



Do energy efficiency policies save energy? A new approach based on energy policy indicators (in the EU Member States)

Paolo Bertoldi ^{a,*}, Rocco Mosconi ^b

^a European Commission, Joint Research Centre, Italy

^b Politecnico di Milano, Italy

ARTICLE INFO

JEL classification:

Q4
Q48
C23

Keywords:

Energy efficiency policies
Energy policy indicators
Dynamic panel model
Energy efficiency policy evaluation

ABSTRACT

Over the last two decades, the European Union and its Member States have introduced policies aimed at improving energy efficiency. The Energy Service Directives (ESD) introduced the concept of measurement of energy savings attributed to policies. Two different and complementary methodologies for the evaluation of energy savings have been developed under the ESD: the bottom-up (BU) approach, based on a technical analysis of each measure, and the top-down (TD) approach, based on the analysis of how energy intensity changes over time. BU methods can hardly take into account policy-induced behavioural changes, whereas TD methods have difficulties in disentangling policy-induced savings from other savings. Econometric models have been proposed as a viable alternative to deal with both drawbacks. The purpose of this article is to present an econometric model aimed at estimating the energy savings induced by energy efficiency policies in the EU Member States in the period 1990–2013. We introduce an explicit measure of Energy Policy Intensity based on the MURE database, which is used as explanatory variable in a dynamic panel model for 29 European countries. Our results suggest that energy consumption in 2013 in Europe would have been about 12% higher in the absence of energy efficiency policies.

1. Introduction

The recent IPCC Special Report on 1.5 °C has highlighted the role of end-use energy efficiency for climate mitigation in order to stabilise the temperature increase at 1.5 °C by the end of the century (IPCC, 2018; Bertoldi, 2018). In addition to climate change mitigation, energy efficiency provides an important contribution to security of energy supply and in increasing business competitiveness and citizen welfare (Fawcett and Killip, 2019; Mallaburn and Eyre, 2014). During the last three decades, the European Union (EU) and its Member States (MSs) have introduced policies to reduce energy demand and improve energy efficiency (Bertoldi, 2018). The EU has adopted in 2007 the EU Climate and

Energy targets for 2020.¹ The targets include a GHG emission reduction target of 20% compared to 1990, a renewable energy target of 20% and an energy consumption reduction target of 20%² (Carvalho, 2012; Streimikiene et al., 2012). In 2014 the EU adopted Energy and Climate targets for 2030,³ as follows: 40% reduction in greenhouse gas emissions compared to 1990 levels; at least a 27%⁴ share of renewable energy consumption; and at least 27%⁵ improvement in energy efficiency (deLlano-Paz et al., 2016).

In recent years, the EU has intensified its efforts of improving for energy efficiency. The Energy Service Directives (ESD - 2006/32/EC) introduced the indicative end-use efficiency target of at least 9% by 2016 for EU MSs (Apajalahti et al., 2015; Streimikiene et al., 2012; Hull

* Corresponding author.

E-mail addresses: Paolo.Bertoldi@ec.europa.eu (P. Bertoldi), rocco.mosconi@polimi.it (R. Mosconi).

¹ Presidency Conclusions of the Brussels European Council (8/9 March 2007), available at https://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/93135.pdf.

² Compared to a Business as Usual scenario modelled with Primes in 2007 - <http://www.energyplan.eu/othertools/national/primes/>.

³ European Council (23 and 24 October 2014) Conclusions on 2030 Climate and Energy Policy Framework, available at https://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/145356.pdf.

⁴ On December 2018 the final agreement was reached among the EU institutions to set the renewable 2030 target at 32.0%.

⁵ On 19 June 2018 a political agreement was reached among the EU institutions for an energy efficiency target for 2030 at 32.5% with an upwards revision clause by 2023, as for the 2020 target it is an energy reduction target compared to the same BaU scenario.

et al., 2009; Thomas et al., 2012; Boonekamp, 2011). Under the ESD, MSs were obliged to prepare National Energy Efficiency Plans (NEEAPs) every three years, starting from 2008 (Ringel and Knodt, 2018; Bertoldi and Economidou, 2018). The NEEAPs provided an overview of energy efficiency activities in each MS, including descriptions of national energy efficiency measures and quantification of achieved and forecast energy savings (Bertoldi and Economidou, 2018; Hull et al., 2009).

To reinforce the progress in energy efficiency, the Energy Efficiency Directive (EED - 2012/27/EU) was adopted in December 2012 (Rosenow et al., 2017). The EED contains a set of binding measures such as: legal obligations to establish energy saving schemes in MSs (Malinauskaite et al., 2019; Fawcett et al., 2019), public sector to lead by example, energy audits (Nabitz and Hirzel, 2019), energy services (Bertoldi and Boza-Kiss, 2017), energy efficiency funds, efficient CHP, metering and billing information (Zangheri et al., 2019), consumer behavior, etc.

Other EU energy efficiency policies are: the Energy Performance of Buildings Directive (EPBD - 2010/31/EU),⁶ which is the main EU legislative instruments for reducing the energy consumption in new and existing buildings (Dascalaki et al., 2012; Burman et al., 2014); the Eco-design Directive (2009/125/EC) to improve efficiency in energy related products (e.g. domestic appliances) (Bundgaard et al., 2017; Hinchliffe and Akkerman, 2017); the Energy Labelling Regulation (2010/30/EU) for products (Bjerregaard and Møller, 2019) and the Regulations for the Reduction of CO₂ Emissions of vehicles (443/2009/EC and 510/2011/EC) (Thiel et al., 2016). It is important to note that the above EU policies are complemented by MSs policies as described in the NEEAPs (2008, 2011 and 2014) and in the MURE database (see www.measures-odyssee-mure.eu) These policies include: taxation, financial incentives, regulation, voluntary programmes for industry, information campaigns, etc. (Bertoldi and Economidou, 2018).

The purpose of this article is to estimate with an econometric model the effect of energy efficiency policies on energy consumption in the EU MSs and Norway in the period 1990–2013. The aim of the model is to answer two research questions:

1. Are EU and national energy efficiency policies effective in reducing aggregate energy consumption? Can we estimate the policy-induced energy savings in each year from 1990 to 2013, measured as a percentage of the energy consumption as it would have been in the absence of energy policies?
2. Are sector-specific (household, services, industry, transport) energy efficiency policies effective in reducing sector's energy consumption? Can we measure effectiveness of energy policies in reducing consumption of energy in each sector?

The article's structure is the following: Section 2 reviews the existing literature on measurement methods of energy savings resulting from energy efficiency policies. Section 3 illustrates the methodology for constructing the Energy Policy Indicators. Section 4 outlines the structure of the econometric models. Section 5 illustrates the estimates, the simulation methodology used to isolate the effect of energy policy from the contribution of the other determinants of energy demand and reports the results of the simulations. Conclusions are drawn in Section 6, including some directions for further research. An Appendix provides details on the database created for this article.

2. Measuring energy savings induced by policies: a literature review

A key issue in energy efficiency policy analysis is the measurement of energy savings attributed to policies, in order to evaluate the impact of policies (Boonekamp, 2006). The ESD introduced the obligation for MSs

to measure energy savings induced by policies (Hull et al., 2009). The European EMEES project identified two complementary methodologies for evaluating the savings: the bottom-up (BU) and the top-down (TD) approaches (Thomas et al., 2012; Reichl and Kollmann, 2010, 2011; Bukarica and Tomsic, 2017). The BU assesses the energy savings in each individual project covered by the policy and then sums the individual savings (Boonekamp, 2006, 2011; Thomas et al., 2012). BU methods do not adequately capture behavioural changes, which may increase or decrease the calculated energy savings (Reichl and Kollmann, 2011). TD methods use an aggregate measure of energy consumption, normalized by an exogenous variable that adjusts for scale across cross-section observations (e.g. kWh/m²), usually derived from national statistical data (Boonekamp, 2011; Thomas et al., 2012). To calculate the energy savings, the aggregate measure is multiplied by the activity level (e.g. total floor area in m²) in different years. TD methods include all the policies covering the sector/equipment, the autonomous effects (e.g. technologies improvements not induced by specific policies) and structural effects (e.g. changes in activity) (Boonekamp, 2011). Therefore, TD methods capture all savings and corrections to calculate only the policy-induced savings are thus difficult (Horowitz, 2008).

Researchers have proposed the use of econometric models as an alternative to the BU and TD methods (Horowitz, 2011; Filippini and Hunt, 2012; Lang and Siler, 2013), to overcome the limitation of BU and TD methods. The objective of econometric models is to identify the energy savings induced by policies as compared to other factors such as economic growth, structural changes, populations, production levels, energy prices, etc. Filippini and Hunt (2012) estimated the efficiency of US residential energy consumption by using an econometric energy demand model with a stochastic frontier function. In another article, Filippini et al. (2014) assessed the level of energy efficiency of the EU residential sector against the potential for energy savings and estimated the impact of policies. Laes et al. (2018) reviewed the effectiveness of individual policies or policy packages for CO₂ emission reduction and/or energy savings by using a panel econometric model. Aydin and Brunen (2019) have assessed the impact of specific policies on electricity and non-electricity energy consumption by focusing on two types of regulatory measures: mandatory energy efficiency labels for household appliances and building standards.

A report by the European Commission JRC (Bertoldi and Hirl, 2013) has calculated the policy energy savings resorting to the counterfactual simulation approach proposed in Horowitz (2011). This is based on dividing the observed time span in a "pre-policy period", where policies are essentially absent, and a "policy period", characterized by the existence of relevant policies; the difference between actual energy demand and the forecasted energy demand is regarded as the (estimated) savings induced by policy (Horowitz, 2011).

Horowitz and Bertoldi (2015) introduced an explicit measure of energy policy as an explanatory variable in the econometric model. In order to evaluate policy effectiveness, the estimated model is analyzed through simulation techniques to isolate the contribution of energy policy from the impact of other determinants (prices, level of activity, technology, etc.).

In this paper we introduce an explicit measure of energy policy, based on the MURE database. The idea of introducing this type of variable in the econometric analysis of effectiveness of energy policies is not new (Bigano et al., 2011; Saussay et al., 2012; Filippini et al., 2014). In those papers, the adoption of policy measures is essentially measured through dummies. Building on those seminal contributions, Ó Broin et al. (2015) proposed a methodology to construct time series indexes, which increase as more policies are introduced and decrease as policies become obsolete.

The approach in this article shares, in line with the work done by Horowitz (2011), Bertoldi and Hirl (2013) and Horowitz and Bertoldi (2015), the idea of using a panel econometric model to evaluate policy effectiveness. While the energy policy indicator used in Horowitz and Bertoldi (2015) is based on a methodology for transforming the ODEX

⁶ Recently the EPBD was amended and entered into force on 9 July 2018.

bottom-up energy efficiency indicators (Cahill and Ó Gallachóir, 2010) into empirical estimates of policy impacts, in this paper we develop a direct indicator based on the MURE database of energy policy measures, as in Ó Broin et al. (2015).

3. The energy policy indicators

The purpose of this article is to analyze the effectiveness of energy policies by developing econometric models for energy demand where an indicator of energy policy intensity is introduced as a focal explanatory variable, along with control variables.

All the previous studies in this research line (Bigano et al., 2011; Saussay et al., 2012; Filippini et al., 2014, Ó Broin et al., 2015), are based on the MURE database. Ó Broin et al. (2015) discussed how their indicator relates to previous research: the differences of our indicator with respect to theirs are highlighted below. The MURE database classifies about 2000 energy policy measures adopted by European countries since 1970, reporting the year of adoption and some stylized characteristics (type of measure, expected impact, ...), along with a description of each measure. The database is structured by final energy consumption sectors (household, tertiary, industry, transport). In the MURE database, each measure is also classified according to a semi-quantitative impact indicator based on experts' judgement: low, medium or high impact (see below). In the next three subsections, we define several Energy Policy Indicators which will be used in our econometric models, and provide some descriptive statistics.

3.1. Base indicator (unweighted, permanent)

The basic version of our indicators simply counts the measures. These indicators are denoted by $pol_{h,i,t}^{U,P}$ where U stands for "unweighted", P stands for "permanent", and h indicates the "sector" ($h = 1, \dots, 4$: 1 = Household, 2 = Tertiary, 3 = Industry, 4 = Transport). Assuming that country i has adopted $N_{h,i}$ measures in sector h over the period $t = 1, \dots, T$, let us number them increasingly by adoption date by $n = 1, \dots, N_{h,i}$ and define.

- $a_{h,i}^n$: year of adoption of the n -th policy measure in sector h , country i
- $s_{h,i,t}^n$: dummy indicating if the n -th policy measure in sector h , country i started in year t (i.e. if $a_{h,i}^n = t$)

Our proposed energy policy indicator is then given by

$$pol_{h,i,t}^{U,P} = \sum_{\tau=1}^t \left(\sum_{n=1}^{N_{h,i}} s_{h,i,\tau}^n \right) \quad (1)$$

To understand the formula, notice that the first difference

$$\Delta pol_{h,i,t}^{U,P} = \sum_{n=1}^{N_{h,i}} s_{h,i,t}^n \quad (2)$$

is a simple measure of the energy policy effort in country i , year t , sector h , obtained by counting the measures implemented in that country, that year, that sector.

The policy indicator $pol_{h,i,t}^{U,P}$ simply cumulates the measures over time. It is important to remark that the implicit assumption underlying the idea of cumulating is that energy policy measures have a permanent effect, so that $pol_{h,i,t}^{U,P}$ is a non-decreasing function of time. This aspect of the indicator is better discussed below.

The methodology has been implemented for all 29 countries, using the energy policy measures listed in the MURE database, from 1975 to 2013. In order to provide some evidence about general tendencies in the

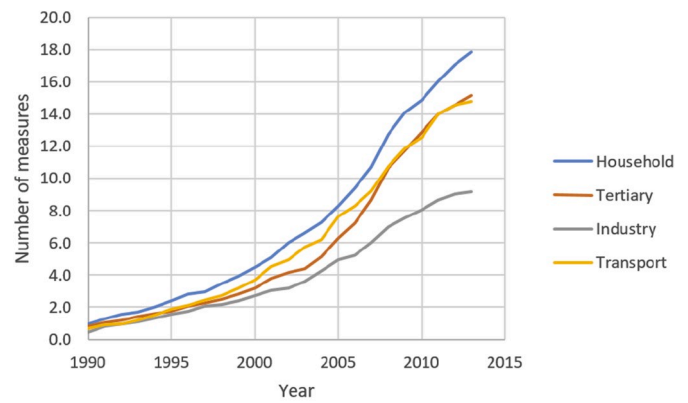


Fig. 1. Sectoral Energy Policy Indicators $pol_{h,i,t}^{U,P}$, average of 29 countries, 1990–2013.

EU, we illustrate below some descriptive statistics based on averages of $pol_{h,i,t}^{U,P}$ taken on all 29 countries. Fig. 1 illustrates the dynamics of the average across countries of the sectoral energy policy indicators in the period 1990–2013.

The household sector has the highest number of the energy policy measures: the final value means that on average, European countries have adopted 18 energy policy measures in the household sector between 1975 and 2013. Notice that the series equals about 1 in 1990 and 18 in 2013, meaning that the number of measures adopted (on average) between 1975 and 1990 has been about 1, while 17 measures have been adopted between 1990 and 2013: this is the reason why, in the econometric analysis carried on in Section 5, we will start from 1990.

In Table 1, we provide a summary of energy policy intensity in each country considering the entire period 1975–2013, given by the last value of $pol_{h,i,t}^{U,P}$. In practice, the numbers in the table correspond to the total number of measures in each country and each sector.

3.2. Weighting by policy relevance

In the base indicator measures are simply counted. This does not mean that we have assumed that all measures have the same effect, but rather that our purpose is to estimate the average effect, without focusing on individual measures. However, having some reliable evaluation of the relevance of each measure, we can create a weighted version of the Energy Policy Index, where such information is embodied. If the weighting system is correct, the weighted indicator should outperform the unweighted one in terms of explanatory power (i.e. goodness of fit) in the econometric model. Let us define the weights as follows.

- $w_{h,i}^n$: weight of the n -th policy measure in sector h , country i

The weighted policy indicator may be obtained as

$$pol_{h,i,t}^{W,P} = \sum_{\tau=1}^t \left(\sum_{n=1}^{N_{h,i}} w_{h,i}^n s_{h,i,\tau}^n \right) \quad (3)$$

Of course, when $w_{h,i}^n = 1$ for all n, h and i , the weighted indicator collapses into the unweighted one. If instead the weights are different, then different policies have a different contribution to the policy indicator. Ideally each policy might have its own weight, although in practice it is more reasonable to group policies into categories of equal

Table 1

Value of the Energy Policy Indicator in each sector and each country in 2013. ⁺ = high intensity (more than 1.5 times the average), ⁻ = low intensity (less than 0.5 times the average).

	Household	Tertiary	Industry	Transport
Austria	10	9	3 ⁻	11
Belgium	18	21	7	10
Bulgaria	22	16	15 ⁺	12
Croatia	16	17	5	27 ⁺
Cyprus	7	5 ⁻	4 ⁻	12
Czech Republic	15	6 ⁻	6	9
Denmark	11	3 ⁻	2 ⁻	8
Estonia	19	21	11	13
Finland	23	29 ⁺	15 ⁺	22
France	47 ⁺	26 ⁺	18 ⁺	29 ⁺
Germany	27 ⁺	25 ⁺	10	18
Greece	11	10	6	11
Hungary	16	5 ⁻	6	11
Ireland	25	23 ⁺	15 ⁺	24 ⁺
Italy	22	17	13	25 ⁺
Latvia	12	10	10	11
Lithuania	13	22	6	11
Luxembourg	12	4 ⁻	8	6 ⁻
Malta	26	16	5	5 ⁻
Netherlands	25	13	19 ⁺	16
Norway	23	22	19 ⁺	10
Poland	3 ⁻	7 ⁻	7	9
Portugal	15	12	3 ⁻	18
Romania	11	12	7	9
Slovakia	17	18	12	7 ⁻
Slovenia	13	11	7	9
Spain	32 ⁺	41 ⁺	15 ⁺	48 ⁺
Sweden	12	5 ⁻	6	14
United Kingdom	16	14	7	15
AVERAGE	17.9	15.2	9.2	14.8

weight according to their relevance.⁷

Although alternative weighting systems are possible, in this paper we will analyze the weighting scheme adopted by MURE. As mentioned above, MURE provides a semi-quantitative evaluation of the impact of each measure, based on quantitative evaluations or expert estimates; the following limits (in each case in % of the overall final energy consumption of a sector) are defined for the three impact levels: low = less than 0.1%, medium = 0.1%–0.5% and high = greater than 0.5% savings. We have therefore considered $w_i^l = 0.05\%$, or 0.0005, for all measures whose semi-quantitative impact is low (the same weight is given also to the measures whose semi-quantitative impact is unknown), $w_i^m = 0.3\%$, or 0.003, for all measures whose semi-quantitative impact is medium, $w_i^h = 0.7\%$, or 0.007, for all measures whose semi-quantitative impact is high. In practice, using this weighting scheme, the weighted indicator may be interpreted as the percentage decrease in energy consumption expected to be achieved by the policy measures, according to the MURE's impact evaluation, as compared to the energy consumption the sector would have experienced in the absence of policies. We leave for future analysis different alternatives, where, for example, a different weight is given to different “types” of measures, which are instead

⁷ Notice that giving all policies the same non-unitary weight (for example 0.5 instead of 1) would have the only effect to multiply the energy policy indicator by a constant, with no other consequence than dividing the corresponding coefficient by the same constant when the indicator is used within a linear econometric model. Therefore, it is the heterogeneity of weights, rather than their absolute scale, which makes a difference.

equally weighted here.⁸

It is important to remark that the weighting scheme may alter the “within” and “between” variability of the EP indicators substantially. As an illustration of this point, Fig. 2 compares the Energy Policy Indicators $pol_{h,i,t}^{U,P}$ and $pol_{h,i,t}^{W,P}$ for Italy and Germany for the household sector. The time series are illustrated from 1990 to 2013, even if they are computed starting from 1975: notice that in Germany energy policies started before 1990, that is why the German series do not start at zero. Using the unweighted indicator (left), where the measures are simply counted and cumulated over time, the number of measures adopted up to 2013 in the two countries is not so different (22 in Italy, 27 in Germany). Conversely based on the weighted indicator (right), where the MURE semi-quantitative impact indicator is used for weighting the measures, energy policy in the household sector in Germany in 2013 seems to be about three times more relevant than in Italy. The difference is due to the fact that the impact of most of the policies adopted in Italy before 2005 is classified by MURE as low, so that their weight is very close to zero, whereas the impact of most of the policies in Germany is classified as high. To clarify the metric on the vertical axis of the right plot of Fig. 2, it should be noted that using the “MURE inspired” weights, the weight is 0.05% for all measures whose semi-quantitative impact is low. Similarly, the weight is 0.3% for all measures whose semi-quantitative impact is medium, and 0.7% for all measures whose semi-quantitative impact is high. Cumulating the “weights” for all the 27 measures adopted in Germany in the household sector by 2013, one gets 11%, whereas all 22 measures adopted in Italy sum up to 3.5%.

In other words, trusting MURE's assessment of the impact of the policies, and assuming that policies have a permanent effect on energy consumption, one might argue based on the right plot of Fig. 2 that energy consumption in the household sector in Germany would have been 11% higher in 2013 in the absence of policies (3.5% higher in Italy). Our idea is not to use the indicator this way, but rather to introduce it as an explanatory variable in an econometric model where energy consumption is the dependent variable (see Sections 4 and 5). For example, finding a coefficient of zero would imply that energy policies are not effective, whereas a negative and significant coefficient, but different from -1 , would imply that energy policies are effective, but MURE under or over rate their impact on energy consumption.

3.3. Permanent versus transitory effects

As remarked above, in the $pol_{h,i,t}^{U,P}$ indicator, policy measures are assumed to have a permanent effect, since they are counted in every year since MURE's “Starting date”, even after the “Ending date”. This is one of the differences with respect to Ó Broin et al. (2015), where the measures are assumed to have transitory effect, since their “policy counter” decreases when the measure is discontinued. To compare the approaches, let us define.

⁸ The MURE database also classifies the measure by “type” (financial, normative, information, ...). In Bertoldi and Mosconi (2015), the energy policy indexes illustrated here are further disaggregated into five sub-indexes, measuring the intensity of each “type”, but we do not present this here since the econometric models illustrated in the following do not use these sub-indexes. Notice that the classification of policies by type is very analytic in the MURE database, therefore different aggregations of MURE “types” could be considered. For example, Ó Broin et al. (2015), building on Yearwood-Travezan et al. (2013), categorize the MURE types into three groups, according to their “degree of authoritative force” (in decreasing order of authoritative force: Regulatory, Financial, Informative). We refer to the MURE database for a more detailed illustration of the classification. We are currently exploring the usefulness of sub-indexes in an ongoing research project focused on the household sector.

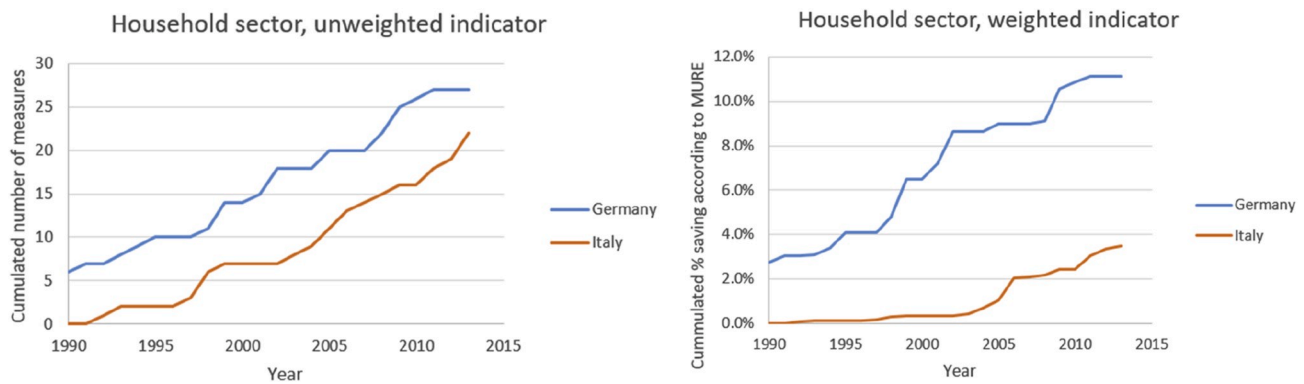


Fig. 2. Energy Policy Indicators in Germany and Italy, comparing $pol_{h,i,t}^{U,P}$ (left) and $pol_{h,i,t}^{W,P}$ (right).

- $d_{h,i}^n$: year in which the n -th policy measure in sector h , country i is discontinued ($d_{h,i}^n = 3000$ if the measure is in force)
- $e_{h,i,t}^n$: dummy indicating if the n -th policy measure in sector h , country i has been discontinued in year $t-1$ (i.e. if $d_{h,i}^n = t-1$): in practice this dummy indicates the first (complete) year in which the measure is not active

The “unweighted” “transitory” indicator $pol_{h,i,t}^{U,T}$ is given by

$$pol_{h,i,t}^{U,T} = \sum_{\tau=1}^t \left(\sum_{n=1}^{N_{h,i}} S_{h,i,\tau}^n - e_{h,i,\tau}^n \right) \quad (4)$$

In practice, $pol_{h,i,t}^{U,T}$ excludes the ended policies, and represents the count of the measures *in force* in year t : therefore it embodies the implicit assumption that when the measure is ceased, energy consumption goes suddenly back to the same level where it used to be before the measure started: the effect of energy policies is therefore assumed to be transitory, whereas $pol_{h,i,t}^{U,P}$ assumes is that the measure has a permanent effect even after it is discontinued.

Perhaps, neither of the extreme assumptions is always completely valid, and possibly different policies have a different level of persistence. For example, grants or subsidies for the installation of heat pumps might not last forever, but as they are discontinued some old and inefficient equipment has been replaced, so that energy consumption has been reduced. One might argue that also the new equipment will eventually become obsolete, and therefore in the absence of incentives consumers might switch back to the old and inefficient technology. However, this rebound might be partly or completely offset since, while incentives existed, the increased level of demand for efficient products might have reduced the price of the new efficient technologies, so that they remain competitive even when grants or subsidies are discontinued. Similarly, if an information campaign is successful in modifying consumer’s behavior, it is not obvious that people will return to previous habits when the campaign ends.

It is important to remark that different assumptions about the degree of persistence may alter the “within” and “between” variability of the EP indicators substantially. As an illustration of this point, Fig. 3 compares the Energy Policy Indicators $pol_{h,i,t}^{U,P}$ and $pol_{h,i,t}^{U,T}$ for Finland and the Netherlands for the household sector. According to MURE, 10 out of the 25 measures adopted in the Netherlands have been discontinued, mainly between 2000 and 2005, whereas almost all of the 23 measures adopted in Finland were still active in 2013. As a result, the ranking of the two countries in 2013 is different depending on which of the two version of the indicator is used.

In this article we will only analyze the two polar cases of infinite persistence and no persistence, to check if any of the two alternatives outperform the other in terms of goodness of fit in the econometric model: a detailed analysis of the issue is left for further research.

4. The econometric model: structure of the model and methodological aspects

Previous studies using explicit measures of energy policy intensity have been mainly focused on the household (residential) sector, see Bigano et al. (2011), Saussay et al. (2012), Filippini et al. (2014), Ó Broin et al. (2015). In this article, in order to get a more general picture of the effectiveness of energy policies on the whole economy, we consider a model with 4 equations, one per sector (Households, Services, Industry, Transport, labeled by $h = 1, \dots, 4$ respectively). Our approach allows us to measure the magnitude of policy-induced energy savings in each sector and hence, by aggregation, for the whole economy.

Below we illustrate the variables introduced in each equation, the mathematical structure of the econometric models, and some important methodological issues leading to the choice of the appropriate estimator.

4.1. The variables involved in the model

The *dependent variable* in each equation is (the natural logarithm of) $q_{h,i,t}^3$, i.e. the total energy consumption in sector h , country i , year t , measured in TJ, where the superscript 3 means that we consider the total consumption for 3 energy sources, namely electricity, oil and gas.⁹ Details of this variable are given in the appendix along with the data sources and construction methodology. As illustrated in the Appendix, the average coverage of $q_{h,i,t}^3$ in 1990–2013 is more than 80% in most countries, but in a few countries it is as low as 50% (especially in former planned economies and Nordic countries). Moreover, the quota of other sources has changed dramatically over the period considered in our study (usually decreasing). Of course, countries where the consumption of energy from other sources was initially large, and then decreased, have experienced a corresponding growth in consumption of oil, gas and electricity, not accounted for by other covariates. To overcome this problem, we have introduced in the model the control variable $other_{h,i,t}^3$, see below for a discussion.

As for the other explanatory variables, a short description is provided in Table 2, along with the expected sign in each sectoral equation. Details on the sources and the methodology are provided in the appendix.

The *focal explanatory variable* in each equation of the model is the sectoral Energy Policy Indicator, constructed along the lines illustrated in Section 3. We have considered alternative versions of the EP indicators in the model, as discussed in Subsections 3.1, 3.2 and 3.3. It is important to remark that the indicator introduced in the equation for each sector h is based exclusively on the measures that are relevant for the sector. Another key variable when the model is used to analyze the

⁹ We have also tried to analyze $q_{h,i,t}^4$, which includes also solid fuels: the results are qualitatively similar, but slightly worst.

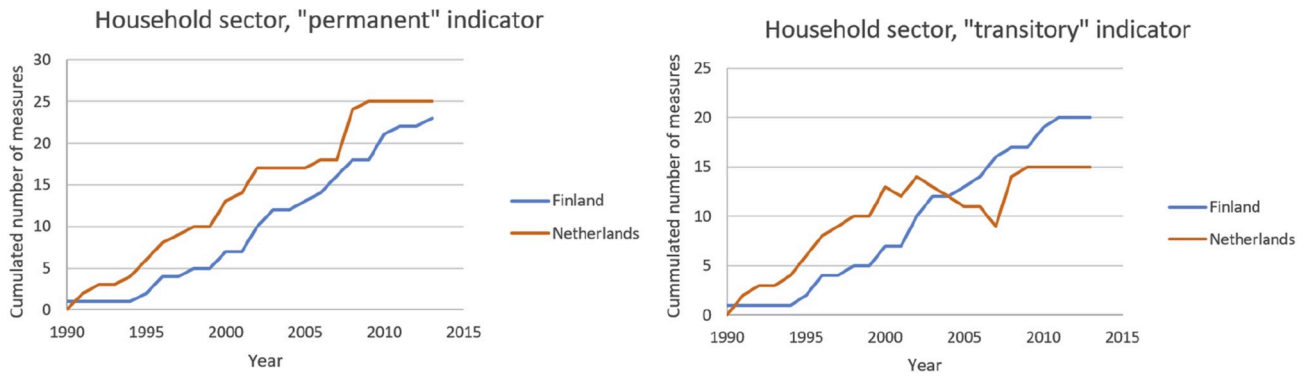


Fig. 3. Energy Policy Indicators in Finland and the Netherlands, comparing $pol_{h,i,t}^{U,P}$ (left) and $pol_{h,i,t}^{U,T}$ (right).

Table 2

Short description of the explanatory variables introduced in the model (more details in the Appendix).

	VARIABLE	SHORT DESCRIPTION	EXPECTED SIGN			
			Hou	Serv	Ind	Tra
Focal variables	$\ln(q_{h,i,t-1}^3)$	Lagged dependent variable (log)	+	+	+	+
	$pol_{h,i,t}^{U,P}$ or $orpo_{h,i,t}^{U,P}$	Sectoral Energy Policy Indicator	-	-	-	-
	$orpo_{h,i,t}^{U,T}$					
Control variables included in all equations	$\ln\left(\frac{p_{h0,i,t}^3}{def_{i,t}}\right)$	Relative price of energy for the sector (log)	-	-	-	-
	$other_{h,i,t}^3$	Share of total sectoral energy consumption from other sources (solid + derived heat + renewable energies + nuclear heat + non-renewable waste)	-	-	-	-
	$\ln(pop_{i,t})$	Population (log)	+	+	+	+
	$\ln(rgdp_{i,t})$	Real GDP (log)	+	+	+	+
	$hdd_{i,t}$	Heating degree days	+	+	+	?
	t	Quadratic trend, capturing omitted variables behaving as a smooth function of time (e.g. technology, culture, ...)	?	?	?	?
	t^2		?	?	?	?
Sector specific control variables	$\ln(area_s1_{i,t})$	Total floor area of dwellings in km^2 (log)	+			
	$perc_equip_s1_{i,t}$	Average of the rate of ownership (%) for freezers, washing machines and dishwashers	+			
	$\ln(rcons_s1_{i,t})$	Real private consumption in M€2005 (log)	+			
	$\ln(rva_s2_{i,t})$	Real value added of tertiary sector in M€2005 (log)		+		
	$\ln(empl_s2_{i,t})$	Employment of tertiary sector, thousand (log)		+		
	$\ln(rva_s3_{i,t})$	Real value added of industry in M€2005 (log)			+	
	$\ln(rginv_s3_{i,t})$	Real gross investment of industry in M€2005 (log)			?	
	$\ln(cars_s4_{i,t})$	Stock of cars, millions (log)				+
$\ln(goods_s4_{i,t})$	Traffic of goods, tkm (log)				+	

effect of energy policies is the lagged dependent variable, whose role and interpretation is discussed in Subsections 4.2, 4.3, 5.2 and 5.3.

The other variables introduced in the model play the role of control variables, essentially chosen among the classical variables introduced in energy demand studies. These variables are used in the model essentially to avoid omitted variables bias potentially arising from correlation with due focal variable, i.e. the energy policy indicator. Therefore, we are not so interested in estimating or interpreting their coefficients, although it is clearly reassuring that the sign and magnitude of the coefficients is in line with the theoretical expectation and the typical finding in other studies. For the residential sector, an extensive literature review and comprehensive list of the variables may be found for example in Tsemekidi Tzeiranaki et al. (2019): the typical variables are sectoral energy prices, socio-economic indicators (e.g., population, income, etc.), climate, and more specific sectoral variables such as dwelling characteristics and household attributes. For the industry sector, models are based on the production theory; a review of the econometric literature on energy demand with the list of explanatory variables may be found in Adeyemi and Hunt (2007) and Khayyat (2015). The key variables are the level of activity and the sectoral price of energy; most studies add some indicator of capital formation (investment) and the price of the other production factors (capital and labour). For the services sector, we

did not find many econometric studies on energy demand, although given the nature of the sector, one can get inspiration from the studies focusing on industry (since the economic theory of production is central to both sectors) and to some extent household (given the relevance of climate for both sectors). Finally, for the transport sector, a rather complete list of variables may be found for example in Gupta et al. (2019). The variables in this sector are sectoral energy prices, GDP, population, along with sector specific variables measuring the amount of transported passengers and freight. Linear or quadratic trend is usually introduced in all sectors to account for hardly measured variables which can be assumed to be smooth functions of time (technology, habits, ...).

4.2. Structure of the econometric model

The structure of the four equations is identical, namely:

$$\ln(q_{h,i,t}^3) = \beta_{0,i}^h + \rho^h \ln(q_{h,i,t-1}^3) + \gamma^h pol_{h,i,t} + \beta_1^h \ln\left(\frac{p_{h0,i,t}^3}{def_{i,t}}\right) + \beta_2^h other_{h,i,t}^3 + \beta_3^h \ln(pop_{i,t}) + \beta_4^h \ln(rgdp_{i,t}) + \beta_5^h \ln(hdd_{i,t}) + \beta_6^h t + \beta_7^h t^2 + \delta_h^h SSV_{h,i,t} + \epsilon_{i,t}^h \tag{5}$$

where $h = 1, \dots, 4$ indicates the sector. Each equation represents a dynamic panel model to be estimated on data for 29 countries ($i = 1, \dots, 29$) and 24 years ($t = 1, \dots, 24$, where 1 stands for 1990 and 24 stands for 2013).

The variables are introduced and discussed in the previous subsection (see Table 2). $SSV_{h,i,t}$ is a vector of sector specific variables. Specifically, as seen in Table 2, for sector 1 (household) SSV includes: $\ln(\text{area_s1}_{i,t})$, $\text{percequip_s1}_{i,t}$ and $\ln(\text{rcons_s1}_{i,t})$. For sector 2 (services) SSV includes: $\ln(\text{rva_s2}_{i,t})$ and $\ln(\text{empl_s2}_{i,t})$. For sector 3 (industry) SSV includes: $\ln(\text{rva_s3}_{i,t})$ and $\ln(\text{rginv_s3}_{i,t})$. Finally, for sector 4 (transport) SSV includes: $\ln(\text{cars_s4}_{i,t})$ and $\ln(\text{goods_s4}_{i,t})$.

Each equation has a country effect $\beta_{0,i}^h$, accounting for country specific unobservable time invariant characteristics. This country effect is not handled by using fixed or random effects, but rather by differencing, using the Arellano-Bond estimator (see below).

The dependent variable $q_{h,i,t}^3$ is log transformed to account for heteroscedasticity (clearly visible in preliminary estimates where the variable was not log transformed). We also introduced the lagged dependent variable $\ln(q_{h,i,t-1}^3)$ as a right hand side variable in the model. This is an important difference with respect to previous papers introducing energy policy indicators in panel data models for EU countries; in fact, most of the aforementioned literature has used static models, mainly the fixed effect model or frontier techniques. Preliminary analysis carried out with our data using static models has shown clear evidence of residuals autocorrelation, and this is indeed one of the motivations for considering dynamic models. However, the role of the lagged dependent variable is not only to get uncorrelated residuals. It also represents the idea that the adjustment of energy consumption to changes in the policies (and in the other variables) is not instantaneous, but takes time. Dynamic panel models have been proposed in Balestra and Nerlove (1966) in the framework of gas demand analysis. They developed a panel model for total gas demand, where the key equation is $G_{i,t} = G_{i,t}^* + (1 - r)G_{i,t-1}$. Here $G_{i,t}$ represents the current total demand in year t and country i , which is decomposed into new demand, $G_{i,t}^*$, originated by new appliances, plus old demand, which is a proportion of the previous period's demand, r being the depreciation rate for gas using appliances (r is therefore assumed to be a real number between zero and one, so that $\rho = 1 - r$ is also between zero and 1). New demand $G_{i,t}^*$ is then modelled as a linear function of suitable vector of covariates, $G_{i,t}^* = x_{i,t}'\beta + \varepsilon_{i,t}$: hence, the final model is $G_{i,t} = x_{i,t}'\beta + \rho G_{i,t-1} + \varepsilon_{i,t}$. The presence of the lagged dependent variable induces inertia in the adjustment process, so that a permanent change in the x 's takes time to transfer entirely on G , since only a proportion r of the appliances is replaced every year. A similar argument can be applied to buildings or, more generally, to habits. For more details and further references on the interpretation of the adjustment process in dynamic panel models see, for example, Greene (2012), Section 11.11.3, and Baltagi (2005), Chapter 8.¹⁰

In order to estimate the possible effect of energy policies on energy consumption using the model, the parameter of interest is γ^h . This parameter is expected to be negative. The parameter ρ^h also plays an important role, discussed in Subsection 5.2, where the adjustment process implied by (5) is discussed through a simulation exercise.

¹⁰ Ó Broin et al. (2015) also address the issue of the delay needed for a policy measure to induce a reduction in energy consumption. However, their approach is different, since they introduce a lag in the energy policy indicator (up to 7 years) in their econometric model, selecting the best fitting lag based on testing. Their approach implies that the policy has absolutely no effect before the selected lag, after which it suddenly becomes fully effective. Conversely, with our approach we assume a gradual effect, starting on the same year where the policy is implemented: our empirical analysis suggests that this is very effective in capturing the "slow adjustment" features.

4.3. Econometric issues and estimation technique

Evaluating the effect of policies on macro aggregates is problematic, (see for example Athey and Imbens, 2017), since the policy variable is potentially endogenous, i.e. correlated with the error term. In our framework, endogeneity might originate from omitted variables possibly affecting both energy policy and energy demand, and from reverse causality (policies might depend on past energy consumption). In our opinion simultaneity, which is another potential source of endogeneity bias, is less of a concern here. Using the right inferential tool is therefore crucial in order to avoid biased estimates, which would only identify correlations rather than causal effects. The paradigm of the statistical analysis of causal links, namely randomized controlled experiments, is not viable in macro settings (Cooper, 2018). Other tools like regression discontinuity design are also essentially designed for evaluating policies effectiveness using micro data. Conversely, the difference-in-differences method can be used in macroeconomic analysis (Giavazzi and Tabellini, 2005; Papaioannou and Siourounis, 2004; Persson, 2005; Persson and Tabellini, 2006; Rodrik and Wacziarg, 2005). This method has been used for assessing the impact of specific carbon and energy policies (Filippini and Zhang, 2019; Meng et al., 2017; Lin and Zhu, 2019). However, differences-in-differences assumes that the policy intervention may be represented as a binary variable (on/off), which is not the case here.

Therefore, the appropriate tool in our case is the Instrumental Variables (IV) Estimator, whose importance for causal inference is illustrated, for example, in Angrist and Pischke (2009), Section 4.1. Before introducing IV in our case, let us first remark that in general, and even more so in the macro settings, an appropriate and complete set of control variables is the best antidote against omitted variables bias. The panel setting is to some extent helpful in this sense, since unobserved heterogeneity potentially correlated with the policy variable can be dealt with either by fixed effects or by differencing: therefore, any problem originating from omitted time-fixed variables is neutralized. However, introducing appropriate time varying control variables is extremely important. In the present research, we have included all the variables which are usually introduced in the energy demand literature. We also added two control variables that are not so usual, namely the lagged dependent variable (i.e. energy consumption in focal sector in the previous year), and the share of other sources on total energy consumption. We believe that including these two variables is extremely important, since.

- (i) they might have an impact on energy policy, which can be expected to be more intense in a given year, everything else being fixed, if energy consumption in the previous year increased (due to reverse causality) and if the share of other sources increased (since our energy policy indicator is sector specific, but not source specific)
- (ii) they also might have an impact on the current consumption of energy: as most aggregate economic variables, energy consumption is likely to adjust slowly, and is therefore likely to be correlated with its own past; moreover, everything else being fixed consumption of a given group of energy sources is likely to be instantaneously negatively correlated to the share of other sources.

Excluding these variables from the analysis might therefore potentially bias the coefficient of energy policy.

Even when the set of control variables is generous, it is still important to test the null hypothesis (H_0) that the policy variable is exogenous (i.e. uncorrelated with the error term), versus the alternative (H_1) that it is endogenous (i.e. correlated with the error term). This can be done through the t version of the Hausman specification test (see Hausman, 1978; Wooldridge, 2002, p. 290). The test is based on the difference

between the two estimates, $\hat{q} = \hat{\gamma}_1^h - \hat{\gamma}_0^h$ where $\hat{\gamma}_0^h$ is obtained using the policy indicators as instruments for themselves, and is therefore efficient under H_0 but inconsistent under H_1 , while $\hat{\gamma}_1^h$ is obtained introducing suitable instrumental variables for the policy indicators, and therefore, provided that the instruments are appropriate, it is a consistent estimator under both H_0 and H_1 , although inefficient under H_0 . If no misspecification is present, the probability limit of \hat{q} is zero: hence, when the two estimates are similar, the evidence is in favour of H_0 , and the practical implication is that $\hat{\gamma}_0^h$ is consistent and efficient, and therefore its causal interpretation makes sense. Conversely, if the two estimators are too different, we reject H_0 , and the practical implication is that only $\hat{\gamma}_1^h$ is consistent and can therefore be given a causal interpretation. It is important to remark, that this procedure rests upon an “appropriate” choice of the instrumental variables (see Angrist and Pischke, 2009); instruments should be

- (i) correlated with the potentially endogenous regressor (i.e. the energy policy indicator), otherwise we say that the instruments are “weak”, and IV estimates will be very unstable
- (ii) uncorrelated with any omitted variable in the equation of interest (this assumption is sometimes called exclusion restriction or orthogonality condition), otherwise the IV estimates will be biased

Following Keane (2010), an appropriate set of instruments can be determined based on an economic interpretation of the determinants of energy policy. In fact, instruments may be selected among the determinants of energy policy (this would avoid weak instruments) which the economic theory would exclude from the focal equation (so that the exclusion restriction is fulfilled). Energy policies have multiple objectives, not just energy savings, see for example Haydt et al., (2014). Among the goals of energy policy: reducing dependence on imported energy, preserving natural resources and minimize environmental impacts (reducing CO2 emissions and possible climate changes), reducing dependence on non-renewable sources, diversifying sources to reduce dependence on suppliers, increasing national production of energy, improving efficiency. Based on data availability, we have therefore introduced as instruments an indicator of energy dependence and an indicator of greenhouse gas emission, both evaluated in the previous year, $ed_{i,t-1}$, and $ghge_{i,t-1}$ (see the appendix for details on these variables). One might argue that energy consumption in a given year might be correlated with energy dependence and greenhouse gas emissions in the previous year: however, it is reasonable to assume that such dependence vanishes once energy consumption in the previous year is controlled for, as we do.

Taking into account the previous argument, in this article, all equations have been estimated separately using the Arellano and Bond (1991) estimator for dynamic panel models, in short AB¹¹; AB estimator is specifically designed to deal with the econometric problems arising from the presence of the lagged dependent variable. In the panel data framework this estimation technique has several advantages over alternative candidates (see for example Wintoki et al., 2012):

- Since AB is based on differencing, it is not affected by bias due to omitted country effects. In other words, (time-fixed) omitted country characteristics influencing both energy policy and energy demand are not expected to create any bias. In fact, estimators based on differencing, exactly as the standard fixed effect estimator, makes

use of the within variability only, neglecting the between variability which is the source of bias in this case.

- If the model includes the lagged dependent variable to account for partial adjustment, which is our case, GMM based estimators using “internal” instruments (i.e. lagged differences of the dependent variable), such as AB, are preferable to OLS based estimators, such as fixed effects. In fact, fixed effects estimates of the lagged dependent variable parameter would be biased towards zero (even if the bias vanishes as T goes to infinity, it can still be substantial for T = 23 as in our case).¹²
- We strongly believe that estimator based on static models, such as CCEMG (Pesaran, 2006) or AMG (Eberhardt and Teal, 2010) should be avoided in this case: as clearly illustrated in Wooldridge (2002), Roodman (2008), Wintoki et al. (2012), when some of the regressors are potentially influenced by past values of the dependent variable (reverse causality), the lagged dependent variable also plays the role of control variable, and excluding it might severely bias the coefficient of the policy indicator. This is likely to be the case for the policy variable, which is the focal variable in our analysis.
- Although possibly omitted time-fixed determinants are not creating any problem, since they are dealt with by differencing, the standard version of AB has the problem of not being robust to omitted time-varying variables possibly correlated with the policy variables (the motivation which stands behind CCEMG and AMG). However, we remark that, on top of the “GMM-style” instruments used for the lagged dependent variable, Stata command xtabond2 allows to introduce ordinary “IV-style” instrumental variables, to avoid inconsistency of the parameter estimates for the variables suspected of endogeneity. It is therefore possible to perform the Hausman specification test as illustrated above.

5. Parameter estimates and evaluation of energy policy effectiveness

In this Section we illustrate the estimates of the parameters in model (5), and discuss how the estimated dynamic econometric model can be used to evaluate the effectiveness of energy policy in each Sector and to estimate the total energy saving induced by energy policy in each country and in Europe over the period 1990–2013.

5.1. Parameter estimates

The AB estimates of model (5), using the unweighted permanent Energy Policy Indicator $pol_{h,i,t}^{U,P}$ are summarized in Table 3.¹³ The table also reports the results of the usual misspecification tests: the robust m_2 statistics (see Arellano and Bond, 1991) supports the assumption of

¹² It is sometimes claimed that AB has been developed for large N and short T and not for moderate N and large T as the data set used in this paper, and that therefore AB would not be ideal in this setting. However, the only problem of AB with large T is that the number of instruments is quadratic in T, so that, when T is large, the default number of instruments might be extremely large, possibly overfitting the endogenous regressors. However, Roodman (2009) xtabond2 allows for the specification of the particular lags to be included in estimation, rather than relying on the default strategy to include them all: therefore, large T is not a problem in our case.

¹³ We have used the “difference GMM” estimator, introducing all available lagged differences of $\Delta \ln(q_{h,i,t}^3)$ as instruments for the lagged dependent variable (the results actually do not change if we use just a few lags as instruments). All other regressors have been treated as exogenous, an assumption which seems to be reasonable in the light of the results of the “Difference-in-Hansen statistic”, whose null hypothesis is that the variables which are assumed to be exogenous are in fact uncorrelated with the error term (see below for more discussion on this point). The “robust” option of xtabond2 has been used to get standard errors that are robust to heteroskedasticity and arbitrary patterns of autocorrelation within individuals.

¹¹ For a concise illustration of the Arellano Bond estimator, the assumptions behind this estimation technique, and how the estimator addresses the endogeneity problems see, for example, Greene (2012), Section 11.8.3. Computation has been performed using xtabond2 in Stata® 15, see Roodman (2009).

serially uncorrelated errors in (5) for Industry and Transport, whereas for Household and Services a moderate correlation is detected (although lack of correlation is accepted at 1% significance level). The robust Hansen J statistics (Hansen, 1982) confirm that, the instruments, when jointly analyzed for over-identifying orthogonality restrictions, may be regarded as exogenous. Moreover, the difference-in-Hansen statistics,¹⁴ which is designed to test the null hypothesis that the supposedly exogenous regressors are uncorrelated with the error term, accepts the null.

Before discussing the estimates of the key parameters, let us briefly observe that the coefficients of the control variables are all correctly signed, with the only exception of the negative sign on $\ln(pop_{i,t})$ and $\ln(rgdp_{i,t})$ in the household equation, although these negative parameters are compensated by the positive parameter on the other “scale” variable $\ln(rcons_{i,t})$. This suggests that the scale effect in the household sector is small. In the other sectors the scale effect (essentially measured by the sum of all the parameters of the scale related variables) is higher, especially for services and transport. The price elasticities are always negative, ranging from -0.054 (transport) to -0.40 (services). These figures might, at first sight, appear low as compared to other studies, but it has to be remarked that β_1^h is the impact elasticity, whereas the long run elasticity is given by $\frac{\beta_1^h}{1-\rho}$ (see the discussion in subsection 5.2), and it is therefore almost twice as much, which is in line with usual results. The sector specific variables for services and industry are correctly signed but very insignificant, as well as floor area for household. The number of cars also appears insignificant, with a wrong sign and very small magnitude: we also tried to include passenger-km (source Enerdata) as an alternative way to account for passenger transport, but the result is similar; conversely, freight transport seems much more important and significant.

Focusing on the Energy Policy Indicator, the coefficient is negative in all four equations - which is consistent with effectiveness of energy efficiency policies. Before discussing these estimates in detail, as anticipated in Section 4.3, we wish to test for exogeneity of the Energy Policy Indicators by re-estimating the model treating this variable as endogenous, and therefore introducing suitable instrumental variables. As illustrated in Section 4.3, we use as instruments an indicator of energy dependence and an indicator of greenhouse gas emission, both evaluated in the previous year, $ed_{i,t-1}$, and $ghge_{i,t-1}$ (see the appendix for details). Table 4 illustrate the result of the Hausman t -test, see section 4.3 for details and references on the test; the estimates when $pol_{h,i,t}^{U,P}$ is treated as endogenous (and therefore instrumented) are very similar to the estimates reported in Table 3, and the Hausman test does not reject exogeneity¹⁵: therefore, in the following we will consider $pol_{h,i,t}^{U,P}$ as exogenous, and analyze the estimates in Table 3, since they are theoretically more efficient.

The estimated coefficient for the energy policy indicator is negatively

¹⁴ This test is also called C statistic in the literature, see Baum et al. (2003) and Roodman (2009). As Roodman (2009) points out, “the Hansen test should not be relied upon too faithfully, because it is prone to weakness ... the test actually grows weaker the more moment conditions there are”. Since we have as many as 253 instruments, we tried to severely reduce the instruments count by reducing the number of lags of the differenced dependent variable to be used as instruments: the results remain essentially the same.

¹⁵ The Hausman test could not be computed for the Industry sector, since in that case the standard error associated to the difference of estimates turns out to be negative (even using the option `sigmamore` in Stata command `Hausman`). This circumstance is usually referred to as “Heywood case”. This is usually interpreted as evidence in favour of H_0 . However, following Schreiber (2008) who is sceptical about this recommendation, we also considered the absolute value of the statistic, which is equal to 2.15 (p-value 0.03): this would marginally reject H_0 for the Industry sector. Notice however that, the estimates based on Instrumental Variables for the Industry sector would indicate that the impact of the policy indicator is higher and more significant than what emerges from Table 3, therefore our conclusion would be unchanged.

signed, as expected, in all sectors, although with different magnitude and significance. In particular, it is insignificant for Services (p-value 0.846), almost significant for Household (p-value 0.162), weakly significant for Industry and Transport (p-value 0.058 and 0.048 respectively). Low significance was expected, since clearly the current version of our energy policy indicator is a poorly measured proxy of the “true” policy effort. However, the fact that the sign of the coefficient is in line with expectations is very encouraging. As discussed in Section 3, the quality of the indicators can be improved, and our preliminary results in a follow up study limited to the household sector and a subset of countries seem to suggest that a better proxy increases magnitude and significance of the estimated coefficient, without altering its sign. Moreover, as discussed in Wasserstein and Lazar (2016), a p-value >0.05 does not imply that the null hypothesis is true, and “scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold” (see also McShane and Gal (2017), Amrhein et al. (2019), and most of the contributions to the 2019 special issue of The American Statistician “Statistical Inference in the 21st Century: A World Beyond $p < 0.05$ ”).

We also checked whether the alternative versions of the Energy Policy Indicators discussed in Section 3, namely $pol_{h,i,t}^{W,P}$ and $pol_{h,i,t}^{U,T}$, improve the fit of the model. Due to space constraints we do not report the results (available upon request), but essentially the performance of the two alternative indexes is quite similar in terms of sign, magnitude and significance as compared to the base version of the indicator.¹⁶

5.2. Interpretation of the policy coefficients

The EP indicators have negative sign, which supports the effectiveness of energy policies in all sectors except services, where the coefficient is very close to zero and insignificant. To interpret the magnitude of the coefficients, notice that the EP indicators in this version of the model are the unweighted ones. Therefore, a unit increase in $pol_{h,i,t}$ corresponds to the adoption of a single policy measure. In the household sector, the estimated coefficient $\hat{\gamma}^1 = -0.0017$ implies that when $pol_{h,i,t}$ is increased by 1 (i.e. a new measure is adopted in the household sector), the logarithm of energy consumption is reduced on average by 0.0017, which implies that energy consumption is reduced by about 1.7 per thousand in the same year (80% significant). $\hat{\gamma}^1$ is sometimes referred to as short-run elasticity. Due to the autoregressive component, the change in $pol_{h,i,t}$ induces a dynamic adjustment leading, in the long run, to a reduction in the logarithm of energy consumption in the household sector equal to $\frac{\hat{\gamma}^1}{1-\rho} = -0.00354$, i.e. about 3.5 per thousand, so that on average it takes about 3 measures - and some time - to reduce consumption by 1%. A similar argument may be made for each sector, leading to the short and long run saving associated to one (typical) measure, reported in Table 5.

The estimates suggest a much stronger effectiveness of energy policies adopted in the industrial sector, where the percentage saving associated to each measure is estimated to be, in the long run, about 2%.

¹⁶ In other words, the MURE semi quantitative impact assessment (low, medium, high impact) as a basis for weighting is not really improving the quality of the indicator, although it is not even worsening it: this might suggest that the method used to attribute low, medium or high importance to the energy policy measures is not completely coherent across countries or over time. As for the comparison of the “permanent effect” and “transitory effect” versions of the indicator, the evidence that none of the versions outperforms the other seems to suggest that both extreme versions of the indicator have merits and flaws, which probably balance. More research on the issue would be needed.

Table 3

Estimates of the sectoral model (p-value of the two tailed t-test in parenthesis; ***:p-value<0.01, **:p-value<0.05, *:p-value<0.1, ° p-value<0.2).

VARIABLE	MODEL (5) Households	MODEL (5) Services	MODEL (5) Industry	MODEL (5) Transport
$\ln(q_{h,i,t-1}^3)$	0.52 (0.001)***	0.47 (0.000)***	0.59 (0.000)***	0.56 (0.000)***
$pol_{h,i,t}^{U,P}$	- 0.0017 (0.162)°	- 0.00053 (0.846)	- 0.0083 (0.058)*	- 0.0026 (0.048)**
$\ln\left(\frac{p_{h0,it}^3}{def_{it}}\right)$	- 0.11 (0.001)***	- 0.40 (0.000)***	- 0.080 (0.004)***	- 0.054 (0.061)*
$other_{h0,it}$	- 0.83 (0.000)***	- 0.46 (0.037)**	- 0.55 (0.013)**	- 1.44 (0.001)***
$\ln(pop_{it})$	- 0.18 (0.414)	0.58 (0.148)°	0.087 (0.790)	0.44 (0.005)***
$\ln(rgdp_{it})$	- 0.14 (0.088)*	0.37 (0.034)**	0.13 (0.409)	0.32 (0.001)***
hdd_{it}	0.00014 (0.000)***	0.00013 (0.002)***	0.000022 (0.092)*	0.000018 (0.165)°
t	- 0.0033 (0.453)	- 0.0085 (0.303)	0.0049 (0.430)	0.0023 (0.545)
t^2	0.00016 (0.115)°	0.00055 (0.049)**	- 0.00022 (0.247)	0.000032 (0.800)
$\ln(area_s1_{it})$	0.060 (0.565)			
$perc_equip_s1_{it}$	0.20 (0.048)**			
$\ln(rcons_s1_{it})$	0.26 (0.005)***			
$\ln(rva_s2_{it})$		0.12 (0.455)		
$\ln(empl_s2_{it})$		0.16 (0.286)		
$\ln(rva_s3_{it})$			0.027 (0.763)	
$\ln(rginv_s3_{it})$			- 0.019 (0.571)	
$\ln(cars_s4_{it})$				- 0.062 (0.575)
$\ln(goods_s4_{it})$				0.08 (0.010)***
Arellano-Bond m_2 test	$z = 2.7, p_{val} = 0.01$	$z = 2.3, p_{val} = 0.02$	$z = 0.0, p_{val} = 0.99$	$z = -1.8, p_{val} = 0.07$
Hansen test	$\chi^2_{251} = 24.7, p_{val} = 1.00$	$\chi^2_{252} = 25.7, p_{val} = 1.00$	$\chi^2_{252} = 23.8, p_{val} = 1.00$	$\chi^2_{252} = 23.2, p_{val} = 1.00$
Difference-in-Hansen	$\chi^2_{10} = 1.0, p_{val} = 1.00$	$\chi^2_{10} = 4.0, p_{val} = 0.94$	$\chi^2_{10} = 0.2, p_{val} = 1.00$	$\chi^2_{10} = 0.6, p_{val} = 1.00$
Number of obs	515	521	522	565

Table 4

Hausman t-test for the exogeneity of the Energy Policy Indicator.

	MODEL (5) Households	MODEL (5) Services	MODEL (5) Industry	MODEL (5) Transport
$\hat{\gamma}^h$ ($pol_{h,i,t}^{U,P}$ endogenous)	- 0.0019	- 0.0010	- 0.010	- 0.0022
$\hat{\gamma}^h$ ($pol_{h,i,t}^{U,P}$ exogenous)	- 0.0017	- 0.00053	- 0.0083	- 0.0024
Difference of estimates	- 0.00017 (st. err. : 0.0012)	- 0.00050 (st. err. : 0.0017)	- 0.0028 (st. err. : na)	0.00019 (st. err. : 0.00026)
Hausmann t-test	- 0.148 (pval : 0.88)	- 0.295 (pval : 0.77)	na (pval : na)	0.720 (pval : 0.47)

5.3. Estimating policy-induced energy saving in each country and in Europe

Estimation of the policy-induced energy saving in each country in 1990–2013 may be obtained through a simulation exercise.¹⁷ To isolate the effect of the policy variable, we consider the estimated counterpart of equation (5), dropping all other regressors, the intercept and the error term. This gives the following simple equation, which is all we need to discuss the dynamic effect of $pol_{h,i,t}^{U,P}$ when the other variables are fixed:

$$\ln(q_{h,i,t}^3) = \rho^h \ln(q_{h,i,t-1}^3) + \gamma^h pol_{h,i,t}^{U,P} \tag{6}$$

Based on (6) and assuming $0 < \rho^h < 1$, it can be easily seen that when $pol_{h,i,t-1}^{U,P}$ takes on some value, and $\ln(q_{h,i,t-1}^3) = \frac{\gamma^h pol_{h,i,t-1}^{U,P}}{1-\rho^h}$, then the system is in steady state, which means that if $pol_{h,i,t}^{U,P} = pol_{h,i,t-1}^{U,P}$ (i.e. it is not changed), then $\ln(q_{h,i,t}^3)$ also remains fixed. The equation is initialized with the policy indicator equal to its true historical value in 1989, i.e. $pol_{h,i,1989}^{U,P}$, while $\ln(q_{h,i,t}^3)$ is initialized at the corresponding steady state value $\ln(q_{h,i,t}^3) = \frac{\gamma^h pol_{h,i,1989}^{U,P}}{1-\rho^h}$; then, the entire historical path of the policy variable from 1990 to 2013 is given as an input, and the response of

energy consumption is measured and analyzed. Assuming that the future values of all regressors, as well as the error term, are not influenced by the current value of the energy policy (otherwise neglecting them would not be appropriate), the final step of the simulation is a measure of the energy saving in 2013 induced by all energy policies introduced from the beginning to 2013. Moreover, the simulation provides a measure, in each year, of the percentage energy saving induced by energy policies adopted up to that year. This approach is like a negative image of the counterfactual simulation approach adopted, for example, in Horowitz (2011): there, the model is estimated without the policy variable using the pre-policy period, and then the energy policy is set to zero in the simulated period, leaving the other variables at their historical level. Conversely, here we estimate the model using the entire period (we do not need a “policy free” period for estimation), and then we simulate the entire period as if the other variables are fixed, allowing only the policy variable to change.

As an example, Table 6 provides the details of the dynamic

Table 5

Estimated short and long run elasticities of policy measures on energy consumption in each sector.

SECTOR	SHORT RUN ELASTICITY	LONG RUN ELASTICITY
Household	-0.17%	-0.35%
Services	-0.05%	- 0.10%
Industry	-0.83%	-2.02%
Transport	-0.26%	-0.59%

¹⁷ The exercise is a variant of the step response analysis, see for example Ogata (1995).

simulation exercise for the Household sector in France, Germany and Italy; the simulations are then illustrated in Fig. 4.

The point estimates suggest that the consumption of energy the household sector in France in 2013, due to all policies adopted since the nineties, was 15.9% less than what it would have been in the absence of policies. In Germany, the percentage of energy savings was 9.4% in 2013, whereas in Italy it was 7.0%. Of course, confidence intervals around these figures would be useful, but the computation is not easy, and we leave it for future research. The differences between France, Germany and Italy are due to the fact that, according to MURE’s database, the number of measures adopted in the household sector are different in the three countries (47 in France, 27 in Germany and 22 in Italy).

It may be worth noticing that these figures are likely to slightly underestimate the effect of policies. It is well known that measurement error in the independent variables determine an “attenuation bias”, i.e. a bias towards zero, in the estimates (see for example Fuller, 1987), and clearly, the Energy Policy indicators proposed here suffer from measurement errors. Investing on the indicators to improve their quality might reduce the problem.

Table 7 reports the final value of the simulation for each country and each sector, i.e. the estimated percentage policy-induced energy savings in 2013 (we have changed the sign from negative to positive for readability). By multiplying the percentage savings by the actual consumption in 2013 one gets the absolute saving (in TJ), which may then be aggregated across countries to obtain the total saving in 2013 for EU28 plus Norway, and across sectors, to obtain the estimated savings in each country for all sectors.

The estimated percentage policy-induced energy savings aggregated on EU28 plus Norway are about 10% for Households and Transport, about 20% for Industry, and about zero for Services, which corresponds to 12.1% when we aggregate all sectors, and is equivalent to about 4.9 million TJ. The aggregate figure for EU is consistent with other studies (Horowitz and Bertoldi, 2015), whereas the comparison among MSs sometimes seems to contradict the common sense on the level of commitment of national governments with respect to energy policies. As we already pointed out, the results reported in Table 7 are mainly related to the number of measures reported in the MURE database for

Table 6
Dynamic simulation of the estimated effect of energy policies in the household sector in France, Germany and Italy.

t (year)	pol_t			$\ln(q_t) = 0.52\ln(q_{t-1}) - 0.0017pol_t$		
	France	Germany	Italy	France	Germany	Italy
1989	7	6	0	-0.025	-0.021	0.000
1990	8	6	0	-0.026	-0.021	0.000
1991	9	7	0	-0.029	-0.023	0.000
1992	9	7	1	-0.030	-0.024	-0.002
1993	9	8	2	-0.031	-0.026	-0.004
1994	10	9	2	-0.033	-0.029	-0.006
1995	12	10	2	-0.038	-0.032	-0.006
1996	13	10	2	-0.042	-0.034	-0.007
1997	13	10	3	-0.044	-0.034	-0.009
1998	14	11	6	-0.047	-0.037	-0.015
1999	16	14	7	-0.051	-0.043	-0.020
2000	17	14	7	-0.056	-0.046	-0.022
2001	19	15	7	-0.061	-0.049	-0.023
2002	20	18	7	-0.066	-0.056	-0.024
2003	20	18	8	-0.068	-0.060	-0.026
2004	22	18	9	-0.073	-0.062	-0.029
2005	22	20	11	-0.075	-0.066	-0.034
2006	26	20	13	-0.083	-0.068	-0.040
2007	30	20	14	-0.094	-0.070	-0.044
2008	35	22	15	-0.109	-0.074	-0.049
2009	41	25	16	-0.126	-0.081	-0.052
2010	43	26	16	-0.139	-0.086	-0.054
2011	44	27	17	-0.147	-0.091	-0.059
2012	45	27	19	-0.153	-0.093	-0.063
2013	47	27	22	-0.159	-0.094	-0.070

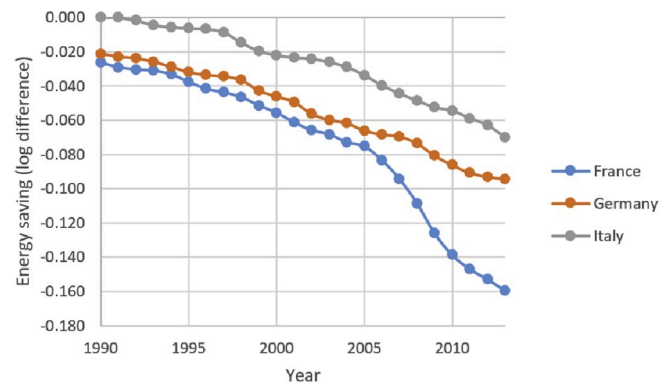


Fig. 4. Dynamic simulation exercise for France, Germany and Italy, Household sector.

each country, which deserves further investigation. For example, according to the MURE database the average number of energy policy measures adopted by European countries between 1990 and 2013 was 66; Spain is reported to have taken 139 measures while Denmark only 27. Even worse if we focus on “high impact” measures: according to the MURE database, the average number of “high impact” energy policy measures between 1990 and 2013 was 21; Spain is reported to have taken 91 “high impact” measures while Denmark only 3.

This discussion underlines the problem of treating a policy (as defined in MURE) as equivalent to another: how does this affect our model results and thus our simulations reported in Table 7? To answer this, let us first restate that the purpose of our model is not to analyze which policy is more effective: what we try to do is to estimate the average effect of one policy, being aware that individual policies are different, and therefore might have different effect. Clearly, what we mean by “one policy” has to be defined properly, and our paper leaves this burden to MURE and their guidelines (Schlomann et al., 2016): our attempt is to stimulate the discussion on this, pointing out that, quite likely, national MURE teams interpret the guidelines, and the notion of “one measure”, differently. Although this is not the perfect situation, we think that the results in Table 7 are interesting, although preliminary. It is worth remarking that systematic differences in the interpretation of the guidelines among countries are to some extent accounted for by the “country effect” $\beta_{0,i}^h$ in equation (5). In other words, countries interpreting the notion of “one measure” more restrictively, hence reporting fewer measures, are likely to have (everything else being fixed), a lower country effect. The opposite is true for countries adopting a more extensive notion of measure. Unfortunately, country effects also take into account other omitted variables, therefore disentangling the specific component of $\beta_{0,i}^h$ associated with country specific mis-measurement of the policy indicator is difficult (although country-specific heterogeneity of the parameter γ^h might be considered in the future). However, assuming that extensive and restrictive interpretations of the notion of measure balance across countries, the estimate of the overall European policy-induced energy savings (12.1%) can be regarded as plausible (although possibly underestimating the real effect due to attenuation), even if the contribution of the individual countries to the overall result has to be taken with caution. Indeed, a more careful analysis of the MURE database, possibly correcting this unbalance using the IEA database, might lead to a better and more reliable energy policy database, resulting in a more precise assessment of the amount of the policy-induced energy savings in the individual MSs.

6. Conclusion and policy implications

In this article, we have developed new econometric models aimed at providing some suggestive evidence of the effect of energy efficiency

Table 7
Estimated Policy-Induced Energy Savings based on the model (percentage and absolute value, in TJ).

COUNTRY	Household		Services		Industry		Transport		All Sectors	
	Saving in 2013		Saving in 2013		Saving in 2013		Saving in 2013		Saving in 2013	
	%	TJ	%	TJ	%	TJ	%	TJ	%	TJ
Austria	3.5%	5973	0.9%	597	6.0%	16422	6.5%	23532	5.3%	46525
Belgium	6.2%	22187	2.0%	4063	14.1%	49016	5.9%	23841	7.6%	99107
Bulgaria	7.0%	2976	1.5%	491	24.7%	19949	5.8%	6654	11.1%	30070
Croatia	5.2%	2642	1.6%	456	10.0%	3959	13.8%	12425	9.3%	19482
Cyprus	2.2%	226	0.5%	37	7.9%	626	5.8%	2146	4.9%	3034
Czech Republic	5.3%	7486	0.6%	633	11.6%	24141	5.3%	13063	6.4%	45322
Denmark	3.4%	2696	0.3%	145	2.9%	2211	4.6%	9002	3.5%	14054
Estonia	6.4%	608	2.1%	263	22.0%	4091	7.2%	2356	9.9%	7318
Finland	7.8%	7709	2.8%	2167	29.0%	76285	12.0%	24604	17.2%	110766
France	15.9%	247717	2.4%	22685	34.4%	431977	16.4%	346315	17.8%	1048694
Germany	9.4%	200437	2.4%	32011	20.1%	410471	10.3%	272638	11.3%	915558
Greece	3.5%	4062	0.9%	711	10.6%	11658	6.2%	16554	5.8%	32985
Hungary	5.6%	8277	0.5%	437	12.0%	15536	6.4%	9717	6.6%	33966
Ireland	8.5%	8509	2.3%	1235	30.0%	27986	13.8%	25638	14.6%	63368
Italy	7.0%	80730	1.6%	10193	26.1%	268379	14.7%	247598	13.4%	606900
Latvia	3.8%	498	0.9%	138	18.4%	3044	6.1%	2756	7.2%	6436
Lithuania	4.5%	756	2.1%	298	11.3%	2747	5.9%	3859	6.3%	7659
Luxembourg	4.0%	734	0.4%	78	14.2%	2903	3.5%	3714	4.5%	7429
Malta	6.9%	237	1.5%	35	9.4%	204	2.7%	333	4.0%	809
Netherlands	8.8%	39151	1.3%	4487	38.2%	214580	9.4%	58959	16.0%	317178
Norway	8.0%	11619	2.2%	2237	35.6%	88041	5.8%	13052	16.0%	114949
Poland	1.0%	2784	0.6%	1580	12.0%	47408	5.2%	34588	5.5%	86360
Portugal	5.3%	4117	1.2%	852	6.0%	8045	10.4%	28179	7.4%	41193
Romania	3.8%	5982	1.2%	761	13.6%	30781	5.3%	11576	7.4%	49100
Slovakia	5.8%	3919	1.8%	1276	23.7%	28075	4.0%	3839	10.5%	37109
Slovenia	4.5%	1118	1.0%	180	11.9%	5357	5.1%	3985	6.4%	10639
Spain	10.2%	54897	3.7%	15089	29.1%	255559	26.1%	387900	21.6%	713447
Sweden	4.0%	5716	0.5%	638	11.7%	29835	7.8%	26099	7.2%	62288
United Kingdom	5.4%	90806	1.4%	10043	12.9%	116562	8.8%	190396	7.5%	407808
All countries	8.5%	824569	1.9%	113816	22.4%	2195848	11.9%	1805319	12.1%	4939552

policies on energy consumption in the EU MSs plus Norway in the period 1990–2013. The novelty of the approach lies in the use of MURE's database to produce panel Energy Policy Indicators at the sector level for each EU MSs plus Norway. These policy indicators are then included, along with control variables, in dynamic panel models for each sector. The models are then used to try to estimate a quantitative measure of the policy-induced energy savings from 1990 to 2013, measured as a percentage of the energy consumption as it would have been in the absence of energy policies.

In order to avoid any bias in the estimates, we have introduced a suitable set of control variables and, following a standard approach in the macro-econometric literature, we have used suitable Instrumental Variables (IV) for the energy policy indicator (an indicator of energy dependence and an indicator of greenhouse gas emission). IV might fail to deliver unbiased estimates if the instruments are correlated with the error term in the focal equation, and if the instruments are weak: however, proper tests suggest that neither of these problems is likely to affect the proposed model. Therefore the figures given in [Table 7](#) can be tentatively interpreted as estimates of policy induced energy saving, confirming suggestive evidence of the effect of policy measures on energy demand. This paper hopefully will stimulate further research providing additional evidence on the issue.

Although the energy policy indicator is very aggregate, being given by the total number of policies adopted by each country in each year, the estimated parameters suggest that the “average” energy policy measure is associated to a non-negligible percentage decrease in energy consumption. In order to assess the effectiveness of different policies we have replaced the basic policy intensity indicator with a weighted version, where the weights are given by the MURE “impact indicator”, however this did not prove superior to the basic one in estimating energy savings. This highlights that more research is needed to find out rigorous and operational impact evaluation criteria. Alternative weighting schemes may be tested (for example based on the “policy type”).

In summary, the results are the following: (i) Energy policies contribute to reducing energy consumption. In the absence of energy policies consumption in EU28 plus Norway would have been approximately 12% higher in 2013, in line with findings by other authors. (ii) The energy savings induced by policies seems to be higher in Industry (20.4% savings in 2013 for EU28 plus Norway), intermediate for Household (8.5%) and Transport (11.9%),¹⁸ while for Services the magnitude and significance of the effect seems negligible. (iii) For most MSs the ranking based on energy policy-induced energy saving seems in line with previous findings ([Filippini et al., 2014](#)), while in some cases the results diverge. The possible explanation is linked to the number of policy measures reported in the MURE database, which for some countries seems too low or too high.

The present research points to some preliminary policy conclusions: the policies adopted by EU MSs have been more effective in the industrial sector ([Malinauskaitė et al., 2019](#)), composed by a relative small number of operators and dominated by large energy intensive plants. Active collaboration between private organizations and public authorities, e.g. voluntary agreements in the Netherlands and Finland ([Rezessy and Bertoldi, 2011](#)), and clear competitiveness advantages for companies have contributed to successful implementation of policies. The service sector results confirm the difficulties to have effective policies, due to the large number of SMEs in the sector, the low economic benefits for private operators, the landlord-tenant dilemma and the large number of public buildings ([Schloman and Schleich, 2015](#)). In the residential and transport sectors, the analysis confirms the important contribution of policies on mitigation of energy demand despite the large number and

¹⁸ The average impact of one policy measure in the transport sector is estimated to be higher than in the household sector (see [Table 5](#)), but the number of policies is higher in the household sector in most countries (see [Table 1](#)): this results in a similar energy saving in the two sectors.

fragmentation of decision makers.

However, there are some limitations on the policy implication derived by these preliminary results given the aggregate nature of the indicators containing all policy measures, for the example the impact of different type of policies (e.g. market based instruments, regulation, information, etc.) cannot be derived.

In this article, we have identified possible areas for future research focusing on energy policy indicators and models:

1. Improving Energy Policy Indicators: this is a key issue, since a more reliable and accurate database, might reduce the attenuation effect, resulting in a more precise assessment of the amount of policy-induced energy savings in the individual countries. The results obtained in this research are encouraging, since the sign and magnitude of the indicators based on the MURE database “as is” are coherent. Improvements might come from using the information about the type of measure (financial, regulatory, information, etc.).
2. Adapting the model to address further issues: It would be interesting to analyze whether the impact of sector specific energy efficiency policies changes according to the type of energy source. Ideally, it would be possible to introduce “energy source specific” EP Indicators. With these indicators, it would be possible to estimate source specific equations (e.g. electricity consumption in the household sector), and within this class of disaggregate models one might also explore the empirical relevance of inter-fuel substitution between coal, oil, gas, and electricity, see Hall (1986), Urga and Walters (2003), Stern (2009). Another interesting aspect to be considered for the industrial sector is the presence of composition effects. The changes of energy intensity in the industrial sector is also due to the relative decline of energy intensive manufacturing. This shift has also increased the use of electricity with respect to oil, coal, gas. Estimated input demand function at an aggregate level should therefore include an indicator accounting for changes in the weight of energy intensive industry over time.
3. Improving the econometric methodology: one interesting issue in the dynamic analysis of energy panels, pointed out for example in Madlener et al. (2011), is the non-stationarity of most of the

variables involved in energy studies. To exclude possible inconsistency in the estimates and spurious regression problems, it would be important to carry out appropriate unit roots and cointegration analysis within the panel framework (see among others Baltagi et al., 2001). It would also be possible to explore possible heterogeneity of the coefficients across countries, as in the case of Greene (2012). The evidence from single country energy demand suggests for example that price elasticities might differ substantially across countries. Hsiao and Pesaran (2008) illustrates how panel models may allow for heterogeneous coefficients. Allowing for parameters heterogeneity would enable researchers to explore whether energy policy effectiveness is equal in all countries. Finally, it would be interesting to explore the use of additional or alternative instrumental variables to the ones used in this paper, such as political variables (see for example Datta and Filippini, 2016).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRedit authorship contribution statement

Paolo Bertoldi: Conceptualization, Methodology. **Rocco Mosconi:** Conceptualization, Methodology.

Acknowledgments

The authors would like to thank Marvin J. Horowitz for his insights and advice, Francesca Diluiso and Marina Economidou for valuable comments on an earlier version of the paper, and three anonymous reviewers for their constructive comments and suggestions in preparing the final version of the manuscript; in particular, we are indebted to the anonymous reviewers for their comments and suggestions on causality. The authors take sole responsibility for any errors and all opinions found within.

Appendix. The dataset

In this project we have created a dataset covering 29 European countries (EU28 + Norway, $i = 1, \dots, 29$). The data are annual, from 1990 to 2013 ($t = 1, \dots, 24$). The main sources are Eurostat (ec.europa.eu/eurostat/data/database), Enerdata (Global Energy and CO₂ data, services.enerdata.net) and the MURE database for energy policy measures (www.measures-odyssee-mure.eu).

Whenever possible, the data are disaggregated by sector ($h = 1, \dots, 4$: 1 = Household, 2 = Tertiary, 3 = Industry, 4 = Transport) and by energy source ($s = 1, \dots, 4$: 1 = Electricity, 2 = Gas, 3 = Oil, 4 = Solid fuels). We excluded all other sectors (agriculture) and all other sources (derived heat, renewable energies, nuclear heat, non-renewable waste). As we will show below, the weight of the excluded sectors and sources in terms of energy consumption varies across countries, ranging from about 5% to about 40%.

We can group the variables in the dataset in five groups, introduced in the following Subsections.

Quantity variables

The source is Eurostat, and all quantities are in TJ; no missing values. We have also collected quantities from Enerdata and checked the coherence with Eurostat. Enerdata provides data for a longer time period, but the data on oil products and solid sources are disaggregate and the components are measured in different units: the attempt to convert and aggregate them provides time series which do not match Eurostat very much, so we decided to use Eurostat for quantities and Enerdata for prices. The quantity variables in our data set are:

- $q_{h,s,i,t}$: energy consumption in sector h (1 = household, 2 = services, 3 = industry, 4 = transport) from source s (1 = electricity, 2 = gas, 3 = oil, 4 = solid fuel), country i year t .
- $q_{h,i,t}^4 = \sum_{s=1}^4 q_{h,s,i,t}$: energy consumption of sector h from the 4 sources mentioned above; it is not total energy consumption in sector h , since it excludes other sources
- $q_{h,i,t}^3 = \sum_{s=1}^3 q_{h,s,i,t}$: (excludes solid fuels also)

- $q_{i,t}^4 = \sum_{h=1}^4 q_{h,i,t}^4$: total energy consumption in the four sector mentioned above from the 4 sources mentioned above in country i year t ; it is not total energy consumption in the country, since it excludes other sources and other sectors
- $q_{i,t}^3 = \sum_{h=1}^4 q_{h,i,t}^3$: (excludes solid fuels also)
- $q_{h,i,t}$: total energy demand (all sources) for sector h , provided by Eurostat
- $q_{i,t}$: total energy demand (all sources, all sectors), provided by Eurostat

The average coverage of $q_{i,t}^4$ and $q_{i,t}^3$ on total (including all sectors and all sources) energy consumption $q_{i,t}$ in each country is given in Table 8.

Table 8

Average coverage (1990-2013) of $q_{i,t}^4$ and $q_{i,t}^3$ on total energy consumption $q_{i,t}$.
 = low coverage (less than 70%)

	Coverage $q_{i,t}^4$	Coverage $q_{i,t}^3$
Austria	82%	76%
Belgium	94%	87%
Bulgaria	74%	65%
Croatia	85%	82%
Cyprus	95%	93%
Czech Republic	80%	61%
Denmark	73%	71%
Estonia	58%	53%
Finland	65%	61%
France	88%	84%
Germany	91%	85%
Greece	88%	84%
Hungary	83%	78%
Ireland	96%	87%
Italy	94%	91%
Latvia	54%	51%
Lithuania	63%	58%
Luxembourg	97%	91%
Malta	96%	96%
Netherlands	88%	85%
Norway	89%	84%
Poland	73%	49%
Portugal	80%	79%
Romania	73%	67%
Slovakia	88%	70%
Slovenia	85%	83%
Spain	91%	89%
Sweden	72%	69%
United Kingdom	96%	92%
AVERAGE	82%	77%

Price variables

The source is Enerdata, and all prices have been converted in KEuro/TJ from the original unit. Despite the effort in reconstructing many points based on reasonable assumptions, many missing values remain (mainly in the 90's in former planned economies countries).

- $p_{h,s,i,t}$: price of energy, sector h (1 = household, 2 = services, 3 = industry, 4 = transport) source s (1 = electricity, 2 = gas, 3 = oil, 4 = solid fuel). More specifically, we have chosen the following series from the Enerdata database:
 - $p_{1,1,i,t}$: "Price per toe in € of electricity for households (taxes incl.)", divided by 0.01163 to convert in KEuro/TJ. This series from Enerdata is very similar to the Eurostat series for household consumer band DC, which is the median band with the highest number of electricity and gas consumers in the majority of Member States.¹⁹ We have opted for the Enerdata series since it has fewer missing values.
 - $p_{1,2,i,t}$: "Price per toe in € of natural gas for households (taxes incl.) NCV", divided by 0.01163 to convert in KEuro/TJ. This series from Enerdata is very similar to the Eurostat series for household consumer band D2, which is the median bands with the highest number of gas consumers in the majority of Member States.²⁰ We have opted for the Enerdata series since it has fewer missing values.
 - $p_{1,3,i,t}$: "Price per toe in € of light fuel oil for households (taxes incl.)", divided by 0.9×0.01163 to convert in KEuro/TJ.
 - $p_{1,4,i,t}$: "Price per toe in € of bituminous coal for households (taxes incl.)", divided by 0.01163 to convert in KEuro/TJ. We have not considered the price of other solid fuels since the series are too incomplete. The coverage of bituminous coal on solid fuels seems high. The series has many missing values, especially in those countries where the weight of solid fuels for household is low.
 - $p_{2,1,i,t}$: There is no official time series for the price of electricity for services. Therefore, we use the average (equally weighted) of the prices for household ($p_{1,1,i,t}$) and the prices for industry ($p_{3,1,i,t}$).

¹⁹ The limiting values for the consumer band DC is: 2,500 kWh < Consumption <5,000 kWh.

²⁰ The limiting values for the consumer band D2 are: 20 GJ < Consumption <200 GJ equivalent to 5560 kWh < Consumption <55 560 kWh.

- $p_{2,2,i,t}$: There is no official time series for the price of gas for services. Therefore, we use the average (equally weighted) of the prices for household ($p_{1,2,i,t}$) and the prices for industry ($p_{3,2,i,t}$).
- $p_{2,3,i,t}$: There is no official time series for the price of oil products for services. Therefore, we use the average (equally weighted) of the prices for household ($p_{1,3,i,t}$) and the prices for industry ($p_{3,3,i,t}$).
- $p_{2,4,i,t}$: There is no official time series for the price of solid fuels for services. Therefore, we use the average (equally weighted) of the prices for household ($p_{1,4,i,t}$) and the prices for industry ($p_{3,4,i,t}$).
- $p_{3,1,i,t}$: “Price per toe in € of electricity in industry (taxes incl.)”, divided by 0.01163 to convert in KEuro/TJ. This series from Enerdata is very similar to the Eurostat series for industrial sector band IC, which typically represent medium size enterprises.²¹ We have opted for the Enerdata series since it has fewer missing values.
- $p_{3,2,i,t}$: “Price per toe in € of natural gas in industry (taxes incl.) NCV”, divided by 0.01163 to convert in KEuro/TJ. This series from Enerdata is very similar to the Eurostat series for industrial sector band I3, which typically represent medium size enterprises.²² We have opted for the Enerdata series since it has fewer missing values.
- $p_{3,3,i,t}$: We use the average (equally weighted) of “Price per toe in € of heavy fuel oil in industry (taxes incl.)” and “Price per toe in € of light fuel oil in industry (taxes incl.)”. We have not considered the price of other oil products since the series are too incomplete. The coverage of these two products seems high, and the weight, although varying across countries and years, is similar. In most countries, the price of light fuel is approximately twice the price of heavy fuel. We have then divided by 0.01163 to convert in KEuro/TJ.
- $p_{3,4,i,t}$: “Price per toe in € of bituminous coal in industry (taxes incl.)”, divided by 0.01163 to convert in KEuro/TJ. We have not considered the price of other solid fuels since the series are too incomplete. The coverage of bituminous coal on solid fuels in industry seems high. The series has many missing values, especially in those countries where the weight of solid fuels for industry is low.
- $p_{4,1,i,t}$: We have not collected any price, since the weight of electricity for transport is extremely low.
- $p_{4,2,i,t}$: We have not collected any price, since the weight of gas for transport is extremely low.
- $p_{4,3,i,t}$: “Price per toe in € of premium gasoline (taxes incl.)”, divided by 0.01163 to convert in KEuro/TJ. We have not considered the price of other fuels since the series are too incomplete. The coverage of premium gasoline on oil products for transport is high, and the price of other fuels, when available, is highly correlated.
- $p_{4,4,i,t}$: We have not collected any price, since the weight of solid fuel for transport is extremely low.
- $p_{h,i,t}^4 = \sum_{s=1}^4 \alpha_{h,s,i,t}^4 p_{h,s,i,t}$: reference price for $q_{h,i,t}^4$, obtained as a (time varying weighted) average of energy prices for sector h . The weights represent the relevance of source s in sector h (4 sources, excluding the others), and are therefore given by $\alpha_{h,s,i,t}^4 = \frac{q_{h,s,i,t}}{\sum_{j=1}^4 q_{h,j,i,t}}$. If the weight $\alpha_{h,s,i,t}^4 < 0.1$ then source s is excluded for that year and that country, and $p_{h,i,t}^4$ is computed as a weighted average of just the other prices (the small weight is set to zero and the others are readjusted to sum up to one). The reason for excluding sources with small weight is that, in many countries, the price of sources whose weight is small are missing (or unreliable). If the prices of sources whose weight is larger than 0.1 is missing, $p_{h,i,t}^4$ is also considered as missing.
- $p_{h,i,t}^3 = \sum_{s=1}^3 \alpha_{h,s,i,t}^3 p_{h,s,i,t}$: reference price for $q_{h,i,t}^3$, where solid fuels are also excluded. The weights represent the relevance of source s in sector h (3 sources, excluding the others), and are therefore given by $\alpha_{h,s,i,t}^3 = \frac{q_{h,s,i,t}}{\sum_{j=1}^3 q_{h,j,i,t}}$, used as illustrated above.
- $p_{i,t}^4 = \sum_{h=1}^4 \omega_{h,i,t}^4 p_{h,i,t}^4$: reference price for $q_{i,t}^4$, obtained as a (time varying weighted) average of sectoral energy prices. The weights represent the relevance of sector h in “the whole economy” (meant as 4 sectors, 4 sources), and are therefore given by $\omega_{h,i,t}^4 = \frac{q_{h,i,t}^4}{\sum_{j=1}^4 q_{j,i,t}^4}$.
- $p_{i,t}^3 = \sum_{h=1}^4 \omega_{h,i,t}^3 p_{h,i,t}^3$: reference price for $q_{i,t}^3$, obtained as a (time varying weighted) average of sectoral energy prices. The weights represent the relevance of sector h in “the whole economy” (meant as 4 sectors, 3 sources), and are therefore given by $\omega_{h,i,t}^3 = \frac{q_{h,i,t}^3}{\sum_{j=1}^4 q_{j,i,t}^3}$.

Policy variables

The source is the MURE database, and the methodology is illustrated in Section 3.

Other control variables

We divide the other control variables in 5 groups.

1. Control variables included in all models

- $pop_{i,t}$: population, source Eurostat, no missing values. For France, we have considered metropolitan France only (i.e. excluding overseas territories).
- $rgdp_{i,t}$, $ngdp_{i,t}$, $def_{i,t}$: real GDP, nominal GDP, GDP deflator, source Eurostat, available for all 29 countries from 1995 with few missing, available for a subset of countries before 1995. For some countries (Bulgaria, Hungary, Latvia, Lithuania, Luxembourg, Malta, Romania, Slovenia) the initial part of the time series is available for either RGDP or NGDP, but not both. For these countries, we have reconstructed DEF by applying the average inflation in Europe, and then we have used DEF and the available time series to work out the other. Finally, RGDP and NGDP are

²¹ The limiting values for the consumer band IC are 500 MWh < Consumption < 2000 MWh.

²² The limiting values for the consumer band I3 are 10 000 GJ < Consumption < 100 000 GJ equivalent to 2780 MWh < Consumption < 27 780 MWh.

unavailable for Greece in 2013, and have been reconstructed by applying the average growth rate for Greek RGDP and NGDP in the period 2006–2012 to the 2012 value.

- $hdd_{i,t}$: source Eurostat, availability 1980–2009. Heating degree day (HDD) is a measurement designed to reflect the demand for energy needed to heat buildings. Eurostat calculates heating degree days as $(18\text{ }^{\circ}\text{C} - T_{mean})$ if T_{mean} is lower than $15\text{ }^{\circ}\text{C}$ (heating threshold) and zero if T_{mean} is greater than or equal $15\text{ }^{\circ}\text{C}$; T_{mean} is the mean daily outdoor temperature, calculated as $T_{mean} = \frac{T_{min} + T_{max}}{2}$. Unfortunately, at the time this research has been concluded Eurostat did not provide cooling degree days which would be useful for the regression analysis for countries in Southern Europe. According to the European Environment Agency (www.eea.europa.eu/data-and-maps/indicators/heating-degree-days-1/assessment), “the number of heating degree days has decreased by 13% over the last 3 decades, yet with substantial inter-annual variation. The decrease in HDD has not been homogeneous across Europe: the absolute decrease has been largest in the cool regions in northern Europe where heating demand is highest. Temperatures in Europe are projected to continue to increase. Hence, the trend of decreasing numbers of HDD is very likely to continue, and most likely to accelerate. For example, the heat demand for space heating in 2050 was projected to decrease by 25% in the UK, and by 9% in the EU”. Therefore, since HDD IS available at Eurostat for the period 1980 until 2009, the series has been extrapolated up to 2013 using an ARMA(1,1) model with constant and trend estimated for each country
 - $other_{h,i,t}^4 = \frac{q_{h,i,t} - q_{h,i,t}^4}{q_{h,i,t}}$: share of sources different from (1 = electricity, 2 = gas, 3 = oil, 4 = solid fuel) for sector h , based on Eurostat quantities
 - $other_{h,i,t}^3 = \frac{q_{h,i,t} - q_{h,i,t}^3}{q_{h,i,t}}$: share of sources different from (1 = electricity, 2 = gas, 3 = oil) for sector h , based on Eurostat quantities
2. Other control variables (household):
- $dwell_{i,t}$: stock of dwellings (thousand), source enerdata
 - $floor_{i,t}$: average floor area of dwellings (m^2), source enerdata
 - $area_s1_{i,t} = \frac{dwell_{i,t} \times floor_{i,t}}{1000}$: total floor area of dwellings (km^2)
 - $percfreez_{i,t}$: Rate of equipment ownership for freezers (%), source enerdata (interpolated)
 - $percwash_{i,t}$: Rate of equipment ownership for washing machine (%), source enerdata (interpolated)
 - $percdish_{i,t}$: Rate of equipment ownership for dishwasher (%), source enerdata (interpolated)
 - $percequip_s1_{i,t} = \frac{percfreez_{i,t} + percwash_{i,t} + percdish_{i,t}}{3}$: Average rate of equipment ownership (%)
 - $rcons_s1_{i,t}$: Real private consumption (M€2005), source enerdata
3. Other control variables (services):
- $rva_s2_{i,t}$: Real value added of tertiary sector (M€2005), source enerdata
 - $empl_s2_{i,t}$: Employment of tertiary sector (thousand), source enerdata
4. Other control variables (industry):
- $rva_s3_{i,t}$: Real value added of industry (M€2005), source enerdata
 - $rginv_s3_{i,t}$: Real gross investment of industry (M€2005), source enerdata
5. Other control variables (transport):
- $cars_s4_{i,t}$: stock of cars (millions), source enerdata
 - $goods_s4_{i,t}$: traffic of goods (tkm), source enerdata
- $goods_s4_{i,t}$: traffic of goods (tkm), source enerdata Instrumental variables

The following variables are used as instruments for the policy indicator in Table 4:

- $ed_{i,t}$: energy dependence, source Eurostat. The indicator is calculated as energy imports minus energy export divided by the sum of gross inland energy consumption.²³ Energy dependence may be negative in the case of net exporter countries while positive values over 100% indicate the accumulation of stocks during the reference year.
- $ghge_{i,t}$: greenhouse gas emission intensity, source Eurostat and United Nations. The indicator is computed dividing greenhouse gas emissions by real GDP. Eurostat provides an index representing annual total emissions in relation to 1990 emissions; the absolute values in 1990 (and every 5 years) are provided by UN (the data appear to be coherent, since applying the growth rate derived from Eurostat index to UN 1990 starting points one gets almost exactly the subsequent UN values). The “Kyoto basket” of greenhouse gases includes: carbon dioxide (CO₂), methane (CH₄), nitrous oxide (N₂O), and the so-called F-gases (hydrofluorocarbons, perfluorocarbons and sulphur hexafluoride (SF₆)). These gases are aggregated into a single unit using gas-specific global warming potential (GWP) factors. The aggregated greenhouse gas emissions are expressed in units of CO₂ equivalents.

References

- Adeyemi, O.I., Hunt, L.C., 2007. Modelling OECD industrial energy demand: asymmetric price responses and energy-saving technical change. *Energy Econ.* 29, 693–709.
- Amrhein, V., Greenland, S., McShane, B.B., 2019. Retire statistical significance. *Nature* 567, 305–307.
- Angrist, J.D., Pischke, J.-S., 2009. *Mostly Harmless Econometrics: an Empiricist's Companion*. Princeton University Press.
- Apajalahti, E.-L., Lovio, R., Heiskanen, E., 2015. From demand side management (DSM) to energy efficiency services: a Finnish case study. *Energy Pol.* 81, 76–85.
- Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. Econ. Stud.* 58, 277–297.
- Athey, S., Imbens, G.W., 2017. The state of applied econometrics: causality and policy evaluation. *J. Econ. Perspect.* 31 (Number 2), 3–32.
- Aydin, E., Brounen, D., 2019. The impact of policy on residential energy consumption. *Energy* 169, 115–129. <https://doi.org/10.1016/j.energy.2018.12.030>.
- Balestra, P., Nerlove, M., 1966. Pooling cross section and time series data in the estimation of a dynamic model: the demand for natural gas. *Econometrica* 34, 585–612.
- Baltagi, B., 2005. *Econometric Analysis of Panel Data*, third ed. Wiley & Sons.

²³ Gross inland consumption is calculated as follows: primary production + recovered products + total imports + variations of stocks - total exports - bunkers. It corresponds to the addition of final consumption, distribution losses, transformation losses and statistical differences.

- Baltagi, B., Fomby, T., Carter Hill, R. (Eds.), 2001. Nonstationary Panels, Panel Cointegration, and Dynamic Panels. Emerald Group Publishing Limited, Bingley (Advances in Econometrics, Vol. vol. 15).
- Baum, C.F., Shaffer, M.E., Stillman, S., 2003. Instrumental variables and GMM: estimation and testing. *STATA J.* 3 (1), 1–31.
- Bertoldi, P., 2018. Editorial: ten-year anniversary. *Energy Efficiency* 11 (8), 1913–1915.
- Bertoldi, P., Boza-Kiss, B., 2017. Analysis of barriers and drivers for the development of the ESCO markets in Europe. *Energy Pol.* 107, 345–355.
- Bertoldi, P., Economidou, M., 2018. EU member states energy efficiency policies for the industrial sector based on the NEEAPs analysis. In: ECEEE Industrial Summer Study Proceedings, 2018-June, pp. 117–127.
- Bertoldi, P., Hirtl, B., 2013. Design of Database and Econometric Model to Assess the Energy Savings in Residential and Tertiary Gas and Electricity Consumption. JRC Scientific and Policy Reports (JRC78724 EUR 25795 EN - Institute for Energy and Transport).
- Bertoldi, P., Mosconi, R., 2015. The Impact of Energy Efficiency Policies on Energy Consumption in the EU Member States: a New Approach Based on Energy Policy Indicators. JRC Science for Policy Reports (JRC98236 EUR 27664 EN - Institute for Energy and Transport).
- Bigano, A., Arigoni Ortiz, R., Markandya, A., Menichetti, E., Pierfederici, R., 2011. The linkages between energy efficiency and security of energy supply in Europe. In: Galarraga, I., González-Eguino, M., Markandya, A. (Eds.), *Handbook of Sustainable Energy*. Edward Elgar Publishing Ltd., Northampton, pp. 60–83.
- Bjerregaard, J., Møller, N.F., 2019. The impact of EU's energy labeling policy: an econometric analysis of increased transparency in the market for cold appliances in Denmark. *Energy Pol.* 128, 891–899.
- Boonekamp, P.G.M., 2006. Evaluation of methods used to determine realized energy savings. *Energy Pol.* 34 (18), 3977–3992.
- Boonekamp, P.G.M., 2011. How much will the energy service directive contribute to the 20% EU energy savings goal? *Energy Efficiency* 4 (2), 285–301.
- Bukarica, V., Tomsic, Z., 2017. Design and evaluation of policy instruments for energy efficiency market. *IEEE Transactions on Sustainable Energy* 8 (1), 354–362.
- Bundgaard, A.M., Alberg Mosgaard, M., Remmen, A., 2017. From energy efficiency towards resource efficiency within the Ecodesign Directive. *J. Clean. Prod.* 144, 358–374.
- Burman, E., Mumovic, D., Kimpan, J., 2014. Towards measurement and verification of energy performance under the framework of the European directive for energy performance of buildings. *Energy* 77, 153–163.
- Cahill, C.J., Ó Gallachóir, B.P., 2010. Monitoring energy efficiency trends in European industry: which top-down method should be used? *Energy Pol.* 38, 6910–6918. <https://doi.org/10.1016/j.enpol.2010.07.006>.
- Carvalho, M.d.G., 2012. EU energy and climate change strategy. *Energy* 40 (1), 19–22.
- Cooper, A.C.G., 2018. Evaluating energy efficiency policy: understanding the 'energy policy epistemology' may explain the lack of demand for randomised controlled trials. *Energy Efficiency* 11, 997–1008.
- Dascalaki, E.G., Balaras, C.A., Gaglia, A.G., Droutsa, K.G., Kontoyiannidis, S., 2012. Energy performance of buildings-EPBD in Greece. *Energy Pol.* 45, 469–477.
- Datta, S., Filippini, M., 2016. Analysing the impact of ENERGY STAR rebate policies in the US. *Energy Efficiency* 9 (3), 677–698.
- deLlano-Paz, F., Martínez Fernández, P., Soares, I., 2016. Addressing 2030 EU policy framework for energy and climate: cost, risk and energy security issues. *Energy* 115 (2), 1347–1360.
- Eberhardt, M., Teal, F., 2010. Productivity Analysis in Global Manufacturing Production. Department of Economics, University of Oxford. Discussion Paper 515.
- Fawcett, T., Killip, G., 2019. Re-thinking energy efficiency in European policy: practitioners' use of 'multiple benefits' arguments. *J. Clean. Prod.* 210, 1171–1179.
- Fawcett, T., Rosenow, J., Bertoldi, P., 2019. Energy efficiency obligation schemes: their future in the EU. *Energy Efficiency* 12 (1), 57–71.
- Filippini, M., Hunt, L.C., 2012. US residential energy demand and energy efficiency: a stochastic demand frontier approach. *Energy Econ.* 34 (5), 1484–1491.
- Filippini, M., Zhang, L., 2019. Impacts of heat metering and efficiency retrofit policy on residential energy consumption in China. *Environ. Econ. Pol. Stud.* 21, 203–216. <https://doi.org/10.1007/s10018-018-0227-8>.
- Filippini, M., Hunt, L.C., Zoric, J., 2014. Impact of energy policy instruments on the estimated level of underlying energy efficiency in the EU residential sector. *Energy Pol.* 69, 73–81.
- Fuller, W., 1987. *Measurement Error Models*. Wiley & Sons.
- Giavazzi, F., Tabellini, G., 2005. Economic and political liberalizations. *J. Monetary Econ.* 52 (7), 1297–1330.
- Greene, W., 2012. *Econometric Analysis*, seventh ed. Prentice Hall.
- Gupta, M., Bandyopadhyay, K.R., Singh, S.K., 2019. Measuring effectiveness of carbon tax on Indian road passenger transport: a system dynamics approach. *Energy Econ.* 81, 341–354.
- Hall, V.B., 1986. Major OECD country industrial sector interfuel substitution estimates, 1969–79. *Energy Econ.* 8, 74–89.
- Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.
- Hausman, J., 1978. Specification tests in econometrics. *Econometrica* 46, 1251–1271.
- Haydt, G., Lealb, V., Dias, L., 2014. A multi-objective approach for developing national energy efficiency plans. *Energy Pol.* 67, 16–27.
- Hinchliffe, D., Akkerman, F., 2017. Assessing the review process of EU Ecodesign regulations. *J. Clean. Prod.* 168, 1603–1613.
- Horowitz, M.J., 2008. The trouble with energy efficiency indexes: la aritmetica non e opinione. *Energy Efficiency* 1, 199–210. <https://doi.org/10.1007/s12053-008-9015-9>.
- Horowitz, M.J., 2011. Measuring the savings from energy efficiency policies: a step beyond program evaluation. *Energy Efficiency* 4, 43–56.
- Horowitz, M.J., Bertoldi, P., 2015. A harmonized calculation model for transforming EU bottom-up energy efficiency indicators into empirical estimates of policy impacts. *Energy Econ.* 51 (September 2015), 135–148.
- Hsiao, C., Pesaran, M.H., 2008. Random coefficient models. In: Matyas, L., Sevestre, P. (Eds.), *The Econometrics of Panel Data Fundamentals and Recent Developments in Theory and Practice*, third ed. Springer Verlag.
- Hull, D., Ó Gallachóir, B.P., Walker, N., 2009. Development of a modelling framework in response to new European energy-efficiency regulatory obligations: the Irish experience. *Energy Pol.* 37 (Issue 12), 5363–5375.
- IPCC, 2018. *Global Warming of 1.5 °C. The Intergovernmental Panel on Climate Change Reports*. <https://www.ipcc.ch/sr15/>.
- Keane, M.P., 2010. Structural vs. Atheoretic approaches to econometrics. *J. Econom.* 156 (1), 3–20.
- Khayyat, N.T., 2015. *Energy Demand in Industry: what Factors Are Important*. Springer Verlag.
- Laes, E., Mayeres, I., Renders, N., Valkering, P., Verbeke, S., 2018. How do policies help to increase the uptake of carbon reduction measures in the EU residential sector? Evidence from recent studies. *Renew. Sustain. Energy Rev.* 94, 234–250.
- Lang, C., Siler, M., 2013. Engineering estimates versus impact evaluation of energy efficiency projects: regression discontinuity evidence from a case study. *Energy Pol.* 61, 360–370.
- Lin, Boqiang, Zhu, Junpeng, 2019. Is the implementation of energy saving and emission reduction policy really effective in Chinese cities? A policy evaluation perspective. *J. Clean. Prod.* 220, 1111–1120.
- Madlener, R., Bernstein, R., Alva González, M.Á., 2011. *Econometric Estimation of Energy Demand Elasticities*, E.ON Energy Research Center Series, vol. 3. Issue 8, October (ISSN: 1868-7415).
- Malinauskaitė, J., Jouhara, H., Ahmad, L., Milani, M., Montorsi, L., Venturini, M., 2019. Energy efficiency in industry: EU and national policies in Italy and the UK. *Energy* 172, 255–269.
- Mallaburn, P.S., Eyre, N., 2014. Lessons from energy efficiency policy and programmes in the UK from 1973 to 2013. *Energy Efficiency* 7, 32–41.
- McShane, B.B., Gal, D., 2017. Statistical Significance and the Dichotomization of Evidence. *J. Am. Stat. Assoc.* 112 (519), 885–895. <https://doi.org/10.1080/01621459.2017.1289846>.
- Meng, T., Hsu, D., Han, A., 2017. Estimating energy savings from benchmarking policies in New York City. *Energy* 133, 415–423.
- Nabitz, L., Hirzel, S., 2019. Transposing the requirements of the energy efficiency directive on mandatory energy audits for large companies: a policy-cycle-based review of the national implementation in the EU-28 member states. *Energy Pol.* 125, 548–561.
- Ó Broin, E., Nässén, J., Johnsson, F., 2015. Energy efficiency policies for space heating in EU countries: a panel data analysis for the period 1990–2010. *Appl. Energy* 150, 211–223.
- Ogata, K., 1995. *Discrete Time Control Systems*, second ed. Prentice-Hall.
- Papaioannou, E., Siourounis, G., 2004. *Democratization and Growth*. London Business School Economics Working Paper. <https://doi.org/10.2139/ssrn.564981>.
- Persson, T., 2005. *Forms of Democracy, Policy and Economic Development*. NBER Working paper 11171.
- Persson, T., Tabellini, G., 2006. Democracy and development: the devil in the details. *Am. Econ. Rev.* 96 (2), 319–324.
- Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74, 967–1012.
- Reichl, J., Kollmann, A., 2010. Strategic homogenisation of energy efficiency measures: an approach to improve the efficiency and reduce the costs of the quantification of energy savings. *Energy Efficiency* 3 (3), 189–201, 2010.
- Reichl, J., Kollmann, A., 2011. The baseline in bottom-up energy efficiency and saving calculations – a concept for its formalisation and a discussion of relevant options. *Appl. Energy* 88 (2), 422–431, 2011.
- Rezessy, S., Bertoldi, P., 2011. Voluntary agreements in the field of energy efficiency and emission reduction: review and analysis of experiences in the European Union. *Energy Pol.* 39 (11), 7121–7129.
- Ringel, M., Knodt, M., 2018. The governance of the European energy union: efficiency, effectiveness and acceptance of the winter package 2016. *Energy Pol.* 112, 209–220.
- Rodrik, D., Wacziarg, R., 2005. Do democratic transitions produce bad economic outcomes? *Am. Econ. Rev.* 95 (2), 50–56.
- Roodman, D., 2008. Through the looking glass, and what OLS found there: on growth, foreign aid, and reverse causality. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1100142>.
- Roodman, D., 2009. How to do xtabond2: an introduction to difference and system GMM in Stata. *STATA J.* 9 (1), 86–136.
- Rosenow, J., Cowart, R., Bayer, E., Fabbri, M., 2017. Assessing the European Union's energy efficiency policy: will the winter package deliver on 'Efficiency First'? *Energy Research & Social Science* 26, 72–79.
- Saussay, A., Saheb, Y., Quirion, P., 2012. The impact of building energy codes on the energy efficiency of residential space heating in European countries – a stochastic frontier approach. In: *International Energy Program Evaluation Conference*, 12–14 June 2012, Rome.
- Schlomann, B., Schleich, J., 2015. Adoption of low-cost energy efficiency measures in the tertiary sector—an empirical analysis based on energy survey data. *Renew. Sustain. Energy Rev.* 43, 1127–1133.
- Schlomann, B., Eichhammer, W., Reuter, M., 2016. *Internal Guidelines for the Measure Descriptions in the MURE Database*. Version 5.1. September 2016.

- Schreiber, S., 2008. The hausman test statistic can be negative even asymptotically. *Jahrb. Natl. Stat.* 228 (4), 394–405.
- Stern, D.I., 2009. Interfuel Substitution: A Meta-Analysis. MPRA papers 15792. <https://mpra.ub.uni-muenchen.de/15792/>.
- Streimikiene, D., Volochovic, A., Simanaviciene, Z., 2012. Comparative assessment of policies targeting energy use efficiency in Lithuania. *Renew. Sustain. Energy Rev.* 16 (6), 3613–3620.
- Thiel, C., Nijs, W., Simoes, S., Schmidt, J., van Zyl, A., Schmid, E., 2016. The impact of the EU car CO₂ regulation on the energy system and the role of electro-mobility to achieve transport decarbonisation. *Energy Pol.* 96, 153–166.
- Thomas, S., Boonekamp, P., Vreuls, H., Broc, J.-S., Bosseboeuf, D., Lapillonne, B., Labanca, N., 2012. How to measure the overall energy savings linked to policies and energy services at the national level? *Energy Efficiency* 5, 19–35. <https://doi.org/10.1007/s12053-011-9122-x>.
- Tsemekidi Tzeiranaki, S., Bertoldi, P., Diluiso, F., Castellazzi, L., Economidou, M., Labanca, N., Ribeiro Serrenho, T., Zangheri, P., 2019. Analysis of the EU residential energy consumption: trends and determinants. *Energies* 12, 1065. <https://doi.org/10.3390/en12061065>.
- Urga, G., Walters, C., 2003. Dynamic translog and linear logit models: a factor demand analysis of interfuel substitution in US industrial energy demand. *Energy Econ.* 25, 1–21.
- Wasserstein, R.L., Lazar, N.A., 2016. The ASA's statement on p-values: context, process, and purpose. *Am. Statistician* 70 (2), 129–133.
- Wintoki, B., Linck, J.S., Netter, J.M., 2012. Endogeneity and the dynamics of internal corporate governance. *J. Financ. Econ.* 105 (3), 581–606.
- Wooldridge, J.M., 2002. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge.
- Yearwood-Travezan, J., Harmsen, R., van Toledo, G., 2013. Policy analysis for energy efficiency in the built environment in Spain. *Energy Pol.* 61, 317–326.
- Zangheri, P., Serrenho, T., Bertoldi, P., 2019. Energy savings from feedback systems: a meta-studies' review. *Energies* 12, 3788. <https://doi.org/10.3390/en12193788>.