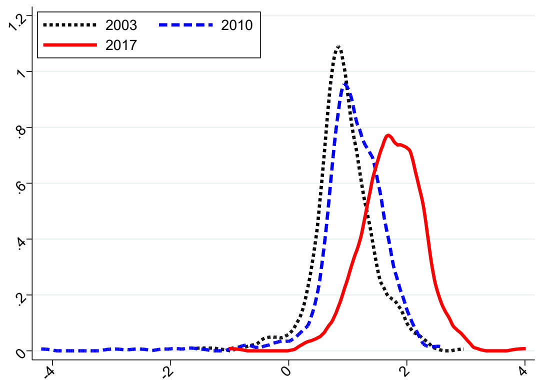
[19] Appendix D shows the distribution of productivity index estimated by the OLS with fixed effects and the dynamic panel approaches.

[20] Appendix E reports all estimates, including LATE of the policy impacts on productivity and where productivity is estimated using OLS with fixed effects and the dynamic panel approach.

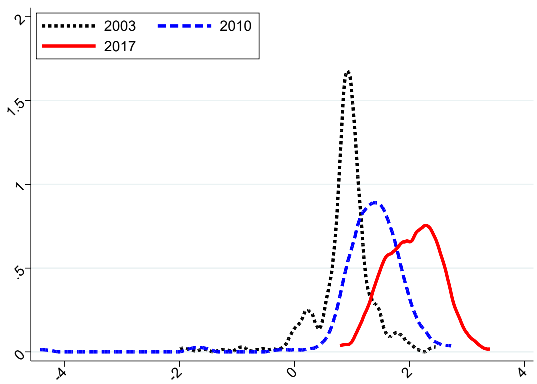
(a) productivity in licensed length 0–10.9m



(b) productivity in licensed length 11–14.9m



(c) productivity in licensed length 15–20.9m



(d) productivity in licensed length 21–27.9m

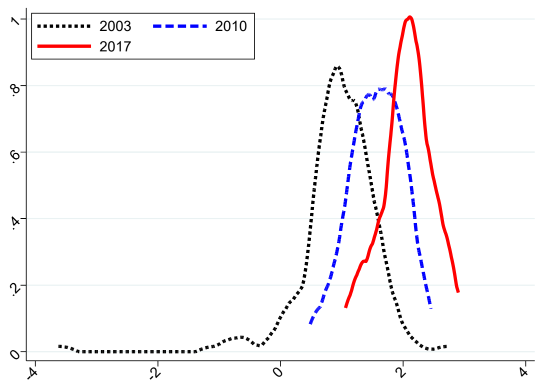
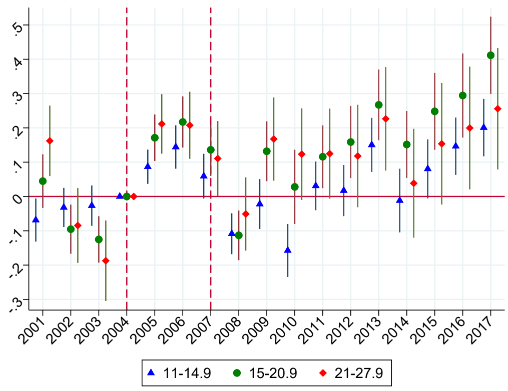


Figure 6: Distribution of productivity (proxy variable approach)

(a) ITT of trading policy



(b) ATT of quota acquisition

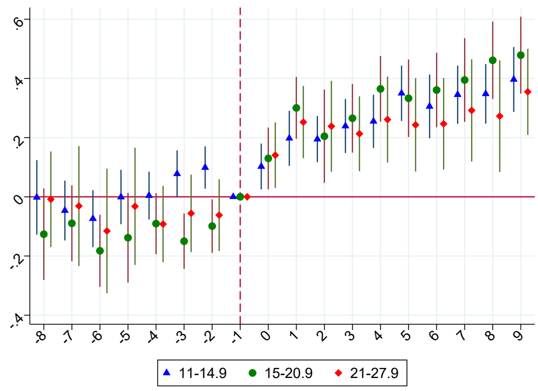


Figure 7: Impacts of the trading scheme and quota acquisition on productivity

Note: Productivity is estimated using proxy variable approach. Panel A plots the event study coefficients of ITT of the trading policy. Year 2004 and the non-tradable group (0–10.9m) are normalized. From 2004, the groups 15–20.9m and 21–27.9m may trade. From 2007, the group 11–14.9m may trade. Panel B plots the event study coefficients of ATT of quota acquisition, for years before and after the acquisition; vessels in the non-tradable group and that are in the tradable group and do not trade are the base group.

Table 4: Effects of trading policy on logged productivity

coefficients	(1) staggered estimate	(2) estimate by group	(3)	(4)
		11-14.9m	15-20.9m	21-27.9m
Panel A: ITT				
Trade qualified	0.025	-0.039	0.145***	0.129***
	(0.020)	(0.024)	(0.033)	(0.041)
Panel B: ATT				
Quota acquisition	0.276***	0.225***	0.381***	0.292***
	(0.022)	(0.028)	(0.039)	(0.048)

Note: Panels represent specifications. Columns reports coefficients. All specifications use yearly observations and include year fixed effects, vessel fixed effects, and vessels' owner fixed effects. Panel A estimates ITT of the trading policy: $Y_{it} = \beta_{ITT} \textit{Trade Qualified}_{it} + \eta_i + \tau_t + \epsilon_{it}$. Panel B estimates ATT of quota acquisition: $Y_{it} = \beta_{ATT} \textit{Quota Acquisition}_{it} + \eta_i + \tau_t + \epsilon_{it}$. Standard errors in parentheses are clustered by vessel's municipality. Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

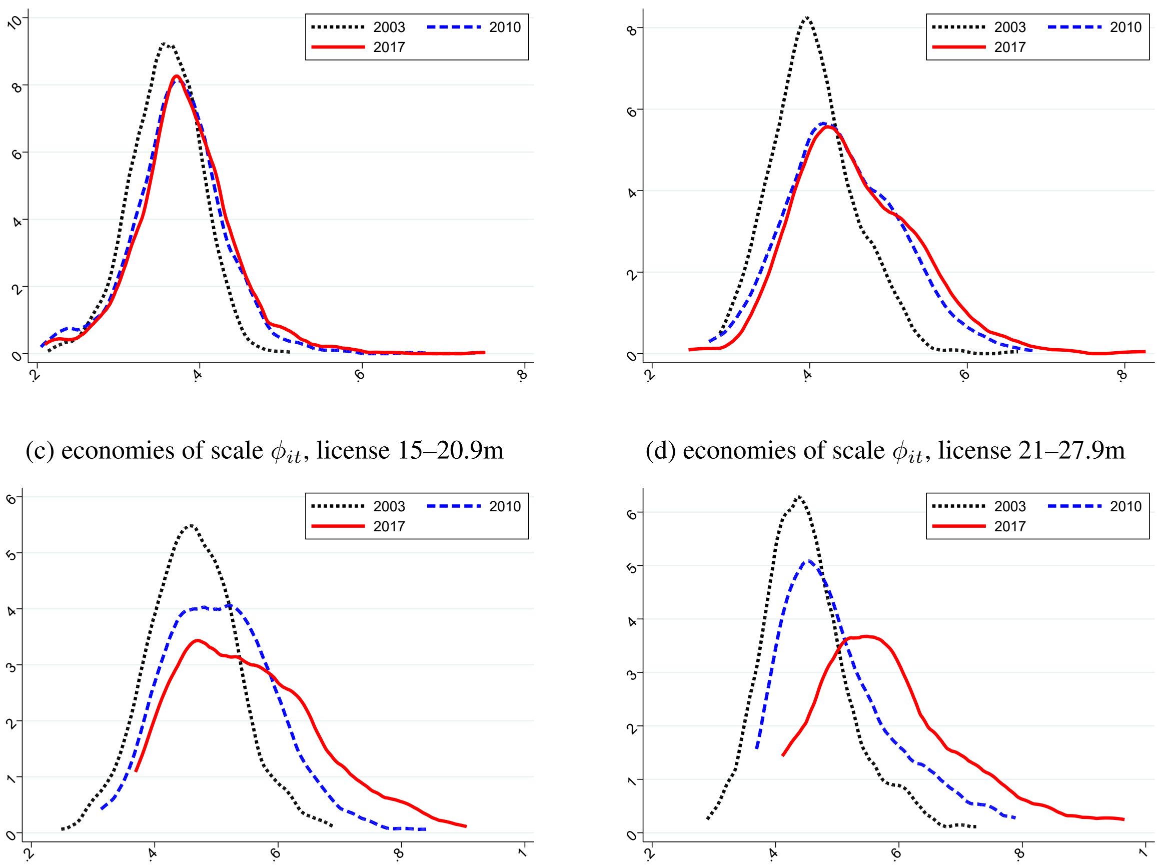
6.3 The Impacts of Trading on Economies of Scale

Table 5: Summary statistics of economies-of-scale index by licensed length group

	Pre-trade-program					Post-trade-program				
	count	mean	sd	min	max	count	mean	sd	min	max
0–10.9m	8,470	0.372	0.049	0.203	0.602	8,362	0.377	0.060	0.203	0.849
11–14.9m	3,590	0.433	0.061	0.244	0.677	3,637	0.449	0.074	0.220	0.827
15–20.9m	995	0.458	0.066	0.249	0.703	2,367	0.514	0.089	0.278	0.918
21–27.9m	433	0.456	0.068	0.289	0.738	1,016	0.524	0.097	0.313	1.014

Note: The index is the output elasticity of total cost ϕ_{it}, calculated using estimates of the production function coefficients using the proxy variable approach. Pre-trade-program period and post-trade period for licensed groups 15–20.9m and 21–27.9m are 2001–2003 and 2005–2017. For licensed length group 11–14.9m, they are 2001–2007 and 2009–2017. Licensed length group 0–10.9m is not allowed to trade during 2001 and 2017, but we compare period 2001–2007 to period 2009–2017.

Table 5 contrasts the means and ranges of economies of scale between before and after the trading policy. The licensed length group 0–10.9m is not allowed to trade, but we contrasts its range of the index in the 2001–2007 period to the level in the 2009–2017 period. The table reports the scale index using the proxy variable estimator for the production function. Appendix D reports results using other approaches for the production function estimation. The first important point of the results is that the index in the pre-trade program are less than 1. This suggests vessels were having economies of scale before the trading program (marginal cost below average cost). The second point to notice is within each license group the mean and the max levels become higher in the post-trade period. These increases reveal that vessels are moving along the average cost curves starting on the left side of the curves toward where marginal cost reaches average cost or the minimum levels of average cost.



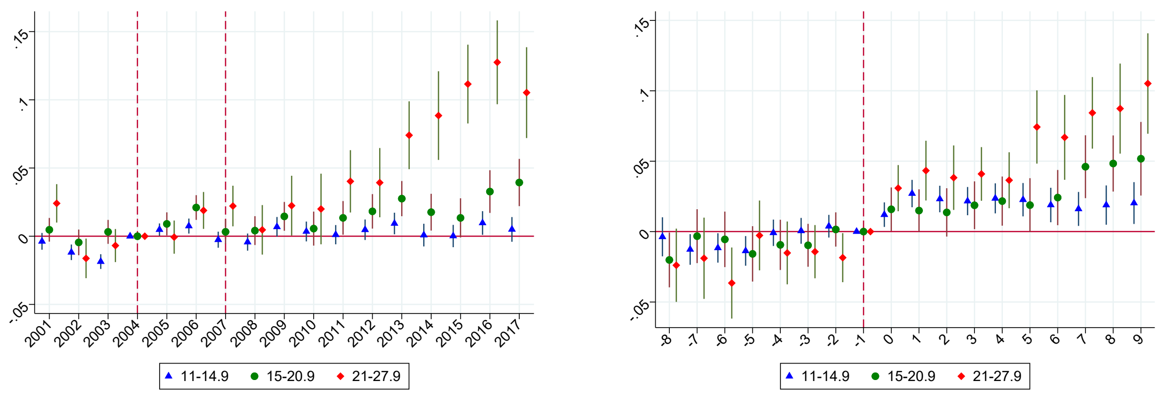
(a) economies of scale ϕ_{it}, license 0–10.9m

(b) economies of scale ϕ_{it}, license 11–14.9m

(c) economies of scale ϕ_{it}, license 15–20.9m

(d) economies of scale ϕ_{it}, license 21–27.9m

Figure 8: Distribution of economies of scale ϕ_{it} (proxy variable approach)



(a) ITT of the trading policy

(b) ATT of quota acquisition

Figure 9: Impacts of the trading policy and quota acquisition on economies of scale

Note: Economies of scale ϕ is recovered from the estimates of the production function coefficients that are estimated using a proxy variable approach. Panel A plots the event study coefficients of ITT of the trading policy. Year 2004 and the non-tradable group (0–10.9m) are normalized. From 2004, the groups 15–20.9m and 21–27.9m may trade. From 2007, the group 11–14.9m may trade. Panel B plots the event study coefficients of ATT of quota acquisition, for years before and after the acquisition. Vessels in the non-tradable group and that are in the tradable group and do not trade are the base group.

Figure 8 examines the change in economies of scale (or cost elasticity) over time within a license group by exploring the distribution of the cost elasticity over years by licensed length group.[21] Two features are noted. First, the distribution of cost elasticity in the lowest group (0–10.9m) almost remains unchanged from 2003 to 2017, suggesting the production costs of vessels in this group nearly do not change. On the other hand, the distributions of cost elasticities in the other three upper groups significantly become fattened and more positively skewed. So, vessels in these groups on average have substantially higher cost elasticities over time. There is also sizable variance in cost elasticity among vessels within each upper group. The second noticeable feature is that it is the upper tail that drives the increase in cost elasticity of vessels in the three upper group. For the two highest upper groups, we also see the left tails have been cut, implying a number of high-average-cost vessels (of which cost elasticities are far much smaller than 1) exits the fishery.

Table 6: Effects of trading policy on economies of scale (output elasticity of cost)

coefficients	(1) staggered estimate	(2) estimate by group	(3)	(4)
		11-14.9m	15-20.9m	21-27.9m
Panel A: ITT				
Trade qualified	0.004**	0.003	0.004	0.012*
	(0.002)	(0.002)	(0.004)	(0.006)
Panel B: ATT				
Quota acquisition	0.033***	0.025***	0.034***	0.062***
	(0.003)	(0.003)	(0.006)	(0.007)

Note: Panels represent specifications. Columns reports coefficients. All specifications use yearly observations and include year fixed effects, vessel fixed effects, and vessels' owner fixed effects. Panel A estimates ITT of the trading policy: $Y_{it} = \beta_{ITT} \textit{Trade Qualified}_{it} + \eta_i + \tau_t + \epsilon_{it}$. Panel B estimates ATT of quota acquisition: $Y_{it} = \beta_{ATT} \textit{Quota Acquisition}_{it} + \eta_i + \tau_t + \epsilon_{it}$. Standard errors in parentheses are clustered by vessel's municipality. Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To formally test whether the change in economies of scale is caused by the cap-and-trade program, I examine the change in cost elasticity within a vessel over time using the DID approach. Table 6 and Figure 9 report the estimates and the event study. I find that the trading program plays an important role in raising the output elasticity of cost, thereby pushing the fishing operation toward the minimum-average-cost operation. The push happened strongest in the highest group (21–27.9m), because this group offers the largest room to acquire additional quota amount, offering the vessels possible biggest move in harvest expansion. Vessels that acquire additional quotas in this group incurred additional 6–8 percentage point in the output elasticity of cost. A back-of-the-envelope calculation uses these estimates finds that a vessel in the licensed group of 11–14.9m, 15–20.9m, and 21–27.9m saves 0.42%, 4.07%, and 15.60%

[21]These estimated economies of scale are calculated from the proxy variable approach of estimating a production function. Appendix D shows the distribution that uses the OLS with fixed effects and dynamic panel approaches of the production function.

of average cost, respectively.

7 Discussion

7.1 Decomposing the Value of Trading

I perform the decomposition method in Proposition (2) to entire vessels in the fishery to investigate the dynamic of output change ($\ln Q_{it}$) as well as the value of trading of a vessel by licensed length group. Figure 10 plots the decomposition of the change by year (relative to 2001). The year index is the vessel-change index weighted by a vessel's share of catch in the licensed length group. Output generally increases over years for all licensed length groups but is underlied by different patterns of the decomposing effects. Vessels in the non-tradable group (0–10.9m) have little economies-of-scale effect. The cost shifting plays the main role in explaining the change in output in this group. For vessels in the tradable group, both the two effects contribute to the change in output, but the relative contribution of the two decomposing effects differs in different periods. Between 2001 and 2003, the cost-shifting effect contributes more to the change in output. From 2004 through 2007, almost only the economies-of-scale effect contributes to the output growth; the cost shifting has no effect at all. Since 2007, the cost-shifting effect dramatically increases and outweighs the economies-of-scale effect, except the vessels in the group 21–27.9m. The output growth in vessels in the biggest licensed length group (21–27.9m) is generally attributed to the econonomies of scale effect. In recent years from 2014, the two effects and the output decrease, except for vessels in the licensed length 15–20.9m.

Figure 11 plots the decomposition of output change relative to the first trading time for vessels that ever acquire tradable quotas. Note that this is the change in output within a vessel relative to its own first trading time and does not take the change in output of other vessels into account. Not surprisingly, vessels increase their catches dramatically after acquiring quotas. However, whereas the economies-of-scale effect does not contribute to the output growth of vessels in the group 11–14.9m, it explains the output growth for vessels in the two upper groups (15–20.9m and 21–27.9m) in the first two years after acquiring quotas. Three years after acquiring quotas, cost shifting plays the main role to increase output of vessels in the group 15–20.9m, but it shares quite an equal contribution to the output growth of vessels in the group 21–27.9m, compared to the role of economies of scale.

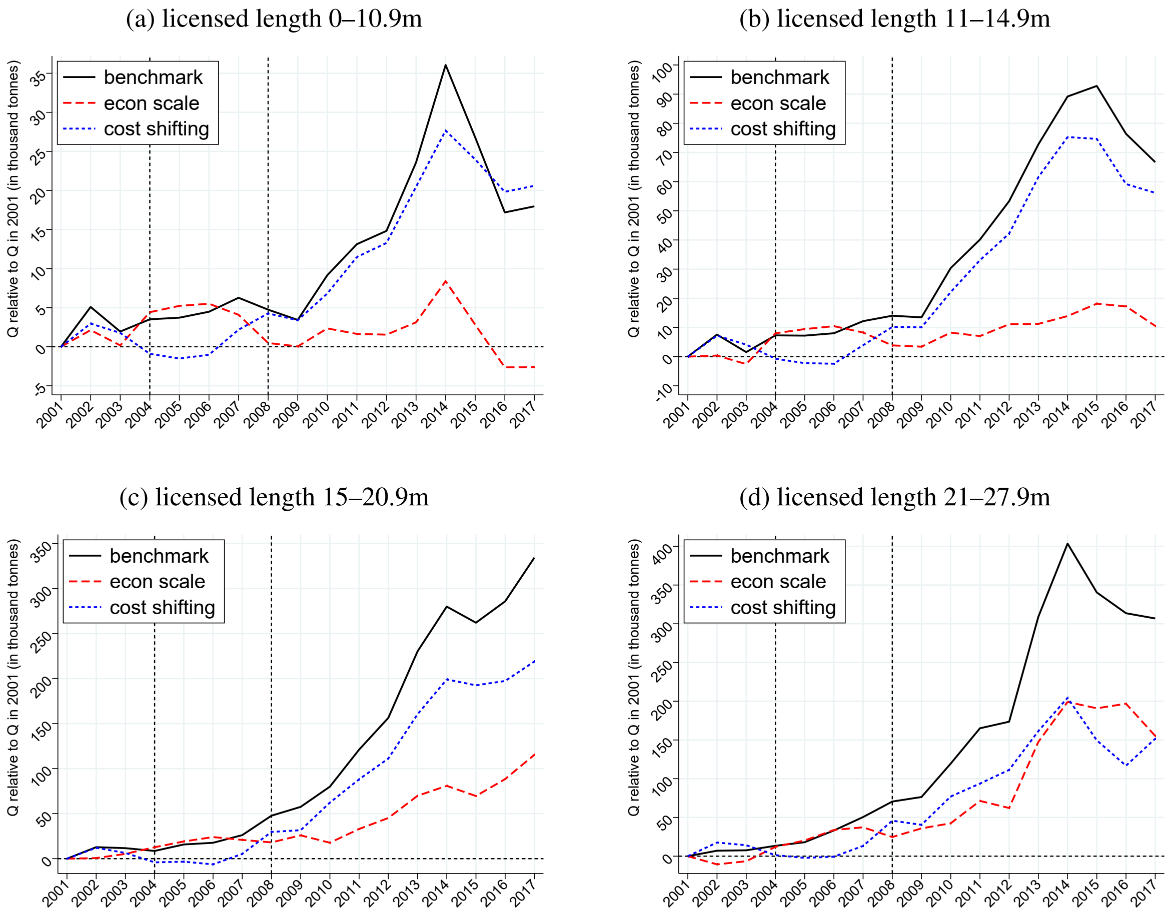


Figure 10: Decomposing the change in Q_{it} (thousand tonnes) by year

Note: The figure plots the change in output within a vessel from 2001 to each year. The average change number for each year is weighted by the vessel share of group catch.

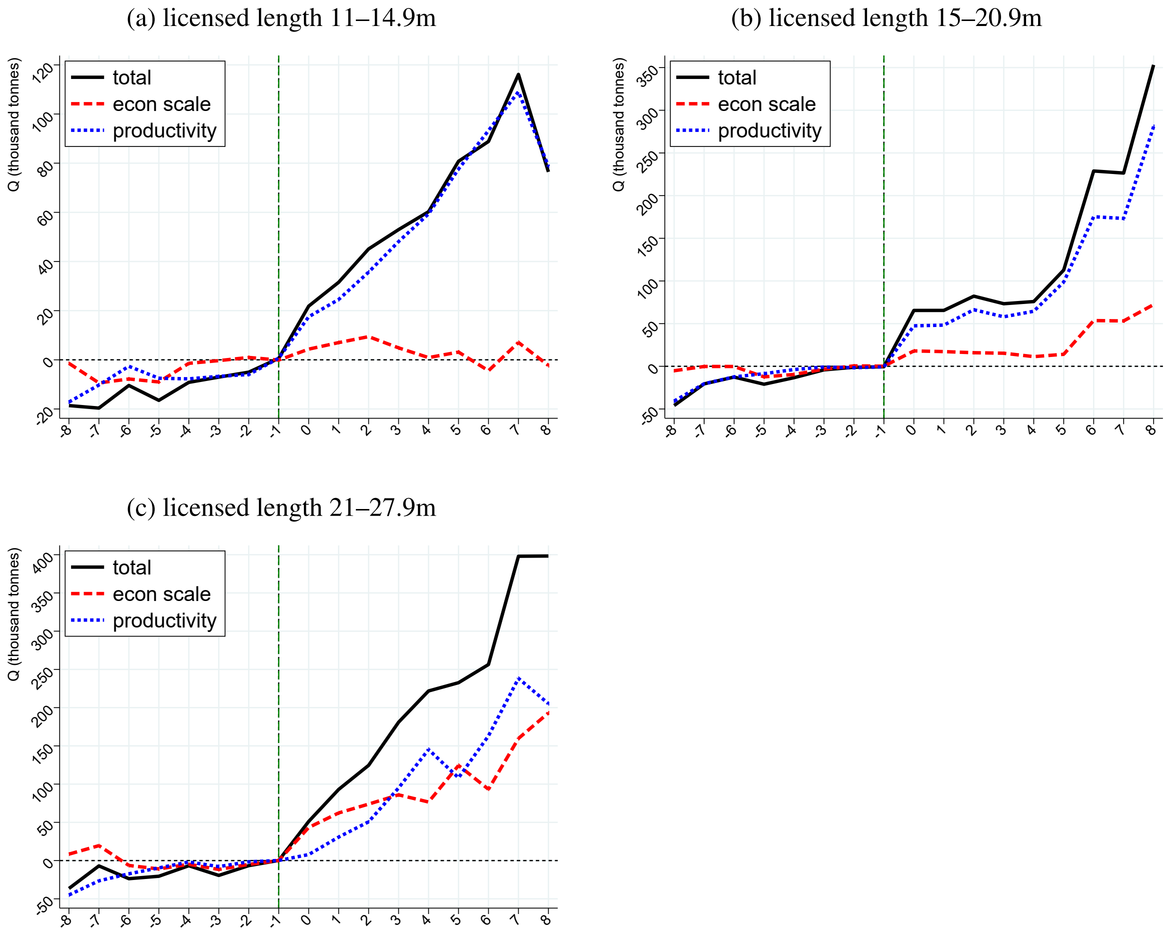


Figure 11: Decomposing the value of trading (in thousand tonnes of catch) by years from the first time a vessel acquires traded quotas

Note: The figure plots the change in output within a vessel over years. All changes are relative to the year when a vessel acquires traded quotas in its first time.

7.2 Where Do Economies of Scale Come From?

Because economies of scale is the reciprocal sum of all output elasticity of inputs, it can be interpreted as the a sufficient statistics of all changes in inputs. To understand how fishermen exploit economies of scale, I investigate the impacts of trading on input factors—boat size, crew size, fishing distance (distance from the fishers' municipality to major catch location), and the number of trips in a year. Figure 12 plots the event studies of the impacts. Table E3 in Appendix E reports the ITT, ATT, and LATE estimates of the effects.

Among the four production factors, we see the clear and significant increases in vessel length and the number of fishing trips. The trading policy seems to reverse the decreasing trend in crew size, and may imply a slight increase in labor usage. However, compared to the patterns in vessel length and the number of trip, I conclude the policy did not change labor size. Neither does it change fishing distance. Hence, fishermen have exploited economies of scale to significantly reduce fishing costs by investing in bigger boats and going fishing more often.

Because different licensed length groups are subject to different quota ceilings, boat size investment and fishing trip adjustment follow heterogeneous patterns for different license groups. Whereas the biggest licensed length group (21–27.9m) invested in 9% longer vessels and went fishing more frequently (by 24%–46%), the other two tradable licensed length group (11–14.9m and 15–20.9m) did not expand their vessels and instead, only went fishing more frequently to utilize larger fishing quotas.

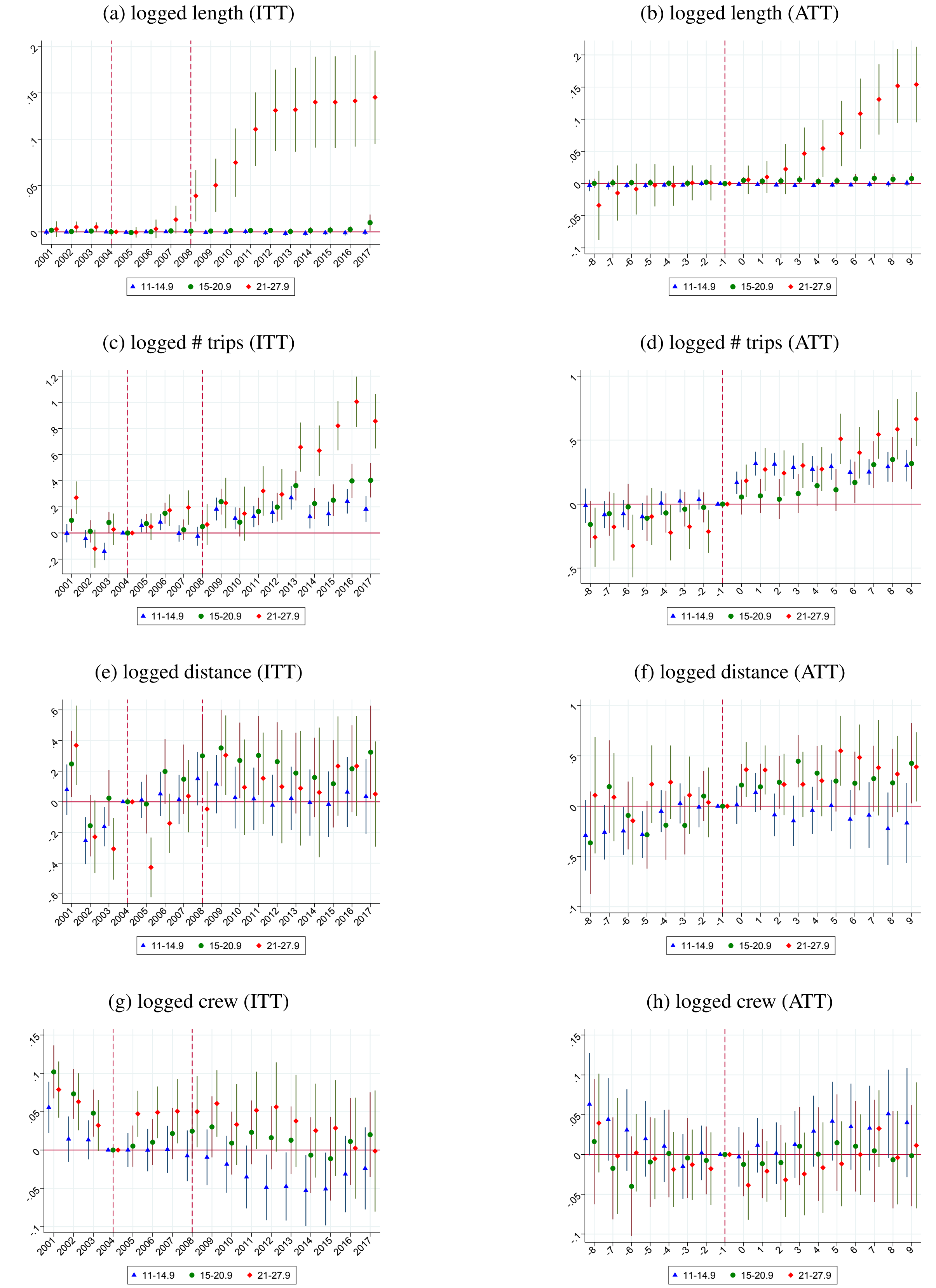


Figure 12: Changes in input factors that contribute to economies of scale due to quota trading

Note: The figure plots the event studies of the impacts of trading on fishing input factors: vessel length (meter), number of fishing trips in a year, fishing distance from the major catch location to the vessel's home municipality (mile), and crew size (person). Results imply vessels' owners exploit economies of scale by using bigger boats and going fishing more often.

7.3 Consolidation, Economies of Scale, and Market Power

Although CAT is popular in emission regulation, tradable caps have remained controversial in fishery management due to consolidation concerns (Adelaja, Menzo and McCay, 1998; Grainger and Parker, 2013; Byrne et al., 2020). Under a tradable quota scheme, large fishing firms can buy out quotas from small fishermen. If the ownership of quotas and catch rights become concentrated into a few players, certain oligopolists may exert market power and keep high prices of fish sales or high markups. On the other hand, those who support CAT in fishery of course rely on the cost effectiveness of trading in which the distribution of quotas is achieved at the least cost. My study contributes to the debate on costs and benefits of CAT and the studies of consolidation in three ways.

First, I provide another benefit of CAT for a policy maker to weigh the costs and benefits. I show CAT offer cost reduction owing to productivity improvement and economies of scale and trading is the causal factor that induces these benefits.

Second, the cost reduction due to economies of scale goes beyond an additional benefit of trading: it reverses the consolidation concern. That is, consolidation may be a benefit rather than a cost. The reason is economies of scale are exploited by only expanding operation, which is strongly and directly associated with consolidation. This argument establishes the tradeoff between market power (or anticompetition) and efficiency when it comes to consolidation concern, connecting to the long-standing arguments in merger and acquisition cases in IO and law studies.

For a long time, efficiency has been used to justify the permission of horizontal mergers (Farrell and Shapiro, 2001; Gugler and Siebert, 2007; Kaplow, 2021). However, "efficiencies in merger analysis are an enigma" and "the economic analysis of merger efficiencies lags far behind that of anticompetitive effects" (Kaplow, 2021). My analysis on economies of scale provides a framework to solve this puzzle: I provide a tool to examine and explain the presence of economies of scale as a source of efficiency. Moreover, clearing mergers on an efficiency defense requires efficiency be merger-specific and outweigh the market power effects, as similarly as should a policy-maker approve a cap-and-trade program. Verifying such efficiency is an empirical question, which can be answered in a similar way to this paper. I have shown three specific points in the analysis of trading in cap and trade: (i) vessels had economies of scale before trading was allowed and they were far to reach their minimum average cost levels, (ii) vessels does expand their catch toward reaching their minimum average cost levels after trading, and (iii) trading is the causal factor of the movement (or the increase in economies of scale index) and the causal effect is significant. Hence, this paper provides an empirical framework to investigate economies of scale as a source of efficiency for consolidation and merger related issues in future.

Third, the method to measure and estimate economies of scale in this paper can shed some light on the implied change in market power. If one observes the share of expenditure on input in total sales or the revenue cost ratio, one can estimate markup and directly investigate the effect of trading on markup. Otherwise, observing the product prices offer two benefits. We can directly test whether trading leads to higher product prices. We can also combine the results of price change and economies-of-scale change to discuss the markup change, if markup is a concern rather than price.

For example, in the cap and trade in Norwegian fishery, I observe the fish sales price (ex-vessel price). Hence, I use DID framework to explore the impacts of trading on price. Table 7 show either insignificant or very weak effects of consolidation on prices. ITTs are positive for the two biggest licensed length groups whereas ATTs are around zero. Figure 13 that plots event studies for the DID analysis reveals that trip-level transacted fish prices follow stable trends over years both before and after the quota acquisition month. Overall, the DID strategy shows no evidence for a change in prices due to the trading program and the quota acquisition.

Table 7: Effects of trading policy on logged price (trip-level price NOK/kg)

coefficients	(1) staggered estimate	(2) estimate by group	(3)	(4)
(logged NOK/kg)		11-14.9m	15-20.9m	21-27.9m
Panel A: ITT				
Trade qualified	0.005	-0.003	0.026**	0.019**
	(0.004)	(0.004)	(0.010)	(0.008)
Panel B: ATT				
Quota acquisition	-0.006	-0.011*	0.006	-0.009
	(0.005)	(0.006)	(0.009)	(0.010)

Note: Panels represent specifications. Columns reports coefficients. All specifications use yearly observations and include year fixed effects, vessel fixed effects, and vessels' owner fixed effects. Panel A estimates ITT of the trading policy: $Y_{it} = \beta_{ITT}\textit{Trade Qualified}_{it} + \eta_i + \tau_t + \epsilon_{it}$. Panel B estimates ATT of quota acquisition: $Y_{it} = \beta_{ATT}\textit{Quota Acquisition}_{it} + \eta_i + \tau_t + \epsilon_{it}$. Standard errors in parentheses are clustered by vessel's municipality. Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Note that the difference-in-difference strategy may underestimate the effects of consolidation on prices because of the spillover threats. The trading program forced the acquired vessels to exit the fishery, reducing the number of vessels in the whole fishery. At the same time, vessels less than 15 meters are subject to the decommissioning policy during 2004–2008, which reduces the number of vessels that are less than 15 meters. Hence, the non-tradable group that has licensed length 0–10.9m and mostly has actual length less than 11 meters might experience a unit price gain from 2004, causing the DID estimator to underestimate the effect of consolidation on prices.

However, the fact that price does not change may be attributed to the substantial competition in this industry. Figure 14 indicates the Herfindahl-Hirschman Index, despite the increasing consolidation over years in the two upper licensed length groups, is still very low (less than 0.03). The fishing market remains very competitive. Furthermore, cod is a valuable specie that

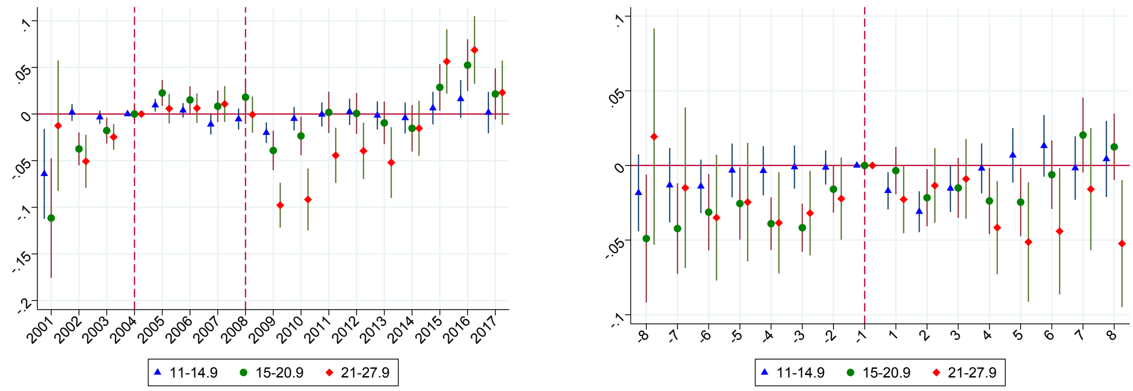


Figure 13: Event studies of ITT of trading policies and ATT of quota acquisition on price
Note: Panel A plots the coefficients of the ITTs, except the indicator *Trading Qualified* is interacted with dummies for years before and after the program started. The base group is the licensed length below 11m and year 2004. Panel B plots the coefficients of the ATTs, except the indicator *Quota Acquisition*$_{it}$ is interacted with dummies for years before and after the acquisition. The year prior to the quota acquisition year is normalized.

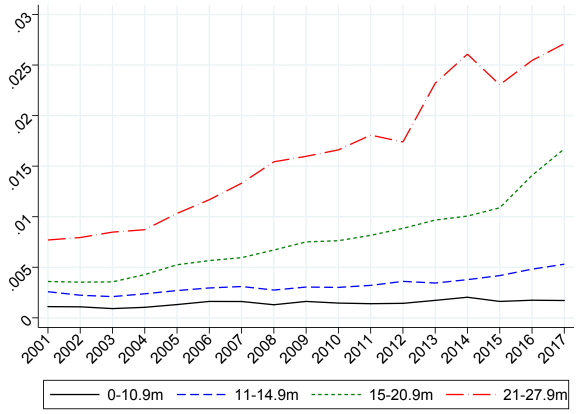


Figure 14: Change in Herfindahl-Hirschman Index over years by licensed length group

Norway exports to the global market, in which Norwegian vessels may be price takers. Overall, I conclude that the quota trading program and quota acquisition did not affect fish sales prices at all.

The fact that price has remained stable after trading and fishermen have significant cost savings owing productivity improvement and economies of scale raises a question on the implicit increase in markups. Fishermen could have lowered prices but they did not. However, such assertion should be discussed with caution. The reason is the cost savings owing to economies of scale refers to the decrease in average cost rather than marginal cost. It is possible that marginal cost has increased while average cost has decreased, especially when they are about to reach at the minimum average cost level. Note that claiming marginal cost has increased to show fishermen had no benefits is not a credible statement either. The reason is the observed fish sales price is likely average price rather than marginal price. Consequently, examining the change in markup should be clear about whether it refers to average price, marginal price, average cost, or marginal cost. The analysis of economies of scale in this paper can discuss the change in average cost with the least data. Of course, the welfare evaluator may simply care about the change in price rather than in markup when the matter regards market power.

8 Conclusion

This paper makes two primary contributions. First, I provide a new value of CAT: cost efficiency within firms owing to productivity improvement and economies of scale. I exploit the unique setting of the CAT in Norwegian cod fishery to show trading is the causal factor for such value. The new additional value of CAT not only requires policy makers re-evaluate the costs and benefits of CAT, but it also casts caveat on the consolidation concerns of trading. Consolidation may offer benefits because it facilitates the realization of economies of scale. Hence, the cost efficiency owing to economies of scale should be weighted against the market power concern when consolidation matters.

Second, I offer a method to empirically estimate the economies of scale using production data. The method is applicable beyond fishery context and offers a tool for future empirical applications beyond fishery CAT. For example, examining the cost efficiency value in CAT in air pollution is interesting and important. Although my study shows the significant efficiency value in the fishery CAT, whether CAT or an environmental regulation, in general, offers cost reduction owing to productivity improvement and economies of scale depends on the context and is an empirical question. Calel (2020) and Calel and Dechezleprêtre (2016) recently show that European Emission Trading (ETS) has increased low-carbon innovation among regulated firms. The finding brings up a question on how much of such innovation contributes to cost efficiency owing to productivity and economies of scale.

References

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer. 2015. "Identification properties of recent production function estimators." *Econometrica*, 83(6): 2411–2451.

Adelaja, Adesoji, Julia Menzo, and Bonnie McCay. 1998. "Market power, industrial organization and tradeable quotas." *Review of Industrial Organization*, 13(5): 589–601.

Ambec, Stefan, Mark A Cohen, Stewart Elgie, and Paul Lanoie. 2013. "The Porter hypothesis at 20: can environmental regulation enhance innovation and competitiveness?" *Review of environmental economics and policy*.

Arellano, Manuel, and Olympia Bover. 1995. "Another look at the instrumental variable estimation of error-components models." *Journal of Econometrics*, 68(1): 29–51.

Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies*, 58(2): 277–297.

Armstrong, Claire W, and Derek J Clark. 1997. "Just fishing? Equity and efficiency in fisheries management regimes." *Marine Resource Economics*, 12(3): 203–220.

Armstrong, Claire W, Arne Eide, Ola Flaaten, Knut Heen, and Inga Wigdahl Kaspersen. 2014. "Rebuilding the Northeast Arctic cod fisheries–economic and social issues." *Arctic Review*, 5(1).

Asche, Frank, Håkan Eggert, Eyjolfur Gudmundsson, Ayoe Hoff, and Sean Pascoe. 2008. "Fisher's behaviour with individual vessel quotas—Over-capacity and potential rent: Five case studies." *Marine Policy*, 32(6): 920–927.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen. 2020. "The fall of the labor share and the rise of superstar firms." *The Quarterly Journal of Economics*, 135(2): 645–709.

Becker, Randy, and Vernon Henderson. 2000. "Effects of air quality regulations on polluting industries." *Journal of Political Economy*, 108(2): 379–421.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile prices in market equilibrium." *Econometrica: Journal of the Econometric Society*, 841–890.

Berry, Steven T. 1994. "Estimating discrete-choice models of product differentiation." *The RAND Journal of Economics*, 242–262.

Blundell, Richard, and Stephen Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics*, 87(1): 115–143.

Blundell, Richard, and Stephen Bond. 2000. "GMM estimation with persistent panel data: an application to production functions." *Econometric Reviews*, 19(3): 321–340.

Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch. 2021. "Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data." *Journal of Monetary Economics*, 121: 1–14.

Braguinsky, Serguey, Atsushi Ohyama, Tetsuji Okazaki, and Chad Syverson. 2015. "Acquisitions, productivity, and profitability: Evidence from the Japanese cotton spinning industry." *American Economic Review*, 105(7): 2086–2119.

Bresnahan, Timothy F. 1989. "Empirical studies of industries with market power." *Handbook of industrial organization*, 2: 1011–1057.

Byrne, Conor, Sveinn Agnarsson, Brynhildur Davidsdottir, and Maartje Oostdijk. 2020. "Species-level quota concentration in the Icelandic harvesting sector." *Marine Policy*, 121: 104108.

Calel, Raphael. 2020. "Adopt or innovate: Understanding technological responses to cap-and-trade." *American Economic Journal: Economic Policy*, 12(3): 170–201.

Calel, Raphael, and Antoine Dechezleprêtre. 2016. "Environmental policy and directed technological change: evidence from the European carbon market." *Review of economics and statistics*, 98(1): 173–191.

Callaway, Brantly, and Pedro HC Sant'Anna. 2020. "Difference-in-differences with multiple time periods." *Journal of Econometrics*.

Carlson, Curtis, Dallas Burtraw, Maureen Cropper, and Karen L Palmer. 2000. "Sulfur dioxide control by electric utilities: What are the gains from trade?" *Journal of Political Economy*, 108(6): 1292–1326.

Chu, Cindy. 2009. "Thirty years later: the global growth of ITQs and their influence on stock status in marine fisheries." *Fish and Fisheries*, 10(2): 217–230.

Cohen, Mark A, and Adeline Tubb. 2018. "The impact of environmental regulation on firm and country competitiveness: A meta-analysis of the porter hypothesis." *Journal of the Association of Environmental and Resource Economists*, 5(2): 371–399.

Costello, Christopher, John Lynham, Sarah E Lester, and Steven D Gaines. 2010. "Economic Incentives and Global Fisheries Sustainability." *Annual Review of Resource Economics*, 2(1): 299–318.

Costello, Christopher, Steven D Gaines, and John Lynham. 2008. "Can catch shares prevent fisheries collapse?" *Science*, 321(5896): 1678–1681.

Dales, John H. 1968. *Pollution Property and Prices University of Toronto Press.* University of Toronto Press.

Dechezleprêtre, Antoine, and Misato Sato. 2017. "The impacts of environmental regulations on competitiveness." *Review of environmental economics and policy*.

De Loecker, Jan. 2011. "Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity." *Econometrica*, 79(5): 1407–1451.

De Loecker, Jan. 2013. "Detecting learning by exporting." *American Economic Journal: Microeconomics*, 5(3): 1–21.

De Loecker, Jan, and Frederic Warzynski. 2012. "Markups and firm-level export status." *American Economic Review*, 102(6): 2437–71.

De Loecker, Jan, and Pinelopi Koujianou Goldberg. 2014. "Firm performance in a global market." *Annual Review of Economics*, 6(1): 201–227.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. "The rise of market power and the macroeconomic implications." *Quarterly Journal of Economics*, 135(2): 561–644.

De Loecker, Jan, Pinelopi K Goldberg, Amit K Khandelwal, and Nina Pavcnik. 2016. "Prices, markups, and trade reform." *Econometrica*, 84(2): 445–510.

Demirer, Mert. 2020. "Production Function Estimation with Factor-Augmenting Technology: An Application to Markups." UPenn. Job Market Paper.

Doraszelski, Ulrich, and Jordi Jaumandreu. 2018. "Measuring the bias of technological change." *Journal of Political Economy*, 126(3): 1027–1084.

Fabra, Natalia, and Mar Reguant. 2014. "Pass-through of emissions costs in electricity markets." *American Economic Review*, 104(9): 2872–99.

Farrell, Joseph, and Carl Shapiro. 2001. "Scale Economies and Synergies in Horizontal Merger Analysis." *Antitrust Law Journal*, 68(3): 685–710.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?" *American Economic Review*, 98(1): 394–425.

Fowlie, Meredith, Stephen P Holland, and Erin T Mansur. 2012. "What do emissions markets deliver and to whom? Evidence from Southern California's NOx trading program." *American Economic Review*, 102(2): 965–93.

Fox, Kevin J, R Quentin Grafton, James Kirkley, and Dale Squires. 2003. "Property rights in a fishery: regulatory change and firm performance." *Journal of Environmental Economics and Management*, 46(1): 156–177.

Frank, Kenneth T, Brian Petrie, Jae S Choi, and William C Leggett. 2005. "Trophic cascades in a formerly cod-dominated ecosystem." *Science*, 308(5728): 1621–1623.

Gandhi, Amit, Salvador Navarro, and David A Rivers. 2020. "On the identification of gross output production functions." *Journal of Political Economy*, 128(8): 2973–3016.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*.

Grainger, Corbett A, and Dominic P Parker. 2013. "The Political Economy of Fishery Reform." *Annual Review of Resource Economics*, 5(1): 369–386.

Gray, Wayne B. 1987. "The cost of regulation: OSHA, EPA and the productivity slowdown." *American Economic Review*, 77(5): 998–1006.

Gugler, Klaus, and Ralph Siebert. 2007. "Market power versus efficiency effects of mergers and research joint ventures: evidence from the semiconductor industry." *Review of Economics and Statistics*, 89(4): 645–659.

Halpern, László, Miklós Koren, and Adam Szeidl. 2015. "Imported inputs and productivity." *American Economic Review*, 105(12): 3660–3703.

Helgesen, Irmelin Slettemoen. 2022. "ITQs, Market Power, and Efficiency Loss." *Marine Resource Economics*, 37(4): 409–435.

Isaksen, Elisabeth Thuestad, and Andries Richter. 2019. "Tragedy, property rights, and the commons: Investigating the causal relationship from institutions to ecosystem collapse." *Journal of the Association of Environmental and Resource Economists*, 6(4): 741–781.

Kaplow, Louis. 2021. "Efficiencies in Merger Analysis." *Antitrust Law Journal*, 83(3): 557–619.

Keohane, Nathaniel O. 2006. "Cost savings from allowance trading in the 1990 Clean Air Act: Estimates from a choice-based model." *Moving to markets in environmental regulation: Lessons from twenty years of experience*, 19: 23.

Kroetz, Kailin, James N Sanchirico, and Daniel K Lew. 2015. "Efficiency costs of social objectives in tradable permit programs." *Journal of the Association of Environmental and Resource Economists*, 2(3): 339–366.

Levinsohn, James, and Amil Petrin. 2003. "Estimating production functions using inputs to control for unobservables." *Review of Economic Studies*, 70(2): 317–341.

Lu, Yi, and Linhui Yu. 2015. "Trade liberalization and markup dispersion: evidence from China's WTO accession." *American Economic Journal: Applied Economics*, 7(4): 221–53.

Montgomery, W David. 1972. "Markets in licenses and efficient pollution control programs." *Journal of Economic Theory*, 5(3): 395–418.

Nærings- og fiskeridepartementet. 2003. "Strukturtiltak i kystfiskeflåten (Structural measures in the coastal fishing fleet)." Nærings- og fiskeridepartementet (Ministry of Trade and Industry). Report No. 20 to Parliament (in Norwegian).

Nærings- og fiskeridepartementet. 2006. "Strukturvirkemidler i fiskeflåten (Structural instruments in the fishing fleet)." Nærings- og fiskeridepartementet (Ministry of Trade and Industry). Public Report No. 16 (in Norwegian).

Nærings- og fiskeridepartementet. 2007. "Strukturpolitikk for fiskeflåten (Structural policy for the fishing fleet)." Nærings- og fiskeridepartementet (Ministry of Trade and Industry). Report No. 21 to Parliament (in Norwegian).

Nærings- og fiskeridepartementet. 2016. "Et fremtidsrettet kvotesystem (A future-oriented quota system)." Nærings- og fiskeridepartementet (Ministry of Trade and Industry). Public Report No. 26 (in Norwegian).

Nærings- og fiskeridepartementet. 2019. "Et kvotesystem for økt verdiskaping – En fremtidsrettet fiskerinæring (A quota system for increased value creation – A future-oriented fishing industry)." Nærings- og fiskeridepartementet (Ministry of Trade and Industry). Report No. 32 to Parliament (in Norwegian).

Natividad, Gabriel. 2016. "Quotas, productivity, and prices: The case of anchovy fishing." *Journal of Economics & Management Strategy*, 25(1): 220–257.

Newell, Richard G, James N Sanchirico, and Suzi Kerr. 2005. "Fishing quota markets." *Journal of environmental economics and management*, 49(3): 437–462.

OECD. 2001. *Domestic Transferable Permits for Environmental Management: Design and Implementation.* OECD Publishing.

OECD. 2013. "OECD Review of Fisheries: Policies and Summary Statistics."

Olley, G Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*, 64(6): 1263–1297.

Raval, Devesh. 2020. "Testing the production approach to markup estimation." *Available at SSRN 3324849*.

Rubens, Michael. 2021. "Market Structure, Oligopsony Power, and Productivity." SSRN.

Rubens, Michael. 2022. "Oligopsony Power and Factor-Biased Technology Adoption." Working paper. Yale Cowles Workshop.

Ryan, Stephen P. 2012. "The costs of environmental regulation in a concentrated industry." *Econometrica*, 80(3): 1019–1061.

Schmalensee, Richard, and Robert N Stavins. 2013. "The SO 2 allowance trading system: the ironic history of a grand policy experiment." *Journal of Economic Perspectives*, 27(1): 103–22.

Standal, Dag, and Bernt Aarset. 2008. "The IVQ regime in Norway: A stable alternative to an ITQ regime?" *Marine Policy*, 32(4): 663–668.

Standal, Dag, and Frank Asche. 2018. "Hesitant reforms: The Norwegian approach towards ITQ's." *Marine Policy*, 88: 58–63.

Stavins, Robert N. 2011. "The problem of the commons: still unsettled after 100 years." *American Economic Review*, 101(1): 81–108.

Stiebale, Joel, and Dev Vencappa. 2018. "Acquisitions, markups, efficiency, and product quality: Evidence from India." *Journal of International Economics*, 112: 70–87.

Tietenberg, Tom. 2003. "The tradable-permits approach to protecting the commons: Lessons for climate change." *Oxford Review of Economic Policy*, 19(3): 400–419.