Optimal carbon tax and fuel switching

Emmanuel Murray Leclair

April 2021

1 Model

1.1 Aggregation:

Fraction $N_d \in (0,1)$ of firms who use dirty fuel and $N_c \in (0,1)$ who use clean fuel. Continuum of firms indexed by productivity φ who produce differentiated goods substitutable at rate $\rho > 1$ to form final good Y which is consumed by a representative consumer:

$$Y = \Big(\int_{\Phi_d} y_d(\varphi)^{\frac{\rho-1}{\rho}} dF(\varphi) + \int_{\Phi_c} y_c(\varphi)^{\frac{\rho-1}{\rho}} dF(\varphi)\Big)^{\frac{\rho}{\rho-1}}$$

The representative consumer chooses the utility-maximizing bundle of each differentiated good that forms Y, taking as given prices and purchasing each good with some exogenous income I that I normalize to 1. for $s = \{c, d\}$ This leads to the following inverse demand function:

$$p_s(\varphi) = \Big(\frac{Y}{y_s(\varphi)}\Big)^{\frac{1}{\rho}} P \tag{1}$$

Where P is the price index of the final good:

$$P = \Big(\int_{\Phi_d} p_d(\varphi)^{1-\rho} dF(\varphi) + \int_{\Phi_c} p_c(\varphi)^{1-\rho} dF(\varphi)\Big)^{\frac{1}{1-\rho}}$$

1.2 Firms:

The continuum of firms is initially distributed into two groups: dirty input users: N_d^* and clean input users: N_c^*. Firms only use fuels (e_d, e_c) in production, can switch between fuels by paying a fixed cost κ and engage in monopolistic competition by internalizing their inverse demand. The profit function of dirty fuel users is as follows:

$$\pi_d(\varphi) = \max\Big\{\max_{e_d}\big\{p(\varphi)\varphi e_d - \tilde{p}_d e_d\big\}, \max_{e_c}\big\{p(\varphi)\varphi e_c - \tilde{p}_c e_c - \kappa\big\}\Big\} \tag{2}$$

Where $\tilde{p}_d = p_d + \tau_d$ is the price fuel d inclusive of its potential tax. An analogous profit function can be derived for firms initially using the clean fuel. Note that the solution to problem (2) is mathematically equivalent to a formulation where both fuels are perfect substitutes but where the firm needs to pay a fixed cost for switching between the two:

$$f(\varphi) = \varphi(e_d + e_c)$$

$$\pi_d(\varphi) \equiv \max_{e_d, e_c} \left\{ p(\varphi)\varphi(e_d + e_c) - \tilde{p}_d e_d - \tilde{p}_c e_c - \mathbb{1}(e_c > 0)\kappa \right\}$$

This will have important implications to understand the intuition underlying optimal fossil fuel taxation in this context.

$$\pi_d = \max_{e_d, e_c} \left\{ Y^{1/\rho} P(\varphi(e_d + e_c))^{\frac{\rho-1}{\rho}} - \tilde{p}_d e_d - \tilde{p}_c e_c - \mathbb{1}(e_c > 0)\kappa \right\}$$

1.3 Solution

Demand for both fuels:

$$e_d(\varphi) = \Big(\frac{\rho - 1}{\rho} \frac{P}{\tilde{p}_d}\Big)^{\rho} Y \varphi^{\rho-1}$$
$$e_c(\varphi) = \Big(\frac{\rho - 1}{\rho} \frac{P}{\tilde{p}_c}\Big)^{\rho} Y \varphi^{\rho-1}$$

Profit for not switching:

$$\pi_{dd}(\varphi) = \frac{Y P^{\rho} \varphi^{\rho-1}}{p_d^{\rho-1}} \underbrace{\Big[\Big(\frac{\rho-1}{\rho}\Big)^{\rho-1} - \Big(\frac{\rho-1}{\rho}\Big)^{\rho}\Big]}_{\Gamma(\rho)>0}$$

Profits for switching:

$$\pi_{dc}(\varphi) = \frac{Y P^{\rho} \varphi^{\rho-1}}{p_c^{\rho-1}} \Gamma(\rho) - \kappa$$

Fuel switching rule:

By comparing profits for using both fuels if the firms was initially using dirty fuel will lead to the following decision rule a the extensive margin based on a productivity threshold:

Let Ω define a threshold for productivity:

$$\Omega = \left[\frac{\kappa}{P^\rho Y \Gamma(\rho)} \frac{1}{p_c^{1-\rho} - p_d^{1-\rho}}\right]^{\frac{1}{\rho-1}}$$

Where $\Gamma(\rho) = \left(\frac{\rho-1}{\rho}\right)^{\rho-1} - \left(\frac{\rho-1}{\rho}\right)^{\rho}$

Let N_{dc} denote the switching for dirty fuel users switching to clean fuel ($d \to c$), and likewise for N_{dd}, N_{cd}, N_{cc}, there are two cases to consider:

Case 1: $p_d > p_c, \Omega > 0$

		New fuel	
		d	c
Original fuel	d	$\varphi < \Omega$	$\varphi > \Omega$
	c	$\varphi < -\Omega$	$\varphi > -\Omega$

$$E_d = \underbrace{\left[N_d^* \int_0^\Omega \varphi^{\rho-1} dF(\varphi) + N_c^* \int_0^{-\Omega} \varphi^{\rho-1} dF(\varphi)\right]}_{\tilde{\varphi}_d^{\rho-1}} \left(\frac{\rho-1}{\rho}\frac{P}{\tilde{p}_d}\right)^\rho Y$$

$$E_c = \underbrace{\left[N_d^* \int_\Omega^\infty \varphi^{\rho-1} dF(\varphi) + N_c^* \int_{-\Omega}^{\infty} \varphi^{\rho-1} dF(\varphi)\right]}_{\tilde{\varphi}_c^{\rho-1}} \left(\frac{\rho-1}{\rho}\frac{P}{\tilde{p}_c}\right)^\rho Y$$

Case 2: $p_c > p_d, \Omega < 0$

		New fuel	
		d	c
Original fuel	d	$\varphi > \Omega$	$\varphi < \Omega$
	c	$\varphi > -\Omega$	$\varphi < -\Omega$

$$E_d = \underbrace{\left[N_d^* \int_\Omega^\infty \varphi^{\rho-1} dF(\varphi) + N_c^* \int_{-\Omega}^{\infty} \varphi^{\rho-1} dF(\varphi)\right]}_{\tilde{\varphi}_d^{\rho-1}} \left(\frac{\rho-1}{\rho}\frac{P}{\tilde{p}_d}\right)^\rho Y$$

$$E_c = \underbrace{\left[N_d^* \int_0^\Omega \varphi^{\rho-1} dF(\varphi) + N_c^* \int_0^{-\Omega} \varphi^{\rho-1} dF(\varphi)\right]}_{\tilde{\varphi}_c^{\rho-1}} \left(\frac{\rho-1}{\rho}\frac{P}{\tilde{p}_c}\right)^\rho Y$$

Proposition 1. *There is no two-way switching (There is always an integration region that is empty and one that has full support)*

Indeed, in case 1 where $p_d > p_c$, $\Omega > 0$ and $-\Omega < 0$ such that no firms switches from the clean fuel to the dirty fuel. This is intuitive because they face a higher unit cost to use the dirty fuel while having to pay a fixed cost on top of that, hence it is never profitable for them to switch. An analogous argument can be made for case 2.

Aggregate price index

Regardless of which fuel firms use, individual pricing decisions are always a constant markup over marginal costs:

$$p_s(\varphi) = \frac{\rho}{\rho - 1}\frac{\tilde{p}_d}{\varphi}$$

Let Φ_d and Φ_c denote the integration regions for using fuels d and c, respectively, then the aggregate price index in this economy is:

$$\begin{aligned} P &= \frac{\rho}{\rho - 1}\left(\tilde{p}_d^{1-\rho}\int_{\Phi_d}\varphi^{\rho-1}dF(\varphi) + \tilde{p}_c^{1-\rho}\int_{\Phi_c}\varphi^{\rho-1}dF(\varphi)\right)^{\frac{1}{1-\rho}} \\ &= \frac{\rho}{\rho - 1}\left(\tilde{p}_d^{1-\rho}\tilde{\varphi}_d^{\rho-1} + \tilde{p}_c^{1-\rho}\tilde{\varphi}_c^{\rho-1}\right)^{\frac{1}{1-\rho}} \end{aligned}$$

1.4 GHG emission and externality

For a given firm, GHG emissions are defined as follows, where $\gamma_d > \gamma_c$ defines the emission intensity of one unit (in equivalent energy units) of the dirty and clean fuel, respectively:

$$ghg(\varphi) = \gamma_d e_d(\varphi) + \gamma_c e_c(\varphi)$$

Which aggregates to:

$$\begin{aligned} ghg &= \gamma_d \int_{\Phi_d} e_d(\varphi)dF(\varphi) + \gamma_c \int_{\Phi_c} e_c(\varphi)dF(\varphi) \\ &= \gamma_d E_d + \gamma_c E_c \end{aligned}$$

To get some intuition on how variation in fuel prices affect ghg emissions, let ghg_d be part of aggregate emissions from using fossil fuel d, where $ghg = ghg_d + ghg_c$. Then,

$$\frac{d\log ghg_d}{d\log \tilde{p}_d} = \underbrace{\frac{1}{1-\rho}\frac{d\log\tilde{\varphi}_d}{d\log\tilde{p}_d}}_{\text{Agg. prod. ("switching") channel <0}} + \underbrace{\rho\frac{d\log P}{d\log\tilde{p}_d}}_{\text{Agg price ("Competitive") channel >0}} - \underbrace{\rho\frac{d\log\tilde{p}_d}{d\log\tilde{p}_d}}_{\text{own price channel}}$$

$$\frac{d\log ghg_c}{d\log \tilde{p}_d} = \underbrace{\frac{1}{1-\rho}\frac{d\log \tilde{\varphi}_c}{d\log \tilde{p}_d}}_{\text{Agg. prod. ("switching") channel >0}} + \underbrace{\rho\frac{d\log P}{d\log \tilde{p}_d}}_{\text{Agg price ("Competitive") channel >0}}$$

I then define the externality as multiplicative damages following the work of Nordhaus and Golosov et. al (2014), where the function $D(ghg) \in (0,1)$ maps damages from CO2e emissions in units of the final good and where γ_g define the extent of damages per units of CO2e. Hence, real output net of the pollution externality is given by:

$$\tilde{Y} = \big(1 - D(ghg)\big)Y$$
$$1 - D(ghg) = \exp\left(-\gamma_g \times ghg\right)$$

1.5 Optimal tax on fossil fuels

In the spirit of Mirless (1972) I study optimal taxation by defining the problem of a government to choose a tax on both fuels to maximize the representative worker's utility, taking into account the externality and all decisions made in a competitive equilibrium. I first study optimal taxation in partial equilibrium, which means that the government effectively can choose the price of each fuel $\tilde{p}_s$ which pins down the tax rate τ_s (Mirless 1972). This is because fuel prices are exogenous from the model, and from the perspective of the government, any price that isn't optimal can be corrected for. General equilibrium will be considered later. Moreover, the interest here is on the relative tax rates between both fuels rather than the level of the tax rates. The government then solves the following problem, taking into account the externality:

$$\max_{\tilde{p}_d,\tilde{p}_c}\{U(C)\}$$
$$s.t.\ PY = I$$
$$C = \underbrace{\exp\left(-\gamma_g \times ghg\right)Y}_{\tilde{Y}}$$

Where the representative worker's income is normalized to 1. Moreover, I assume that the representative worker consumes the proceeds Hence, this problem becomes:

$$\max_{\tilde{p}_d,\tilde{p}_c}\Big\{U\Big(\frac{\exp\left(-\gamma_g \times ghg\right)}{P}\Big)\Big\}$$

Where the fundamental economic trade-off from the government's perspective is that increasing taxes on fossil fuels increases real output by decreasing the externality but also decreases real output due to the increase in the aggregate price index. First conditions are as follows:

$$U'(C)\exp(-\gamma_g ghg)\Big(\frac{-1}{P^2}\frac{\partial P}{\partial \tilde{p}_s} - \frac{\gamma_g}{P}\frac{\partial ghg}{\partial \tilde{p}_s}\Big) = 0$$

$$\frac{1}{P^2}\frac{\partial P}{\partial p_s} = -\frac{\gamma_g}{P}\frac{\partial ghg}{\partial \tilde{p}_s}$$

To learn about the optimal relative tax rate, most of the intuition can be derived from the ratio of FOCs for both fuels. The government's objective is to find relative fuel prices that equate relative marginal losses (by increasing the aggregate price index which decreases real income) to relative marginal gains (by decreasing aggregate GHG emissions):

$$\underbrace{\frac{\partial P/\partial \tilde{p}_d}{\partial P/\partial \tilde{p}_c}}_{\text{rel. marginal losses}} = \underbrace{\frac{\partial ghg/\partial \tilde{p}_d}{\partial ghg/\partial \tilde{p}_c}}_{\text{rel. marginal gains}}$$

A useful benchmark

The first case to consider is the case where there is no switching in the model, that is where firms always use the technology they start with. In such a case, there is no selection into different fuels based productivity, and the economy is equivalent to an economy with an aggregate CES production that takes two inputs (the dirty and the clean fuel) which can be substituted at rate ρ:[1]

$$Y = \tilde{\varphi}\Big[e_1^{\frac{\rho-1}{\rho}} + e_2^{\frac{\rho-1}{\rho}}\Big]^{\frac{\rho}{\rho-1}}$$

Where

$$\tilde{\varphi} = \Big(\int_0^\infty \varphi^{\rho-1} dF(\varphi)\Big)^{\frac{1}{\rho-1}}$$

This is important because this is the type of aggregate production functions that Golosov et al. (2014) use to study optimal carbon taxation, and this allows me to highlight how the presence of fuel switching departs from canonical assumptions in the literature. Under the no switching assumptions, the solution to the optimal tax problem is standard, and the relative difference in prices reflects the relative emission intensity of both fuels, which can be seen in the following graph:

$$\frac{\tilde{p}_d}{\tilde{p}_c} = \frac{\gamma_d}{\gamma_c}$$

[1] See appendix for proof.

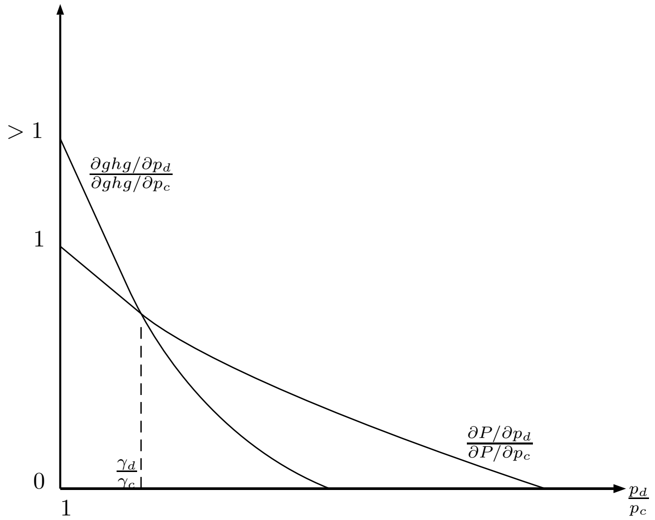

Figure 1: Optimal relative fuel prices (taxes) - no switching benchmark

Note that the marginal relative gains in reducing emissions are greater than one when prices are equal to one due to the difference in the emission intensity of both fuels $\gamma_d > \gamma_c$, whereas marginal effects on the aggregate price index is the same when prices are equal.

Case with switching

In the presence of fuel switching, it is still optimal for the government to choose fuel prices (tax rates) in the range $p_d > p_c$ because $\gamma_d > \gamma_c$. However, conditional on $p_d > p_c$, there are two new effects on the relative marginal gains/losses from increasing p_d relative to increasing p_c that are going to shift the optimal relative prices.

First, increasing the price of the dirty fuel *increases* the output price of firms that are not productive enough to switch to the clean fuel, but *decreases* the output price of firms that do switch because they are now facing lower input costs. Overall, increasing p_d increases P but the magnitude of this change is smaller than in the no switching benchmark whereas the opposite is true for increasing p_c. This essentially shift down the $\frac{\partial P/\partial p_d}{\partial P/\partial p_c}$ in figure 1:

$$\frac{\partial P}{\partial p_d} = \underbrace{\text{benchmark effect}}_{>0} + \underbrace{\frac{1}{1-\rho}\Big[(p_d^{1-\rho} - p_c^{1-\rho})N_d \Omega^{\rho-1}\frac{\partial \Omega}{\partial p_d}\Big]}_{\text{Selection effect due to switching <0}}$$

Second, increasing the price of the dirty fuel decreases ghg emissions for firms that don't switch due to reduction in scale of production (as in the benchmark), but decreases even more ghg emissions for firms that switch to the clean fuel and now pollute at rate γ_c rather than γ_d:

$$\frac{\partial ghg}{\partial p_d} = \underbrace{\text{benchmark effect}}_{<0} + \underbrace{\left(\frac{\rho-1}{\rho}\right)^{\rho}\Big[N_d\Omega^{\rho-1}\frac{\partial\Omega}{p_d}\Big(\frac{1}{1-\rho}(p_d^{1-\rho}-p_c^{1-\rho})+(\gamma_d-\gamma_c)\Big)\Big]}_{\text{Selection effect due to switching}<0}$$

Overall, the net effect of introducing switching is that relative marginal losses are smaller when increasing p_d relative to p_c while relative marginal gains are larger when increasing p_d relative to p_c. This effectively leads to optimal relative fuel prices that are much larger than with a carbon tax:

$$\frac{\partial P/\partial\tilde{p}_d}{\partial P/\partial\tilde{p}_c}\Big|_{\text{switching}} < \frac{\partial P/\partial\tilde{p}_d}{\partial P/\partial\tilde{p}_c}\Big|_{\text{no switching}}$$
$$\frac{\partial ghg/\partial\tilde{p}_d}{\partial ghg/\partial\tilde{p}_c}\Big|_{\text{switching}} > \frac{\partial ghg/\partial\tilde{p}_d}{\partial ghg/\partial\tilde{p}_c}\Big|_{\text{no switching}}$$

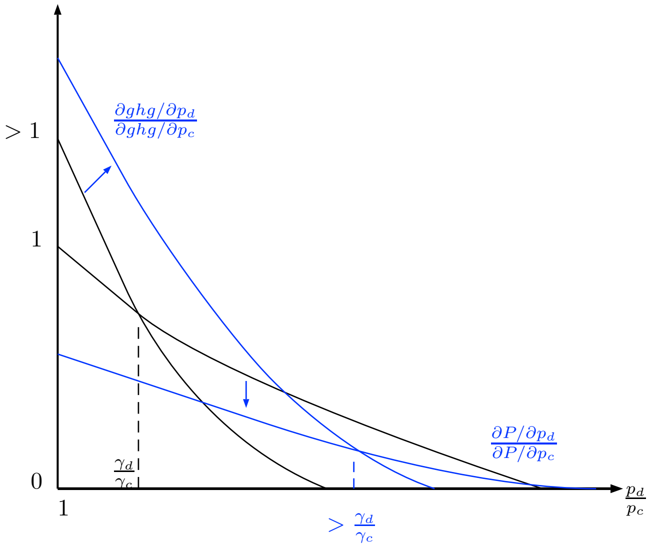

Figure 2: Optimal relative fuel prices (taxes) - switching

Moreover, it can be seen that as the fixed switching cost increases, the optimal relative prices converge to the no-switching benchmark of $\frac{\gamma_d}{\gamma_c}$:

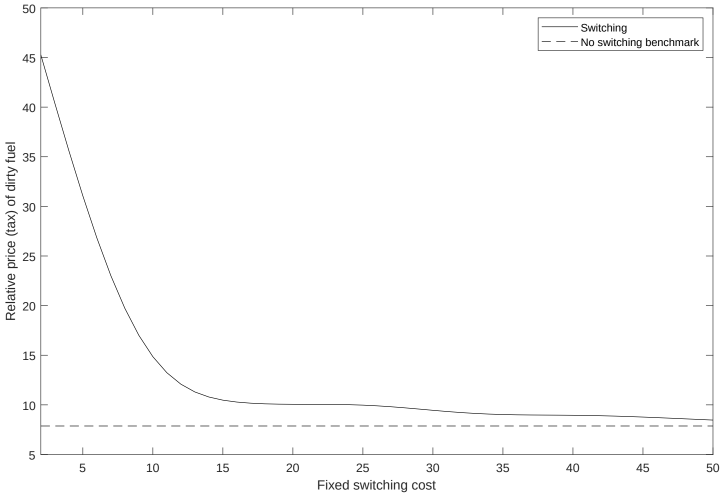

Figure 3: Optimal relative fuel prices across different values of the switching cost