Carbon Tax vs. Emission Trading in a Monopolistically Competitive Market with Heterogeneous Firms*

Kefu Lin[†] Rui Pan[‡] Dao-Zhi Zeng[§]

October 2, 2023

Abstract

We establish a general-equilibrium model to compare the efficiency of two emission regulation policies used worldwide: the carbon tax (CT) and the emission trading scheme (ETS). Assuming monopolistic competition and heterogeneous firms, we show that the ETS is better in an economy with a high degree of heterogeneity, and the CT is better otherwise. We also explore how the market distortions under these two regulation policies are different. Moreover, we find that the excessive input of an immobile resource in manufacturing production may result in market inefficiency.

Keywords: carbon tax, emission trading scheme, heterogeneous firm, market distortions, general equilibrium
JEL Codes: H23, L11, Q50, Q52, Q58

---

*The authors are grateful to editors Xiwei Zhu and Jian Wang, three referees, Naoya Fujiwara, Akira Hibiki, Ryo Ito, Tatsuhito Kono, and Yasuhiro Sato for their comments on earlier versions. The authors acknowledge the financial support from JSPS KAKENHI (Grant no. 20H01485) and the Environment Research and Technology Development Fund (S-19-2: JPMEERF21S11920) of Japan.

[†]Department of Economics, School of Political Science and Economics, Tokai University, 259-1292, Japan. E-mail: lin@se.is.tohoku.ac.jp

[‡]Graduate School of Information Sciences, Tohoku University, 980-8579, Japan. E-mail: pan@se.is.tohoku.ac.jp

[§]Corresponding author. Graduate School of Information Sciences, Tohoku University, 980-8579, Japan. E-mail: zeng@se.is.tohoku.ac.jp

1 Introduction

Emission reduction has been one of the most concerning issues in global climate change since the mid-20th century. Among the various regulation approaches, the carbon tax (CT) and emission trading scheme (ETS) are two of the most commonly adopted policy instruments around the world. This paper builds a two-sector general-equilibrium model with monopolistic competition and heterogeneous firms to compare the efficiency of these two policies.

In contrast to the long history of the CT, the world's first emission cap-and-trade system,[1] named SO_2 allowance-trading system, was established in the US in 1994-2010 (Stavins 2019). Meanwhile, the first multilateral trading scheme for multi-greenhouse gas emissions, named the EU ETS, was formally launched in Europe in 2005.[2]

There has been a lot of controversy over the efficiency comparison between a CT and an ETS in the literature from various perspectives.[3] One of the important arguments attributes the reasons for non-equivalence to the uncertainty and asymmetric information, which can be traced back to Weitzman (1974). He theoretically explains the difference in efficiency between the CT and ETS if the market has uncertainty and asymmetric information. His model has a single firm, and the equilibrium is considered to be in the short run. Later Spulber (1985) further compares the two policies in the long run with entry/exit of firms and shows that the effluent tax and tradeable permit are equivalent under perfect foresight and certainty. Spulber's result is amended by Shinkuma and Sugeta (2016) by incorporating asymmetric information and uncertainty. They find that in the long-run, with an entry cost, the ETS induces insufficient market entry, whereas the CT can induce either excessive or insufficient market entry. Additionally, they show that when the entry cost is sufficiently low, the ETS is always superior to the CT; otherwise, either policy can be more efficient. Meanwhile, the advantages of the ETS will be amplified by the magnitude of uncertainty and asymmetry of information.

The analysis of Shinkuma and Sugeta (2016) is based on a perfectly competitive market with a homogeneous good. However, in the real world, most dirty goods are imperfectly substitutable. In addition, emission-intensive industries are generally characterized by increasing return to scale and a large mass of firms (Zeng and Zhao 2009; Batrakova and Davies 2012; Kreickemeier and Richter 2014; Konishi and Tarui 2015;[4] Forslid et al. 2018).

[1] The "ETS" and the "cap-and-trade system" are interchangeable in our paper to indicate the trade of emission allowances.

[2] See the website of the European Commission (https://reurl.cc/eWvLVx).

[3] Many studies investigate the differences in specific design and operating elements between two policies, such as the government's acting strategy (Ishikawa and Kiyono 2006; Wirl 2012; Kiyono and Ishikawa 2013; Eichner and Pethig 2015), transaction cost (Stavins 1995; Baudry et al. 2021), and footloose capital (Lai 2022).

[4] Some specific examples and empirical facts are given in their paper (pp. 6-7).

Thus, we build a two-sector general-equilibrium model with monopolistic competition to compare the policies. As emphasized by Konishi and Tarui (2015), firm heterogeneity of Melitz (2003) is helpful for us to investigate the policy-induced effect. It can also be used to address the issue of uncertainty and asymmetric information. Therefore, we follow Shunkuma and Sugeta (2016) by assuming that the productivity of each firm is known to this firm after its entry, but unknown to the regulatory authority. We are able to extend the study of Shinkuma and Sugeta (2016) by exploring the selection effect,[5] which allows us to examine how each policy affects the market allocation and analyze the potential margins of policy inefficiency. We verify that the CT/price control can achieve an optimal mass of active firms but an insufficient mass of entrants, while the ETS/quantity control performs just the opposite. Meanwhile, the distinction of market outcome under each policy induces different labor allocations across sectors. Specifically, our study leads to the following two results.

First, the degree of productivity heterogeneity is a crucial determinant of the superiority between the two policies. Given the total emission cap, an economy with high heterogeneity does better to adopt the ETS, whereas the CT is superior in a low-heterogeneity economy. Under the CT, the government can adjust the mass of active firms, but not the mass of entrants, by controlling the lump-sum tax (subsidy). Conversely, under the ETS, the government can adjust the mass of entrants rather than the mass of active firms, by controlling the initial permit allocation. These mechanisms lead to different market outcomes, embodied by fewer/more active firms, fewer/more entrants, and a stronger/weaker selection effect. They also induce different resource allocations across sectors, resulting in different market efficiencies.

Second, we find that both the CT and ETS fail to reach the social optimum. After comparing the market allocation of either policy with the optimum, various market distortions are disclosed. Our analysis shows that under the CT, the market has a proper mass of active firms but too few entrants, which results in low average productivity. In contrast, the market under the ETS has a proper mass of entrants but too few active firms, which leads to insufficient varieties and a resource loss in the entry costs. Moreover, we verify that under either policy, excessive labor resources are allocated to the non-polluting sector, which also induces excessive emissions in the polluting production. Our results are closely related to the literature studying the market distortions in an imperfectly competitive market with one production factor (Nocco et al. 2014; Dhingra and Morrow 2019; Behrens et al. 2020). In contrast, taking emission as a resource, our model shows that the resource-allocation share has an impact on equilibrium output.

[5]The selection effect means that entrants with lower productivity are driven out of the market (Melitz 2003; Melitz and Ottaviano 2008). In this paper, a tougher selection effect indicates a higher average productivity of active firms as more low-productivity entrants are eliminated, allowing the mass of entrants and active firms to be endogenously determined. The selection effect does not work in Shinkuma and Sugeta (2016) as they assume all the entrants can produce and no firms exit.

The remainder of this paper is organized as follows. Section 2 introduces the setting and framework of the model. Section 3 analyzes and compares the equilibrium with different policies, and Section 4 calculates the optimal allocation and investigates the potential margins of distortion. Finally, Section 5 concludes the paper.

2 Model

2.1 Demand

There are two types of goods in the economy: a continuum of differentiated goods in a polluting manufacturing sector M and a homogeneous good in a clean sector A. All individuals have the same preferences, characterized by the following quasi-linear utility function (Pflüger 2004):[6]

$$U = \alpha \ln C^M + C^A, \quad C^M = \left[\int_0^n x(i)^{\frac{\sigma-1}{\sigma}} di \right]^{\frac{\sigma}{\sigma-1}}, \ \alpha \in (0,1), \ \sigma > 1, \tag{1}$$

where C^M is the manufacturing aggregate, C^A is the consumption of the composite good in sector A, n is the mass of available varieties, $x(i)$ is the consumption of variety i, and σ is the substitute elasticity between two varieties. The budget constraint of each individual is written as

$$\int_0^n p(i)x(i)dj + p_A C^A = y,$$

where $p(i)$ is the price of variety i, p_A is the price of the composite good, and y is individual income, including wages w and the transfer payment from the government. The transfer payment differs under different policies, which we will introduce in detail later.

The utility maximization yields the demand functions

$$x(i) = \alpha \frac{p(i)^{-\sigma}}{P^{1-\sigma}}, \quad C^M = \frac{\alpha}{P}, \quad C^A = \frac{y-\alpha}{p_A}, \tag{2}$$

where

$$P \equiv \left[\int_0^n p(i)^{1-\sigma} di \right]^{\frac{1}{1-\sigma}} \tag{3}$$

is the price index.

[6] Some papers, like Shinkuma and Sugeta (2016), introduce a term of environmental cost as a negative externality in the utility. Here we temporarily suppress this term and compare policies to reach the same emission target. We will show that including such an externality term in the utility does not change our results in Section 3.3.

2.2 Production

The labor endowment of the economy is L. The composite good C^A is produced with a constant return to scale technology in a competitive market and does not generate emissions. Choosing C^A as the numeraire, we have $p_A = w = 1$. Varieties in M are produced under increasing returns to scale in a monopolistic competition market, and each firm produces one variety. Specifically, after sinking f_e units of labor as an entry cost, each firm randomly draws its marginal input level $\varphi \in (0, \bar{\varphi}]$ from a Pareto distribution $G(\varphi) = (\varphi/\bar{\varphi})^k$ with density function $g(\varphi) = (k\varphi^{k-1})/\bar{\varphi}^k$. The positive constant k determines the shape of the marginal input distribution, where a smaller k indicates a higher heterogeneity. To start production, each firm needs a fixed input of F units of labor. Emission is generated when manufacturing goods are produced. Since firms can input labor for emission abatement, we follow Copeland and Taylor (1994) and Forslid et al. (2018) to treat emission as input and output simultaneously, and write the production function as

$$q(e_i, l_i, \varphi_i) = \begin{cases} \dfrac{e_i^{\beta} l_i^{1-\beta}}{\beta^{\beta}(1-\beta)^{1-\beta}\varphi_i} & \text{if } e_i \leq \varepsilon l_i, \\ 0 & \text{if } e_i > \varepsilon l_i, \end{cases}$$

leading to the emission-related cost function

$$c_i(q_i, \varphi_i) = (p^e)^{\beta} \varphi_i q_i,$$

where e_i is the emission discharge, l_i is the labor input, $\varepsilon > 0$ is the the bound on the substitution possibility between labor and emission input, p^e is the cost of per-unit emission, and β denotes the input share of emission. The government has two available policies, a CT and an ETS, under which p^e takes different forms. It equals to the tax rate t under the CT, determined by the government; whereas equals to the emission price s under the ETS, determined by the emission market. Additionally, the emission of firms with marginal cost φ_i is $e(\varphi_i) = \beta(p^e)^{\beta-1}\varphi_i q_i$.

The profit function of a firm with marginal input level φ_i can be written as

$$\pi(\varphi_i) = p(\varphi_i)q(\varphi_i) - \mathcal{T}(q, \varphi_i), \quad \mathcal{T}(q, \varphi_i) = \begin{cases} t^{\beta}\varphi_i q + f + F, & \text{CT} \\ s^{\beta}\varphi_i q - \bar{e}s + F, & \text{ETS} \end{cases}$$

where $\mathcal{T}$ is the total cost consisting of production cost and transfer payment from the government, $t(= p^e)$ is the carbon tax charged for each unit of emission and f is the lump-sum carbon tax. When the government adopts the ETS, no extra tax is charged. $\bar{e}$ is the initial emission quota of allowances allocated to each entrant, and $s(= p^e)$ is the emission price for unit emission allowance in the ETS. Moreover, we assume that the total emission target of the government is $\bar{E}$. We impose the following assumptions for the validity of our model.

Assumption 1. *Parameters are assumed to satisfy the following inequalities:*

$$\varepsilon \geq \frac{\sigma \bar{E}}{\alpha L(1-\beta)(\sigma-1)}, \tag{4}$$

$$k+1-\sigma > 0, \tag{5}$$

$$f_e < \frac{F(\sigma-1)^2}{\sigma(k+1-\sigma)}, \tag{6}$$

$$\max\left\{\frac{1+k-\sigma}{k(\sigma-1)^2}, \frac{1}{k(\sigma-1)}\right\} < \beta < 1. \tag{7}$$

Inequality (4) indicates that labor and emission are sufficiently substitutable so that the manufacturing sector is active in accordance with the production function. Inequality (5) ensures that there exist sufficient high-productivity firms. Otherwise, no firms produce in the market. Inequality (6) excludes the case that the entry cost is over-high, in which no potential entrants are willing to enter the market. Inequalities in (7) limit the intensity of emissions. If the intensity of emissions is too low in the production, the low demand for emission allowances will lead to a lower carbon price and initial transfer payment $\bar{e}s$. Consequently, the government can not attract sufficient entrants into the market, leading to the failure of the ETS policy. We impose this assumption as we want to compare the efficiency of two policies when both fully work.

3 Equilibrium under two policies

3.1 The CT policy

Following Shinkuma and Sugeta (2016), the government imposes two kinds of carbon taxes. One is a per unit emission tax t, and the other is a lump-sum tax (subsidy) f, which can be either positive or negative. We use a subscript "t" to denote the CT case. Therefore, the total tax revenue of the economy is written as

$$T = t\bar{E} + fN_t, \tag{8}$$

where N_t is the mass of active firms under a CT. The total tax revenue will be evenly redistributed to individuals.

Together with (2), the profit maximization yields

$$p_t(\varphi_i) = \frac{\sigma}{\sigma-1}t^{\beta}\varphi_i,\ q_t(\varphi_i) = \alpha L\frac{p_t(\varphi_i)^{-\sigma}}{P_t^{1-\sigma}}.$$

Defined by (3), the price index P_t under the CT policy can be rewritten as

$$P_t = \left[\int_0^{\varphi_t^*} p_t(\varphi_i)^{1-\sigma}N_t\mu_t(\varphi_i)d\varphi_i\right]^{\frac{1}{1-\sigma}} = \frac{\sigma t^{\beta}}{\sigma-1}\left(\frac{kN_t}{k-\sigma+1}\right)^{\frac{1}{1-\sigma}}\varphi_t^*,$$

where $k-\sigma+1$ is assumed to be positive according to (5), $\mu_t(\varphi_i)$ is the distribution of active firms when the government imposes tax

$$\mu_t(\varphi_i) = \frac{g(\varphi_i)}{G(\varphi_t^*)} = \frac{k\varphi_i^{k-1}}{\varphi_t^{*k}},$$

and φ_t^* is the marginal cost cutoff of active firms under the CT policy.

The zero cutoff profit condition (ZCP) is

$$\pi_t(\varphi_t^*) = \alpha L \frac{k-\sigma+1}{\sigma k N_t} - F - f = 0, \tag{9}$$

which yields

$$N_t = \frac{\alpha L(1+k-\sigma)}{\sigma k(f+F)}. \tag{10}$$

The zero expected profit (ZEP) condition indicates that

$$f_e = \int_0^{\varphi_t^*} \pi_t(\varphi_i) g(\varphi_i) d\varphi_i = \left(\frac{\varphi_t^*}{\bar{\varphi}}\right)^k \left(\frac{\alpha L}{\sigma N_t} - F - f\right). \tag{11}$$

Combining this with (10), we obtain

$$\varphi_t^* = \bar{\varphi} \left[\frac{f_e(k-\sigma+1)}{(f+F)(\sigma-1)}\right]^{\frac{1}{k}}.$$

Additionally, the mass of entrants can be derived as

$$M_t = \frac{N_t}{G(\varphi_t^*)} = \frac{\alpha L(\sigma-1)}{k\sigma f_e}.$$

Note that the emission amount of the firm with marginal cost φ_i under the CT is $e_t(\varphi_i) = \beta t^{\beta-1}\varphi_i q_t(\varphi_i)$. Therefore, the aggregate emission is written as

$$E = \int_0^{\varphi_t^*} e_t(\varphi_i) N_t \mu_t(\varphi_i) d\varphi_i = \alpha\beta L \frac{\sigma-1}{t\sigma}. \tag{12}$$

The government sets tax rate t to achieve the emission target $\bar{E}$. Equation (12) gives

$$t = \frac{\alpha\beta L(\sigma-1)}{\sigma\bar{E}}. \tag{13}$$

Interestingly, the two types of taxes play different roles in the policy. The per unit emission tax t is used to control the total emission, which does not affect the market outcome, whereas the lump-sum tax f affects the level of selection. Moreover, the CT policy can only adjust the mass of active firms through the selection effect, whereas the mass of entrants is independent of the tax.

The government chooses the optimal f to maximize the utility of a representative resident:

$$
\begin{aligned}
W_t(f) =& \alpha \ln \frac{\alpha}{P_t} + 1 - \alpha + \frac{T}{L}, \\
=& \alpha \Big(\frac{1}{k} + \frac{1}{1-\sigma} \Big) \ln(F+f) + \frac{\alpha f (k+1-\sigma)}{k\sigma(F+f)} + \frac{\alpha\beta(\sigma-1)}{\sigma} \\
& + 1 - \alpha + \alpha \ln \alpha - \alpha \ln \frac{\bar{\varphi}\sigma}{\sigma-1} \\
& - \frac{\alpha}{k} \ln \frac{f_e(k+1-\sigma)}{\sigma-1} + \frac{\alpha}{\sigma-1} \ln \frac{\alpha L}{\sigma} - \alpha\beta \ln \frac{\alpha\beta L(\sigma-1)}{\sigma\bar{E}},
\end{aligned}
$$

where the second equality comes from (8), (10), and (13). The FOC is

$$
W_t'(f) = -\frac{\alpha(F+\sigma f)(1+k-\sigma)}{\sigma k(\sigma-1)(F+f)^2} = 0 \quad \text{giving} \quad f^* = -\frac{F}{\sigma}.
$$

Moreover, we have

$$
W_t''(f^*) = -\frac{\alpha(k+1-\sigma)\sigma^2}{F^2 k(\sigma-1)^3} < 0.
$$

Therefore, the CT equilibrium is described as[7]

$$
\varphi_t^*(f^*) = \bar{\varphi} \Big[\frac{\sigma f_e(k+1-\sigma)}{F(\sigma-1)^2} \Big]^{\frac{1}{k}}, \quad N_t(f^*) = \frac{\alpha L(k+1-\sigma)}{kF(\sigma-1)}.
$$

The equilibrium welfare under the optimal CT is

$$
\begin{aligned}
W_t(f^*) =& \frac{\alpha}{k} \ln \frac{F(\sigma-1)^2}{\sigma f_e(k+1-\sigma)} + \frac{\alpha}{\sigma-1} \ln \frac{\alpha L}{F(\sigma-1)} - \frac{\alpha(k+1-\sigma)}{k\sigma(\sigma-1)} \\
& + \frac{\alpha\beta(\sigma-1)}{\sigma} + 1 - \alpha + \alpha \ln \frac{\alpha(\sigma-1)}{\sigma\bar{\varphi}} - \alpha\beta \ln \frac{\alpha\beta L(\sigma-1)}{\sigma\bar{E}}.
\end{aligned}
$$

The results show that the government needs to provide a subsidy to firms to encourage them to produce. Intuitively, taxation limits the emission level and increases firms' production costs simultaneously, resulting in insufficient producers. To reduce part of the bias, the government has to transfer some taxation into lump-sum subsidy and encourage more entrants to produce. This approach is common in the real world. For example, Bourgeois et al. (2021) find that subsidy recycling has some advantages.

3.2 The ETS policy

It is noteworthy that neither the government nor the entrants know exactly their productivity before entry. Although firms know their productivity later, such kind of private information is not available to the government. Accordingly, under the ETS policy, the

[7] Assumptions (5) and (6) ensure that $\varphi_t^* < \bar{\varphi}$ and $T > 0$ when $f = f^*$.

government allocates the free emission allowances to all entrants. Given the emission target $\bar{E}$ as the total amount of initial allowances, the government controls the mass of entrants by evenly allocating $\bar{e}$ units of free emission allowances to each of them.[8] Specifically, the mass of entrants is

$$M_e = \frac{\bar{E}}{\bar{e}}, \tag{14}$$

where a subscript "e" indicates the case under an ETS.[9]

Using emission price s in the ETS, the profit maximization yields

$$p_e(\varphi_i) = \frac{\sigma}{\sigma - 1} s^\beta \varphi_i, \quad q_e(\varphi_i) = \alpha L \frac{p_e(\varphi_i)^{-\sigma}}{P_e^{1-\sigma}}.$$

The price index defined in (3) under an ETS can be rewritten as

$$P_e = \frac{\sigma s^\beta}{\sigma - 1} \left(\frac{k N_e}{k - \sigma + 1} \right)^{\frac{1}{1-\sigma}} \varphi_e^*,$$

where $k - \sigma + 1$ is positive according to (5), N_e is the mass of active firms, and φ_e^* is the marginal cost cutoff of active firms in the ETS.

The zero cutoff profit condition becomes

$$0 = \pi(\varphi_e^*) - \bar{e}s = \alpha L \frac{k - \sigma + 1}{\sigma k N_e} - F, \tag{15}$$

which yields

$$N_e = \alpha L \frac{k - \sigma + 1}{\sigma k F}. \tag{16}$$

Note that the sales of emission allowances are not included in the operating profit, which does not affect firm's decision on production.

The distribution of active firms when government adopts the ETS is written as

$$\mu_e(\varphi_i) = \frac{g(\varphi_i)}{G(\varphi_e^*)} = \frac{k \varphi_i^{k-1}}{\varphi_e^{*k}}.$$

Moreover, we have

$$N_e = M_e G(\varphi_e^*) = \frac{\bar{E}}{\bar{e}} \left(\frac{\varphi_e^*}{\bar{\varphi}} \right)^k.$$

[8] Some papers assume that the government will allocate a fraction of initial allowances to firms freely and the rest are auctioned (Shinkuma and Sugeta 2016; Lai 2022). However, in this research, the equilibrium market outcome and social welfare remain unchanged regardless of the initial allowances allocation among the entrants or auctioned. The proof is given in Appendix A.

[9] We consider that the initial allowances are allocated to the entrants rather than active firms only to capture the essence of asymmetric information. Firm productivity is private information. The government knows the productivity distribution of all firms, but not the specific productivity level of each firm. Identifying the productivity of active firms later might incur additional costs, which is not the focus of our research. On the other hand, when the productivity information is known to both firms and the government, the market outcomes will vary with different initial allowance allocations, which have been discussed in detail by Konoshi and Tarui (2015).

Combining this with (16), we obtain the cutoff

$$\varphi_e^* = \bar{\varphi}\left[\frac{\alpha \bar{e} L(k-\sigma+1)}{\sigma k F \bar{E}}\right]^{\frac{1}{k}}.$$

The mass of firms, N_e in (16), is independent of the initial allowances allocation under the ETS. The government only controls the mass of entrants to adjust the productivity level of active firms.

Note that the emission output of firms with marginal cost φ_i in the ETS is $e_e(\varphi_i) = \beta s^{\beta-1}\varphi_i q_e(\varphi_i)$. Therefore, the emission-clearing condition under the ETS is written as

$$\bar{E} = \int_0^{\varphi_e^*} e_e(\varphi_i) N_e \mu_e(\varphi_i) d\varphi_i = \alpha\beta L \frac{\sigma-1}{s\sigma},$$

from which we can obtain the emission price in the ETS:

$$s = \frac{\alpha\beta L(\sigma-1)}{\sigma \bar{E}}.$$

The above value is identical to the tax rate (13). This equality is attributed to properties of our CES setup.

Since the mass of entrants is determined by the government, there is no free entry under the ETS. Therefore, firms may have positive net profits, which are evenly redistributed to the individuals. The total profit is

$$\begin{aligned}\Pi =& \int_0^{\varphi_e^*}\left[\frac{p_e(\varphi_i)q_e(\varphi_i)}{\sigma} - F\right] N_e \mu_e(\varphi_i) d\varphi_i - M_e f_e + \bar{E}s \\ =& \frac{\alpha L}{\sigma} - F N_e - \frac{\bar{E} f_e}{\bar{e}} + \bar{E}s.\end{aligned}$$

The government determines the initial allowances $\bar{e}$ to maximize the utility of a representative resident:

$$\begin{aligned}W_e(\bar{e}) =& \alpha \ln \frac{\alpha}{P} + 1 - \alpha + \frac{\Pi}{L} \\ =& 1 - \alpha + \alpha\beta - \frac{f_e \bar{E}}{\bar{e} L} + \frac{\alpha(\sigma - \beta k - 1)}{k\sigma} \\ & - \alpha \ln\left\{\frac{\beta L \bar{\varphi}}{\bar{E}}\left(\frac{\sigma F}{\alpha L}\right)^{\frac{1}{\sigma-1}}\left[\frac{\alpha L \bar{e}(k-\sigma+1)}{\sigma k F \bar{E}}\right]^{\frac{1}{k}}\left[\frac{\alpha\beta L(\sigma-1)}{\sigma\bar{E}}\right]^{\beta-1}\right\}.\end{aligned}$$

The FOC is

$$W_e'(\bar{e}) = \frac{f_e \bar{E}}{\bar{e}^2 L} - \frac{\alpha}{k\bar{e}} = 0 \quad \text{giving} \quad \bar{e}^* = \frac{k f_e \bar{E}}{\alpha L},$$

and the SOC is

$$W_e''(\bar{e}) = -\frac{L^2\alpha^3}{\bar{E}^2 f_e^2 k^3} < 0.$$

Thus, the equilibrium of the optimal ETS is solved out:[10]

$$\varphi_e^*(\bar{e}^*) = \bar{\varphi}\left[\frac{f_e(k+1-\sigma)}{\sigma F}\right]^{\frac{1}{k}}, \quad M_e(\bar{e}^*) = \frac{\alpha L}{k f_e}.$$

The equilibrium welfare in the ETS with the optimal initial allocation is

$$\begin{aligned} W_e(\bar{e}^*) =& 1-\alpha+\alpha\beta-\frac{\alpha(\beta k+1)}{k\sigma} \\ &-\alpha\ln\left\{\frac{\beta L\bar{\varphi}}{\bar{E}}\left(\frac{\sigma F}{\alpha L}\right)^{\frac{1}{\sigma-1}}\left[\frac{f_e(k-\sigma+1)}{\sigma F}\right]^{\frac{1}{k}}\left[\frac{\alpha\beta L(\sigma-1)}{\sigma\bar{E}}\right]^{\beta-1}\right\}. \end{aligned}$$

3.3 Comparison between the CT and the ETS

First, we compare two market outcomes:

$$\frac{\varphi_t^*}{\varphi_e^*} = \left(\frac{\sigma}{\sigma-1}\right)^{\frac{2}{k}} > 1, \; \frac{N_t}{N_e} = \frac{\sigma}{\sigma-1} > 1, \; \frac{M_t}{M_e} = \frac{\sigma-1}{\sigma} < 1. \tag{17}$$

The results are summarized as follows:

Proposition 1. *Compared to the CT policy, the economy under the ETS has more entrants, fewer active firms, and a stronger selection effect (higher average productivity).*

Proof. See (17). □

This result indicates that two policies can shape the market allocation in different ways. Under the CT policy, the government charges the per unit emission tax to achieve the emission target and transfers part of this tax revenue as subsidies to firms to reduce the distortion. Although the CT policy encourages more firms to produce compared to the ETS policy, more low-productivity firms survive, leading to a lower average productivity. In contrast, in the ETS, the government directly allocates all revenue from emissions to the entrants, which increases the expected profits and attracts more potential entrants. However, the ETS fails to encourage more firms to produce, inducing an insufficient mass of active firms.

Next, we investigate how the different market outcomes affect the equilibrium welfare:

$$\begin{aligned} \Delta W \equiv& W_t - W_e = \underbrace{\alpha\ln\frac{P_e}{P_t}}_{\text{price index gap}} + \underbrace{\frac{T-\Pi}{L}}_{\text{redistribution gap}} \\ =& \frac{\alpha}{\sigma-1}\ln\frac{N_t}{N_e} + \alpha\ln\frac{\varphi_e^*}{\varphi_t^*} + \frac{T-\Pi}{L} \\ =& \frac{\alpha(k+2-2\sigma)}{k\sigma(\sigma-1)}\left(\sigma\ln\frac{\sigma}{\sigma-1} - 1\right). \end{aligned} \tag{18}$$

[10] Assumptions (5) and (6) ensure that $\varphi_e^* < \bar{\varphi}$, while (7) ensures that $\Pi > 0$ when $\bar{e} = \bar{e}^*$.

The welfare gap can be divided into two parts: the price index gap and the redistribution gap. The price index gap results from the differences between the mass of varieties and the average productivity level. As we discussed before, the ETS is superior in the average productivity but inferior in the mass of varieties. The redistribution gap indicates the difference between the total income levels. Intuitively, the government needs to choose whether to use the revenue from emission regulations to encourage manufacturing production or to directly redistribute the revenue to households.

Interestingly, according to (18), we have $d(\Delta W)/d\bar{E} = 0$ (i.e., $dW_t/d\bar{E} = dW_e/d\bar{E}$), which indicates that the welfare gap is independent of the total emission target in our model. This property allows us to show that our results can be extended to include a term of environmental cost in the utility function of (1) (see footnote 6).

Let E be the total emission amount to be determined by the government to maximize social welfare. Denote by $U_i(E)$ the utility level of (1) under policy $i \in \{\text{CT, ETS}\}$ when the emission cap is E. We now consider the following alternative utility function

$$V_i(E) = U_i(E) + \psi(E),$$

where $\psi(E)$ is an externality term representing the emission cost of target E. The FOC for the optimal emission level under policy CT gives

$$0 = V'_{\text{CT}}(E) = U'_{\text{CT}} + \psi'(E) = U'_{\text{ETS}}(E) + \psi'(E) = V'_{\text{ETS}}(E),$$

where the 3rd equality is from (18). The last equality above suggests that the FOC for the optimal emission level under the CT policy is equivalent to that under the ETS policy. Therefore, assuming an exogenously given emission target $\bar{E}$ for both policies is consistent with the welfare-maximization behavior of the regulatory authority.[11]

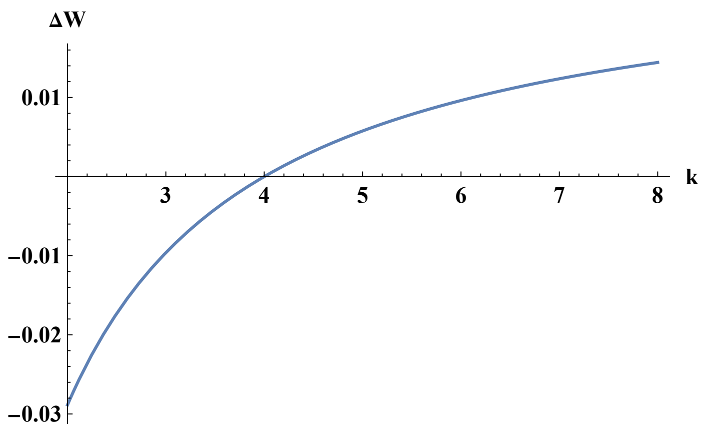
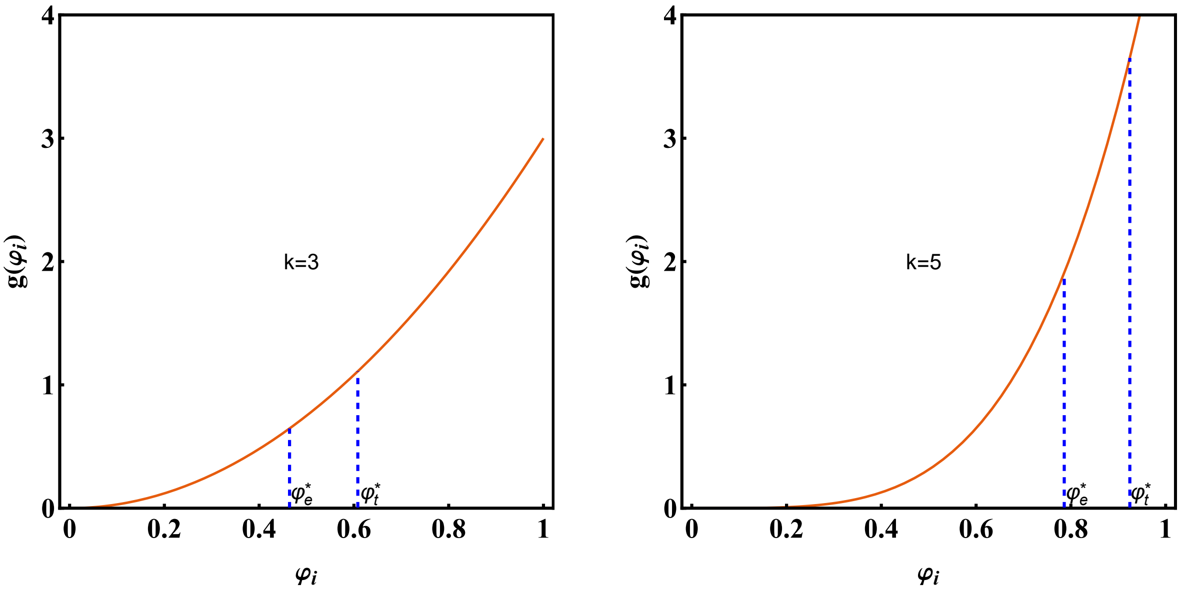


Figure 1: The welfare difference with heterogeneity

The degree of firm heterogeneity plays a distinct role in determining the relative efficiency of the two policies. We give a numerical example in Figure 1 with parameter

[11] The authors are indebted to an anonymous referee for bringing this issue to their attention.



(a) High degree of heterogeneity (b) Low degree of heterogeneity

Figure 2: The cost distribution and cutoffs with different degrees of heterogeneity

value $\alpha = 0.8,\ \sigma = 3$, where the horizontal axis is the degree of heterogeneity (a larger k indicates a lower degree of heterogeneity), and the vertical axis is the welfare difference. This result is summarized as follows.

Proposition 2. *The ETS is more efficient than the CT policy when $k \in (\sigma - 1, 2(\sigma - 1))$ and is less efficient when $k > 2(\sigma - 1)$.*

Proof. It is easy to verify that $\sigma \ln \frac{\sigma}{\sigma-1} - 1 > 0$ always holds. Therefore, $\Delta W \lesseqgtr 0$ holds when $k \lesseqgtr 2(\sigma - 1)$. □

The sharp result of Proposition 2 tells us that the ETS is better if the pollution sector has a high degree of heterogeneity, while the CT policy is better otherwise. As shown in (18), the overall welfare gap depends on the shape of the marginal cost distribution. This fact is illustrated in Figure 2. When the degree of heterogeneity is sufficiently large, the gap in the productivity level is enlarged,[12] which results in the superiority of the ETS; otherwise, the CT performs better due to a larger mass of active firms. Our result is consistent with Shinkuma and Sugeta (2016), who find that the ETS is more likely to be superior to the CT when the variance of uncertainty increases.

Moreover, we find that when a policy performs better, it may be superior in the price index gap but inferior in the redistribution gap. This indicates that apart from the distinction in market outcome, the two policies also lead to different labor allocations

[12]The gap here is a relative value. Note that φ_i denotes the marginal input level. The ratio of productivity levels of ETS to CT is written as φ_t^*/φ_e^*. Figure 2 shows that $\varphi_t^*/\varphi_e^*|_{k=3} > \varphi_t^*/\varphi_e^*|_{k=5}$, indicating a larger productivity gap between two policies when heterogeneity increases.

between sectors. This is consistent with Behrens et al. (2020), who find that one of the inefficiencies in imperfectly competitive markets comes from the mis-allocation of labor between sectors. We will further examine whether the labor allocation reaches the optimum under the more efficient policy.

4 Optimal allocation

Since either of these two policies can be better, the market distortions are not completely removed even in their best equilibria. To understand where the distortions come from, we consider the optimal allocation in this section. We use a subscript "o" in notations to indicate this optimal case. The social planner chooses the labor input, the emission input, cutoffs, and the mass of entrants to maximize the following representative utility:

$$\begin{aligned} \max_{e_o(\varphi_i), l_o(\varphi_i), M_o, \varphi_o^*} \quad & W_o = \frac{\alpha\sigma}{\sigma-1}\left\{ \ln \int_0^{\varphi_o^*} \left[\frac{q(e_i, l_i, \varphi_i)}{L}\right]^{\frac{\sigma-1}{\sigma}} M_o dG(\varphi_i) \right\} + C_o^A, \\ \text{s.t.} \quad & q(e_i, l_i, \varphi_i) = \frac{e_o(\varphi_i)^\beta l_o(\varphi_i)^{1-\beta}}{\beta^\beta (1-\beta)^{1-\beta}\varphi_i}, \\ & M_o\left\{ \int_0^{\varphi_o^*} [l_o(\varphi_i) + F] dG(\varphi_i) + f_e \right\} + C_o^A L = L, \\ & M_o\left[\int_0^{\varphi_o^*} e_o(\varphi_i) dG(\varphi_i) \right] = \bar{E}. \end{aligned} \tag{19}$$

The planner has no control over the uncertainty in drawing φ_i although he/she knows the underlying distributions $G(\varphi_i)$. In Appendix B, we apply the first-order conditions and derive the optimal solution.

First, we compare the optimal conditions with the equilibrium conditions under two policies. Incorporating $\lambda = (\alpha\beta)/\bar{E}$ from (B.10), we rewrite condition (B.7) as follows:

$$\frac{\alpha L}{(\sigma-1)M_o} = F\left(\frac{\varphi_o^*}{\bar{\varphi}}\right)^k + f_e. \tag{ZESP}$$

This condition equates the marginal social benefit of entry in M sector to its marginal social cost (Behrens et al. 2020), which can be comprehended as a zero expected social profit (ZESP) condition analogous to the ZEP conditions (11) and (14). Furthermore, plugging the results of $l(\varphi_o^*)$ and $e(\varphi_o^*)$ into (B.8), we obtain a zero cutoff social profit (ZCSP) condition similar to the ZCP conditions (9) and (15):

$$\alpha\frac{k+1-\sigma}{kM_o}\left(\frac{\bar{\varphi}}{\varphi_o^*}\right)^k = \frac{F(\sigma-1)}{L}. \tag{ZCSP}$$

We also rewrite the ZEP$_j$ ($j = t, e$) and ZCP$_j$ ($j = t, e$) conditions under two policies after plugging the results of t^*, f^*, and $\bar{e}^*$ back into (9), (11), (14), and (15) as follows (in which subscript t represents CT while subscript e represents ETS):

$$\frac{\alpha L}{\sigma M_t} = \frac{(\sigma-1)F}{\sigma}\left(\frac{\varphi_t^*}{\bar{\varphi}}\right)^k + f_e, \tag{ZEP$_t$}$$

$$\alpha\frac{k+1-\sigma}{kM_t}\left(\frac{\bar{\varphi}}{\varphi_t^*}\right)^k = \frac{F(\sigma-1)}{L}, \tag{ZCP$_t$}$$

$$M_e = \frac{\alpha L}{k f_e}, \tag{ZEP$_e$}$$

$$\alpha\frac{k+1-\sigma}{kM_e}\left(\frac{\bar{\varphi}}{\varphi_e^*}\right)^k = \frac{F}{L}. \tag{ZCP$_e$}$$

We depict the conditions for the optimum and the policy equilibria in Figure 3 and compare their outcomes by a numerical example with following parameters

$$F = 1,\ f_e = 0.3,\ \alpha = 0.8,\ \beta = 0.3,\ \bar{\varphi} = 1,\ \bar{E} = 4,\ L = 20,\ k = 3,\ \sigma = 3.$$

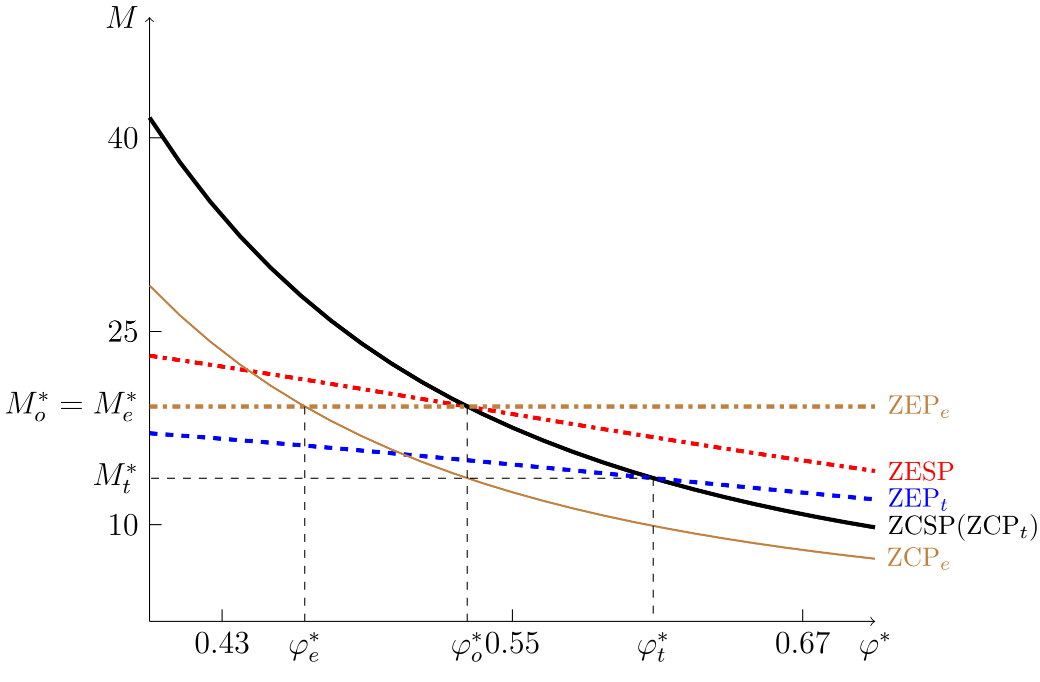


Figure 3: Comparison between equilibria and optimal conditions

Figure 3 depicts the above curves. We observe that the ZCP_t curve coincides with the ZCSP curve, while the ZEP_t curve is lower than the ZESP curve, leading to a smaller mass of entrants (i.e., $M_t^* < M_o^*$) and a larger cut-off (i.e., $\varphi_t^* > \varphi_o^*$). Notice that there is no free entry under the ETS, so that the mass of entrants is directly controlled by the government and the ZEP_e curve is a horizontal line in the figure. This helps the ETS policy achieve the optimal mass of entrants (i.e., $M_e^* = M_o^*$), while the ZCP$_e$ curve is lower than the ZCSP curve, resulting in a smaller cut-off (i.e., $\varphi_t^* > \varphi_o^*$).

Furthermore, we are able to provide analytical results for their relationships.

$$N_o : N_t : N_e = 1 : 1 : \frac{\sigma - 1}{\sigma} \ (< 1), \tag{20}$$

$$\varphi_o^* : \varphi_t^* : \varphi_e^* = 1 : \left(\frac{\sigma}{\sigma - 1}\right)^{\frac{1}{k}} \ (> 1) : \left(\frac{\sigma - 1}{\sigma}\right)^{\frac{1}{k}} \ (< 1), \tag{21}$$

$$M_o : M_t : M_e = 1 : \frac{\sigma - 1}{\sigma} \ (< 1) : 1. \tag{22}$$

Proposition 3. *Compared with the optimal allocation, (i) the economy under the CT has fewer entrants and a weaker selection, while the mass of active firms reaches the optimum; (ii) the economy under the ETS has an optimal mass of entrants but a stronger selection and a smaller mass of firms; (iii) the distortion in selection increases with the heterogeneity under the CT, and decreases under the ETS.*

Proof. (i), (ii): See (20), (21), and (22). (iii) It is straightforward that $d(\varphi_o^*/\varphi_t^*)/dk < 0$ and $d(\varphi_o^*/\varphi_e^*)/dk > 0$ hold from (21). □

In the CT case, although the mass of active firms is identical to that in the optimal allocation, the market has too few entrants, and the government has to allow more low-productivity firms to produce. Thus, the average productivity level under the CT is lower than the optimum, which becomes a main distortion. In contrast, in the ETS case, although the mass of entrants is identical to that in the optimal allocation, the government cannot force more firms to produce. Even if the average productivity level under the ETS is higher than the optimum, a smaller mass of active firms and the too-large sunk costs generate market distortions. Moreover, we find that only the distortion in cutoff is affected by the degree of heterogeneity. This bias is enlarged by the heterogeneity under the CT, whereas it is diminished under the ETS.

Then we compare the equilibrium input and output with the optimal allocation. The results are as follows:

$$\frac{e_o(\varphi_i)}{l_o(\varphi_i)} : \frac{e_t(\varphi_i)}{l_t(\varphi_i)} : \frac{e_e(\varphi_i)}{l_e(\varphi_i)} = 1 : \frac{\sigma}{\sigma - 1} : \frac{\sigma}{\sigma - 1}, \tag{23}$$

$$q_o(\varphi_i) : q_t(\varphi_i) : q_e(\varphi_i) = 1 : \left(\frac{\sigma - 1}{\sigma}\right)^{1 - \frac{\sigma - 1}{k} - \beta} : \left(\frac{\sigma - 1}{\sigma}\right)^{\frac{\sigma - 1}{k} - \beta}. \tag{24}$$

Finally, the welfare levels have the following relationships:

$$
\begin{aligned}
W_o - W_t =& \alpha(1 + \frac{1}{k} - \beta)\ln\frac{\sigma}{\sigma-1} - \frac{\alpha\,(1+k-k\beta)}{k\sigma} \\
=& \frac{\alpha}{k\sigma}\Big[(1+k-k\beta)\Big(\sigma\ln\frac{\sigma}{\sigma-1} - 1\Big)\Big] > 0,
\end{aligned} \tag{25}
$$

$$
\begin{aligned}
W_o - W_e =& \alpha\Big(\frac{1}{\sigma-1} - \frac{1}{k} + 1 - \beta\Big)\ln\frac{\sigma}{\sigma-1} \\
& - \alpha\Big[\frac{(1-\beta)(k\sigma-\sigma+1)+\beta(k+1-\sigma)}{k\sigma(\sigma-1)}\Big] \\
=& \frac{\alpha[(1-\beta)(k\sigma-\sigma+1)+\beta(k+1-\sigma)]}{k\sigma(\sigma-1)}\Big(\sigma\ln\frac{\sigma}{\sigma-1} - 1\Big) > 0.
\end{aligned} \tag{26}
$$

In summary, we have the following proposition:

Proposition 4. *Compared with the optimal allocation, (i) both CT and ETS policies result in more emissions per unit of production; (ii) the welfare difference under the CT policy is enlarged by the heterogeneity, while the welfare difference under the ETS policy is diminished; (iii) the two policies allocate less labor in the manufacturing sector, leading to a lower welfare level.*

Proof. (i) See (23), (25), and (26). (ii) Equalities of (25) and (26) give

$$
\begin{aligned}
\frac{d(W_o - W_t)}{dk} &= -\frac{\alpha}{k^2}\Big(\ln\frac{1}{1-1/\sigma} - \frac{1}{\sigma}\Big) < 0, \\
\frac{d(W_o - W_e)}{dk} &= \frac{\alpha}{k^2}\Big(\ln\frac{1}{1-1/\sigma} - \frac{1}{\sigma}\Big) > 0.
\end{aligned}
$$

(iii) We use C_j^A and L_j^A $(j = t, e, o)$ to denote the individual demand for the composite good and the labor allocated to the composite good sector, respectively. They are given by

$$
\begin{aligned}
L_t^A &= LC_t^A = (1-\alpha)L + T = L - \alpha L\Big(1 - \beta + \frac{1}{\sigma-1} - \frac{1-\beta}{\sigma} - \frac{1}{k\sigma}\Big), \\
L_e^A &= LC_e^A = (1-\alpha)L + \Pi = L - \alpha L\Big(1 - \beta + \frac{\beta}{\sigma} + \frac{1}{k\sigma}\Big), \\
L_o^A &= L - M_o\Big\{\int_0^{\varphi_o^*}[l_o(\varphi_i) + F]dG(\varphi_i) + f_e\Big\} = L - \alpha L\Big(\frac{\sigma}{\sigma-1} - \beta\Big).
\end{aligned} \tag{27}
$$

It is straightforward to verify that

$$
L_o^A < L_t^A, \quad L_o^A < L_e^A.
$$

□

This result shows that labor mis-allocation does occur in the market equilibrium even it is regulated by policies, indicating that both policies allocate too few labor resources in the manufacturing sector. Meanwhile, the bias varies with the degree of productivity heterogeneity. The ETS policy benefits from the heterogeneity, whereas the CT is harmed. This result also gives an explanation to Proposition 2 why the ETS works better in a high-heterogeneity economy.

After taking emissions into account, we find an improper proportion of input factors in the manufacturing production compared to the optimal allocation. The intuition is straightforward. We take the emission regulation as a type of resource that is immobile across sectors. Note that labor is mobile across sectors. As there are not enough labor resources in the polluting sector, firms have to input more emissions into unit production and become more emission intensive. Assuming one production factor, Nocco et al. (2014), Dhingra and Morrow (2019), and Behrens et al. (2020) disclose the potential margins of market distortion in three parts: the proper selection of active firms, the proper output of each firm, and the mis-allocation of labor resources between sectors. In our model, (27) gives a close relationship between emission intensity β and the labor allocation across sectors, showing that resource-allocation parameter β has an impact on the bias of equilibrium output when multiple factors are input. Equation (24) also shows that firms might be either over- or under-producing, which is highly dependent on the value of emission input intensity β. In the case of $\beta = 0$, labor becomes the only factor in production, in which all the manufacturing firms under-produce in market equilibria according to (24). Thus, our result is consistent with Behrens et al. (2020).

5 Conclusion

As an emerging policy, there is no doubt that an ETS can enrich the government's policy choices and bring different impacts on both economic activities and emission regulation. However, whether the market-based instrument can perform more efficiently than the conventional tax scheme remains a controversial topic in the literature. We introduce a new perspective from the degree of productivity heterogeneity, which can be taken as the inequality level of technological advancement within a nation. Our findings suggest that countries should adopt distinct emission regulation policies in their specific stages of development.

We construct a two-sector (one polluting and one clean) general equilibrium model with heterogeneous firms in a monopolistically competitive market to explore the market allocation and welfare level in both policy equilibria. The two policies result in different market outcomes. The CT/price control can adjust the cutoff of production to reach an optimal mass of active firms, but it leads to low average productivity. In contrast, the ETS/quantity control adjusts the mass of entrants to an optimal level, but it allows too few firms to produce. Our results show that in a country with a low degree of heterogeneity, it is more efficient to charge a carbon tax; otherwise, the ETS is better.

We further compare the policy equilibria with an optimal allocation and find that both policies fail to reach the social optimum. Apart from the biases in the market outcomes of the polluting sector, we verify that the mis-allocation of labor between sectors also induces the inefficiency of policy equilibria.

Compliance with Ethical Standards

Conflict of interests

The authors declare that they have no conflict of interest.

Appendices

A Initial allocation of emission allowances

In this Appendix, we assume only part of the initial allowances are allocated to the firms, while the rest are auctioned by the government. We use ξ to denote the share of allocated initial allowances. Therefore, the mass of entrants becomes

$$M_e = \frac{\xi \bar{E}}{\bar{e}}.$$

The price index is given by

$$P_e = \frac{\sigma s^{\beta}}{\sigma - 1}\left(\frac{kN_e}{k-\sigma+1}\right)^{\frac{1}{1-\sigma}} \varphi_e^*.$$

The zero cutoff profit condition remains

$$0 = \pi(\varphi_e^*) - \bar{e}s = \alpha L \frac{k-\sigma+1}{\sigma k N_e} - F,$$

which yields

$$N_e = \alpha L \frac{k-\sigma+1}{\sigma k F}. \tag{A.1}$$

The distribution of active firms is written as

$$\mu_e(\varphi_i) = \frac{g(\varphi_i)}{G(\varphi_e^*)} = \frac{k\varphi_i^{k-1}}{\varphi_e^{*k}}.$$

Moreover, we have

$$N_e = M_e G(\varphi_e^*) = \frac{\xi \bar{E}}{\bar{e}}\left(\frac{\varphi_e^*}{\bar{\varphi}}\right)^k.$$

Combining this with (A.1), we obtain the cutoff

$$\varphi_e^* = \bar{\varphi}\left[\frac{\alpha\bar{e}L(k-\sigma+1)}{\xi\sigma kF\bar{E}}\right]^{\frac{1}{k}}.$$

The emission-clearing condition is rewritten as

$$\bar{E} = \int_0^{\varphi_e^*} e_e(\varphi_i)N_e\mu_e(\varphi_i)d\varphi_i = \alpha\beta L\frac{\sigma-1}{s\sigma},$$

from which we can obtain the emission price in the ETS:

$$s = \frac{\alpha\beta L(\sigma-1)}{\sigma\bar{E}}.$$

Note that the total supply of the emission allowances still equals $\bar{E}$. The price of auctioned allowances should be equal to the transaction price in the emission market. The difference is that the auctioned revenue belongs to the government, which is redistributed to the individuals later.

The total profit of firms is

$$\begin{aligned}\Pi =& \int_0^{\varphi_e^*}\left[\frac{p_e(\varphi_i)q_e(\varphi_i)}{\sigma} - F\right]N_e\mu_e(\varphi_i)d\varphi_i - M_e f_e + \xi\bar{E}s\\ =&\frac{\alpha L}{\sigma} - FN_e - \frac{\xi\bar{E}f_e}{\bar{e}} + \xi\bar{E}s.\end{aligned}$$

The government determines the initial allowance $\bar{e}$ to maximize the utility of a representative resident:

$$\begin{aligned}W_e(\bar{e}) =&\alpha\ln\frac{\alpha}{P_e} + 1 - \alpha + \frac{\Pi}{L} + \frac{(1-\xi)\bar{E}s}{L}\\ =&1 - \alpha + \alpha\beta - \frac{f_e\xi\bar{E}}{\bar{e}L} + \frac{\alpha(\sigma-\beta k-1)}{k\sigma}\\ &- \alpha\ln\left\{\frac{\beta L\bar{\varphi}}{\bar{E}}\left(\frac{\sigma F}{\alpha L}\right)^{\frac{1}{\sigma-1}}\left[\frac{\alpha L\bar{e}(k-\sigma+1)}{\sigma kF\xi\bar{E}}\right]^{\frac{1}{k}}\left[\frac{\alpha\beta L(\sigma-1)}{\sigma\bar{E}}\right]^{\beta-1}\right\}.\end{aligned}$$

The FOC is

$$W_e'(\bar{e}) = \frac{f_e\xi\bar{E}}{\bar{e}^2L} - \frac{\alpha}{k\bar{e}} = 0 \quad \text{giving} \quad \bar{e}^* = \frac{kf_e\xi\bar{E}}{\alpha L},$$

and the SOC is

$$W_e''(\bar{e}) = -\frac{L^2\alpha^3}{\bar{E}^2f_e^2k^3\xi^2} < 0.$$

Thus, the equilibrium of the optimal ETS is solved out:

$$\varphi_e^*(\bar{e}^*) = \bar{\varphi}\left[\frac{f_e(k+1-\sigma)}{\sigma F}\right]^{\frac{1}{k}}, \quad M_e(\bar{e}^*) = \frac{\alpha L}{kf_e}.$$

Therefore, the market outcome and social welfare are independent of the initial allocation share ξ.

B Optimal allocation

For convenience, we rewrite the optimal problem (19) below:

$$\max_{e_o(\varphi_i),l_o(\varphi_i),M_o,\varphi_o^*} W_o = \frac{\alpha\sigma}{\sigma-1}\left\{\ln\int_0^{\varphi_o^*}\left[\frac{q(e_i,l_i,\varphi_i)}{L}\right]^{\frac{\sigma-1}{\sigma}} M_o dG(\varphi_i)\right\} + C_o^A, \tag{B.1}$$

$$\text{s.t.} \qquad q(e_i,l_i,\varphi_i) = \frac{e_o(\varphi_i)^\beta l_o(\varphi_i)^{1-\beta}}{\beta^\beta(1-\beta)^{1-\beta}\varphi_i}, \tag{B.2}$$

$$M_o\Big\{\int_0^{\varphi_o^*}[l_o(\varphi_i)+F]dG(\varphi_i)+f_e\Big\} + C_o^A L = L, \tag{B.3}$$

$$M_o\Big[\int_0^{\varphi_o^*} e_o(\varphi_i)dG(\varphi_i)\Big] = \bar{E}. \tag{B.4}$$

We solve for C_o^A from (B.3) and plugging the result and (B.2) into (B.1). Let λ denote the Lagrange multiplier associated with (B.4). The first-order conditions are written as

$$\frac{dW_o}{dl_o(\varphi_i)} = \alpha(1-\beta)M_o\left[\frac{e_o(\varphi_i)^\beta l_o(\varphi_i)^{\frac{(1-\beta)(\sigma-1)-\sigma}{\sigma-1}}}{\beta^\beta(1-\beta)^{1-\beta}\varphi_i LC^M}\right]^{\frac{\sigma-1}{\sigma}} - \frac{M_o}{L} = 0, \tag{B.5}$$

$$\frac{dW_o}{de_o(\varphi_i)} = \alpha\beta M_o\left[\frac{e_o(\varphi_i)^{\frac{\beta(\sigma-1)-\sigma}{\sigma-1}} l_o(\varphi_i)^{1-\beta}}{\beta^\beta(1-\beta)^{1-\beta}\varphi_i LC^M}\right]^{\frac{\sigma-1}{\sigma}} - \lambda M_o = 0, \tag{B.6}$$

$$\begin{aligned}\frac{dW_o}{dM_o} =& \frac{\alpha\sigma}{(\sigma-1)M_o} - \frac{1}{L}\Big[\int_0^{\varphi_o^*}\left(l_o(\varphi_i)+F\right)dG(\varphi_i)+f_e\Big] \\ & - \lambda\Big[\int_0^{\varphi_o^*} e_o(\varphi_i)dG(\varphi_i)\Big] \\ =& \frac{\alpha\sigma}{(\sigma-1)M_o} - \frac{1}{L}\Big[\frac{\alpha(1-\beta)L}{M_o} + F\Big(\frac{\varphi_o^*}{\bar{\varphi}}\Big)^k + f_e\Big] - \lambda\frac{\bar{E}}{M_o} = 0,\end{aligned} \tag{B.7}$$

$$\begin{aligned}\frac{dW_o}{d\varphi_o^*} =& \frac{k\alpha\sigma M_o\varphi_o^{*k-1}}{\bar{\varphi}^k(\sigma-1)}\left[\frac{e_o(\varphi_o^*)^\beta l_o(\varphi_o^*)^{1-\beta}}{\beta^\beta(1-\beta)^{1-\beta}\varphi_o^* LC^M}\right]^{\frac{\sigma-1}{\sigma}} \\ & - M_o\frac{k\varphi_o^{*k-1}}{\bar{\varphi}^k}\frac{[l_o(\varphi_o^*)+F]}{L} - \lambda M_o\frac{k\varphi_o^{*k-1}}{\bar{\varphi}^k}e_o(\varphi_o^*) = 0.\end{aligned} \tag{B.8}$$

Since (B.5) and (B.6) hold for any firm with $\varphi_i \le \varphi_o^*$, we then have

$$\left[\frac{l_o(\varphi_i)}{l_o(\varphi_j)}\right]^{\frac{(1-\beta)(\sigma-1)-\sigma}{\sigma}}\left[\frac{\frac{e_o(\varphi_i)^\beta}{\varphi_i}}{\frac{e_o(\varphi_j)^\beta}{\varphi_j}}\right]^{\frac{\sigma-1}{\sigma}} = 1,$$

$$\left[\frac{\frac{l_o(\varphi_i)^{1-\beta}}{\varphi_i}}{\frac{l_o(\varphi_j)^{1-\beta}}{\varphi_j}}\right]^{\frac{\sigma-1}{\sigma}}\left[\frac{e_o(\varphi_i)}{e_o(\varphi_j)}\right]^{\frac{\beta(\sigma-1)}{\sigma}-1} = 1,$$

which can be used to derive

$$\frac{e_o(\varphi_i)}{e_o(\varphi_j)} = \frac{l_o(\varphi_i)}{l_o(\varphi_j)} = \left(\frac{\varphi_i}{\varphi_j}\right)^{1-\sigma}.$$

Substituting this condition back into the emission market clearing (B.4), we have

$$e_o(\varphi_i) = \frac{k+1-\sigma}{k}\,\frac{\varphi_i^{1-\sigma}\bar{\varphi}^k\bar{E}}{M_o\varphi_o^{*k+1-\sigma}}. \tag{B.9}$$

Multiplying l_i and integrating both sides of (B.5), we derive

$$\alpha(1-\beta) - \frac{1}{L}\int_0^{\varphi_o^*} l_o(\varphi_i)M_o dG(\varphi_i) = 0.$$

Multiplying e_i and integrating both sides of (B.6), we obtain

$$\alpha\beta - \lambda\int_0^{\varphi_o^*} e_o(\varphi_i)M_o dG(\varphi_i) = 0, \quad \text{which gives} \quad \lambda = \frac{\alpha\beta}{\bar{E}}. \tag{B.10}$$

Substituting λ back into (B.6) and combining that with (B.5), we can get the ratio of optimal labor and emission input for each variety

$$\frac{l_o(\varphi_i)}{e_o(\varphi_i)} = \frac{\alpha L(1-\beta)}{\bar{E}}. \tag{B.11}$$

Equations (B.9) and (B.11) imply that

$$l_o(\varphi_i) = \alpha L(1-\beta)(k+1-\sigma)\frac{\varphi_i^{1-\sigma}\bar{\varphi}^k}{kM_o\varphi_o^{*k+1-\sigma}}.$$

After substituting $l(\varphi_i)$, $e(\varphi_i)$, and λ into (B.7) and (B.8), we can solve out N_o and φ_o^*. Finally, the optimal value of endogenous variables is rewritten as

$$\begin{aligned}
e_o(\varphi_i) &= \varphi_i^{1-\sigma}\frac{\bar{E}F(\sigma-1)}{\alpha L}\left[\frac{f_e\bar{\varphi}^k(k+1-\sigma)}{F(\sigma-1)}\right]^{\frac{\sigma-1}{k}}, \\
l_o(\varphi_i) &= \varphi_i^{1-\sigma}F(1-\beta)(\sigma-1)\left[\frac{f_e\bar{\varphi}^k(k+1-\sigma)}{F(\sigma-1)}\right]^{\frac{\sigma-1}{k}}, \\
q_o(\varphi_i) &= \varphi_i^{-\sigma}\frac{F\bar{E}^\beta(\sigma-1)}{(\alpha\beta)^\beta L^\beta}\left[\frac{f_e\bar{\varphi}^k(k+1-\sigma)}{F(\sigma-1)}\right]^{\frac{\sigma-1}{k}}, \\
M_o &= \frac{\alpha L}{kf_e},\ \varphi_o^* = \bar{\varphi}\left[\frac{f_e(k+1-\sigma)}{F(\sigma-1)}\right]^{\frac{1}{k}},\ N_o = \frac{\alpha L(k+1-\sigma)}{kF(\sigma-1)}.
\end{aligned}$$

References

Batrakova, S., & Davies, R. B. (2012), "Is there an environmental benefit to being an exporter? Evidence from firm-level data", *Review of World Economics*, 148, 449-474.

Baudry, M., Faure, A., & Quemin, S. (2021), "Emissions trading with transaction costs", *Journal of Environmental Economics and Management*, 108, 102468.

Behrens, K., Mion, G., Murata, Y., & Suedekum, J. (2020), "Quantifying the gap between equilibrium and optimum under monopolistic competition," *Quarterly Journal of Economics* 135, 2299-2360.

Bourgeois, C., Giraudet, L.G., & Quirion, P. (2021), "Lump-sum vs. energy-efficiency subsidy recycling of carbon tax revenue in the residential sector: A French assessment", *Ecological Economics*, 184, 107006.

Copeland, B.R. & Taylor, M.S. (1994), "North-South trade and the environment", *The quarterly journal of Economics* 109(3), 755-787.

Dhingra, S. & Morrow, J. (2019), "Monopolistic competition and optimum product diversity under firm heterogeneity", *Journal of Political Economy* 127(1), 196-232.

Eichner, T. & Pethig, R. (2015), "Self-enforcing international environmental agreements and trade: taxes versus caps", *Oxford Economic Papers*, 67(4), 897-917.

Forslid, R., Okubo, T., & Ulltveit-Moe, K.H. (2018), "Why are firms that export cleaner? International trade, abatement and environmental emissions", *Journal of Environmental Economics and Management*, 91, 166-183.

Ishikawa, J. & Kiyono, K. (2006), "Greenhouse-gas emission controls in an open economy", *International Economic Review*, 47(2), 431-450.

Lai, Y.B, (2022), "Capital mobility and environmental policy: Taxes versus TEP", *International Tax and Public Finance*, 1-25.

Kiyono, K. & Ishikawa, J. (2013), "Environmental management policy under international carbon leakage", *International Economic Review*, 54(3), 1057-1083.

Konishi, Y. & Tarui, N. (2015), "Emissions trading, firm heterogeneity, and intra-industry reallocations in the long run", *Journal of the Association of Environmental and Resource Economics*, 2(1), 1-42.

Kreickemeier, U., & Richter, P. M. (2014), "Trade and the environment: The role of firm heterogeneity", *Review of International Economics*, 22(2), 209-225.

Melitz, M.J. (2003). "The impact of trade on intra-industry allocation and aggregate industry Productivity", *Econometrica*, 71, 1695-1725.

Melitz, M.J. & Ottaviano, G.I. (2008), "Market size, trade, and productivity", *The review of economic studies* 75(1), 295-316.

Nocco, A., Ottaviano, G.I., & Salto, M. (2014), "Monopolistic competition and optimum product selection", *American Economic Review* 104(5), 304-09.

Pflüger, M. (2004), "A simple, analytically solvable, Chamberlinian agglomeration model", *Regional science and urban economics*, 34(5), 565-573.

Shinkuma, T. & Sugeta, H. (2016), "Tax versus emissions trading scheme in the long run", *Journal of Environmental Economics and Management* 75, 12-24.

Spulber, D.F. (1985), "Effluent regulation and long-run optimality", *Journal of Environmental Economics and Management* 12(2), 103-116.

Stavins, R.N. (1995), "Transaction costs and tradeable permits", *Journal of Environmental Economics and Management*, 29: 133-146.

Stavins, R.N. (2019), "Carbon taxes vs. cap and trade: theory and practice", *Cambridge, Mass.: Harvard Project on Climate Agreements.*

Weitzman, M.L. (1974), "Prices vs. quantities", *Review of Economic Studies*, 41(4): 477-491.

Wirl, F. (2012), "Global warming: prices versus quantities from a strategic point of view", *Journal of Environmental Economics and Management*, 64(2), 217-229.

Zeng, D.-Z., & Zhao, L. (2009), "Pollution havens and industrial agglomeration", *Journal of Environmental Economics and Management*, 58(2), 141-153.