Research Paper 55 | **2017**

THE HETEROGENEOUS IMPACT OF COAL PRICES ON THE LOCATION OF DIRTY AND CLEAN STEEL PLANTS

François COHEN, Giulia VALACCHI

The Heterogeneous Impact of Coal Prices on the Location of Dirty and Clean Steel Plants

Francois Cohen[a] and Giulia Valacchi[b]

March 2018

Abstract:

Climate policy will predominantly affect industries that primarily rely on fossil fuels, such as steelmaking. Within these industries, exposure may be different by country according to the energy-intensity of national plants. We estimate the effect of coal prices on steel plant location worldwide and production preferences for BOF, a polluting technology, and EAF, a greener one. A 1% increase in national coal prices reduces BOF and EAF installed capacity by around 0.51% and 0.34% respectively. We simulate the implementation of a stringent European carbon market with no border adjustment and find a non-negligible shift in steel production outside Europe, with limited impact on the technologies employed to produce steel. If applied worldwide, the same policy would primarily affect production in Asia, which relies on BOF and currently benefits from lower coal prices.

Keywords: Steel industry; firm relocation; technological change; energy prices; carbon market.

JEL Classification: O14, O33, Q41, Q42

Acknowledgements: For useful comments, we thank Richard Baldwin, Anthony Decarvalho, Antoine Dechezlepretre, Florian Egli, Cameron Hepburn, Joëlle Noailly, Jacquelyn Pless, Filipe Silva, Suchita Srinivasan, Banban Wang and Yuan Zi. We thank Beni Suryadi, Tania Theodoluz and Xiaojing Zhou for precious data contribution and Helena Ting for her work as research assistant on this project. We also thank seminar participants at the University of Oxford. Furthermore, we thank James King, who accepted that his data on steel plants is used in this paper. This research was funded by the Swiss National Science Foundation under the Sinergia programme, Project "Innovation, Diffusion and Green Growth" No CRSII1_147612.

a: Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, United Kingdom. Corresponding author. Address: INET, Eagle House, Walton Well Road, Oxford, OX2 6ED, United Kindgdom. Email: francois.cohen@smithschool.ox.ac.uk. Tel: +44 78 80 75 65 80. Fax: n.a.

b: World Intellectual Property Organization, Geneva, Switzerland.

1. Introduction

The Paris Agreement (2015) has set the ambitious objective of limiting global warming below 2°C. However, while economists have long argued that the main policy to achieve this should be either a global price of carbon or a global carbon market (e.g. Stoft et al., 2013; Weitzman, 2014), extensive international discussions have so far been unsuccessful in delivering it.[1] To date, only a small number of independent carbon schemes are operative (e.g. the EU, Japan, California). In all these schemes, the price of carbon is far below its social cost as estimated by integrated impact assessment models.[2] The main reason why carbon prices are low on these markets is political and comes from the risk that regulated industries lose competitiveness. In energy-intensive sectors, largely exposed to international competition, unilaterally implementing a carbon tax or trading scheme may push industries to relocate elsewhere.

In this paper, we estimate the effect of changes in coal prices on steel plant location worldwide. Steel represents 27% of all greenhouse gas emissions (GHG) from industry (IEA, 2017b). We explicitly focus on the impact of coal prices on the steel industry because 75% of all its CO_2 emissions come from the burning of coking coal in Basic Oxygen Furnaces (BOF) (Columbia Climate Center, 2012). However, we can expect that coal prices have heterogeneous impacts across production processes because a less coal-intensive process, Electric Arc Furnaces (EAF), can also be used to produce steel. EAF is a recycling technology that cuts GHG emissions by 75% with respect to BOF.[3]

Looking jointly at the effect of energy price shocks on plant location and production preferences constitutes the main contribution of this paper. The two are likely to interact. Even though changes of location may be encouraged by the availability of low coal prices in some countries, the choice

[1] Leading emitters (e.g., US, China, India) diverge on the concept of "differentiated responsibilities" as proposed initially in the Berlin Mandate of 1995. While developing countries argue that severe emission caps may hinder the economic growth trajectory of such countries, developed nations, mainly the United States, argue that having differentiated targets defeats the purpose of a common market, as global pollution levels would continue to grow as usual through the pathway of leakage.(Bosetti et al., 2014).

[2] Nordhaus (2017) estimates that the social cost of carbon is around \$31/tCO_2 eq. for the current period. Yet, in 2015, the traded price of carbon ranged from \$1 to \$13/tCO_2 eq. worldwide, and 90% of carbon permits were traded at less than \$10/tCO_2 eq. (WBG and ECOFYS, 2015). In the EU, the carbon price oscillated between \$3 and \$10/tCO_2 eq. between 2013 and 2015 (Marcu, 2016).

[3] This process also relies indirectly on coal since it employs electricity that is generated from coal: globally, more than 70% of the GHG emissions from power generation are caused by coal (IEA, 2017a).

to relocate will be ultimately dependent on the cost of staying, which depends on the availability of low-pollution technologies or the potential for their development. Although several studies have looked at the impact of environmental regulation (or energy taxation) on firm location, they do not address the concomitant effect of environmental stringency on technology development and diffusion. In the case of steel, this paper shows that the low-polluting technology is constrained by the availability of scrap. This reduces the potential for changes in coal prices to lead to higher technology adoption. We find relocation effects of steel firms following increases in coal prices.

This paper relies on steel plant data (1960-2014) collected by James King and merged with data on coal prices. We use a national-level econometric model to correlate the size of national industries to coal price shocks. Our setting circumvents several major identification issues. First, we account for the difference between current and expected coal prices by approximating coal price expectations with autoregressive integrated moving average models (ARIMA). We then rely on a pre-sample mean estimator (Blundell et al., 2002) to account for coal price endogeneity and the risk of small sample bias in our estimates. Complementary robustness checks include tests for instrument exogeneity and several specification changes, for example in the definition of coal prices. We also make sure that our results are not driven by the concomitant evolution of the prices of the other main steel production inputs: iron ore, electricity and scrap. Finally, we crosscheck that the national-level econometric results are not biased by data aggregation: we disaggregate the data at plant level and run linear dynamic panel data models that confirm the national-level findings.

We find that an increase in coal prices at national level has a negative effect on the size of steel manufacturing in a country. In our preferred specification, a 1% increase in coal prices reduces BOF production capacity by around 0.51% and EAF capacity by around 0.34%. As a result, a 1% rise in coal prices increases the share of EAF capacity over total national steel capacity by 0.22%.

We indirectly assess the effect of the introduction of ambitious climate policies on national steel industries by making the simplifying assumption that a carbon market is equivalent to a coal price increase. We simulate the impact of the implementation of two climate policies: a multilateral carbon market in the EU with a more stringent carbon price than today and no border adjustment; and a multilateral carbon market that would apply to all the countries that we cover (around 80%

of the steel produced worldwide)[4]. In these carbon markets, we set the price of GHG emissions at \$31/tCO_2 eq. This is the estimate of the current social cost of carbon in Nordhaus (2017). We find that redistributive effects across countries are larger than redistributive effects across technologies. The portion of world capacity that is BOF would only decrease by 0.6% with a stringent European carbon market, and by 4.7% if the carbon market was implemented everywhere. On the other side, the share of BOF capacity in European countries would drop by 17.5% if they were the only ones to implement this carbon market. On the opposite, if the carbon market was global, we find that Asian countries would reduce their overall market share by 8% because Asian firms are more coal intensive and have built their steel industry on relatively cheaper energy.

This paper complements a large body of economic literature that has looked at the effect of energy prices or environmental regulation on firm performance and location. Recent studies have shown that environmental regulation tends to decrease output and profits (Aldy and Pizer, 2015b; Ho et al., 2008; List et al., 2003; Greenstone, 2002)[5] and/or reduce exports and increase imports (Aldy and Pizer, 2015a; Levinson and Taylor, 2008; Ederington et al., 2005). Therefore, the location of plants seems impacted by environmental regulation and/or energy prices (Wagner and Timmins, 2009; Kellenberg, 2009; Kahn and Mansur, 2013). Some recent papers have studied the impact of the EU-ETS on firm relocation (Dechezlepretre et al., 2014, Borghesi et al., 2016, Koch and Basse Mama, 2016). Both Borghesi et al. (2016) and Koch and Basse Mama (2016) find evidence of relocation caused by the EU-ETS particularly for these sector exposed to international competition. On the other hand, Dechezlepretre et al. (2014) find no evidence of carbon leakages triggered by the introduction of the carbon scheme.[6]

Evidence that energy prices foster the adoption of cleaner technologies has been found in very diverse industry contexts (e.g. Cohen et al., 2017; Aghion et al., 2016; Dechezlepretre et al., 2011; Popp, 2006; Brunnermeier and Cohen, 2003; Popp, 2002; Newell et al., 1999; Jaffe and Palmer, 1997; Lanjouw and Mody, 1996). Yet, none of the above-mentioned studies considered plant

[4] This figure is obtained from the estimates given by the World Steel Association for the year 2014.
[5] Earlier studies attempting to measure the effect of environmental regulation on net exports, overall trade flows, job creation or plant-location decisions produced estimates that were either small or statistically insignificant (Jeppesen et al., 2002; Morgenstern et al., 2002; Jaffe et al., 1995).
[6] A limitation of their study is that they focus only on the initial period of activity of the EU-ETS, up to 2009, and they do not rule out the possibility of finding different results in case of an increase of the policy stringency by the EU policy makers.

location and technological choice in the same framework.[7]

The interdependency between relocation and lack of technological options, found in the case of steel manufacturing, may apply to other highly energy-intensive industries that are also strongly exposed to international competition, such as the cement industry or the chemical industry. They may largely explain the reticence of EU countries to increase the stringency of the EU ETS, or to allow for exemptions in the steel sector due to the risk of carbon leakage. National industrial interests are one of the main reasons why multinational carbon markets are not being put forward. Since countries will be asymmetrically affected by it, systems with quota allocations that take into account the current distribution of firms across countries and their energy intensity may be necessary if a multilateral agreement on a carbon market is to be found.

The rest of the paper is structured as follows. Section 2 presents the data while providing a brief overview of the steel industry. Section 3 presents our estimation method. Section 4 comments on the results and the main robustness checks performed. Section 5 presents our simulation exercise and section 6 concludes.

2. Data

2.1 James King data on the steel industry

The steel plant data has been gathered by James King and provides information on the location of steel plants over the world. For a few countries, the data is available since the beginning of the 20th century but it starts by around 1960 for most of them. The most disaggregated layer of observation in the database is the production unit: a steel plant is composed of several units, which may become operative or close down at different moments in time, even if they are on the same site. The data records the opening and closing year of these different units.

Units may also use different production technologies. The dataset records whether a unit is EAF or

[7] For the steel sector, Reppelin-Hill (1999) and Schleich (2007) have looked at the impact of energy prices on firm production choices. However, they do not properly address the endogeneity caused by the simultaneous determination of production choices and input prices. Mathiesen and Maestad (2004) aims to jointly account for the locational and technological choices of manufacturers in a static numerical partial equilibrium model of the world steel industry. These authors however rely on expert opinion to assess the magnitude of the elasticities relevant to their modelling exercise, in particular the elasticity of steel production technologies to changes in energy prices.

BOF. BOF is a steel-making technology that came into wide adoption in the 1960s.[8] It produces steel with iron ore and coking coal.[9] 88% of CO_2 emissions associated with plants using BOF are due to the combustion of coking coal to obtain coke and then the mixing of iron ore with coke to obtain steel. The remaining CO_2 emissions indirectly come from electricity usage, usually generated with on-site coal-fired power generators (EPA, 2012; OECD, 2013; IEA, 2012). On the other hand, EAF is a recycling process that uses electricity from the grid to convert ferrous scrap into new steel (Giarratani et al., 2013).[10, 11] Emissions are mostly due to electricity usage in the EAF process (EPA, 2012).

As EAF uses recycled ferrous scrap and bypasses the coke production process, it is much less energy intensive. Switching from BOF to EAF is by far the option of greater impact on the emissions performance of steelmaking. EAF units emit, for the same amount of steel produced, 4 times less GHG emissions than their BOF counterparts (OECD, 2013 IEA, 2012). Other options have a much smaller emissions abatement potential.[12] However, since EAF resorts to scrap, it cannot fully substitute BOF. It is more frequently used in countries where scrap is sufficiently available to sustain production.

[8] In the 19th century and early 20th century, the mostly-used steel-making process was Open Heart Furnace (OHF). It employed as main inputs raw materials such as iron ore, natural gas, oil or coal. It was a slow and inefficient procedure which, from the 1960s, has been completely replaced by the more efficient Blast Oxygen Furnace (BOF), which uses the same inputs, but exhibits big improvements in efficiency: approximately 1 BOF is required to replace 6 OHFs. Given this phase-out we exclude OHF from the analysis.

[9] BOF can use scrap to produce steel, but up to a maximum of 25% of the amount of total inputs.

[10] The first EAF plant was established in the US in 1907 but, initially, the quality of the steel produced was lower than the one obtained through BOF and not enough scrap was around to make it cheaper to produce only through recycling. At the beginning of the 20th century, it was difficult to control the quality of the scrap, therefore, EAF steel was considered a byproduct. With technology advancement there has been improvements in the quality of steel produced via EAF and in the second half of the 20th century it started to spread as an almost-perfect substitute to BOF.

[11] Steel scrap can have different sources: "home scrap" generated within the plant (nowadays it is not sufficient any more to produce steel due to the requirement of very high volumes of materials, therefore it needs to be integrated with the one purchased outside the firm); "new/prompt scrap" which is produced within the industrial activity of other firms (it is the same as home scrap, but it is not produced within the firm); "post-consumer scrap" which returns in the market after it ends its useful life (it could be very quick, as for cans, but it could take up to some years, as in the case of cars) (Yellishetty et al., 2011).

[12] Worrell et. al. (2001) discusses in details many alternative measures to improve the energy efficiency of steel manufacturing (e.g. preventative maintenance, cogeneration, heat recovery, etc). With 1994 data, assuming a payback period of 3 years and analyzing 47 specific energy efficiency technologies and measures, they found that energy efficiency could be improved cost-effectively in most iron and steel plants by 18%. In this paper, we have no data on the efficiency of active EAF and BOF units and we cannot evaluate the extent to which an increase in coal prices may lead to changes in the efficiency of installed EAF and BOF units. Still, our results are informative on the impact of coal prices on the environmental footprint of steel manufacturing since EAF production is far less carbon intensive.

Units may substantially differ according to capacity. For example, BOF units tend to be much larger than EAF units. The James King data includes information on unit capacity. However, this information is only available for 2014. The information is missing for the units that closed down prior to 2014. In addition, the historical capacity of operative units could be different from their 2014 capacity if modifications were performed prior to 2014. We have recovered unit capacity for all units and time periods with a single imputation method described in Appendix A.[13]

Table 1 provides the descriptive statistics for installed capacity by technology, region and period after the imputation on the missing capacity data. We only display the data used in the econometric estimation, i.e. the one that could be matched with coal price data. We cover 22 countries representing around 80% of world production. In Appendix A, we also report information on the number of units located in each region, with a breakdown by technology, before the imputation.

Table 1: Descriptive statistics of James King data on steel capacity (in million tonnes) by location and main technology (1982-2014)

Period	1982-1990		1991-2000		2001-2014	
Technology	BOF	EAF	BOF	EAF	BOF	EAF
North America	84.6	54.4	66.7	64.8	56.5	78.4
Europe	160.8	41.4	148.6	41.8	70.8	29.0
Asia	273.0	34.5	366.6	97.6	459.3	139.0
Other	11.0	0.7	N/A	N/A	1.4	0.4
Total	**529.4**	**131.0**	**581.9**	**204.2**	**588.0**	**246.9**

Notes: The table reports the average installed capacity of steel-making production in each region of the world over three different periods. Note that the data was not available for all countries and all years. Also, this information is based on imputed capacity levels. Note that we only report figures for the restricted list of 22 countries used in the regressions hereafter. North America includes Canada and the US; Europe includes Germany, Spain, Italy, France, Poland, Belgium, Czech Republic, Sweden, Finland, the Netherlands, Portugal, Switzerland and Norway; Asia includes Turkey, China, Japan, Russia and India; Other consists of Australia (1982-1989) and Chile (2006-2014). We have no coal price data for any of these two countries between 1991 and 2000.

For 2001-2014, we estimate that 70% of steel manufacturing capacity is BOF in our data. While production capacity has not evolved in Europe and the US, it has significantly increased in Asia. For 2001-2014, 70% of production capacity is in Asia according to our data. These capacity figures are indicative of production levels across countries and regions but are different from production data. Production is likely to be more responsive to immediate increases in coal prices. However, installed capacity should consistently reflect long-term trends in production levels resulting from changes

[13] Our preference for a single imputation method over multiple imputations is because we use IV regressions later on. This process is not compatible with multiple imputations.

in expected coal prices.

2.2 Coal price data

We use industrial coal price[14] data from the International Energy Agency, converted to constant 2010 USD,[15] except for China, for which we use the data from the Chinese Ministry of Coal, integrated with the Bohai-Rim Price Index. These sources report both coking coal and steam coal prices from 1978.

Steam coal, or thermal coal, is used primarily in electricity generation, while coking coal, or metallurgical coal, is used in steelmaking. In BOF processes, coking coal is converted into coke by eliminating virtually all impurities and leaving close to pure carbon. Around 600 kg of coke is used to produce a tonne of steel, which means that around 770 kg of coking coal is used to produce a tonne of steel (WCA, 2017). The EAF processes do not involve mixing coke with iron-ore. However, they are reliant on electricity generated by coal-fired power plants. With EAF, around 150 kg of steam coal is used to produce a tonne of steel (WCA, 2017).

Figure 1 presents the evolution of average coking coal prices in North America, Europe and Asia for 1978-2014. Coking coal prices declined strongly in all regions in 1988 because of the concomitant oil price collapse (King and Tang, 1988). After 2000, they then abruptly increased because of the commodities super-cycle, mainly due to the rising demand from emerging markets such as the BRIC countries (Schwartz and Creswell, 2015). Coking coal prices have been consistently lower in Asia compared to Europe or North America.

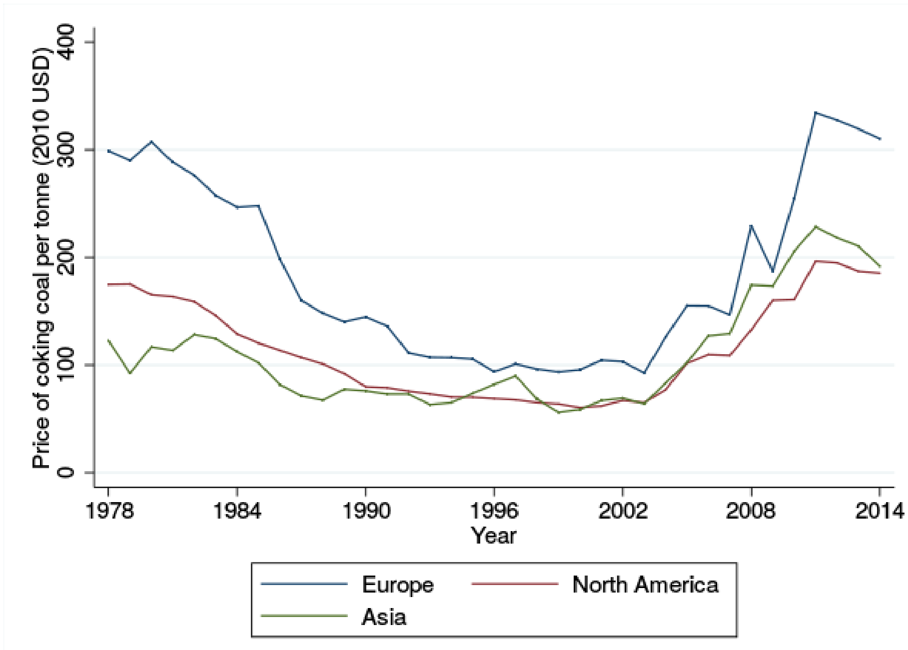
Even though EAF relies on steam coal and BOF on coking coal, we only include coking coal prices as an independent variable in our main specifications. We expect that shifts in coking coal price will capture the correlation between coking coal and steam coal prices, and therefore the impact of a change in steam coal prices on electricity prices. In Appendix D.3, we use steam coal prices and

[14] We collected the total price charged to the industrial sector.

[15] Coal prices are initially extracted in national currency/tonne. To convert them in constant 2010 USD/tonne, we apply the formula: $P^{cons2010} = \frac{P_t^{LC}}{deflator_t^{GDP}} \frac{deflator_{2010}^{GDP}}{r_{2010}^{x}}$, where $P^{cons2010}$ is the price of coal expressed in constant 2010 USD/tonne, P_t^{LC} is the price of coal expressed in local currency/tonne at time t, $deflator_t^{GDP}$ is the GDP deflator for that country at time t, $deflator_{2010}^{GDP}$ is the GDP deflator of that country in 2010, and r_{2010}^{x} is the exchange rate between the local currency and USD in 2010. Prices are inclusive of taxes.

not coking coal in alternative regressions. Results with steam coal are very similar. This is not surprising since these prices strongly correlate. Also, coal prices are instrumented in our estimations, which corrects for measurement errors. In additional specifications (in section 4.2), we control for other input prices, among which electricity. As soon as electricity prices are controlled for, the effect of coking coal prices and electricity prices are separately estimated, making redundant the use of steam coal prices in the model.

Figure 1: Evolution of coking coal prices in major geographical areas (1978-2014)



Notes: North America includes Canada and the US; Europe includes Germany, Spain, Italy, France, Poland, Belgium, Czech Republic, Sweden, Finland, the Netherlands, Portugal, Switzerland and Norway; Asia includes Turkey, China, Japan, Russia and India. We do not report prices for Chile and Australia even though they are used later on. Average prices are unweighted (i.e. they do not take into account the relative size of each country or market).

2.3 Supplementary data

Some of our robustness checks resort to price data of the other inputs used in production: iron ore for BOF, electricity and scrap for EAF.

For the electricity price we use the total price per MWh charged to the industry. The data has been obtained from multiple sources and merged in a unique database. Prices in national currency have been converted to constant 2010 USD. The sources for electricity prices are EUROSTAT statistics for the Eurozone countries[16], the Annual Report of the Chinese National Energy Administration for China, the Energy Price Report of South Africa for South Africa and the International Energy Agency for the remaining countries.

The data for iron ore and scrap comes from UN COMTRADE and has been calculated based on the import value and weight of goods in a given country.[17] The data availability differ from one country to the other but goes back to the 1960s for a few countries. Measurement errors can be quite large because the COMTRADE nomenclature aggregates goods of different nature and because the price statistics is obtained by dividing value over weight. To reduce the risk of measurement errors, the data has been cleared. Only the observations with annual shipments of over 10,000,000 tonnes were considered to ensure higher homogeneity of goods. Furthermore, we excluded prices below the 5th and 95th percentiles. Appendix B provides summary statistics for the evolution of the price of electricity, iron ore and scrap over time. These price series have been deflated using the US consumer price index.

3. Empirical methods

We develop an empirical strategy that allows estimating the impact of expected coal prices on the size of steel manufacturing in a country. Our main dependent variable is the share N of installed capacity using technology $s \in \{BOF; EAF\}$ in country i during year t over the total worldwide installed capacity: e.g. 10% of global capacity is located in country i at time t and use technology s. The expected price of coal, denoted $p^*_{i,t}$, constitutes the independent variable of interest of our model. Therefore, we aim to estimate a reduced-form model such that:

$$N_{s,i,t} = f(p^*_{i,t})$$

To estimate such a model, we are confronted with four difficulties. The first difficulty is that we do not observe the expected price of coal in country i at time t, but its actual realization, which we

[16] EUROSTAT reports electricity prices charged to different categories of firms according to their annual electricity consumption. There are 9 categories before 2007, and 7 categories after 2007. Prices are averaged across all firms' categories.
[17] For iron ore, we use code 281 ("iron ore and concentrates") of the standard international trade classification (SITC), and cross check the data with code 2601 of the harmonised system ("iron ores and concentrates, including roasted iron pyrites"). For scrap, we used code 282 ("iron and steel scrap") of the SITC.

denote $p_{i,t}$. However, the right variable of interest is the expected price of coal because manufacturing companies are likely not to base their investment decision on current inputs' prices. They will rather take in consideration their expected value over the lifetime of their investment. This is particularly true for a sector like the steel one because investments are costly and have a long lifetime.

The second difficulty is that expected coal prices constitute an endogenous variable. The total demand for coal in country i at time $\underline{t}$ depends on the demand for coal from the steel industry. Therefore, the installed capacity using technology s in country i and at time t is a determinant of the price of coal in country i at time t. Since $N_{s,i,t}$ has an impact on $p_{i,t}$ while it also depends on $p_{i,t}$, these two factors are simultaneously determined. On the other hand, expectations about future coal prices necessarily depend on the current price of coal. It follows that $N_{s,i,t}$ and $p^*_{i,t}$ are also simultaneously determined.

Two additional difficulties arise because we rely on a country-level panel. Country panels typically correspond to small samples. Depending on the estimation method used, small sample biases might arise. Finally, country level observations might insufficiently reflect heterogeneity between plants. Aggregation bias could affect our results if unaccounted plant characteristics systematically correlate with changes in coal prices.

We handle all these difficulties as follows.

3.1. Computing expected coal prices

To proxy expected coal prices, we consider that a perfectly rational agent forecasts future prices based on the information that s/he has. We assume that past prices constitute all the information easily available to economic agents, such that their expectations about future prices are a function of past prices. In this framework, expectations can be recovered with a time-series model that produces a forecast of the coal price at time t+1, t+2, t+3, etc. with the coal price at time t, t-1, t-2, etc. We recursively apply autoregressive integrated moving-average models (ARIMA) on national annual time-series of real coal prices. The ARIMA model that we use to make the predictions is a first-order autoregressive model. This is because this model proved to be a good fit to the data. This is a mean-reversing model: agents assume that price spikes or lows are unusual and expect that prices will align with longer-term trends. We use the following specification:

$$p_{i,t} = \mu_i + \gamma_i p_{i,t-1} + \varepsilon_{i,t}$$

In this model, the price of coal in each country *i* is regressed on its first lag. $\varepsilon_{i,t}$ is a term of error and μ_i and γ_i are parameters to be estimated. For each period *t* and country *i*, the ARIMA model is estimated with data from the previous 10 years. When we lack information on the previous 10 years, we estimate the model with all available years, provided that we have data for the previous 4 years at least. In each case, we then take the 10-year average of forecasted prices (for time *t+1*, *t+2*, *t+3*, etc.) as our value for the expected coal price, $p^*_{i,t}$. The ARIMA model is re-estimated for each time period *t* present in the sample, except for the first four years of our data (1978-1981), for which we did not have enough data to run the ARIMA model. We make out-of-sample predictions and, year after year, we allow steel companies to update their beliefs about future prices as soon as new information is available. Detailed statistics on the output of the ARIMA model and the predictions obtained are reported in Appendix C.

We performed several robustness checks relative to the choice of the coal price variable. They are reported in Appendix D and also briefly discussed in the results section.

3.2. Endogeneity of expected coal prices

Our econometric strategy needs to deal with the endogeneity of the expected price of coal. In the literature, the most frequently suggested technique consists of using cost shifters as instruments, i.e. factors that are correlated with the cost of producing coal, and not with the demand for coal (e.g. Berry, 1994). However, the context of coal and steel production makes it hard to find strictly exogenous instruments: the two sectors are closely related. For example, steel production uses another output of the mining industry as an input, namely iron ore. Therefore, supply shocks on the extraction of coal may also affect the extraction of iron. If such cost shifters cannot be used as instruments, an alternative would consist in using shocks on the demand for coal that are not correlated with the demand from the steel industry as instruments. This approach is the one of Hausman et al. (1994), who instruments the price of a product in a given market with the price of this same product on other markets. Provided that demand shocks are not correlated across markets, this instrumentation strategy is valid. Yet, the steel industry is the main demander for coal and markets are integrated across regions. When not used to make steel, coal is used to produce electricity. The assumption that demand shocks on the electricity market are not correlated with demand shocks on the steel market is not serious: EAF units directly use electricity from the grid to produce steel.

We adopt a conservative approach and consider that most of the (demand and supply) shocks on coal prices are likely to be correlated with contemporaneous shocks on the steel industry. When contemporaneous correlations are strong, the usual approach consists in exploiting past information that is not correlated with contemporaneous shocks to produce valid instruments. Namely, it is possible to use the lags of the endogenous variables as instruments, i.e. instrumenting $p^*_{i,t}$ with $p^*_{i,t-1}$.

However, using pre-determined variables as instruments (e.g $p^*_{i,t-1}$ for $p^*_{i,t}$) is not possible with fixed effect models which rely on the assumption of strict exogeneity of the instruments. To relax this assumption, applied economists have usually relied on models in first differences (Roodman, 2008). However, these models lack efficiency, in particular when time-persistent processes are studied. Models in first differences are also subject to small sample bias. With a total sample of around 700 observations and an industry that relies on long-term investments, first difference models are likely to provide inefficient estimates, if not inconsistent, for the national-level analysis.

3.3. Estimation method circumventing small sample biases

To avoid small sample biases while instrumenting endogenous variables with pre-determined regressors, Blundell, Griffith and Windmeijer (2002) recommend using pre-sample mean estimators. A specific interval of time $\{t_1; t_f\}$ is set to be the time of the analysis. In this interval, the set of information about dependent and independent variables is complete. There is, additionally, a pre-sample interval $\{t_i; t_0\}$ where only information on the dependent variable is available. A mean of the dependent variable is estimated over $\{t_i; t_0\}$ and included as a control variable in the estimation. With small samples, Blundell, Griffith and Windmeijer (2002) show that pre-sample mean estimators are more efficient and significantly less biased than first difference estimators, even when time persistence is not extreme and for a small number of pre-sample observations. The intuition why pre-sample mean estimators are superior to first difference estimators is quite simple: they incorporate additional information into the model, namely the pre-sample mean of the dependent variable. The main limitation explaining why these estimates are barely used is that they can only be applied if pre-sample information on the dependent variable is available to the econometrician.[18]

We estimate the following log-log specification:

$$\ln(N_{s,i,t}) = \alpha + \beta \ln(p^*_{i,t}) + \sigma \ln(\bar{N}_{s,i,t_i}) + \theta_t + e_{s,i,t} \quad (1)$$

Where $\bar{N}_{s,i,t_i}$ is the pre-sample mean of $N_{s,i,t}$, θ_t is a year fixed effect and $e_{s,i,t}$ is the error term. α, β and σ are parameters to be estimated.

In the present case, we have data on $N_{s,i,t}$ since the 1960s, while our data on $p^*_{i,t}$ starts in 1982. We have 22 years of pre-sample information on $N_{s,i,t}$ which can be used to run pre-sample mean estimators in the fashion of Blundell, Griffith and Windmeijer (2002). However, we use the period from 1970 to 1981 to calculate the pre-sample mean, and not the full period (1960-1981). The reason is simple. Our data is on unit openings and closings since 1960. Therefore, the 1960 data does not correspond to the stock of units that were active in 1960, but only to the flow that opened this year. After 10 years of data, i.e. in 1970, we can compute a value of $N_{s,i,t}$ that includes all the units that became active over the past 10 years and are still active in 1970. The pre-sample mean between 1970 and 1981 is therefore more representative of the stock of units operating in the pre-sample period. [19]

Equation (1) can be estimated using two-stage least squares (2SLS). To instrument for $\ln(p^*_{i,t})$, the lags of $\ln(p^*_{i,t})$ are valid instruments provided that contemporaneous shocks on $p^*_{i,t}$ are not correlated with previous shocks. The latter means that shifts in expectation between time *t* and *t-1* arise from the inclusion of new information about $p^*_{i,t}$. This is the case if expectations are rationally formed, or if they follow a random walk. In our base specification, we simply use $\ln(p^*_{i,t-1})$ to instrument for $\ln(p^*_{i,t})$: our instrumentation strategy assumes that past shocks on expected coal

[18] The also require that the dependent variable is mean stationary. In the present context, we have normalised the dependent variable as the share of units in country *i* with respect to all the units present in the world. This share is naturally bounded between 0 and 1.

[19] In Appendix E.1, we modify the pre-sample period used for estimation to 1960-1981instead of 1970-1981. Results are similar to our preferred specifications.

prices do not correlate with the error term ($e_{s,i,t}$), and they only determine the value of expected coal prices at time t. In mathematical terms, our instrument is valid if $\ln(p^*_{i,t}) = \delta \ln(p^*_{i,t-1}) + v_{i,t}$, with $corr(v_{i,t}, v_{i,t-1}) = 0$. The pre-sample mean estimator tolerates that $v_{i,t}$ is correlated with $e_{i,t-x}$ with $x \geq 1$, whereas a fixed effect model would require $corr(v_{i,t}, e_{i,t-z}) = 0$ with $z \geq 0$.

In an alternative specification, we expand the number of lags used as instruments, and run the model with the first five lags of $\ln(p^*_{i,t})$ as instruments. This allows us to run an over-identification test and make sure that all lags provide the same results when used as instruments. The general idea of this process is that deeper lags are even less likely to be correlated with $e_{s,i,t}$ than the first lag. If the first lag is not an exogenous instrument, then it should be providing different estimates than the ones obtained with deeper lags. Using up to 5 lags, we can apply this same logic to the 2nd lag, the 3rd lag and so on, assuming that, at least, shocks on the 5th lag of expected coal prices should not be correlated with $e_{s,i,t}$. The sufficient condition for the 5th lag to be a valid instrument is that $corr(v_{i,t}, v_{i,t-5}) = 0$. The result of this test, presented in section 4, corroborates our assumption that there is no correlation between $p^*_{i,t-1}$ and $p^*_{i,t}$ creating endogeneity.

In addition, we report the results obtained with a fixed effect model with no treatment of endogeneity in Appendix E.2 and the results obtained with a first difference estimator in Appendix E.3. Unsurprisingly, the results of the fixed effect model are inconsistent. The results with the first difference estimator are very inefficient due to weak instrumentation. They should also suffer from small sample bias (Staiger and Stock, 1997; and Blundell and Bond, 1998). The bias of the first difference estimator leads the estimates to be similar to the ones provided by a first differenced model where the expected price of coal would be treated as exogenous. We also provide the results of such a model in Appendix E.3.

3.4. Considerations regarding aggregation bias

When relying on national data, we implicitly assume that the national capacity aggregates by technology are homogenous. If there are strong differences between BOF or between EAF units, then our results could be biased. For example, if BOF plants that are in areas with high coal prices also produce high value-added steel, our econometric results could be biased since we do not control for plant characteristics within a country (beyond their use of BOF or EAF technologies).

In Appendix F, we complement the national-level analysis with a plant-level analysis. The plant-level analysis includes plant-specific fixed effects that mechanically tackle the above-mentioned issue. We show that plant-level results are aligned with national-level results.

In this context, we could indifferently use the analysis at national or plant level. For our base model, we prefer to rely on the national-level analysis for two reasons. The first reason is that unit openings and closures are frequent when observed at national level, but rather rare at plant level. Consequently, the plant-level analysis requires a dynamic specification and this increases the number of assumptions made for the econometric model to be consistent. The second reason is that the time persistency of the dynamic specification is high because unit lifetimes are long. This leads to an imprecise estimation of long-term effects in the dynamic, plant-level setting.[20]

4. Results

Table 2 presents the results of the estimation of equation 1. Column 1 reports results on the impact of coal prices on installed capacity for the two technologies jointly, Column 2 for BOF capacity and column 3 for EAF capacity. Finally, column 4 uses the relative share of EAF over the sum of BOF and EAF capacity as the main dependent variable. An increase in this share therefore pinpoints that national production resorts more to the EAF technology.

Overall, we find that a 1% increase in the expected price of coal in country *i* would cause a decrease in the capacity of steel plants by about 0.43%. The impact is however 40% larger for BOF capacity (-0.51%) as compared to EAF capacity (-0.34%). Column 4 logically suggests that national

[20] Take the following model:

$$Y_{i,t} = a.Y_{i,t-1} + bX_{i,t}$$

Where $Y_{i,t}$ is the dependent variable and $X_{i,t}$ is the independent variable. The long-run effect of $X_{i,t}$ on $Y_{i,t}$ is $b/(1-a)$. If a is close to 1, then the long-run effect is very imprecisely estimated since:

$$\frac{d[b/(1-a)]}{da} = \frac{b}{(1-a)^2}$$

If $a = 0.98$, the derivative in a of $b/(1-a)$ is $2500b$. We face this problem in Appendix F.

production shifts towards EAF when expected coking coal prices increase: this effect is statistically significant at 5%.

Table 2: Pre-sample mean estimation of the impact of expected coal prices on the share of installed capacity located in a given country

Dependent Variable:	Log-log model			
Log. share of installed	BOF & EAF	BOF	EAF	EAF/(EAF+BOF)
capacity worldwide	(1)	(2)	(3)	(4)
Log. expected coking	-0.427***	-0.508***	-0.343**	0.216**
coal price	(0.110)	(0.077)	(0.167)	(0.099)
Log. pre-sample mean	0.683***	0.651***	0.800***	0.431***
(1970-1981)	(0.046)	(0.096)	(0.056)	(0.112)
Year fixed effects	Yes	Yes	Yes	Yes
Observations	707	348	359	359
Countries	22	21	21	21

Notes: The model is estimated with 2SLS. The logarithm of the expected price of coking coal is instrumented with its one-year lag. Cluster-robust standard errors in parentheses. Clusters are set at country level and *, ** and *** respectively denote significance at 10%, 5% and 1% levels.

The results displayed in Table 2 rely on the validity of our instrumentation strategy. We jointly use the first five lags of $p_{i,t}^{*}$ as instruments in Table 3 and run over-identification tests. In Table 3, results are very close to the ones obtained with only one lag and the four over-identification tests of columns 1–4 corroborate the validity of our instrumentation strategy.[21]

The results of Tables 2–3 are robust to many specification changes. In Table 4, we reproduce the econometric estimation with the number of BOF and EAF units instead of installed capacity. This allows checking that these results are not artificially caused by the imputation of capacity levels on missing data earlier on. Results are very similar to our base model: a 1% increase in the expected price of coal in country i would cause a decrease in the number of units by about 0.49% for BOF and 0.27% for EAF.[22]

We also check that these results do not rely on a specific functional form between coal prices and

[21] For concision, weak identification tests are not reported: instrumenting with lags necessarily ensures high strength of the instruments.

[22] As described before, installed capacity at unit level has been imputed based on the information available for 2014. Therefore, the imputation method creates measurement error in the dependent variable and may reduce efficiency. It has however no impact on model consistency under two assumptions: 1) unit capacities did not significantly differ between the units that closed down and the units that are still operative, conditional on technology, country and installation year; and 2) unit capacity did not go through substantial adjustments prior to 2014. Even though these assumptions are likely to hold, we provide complementary regression results in Table 4 where we do not rely on any capacity estimate. Results are similar to Table 2 even though less efficient, which suggests that the imputation method has not created any bias.

our dependent variable (see Appendix D.1). Results are not altered when we directly use expected coal prices instead of their logarithm in the model. When we add a quadratic term, results suggest that the marginal impact of an increase in coal prices is higher when coal prices are low, which is consistent with a logarithmic form.

Table 3: Specifications using the five first lags of the expected coal price as instruments

Dependent Variable: Log. share of installed capacity worldwide	Log-log model BOF & EAF (1)	BOF (2)	EAF (3)	EAF/(EAF+BOF) (4)
Log. expected coking coal price	-0.558***	-0.624***	-0.450*	0.209
	(0.170)	(0.117)	(0.255)	(0.134)
Log. pre-sample mean (1970-1981)	0.641***	0.653***	0.751***	0.415***
	(0.0607)	(0.0934)	(0.0769)	(0.0992)
Year fixed effects	Yes	Yes	Yes	Yes
Over-identification test (p-value of Hansen J statistic)	0.21	0.16	0.24	0.62
Observations	540	265	275	275
Countries	19	19	18	18

Notes: The model is estimated with 2SLS. The five first lag of the logarithm of expected coal price are used as instruments. Cluster-robust standard errors in parentheses. Clusters are set at country level and *, ** and *** respectively denote significance at 10%, 5% and 1% levels.

Table 4: Specifications using the number of units instead of installed capacity to construct the dependent variable

Dependent Variable: Log. share of units	Log-log model BOF (1)	EAF (2)
Log. expected coking coal price	-0.491***	-0.268
	(0.0972)	(0.164)
Log. pre-sample mean (1970-1981)	0.619***	0.905***
	(0.124)	(0.0619)
Year fixed effects	Yes	Yes
Observations	348	359
Countries	21	21

Notes: The model is estimated with 2SLS. The five first lag of the logarithm of expected coal price are used as instruments. Cluster-robust standard errors in parentheses. Clusters are set at country level and *, ** and *** respectively denote significance at 10%, 5% and 1% levels. We do not provide pooled results with BOF and EAF units altogether since the average capacity of BOF and EAF units is different.

Also, our general results do not rely on the way expectations were computed. We obtain similar results when we use contemporaneous prices ($p_{i,t}$) instead of expected prices ($p_{i,t}^*$) as the main independent variable (see Appendix D.2), implying that our results hold if we assume that manufacturers are myopic or think that coal prices follow a random walk.

4.2 Interpretation

So far, we have exclusively looked at the correlation between expected coal prices and installed production capacity. We found a stronger impact of expected coking coal prices on BOF installations as compared to EAF installations. The interpretation of this result is not as straightforward as it seems. The national-level impacts may encompass a large list of effects, both direct and indirect, of coal prices on installed capacity.

In the case of BOF units, the price of coking coal should influence installed capacity because BOF units directly use coking coal as an input. The national level model may also capture three additional effects. First, an increase in the price of coal may modify the relative competitiveness of BOF plants as compared to EAF plants. Therefore, when coal prices increase, EAF plants may cannibalize part of the market traditionally supplied by BOF plants. Another effect comes from a change in the price of electricity due to a change in the price of coal. This effect is likely to be small but may increase production costs for BOF plants. Finally, an increase in the price of coal could be correlated with the price of iron ore, since iron extraction is also reliant on energy.

In the case of EAF, there should be no direct impact of coking coal prices on production costs since EAF units do not use coking coal. Yet, we find an impact of coal prices on installed capacity. The econometric specification may exclusively capture indirect effects. First, EAF uses electricity and we could be capturing the effect that an increase in (steam) coal prices has on electricity prices. Furthermore, an important phenomenon is the interdependency between EAF and BOF production. EAF uses scrap as a main input to produce steel. In practice, the value of scrap should be correlated with the price of steel and therefore the price of BOF inputs. Besides, home market effects might also be important determinants. Both EAF and BOF technologies rely on a large downstream demand for raw steel. The downstream demand is likely to locate where steel manufacturing is to reduce transportation costs. Likewise, steel manufacturers have an interest to be geographically close to their clients. A downsizing of the BOF industry, which represents the largest share of the market, could significantly affect the smaller share of the market serviced by the EAF industry.

In Table 5, we aim to separately control for a series of effects that might explain the correlation between coal prices and BOF and EAF capacity levels. First, the pre-sample mean estimator used previously can easily accommodate for the inclusion of the lagged dependent variable as an additional control variable. This addition takes out the correlation between previous market size and previous price values from the impact of expected coal prices on current market size. As such, it provides a narrower assessment of the impact of coal prices on openings and closures. In addition, we add three additional control variables for the other inputs likely to determine the level of installed BOF and EAF capacity: the expected price of iron ore (used in BOF), electricity (mostly used in EAF) and scrap (mostly used in EAF).

Table 5: Pre-sample mean estimation with lag dependent variable and additional inputs

Dependent Variable:	Log-log Model			
Log. share of installed	BOF & EAF	BOF	EAF	EAF/(EAF+BOF)
capacity worldwide	(1)	(2)	(3)	(4)
Log. expected coking	-0.448**	-0.641***	-0.230	0.263**
coal price	(0.192)	(0.213)	(0.206)	(0.134)
Log. expected iron ore	0.249	0.0716	0.727*	0.603***
price	(0.293)	(0.257)	(0.412)	(0.220)
Log. expected	0.175	0.152	0.282	0.0241
electricity price	(0.223)	(0.188)	(0.267)	(0.122)
Log. expected scrap	-0.0183	0.388*	-0.476	-0.593**
price	(0.245)	(0.229)	(0.317)	(0.254)
Lagged dependent	0.359***	0.253**	0.485***	0.275***
variable	(0.0974)	(0.0988)	(0.111)	(0.0602)
Log. pre-sample mean	0.503***	0.598***	0.393***	0.303***
(1970-1981)	(0.0525)	(0.0875)	(0.0662)	(0.0629)
Year fixed effects	Yes	Yes	Yes	Yes
Observations	497	247	250	250
Countries	18	16	17	17

Notes: The model is estimated with 2SLS. The first lag of the logarithm of the four expected input prices are used as instruments. Cluster-robust standard errors in parentheses. Clusters are set at country level and *, ** and *** respectively denote significance at 10%, 5% and 1% levels.

Table 5, column 1, confirms that expected coking coal prices is the main input determining the total capacity present in a given country. Column 2 shows that this effect is mostly driven by BOF production. On the other hand, the impact of expected coking coal prices on EAF production is no longer statistically significant when other inputs and the lagged dependent variable are included (see Table 5, column 3). Finally, column 4 consistently shows that national production relies more on EAF when the input prices of BOF processes go up, and if scrap prices are low.

When a dynamic setting is used on plant-level data (see Appendix F), we likewise find that the expected coking coal price is a strong determinant of the opening and closing of steel plants, in particular BOF plants.

5. Simulation of multilateral carbon markets

We now run a simulation where we quantify the effect of the introduction of multinational carbon markets on the location of steel plants and the share of EAF units in a given country. We assume that if country *i* adopts the carbon market, it raises the price of carbon by an additional $ 31 per tonne of CO_2 equivalent. This figure corresponds to the current social cost of carbon as estimated in Nordhaus (2017). For simplicity, we assume that this increase in the price of carbon would be additional to any existing policies, and that it would exclusively translate into an increase in the price of coal. We convert the carbon tax of $ 31/tCO_2 eq. into a coal price increase by assuming that a tonne of coal emits 2.457/tCO_2 eq.[23]

We therefore raise the expected price of coal in regulated countries by about $ 76 per tonne. This is a significant increase of more than 80%, given that the average price of coking coal is around $ 90 in our data. This implies that our simulation results will only be indicative of the direction that the market could take if coking coal prices increased significantly: the marginal effects estimated with our econometric models have limited value to be extrapolated to large changes in expected prices.

We produce three scenarios with different combinations of countries implementing the scheme. In the business as usual (BAU) scenario, the policy is implemented nowhere. In the second scenario, the carbon market is adopted by the Member States of the European Union and Associated Countries (Norway and Switzerland in the simulation), with no border adjustment of steel prices with third countries. This scenario corresponds to the one where the steel industry would be strictly regulated under the EU Emissions Trading Scheme (ETS), with a much higher carbon price (of $ 31/tCO_2 eq.) than the one we can be observe today. We assume that Switzerland and Norway would be linked to this trading scheme, but not Russia or Turkey. The last scenario assumes that all countries would implement the carbon market.

To assess policy impacts, the simulation uses the specifications displayed in columns 2 and 3 of Table 2: we therefore consider that the impact of increasing coking coal prices would be different for BOF and EAF units. We choose Table 2 since we know that the results of Table 3 are globally

[23] https://www.carbontrust.com/home/. Consulted June 2017.

robust throughout Tables 3–5 and the appendices, in particular when the data is disaggregated at plant-level. We also know that the output of Table 2 reflects the overall effect of an increase of coal prices on BOF and EAF processes, inclusive of many direct and indirect effects.

Based on Table 2, we compute the share of BOF and EAF units that would have operated in country *i* over the sample period (1982-2014) under the two policy scenarios, and compare this share to the one recorded in our data (which corresponds to the business-as-usual scenario). We are not making any general equilibrium adjustment.

Results are reported in Table 6. We are, above all, interested in shifts in the market share of BOF vs. EAF, and their location in either North America, Europe or Asia. BOF would represent 70% of installed capacity in the BAU. This would slightly decrease by 0.4 percentage points with a higher carbon market in Europe, and by 3.3 percentage points (i.e. by 4.7% in relative terms) if the carbon market is implemented everywhere. These values are quite small if we consider that this would be for an 80% coal price increase and an estimated reduction of total installed capacity worldwide by around 30% if all countries implemented the carbon market.

Table 6: Simulation results

Indicator	World capacity that is BOF (%)			World capacity that is EAF (%)		
Scenario	BAU	Europe	All	BAU	Europe	All
World	73.3	72.9	70.0	26.7	27.1	30.0
North America	10.9	11.4	11.2	10.6	11.0	12.1
Europe	16.5	13.6	18.3	4.8	4.2	5.7
Asia	45.3	47.3	39.9	11.3	11.8	12.2
World capacity (BAU = 100)	73.3	69.8	50.0	26.7	26.0	21.5

Notes: North America includes Canada and the US; Europe includes Germany, Spain, Italy, France, Poland, Belgium, Czech Republic, Sweden, Finland, the Netherlands, Portugal, Switzerland and Norway; Asia includes China, Japan, Russia, Turkey and India. Other countries included in the world average are Australia (1982-1989) and Chile (2006-2014).

Redistributive effects across countries and technologies are larger. The share of BOF capacity in European countries would drop by 17.5% (from 16.5 to 13.6 percentage point) if they were the only ones to implement this carbon market. This could also undermine EAF production in Europe (decreasing from 4.8 to 4.2 percentage points, i.e. by 12.5%). Therefore, the reduction in market share appears to be disproportionately high compared to the gains obtained in greening the domestic production of steel in Europe.

If the policy was implemented everywhere, we find that Asian countries would reduce their market share by 8% (minus 4.5 percentage points). The decrease would be due to a 12% reduction in BOF capacity (from 45.3 to 39.9 percentage points), barely compensated by the increase in market share from EAF processes, which would be relatively large in relative terms (8%), but small in absolute terms (only 0.9 percentage points). The main reason why Asian countries would lose market share is because Asian firms are more coal intensive and they have built their steel industry on relatively cheaper energy.

In the above simulation, we considered that marginal effects were homogeneous across regions. This assumption is relaxed in Appendix G, where we estimate region-specific response functions. With region-specific responses, we globally find the same trends as above. However, the effect of coal prices on steelmaking is more equally shared between relocation and technological change. The negative effect of a global carbon market on Asian firms persists but is attenuated. We prefer the model with homogenous responses because region-specific effects are estimated from a small sample of countries for North America (2) and Asia (5), reducing efficiency.

Overall, this simulation is informative about the upcoming difficulties in reaching agreements on carbon markets across the globe. The steel industry would be sensitive to the implementation of a carbon market as a whole, since we predict a significant decrease in installed capacity if the carbon price equaled the social cost of carbon of Nordhaus (2017). However, a higher carbon price would not affect countries equally. Our figures suggest that some European firms would relocate if the European steel industry was to comply with a more stringent carbon price with no border adjustments, e.g. in the framework of the EU ETS. Asian countries have benefited, so far, from a competitive advantage in cheap energy, which allows them to rely on polluting technologies to produce steel. They would lose part of this advantage if emissions were uniformly priced worldwide as they specialized more in BOF technologies. Finally, cross-country relocation effects appear to be at least as strong as within-country changes in steel-making technologies, plausibly because technology shifts are constrained by the availability of scrap in the first place.

6. Conclusion

The Paris Agreement, in 2015, has shown the ambition of countries to undertake a global action against climate change. The steel sector is particularly sensitive to this topic as it is a major emitter of GHG emissions in the atmosphere. Given the availability of two substitutable steel-making technologies, the interest lies in the possibility, through environmental policies, to shift steel production from BOF to EAF, making it less carbon intensive.

This study aims at shedding light on the impact of an increase in coal prices on the location of steel plants and the technologies that the firms would choose to produce steel. We find that, on average, an increase of 1% in the expected price of coal reduces the share of steel-making capacity active in one country, compared to the rest of the world, by 0.43%. This effect seems to differ across types of production technologies: BOF units are more sensitive (minus 0.51%) than EAF ones (minus 0.34%). When we simulate the effect of regional increases in coal prices, we find that European firms would lose competitiveness if they were to unilaterally set a binding carbon price on their firms. Therefore, the risk of relocation is real for EU firms if steel making processes were more heavily taxed. Another important finding is that a uniform increase of coal prices across the globe would also have an impact on the location of steel plants. Asian firms would be more severely touched because they are the ones that rely on BOF technologies the most.

The simulation on the European market shows that unilateral agreements may prove detrimental to national industries in industrialized countries, making the search for a multilateral agreement all the more necessary. Yet, a multilateral agreement could also have strong redistributive consequences for the steel sector. Our results suggest a multilateral agreement on taxing energy use in the steel sector will be much easier to reach if measures were taken to guarantee that reducing GHG emissions in the industry would not strongly modify the location of production worldwide. Whereas these results mostly apply to the steel sector, they are likely to be valid for other energy-intensive sectors open to international competition, such as the cement industry. These redistributive effects would come on top of the reduction in demand that would stem from the increase in prices for goods that require a lot of energy to be produced.

Having said so, the reader should be careful that the general figures provided in the simulation are only indicative, since we are making extrapolations of effects for a non-marginal increase in coal prices, and do not account for macroeconomic adjustments and interactions at world level resulting from such an increase. As such, these results are however revelator of the sensitiveness of the steel industry to the implementation of ambitious climate policies.

References

Aghion, P., Dechezlepretre, A., Hemous, D., Martin, R. and VanReenen, J. (2016). Carbon taxes, path dependency, and directed technical change: evidence from the auto industry. *Journal of Political Economy.* 124(1), 1–51.

Aldy, Joseph E., and William A. Pizer (2015a). The competitiveness impacts of climate change mitigation policies. *Journal of the Association of Environmental and Resource Economists*, 2(4), 565-595.

Aldy, Joseph E., and William A. Pizer (2015b). The impacts of power sector carbon regulations on U.S. manufacturing employment. Harvard Kennedy School Working paper

Berry, S. T. (1994). Estimating Discrete-Choice Models of Product Differentiation. *RAND Journal of Economics.* 25(2), 242–262.

Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115-144.

Blundell, R., Griffith, R. and Windmeijer, F. (2002). Individual effects and dynamics in count data models. *Journal of Econometrics.* 108(1), 113–131.

Bosetti, V., Carraro, C., Duval, R., De Cian, E., Massetti, E. and Tavoni, M. (2014). The incentives to participate in and the stability of international climate coalitions: a game theoretic approach using the witch model. OECD Economics Department Working Papers (1177) .

Borghesi, S., Franco, C., and Marin, G., (2016). Outward foreign direct investments: pattern of Italian firms in the EU-ETS. SEEDS Working Paper.

Brunnermeier, S. and Cohen, M. (2003). Determinants of environmental innovation in us manufacturing industries. *Journal of Environmental Economics and Management.* 45(2), 278–293.

Chamberlain, G., 1992. Sequential moment restrictions in panel data. *Journal of Business & Economic Statistics* 10, 20-26.

Cohen, F., Glachant, M. and Soderberg, M. (2017). The impact of energy prices on product innovation: evidence from the UK refrigerator market. *Energy Economics*.

Columbia Climate Center, 2012. *Mitigating Iron and Steel Emissions*. The GHCN Factsheets.

Dechezlepretre, A., Glachant, M., Hascic, I., Johnstone, N. and Meniere, Y. (2011). Invention and transfer of climate change-mitigation technologies: a global analysis. *Review of Environmental Economics and Policy.* 5(1). 109–130.

Dechezlepretre, A., Gennaioli, C., Martin, R., and Muuls, M. (2014). Searching for carbon leaks in multinational companies. Grantham Research Institute on Climate Change and the Environment Working Paper (165).

Ederington, J., A. Levinson, and J. Minier. 2005. Footloose and pollution-free. Review of Economics and Statistics 87 (1): 92 – 99.

EPA (2012). Available and emerging technologies for reducing greenhouse gas emissions from the iron and steel industry. Environmental Protection Agency.

Giarratani, F., Hewings, G. J. and McCann, P. (2013), Handbook of Industry Studies and Economic Geography. Edward Elgar Publishing.

Greenstone, Michael. 2002. The impacts of environmental regulations on industrial activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufactures. *Journal of Political Economy*. 110 (6): 1175-1219.

Hausman, J., Leonard, G. and Zona, J. D. (1994). Competitive Analysis with Differentiated Products. *Annals of Economics and Statistics.* (34), 143–157.

Ho, M. S., Morgenstern, R. D., & Shih, J. S. (2008). Impact of carbon price policies on US industry. *Resources for the Future Discussion Paper No. 08-37.*

IEA (2012). Profiles- co2 abatement in the iron and steel industry. International Energy Agency.

IEA (2016). Energy technology perspective 2016: Towards sustainable urban energy systems. International Energy Agency.

IEA (2017a). *CO2 emissions from fossil fuel combustion: highlights*.

IEA (2017b). *Tracking clean energy progress 2017*.

Jaffe, Adam B., Steven R. Peterson, Paul R. Portney, and Robert N. Stavins. 1995. Environmental regulation and the competitiveness of U.S. manufacturing: What does the evidence tell us? *Journal of Economic Literature.* 33 (1): 32–63.

Jaffe, A. and Palmer, K. (1997). Environmental regulation and innovation: a panel data study. *The Review of Economics and Statistics*. 79(4), 610–619.

Jeppesen, Tim, John A. List, and Henk Folmer. 2002. Environmental regulations and new plant location decisions: evidence from a meta-analysis. *Journal of Regional Science.* 42 (1): 19–49.

Kahn, Matthew E., and Erin T. Mansur. 2013. Do local energy prices and regulation affect the geographic concentration of employment? *Journal of Public Economics.* 101:105–14.

Kellenberg, D. K. (2009). An empirical investigation of the pollution haven effect with strategic environment and trade policy. *Journal of international economics*, *78*(2), 242-255.

King, R.E., Tang, H. (1988). International Macroeconomic Adjustment, 1987-1992 : a World Model Approach (The World Bank).

Koch, N. and Basse Mama, H. (2016). European climate policy and industrial relocation: evidence from German multinational firms. Available at SSRN: https://ssrn.com/abstract=2868283 or http://dx.doi.org/10.2139/ssrn.2868283

Lanjouw, J. and Mody, A. (1996). Innovation and the international diffusion of environmentally responsive technology. *Research Policy.* 25(4), 549–571.

Levinson, Arik, and M. Scott Taylor. (2008). Unmasking the pollution haven effect. *International Economic Review*. 49 (1): 223-54.

List, J. A., Millimet, D. L., Fredriksson, P. G., & McHone, W. W. (2003). Effects of environmental regulations on manufacturing plant births: evidence from a propensity score matching estimator. *The Review of Economics and Statistics*, *85*(4), 944-952.

Marcu, A. (2016). Carbon market provisions in the Paris agreement (article 6). CEPS Special Report, Centre for European Policy Studies 128.

Mathiesen, L. and Maestad, O. (2004). Climate Policy and the Steel Industry: Achieving Global Emission Reductions by an Incomplete Climate Agreement. *The Energy Journal.* (4), 91–114.

Morgenstern, R. D., Pizer, W. A., & Shih, J. S. (2002). Jobs versus the environment: an industry-level perspective. *Journal of Environmental Economics and Management*, *43*(3), 412-436.

Newell, R., Jaffe, A. and Stavins, R. (1999). The induced innovation hypothesis and energy-saving technical change. *Quarterly Journal of Economics.* 114(3), 941–975.

Nordhaus, W. (2017). Evolution of modeling of the economics of global warming: changes in the dice model, 1992-2017. Cowles Foundation Discussion Paper No. 2084.

OECD (2013). Impacts of energy market developments on the steel industry.

Popp, D. (2002). Induced innovation and energy prices. *The American Economic Review.* 92(1), 160–180.

Popp, D. (2006). International innovation and diffusion of air pollution control technologies: the effects of nox and so2 regulations in the US, Japan and Germany. *Journal of Environmental Economics and Management.* 51(1), 46–71.

Reppelin-Hill, V. (1999). Trade and Environment: An Empirical Analysis of the Technology Effect in the Steel Industry. *Journal of Environmental Economics and Management.* 38(3), 283–301.

Roodman, D. (2008). Through the looking glass, and what ols found there: on growth, foreign aid, and reverse causality. Center for Global Development, Working Paper No. 137.

Schleich, J. (2007). Determinants of structural change and innovation in the German steel industry: an empirical investigation. *International Journal of Public Policy.* 2(1/2), 109–123.

Schwartz, N. D., Creswell, J. (23 October 2015). A Global Chill in Commodity Demand Hits America's Heartland. The New York Times.

Staiger, D., Stock, J. H., 1997. Instrumental variables regression with weak instruments. *Econometrica* 65, 557-586.

Stoft, S., Ockenfels, A. and Cramton, P. (2013). How to negotiate ambitious global emissions abatement. Working Paper .

Wagner, U. J., & Timmins, C. D. (2009). Agglomeration effects in foreign direct investment and the pollution haven hypothesis. *Environmental and Resource Economics*, *43*(2), 231-256.

WBG and ECOFYS (2015) *State and trends of carbon pricing*, 2015. World Bank Group.

WCA (2017). How is steel produced? World Coal Association. https://www.worldcoal.org/coal/uses-coal/how-steel-produced. Site consulted in June 2017.

Weitzman, M. L. (2014). Can negotiating a uniform carbon price help to internalize the global warming externality?. *Journal of the Association of Environmental and Resource Economists*. 1, 29–49.

Worrell, E., Price, L., and Martin, N. (2001) Energy efficiency and carbon dioxide emissions reduction opportunities in the US iron and steel sector. Energy, 26(5), 513-536.

Wooldridge, J. M., 1997. Multiplicative panel data models without the strict exogeneity assumption. *Econometric Theory* 13, 667-678.

Yellishetty, M., Mudd, G. M., Ranjith, P. and Tharumarajah, A. (2011). Environmental life-cycle comparisons of steel production and recycling: sustanaibility issues, problems and prospects. *Environmental Science & Policy*. 4, 650–663.

Appendices

A. Imputation method for plant capacity

For all the units for which we have capacity levels in 2014, we have assumed that historical capacity levels are equal to the 2014 capacity level. This is not a very strong assumption since we look at unit-specific capacity levels, not plant-specific levels. For all the units that closed down prior to 2014, we have imputed a capacity value based on technology (BOF vs. EAF), country and year of installation. More precisely, we ran a regression on all operative units in 2014. We took the logarithm of their capacity as the dependent variable. The explanatory variables used in the regression were technology fixed effects (BOF vs. EAF), technology by country fixed effects, and technology by year of installation fixed effects. This regression explained 67% of the variation in capacity across production units (R-squared of 0.67). We produced out-of-sample capacity estimates for the units that already closed down according to their technology, country and installation year.

Instead of relying on installed capacity, we directly use the number of units in a robustness check displayed in Table 4. Table A.1 below provides information on the number of steel units recorded in the data for three periods (1982-1990, 1991-2000, 2001-2014), with a breakdown by technology and geographical location. We only display the data used in the econometric estimation, i.e. the one that could be matched with coal price data. We cover 22 countries representing around 80% of world production.

Table A1: Descriptive statistics of James King data on steel units by country and main technology used in production (1982-2014)

Period	1982-1990		1991-2000		2001-2014	
Technology	BOF	EAF	BOF	EAF	BOF	EAF
North America	69.3	234.3	51.6	192.1	42.3	164.4
Europe	93.7	134.3	87.2	90.9	44.5	55.7
Asia	209.1	183.9	261.6	316.3	297.8	335.0
Other	7.0	3.9	N/A	N/A	2.0	2.0
Total	**379.1**	**556.4**	**400.4**	**599.3**	**386.6**	**587.2**

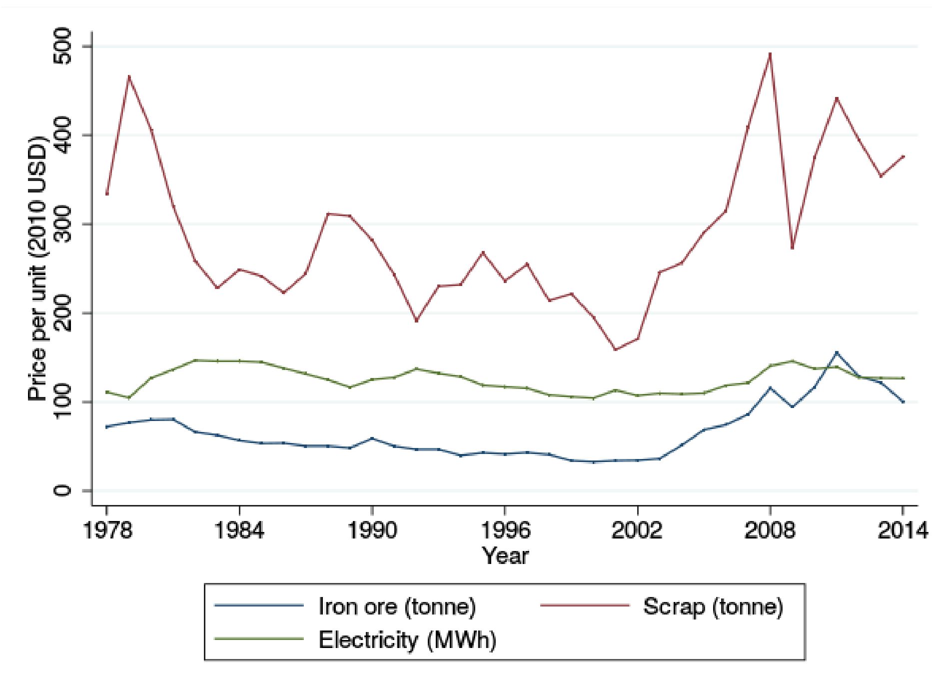
Notes: The table reports the average number of active units of steel-making production in each region of the world over three different periods. Note that the data was not available for all countries and all years. We only report figures for the restricted list of 22 countries used in the regressions. North America includes Canada and the US; Europe includes Germany, Spain, Italy, France, Poland, Belgium, Czech Republic, Sweden, Finland, the Netherlands, Portugal, Switzerland and Norway; Asia includes Turkey, China, Japan, Russia and India; Other consists of Australia (1982-1989) and Chile (2006-2014). We have no coal price data these two countries between 1991 and 2000.

The majority of steel units are in Asia and there are about 50% more EAF units than BOF units. However, this information is misleading because BOF units have a much larger production capacity.

B. Price of other inputs

Figure B1 below provide the average values of electricity, scrap and iron ore prices for the observations used in the base estimation (Table 4). The graph goes back to 1978 and therefore includes the minimum period of 4 years used to compute price expectations. The price of electricity has remained stable whereas the price of scrap and iron ore appear to have more volatility.

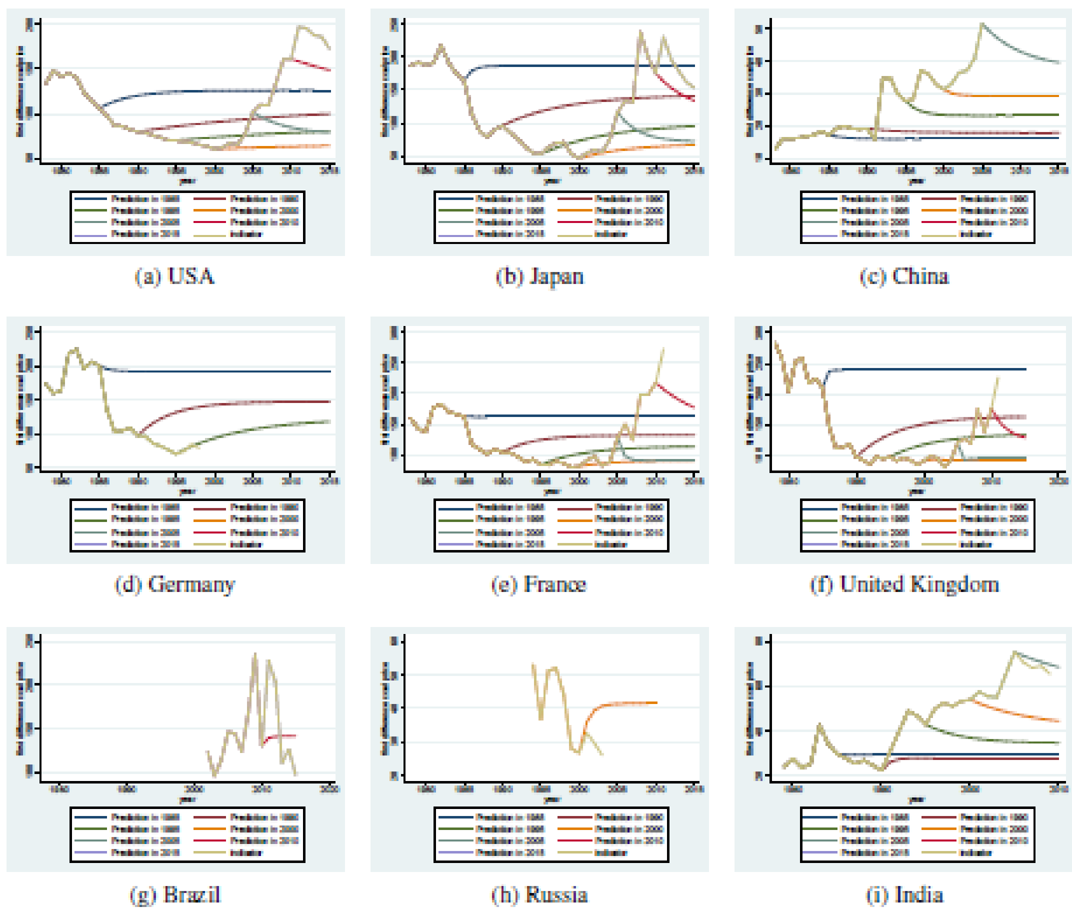
Figure B1: Expected coking coal prices as predicted with the ARIMA model



C. Output of ARIMA model to produce coal price expectations

Figure C1 shows, for a selection of countries, the expectations on coking coal prices computed with the ARIMA model presented in section 3. Predictions are only displayed for specific years (1995, 2000, 2005, 2010 and 2015) even though they were computed for each year and country used to run our econometric model.

Figure C1: Expected coking coal prices as predicted with the ARIMA model



(a) USA (b) Japan (c) China

(d) Germany (e) France (f) United Kingdom

(g) Brazil (h) Russia (i) India

D. Alternative specifications of coal prices

1. Using coal prices in levels and not in logarithm

Table D1.a presents the estimation where the expected coal prices are included in levels instead of the logarithmic form. Logically, the magnitude of the estimated betas is around 100 times smaller, but the relative effects remain stable.

Table D1.a: Specifications using expected coal prices in levels

Dependent Variable:		Log-log model		
Log. share of installed	BOF & EAF	BOF	EAF	EAF/(EAF+BOF)
capacity worldwide	(1)	(2)	(3)	(4)
Expected coking	-0.00331***	-0.00408***	-0.00238	0.00193*
coal price (in levels)	(0.00101)	(0.000864)	(0.00147)	(0.00101)
Log. pre-sample mean	0.678***	0.647***	0.793***	0.427***
(1970-1981)	(0.0474)	(0.0990)	(0.0614)	(0.117)
Year fixed effects	Yes	Yes	Yes	Yes
Observations	707	348	359	359
Countries	22	21	21	21

Notes: The model is estimated with 2SLS. The expected price of coking coal is instrumented with its one-year lag. Cluster-robust standard errors in parentheses. Clusters are set at country level and *, ** and *** respectively denote significance at 10%, 5% and 1% levels.

In Table D1.b, we have added a quadratic terms. Results suggest that plants are more sensitive to marginal increases in coal prices when they are low. This corroborates our preference for a model in logarithm.

Table D1.b: Specifications using expected coal prices in level and their squared value

Dependent Variable:		Log-log model		
Log. share of installed	BOF & EAF	BOF	EAF	EAF/(EAF+BOF)
capacity worldwide	(1)	(2)	(3)	(4)
Expected coking	-0.00506*	-0.00551**	-0.00473	0.00202
coal price (in levels)	(0.00289)	(0.00236)	(0.00409)	(0.00237)
squared	0.00000398	0.00000320	0.00000535	-0.000000206
	(0.00000470)	(0.00000419)	(0.00000643)	(0.00000344)
Log. pre-sample mean	0.677***	0.644***	0.796***	0.426***
(1970-1981)	(0.0461)	(0.0971)	(0.0580)	(0.115)
Year fixed effects	-0.00506*	-0.00551**	-0.00473	0.00202
Observations	707	348	359	359
Countries	22	21	21	21

Notes: The model is estimated with 2SLS. The expected price of coking coal is instrumented with its one-year lag. Its squared value is also instrumented with its first lag. Cluster-robust standard errors in parentheses. Clusters are set at country level and *, ** and *** respectively denote significance at 10%, 5% and 1% levels.

2. Using the contemporaneous price of coal

Table D2 reports the estimation results obtained when using the current coal prices instead of the expected ones. Results are similar, since column 4 shows that an increase in the price of coking coal leads to a shift in production capacity towards EAF.

Table D2: Specifications using contemporaneous coal prices

Dependent Variable:	Log-log model			
Log. share of installed capacity worldwide	BOF & EAF (1)	BOF (2)	EAF (3)	EAF/(EAF+BOF) (4)
Log. coking coal price	-0.399***	-0.417***	-0.381**	0.222*
	(0.126)	(0.101)	(0.188)	(0.116)
Log. pre-sample mean	0.694***	0.652***	0.810***	0.407***
(1970-1981)	(0.0443)	(0.111)	(0.0472)	(0.122)
Year fixed effects	Yes	Yes	Yes	Yes
Observations	775	383	392	392
Countries	22	21	21	21

Notes: The model is estimated with 2SLS. The price of coking coal is instrumented with its one-year lag. Cluster-robust standard errors in parentheses. Clusters are set at country level and *, ** and *** respectively denote significance at 10%, 5% and 1% levels.

3. Using steam coal prices instead of coking coal prices

In Table D3, we use expected steam coal prices instead of expected coking coal ones. The model instruments for coal prices and, therefore, may correct for measurement errors. The point estimate for BOF is negative and similar to the one found in the base model. The same applies to EAF.

Table D3: Specifications using expected stream coal prices

Dependent Variable:	Log-log model			
Log. share of installed capacity worldwide	BOF & EAF (1)	BOF (2)	EAF (3)	EAF/(EAF+BOF) (4)
Log. expected steam	-0.622***	-0.670***	-0.487*	0.0832
coal price	(0.190)	(0.198)	(0.249)	(0.154)
Log. pre-sample mean	0.732***	0.669***	0.857***	0.623***
(1970-1981)	(0.0337)	(0.0823)	(0.0454)	(0.0765)
Year fixed effects	Yes	Yes	Yes	Yes
Observations	803	376	427	427
Countries	22	21	21	21

Notes: The model is estimated with 2SLS. The log. expected price of steam coal is instrumented with its one-year lag. Cluster-robust standard errors in parentheses. Clusters are set at country level and *, ** and *** respectively denote significance at 10%, 5% and 1% levels.

E. Choice of estimation method

1. Using a different pre-sample period

Using a longer pre-sample period starting in 1960 has little influence on the results obtained (see Table E1 below).

Table E1: Specifications with a pre-sample starting in 1960

Dependent Variable:	Log-log model			
Log. share of installed	BOF & EAF	BOF	EAF	EAF/(EAF+BOF)
capacity worldwide	(1)	(2)	(3)	(4)
Log. expected coking	-0.471***	-0.609***	-0.358*	0.268**
coal price	(0.151)	(0.195)	(0.202)	(0.123)
Log. pre-sample mean	0.629***	0.578***	0.707***	0.350***
(1960-1981)	(0.0484)	(0.107)	(0.0494)	(0.0899)
Year fixed effects	Yes	Yes	Yes	Yes
Observations	707	348	359	359
Countries	22	21	21	21

Notes: The model is estimated with 2SLS. The logarithm of the expected price of coking coal is instrumented with its one-year lag. Cluster-robust standard errors in parentheses. Clusters are set at country level and *, ** and *** respectively denote significance at 10%, 5% and 1% levels.

2. Results with a fixed effect model

In Table E2, we present the results obtained when using a fixed-effect estimator. This model assumes that expected coking coal prices are fully exogenous. Results are biased and the model is inconsistent.

Table E2: Specifications with a fixed effect model

Dependent Variable:	Log-log model			
Log. share of installed	BOF & EAF	BOF	EAF	EAF/(EAF+BOF)
capacity worldwide	(1)	(2)	(3)	(4)
Log. expected coking	0.527**	0.429**	0.672**	0.234*
coal price	(0.226)	(0.161)	(0.279)	(0.134)
Year fixed effects	Yes	Yes	Yes	Yes
Observations	750	369	381	381
Countries	22	21	21	21

Notes: Cluster-robust standard errors in parentheses. Clusters are set at country level and *, ** and *** respectively denote significance at 10%, 5% and 1% levels.

3. Results with a first difference estimator

Tables E3.a and E3.b feature first-difference estimators. Table E3.a uses a transformed model based on Chamberlain (1992) and Wooldridge (1997). We use the second lagged difference of the independent variable as an instrument, but instrumentation is very weak. All effects are very large, yet not statistically different from 0 because very imprecisely estimated.

Table E3.a: Specifications with a FD model based on Chamberlain (1992) and Wooldridge (1997)

Dependent Variable: Log. share of installed capacity worldwide	Log-log model BOF & EAF (1)	BOF (2)	EAF (3)	EAF/(EAF+BOF) (4)
Log. expected coking coal price	-1.648	3.525	-1.070	-0.618
	(2.365)	(12.37)	(1.126)	(0.545)
Year fixed effects	Yes	Yes	Yes	Yes
Weak identification test				
Maximal IV size	>25%	>25%	>25%	>25%
Observations	622	306	316	316
Countries	22	21	21	21

Notes: The model is estimated in first differences and with 2SLS. The logarithm of the difference in the expected price of coking coal is instrumented with its second year lag. Cluster-robust standard errors in parentheses. Clusters are set at country level and *, ** and *** respectively denote significance at 10%, 5% and 1% levels.

The results of table E3.a are also likely to be biased towards a first differenced model where we would assume that the coal price is exogenous (as shown in Staiger and Stock, 1997; and Blundell and Bond, 1998). Such a first differenced model is reported in Table E3.b. Results are inconsistent due to the violation of the exogeneity assumption.

Table E3.b: Specifications with a FD model with assumption of exogenous coal prices

Dependent Variable: Log. share of installed capacity worldwide	Log-log model BOF & EAF (1)	BOF (2)	EAF (3)	EAF/(EAF+BOF) (4)
Log. expected coking coal price	0.0883	0.0649	0.100	0.0604
	(0.0563)	(0.0452)	(0.0789)	(0.0554)
Year fixed effects	Yes	Yes	Yes	Yes
Observations	707	348	359	359
Countries	22	21	21	21

Notes: The model is estimated in first differences and with OLS. Cluster-robust standard errors in parentheses. Clusters are set at country level and *, ** and *** respectively denote significance at 10%, 5% and 1% levels.

F. Analysis with disaggregated data

We disaggregate the data and take plants as the unit of reference. Within each plant, we calculate the installed capacity in BOF and EAF processes. To avoid any bias caused by the relative increase in steel production over time, we normalize this variable over the sum of the capacity of all units operative at time *t*. We then merge the steel plant data with the input price data. Table F1 provides summary statistics at plant level, separately for BOF and EAF.

Table F1: Plant-level summary statistics

	Sites with BOF		
Variable	Observations	Mean	Standard deviation
Share of global capacity	5,648	0.22	0.29
Expected coking coal price	5,648	69.6	60.5
Expected iron ore price	4,593	53.6	20.2
Expected electricity price	5,398	93.7	48.3
Expected scrap price	4,076	255.1	85.0
	Sites with EAF		
Variable	Observations	Mean	Standard deviation
Share of global capacity	21,997	0.02	0.04
Expected coking coal price	21,997	97.8	60.7
Expected iron ore price	18,940	60,7	23.7
Expected electricity price	18,913	125.4	58.8
Expected scrap price	17,701	285.7	88.3

Notes: Share of global capacity is given in percentage points and therefore ranges from 0 to 100. Prices are expressed in 2010 USD. Summary statistics are only provided when expected coal price data is available.

Let's denote $M_{j,i,s,t}$ the share of global installed capacity of plant *j* with technology *s*, in country *i* at time *t*. We want to run a model such that:

$$M_{j,s,i,t} = f(p^*_{i,t})$$

We consider two issues in order to choose the right specification. First, $p^*_{i,t}$ is likely to be endogenous because of simultaneity and omitted variable bias. Second, changes at plant level are likely to be much more time-persistent than changes at country-level. Unit openings and closures are frequent when observed at national level, but rather rare at plant level. A dynamic specification is therefore required with disaggregated data.

We estimate the effect of coal prices on plant size with the following econometric estimation:

$$M_{j,s,i,t} = a. M_{j,s,i,t-1} + b \ln\left(p^*_{i,t}\right) + c_t + r_j + e_{j,s,i,t}$$

This model can be estimated with a first difference estimator, using the lags of the dependent variable and the endogenous regressors as instruments.[24] There is no risk of small sample bias in the present case since the disaggregated sample is large enough. We however need to mitigate strong efficiency losses caused by an IV estimation in first differences.

We use the Blundell-Bond panel estimation method. This comes at a cost since this model relies on stronger assumption. First, we must assume that the process is mean-stationary. In this respect, we avoid any bias caused by the relative increase in steel production over time since the dependent variable has been normalized. We furthermore run separately the regressions for EAF and BOF units. Another limitation is that the instrumentation of the lag dependent variable with deeper lags has been criticized because it relies on assumptions regarding the correlation of the error terms from one period to another. While our estimation method suffers from this drawback, we make sure that our instrumentation strategy respects the output of the autocorrelation tests usually run with a linear dynamic panel data model.

Estimation results are displayed in Table F1 and Table F2. In Table F1, we consider only coking coal as an input, while we consider all four inputs in Table F2. In both tables, columns 1 report the results for BOF and EAF jointly, column 2 for BOF and column 3 for EAF separately. For columns 1–3, the dependent variable is the share of the capacity that belongs to a plant over the total capacity registered in the dataset at time *t*. We have expressed it in percentage points (from 0 to 100). The Blundell-Bond estimation is obtained using the first lags of the explanatory variables as instruments in the equations in levels, and their differenced value in the differenced equation. For the lagged dependent variable, the second lag of the dependent variable is being used. This strategy seems valid, considering that the AR(2) tests find no autocorrelation. Column 4 reports the results related to a change in the share of EAF over the total capacity of a plant. It is also expressed in percentage points. For this regression, time persistence seems higher and we use the 3rd lags to instrument for the explanatory variables, and 4th lag for the dependent variable.

[24] We have not reapplied the log-log estimation and pre-sample mean estimator used on national-level data. A couple of issues make this choice less attractive with disaggregated data. Both the log-log specification and pre-sample estimator are useful only if the variable of interest and the pre-sample average are non-null. While this is the case for $N_{s,i,t}$ with national level data, many plants may not be in operation for the full length of the study period. We therefore record several zeros in $M_{j,s,i,t}$. Using logarithms would make these observations disappear while these zero values are relevant to our analysis. Above all, approximating plant-specific fixed effects with a pre-sample mean equal to zero for plants that do not yet exist is problematic.

In Table F2, we find that a 1-unit increase in the log. of expected coal prices reduces BOF capacity by 0.005 points, or about 2.5% of average BOF capacity in the estimation sample for this regression (at 0.22 points – see Table F1). Results for EAF are statistically insignificant but reflect a smaller relative decrease of 1.5%.[25] We globally find that a 1-unit increase in the log. of expected coal prices increases the absolute share of EAF processes by around 0.5%. These effects might appear to be small, since expected coal prices average 92 dollars (their log is 4.27 on average in the sample, with a standard deviation of 0.72). However, long-run effects might be quite large, considering that the process is highly time-persistent with estimates for the lag. dependent variable over 0.95. In Column 4, the long-run multiplier is more than 50 and the extrapolated, long run effect of a 1-unit increase in log. expected prices correspond to a 25 percentage-point increase (0.5% times 50).

Table F3 confirms these findings. We furthermore find that the mix between EAF and BOF is also sensitive to electricity prices in column 4. Overall, we find that the disaggregated results are consistent with the national-level results, as they show that coking coal prices have an incidence on installed capacity, with stronger effect on BOF than on EAF.

Table F2: Blundell-Bond dynamic panel estimation of the installed capacity of plants

Dependent Variable:	Log-log Model			
Share of installed capacity worldwide (0-100 percentage points)	BOF & EAF (1)	BOF (2)	EAF (3)	EAF/(EAF+BOF) (4)
Expected log. coking coal price	-0.002^{***}	-0.0055^{***}	-0.0003	0.517^{**}
	(0.0006)	(0.0016)	(0.0003)	(0.257)
Lagged dependent Variable	0.972^{***}	0.963^{***}	0.988^{***}	0.979^{***}
	(0.003)	(0.005)	(0.018)	(0.007)
Year fixed effects	Yes	Yes	Yes	Yes
AR(1) test (p-value)	<0.01	<0.01	<0.01	<0.01
AR(2) test (p-value)	0.66	0.66	0.79	<0.01
AR(3) test (p-value)	0.25	0.25	0.88	0.18
Observations	26,771	5,437	21,334	8,356
Number of plants	1,099	236	863	462

Notes: The model is estimated with the Blundell-Bond two step estimator. In columns 1–3, the first lags of the explanatory variables as instruments in the equations in levels, and their differenced value in the differenced equation. For column 4, we use the 3rd lags to instrument for the explanatory variables, and 4th lag for the dependent variable. Clusters are set at country level and *, ** and *** respectively denote significance at 10%, 5% and 1% levels.

[25] The average capacity per plant is 0.02 points and the decrease is by about 0.003 percentage points.

Table F3: Blundell-Bond dynamic panel estimation with additional inputs

Dependent Variable: Share of installed capacity worldwide (0-100 percentage points)	Log-log Model BOF & EAF (1)	BOF (2)	EAF (3)	EAF/(EAF+BOF) (4)
Expected log. coking coal price	-0.00352***	-0.00761***	-0.000821***	0.901*
	(0.000685)	(0.00139)	(0.000288)	(0.492)
Expected log. iron ore Price	-0.000317	0.00681	-0.0000773	-0.259
	(0.00152)	(0.00526)	(0.000766)	(0.603)
Expected log. electricity Price	0.00142	0.00590*	0.000255	-0.537**
	(0.00122)	(0.00355)	(0.000563)	(0.228)
Expected log. scrap Price	-0.00204*	0.000871	-0.000606	0.160
	(0.00114)	(0.00337)	(0.000676)	(0.253)
Lagged dependent Variable	0.972***	0.963***	0.988***	0.974***
	(0.00318)	(0.00512)	(0.0180)	(0.0144)
Year fixed effects	Yes	Yes	Yes	Yes
AR(1) test (p-value)	<0.01	<0.01	<0.01	<0.01
AR(2) test (p-value)	0.96	0.83	0.22	0.02
AR(3) test (p-value)	0.03	0.03	0.84	0.25
Observations	18,390	3,784	14,606	6,542
Number of plants	1,031	223	808	422

Notes: The model is estimated with the Blundell-Bond two step estimator. In columns 1–3, the first lags of the explanatory variables as instruments in the equations in levels, and their differenced value in the differenced equation. For column 4, we use the 3rd lags to instrument for the explanatory variables, and 4th lag for the dependent variable. Clusters are set at country level and *, ** and *** respectively denote significance at 10%, 5% and 1% levels.

G. Estimation and simulation with heterogeneous effects across region and technology

With region-specific effects, we find that expected coal prices have a negative impact on BOF capacity, but no effect on EAF. The simulation leads to a smaller decrease in overall demand (by about 15%) and a more even repartition of effects between technology shifts and relocations. We still find redistributive effects between Europe and Asia. In the scenario with a global implementation, the losses of Asian plants are attenuated since they start relying more on EAF.

Table G1: Specifications with region-specific effects

Dependent Variable: Log. share of global capacity		Log-log model	
		BOF	EAF
		(1)	(2)
Log. expected coking coal price			
	North America	-0.381***	0.043
		(0.051)	(0.126)
	Europe	-0.423***	-0.131
		(0.055)	(0.118)
	Asia	-0.298***	0.116
		(0.064)	(0.146)
	Other	-0.510***	-0.156
		(0.080)	(0.148)
Log. pre-sample mean (1970-1981)		0.618***	0.660***
		(0.101)	(0.084)
Year fixed effects		Yes	Yes
Observations		348	359
Countries		21	21

Notes: The model is estimated with 2SLS. The five first lag of the logarithm of expected coal price are used as instruments. Cluster-robust standard errors in parentheses. Clusters are set at country level and *, ** and *** respectively denote significance at 10%, 5% and 1% levels.

Table G2: Simulation results with region-specific effects

Indicator	World capacity that is BOF (%)			World capacity that is EAF (%)		
Scenario	BAU	Europe	All	BAU	Europe	All
World	73.3	72.7	67.2	26.7	27.3	32.8
North America	10.9	11.3	10.3	10.6	10.9	12.8
Europe	16.5	14.0	16.0	4.8	4.6	5.3
Asia	45.3	46.8	40.5	11.3	11.7	14.6
World capacity (BAU = 100)	73.3	70.4	57.0	26.7	26.4	27.8

Notes: North America includes Canada and the US; Europe includes Germany, Spain, Italy, France, Poland, Belgium, Czech Republic, Sweden, Finland, the Netherlands, Portugal, Switzerland and Norway; Asia includes China, Japan, Russia, Turkey and India. Other countries included in the world average are Australia (1982-1989) and Chile (2006-2014).