Manuscript version: Author's Accepted Manuscript
The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:
http://wrap.warwick.ac.uk/136845

How to cite:
Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:
Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

warwick.ac.uk/lib-publications

Information Asymmetry in voluntary environmental agreements (VEAs): Theory and evidence from UK climate change agreements (CCAs)

Morakinyo O. Adetutu[†] and Eleni Stathopoulou[*]

March 2020

Abstract

Voluntary environmental agreements (VEAs) are often plagued by adverse selection problems, because the regulator has imperfect information about firm-specific production technologies and abatement costs. To the best of our knowledge, this paper is the first one to explore this issue using the UK climate change agreement (CCA) as a case study. First, we present a theoretical emulation of the program. Second, we resolve the regulator's asymmetric information problem by estimating unobserved energy efficiency using production theory. Third, we use microdata from three confidential manufacturing surveys to empirically test how limited information impacts resource allocation within the scheme. In line with the problem of limited information about firm production technologies, we find that firms with lower levels of energy efficiency receive higher CCA tax discounts. This finding holds over a range of robustness tests.

JEL Classification: C3; D21; H23; L6; O31; Q54

Keywords: Climate change levy (CCL), Climate Change Agreement (CCA), Hyperbolic Distance Function (HDF), Energy Efficiency, Information Asymmetry, Manufacturing.

[†] Corresponding author: Department of Global Sustainable Development, University of Warwick, Coventry, United Kingdom CV4 7AL. *E-mail address*: Morakinyo.Adetutu@warwick.ac.uk (M. Adetutu)

[*] Nottingham Business School, Nottingham Trent University, Shakespeare Street, Nottingham, United Kingdom NG1 4FQ. *E-mail address*: Eleni.Stathopoulou@ntu.ac.uk (E. Stathopoulou).

We are grateful to Editor Francis Teal and two anonymous referees, for their constructive comments which resulted in significant improvements to this work. We also thank, without implicating, Thomas Weyman-Jones, Piercarlo Zanchettin,, Rob Ackrill, Kayode Odusanya and Nina Boogen for helpful comments and discussions. Earlier versions of this paper were presented at the workshop on Tax Competition, Environmental Competition and Energy Regulation, Loughborough University, UK, March 2016; 41st IAEE international conference, Groningen, Netherlands, June 2018. The authors are thankful to the discussants and participants for their helpful comments and suggestions. The usual disclaimer applies.

1. Introduction

Voluntary environmental agreements (VEAs) between environmental regulators and greenhouse gas (GHG) emitting units often represent a second-best policy tool to combat negative externalities arising from environmental pollution. VEAs have become key environmental policy instruments because they provide greater flexibility and preserve firms' international competitiveness relative to traditional environmental policy instruments. However, these programs are often plagued by adverse selection challenges arising from (i) the voluntary/opt-in nature of the programs and (ii) asymmetric information given the regulator's incomplete information about firm-specific production technologies and abatement costs[1]. Consequently, the regulator faces a trade-off between productive efficiency and information rent extraction (Montero, 2000, p. 275).

A classic example of this trade-off is provided by UK climate change agreements (CCA), one of the two components of the Climate Change Levy (CCL) package- the UK's most important policy instrument for limiting industrial carbon emissions (see HM Government, 2006; Martin et al., 2014). The CCL package was introduced in April 2001 and comprises (i) a climate change levy, i.e., a tax[2] added to energy bills of non-domestic or business users and (ii) climate change agreements (CCA), a program of VEAs negotiated between sector associations of pollution-intensive industries and the UK environmental regulator (See Section 2 for details). When an energy intensive firm joins the CCA scheme, it enters a voluntary agreement to adopt energy efficiency or carbon reduction targets in exchange for a discount on its CCL liabilities.[3] While a healthy strand of the literature evaluates the

[1] For a detailed treatment of adverse selection in regulatory economics, see Laffont and Tirole (1993). For applications to incentive regulation and voluntary environmental programs, see Montero (2000), Hawdon, et al. (2007), Arguedas and Van Soest (2009).

[2] The fuels taxed under the scheme, along with their tax rates, are electricity (10.1%), natural gas (16.5%), coal (6.1%) and non-transport liquefied petroleum gas (LPG) (8.2%) (See Martin et al., 2014, p.3). Hence, the levy is a non-uniform tax, with different fuels having different rates, but which do not vary with their carbon content. Consequently, consistent with the extant literature, we treat the levy as an energy tax or an implicit carbon tax, rather than a pure carbon tax (see Pearce, 2006; Martin et al., 2014).

[3] Presently, participating plants can receive up to 90% CCL discount under a CCA. See http://www.cclevy.com/

incentive and impact of the tax component of the policy package (e.g. Cambridge Econometrics, 2005; Martin, *et al*., 2014), rigorous evaluation of the CCA scheme is scarce, especially in the context of how asymmetric information shapes the stringency and ultimately the resource allocation within the CCA scheme. This is potentially for two issues.

First, a rigorous evaluation of the CCA scheme requires information on participants' energy efficiency and abatement potential. However, this information is not observable to the regulator, underscoring the difficulty in arriving at a convincing identification strategy. Second, a dearth of suitable microeconomic data has limited the scope for much-needed empirical analysis of the UK CCA scheme. The CCA scheme offers an appropriate case to investigate the linkages between asymmetric information, policy stringency and resource allocation. For instance, CCA agreements are negotiated in the face of the regulator's limited knowledge about participating firms' underlying energy efficiency and control costs; and there is therefore ongoing concern about the stringency of the negotiated CCA targets (see Ekins and Etheridge, 2006; Barker et al., 2007; Martin et al. 2014). Besides, the huge social costs[4] associated with the CCAs further justify the evaluation of the potential misallocation within VEAs, as undertaken in this study. Thus, our main contribution is twofold.

First, we present a theoretical emulation of the program to provide insight into the mechanism through which information asymmetry impacts social allocation within the CCA scheme. We argue that the success of the CCA scheme depends on how this efficiency-information trade-off implicates the social allocation problem facing the environmental regulator[5]. Second, we address the regulator's asymmetric information problem by estimating

[4] This social cost is potentially two-fold. First, there is a direct financial cost arising from the loss of government revenue resulting from the tax discounts. Recent estimates suggest a total revenue loss of £940 million during the period 2013- 2017 (HMRC, 2018). A second potential cost is the implicit social cost arising from higher emissions than with a more stringent policy. For instance, Cambridge Econometrics (2005) indicate that "the energy (and therefore carbon) saving and energy-efficiency targets would have been met without the CCAs" (p. 7).

[5] Despite the initial CCA design being a discovery process, it seems plausible that a rational government or environmental regulator would, in principle, prefer to minimize potential social misallocation by allocating higher CCL discounts to more efficient firms, rather than vice versa.

the unobservable levels of energy efficiency using parametric production frontier analysis. This approach permits an informed assessment of the stringency of the CCA targets by gauging the slack between potential energy savings of firm production technologies and the negotiated CCA targets.

The above approach also acknowledges that regulatory stringency is likely to become the crux of future policy discussions. Moreover, the present public perception that UK climate change policies are weak[6] adds another layer of justification for a closer scrutiny of the CCAs' regulatory stringency. A recent paper by the Department for Business, Energy & Industrial Strategy stated inter alia: *"The UK Government is currently evaluating the effectiveness of the CCA Scheme"* (BEIS, 2019, p.54). This research is therefore timely in contributing to the sparse body of evidence assessing the CCA scheme. To do so, we analyse an interesting but hitherto unexplored intersection between three different literatures: asymmetric information, environmental regulation and parametric efficiency analysis.

The remainder of the paper is organized as follows. In section 2, we provide an overview of the CCL scheme. Section 3 present a theoretical emulation of the CCL package using a model that highlights the importance of policy stringency and how this may shape social (mis) allocation. We also resolve the regulator's problem of imperfect information about firm[7] production technologies by describing a production theory approach for identifying unobservable firm-level energy efficiency. Section 4 sets out our econometric methodology for empirically testing our model assertion. In Section 5, we provide details of our dataset, which

---

[6] In a recent YouGov survey, around two-thirds of respondents think that the UK government is not doing enough on climate change. See https://www.documents.clientearth.org/library/download-info/clientearths-climate-snapshot/. This view is also shared by the scientific community: https://www.theguardian.com/tv-and-radio/video/2019/jul/09/we-cannot-be-radical-enough-david-attenborough-climate-emergency-video

[7] Participation in the CCA scheme is at the facilities level, hence the plant-level is the appropriate unit of analysis. In the theoretical analysis that follows, we set out our model in line with production theory using 'firms' rather than 'plants'. However, in our empirical estimation, we follow the existing literature (e.g. Martin et al., 2012, 2014) by employing plant-level microdata. Thus, our reference to firms connotes plant-level production units.

we draw from three confidential UK manufacturing surveys. Section 6 contains the econometric results, while Section 7 concludes.

2. Overview of the UK CCA Scheme

The CCA scheme is a program of voluntary agreements [8] negotiated between sector associations of energy intensive industries and different UK environmental regulators across different time periods. At inception, negotiations were between sector associations and the Department for Environment, Food & Rural Affairs (DEFRA)[9] which originally negotiated 44 umbrella agreements (Martin et al., 2014). At the end of each target period, the sector associations were required to report industry-wide performance to verify that targets had been met. When a sector fell short of the negotiated target, DEFRA verified compliance at the unit or firm level. Subsequently, a non-compliant unit was not re-certified for the rebate in the following period.

At the firm level, it is easy to identify a major challenge for regulatory stringency as participants may be unwilling to yield private information on their true energy efficiency savings. Given the absence of any direct mechanism for eliciting this private information, it is unsurprising that compliance rates were high in the three periods covered in this study: 88%, 98% and 99%, respectively (AEAT, 2004, AEAT 2005, AEAT, 2007). Moreover, the compliance mechanism at the sector level poses an additional challenge to the scheme's stringency, given that the lower bound to compliance cost was zero during the period of interest: plants were re-certified for CCA rebates even when they missed their efficiency targets, provided that sector-level targets as a whole were met (NAO, 2007; Martin et al, 2014).

---

[8] There are two types of agreements ("umbrella agreements" and "underlying agreements") under the CCA, reflecting that the energy efficiency targets are negotiated at two levels. Umbrella agreements pertain to sector-wide targets for energy consumption or emissions while underlying agreements embody reduction targets specific to a firm or facility with an umbrella agreement. See De Muizon and Glachant (2004) for details.

[9] In 2008, the Department of Energy & Climate Change (DECC) (now Department for Business, Energy and Industrial Strategy (BEIS)) assumed responsibility for the CCA negotiations with the sector associations, with The Environment Agency administering the scheme on behalf of the BEIS

Despite the stringency discussions above, we note that the scheme was not initially conceived as a stringent policy intervention aimed at placing binding constraints on participants' production decisions. Specifically, the initial design of the CCA scheme mirrored a discovery process in which plants evaluate their abatement potential towards identifying efficiency savings rather than place binding constraints on their production technology choices[10]. While the scheme aimed to offer some signal of the social marginal cost of energy consumption and pollution activities, in practice, other 'political economy' considerations (e.g. preserving the international competitiveness of affected sectors) featured prominently in the program's design. Nevertheless, the program has evolved over the years[11] to reflect the increasing importance of environmental cost signalling through a vigorous anti-climate change policy stance as evidenced by some key changes to the CCA program.

For instance, CCA eligibility was based on pollution-generating activities under the Pollution Prevention and Control (PPC) Act 1999, but this evolved into a wider eligibility rule to cover energy intensive activities. The eligibility rules were amended in 2006 to include energy intensive sectors that were hitherto not PPC-regulated, increasing the number of sectors with umbrella agreements from 44 to 54[12]. This new arrangement therefore widened sectoral eligibility and participation within the CCA, while also accounting for exposure to international competition[13]. Furthermore, the price signal and incentives available under the scheme have also evolved since inception. For instance, the EU Emissions Trading Scheme (EU ETS), introduced in 2005, potentially allowed compliant and non-compliant plants to trade carbon

[10] We thank an anonymous referee for highlighting this important feature of the CCA scheme.
[11] See Martin et al (2014), McEldowney and Salter (2016) for details.
[12] The current scheme which started in 2013 contains 51 sectors, following the classification merger of some sectors. See sector list at https://www.gov.uk/government/publications/climate-change-agreement-cca-sector-contact-list
[13] These new sectors are subject to relative energy intensity (EI) and international competitiveness tests, such as (i) the share of energy consumed in total sectoral production value must be 3% or greater, and (ii) a sectoral import penetration level of 50% or more is required. The joint importance of energy intensity and import penetration levels in the 2006 eligibility rules reflects the regulator's intention to better reflect the social cost of emissions by widening CCA participation by energy intensive sectors while also attempting to preserve the sectors' international competitiveness. This delicate balance is best seen in the fact that sectors that fail the 50% import penetration requirement must have energy intensity levels of 10% or more, rather than the baseline 3%. We thank an anonymous referee for providing invaluable insights on the evolution of CCA eligibility rules.

savings[14] (Martin et al., 2014). This may well have added another layer to the incentive provided by the CCA scheme. In April 2013, another crucial change raised maximum discount rate from 80% to 90%, to provide further incentive for firms to invest in energy efficient production technologies in the face of rising energy saving targets.

Despite the significant modifications to the CCA program detailed above, we note that two persistent challenges remain. First, the lack of a direct mechanism for eliciting private information about firms' production technologies and true energy efficiency savings is likely to inhibit regulatory stringency. Second, the strong bearing of international competitiveness considerations on the program's design and implementation is likely to further weaken its regulatory stringency. These matters central to this study. While the CCA scheme is still in place today, our analysis focuses on its first three reporting periods covering 2001-2007, data limitations within the UK business microdata (see Section 5 for details). Thus, the empirical analysis does not evaluate potential information-efficiency trade-offs within the CCA scheme after 2007.

3. Theoretical framework

3.1 Emulation of the CCL package

In this section, we emulate the CCL package (the tax liability and the rebate components) to examine the relationship between efficiency and the allocated discounts. Consider an imperfectly competitive market for a homogeneous good supplied by *n* firms. We denote firm i's ($i = 1, \ldots, n$) output by $y_i((z_i(e_i))$, where z_i>0 denotes emission intensity which is negatively affected by the efficiency level $e_i > 0$. Total output is denoted by $Y = \sum_{i=1}^{n} y_i\,(z_i(e_i))$ and total emissions by $Z\ = \sum_{i=1}^{n} z_i\,(e_i)$. Assume that output is increasing in emission intensity so $y_z(z_i(e_i)) > 0$. The inverse demand function illustrating the consumers' benefit from the consumption of the good reads:

---

[14] A similar UK carbon market, the UK Emissions Trading Scheme (UK ETS), operated between 2002 and 2006.

$$P = P(Y) \quad [1]$$

The regulator adopts a pollution tax to reflect the social cost of emissions. However, to also preserve firms' international competitiveness, it offers reduced tax liability through a CCA tax discount $d_i(e_i)$ in exchange for energy efficiency level e_i (CCA scheme). The regulator's problem is to maximise the following welfare function, which is the sum of consumer and producer surplus and tax revenues minus the discount given. This is subject to the constraint that the discount offered cannot exceed the tax:

$$\max_{y_i \geq 0, z_i \geq 0, e_i \geq 0} W = CS + \sum_{i=1}^{n} \Pi_i + t\big(Z(e_{1,\dots,n})\big) - D\big(e_{1,\dots,n}\big) \quad [2]$$

$$\text{s.t.} \qquad t\big(Z(e_{1,\dots,n})\big) - D\big(e_{1,\dots,n}\big) \geq 0$$

where the tax revenues are denoted by $t\big(Z(e_{1,\dots,n})\big)$ and the discounts offered by $D\big(e_{1,\dots,n}\big) = \sum_{i=1}^{n} \mathrm{d}_i(e_i)$. Consumer surplus, CS, which is reduced by the level of pollution (emissions) can be written as:

$$CS = \int_0^Y P(v)dv - P(Y)Y\big(Z_i(e_{1,\dots,n})\big) - Z_i(e_{1,\dots,n})Y\big(Z_i(e_{1,\dots,n})\big) \quad [3]$$

Profits, Π_i, are

$$\Pi_i = \big[P(Y) - C_i\big(z_i(e_i)\big)\big]y_i\big(z_i(e_i)\big) - t\big(z_i(e_i)\big) + d_i(e_i). \quad [4]$$

The firm's cost of production is denoted by C_i and captures the innovation effort and implementation costs for attaining energy efficiency[15]; hence it is assumed to be increasing in efficiency $\frac{\partial C_i}{\partial e_i} > 0$.

Using [3] and [4] the first order condition of [2] w.r.t efficiency gives us[16]:

[15] This implies that attaining energy efficiency savings is costly since this requires some firm effort such as R&D and implementation-related activities.

[16] The Lagrange multiplier is positive so that the constraint is binding.

$$\frac{\partial W}{\partial e_i} = P(Y)\frac{\partial Y_i}{\partial Z_i}\frac{\partial Z_i}{\partial e_i} - \frac{\partial Z_i}{\partial e_i}Y_i\big(Z_i(e_{1,\dots,n})\big) - Z_i(e_{1,\dots,n})\frac{\partial Y_i}{\partial Z_i}\frac{\partial Z_i}{\partial e_i} -$$

$$\frac{\partial C_i}{\partial Z_i}\frac{\partial Z_i}{\partial e_i}Y_i\big(Z_i(e_{1,\dots,n})\big) - C_i\big(Z_i(e_{1,\dots,n})\big)\frac{\partial Y_i}{\partial Z_i}\frac{\partial Z_i}{\partial e_i} + \lambda(\frac{\partial t}{\partial Z_i}\frac{\partial Z_i}{\partial e_i} - \frac{\partial D}{\partial e_i}) = 0$$

or

$$\frac{\partial D}{\partial e_i} = \frac{\frac{\partial Z_i}{\partial e_i}\left\{\frac{\partial Y_i}{\partial Z_i}[P(Y)-C_i\big(Z_i(e_{1,\dots,n})\big)-Z_i(e_{1,\dots,n})]-Y_i\big(Z_i(e_{1,\dots,n})\big)\left(\frac{\partial C_i}{\partial Z_i}+1\right)\right\}+\lambda\frac{\partial t}{\partial Z_i}\frac{\partial Z_i}{\partial e_i}}{\lambda} \quad [5]$$

3.2 Potential misallocation within the CCA

Does the scheme result in the misallocation of discounts? Eqn. [5] could help us explain when inefficient firms take advantage of this policy. In particular, this is the case when the numerator of the fraction on the RHS of eqn. [5] is negative, i.e., when[17]

$$\frac{\partial t}{\partial Z_i} < \frac{Y_i\big(Z_i(e_{1,\dots,n})\big)\left(\frac{\partial C_i}{\partial Z_i}+1\right) - \frac{\partial Y_i}{\partial Z_i}[P(Y)-C_i\big(Z_i(e_{1,\dots,n})\big)-Z_i(e_{1,\dots,n})]}{\lambda}.$$

This leads us to the following proposition.

PROPOSITION 1. *A negative relationship between the discount and energy efficiency is possible under the CCA scheme if*

$$\frac{\partial t}{\partial Z_i} < \frac{Y_i\big(Z_i(e_{1,\dots,n})\big)\left(\frac{\partial C_i}{\partial Z_i}+1\right) - \frac{\partial Y_i}{\partial Z_i}[P(Y)-C_i\big(Z_i(e_{1,\dots,n})\big)-Z_i(e_{1,\dots,n})]}{\lambda}.$$

Proposition 1 says that the designed CCA policy would give higher discounts to the more efficient firms if the tax response to a change in emission intensity is large enough. However, when structural parameters are conducive to a situation where the response of tax to a change in emission intensity is small, then firms can take advantage of the policy, even when they are inefficient. In other words, a less stringent scheme could give opportunity to inefficient firms to take advantage of the program.

[17] See appendix for a detailed derivation of the expression.

It is interesting to understand why (i.e., the mechanism behind) the outcome above may happen. Let us consider, as an example, one firm in two different settings. First, suppose that there is a big enough response of tax to changes in emission intensity. In this case, the firm would have stronger incentives to achieve higher efficiency and decrease its emission intensity further, so as to further reduce its tax liability. Alternatively, when these incentives are weak, the firm would not strive to reduce emission intensity significantly as the tax would not change much. Hence, the tax it would pay would be higher in the latter setting compared to the former. The way this policy is designed seems to play a key role in this outcome, as it implies that an inefficient firm in the latter setting would secure a higher discount, compared to the first setting, since the regulator aims to preserve international competitiveness and compensate for the higher tax by giving a higher discount. While we understand that the initial design of the scheme as a discovery process through target-setting, it is necessary, as mentioned above, to rigorously explore the efficiency-information trade-off and the mechanism behind a potential misallocation in the CCA scheme. This is because the CCA rebates (or any subsidy for that matter) are socially costly and society would prefer to reduce the public expenditure on such subsidies, especially if such subsidies enhance recipients' profitability (Arguedas and van Soest, 2009). Hence, it should be in the interest of a rational regulator to minimize misallocation within the scheme.

3.3 Measuring energy efficiency: a production theory approach

Theoretically, an omniscient regulator can observe firm-specific energy efficiency and assign the CCA discounts accordingly. In practice, however, the regulator cannot observe firm production technology and control costs. The efficiency targets of the CCA scheme are energy or carbon emission reduction targets, relative to a base year[18]. However, we believe that a better

[18] The energy reduction targets depend on "growth rate, hypothesis on technological evolution, market structure, negotiating skills of the sector association" (De Muizon and Glachant, 2004, p.4)

approach would be to design sector and firm-level targets based on *potential* energy efficiency savings within production technologies. Therefore, our empirical analysis aims to uncover the efficiency in the use of one of the factor inputs (energy use) by comparing the observed level of energy consumption with its optimal use. Given that the derivation of energy services is a production process, we propose a measure of (unobservable) energy efficiency using production theory. We assume that producers are optimizers who seek to maximize feasible energy services given their energy consumption and available production technology.

The microeconomic foundation of productive efficiency analysis, which dates to Farrell (1957), provides a starting point for the definition of firm-level efficiency. Within this framework, not all producers attain this optimization objective (e.g. due to poor management practices), i.e. some are inefficient[19], failing to maximize output given their inputs and production technology. This optimization failure can be estimated using production frontier analysis, in which technically inefficient producers lie beneath the (efficient) production frontier. The above theoretical arguments are often implemented empirically using Shepard's (1953, 1970) distance functions that radially contract a multi-input vector to produce multiple outputs[20].

However, in this study, we are interested in the efficiency relating to the use of a single input (i.e., energy consumption), obtained via a non-radial distance function, rather than the radial input distance function which yields a broader measure of technical efficiency in the use of all factor inputs. The reason is twofold. First, the focus on the energy input allows for a more precise measure of energy efficiency (EE) that is consistent with the CCA scheme's objective to improve energy efficiency rather than multi-input technical efficiency. Second, an energy-specific efficiency measure is consistent with the assumption underpinning our theoretical

[19] Leibenstein (1977) provides theoretical arguments treating inefficiency as a production problem relating to "the techniques of an input called management" (p. 312).
[20] See Fare, et al. (1994) and Khumbakar and Lovell (2003) for detailed discussions on distance functions.

model - that energy efficient technologies are costly, i.e., attaining energy efficiency savings requires some R&D effort and implementation-related activities: $\frac{\partial c_i}{\partial e_i} > 0$. Given the foregoing, a radial efficiency measure would be inconsistent with our theoretical model since it yields an efficiency concept that is negatively related with input usage and costs.

There are two potential approaches to estimating our proposed non-radial efficiency measure. The first approach is based on a sub-vector energy distance function[21] in which energy usage decreases for more efficient firms, such that efficiency improvements is costless: $\frac{\partial c_i}{\partial e_i} = 0$. In other words, the production technology is somewhat strongly disposable (i.e. the energy efficiency improvement does not require an increase in effort or other inputs). However, this assumption is inconsistent with our theoretical model and is likely at variance with the more realistic scenario that efficiency savings are costly[22]. A second but more appropriate approach is the hyperbolic distance function (HDF) which permits a weakly disposable production technology. In its traditional formulation, efficiency improvements are achieved by the simultaneous contraction in inputs and bad outputs (usually emissions). Consequently, we adapt the HDF framework by treating energy as an undesirable input, such that reductions in energy use must be accompanied by a proportional increase in efforts or other input cost, holding output constant[23].

3.4. Hyperbolic energy distance function

Consider a production technology in which capital (*K*), labor (*L*), and energy (*E*) are inputs employed to produce output (*Y*):

---

[21] See Filippini and Hunt (2015) for an excellent technical discussion on non-radial energy efficiency. Empirical applications of the sub-vector energy input distance function in the energy and environmental economics literature can be found in Zhou et al. (2012) and Boogen (2017).

[22] Mehdiloo and Podinovski (2019) highlight that the strong disposability assumption may not be suitable for modelling situations where some inputs and/or outputs are closely related. In this case, strong disposability assumption is likely to yield meaningless efficiency measures (*ibid*).

[23] We thank an anonymous referee for providing excellent guidance on this modified hyperbolic distance function, and for aiding our empirical research direction.

$$T = \{(K, L, E, Y) : (K, L, E) \text{ can produce } Y\} \quad [6]$$

The production technology *T* contains all the feasible input-output vectors and is assumed to be non-decreasing in the desirable output and non-increasing in the undesirable output (*E* in this case) and inputs.[24] Using the almost homogeneity property (Cuesta et al., 2009), we can express the traditional hyperbolic distance function (HDF) as:

$$D_H(\boldsymbol{x}, \theta y, \theta^{-1} b) = \theta D_H(\boldsymbol{x}, y, b), \theta > 0 \quad [7]$$

where $\boldsymbol{x}$ is the input vector, *y* denotes desirable output while *b* is the bad or undesirable output. The means that an increase in the desirable output *y* by a given proportion is accompanied by a contraction in the undesirable output *b* by the same proportion for a given input set, so that the HDF increases by that same proportion. However, because our aim is to estimate energy efficiency, we adapt the above HDF by treating energy as an undesirable input (i.e. the only input that requires further reductions). More succinctly, we can specify an energy saving hyperbolic distance function[25] as:

$$D_E(K, L, E, Y) = \sup\left\{\theta > 0 : \left(K\theta, L\theta, \frac{E}{\theta}, Y\right) \in T\right\} \quad [8]$$

The adapted HDF in [8] has a range $0 < D_E \leq 1$ and it embodies the feasible contraction of the energy input *E* and the simultaneous or equi-proportional increase in efforts or other inputs *K* and *L* that places the firm on the boundary of the technology represented by *T*. We note that, while the above specification violates the input distance axiom that inputs are non-increasing, it crucially permits the reality that energy efficiency savings are costly in terms of effort and non-energy input usage[26]. Therefore, this formulation is more appropriate for the empirical

[24] See Färe and Primont (1995) for some technical treatments of the axioms and properties of the technology.

[25] See Cuesta and Zofio (2005), Cuesta (2009), Zhang and Ye (2015) and Adenuga et al. (2019) for applications of the HDF in the empirical estimation of energy and environmental efficiency.

[26] The selective application of the weak disposability assumption has been suggested as a safe and suitable approach for modelling a set of closely related inputs and output (Mehdiloo and Podinovski, 2019).

estimation of our theoretical conception of energy efficiency improvement is Section 3. Using eqn. [3] and adopting the almost homogeneity property (i.e., $\theta = \frac{1}{E}$), we can obtain

$$D_E\left(K\theta, L\theta, \frac{E}{\theta}, Y\right) = \frac{1}{E} D_E(K, L, E, Y) \quad [9]$$

By taking logarithm of both sides of eqn. [9] in conjunction with the elements, $i = 1, \ldots, N; t = 1, \ldots, T$, we can write a panel data energy distance function as:

$$\ln\left(D_{E_{it}}/E_{it}\right) = TL(K^*, L^*Y, t)_{it} + v_{it} \quad [10]$$

where $TL(K^*, L^*Y, t)_{it}$ represents the technology as the translog[27] approximation to the log of the distance function; $K^*_{it} = K_{it} \times E_{it}$ and $L^*_{it} = L_{it} \times E_{it}$, while t is a time index capturing technical progress. v_{it} is the traditional symmetric error term representing sampling, specification and measurement errors. We can estimate the energy distance function using the stochastic frontier analysis (SFA) by Aigner et al. (1977) and Meeusen and Van den Broeck (1977):

$$-\ln E_{it} = TL(K^*, L^*Y, t)_{it} + (v_{it} - u_{it}) \quad [11]$$

where u_{it} represents the non-negative efficiency measure. Plant-specific energy efficiency in each period is then estimated as the conditional expectation of the one-sided error term, $\exp(u)$, given the composed error, $(v - u)$:

$$EE_{it} = E[exp(-u_{it})|\varepsilon_{it}] \quad [12]$$

$$\text{where } \varepsilon_{it} = (v_{it} - u_{it}) \quad [13]$$

A major advantage of SFA is that it eliminates random shocks in the estimation of energy efficiency: unlike other non-parametric approaches (e.g. data envelopment analysis

[27] We follow much of the literature (e.g. see studies in footnote 24 above) by using the Translog functional form because it is a more flexible approximation of the production technology (see Christensen et al., 1971, 1973).

(DEA)); and it is less susceptible to the vagaries of random events because it embodies a stochastic (traditional) error term to capture elements such as measurement and sampling error, as well as other disturbances relating to bad luck or policy shocks. An important issue in applied SFA studies is the possibility to model the specific parameters of the density function for u_{it} as functions of firm characteristics (i.e. conditional heteroscedasticity) using two broad approaches. Under the first approach proposed by Battese and Coelli (1995), u_{it} is assumed to follow the truncated normal distribution with a mean μ_{it} specific to each observation: $u_{it} \sim \mathcal{N}^+(\mu_{it}, \sigma_{it}^2)$, where $\mu_{it} = \boldsymbol{\varphi}'\boldsymbol{F}_{it}$ and $\boldsymbol{F}_{it}$ contains firm-specific factors. In the second approach (Caudill and Ford, 1993), firm-specific effects are introduced into the inefficiency terms by scaling its distribution, so that its variance is a function of $\boldsymbol{F}_{it}$: $u_{it} \sim \mathcal{N}^+(0, \sigma_{u_{it}}^2)$ where $\sigma_{u_{it}}^2 = \exp\,(\boldsymbol{\gamma}'\boldsymbol{F}_{it})$.

We favour the latter variance specification, because first, its scaling property is desirable as changes in $\boldsymbol{F}_{it}$ affect the scale but not the shape of the distribution of u_i, unlike the former approach where $\boldsymbol{F}_{it}$ enters the mean efficiency and alters the shape of its distribution.[28] Second, the scaling property provides an intuitive economic interpretation because u_{it} is treated as the base level efficiency (firms' natural capabilities), so that the extent to which these capabilities are exploited depends on firm-level conditions, captured by F_{it}. Finally, scaling functions (e.g. the exponential function) yield coefficients that are derivatives of the log of inefficiency w.r.t the exogenous variables, i.e., $\gamma = \partial \ln (u_{it})/\partial F_{it}$ for $u_{it} = \exp(F_{it}, \gamma) \cdot u_{it}{}^*$. This allows us to treat the coefficients on F_{it} as the quantitative effects of exogenous variables on inefficiency. This is not the case with the mean efficiency specification, which has no quantitative interpretation in terms of the magnitude of the parameters of F_{it}.

[28] See Alvarez et al. (2006) for a detailed technical discussion on the advantages of the scaling property.

4. Empirical tests

4.1 Econometric model

In principle, the primary objective of UK CCAs is to offer discounted energy or carbon tax liabilities in exchange for energy efficiency targets. Hence, assuming three firms in a sector have varying levels of efficiency: $e_{max} \geq e_i \geq e_{min}$, their respective tax discounts are:

$$D_{max} \geq D_i \geq D_{min} \quad [14]$$

This yields a discriminating incentive function:

$$D'(e) > 0; D''(e) > 0; D(0) = 0; D'(0) = 0 \quad [15]$$

We thus assume that the regulator aims for a regime where the tax discount is increasing in efficiency performance[29], and which also embodies a disincentive function, i.e. the marginal tax discount to an emitting plant is zero if efficiency improvement is zero. Using the information above, we formulate a model to investigate the relationship between the tax discount and energy efficiency:

$$D_{it} = \alpha_0 + \beta_1 D_{it-1} + \beta_2 Eff_{it-1} + \boldsymbol{x}'_{it}\boldsymbol{c} + \mu_i + \boldsymbol{T} + \varepsilon_{it} \quad [16]$$

where i =1,…, N is the plant identifier and t identifies the period of observation. D_{it} is the tax discount of plant i in period t obtained from the CCA scheme, and D_{it-1} is the tax discount of plant i measured in the previous period. This lagged dependent variable D_{it-1} ensures that we capture the dynamics arising from the temporal negotiations between the regulator and the various sector associations, which could theoretically lead to a long-run level of CCA tax discounts. Eff_{it-1} is the lag[30] of the energy efficiency values estimated from the input distance frontier described in Section 3, $\boldsymbol{x}'_{it}$ is a vector of firm-level characteristics such as firm size,

[29] We do not suggest that the program is designed to allocate CCA discounts to the most efficient firms. Rather, we motivate our empirical set-up with the plausible assumption that the environmental regulator aims to maximize social welfare. This assumption is consistent with our initial arguments that a rational regulator would, in principle, prefer to minimize social misallocation by allocating higher CCL discounts to more efficient firms.

[30] Using lagged values of energy efficiency fits with the program design that if a firm meets its efficiency target in one period, it continues to enjoy the CCA discount in the following period. This specification also mitigates problems arising from contemporaneous feedback effects.

age and ownership, μ_i represents unobserved plant-specific effects. while $\boldsymbol{T}$ is a vector of time dummies. ε_{ist} is a random error term that may vary across time periods, firms and sectors. β_1, β_2 and $\boldsymbol{c}$ are parameters to be estimated, but our primary focus is on the extent to which higher energy efficiency performance is rewarded under the CCA scheme, β_2.

μ_i captures unobserved plant managers' ability to achieve optimum output levels, which may be fixed over the timeframe where we observe the plants in our dataset. Because these managers' abilities directly influence productive energy efficiency[31], μ_i is potentially correlated with Eff_{it-1}. Moreover, although using Eff_{it-1} mitigates the problem of reverse causality, this may not fully alleviate the possible simultaneity bias in the empirical model, considering that the energy efficiency targets and the CCA discounts are jointly determined or negotiated simultaneously. Furthermore, the dynamic modelling approach can be complicated by issues such as correlation between D_{it-1} and ε_{it}, especially the firm-specific heterogeneity component[32] (Nickell 1981). Because D_{it} is a function of the unobserved firm-specific heterogeneity embodied in ε_{it}, it follows that D_{it-1}, one of the regressors, is correlated with ε_{it}.

A range of solutions are possible to address the above problems by exploiting the panel structure of our data set. For instance, we can define a set of instrumental variables to solve these problems by invoking the testable assumption that $cov(\varepsilon_{it}, \varepsilon_{it-1}) = 0, l = 1, \dots, T-1$. This approach restricts the random shocks to be uncorrelated over time, conditional on the lagged dependent variable, the plant-specific effect, energy efficiency and the vector $\boldsymbol{x}'_{it}$. Under this first solution, $\Delta D_{it-1}, l \geq 1$ can be employed as instrumental variables for D_{it-1} in eqn. [16] since $cov(D_{it-1}, \mu_i) > 0$ but $cov(\Delta D_{it-1}, \mu_i) = 0, l \geq 1$. Further, we note that

[31] Martin et al (2012) provide empirical evidence on the positive role of management practices and managerial abilities on energy intensity and productivity across a sample of UK manufacturing firms.
[32] This is often referred to as the Nickel bias.

$\Delta Eff_{it-2}, l \geq 1$ can also be used as instrumental variables for Eff_{it-1}. Although $cov(Eff_{it-1}, \varepsilon_{it}) \neq 0$ and $cov(Eff_{it-1}, \mu_i) \neq 0$, $cov(\Delta Eff_{it-2}, \varepsilon_{it}) = cov(\Delta Eff_{it-2}, \mu_i) = 0, l \geq 1$. Alternatively, we can address the endogeneity arising from the time-invariant plant-specific effects μ_i by applying first differences to eqn. [16]:

$$\Delta D_{it} = \alpha_0 + \beta_1 \Delta D_{it-1} + \beta_2 \Delta Eff_{it-1} + \Delta \boldsymbol{x}'_{it}\boldsymbol{c} + \Delta \boldsymbol{T} + \Delta \varepsilon_{it} \qquad [17]$$

While the differencing in eqn. [17] removes the firm-specific effects, the transformed error term is now correlated with the right hand side (RHS) variables, i.e., $\Delta\varepsilon_{it} = (\varepsilon_{it} - \varepsilon_{it-1})$ is correlated with $\Delta D_{it-1} = (D_{it-1} - D_{it-2})$ given that $cov(D_{it-1}, \varepsilon_{it-1}) \neq 0$. Hence, OLS is still inconsistent such that panel data (FE) estimators do not provide consistent estimates. However, we can obtain instrumental variables correlated with $\Delta D_{it-1} = (D_{it} - D_{it-1})$ but orthogonal to $\Delta\varepsilon_{it} = (\varepsilon_{it} - \varepsilon_{it-1})$. For instance, a range of possible IV candidates can arise from the moment conditions on the error terms. In this case, we can combine the levels equation [16] and the first differenced equation [17] using the Generalised Method of Moments (GMM) estimator (Arellano and Bond, 1991). By suppressing the other exogenous variables for now, we can obtain the instrument matrix for each plant as[33]:

$$\mathbf{Z}_i = \begin{bmatrix} Eff_{i2} & 0 & 0 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & Eff_{i2} & \cdots & Eff_{iT-3} \end{bmatrix} \qquad [18]$$

These moment conditions can be approximated by equation:

$$E(\mathbf{Z}_i \Delta \boldsymbol{\varepsilon}_i) = 0; \quad i = 1 \dots N \qquad [19]$$

where $\Delta\boldsymbol{\varepsilon}_i = (\Delta\varepsilon_{i4} \dots \Delta\varepsilon_{iT})'$. By applying a quadratic loss criterion with a weighting matrix that is inversely proportional to the variances of the moments, we can obtain the Arellano-Bond (AB) two-step difference GMM estimator (Arellano and Bond, 1991). In this estimator, the

[33] We can expand the instrument matrix by adding the exogenous variables.

$\Delta\hat{\boldsymbol{\varepsilon}}_i$ are consistent estimates of the first difference residuals from the preceding consistent estimator. The difference estimator employs lagged levels of the dependent variable as instruments for the first difference equation, but these may be weak instruments, especially in cases where the series are highly persistent. Thus, we can combine the equations in levels [16] and in first differences [17] to obtain a more efficient two-step estimator (Arellano and Bover, 1995; Blundell and Bond, 1998). Although this estimator offers greater efficiency than the alternative one-step estimator, it suffers from downward bias in its standard errors. Hence, we apply Windmeijer's (2005) finite sample correction to the standard errors.

5. Data and descriptive statistics

Our model estimations and analysis are based on a unique dataset, which we constructed from the most comprehensive restricted-use business microdata on UK manufacturing firms. Specifically, we gather information about the production technologies of 280 manufacturing firms by matching of data across three different confidential[34] surveys held by the Office for National Statistics (ONS): Quarterly Fuels Inquiry (QFI) SN: 6898, Quarterly Capital Expenditure Survey (QCES) SN: 6708 and the Annual Respondents Database (ARD) SN: 6644.

The QFI is a quarterly survey of over 1000 UK manufacturing plants, containing information on energy consumption, expenditures and CCL payments. The QCES is a quarterly survey that collects capital expenditure data by industry group across over 25,000 UK businesses. The QCES also contains core production function variables such as number of employees (our measure of labour, L) and total revenue which we employed as our measure of output Y. Our capital stock variable is constructed using the standard perpetual inventory method (PIM):

[34] Access was obtained through the UK Data Service secure access program.

$$K_{it} = (1 - \delta)K_{it-1} + I_{it} \quad [20]$$

Essentially, real capital stock in period t is assumed to be the depreciated[35] capital stock in period t-1 plus the investment in period t. The PIM approach is necessitated by the fact that the capital expenditure in the QCES is merely a measure of fixed capital formation or investment.

The energy input variable (E) and the CCA tax discount, D are obtained from QFI data. Energy consumption includes quantities consumed for different fuels (medium and heavy fuel oil, gas oil, liquefied petroleum gas (LPG), coal, natural gas, and electricity). Although our analysis is focused on the VEAs embodied in the CCA scheme, deriving the CCA discount variable requires that we take into account the intertwined nature of the CCL and CCA as two different components of a single policy package. While the CCL is a moderate tax on energy that increased the typical energy bill in the business sector by roughly 15% (NAO, 2007), the CCA is a VEA scheme for plants in certain industries that receive discounted CCL tax liabilities[36] but are also subject to energy use or efficiency target. Hence, we compute the discount variable as $D_{it} = E_{it} \times 0.8\ CCL_{it}$; where E_{it} is energy consumed, CCL_{it} is the carbon tax liability, and 0.8[37] reflects that plants could obtain maximum rebates of 80% on their tax liabilities during our period of analysis. We refer to D_{it} and CCL_{it} as the *effective* levels of the CCA discount and CCL tax liabilities, respectively; given that CCL_{it} pertains to the final tax liability across a range of fuels with different tax rates[38]. This data property confers the much-needed requirement that CCL_{it} varies across plants and over time within our empirical analysis[39]. Figure 1 provides a data plot of the CCL and CCA variables, confirming the

---

[35] We follow much of the literature in estimating capital stock and use a constant depreciation rate of δ= 6%

[36] During our study period, plants could receive up to 80% discount on their CCL liability.

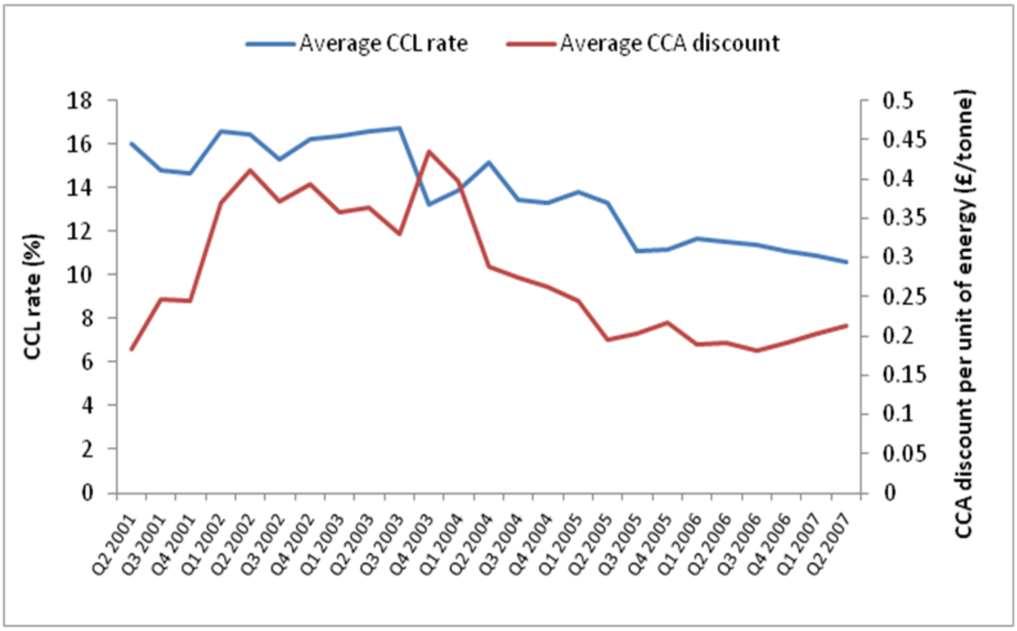
[37] This rebate rate varies across firms and is mostly less than the maximum 0.8.

[38] See Pearce (2006) and Martin et al. (2014) for useful discussions on CCL tax liabilities on energy and carbon content by fuel type.

[39] The CCL variable also embodies additional exogenous variation resulting from normalizing the tax liabilities to real terms using producer price indices.

variation in both variables for our data sample. This variation[40] is consistent with the UK government review of the CCL policy package conducted by the National Audit Office (NAO) during our study period. The NAO review revealed that the average CCL component of end-use electricity (gas) prices ranged from around 12% (17%) in 2001 to 7% (8%) in 2006 (see NAO, 2007, p. 19).

Figure 1: CCL and CCA discount plots



Our sample period covers the period from the introduction of the CCL (Q2:2001) to Q4: 2007. Our dataset ends in 2007 for two main reasons. First, roughly around 800 plants have consistent QFI data across all periods to 2007. That said, when the QFI data are matched with the QCES, our matching scheme yields a drastic fall in the number of plants with consistent data to around 280. Going beyond 2007 sees the number of plants with continuous data falling below 100. This is not surprising given the random sampling in the QCES, which means that we do not have QCES data for all QFI plants.

40 In addition to the visual test afforded by Fig 1, significance tests from our empirical analysis indicate that the level of variation in the CCL and CCA variables is not a problem in our empirical set-up, considering that insufficient variation will likely yield t-values that are too low to be significant.

Given that our final dataset contains information on 280 plants, a potential concern is that this sample may be inadequate to assess the impact of the CCA scheme. This may be further compounded by the high turnover of facilities within the CCA scheme, due in part to 'min-met' (mineralogical and metallurgical) exemptions.[41] That said, we believe that the data sample limitation reflects the difficulty or challenge arising from a severe lack of suitable microdata to evaluate energy and carbon policy programs. This situation is underscored by the dearth of microeconometric evaluation of the CCL/CCA package despite being the UK's single most important climate change policy. We therefore believe that our study still makes an important contribution to the small body of literature on the CCL-CCA package.

Further to this, our analysis covers the early days of the CCA scheme when the number of participants was limited due to the high fixed costs of participating in the CCA scheme[42]. These costs only incentivized large energy-intensive plants that could receive larger absolute discounts on their CCL liability to participate in a CCA (see Martin et al., 2014, p.4). Hence, our data sample is reasonable to analyse the CCA during our study period. Indeed, our sample size is broadly comparable to those in related studies (Martin et al., 2012, 2014). Furthermore, our study period (2001-07) avoids the serious data complications that may arise due to the high churn rates attributable to the 'min-met' exemptions which came into effect in 2014.

Table 1 provides descriptive statistics for the variables in our data sample, which contains 280 firms and 5458 observations. All monetary values are deflated by the manufactured output and input producer price indices, obtained from the ONS database. Firm size is an indicator variable such that firms with 250 employees or more in 2000 were defined as big. We include additional information to account for other observable heterogeneity across

---

[41] For further details, see https://www.envantage.co.uk/what-is-the-min-met-ccl-exemption-scheme.html

[42] As far as we know, the most comprehensive analysis of the CCA's participation to date was conducted by Martin et al. (2014) who merged both the DEFRA and HM Revenue and Customs (HMRC) list of CCA facilities during a period that overlaps with our study period (see their data appendix for detailed discussions). Their uptake calculations indicate that fewer than 2000 plants participated in the CCA during the earlier period.

sampled firms using the Interdepartmental Business Register (IDBR) available through the Annual Respondents Database (ARD) SN: 6644. We calculate firm age using information on year of establishment and we create an ownership dummy variable using the foreign ownership marker.

Table 1: Summary Statistics

Variables	**Mean**	**SD**
Output (£ ‘000)	116,435.80	240,006
Capital stock (£ ‘000)	71555.13	65217.48
Labor (employee headcount)	702.41	863.04
Energy consumption (tonnes)	28879.27	105609.50
CCA discount (£)	27947.77	134626.80
Size (Big=1 if employees>250, 0 otherwise)	0.785	0.411
Age (years)	23.75	7.52
Ownership (=1 if foreign owned, 0 otherwise)	0.549	0.498
Gas share of energy (%)	0.420	0.325
Electricity share of energy (%)	0.451	0.324
R&D (=1 if firm undertook R&D program in previous year)	0.508	0.496
Observations		5458

6. Empirical results

6.1 Hyperbolic distance function (HDF) estimates

Table 2 presents the maximum likelihood estimates (MLE) of the two non-radial approaches described in Section 3.3. In column 1, we estimate the sub-vector energy distance function while Column 2 presents results from our preferred HDF model specification. In both models, we control for firm fixed effects using the True Fixed Effect frontier model (TFE) (Greene, 2005a,b) and heteroscedasticity in u_{it}. Given our theoretical assumption on the contributions of innovation effort towards efficiency improvements, we model the variance of u_{it} as function of lagged R&D activities. The estimation data for the HDF are all log-normalised (i.e., mean-adjusted), allowing us to interpret the first-order coefficients of the translog model as output and input elasticities at the sample mean of the data.

Focusing on the input elasticities, we observe that the coefficients on the labour and capital inputs in the sub-vector energy distance function are negative, reflecting that energy efficiency savings do not necessarily require higher non-energy usage or costs. However, in the HDF estimates in Column 2, we obtain positive coefficients on both inputs, implying that any increase in the value of the labour and capital inputs would reduce distances to the frontier. This ensures the relaxation of the traditional axiom that the HDF is non-increasing in inputs, while crucially permitting our underlying theoretical assumption that energy efficiency improvements require reductions in energy use that are accompanied by proportional increases in both capital and labor[43]. Hence, the subsequent estimations in the second stage of our empirical analysis are based on the HDF.

In terms of the efficiency effect, we estimate negative coefficients on the lagged R&D variable across both model specifications, implying that R&D efforts reduces energy inefficiency. The average energy efficiency score[44] for our preferred model is around 81%, which indicates the feasibility of a 19% reduction in energy consumption by the typical firm, relative to the best practice production technology in the data sample. The distribution of this efficiency index is presented in Figure 2. In comparison to the efficiency scores from the sub-vector specification, the HDF efficiency score is slightly lower. This lower efficiency score is, in a way, expected since the HDF specification employs non-energy input cost expansion to attain efficiency savings; unlike the sub-vector distance function where efficiency savings are costless.

[43] The positive coefficients on both inputs can also be interpreted as positive derivatives of non-energy costs with respect to energy efficiency improvements, i.e. that energy efficiency improvements are costly.

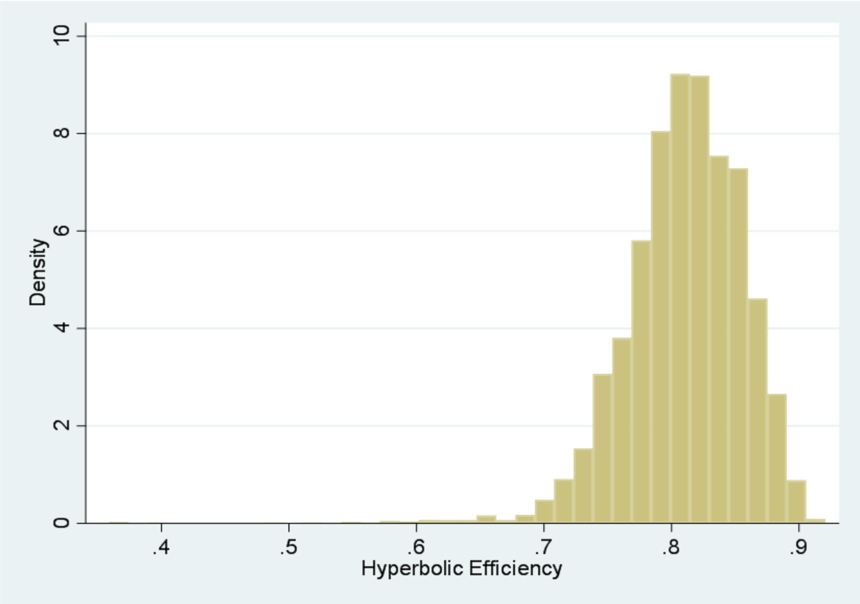
[44] We emphasize that the energy efficiency measure is obtained from inefficiency estimates that are measured in relative terms (i.e. in percentages) rather that in physical units, as in a pure standard additive directional function. Therefore, changes in energy efficiency must be interpreted in relation to the best-practice production activities across the sampled firms.

Table 2: Estimated input distance function parameters

Der var: $-\ln Energy_{it}$	**(1)**	**(2)**
Output	-0.654***	-0.518***
	[0.029]	[0.009]
Capital	-0.113***	0.026***
	[0.010]	[0.009]
Labour	-0.184***	1.008***
	[0.044]	[0.018]
Time	-0.008	-0.022
	[0.021]	[0.015]
Output-squared	-0.030***	-0.025***
	[0.002]	[0.001]
Capital-squared	0.009***	0.004***
	[0.001]	[0.000]
Labour-squared	0.129***	-0.007***
	[0.032]	[0.001]
Time-squared	0.031**	-0.014***
	[0.013]	[0.003]
Output*capital	-0.010	0.004*
	[0.007]	[0.002]
Output*labour	-0.133***	0.008**
	[0.021]	[0.003]
Capital*labour	-0.029***	-0.005***
	[0.010]	[0.001]
Output*time	-0.021	-0.030
	[0.013]	[0.032]
Capital*time	-0.007**	-0.006***
	[0.003]	[0.001]
Labour*time	0.015	0.011***
	[0.021]	[0.002]
Constant	-9.635***	0.583***
	[0.075]	[0.071]
$R\&D_{t-1}$	-1.936*	-1.898*
	[1.180]	[1.158]
Average energy efficiency level	83.02%	81.01%
Log likelihood function	9984.33	9940.21
Number of observations	5458	5458

Notes: This table contains estimations from the input distance function. Column 1 contains estimates from the sub-vector energy distance function, while column 2 presents the results from the hyperbolic distance function (HDF). Both models controls for firm effects and heteroscedasticity in the errors. ***, ** and * denote statistical significance at 1%, 5% and 10%, respectively.

Figure 2: Estimated energy efficiency index



6.2 Is the CCA discount increasing in energy efficiency? Baseline results

Based on the HDF efficiency scores[45], we estimate the model in [16] using pooled Ordinary Least Squares (OLS), Fixed Effects (FE) and System-GMM. The estimation results for the three models are reported in Table 3 for comparison purposes. Columns (1) and (2) are, respectively, pooled OLS and FE estimations of the discount model. Both show a negative relationship between the CCA discount and energy efficiency. In Column (3), we treat energy efficiency as endogenous, and we address the challenges arising from the dynamic panel model formulation using the System-GMM. For our analysis, we will focus on the system-GMM model results, as being more efficient and consistent than the OLS and FE results. The System-GMM results (Table 3, column 3) also indicate a negative relationship between the allotted tax discount and energy efficiency.

[45] For robustness checks, we re-estimate our second stage regressions using the efficiency scores from the sub-vector model specification. The results from these re-estimations are presented in the online appendix and the underlying findings from these alternative estimations are qualitatively similar to those obtained from the HDF specification. Therefore, we argue that our findings are robust to variations in the disposability properties of the estimated production technology.

Table 3: Energy efficiency and CCA discount

Der var.: *Discount*	**OLS**	**FE**	**GMM-SYS**
Discount$_{t-1}$	0.644***	0.204***	0.243***
	[0.018]	[0.020]	[0.031]
Energy efficiency$_{t-1}$	-0.106***	-0.102***	-0.543***
	[0.033]	[0.035]	[0.148]
Size	-0.198	0.485*	-1.755*
	[0.313]	[0.273]	[0.959]
Age	-0.903**	-0.902**	-1.185
	[0.393]	[0.381]	[0.933]
Foreign ownership	-0.629*	-0.764*	-1.907***
	[0.336]	[0.456]	[0.694]
Firm dummies	N	Y	Y
Quarter dummies	N	Y	Y
Year dummies	N	Y	Y
Wald/F test	0.000	0.000	0.000
AR1	-	-	0.000
AR2	-	-	0.144
Hansen p-value	-	-	0.101
No. of instruments	-	-	251
R-squared	0.439	0.479	-
Number of firms	280	280	280
Number of observations	5077	5077	5077

Notes: The dependent variable "Discount" is the log of the CCA discount amount (in £). Energy efficiency is computed from the HDF in column 2 of Table 2. Robust standard errors in parentheses are clustered at the industry level. ***, ** and * denote statistical significance at 1%, 5% and 10%, respectively. For the GMM model, we use the Windmeijer (2005)-corrected standard errors. Wald/F test: *p*-value from test of joint significance of parameter estimates. Arellano-Bond tests for first- and second-order serial correlation in the first-differenced residuals, asymptotically distributed N(0,1) under the null of no serial correlation. Hansen test of the over-identifying restrictions is asymptotically chi sq. distributed under the null that the instruments are orthogonal to the errors.

The AR parameter on the lagged discount variable is 0.243 and it lies between the OLS and the FE estimates of 0.644 and 0.204, respectively; indicating reasonable degree of state dependence and stable convergence towards equilibrium discount values. As Bond (2002) notes, the bounds provided by the OLS and FE estimates of the lagged dependent variable represent a useful check on results from the proposed GMM estimator. Specifically, the OLS model yields a naïve estimate of the lagged discount variable that is upward biased, since the lagged discount variable is positively correlated with the error term, whereas the opposite is the case with the FE estimates. Hence, an informal test of the credibility of the true AR parameter, as theoretically assumed by the SYS-GMM estimator is that it should between the OLS and FE estimates (Roodman, 2009).

The SYS-GMM estimator consists of a system of two simultaneous equations where lagged first differences are used as instruments in the levels equation; and where lagged levels are used as instruments in a first differenced equation. To test the validity of these instruments, we apply a test of over-identifying restrictions under the null hypothesis that there is no correlation between instruments and errors, i.e. that they are properly excluded. Further, we also apply the Arellano-Bond (AR) test under the null hypothesis of no second-order autocorrelation. Using the *p*-values on both specification tests, we fail to reject the overidentifying restrictions, suggesting that instruments are valid, while the Arellano-Bond test confirms that there is no second-order serial correlation in the first differenced error-term.

The coefficient on lagged energy efficiency can be interpreted as the social return on the CCA discount. The point estimate of the SYS-GMM regression is -0.543 which is both economically and statistically significant at the 1%- level. Thus a 10% improvement in energy efficiency is rewarded with approximately a 5.43% reduction in CCA discounts. The benefits of controlling for endogeneity bias is evident from the sizable differences between the GMM estimates and the OLS estimates (-0.106 and -0.102) which are upward biased. Economically, if the CCA efficiency targets are stringent enough or binding on plant energy use, the allotted tax discount ought to be positively correlated with plant energy efficiency. The empirical results in Table 3 indicate that less efficient plants received disproportionately higher tax discounts under the CCA scheme.

Surprisingly, we also find a negative relationship between the CCA discount and plant size, although the estimated coefficient on plant size is only significant at the 10% level. Similarly, we find that older plants received lower discount amounts, albeit this effect is not statistically significant in the GMM model. These results on plant size and plant age are encouraging, as it suggests larger and older plants, likely to have energy intensive production technologies, received lower discounts under the CCA scheme during the study period. In terms of plant ownership, we find that foreign owned plants received significantly lower discounts under the CCA scheme. The GMM coefficient (-1.907) on the foreign dummy variable suggests that foreign firms received around 191% less discount than their British counterparts.

6.3 Robustness test: Exploring an alternative measure of energy efficiency performance

So far, our analysis has relied on productive energy efficiency from an input-based frontier, where we radially contract energy consumption (and other inputs) for a given level of output. However, it is possible that our results are inferred incorrectly because the efficiency measure from the distance function does not properly capture energy efficiency. Therefore, we check the robustness of our results using an alternative measure of energy efficiency performance. We re-estimate the discount model using an output-based measure of energy efficiency: the ratio of output to energy use; see Table 4.

Table 4: Energy productivity and CCA discount

Der var.: *Discount*	**OLS**	**FE**	**GMM-SYS**
Discount$_{t-1}$	0.625***	0.222***	0.272***
	[0.027]	[0.022]	[0.027]
Output per energy unit $_{t-1}$	-0.419***	-0.320**	-0.704***
	[0.150]	[0.135]	[0.180]
Size	0.473	0.977***	0.972
	[0.295]	[0.232]	[0.639]
Age	-0.869**	-1.133***	-1.174
	[0.388]	[0.396]	[0.765]
Foreign ownership	-0.549*	-0.674*	-1.012*
	[0.314]	[0.404]	[0.577]
Firm dummies	N	Y	Y
Quarter dummies	N	Y	Y
Year dummies	N	Y	Y
Wald/F test	0.000	0.000	0.000
AR1	-	-	0.000
AR2	-	-	0.101
Hansen p-value	-	-	0.245
No. of instruments	-	-	251
R-squared	0.445	0.473	-
Number of firms	280	280	280
Number of observations	5077	5077	5077

Notes: The dependent variable "Discount" is the log of the CCA discount amount (in £). Energy productivity is computed as the ratio of output to energy consumption, a proxy for energy productivity. Robust standard errors in parentheses are clustered at the industry level. ***, ** and * denote statistical significance at 1%, 5% and 10%, respectively. For the GMM model, we use the Windmeijer (2005)-corrected standard errors. Wald/F test: *p*-value from test of joint significance of parameter estimates. Arellano-Bond tests for first- and second-order serial correlation in the first-differenced residuals, asymptotically distributed N(0,1) under the null of no serial correlation. Hansen test of the over-identifying restrictions is asymptotically chi sq. distributed under the null that the instruments are orthogonal to the errors.

The regression results presented in Table 4 show that the relationship between the discount and energy productivity remains negative and statistically significant, implying that plants with better energy productivity receive lower CCA discounts. In general, the coefficients on the control variables remain largely similar, except for the size variable where the sign on the coefficient turns positive; albeit its coefficient from the GMM model is not statistically significant. Overall, these results indicate that less efficient plants received disproportionately higher tax discounts under the CCA scheme.

6.4 Robustness test: Accounting for energy intensity in CCA discount received

One could argue that the baseline result is only a simple approximation of the aggregate relationship between the discount variable and energy efficiency, given that our estimations have so far used total discount value received under the CCA program. Consequently, we employ an alternative dependent variable – the ratio of total carbon discount to energy consumed – in a re-estimated model. This allows us to check the robustness of our result accounting for the implicit discount incentive, while also controlling for the intensity of firm-specific production technologies. Table 5 presents the re-estimated regressions based on discount per unit of energy (DPE). The results are qualitatively consistent with those estimated in the baseline specification: the energy efficiency coefficient remains negative.

Table 5: Energy efficiency and CCA discount using discount per unit of energy

Der var.: *Discount per unit of Energy (DPE)*	**OLS**	**FE**	**GMM-SYS**
DPE_{t-1}	0.600***	0.201***	0.214***
	[0.023]	[0.019]	[0.027]
Energy Efficiency$_{t-1}$	-0.084***	-0.060*	-0.365***
	[0.029]	[0.031]	[0.119]
Size	-0.445*	0.112	-1.391*
	[0.267]	[0.236]	[0.848]
Age	-1.019***	-0.973***	-1.524*
	[0.389]	[0.344]	[0.806]
Foreign ownership	-0.558*	-0.698*	-1.668***
	[0.325]	[0.411]	[0.632]
Firm dummies	N	Y	Y
Quarter dummies	N	Y	Y
Year dummies	N	Y	Y
Wald/F test	0.000	0.000	0.000
AR1	-	-	0.000
AR2	-	-	0.297
Hansen p-value	-	-	0.127
No. of instruments	-	-	249
R-squared	0.380	0.419	-
Number of firms	280	280	280
Number of observations	5077	5077	5077

Notes: The dependent variable is log of Discount per unit of Energy (DPE). Energy efficiency is computed from the HDF in column 2 of Table 2. Robust standard errors in parentheses are clustered at the industry level. ***, ** and * denote statistical significance at 1%, 5% and 10%, respectively. We use the Windmeijer (2005)- corrected standard errors. Wald/F test: p-value from test of joint significance of parameter estimates. Arellano-Bond tests for first-and second-order serial correlation in the first-differenced residuals, asymptotically distributed N(0,1) under the null of no serial correlation. Hansen test of the over-identifying restrictions is asymptotically chi sq. Distributed under the null that the instruments are orthogonal to the errors.

6.5 Robustness test: Accounting for variation in energy technology

There is considerable variation in tax rates across different fuels, upon which the CCA discounts are calculated. In particular, given the higher tax rates on gas and electricity, plants with gas or electricity-intensive technologies incur greater tax liabilities, which then shapes the size of the CCA discount. Hence, we re-estimate the baseline regressions using split samples based on “High” (“Low”) electricity and gas intensities, delineated using the pre-sample average in year 2000. Table 6 presents the results from split-sample fuel technology regressions and indicate that misallocation in the CCA discount scheme is common across both fuel technologies, supporting the conclusion that the negative relationship between the CCA discount and energy efficiency is robust to variation in fuel technologies.

Table 6: Discount regressions by fuel intensity: sub-samples (GMM estimates)

	Electricity intensity		Gas intensity	
Der var.: *Discount*	**High**	**Low**	**High**	**Low**
Discount$_{t-1}$	0.181***	0.322***	0.228***	0.206***
	[0.051]	[0.061]	[0.038]	[0.047]
Energy Efficiency$_{t-1}$	-0.366***	-0.438***	-0.387**	-0.306*
	[0.115]	[0.152]	[0.169]	[0.181]
Size	-0.243	-1.925**	-1.616	-0.419
	[1.082]	[0.966]	[1.256]	[1.292]
Age	-0.901	-1.774*	-0.801	-1.947*
	[1.121]	[0.944]	[1.449]	[1.007]
Foreign ownership	-2.107**	-1.361**	-1.552	-2.032***
	[0.978]	[0.682]	[1.058]	[0.767]
Firm dummies	Y	Y	Y	Y
Quarter dummies	Y	Y	Y	Y
Year dummies	Y	Y	Y	Y
Wald/F test	0.000	0.000	0.000	0.000
AR1	0.000	0.000	0.000	0.000
AR2	0.068	0.795	0.396	0.066
Hansen p-value	0.081	0.355	0.229	0.074
No. of instruments	177	177	177	177
Number of firms	204	233	208	218
Number of observations	2231	2846	2739	2338

Notes: The dependent variable "Discount" is the log of the CCA discount amount (in £). Energy efficiency is computed from the HDF in column 2 of Table 2. "High" and "Low" electricity intensities are based on the average pre-sample electricity share (45.1%) of total energy use such that high>45.1% and low<45.1%. For gas intensities these are High>42% and Low<42%

6.6 Evaluating efficiency savings: Reported CCA performance versus SEDF performance

In the final analysis, we collect efficiency savings information from the published CCA sector performance data[46] and conduct a comparative analysis of these energy efficiency savings versus our estimated input distance function (IDF) for three different target periods (TPs); see Table 7. We add a caveat that the estimated efficiency savings under the IDF pertain to information on a random sample of unidentified plants in the UK microdata, while the CCA performance data are based on a larger sample of identified participating plants.

Notice that the efficiency savings reported under the CCA performance data are at least twice as large as the efficiency savings estimated from the IDF. This finding is indicative of the trade-off between productive efficiency and information rent extraction in the CCA scheme, given that the energy efficiency of participating plants is not directly observable from the

[46] World and Scott (2011).

energy use and carbon emissions data reported for compliance monitoring. This is also in line with our theoretical analysis of the scheme's design, which raises the likelihood of a stringency problem.[47] CCA negotiations are based on reductions in energy use/carbon emissions, relative to a base year; and this is likely to be less stringent than negotiations driven by *potential reduction* based on firm production technologies.

7. Policy discussion and conclusion

The UK climate change agreements (CCAs) are voluntary agreements between regulators and sector associations to reduce energy use. Given the susceptibility of such programs to information asymmetry, this paper explores the trade-off between productive efficiency and information rent extraction with the CCA scheme. Using a panel of manufacturing plants drawn from the confidential UK business register during 2001-07, our empirical results indicate that plants with higher levels of energy efficiency received lower CCA discounts. We believe that, with more information about sector or firm-level efficiency savings potential, the UK environmental regulator is very likely to allocate the discounts in a way that better mirrors energy efficiency levels. We therefore conclude that this outcome is largely a result of asymmetric information. There is discussion in the literature (see Ekins and Etheridge, 2006; Barker et al., 2007; Martin et al. 2014) around the role of the stringency of this policy and our finding stresses its importance.

However, given data limitations, our results may be sensitive to uncertainty and should not be interpreted as an unambiguous call for increasing the stringency of the policy as this would require perfect information. In this case, a perfectly informed government will choose a combination of tax discount and reduction targets to induce at least as much abatement as under the full tax rate (Smith and Swierzbinski, 2007). However, in the real world, there is

[47] This stringency problem appears to be supported by the data plots in Fig. 2, which indicate that the average CCL tax rate declined over time, thereby weakening the program's stringency.

information asymmetry (see Montero, 2000) while the regulator may also choose to maintain international competitiveness. Hence, issues such as less efficient firms securing higher discounts, as suggested by our analysis (as in Montero, 1999, who examined the U.S. acid rain program) may arise. Nevertheless, an important implication from this study is that it identifies the trade-offs between stringency and potential energy efficiency, which are embodied within the CCA scheme.

Given current efforts to evaluate the effectiveness of the CCA scheme, there is a strong possibility that the UK government will seek to increase the stringency of environmental regulations. Our reasoning derives from the recent announcement by the government to end UK's contributions to global warming by 2050.[48] In this context, a practical suggestion arising from this study is the possibility for the regulator to conduct its negotiations based on sectoral efficiency benchmarks obtained from production frontiers, rather than assumptions about sectoral growth rates and technological evolution. This benchmark approach has been successfully employed by regulators in other UK sectors[49]. While we concede that a common problem in regulatory benchmarking is that participating firms may be reluctant to share information about strategic variables, this problem can be mitigated by the increasing availability of business microdata.

[48] https://www.gov.uk/government/news/pm-theresa-may-we-will-end-uk-contribution-to-climate-change-by-2050
[49] See for instance: (i) https://www.ofgem.gov.uk/sites/default/files/docs/2003/09/4720-background_cepa_report_and_efficiency_dpcr300903.pdf
(ii) https://www.ofcom.org.uk/__data/assets/pdf_file/0019/69400/benchmarking-report.pdf
(iii) https://www.ofwat.gov.uk/wp-content/uploads/2018/03/Testing-the-use-of-Ofwats-cost-benchmarking-models-in-merger-analysis-Final-report.pdf

Table 7: Comparison of efficiency savings: IDF vs. reported CCA savings (%)

Sector	No of firms	TP1			TP2			TP3		
		HDF	CCA	DIFF	HDF	CCA	DIFF	HDF	CCA	DIFF
Basic metals	26	1.80	31	-29.2	2.35	24	-21.65	6.95	23	-16.05
Chemicals and chemical products	32	2.27	15	-12.73	2.29	20	-17.71	4.59	20	-15.41
Fabricated metal products	22	1.97	10	-8.03	2.94	18	-15.06	4.08	16	-11.92
Food and beverage	36	2.26	10	-7.74	2.19	10	-7.81	7.25	14	-6.75
Machinery and equipment NEC/motor machinery man	31	3.05	17	-13.95	2.61	30	-27.39	2.60	41	-38.4
Other non-metallic mineral products	20	1.04	14	-12.96	2.39	16	-13.61	4.51	19	-14.49
Pulp paper and paper products	21	1.26	32	-30.74	2.17	35	-32.83	9.73	38	-28.27
Rubber and plastic products	22	1.25	15	-13.75	1.82	31	-29.18	7.16	34	-26.84
Textiles	23	0.86	8	-7.14	3.93	19	-15.07	5.26	16	-10.74

Notes: This table contains a comparison of estimated HDF efficiency savings versus the reported efficiency savings by the UK regulator (World and Scott, 2011). The comparisons relate to three target periods (TP) namely TP1 (2001-02), TP2 (2003-04) and TP3 (2005-06)

References

Adenuga, A.H., Davis, J., Hutchinson, G., Donnellan, T. and Patton, M., 2019. Environmental Efficiency and Pollution Costs of Nitrogen Surplus in Dairy Farms: A Parametric Hyperbolic Technology Distance Function Approach. *Environmental and Resource Economics*, 74(3), pp.1273-1298.

AEAT, 2004. Climate change agreements — results of the first target period assessment. Technical Report. AEA Technology, Didcot, UK.

AEAT, 2005. Climate change agreements — results of the second target period assessment. Technical Report. AEA Technology, Didcot, UK.

AEAT, 2007. Climate change agreements — results of the third target period assessment. Technical Report. AEA Technology, Didcot, UK

Aigner, D., Lovell, C.K. and Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics, 6(1), pp.21-37.

Alvarez, A., Amsler, C., Orea, L. and Schmidt, P., 2006. Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. *Journal of Productivity Analysis*, 25(3), pp.201-212.

Arellano, M. and Bover, O., 1995. Another look at the instrumental variable estimation of error-components models. *Journal of econometrics*, 68(1), pp.29-51.

Arguedas, C. and van Soest, D.P., 2009. On reducing the windfall profits in environmental subsidy programs. *Journal of Environmental Economics and Management*, 58(2), pp. 192-205.

Barker, T., Ekins, P. and Foxon, T., 2007. Macroeconomic effects of efficiency policies for energy-intensive industries: the case of the UK Climate Change Agreements, 2000–2010. *Energy Economics*, *29*(4), pp.760-778.

Battese, G.E. and Coelli, T.J., 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical economics*, 20(2), pp.325-332.

Blundell, R. and Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), pp.115-143.

Bond, S., 2002. Dynamic panel data models: a guide to micro data methods and practice; *Portuguese Economic Journal*, 1, 141-162.

Boogen, N., 2017. Estimating the potential for electricity savings in households. Energy Economics, 63, pp.288-300.

Cambridge Econometrics, 2005.Modelling the initial effects of the Climate Change Levy.Technical Report. Cambridge Econometrics, Cambridge, UK.

Caudill, S.B. and Ford, J.M., 1993. Biases in frontier estimation due to heteroscedasticity. *Economics Letters*, 41(1), pp.17-20.

Christensen, L.J., D., and Lau, L.(1971)."Conjugate Duality and the Transcendantal Logarithmic Production Function,". *Econometrica*, 39(4-6), pp.255-6.

Christensen, L.R., Jorgenson, D.W. and Lau, L.J., 1973. Transcendental logarithmic production frontiers. *The Review of Economics and Statistics*, pp.28-45.

Cuesta, R.A., Lovell, C.K. and Zofío, J.L., 2009. Environmental efficiency measurement with translog distance functions: A parametric approach. *Ecological Economics*, 68(8-9), pp.2232-2242.

Cuesta, R.A. and Zofío, J.L., 2005. Hyperbolic efficiency and parametric distance functions: with application to Spanish savings banks. *Journal of Productivity Analysis*, 24(1), pp.31-48.

De Muizon, G. and Glachant, M., 2004. The UK Climate Change Levy Agreements: Combining negotiated agreements with tax and emission trading. *Voluntary Approaches to Climate Protection: An Economic Assessment of Private-Public Partnership, Edward Elgar, Cheltenham (UK)*, pp.231-248.

Department for Business, Energy and Industrial Strategy (BEIS) 2019. The UK's draft Integrated National Energy and Climate Plan (NECP)

Department of Energy and Climate Change., 2016. *Quarterly Fuels Inquiry*, 1998-2015: Secure Access. [data collection]. 3rd Edition. UK Data Service. SN: 6898, http://dx.doi.org/10.5255/UKDA-SN-6898-3.

Ekins, P. and Etheridge, B., 2006. The environmental and economic impacts of the UK climate change agreements. *Energy Policy*, 34(15), pp.2071-2086.

Färe, S. and D. Primont. 1995. Multi-Output Production and Duality: Theory and Application. Kluwer Academic Publishers, Dordrecht.

Fare, R., Grosskopf, S., Lovell, C.K., 1994. Production Frontiers. Cambridge University Press

Farrell, M.J., 1957. The measurement of productive efficiency. *Journal of the Royal Statistical Society*. Series A (General), 120(3), pp. 253-290.

Filippini, M. and Hunt, L.C., 2015. Measurement of energy efficiency based on economic foundations. *Energy Economics*, 52, pp.S5-S16.

Greene, W., 2005a. Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model, *Journal of Econometrics*, 126, 269-303

Greene, W. H., 2005b. Fixed and random effects in stochastic frontier models, *Journal of Productivity Analysis*, 23(1), 7-32.

Hawdon, D., Hunt, L.C., Levine, P. and Rickman, N., 2007. Optimal sliding scale regulation: an application to regional electricity distribution in England and Wales. *Oxford Economic Papers*, 59(3), pp.458-485.

HM Government, 2006.Climate change. The UK Programme 2006 (London, UK).

HMRC, 2018. Estimated costs of principal tax reliefs. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/737597/Dec_17_Main_Reliefs_Final.pdf

Laffont, J.J. and Tirole, J., 1993. *A theory of incentives in procurement and regulation*. MIT press.

Leibenstein, H., 1977. X-efficiency, technical efficiency, and incomplete information use: A comment. *Economic Development and Cultural Change*, *25*(2), pp.311-316.

Martin, R., Muûls, M., de Preux, L.B. and Wagner, U.J., 2012. Anatomy of a paradox: Management practices, organizational structure and energy efficiency. *Journal of Environmental Economics and Management*, 63(2), pp.208-223.

Martin, R., de Preux, L.B., and Wagner, U.J., 2014. The impact of a carbon tax on manufacturing: Evidence from microdata. *Journal of Public Economics*, 117, pp. 1-14.

McEldowney, J. and Salter, D., 2016. Environmental taxation in the UK: the Climate Change Levy and policy making. *Denning LJ*, 28, p.37.

Meeusen, W. and van Den Broeck, J., 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review, 18 (2), pp.435-444.

Mehdiloo, M. and Podinovski, V.V., 2019. Selective strong and weak disposability in efficiency analysis. European Journal of Operational Research, 276(3), pp.1154-1169.

Montero, J.P., 1999. Voluntary compliance with market-based environmental policy: Evidence from the US Acid Rain Program. *Journal of Political Economy*, 107(5), pp. 998-1033.

Montero, J.P., 2000. Optimal design of a phase-in emissions trading program. *Journal of Public Economics*, 75(2), pp. 273-291.

NAO, 2007. The Climate Change Levy and Climate Change Agreements. Technical Report. National Audit Office, London, UK. https://www.nao.org.uk/wp-content/uploads/2012/11/climate_change_review.pdf

Office for National Statistics, Annual Respondents Database, 1973-2008: Secure Data Service Access [computer file]. 2nd Edition. Colchester, Essex: UK Data Archive [distributor], June 2012. SN: 6644 , http://dx.doi.org/10.5255/UKDA-SN-6644-2

Office for National Statistics. (2016). Quarterly Capital Expenditure Survey, 1998-2014: Secure Access [data collection]. 3rd Edition. UK Data Service. SN:6708, http://doi.org/10.5255/UKDA-SN-6708-3

Pearce, D., 2006. The political economy of an energy tax: The United Kingdom's Climate Change Levy. *Energy Economics*, *28*(2), pp.149-158.

Poi, B. and Wiggins, V., 2001. Testing for panel-level Heteroskedasticity and Autocorrelation. *StataCorp LP*.

Roodman, D., 2009. How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal*, *9*(1), pp.86-136.

Windmeijer, F., 2005. A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, *126*(1), pp.25-51.

World, J. and P. Scott, 2011. Climate Change Agreements: Results of the Fifth Target Period, Report AEAT/ED43694/R1, AEA Technology, Didcot, UK.

Zhang, Z. and Ye, J., 2015. Decomposition of environmental total factor productivity growth using hyperbolic distance functions: A panel data analysis for China. *Energy Economics*, 47, pp.87-97.

Zhou, P., Ang, B.W. and Wang, H., 2012. Energy and CO2 emission performance in electricity generation: a non-radial directional distance function approach. *European Journal of Operational Research*, 221(3), pp.625-635.