Title page

Environmental policy performance and its determinants: a application of a three-level random intercept model *

Marzio Galeotti[a,b], Yana Rubashkina[c], Silvia Salini[a], Elena Verdolini[d]

[a] *DEMM, Università degli Studi di Milano*
[b] *IEFE, Università Bocconi, Milano*
[c] *DEFAP, Università Cattolica del Sacro Cuore, Milano*
[d] *Fondazione CMCC and Fondazione Eni Enrico Mattei*

Abstract: This paper contributes to the study of environmental and energy policy by using a three-level random intercept (TLRI) model to rank the performance of different countries. Inspired by the literature on Item Response Theory and multilevel latent models, the TLRI model treats policy commitment as a latent variable which is estimated conditional on the difficulty of the policy portfolio implemented by each country. This approach is characterized by three novel aspects. First, the model results in a ranking of countries which is conditional on the complexity of their chosen policy portfolio. Second, it provides a unified framework in which to construct a policy indicator and to study its determinants through a latent regression approach. The resulting country ranking can thus be cleaned from the effect of economic and institutional observables which affect policy design and implementation. Third, the model estimates parameters which can be used to describe and compare policy portfolios across countries. We apply this methodology to the case of energy efficiency policies in the industrial sectors of 29 EU countries between 2004 and 2011. In the conclusions we highlight the future possible applications of this approach, which are not confined to the realm of environmental and energy policy.

Keywords: Energy policy, environmental policy, ranking, policy portfolios
JEL codes: Q58, O57, C33

Acknowledgements: Silvia Salini gratefully acknowledges funding from the MIUR PRIN project MISURA. Elena Verdolini from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 308481 (ENTRACTE) and the Marzio Galeotti from the MIUR PRIN 2010-11 Project "Climate changes in the Mediterranean area: evolutionary scenarios, mitigation policies and technological innovation"**.**

Email addresses: silvia.salini@unimi.it (Silvia Salini), elena.verdolini@feem.it (Elena Verdolini), yrubashkina@gmail.com (Yana Rubashkina), marzio.galeotti@unimi.it (Marzio Galeotti)

Corresponding author: Marzio Galeotti, Department of Economics, Management and Quantitative Methods, via Conservatorio 7, I-20122 Milano, Italy. E-mail: marzio.galeotti@unimi.it

***Highlights**

- Three-level random intercept (TLRI) model to rank the performance of different countries
- TLRI model treats policy commitment as a latent variable
- Estimated conditional on the difficulty of the policy portfolio implemented by each country
- Application to energy efficiency policies in the industrial sectors of 29 EU countries 2004-2011

1. Introduction

The COP21 Conference in Paris gave new impetus to efforts towards limiting greenhouse gas emissions (GHGs). As testified by the national pledges and the signing of the Paris Treaty, many countries committed to implementing policies supporting sustainable development through the promotion of renewable energy sources and increased energy efficiency. Indeed, European countries have been at the forefront of fighting climate change. For instance, cleaner energy is one of the five objectives of Europe 2020, the sustainable growth strategy that EU member states launched in 2010 as a response to the recent global economic crisis.[1]

In light of this renewed commitment, a major challenge for researchers and policy makers alike is the assessment of past energy and environmental policies, and specifically how countries are performing in this respect. This question is important for both policy evaluation and for research purposes.

First, appropriately describing and understanding the past performance of countries with respect to energy and environmental policies, and their ability to commit to a more or less complex portfolios of policy instruments, is a crucial step in ensuring that future interventions are drafted in a sound and cost-effective way. An in-depth analysis in this respect is currently missing due to lack of appropriate data and to more complex conceptual problems linked with the creation of appropriate indicators.

Second, the availability of sound indicators of policy commitment and stringency would allow for more solid empirical research on the inducement effects of such

[1] The Europe 2020 strategy includes five main objectives: ensuring 75 % employment of 20/64-year-old; Getting 3% of the EUs GDP invested into research and development; limiting greenhouse gas emissions by 20 % or even 30 % compared to 1990 levels, creating 20 % of EU energy needs from renewables and increasing energy efficiency by 20 %; reducing school dropout rates to below 10 %, with at least 40 % of 30/34-year-old completing tertiary education; ensuring 20 million fewer people are at risk of poverty or social exclusion. The environment and energy objective summarizes the so-called 20-20 Climate and Energy Package approved in 2007 by the EC and subsequently translated into a set of five directives approved in 2009.

policies for innovation, competitiveness and economic performance more in general. Indeed, the poor quality of available indicators is often cited as one of the major shortcomings of the empirical literature dealing with such research questions.

As pointed out in Brunel and Levinson (2013, 2016), assessments of environmental and energy policy are characterized by major challenges. First, to address climate and energy concerns countries can choose from a wide array of policy instruments, each of which is characterized by a different level of effectiveness, dynamic efficiency and political acceptability (Fisher and Newell, 2008). This “multidimensionality” translates into the challenge of building a policy indicator able to capture the different aspects of a country’s policy portfolio. Second, the ability of countries to implement certain (lower cost) options might depend crucially on some “initial condition” or on some time varying characteristics. For instance, the complexity and stringency of a country’s policy portfolio at any given point in time is likely higher for those countries which have been committed to sustainable energy for a longer period of time. While these countries may appear has having an overall higher score, their efforts over a given period of time may be lower than that of countries which only recently committed to GHG reductions and energy efficiency. On the one hand, latecomers to climate mitigations may be asked to bridge the gap in environmental protection very swiftly upon joining international organizations (such as the OECD or the EU). On the other hand, forerunners in environmental protection may have already reaped the low hanging fruits, and any increase in policy commitment and stringency may be particularly difficult due to higher marginal costs in terms of economic performance or political and social support.

This state of affairs makes it hard to build a comprehensive indicator to assess countries’ performance in the realm of environmental and energy policy. In addition, the data available to the researcher is poor to begin with, as even collecting information on the stringency of all the different policy instruments put in place is challenging. Actually, data in this respect is scarce or imperfect. Most databases only provide a count of the different policy interventions implemented in any country, often categorizing them by policy instrument, as in MURE (2012) or IEA (2012). The few attempts to provide qualitative scores together with counts of policies have been criticized since they rest on assessments by experts, which are often perceived as arbitrary. Indeed, to date the efforts to produce environmental policy indexes for a large number of countries and for long time frames has been severely limited by lack of data.

This paper is a methodological contribution aimed at showing the potential of a model, which has been largely applied in statistics, for the field of energy and environmental policy assessment. We propose a novel approach to score countries with respect to their commitment to environmental and energy policy. Recognizing the fundamental challenges characterizing data on energy and environmental policy (Brunel and Levinson, 2013; Nesta, Vona, and Nicolli, 2014; Galeotti, Salini, and Verdolini, 2017), we show how a three level random intercept (TLRI) model inspired by Item Response Theory (IRT) can be of help to score the policy performance of different countries in a given sector. This model allows using the count of policy instruments by type, active in a given country in a given year, to characterize the complexity of a country's policy portfolio and its level of policy commitment.

To illustrate the potential of this approach, as well as its limitations, we estimate the model using data on policies promoting energy efficiency in the industrial sector in a sample of 29 European countries over the years 2004-2011. The contribution of our analysis to the literature is fourfold. First, our approach has relatively few data requirements and allows exploring the scarce information available on environmental and energy policies to the fullest. Second, the TLRI model allows building an index to assess and compare countries' environmental policy portfolios and performance addressing the aforementioned issue multidimensionality. The score we build accounts for the type, number and complexity of the policy instruments implemented in each of the countries in our sample. Third, the TLRI model can be augmented with a latent regression. This allows to condition the “raw” country score on specific observables at the beginning and during our sample period, thus addressing the problem of “initial conditions” noted above. Fourth, our methodology provides a unified framework to rate policy commitment and stringency (through a three level random intercept model) and to study its determinants (through a latent regression). It is therefore of potential relevance also for applications on a variety of research questions where a key requirement is the creation of a policy indicator cleaned from reverse causality and from the effect of covariates.

The rest of this paper is organized as follows. Section 2 provides a review of the available literature and highlights the contributions of this paper. Section 3 presents the proposed statistical model. Its empirical application, which focuses on energy efficiency policies in Europe, is presented in Section 4. We describe therein the data and report the empirical results, which include country rankings which account for (a) the complexity of the policy mix put into place and (b) the effect of economic and institutional observables. Section 5 concludes with a summary of main results, policy implications and a list of future research avenues.

2. Literature Review

Assessing the economic impact of policy decisions is of central interest to Economics. As environmental and energy policy has become increasingly active worldwide in the last decades, several efforts were undertaken to ascertain the consequences of decisions concerning energy efficiency, renewable energy sources, emission reductions, and the like, on key variables such as innovation activity, economic growth or overall economic performance. A critical issue is of course the definition of an appropriate indicator of policy commitment and stringency. This is a topic that has received recently increasing attention. Brunel and Levinson (2013) provide a comprehensive review of the literature in this respect.

Popular proxies for regulatory stringency are data on private sector abatement expenditures (Pollution Abatement Costs, or PACs). Such data inform on the level of financial effort a given firm/sector has to face to comply with given standards (Lanjouw and Mody, 1996; Jaffe and Palmer, 1997; Berman and Bui, 2001; Hamamoto, 2006; Rubashkina et al., 2014).[2] Reductions in emissions or pollutants or indicators based on energy use are other popular indicators of choice (Cole and Elliot, 2003; Gollop and Roberts, 1983). Changes in regulation-based measures have also been used to judge the level of policy stringency (Popp, 2003, 2010). A different tack has been taken by the numerous papers which made use of general composite indexes through the use of aggregation techniques. The data used to that end include information on the presence or absence of a given policy (0-1 indicators) or on scores from surveys of government officials or business leaders (Tobey, 1990; Kellenberg, 2009). Finally, many have resorted to *ad hoc* data sets which are tailored to answering a specific research question (Jeppsen and Folmer, 2001).

Brunel and Levinson (2013) nicely describe the main conceptual issues that plague almost all previous efforts to create an index of energy and environmental policy stringency.

First, creating a reliable indicator is challenging due to the issue of "multidimensionality". Governments regulate various aspects of energy production and environmental protection, namely air, water, toxic chemicals, but also energy efficiency

[2] The use of this indicator is based on the assumption that profit maximizing firms typically face marginal abatement costs that are increasing in pollution abatement.

and renewable energy production. Moreover, policy instruments can be aimed at regulating pollution directly, through either a command-and-control or a market-based approach. In addition, environmental and energy policies *per se* can be combined with policies aimed at addressing the knowledge market failure and can stimulate the creation and diffusion of less polluting technologies.[3] Such heterogeneity in policy responses and in the sectors targeted makes it hard to build an indicator that is at the same time comprehensive and detailed enough to capture changes in all different aspects of a country's policy portfolio.

Second, while policy makers and researchers ideally would want to measure the effect of policy on other important outcome variables such as industry location, trade patterns, economic growth or knowledge transfer, the variables measuring the stringency of environmental regulation are plagued by simultaneity and endogeneity.[4] One must therefore bear in mind that policies are often jointly determined with other outcome variables and that they themselves are not exogenous, but are the result of forces within the economic system.

Finally, some countries/sectors might have a "comparative" advantage with respect to others in implementing strict environmental policy. This might be due to their industrial composition, but also to the vintage of capital, to the fact that they are more polluting to begin with, or to the fact that they are latecomers in environmental and energy regulation. This gives them the option to implement more easily low-cost high-

[3] Environmental (and energy) policy directly targets the environmental externality by regulating pollutants or emissions. On the one hand, command-and-control policy instruments include mandates and standards, which set a minimum requirement for firms to comply with. On the other hand, market-based approaches such as taxes and permits allow firms to respond more flexibly to comply with the regulation. Conversely, technology policy targets the knowledge market failure and supports R&D in cleaner and more efficient technologies with, among other options, research subsidies and investments. See, for instance, Perman et al. (2011).

[4] An example of simultaneity relates to the fact that pollution-intensive industries may have more lobbying power the greater their share of a country's economy, they may pressure their governments to enact less stringent regulations. See the discussion in Brunel and Levinson (2013) and the references therein.

reduction (or high-efficiency improvement) policies.

Indeed, all the indicators proposed in the literature so far suffer from one or more of these shortcomings. For instance, PACs are plagued with reverse causality issues and in the presence of market or behavioral failures they no longer successfully measure the level of regulatory pressure (Berman and Bui, 2001).[5] Emission or energy-use based indicators are also likely to mirror changes other than regulatory stringency, such as for example factor prices. Moreover, when used at the disaggregated level, it is often hard to build indicators that can be used in cross sectoral or cross-country analyses due to the heterogeneity of the regulated pollutants. Proxies based on normative prescriptions do not account for the level of actual enforcement of a given policy and might also be subject to issues of reverse causality (Brunnermeier and Cohen, 2003; Shimshack and Ward, 2005). Finally, most composite indicators are built using simple approaches such as the sum of policy instruments.

Among the most recent attempts to overcome some of these shortcomings are statistical aggregation techniques. Nicolli and Vona (2012) propose two different aggregate indicators to measure the level of renewable environmental policies in European countries. First, an average-based indicator which uses information on the timing of adoption of a given policy instrument (namely, an average of dummy variables indicators equal to zero before the instrument is put into place and equal to one after- wards). Second, a more complex indicator is built using principal component analysis (PCA) relying both on dummy variables and on intensity of specific policy instruments such as Renewable Energy Certificates or Feed-in Tariffs. The factor loadings resulting from the PCA in Nicolli and Vona (2012) can be interpreted as

[5] Moreover, if the data are used at the aggregate level, such as sectors or countries, changes in PACs might result from changes due to unobserved heterogeneity rather than from changes in regulatory stringency.

importance weights which vary by item/policy.[6]

Another composite index is the OECD EPS composite indicator (Botta and Koźluk, 2014). It relies on a recent database produced by the OECD which includes fifteen continuous policy indicators which are then categorized as a Likert scale from 0 to 6 by identifying specific bins. These fifteen Likert-scale scores are then aggregated into 6 large macro-instruments: Taxes, Certificates, Limits, Feed-in Tariffs (FIT), Deposit Refund Schemes (DRS), and R&D by using weights. Subsequently, these six indicators are aggregated into a Market-Based (MB) score (Taxes, Certificates, FIT, DRS) and a Non Market-Based (NMB) score (R&D and Standards). The EPS Composite score is then obtained as the average between the MB and NMB scores.

One last composite indicator worth mentioning is the Index of Climate Policy Action (Schaffrin, Sewerin, and Seubert, 2015). This index combines density and intensity of the policy portfolio, with the former accounting for the number of policy instruments of the portfolio, and the latter providing information on the content of those instruments.. Here a weighting scheme is used for the following intensity measures: objectives, scope, integration, budget, implementation, and monitoring. The coding of these measures is based on expert judgement and the index is computed for the energy supply sector of Austria, Germany, and the U.K. over the period 1998-2010.

In this paper, we propose to use a TLRI mode, which is inspired by Item Response Theory (IRT) models which are widely applied in the statistical literature of scoring. To illustrate the potential of this novel approach, which is specifically designed to deal with the presence of several policy instruments as well as with longitudinal data, we apply it to the scoring of policies supporting energy efficiency in Europe. Finally, we show how

[6] In their approach, however, PCA is built using binary indicators of the presence/absence of a given policy instrument and no consideration is given to the number of policies in place at any given time. Moreover, Ferrari and Salini (2011) hold that the presence of dummy variables should require Categorical Principal Components Analysis (CATPCA) which is not based on the assumption of linear correlation.

TLRI models can be used to produce an indicator of policy commitment which is clean from the effect of observables affecting both policy instrument choice and implementation.

3. Methodology

A proper understanding of the empirical approach proposed in this paper requires a short digression on Rasch models, which are a particular class of Item Response Theory models (Rasch, 1980; van der Linden and Hambleton 1997; Baker and Kim 2004). IRT models were developed as a psychometric tool in social sciences to compare the performance of various subjects in questionnaires/tests, i.e. to characterize individual ability along a continuum. IRT models rest on the underlying hypothesis that the phenomenon to be measured represents a latent factor, namely something that cannot be observed directly, but only measured indirectly by many variables whose categories represent different aspects and/or levels of the latent dimension.

The application of IRT models was then extended beyond psychometrics to all those applications in which the variable of interest cannot be measured with conventional means, but rather quantified by assuming a latent variable, such as intelligence, mathematical or verbal ability, racial prejudice, political attitude, consumer preferences. Hence, IRT models found applications in fields such as psychology, education, sociology, marketing, and medicine, among others. Thus, for instance, Bacci (2012) and Bacci and Bartolucci (2012) apply IRT to the scoring of quality of life, Bacci and Caviezel (2011) use it to score teaching evaluation and Gnaldi et al. (2016) assess students' acquired skills and cluster students according to their ability level. These models have found application also in organizational and management studies in particular for financial issues (Soutar and Cronish-Ward, 1997), marketing and consumer behavior (Fischer et al., 2006; Salzberg and Sinkovics, 2006) and tourism management (Oreja-Rodriguez and Yanes-Estevez, 2007). Ferrari et al. (2005) explore the validity and constraints of this approach as a tool to quantify the degree of vulnerability of historical-architectonic buildings in Northern Italy. Finally, Murray and Mills (2012) apply a similar methodology to the scoring of energy insecurity of households in the United States.

Among the several classes of IRT models, the Rasch model (1980) is a statistical model for dichotomous data in which the probability of observing a positive/correct response to each item/question by each individual is modeled as a decreasing function of the item's "difficulty" ("complexity") and as an increasing function of the subject's "ability".[7] As in other IRT models, both these variables are modeled as latent traits and estimated using data on how each individual performs with respect to a given item. Specifically, the probability of a correct response by individual j (j = 1,2,…J) on item i (i = 1,2,…I) , $P(Y_{ij} = 1)$, is a function of the two latent traits, namely it is increasing in the respondent's ability (θ_j)and decreasing in the item difficulty (β_i):

$$(1) \qquad P(Y_{ij} = 1|\theta_j, \beta_i) = \frac{\exp[(\theta_j - \beta_i)]}{1+\exp[(\theta_j - \beta_i)]}$$

By using information on whether a given subject replied correctly to a given question, the Rasch model (RM) is able to estimate the latent traits θ and β.

The original RM built for dichotomous dependent variables has been extended along several dimensions. Firstly, polytomous models have been developed to deal with ordinal data (Mair and Hatzinger, 2007). To this end, the RM model is modified to

[7] In the Rasch model context, the definition of what an "item" represents depends on the focus of the analysis. In those analyses scoring pupils performance, each item can for instance represent a question within a larger questionnaire. In assessments of either physical or psychological wellbeing, an "item" is a particular set of characteristics (for instance, fatigue, ability to concentrate, etc.). In the analysis of Murray and Mills (2012) each item is a different question focusing on ability to pay for electricity bill, access to electricity, etc. As explained more in detail below, in this analysis an item is defined as one specific policy type and a subject is a given country (see below).

estimate k "thresholds" for the m categories of data (for example, with four categories ranging from 0 to 4, the model estimates three thresholds: 0-1, 1-2, 2-3). The "thresholds" are defined as those points in which two adjacent items scores have the same probability of being observed in the specific response under consideration. Within the class of polytomous RMs, in Rating Scale Models (Andersen, 1997) the threshold values are assumed to be equal across all items/questions (namely, each of the item is scored on the same number of categories), even though the distance between two thresholds can differ (Andrich, 1978). In more complex Partial Credit Models (PCM) (Masters and Wright, 1997) the estimated thresholds are allowed to differ also by item/question (namely, each of the items can have a different number of categories). The PCM logistic model reads as follows:

(2) $$P\left(Y_{ij} = y \middle| \theta_j, \tau_{ik}\right) = \frac{\exp\,[\sum_{k=0}^{y}(\theta_j - \tau_{ik})]}{\sum_{y=0}^{m}\exp\,[\sum_{k=0}^{y}(\theta_j - \tau_{ik})]}$$

where $k=0,...,m-1$ indicate the number of thresholds, $y=0,\ldots,m$ indicate the response categories, and τ_{ik} denotes the item difficulty and threshold parameter jointly for each item.

Secondly, Longitudinal RMs have been developed to handle panel data observations, as opposed to cross sections, to study changes in "ability" scores over time (Fischer, 1989).

Thirdly, RMs have been augmented with a latent regression in order to study the determinants of a subject's ability (De Boeck and Wilson, 2004). This allows conditioning the individual's estimated "ability" parameter on a set of observables, thus cleansing the parameter itself from the effect of any socio-economic characteristics so that it truly measures a latent trait.

RMs have been applied in a variety of fields with the aim of scoring performance along a specified dimension, as noted above. Here, we apply a modified version of the RM to score the performance of countries with respect to environmental and energy policy. There are three main features of RMs which are of particular relevance in this respect, and which have not been exploited by any previous study. Firstly, these models were created to handle the treatment of many interrelated variables with the aim of summarizing data and highlighting possible latent factors, scoring subjects along a continuum. Secondly, the data requirements to estimate such models are relatively limited: in theory, it is possible to estimate "item scores" and "ability scores" just knowing whether a country has implemented or not a specific policy instrument (dichotomous model), or the number of regulations for each policy instrument which were implemented (polytomous model). Thirdly, it is possible to create a model which allows to handle longitudinal data as well as to condition the estimated "ability" scores on observables which are likely to affect such scores through the use of a latent regression. Hence, RMs can potentially address the main shortcomings noted before linked with the creation of indexes scoring environmental and energy policy: multidimensionality, need for a time-varying indicator, and issue linked with controlling for differences in "initial conditions".

The Three-Level Random Intercept (TLRI) model we propose here is a refinement of RMs and is a 3-level application of multilevel models which are appropriate for research designs where data are organized in more than one level (Goldstein, 2011).[8] Specifically, it is an ordinal logistic model for adjacent item scores following Bacci and Caviezel (2011) which is augmented by latent regressions as in de Boek and Wilson (2004).

[8] While in this paper we apply a multi-level model to the scoring of countries' environmental and energy policy, the typical example of multilevel models is a two level model of test scores by pupils (lower level) nested within classes (higher level) See Goldstein (2011) for a thorough discussion of such models

In our TLRI model, a country corresponds to a "subject", while different types of policy instruments are modelled as different "items". The score for each item is defined as the number of regulations active in a given country in a given year for each specific policy category. Given that the total scores on each item differ by policy instrument (as explained above), the number of categories for each item in our TLRI are allowed to be different by item. This is similar to the PCM framework discussed above.

By construction, the TLRI the model clusters data in three levels: item scores for each item (environmental policy instrument) and subject (country) are clustered for each year, and all the item scores for each year and each subject are clustered by subject. Hence, the three levels assumed in the model are item (i.e., policy instrument -- first level), year (second level) and subject (i.e. country -- third level). The model reads as follows:

(3) $$P\left(Y_{ij} = y \middle| \theta_{0tj}, \theta_{00j}\right) = \frac{\exp\{\sum_{k=0}^{y}[(\theta_{0tj}+\theta_{00j}+(\beta_i-\tau_{ik})]\}}{\sum_{y=0}^{m}\exp\{\sum_{k=0}^{y}[(\theta_{0tj}+\theta_{00j}+(\beta_i-\tau_{ik})]\}}$$

where $P(.)$ is the probability that the score Y_{itj} on item/policy i ($i = 1, \ldots, I$) in time t ($t = 1,\ldots,T$) for country j ($j = 1, \ldots, J$) assumes a given value y, where y is the number of regulations for the given policy category. The parameter βi describes the average difficulty of the i-th policy category (i.e. item). The parameter τ_{ik} indicates the different threshold in each item/policy score. The two random effects θ_{0tj} and θ_{00j} are the latent variables measuring a subject's ability/performance (i.e. they replace the θ parameter in the RMs described above). θ_{0tj} and θ_{00j} are both obtained as the expected a posteriori (empirical) Bayes predictions, namely the posterior distributions of the country parameters given the policy responses. θ_{0tj} is the second level residuals and indicates the deviation of the latent variable θ for year t and country j from the average value of country j: accordingly, they allow for an analysis of time within each country. θ_{00j} is the third-level residuals indicate the deviation of the latent variable for county j from the average value of the population: they thus allow for a ranking of countries in terms of the mean level of ability/performance. θ_{0tj} and θ_{00j} are independent and normally distributed random variables with zero means and constant variances.

τ_{ik}, θ_{0tj} and θ_{00j} are the parameters of interest for our analysis. The threshold parameters τ_{ik}, vary across different policy categories (items) and can be used to characterize the level of complexity/difficulty of the policy portfolio implemented in each year. The second level residuals, θ_{0tj}, allow tracking the performance of each country throughout the sample period with respect to its average. Finally, the third-level residuals θ_{00j} represent the source of information of greatest use in our case as they allow the ranking of countries with respect to their observed composite indicator of environmental and energy policies over the sample period. Note that these parameters are obtained conditional on the complexity of the policy portfolio in each year and on the overall commitment of each country with respect to policy implementation. Comparing the θ_{00j} for the different countries allows us to characterize how they perform in terms energy and environmental policy.

The TLRI model provides us with: (i) threshold parameters measuring the intrinsic difficulty/probability of observing a given categorical response for each item/policy instrument (τ_{ik}); (ii) time-country specific intercepts for each country over time (θ_{0tj}) and (iii) country-specific parameters which allow for an overall country ranking in the period under consideration (θ_{00j}). The elements (ii) and (iii) are derived conditional on the policy category difficulty levels.

Following de Boeck and Wilson (2004), the TLRI can be augmented with two latent regressions, one for θ_{0tj} and one for θ_{00j}. This allows conditioning the estimation of these two latent traits on certain observable variables which are likely to affect the score/performance of countries in each year and on average. Indeed, the ability of each country to implement, say, energy efficiency policies may be affected by institutional and economic characteristics. For example, certain countries could rank highest because they have better starting conditions (for instance, they might have more room to phase out old capital equipment because they have a higher GDP). We thus include the vectors of covariates X_j for the third-level residuals θ_{00j} for each country j and a vector of covariates Z_{tj} for the second level residuals θ_{0tj} for each year t and for each country j, respectively as:

(4) $$\theta_{00j} = X_j'\gamma + \epsilon_{00j}$$

and:

(5) $$\theta_{0tj} = Z_{ti}'\delta + \epsilon_{0tj}$$

The model (3) hence becomes:

(6) $$P\left(Y_{ij} = y \middle| \theta_{0tj}, \theta_{00j}\right) = \frac{\exp\left\{\sum_{k=0}^{y}\left[\left(X_j'\gamma + Z_{ti}'\delta + \epsilon_{00j} + \epsilon_{0tj} + (\beta_i - \tau_{ik}\right)\right]\right\}}{\sum_{y=0}^{m}\exp\left\{\sum_{k=0}^{y}\left[\left(X_j'\gamma + Z_{ti}'\delta + \epsilon_{00j} + \epsilon_{0tj} + (\beta_i - \tau_{ik}\right)\right]\right\}}$$

By augmenting the TLRI model with latent regressions, the parameters of interest become the residual components ε_{0tj} and ε_{00j}. These residuals are cleaned from the "comparative advantage" effect or any causality between the policy indicator and the covariates. To illustrate the usefulness of this proposed approach, the next Section discusses the application of the TLRI model both with and without latent regressions, to the scoring of energy efficiency policy commitment in the EU28 members and Norway in the manufacturing sector.

4. Empirical Application

To better understand the potential of the model we described above, we use it to study the performance of the 28 EU Member States plus Norway with respect to policies promoting the rational use of energy and end-use renewables in the manufacturing sector over the years 2004-2011. The choice of time window and countries was determined by data availability, as explained below.

4.1. Data and Estimation Method

The data, which consists of a set of policy indicators, are extracted from the MURE (2012) database. We focus on policies for energy efficiency in the manufacturing sector because these are expected to provide a significant contribution to climate change mitigation. The time spell of the analysis is determined by the fact that in the years 2004-2011 the MURE database contains information for all EU28 countries plus Norway.[9] The MURE database includes the national policies targeting energy efficiency that have macro-economic impact, imposing a quality threshold which eliminates low-impact policies. The database provides information on five different environmental policy categories:[10]

[9] The MURE database contains detailed country fact sheets reflecting the adoption time of energy efficiency policy measures in 28 EU Member States plus Norway over the period 1993-2011. Data is available from 1993 for 16 (Old Member) countries, and from 2004 for the other 13 countries.

[10] In theory, we could also consider an additional "item" (policy instrument), namely those policies which support innovation in energy efficient technologies through, for example, R&D investment or incentives. We chose not to do so for two reasons. First, the timing of the impact of innovation policies is very different, as they work on a longer time frame than policies directly regulating the environmental externality. Their impact is thus likely to occur with a lag because the invention of new technologies requires time. Second, given this difference in timing, the availability of novel improved technologies is in fact an enabling factor that would make the implementation of stringent environmental policies more likely. Setting stringent performance standards or costly taxes on energy use is more politically and socially feasible if highly efficient equipment is already available to firms and consumers to purchase.

1. *Regulatory policies* include norms and standards, such as energy efficiency levels for various kinds of equipment and production processes or products, which often are based on the phase out of old technologies.
2. *Voluntary agreements* include the creation of industry/government co-operation, as well as various industry initiatives aimed at promoting higher levels of energy efficiency.
3. *Financial incentives* include investment subsidies and low interest loans, as well as incentives and subsidies for energy audits.
4. *Environmental taxes/Fiscal reductions* include tax credits and exemptions which are put in place to target higher levels of efficiency within industrial sectors.
5. *Informational and educational programs* are aimed at increasing the awareness of technology users and their knowledge about opportunities for efficiency improvements.

The dimensions of our data are therefore: (i) the single policy instrument, (ii) the policy type or category (as spelled above), (iii) the country, and (iv) the year.

To estimate our model, we construct an ordered categorical variable for each policy instrument by counting the specific policies which are active in any given year. The number of categories differs for each instrument. For Regulatory policies we have: 0, 1, 2 or more; Voluntary measures: 0, 1, 2, 3 or more; for Financial instruments: 0, 1, 2, 3, 4 or more; for Fiscal/tax reductions: 1, 2, 3 or more; for Information/Education: 0, 1, 2, 3 or more. Figure 1 provides an overview of the data by type of instrument and country.

[Insert Figure 1 here]

Hence, measures related to increasing the supply of efficiency technologies are considered as entering the latent regression that explains policy implementation, as explained later in this Section. This notwithstanding, the results including items indicating support to technology development (and excluding these variables from the latent regression) are available from the authors upon request, and do not significantly differ from the ones presented here.

As apparent from the Figure, some countries such as Germany, France and Finland, score higher than other countries. However, they also exhibit lower variation throughout the sample period. Most Eastern European countries, such as for example Romania, increase rather significantly their commitment to energy efficiency towards the middle of the sample period. Regulatory measures are among the least implemented across the sample, while Financial instruments are widely used. Voluntary policy and Education are relatively unexploited in most countries.

The TLRI model with and without latent regression presented in Section 3 are estimated on this data using the GLLAMM routine in STATA (Rabe-Hesketh et al., 2004).

4.2. Descriptive Three Level Random Intercept Model

Estimates for each of the parameter thresholds are reported in Table 1. The estimated thresholds give then rise to a category probability curve for each of the policy instruments analyzed. These curves can be visualized as shown in Figure 2.

[Insert Table 1 here]

[Insert Figure 2 here]

The assumption that thresholds differ by policy instrument or category, as in the PCM framework, is confirmed by the empirical findings. The estimated thresholds for the different policy instruments are also generally significant, with the exception of the second and third thresholds associated with financial instruments (i.e. the thresholds between one and two policy instruments and two and three policy instruments). This lack of significance indicates that having either two or three fiscal policies in place (as opposed to none or one, or more than three) is not the most likely response in any portion of the latent continuum. As Fischer and Parzer (1991) argue, this is an indication that the third and fourth categories (two and three fiscal policies in place, respectively) are less popular than the other categories. Indeed, Panel c Figure 1 shows a sort of "bimodal" distribution of countries with respect to the fiscal policy instruments, with most countries engaging either very little (one fiscal policy or none) or significantly (four or more fiscal policies). This is for instance the case of Italy, Lithuania and Latvia. While the model itself cannot provide an explanation of why this is the case, a possible conjecture is that countries offer a portfolio of financial instruments at a time to promote the increase of efficiency in the industrial sectors, or, alternatively, that once countries engage in this direction, the instruments then quickly become one popular approach to support energy efficiency.

Furthermore, with the exception of the third threshold calculated for Voluntary measures, all the significant thresholds follow a sequential ordering. This means that for all policy categories considered, each threshold is higher than the previous one. A possible explanation for the exception, which mirrors the insights from general economic literature and policy debate, is that voluntary policy measures are not stringent, and therefore not very difficult to implement, as they do not require the imposition of any limitation on the economic agents. Hence, once countries engage more extensively in this type of approach, it is not necessarily harder, or more difficult, to implement yet another intervention of this type (going, for instance, from two to three voluntary measures).

Overall, our estimates support the hypothesis of well-behaved thresholds. This means, as explained for instance in Murray and Mills (2012), that each possible response has a portion of the latent continuum where it is the most likely response.

Conditional on the estimated item thresholds, the variance of second level (time) residual θ_{0tj} and of the third level (country) residual θ_{00j} are estimated at 9.46e-21 with a standard error 1.953e-11 and 0.730 with a standard error 0.22, respectively. Both effects are therefore significant. Note however that the second level variance, which indicates deviations over time of each country's aggregate score from its own mean, is extremely low in absolute value, while the third level variance, indicating the variation of country scores from the overall mean, is higher in absolute value. The between-country variation over our sample period is thus much higher than within-country variation over time.

Given the estimated variance of second and third level residuals, we can obtain the latent traits as the posterior Bayes estimates from the model. We focus first on the country latent trait and present a ranking of countries in Figure 3 which also displays confidence intervals for the estimates. The ranking emerging from the descriptive TLRI model indicates that, conditional on the difficulty of the chosen policy portfolio, Luxembourg and Portugal are among the worst performing countries in Europe in terms of addressing energy efficiency concerns. Note that confidence intervals for the different countries in our sample greatly overlap, indicating that the performance of different EU countries is not strikingly different. Exceptions are Germany and France, which are at the top of the ranking and whose confidence intervals do not overlap with that of other countries. This is consistent with both common knowledge and evidence from the general debate regarding the stringency of environmental policies in different countries, which put Germany and France at the frontier of promoting energy efficiency in the industrial sectors (Scholoman et al. 2015; Egger et al. 2013).

[Insert Figure 3 here]

Finally, we plot the posterior Bayes estimates for each year and country in Figure 4. The figure displays the evolution of the time profile (second level residual) which differs between countries.

We note that, among the New Member countries, Bulgaria, Latvia, Lithuania, Poland, Romania and the Slovak Republic considerably increase their policy effort over the sample period. Conversely, the improvements during the early years in Croatia and Cyprus are not sustained over time. As far as the Old Member countries are concerned, Greece, Ireland and Sweden show the highest increase of their policy effort over time. Italy also increases its commitment, but only in the early years of the sample. Several countries that scored low on policy performance according to Figure 3 do not show improvements over time. This is the case for Luxembourg, Poland, Malta, the Czech Republic and Slovenia.

[Insert Figure 4 here]

4.3 Explanatory Three Level Random Intercept Model

The latent traits emerging from the descriptive TLRI model – and hence the time profiles in Figure 2 and the ranking in Figure 3 – are likely affected by observable country characteristics, as discussed in Section 3. For example, Germany and France may score best due to higher commitment to energy efficiency *over the sample period*, or simply because they are among the richest in the sample, being on top of the ranking at the beginning of the sample period, and sustained their position.

In order to clean the estimated country latent traits, we control for several observable characteristics through latent regressions at both the second and the third level. Specifically, we assume that within country variation (namely the likelihood that a countries increases the number of energy efficiency policies in the manufacturing sector year after year) is affected by energy prices, the weight of the manufacturing sector vis-à-vis other sectors in the economy, the past performance of a country in terms of energy intensity of the overall economy, its dependence on energy inputs and the availability of low-carbon technologies.

We proxy for energy prices using the lagged value of the IEA energy price index for industry (IEA, 2017). The higher the cost of energy, the higher the incentives to engage in energy efficiency as a way to reduce production costs.

The share of manufacturing value added in total value added (World Bank, 2013) measures the weight of the manufacturing sector. It is not clear what the effect of this variable is a priori. A positive effect may arise from because policy measures are more likely in countries with larger manufacturing sectors as they result in larger benefits in terms of lower production costs and emissions; or because stronger lobbies put pressure on governments to put in place fiscal incentives to lower the costs of energy inputs. Conversely, a negative impact could arise from a large manufacturing sector resisting policy intervention.

Energy use as a share of GDP (World Bank, 2013) measures the overall energy efficiency of a given country. Countries with higher levels of energy efficiency may have already reaped the best opportunities and either not feel the need to address the issue of energy efficiency or find it harder to implement any additional regulation.

The share of energy imports as a percentage of merchandise imports (World Bank, 2013) measures energy dependence from abroad. We expect a positive relation between this variable and the latent trait, as promoting energy efficiency is a way to achieve energy security and lower energy dependence from abroad.

The number of energy efficiency patent applications to the EPO by applicants in a given country (OECD, 2013) measures the availability of energy efficient technologies.

This is expected to increase the likelihood that pro-efficiency regulation is passed, given that improvement should be more easily reached vis-à-vis countries with no technology available.

Further, we assume that between country variation (namely, the likelihood that the average performance of countries within the sample is different) systematically differs depending on its richness, on institutional characteristics, and on its overall commitment to the diffusion of renewable energy. Richer countries in terms of GDP per capita (World Bank, 2013) may behave systematically different from poorer countries. Also in this case, the direction of the effect is not clear a priori. On the one hand, citizens in richer countries may be more concerned with climate and pollution issues, and/or governments in richer economies may find it easier to enact regulation promoting energy efficiency. On the other hand, richer economies may also be those where the pressure of higher energy prices is felt less because they don't have a tighter budget constraint.

The institutional variables we include in the latent regression are the average level of respect for Law and Order and of Quality of the Bureaucracy (ICRG, 2011) as well as the length of the democratic system (World Bank 2013). The former are compiled by the International Country Risk Guide and measure on a scale between 0 and 6 the degree with which a given country successfully implements the law and respects it, or the quality of its bureaucratic system. The latter index measures the length of the democratic system in years. Our expectation is that higher these variables, the higher the propensity to implement policies targeting energy-efficiency. The inclusion of these variables are meant to account for the fact that the policy variables (scores) we observe do not convey information about stringency or implementation, rather they represent purely a count of policies. The latent regression approach thus conditions the estimated countries' latent traits also on these observables.[11]

Lastly, we proxy the diffusion of renewable energy with the average growth of renewable electricity by country over the sample period (World Bank 2013).

Descriptive statistics of these variables by country and for the overall sample are displayed in Table 2, which shows large variations within countries. We estimate the model specified in (6). The results on covariates coefficients from the explanatory model are presented in Table 3.

All other things equal, higher energy prices are associated with a higher second level residual, while higher energy intensity of the economy or a higher GDP per capita are associated with lower second level residuals. These results are generally in line with expectation, and they confirm that, conditional on all other variables, richer and more efficient countries find it harder than poorer and less efficient countries to implement energy efficient policies. This may very well be due to the fact, as argued above, that richer countries are also those countries who have already benefitted from the cheapest opportunities, and are finding it harder to sustained increase level of energy efficiency because all their options in this respect are very costly. Finally, the availability of energy efficient technologies and the share of energy imports are not precisely estimated, even though the coefficients are positive, as expected.

A larger manufacturing sector, or an established democracy are associated with a higher third level residual, while a higher growth rate of renewable electricity or a higher score in Law and Order with a lower third level residual. The negative coefficient associated with the growth rate of renewable electricity could be explained, for instance, by the fact that renewable energy deployment and improvements in energy efficient are in fact implemented as substitutes, rather than complements in the countries

[11] Ideally we would want to control for institutional characteristics as well as for stringency and implementation at the sectoral level rather than at the country level. This is not feasible due to a fundamental lack of data. We use country-level proxies under the implicit assumptions that these are strongly correlated with the proxies for the manufacturing sector.

and period under consideration. The negative coefficient associated with the Law and Order variable is consistent with the conjecture that countries with higher enforcement may not need to implement an increasingly higher number of policies targeting energy efficiency, rather, they have a few effective ones in place. The coefficient of the Quality of the bureaucracy is not statistically significant and therefore not precisely estimated.

The ranking of countries emerging from the explanatory TRLI model is shown to change in Figure 5. Specifically, controlling for the observable covariates improves the ranking position of those countries which previously scored low because of some inhibiting "initial conditions", for example (i) low energy prices, or (ii) a manufacturing sector accounting for a smaller share of their economy, or (iii) a higher degree of energy intensity and so forth. Most Eastern European countries, for example, improve their ranking, while Northern countries perform worse when their "initial conditions" are taken into account.

A word of caution is due when commenting Figure 5. As a result of the explanatory model countries such as, for instance, Bulgaria scores significantly higher than in Figure 3. Far from being an indication that the policy stringency of this country is higher than that of, say, Germany or France, the results of the explanatory model simply suggest that *over the sample period* Bulgaria committed to the highest number of new policies, given its initial condition (which was clearly worse than that of other EU countries). Hence, while the results of the descriptive model can be taken as representing a snapshot of countries' performances in general, the results of the explanatory model convey information about the performance of countries over the sample period considered, given their initial condition. The usefulness of the TLRI approach adopted in this paper lies precisely in being able to provide both types of information.

Finally, note that also in the explanatory TLRI model, confidence intervals for countries overlap. Even after conditioning of the model on observable country characteristics, the performance of EU countries with respect to commitment towards energy efficiency appears to be very similar, and the clustering of countries in different and separate groups is not possible.

5. Conclusion and Policy Implications

This paper presented a novel approach to assessing and comparing countries' commitment to environmental and energy policy in a given sector. The Three Level Random Intercept model we propose allows using the sparse data on countries' policies to derive a ranking and to characterize the complexity of their portfolio of energy and environmental policies. In addition, the model can be extended to adjust the ranking as a result of country-specific economic and institutional observables which are likely to a affect regulation over the sample period. We illustrate the potential of this methodology by studying the performance of the EU28 plus Norway with respect to promoting energy efficiency over the years 2004-2011.

We believe the TLRI model is a promising approach to assessing countries' commitment to environmental and energy policy since it is able to overcome a number of shortcomings of the previous literature on policy indicators. First, its data requirements are relatively limited, allowing to extend the analysis to a much wider set of countries than previous analyses. Second, this approach allows attributing different weights or difficulty levels to the different policy instruments included in the policy portfolio. Thus, our assessment is conditional on the specific complexity of each country's policy portfolio. Third, we combine the basic TLRI model with a latent regressions, thereby allowing each country's scores to be conditioned on the country's observed characteristics.

Focusing on the results emerging from our application, we show that accounting for economic and institutional characteristics changes the ranking of countries with respect to energy efficient policy. Specifically, the position of those countries with worse “initial conditions” but which choose to regulate energy efficiency nonetheless, demonstrating a higher than average commitment over the sample period, improves.

We are fully aware that, given the nature of our data, namely count of policies which are implemented in each country in any given year, our ranking only informs on the general commitment of a given country and it does not shed any light on the actual level of stringency of the given policy. While we try to control for this shortcoming in our empirical setting by adding covariates capturing the level of respect for law and order, our results should be interpreted accordingly.[12]

This notwithstanding, basically available efforts to assess national performance of countries with respect to energy and environmental policy in general suffer from this shortcoming or, if they don’t, they often rely on arbitrary categorization and expert scoring, which are often less than transparent to the reader.

While we believe our methodology is a step forward in the direction of creating an index of policy commitment, and possibly stringency, we are also aware that it is not completely free from limitations. First, at present our proposed approach allows to study the performance of countries in a given sector/aspect of energy and environmental policy. Indeed, enlarging the analysis to include other end-use sectors, such as households, transport or the tertiary, would entail the development of a four level random intercept model which would nest an additional level (i.e., sector) in the empirical framework. We have left this effort for future endeavors, given that it significantly increases computing time and, most of all, it requires to somehow account

[12] Note that ideally when building indexes of policy performance and stringency one would want to account for the contribution of each policy to the specified country targets. This is indeed a crucial point, but one that at present we cannot address due a fundamental lack of data as well as of a methodology to calculate the contribution of several policies, which are often implemented, to a specified future target. This is clearly a very interesting avenue for future research.

that the policy instruments applied in one sector are necessarily of a different nature from those applied in another sector.

Another natural follow-up of this paper is to assess how our approach fares compared with other indicators of energy and environmental policy stringency proposed in the literature, which we briefly considered in Section 2.[13] More generally, the application of our methodology to other fields of study where similar data on policy implementation is available is also very promising. This might include, among other, labor or monetary policy. The model presented in this paper could also be fruitfully extended to account for the presence of random slopes (as opposed to only random intercepts) and to better study the effect of time. To this end, focusing on a wider sample of countries would be beneficial, since variation in policy responses of the EU member states is necessarily limited given the common framework under which these policies are developed. Our current research is moving in this direction.

[13] A companion paper (Galeotti, Salini, and Verdolini, 2017), whose content could not be presented here due to obvious space constraints, is a first attempt in that direction.

References

Andersen, E. B., 1997. The rating scale model. In: Handbook of modern item response theory. Springer, pp. 67–84. Andrich, D., 1978. Relationships between the Thurstone and Rasch approaches to item scaling. Applied Psychological Measurement 2 (3), 451–462.

Andrich, D., 1978. Application of a psychometric rating model to ordered categories which are scored with successive integers. Applied Psychological Measurement 2 (4), 581–594.

Bacci, S., 2012. Longitudinal data: different approaches in the context of item-response theory models. Journal of Applied Statistics 39 (9), 2047–2065.

Bacci, S., Bartolucci, F., 2012. A multidimensional latent class Rasch model for the assessment of the health-related quality of life. arXiv preprint arXiv:1211.2635.

Bacci, S., Caviezel, V., 2011. Multilevel IRT models for the university teaching evaluation. Journal of Applied Statistics 38 (12), 2775–2791.

Baker, F.B., Kim, S., 2004. Item Response Theory: Parameter Estimation Techniques. Marcel Dekker, New York.

Berman, E., Bui, L. T. M., 2001. Environmental Regulation and Productivity: Evidence from Oil Refineries. Review of Economics and Statistics 83 (3), 498–510.

Botta, E., Koźluk, T., 2014. Measuring environmental policy stringency in OECD countries: a composite index approach. OECD Economics Department Working Papers No. 1177, OECD Publishing, http://dx.doi.org/10.1787/5jxrjnc45gvg-en.

Brunel C., Levinson, A., 2013. Measuring Environmental Regulatory Stringency," OECD Trade and Environment Working Papers 2013/5, OECD Publishing

Brunel, C., Levinson, A., 2016. Measuring Environmental Regulatory Stringency. Review of Environmental Economics and Policy, 10(1) 2016, 47-67.

Brunnermeier, S. B., Cohen, M. A., 2003. Determinants of Environmental Innovation in US Manufacturing Industries. Journal of Environmental Economics and Management 45, 278293.

Cole, M., Elliot, R., 2003. Do environmental regulations influence trade patterns? Testing old and new trade theories. The World Economy 26 (8), 1163–1186.

De Boeck, P., Wilson, M., 2004. Explanatory Item Response Models. A Generalize Linear and Nonlinear Approach. Springer, New York.

Egger, C., Priewasser, R., Kloiber, M., Borg, N., Bourges, D., Schilken, P., 2013. Energy Efficiency in Europe: Assessment of Energy Efficiency Action Plans and Policies in EU Member States. http://www.energy-efficiency-watch.org/fileadmin/eew_documents/Documents/EEW2/France.pdf

Ferrari, P. A., Annoni, P., Salini, S., 2005. A comparison between alternative models for environmental ordinal data: Nonlinear PCA vs Rasch analysis. In: Statistical Solutions to Modern Problems: Proceedings of the 20th international Workshop on Statistical Modelling, eds. AR Francis, KM Matawie, A. Oshlack and GK Smyth, Sydney: University of Western Sydney. pp. 173–177.

Ferrari, P. A., Salini, S., 2011. Complementary use of Rasch models and nonlinear principal components analysis in the assessment of the opinion of Europeans about utilities. Journal of classification 28 (1), 53–69.

Fischer, G. H., 1989. An IRT-based model for dichotomous longitudinal data. Psychometrika 54 (4), 599–624.

Fischer, G.H. & Parzer, P. (1991). An extension of the rating scale model with an application to the measurement of change. Psychometrika, 56, 637-651.

Fischer, G. H., Frewer, L., Nauta, M., 2006. Toward Improving Food Safety in the Domestic Environment: a multi-item Rasch scale for the Measurement of the Safety Efficacy of Food-Handling Domestic Practices. Risk Analysis 26 (5), 1323–1338.

Fisher, C., Newell, R., 2008. Environmental and technology policies for climate mitigation. Journal of Environmental Economics and Management 55, 142–162.

Schlomann, B. , Wolfgang, E., Reuter, M., Frölich, C., Tariq S., 2015. Energy Efficiency trends and policies in Germany, Fraunhofer Institute for Systems and Innovation Research ISI, http://www.odyssee-mure.eu/publications/national-reports/energy-efficiency-germany.pdf

Galeotti, M., Salini, S., Verdolini, E., 2017. Measuring Environmental Policy Stringency: Approaches, Validity, and Impact on Energy Efficiency. Centro Studi Luca D'Agliano Working Paper No. 412, January.

Gnaldi, M., Bacci, S., Bartolucci, F., 2016. A multilevel finite mixture item response model to cluster examinees and schools. Advances in Data Analysis and Classification 10 (1), 53–70.

Goldstein, H. 2011. Multilevel Statistical Models (4th ed.). London: Wiley. ISBN 978-0-470-74865-7.

Gollop, F., Roberts, J., 1983. Environmental regulations and productivity growth: The case of fossil-fueled electric power generation. Journal of Political Economy 91 (4), 654–674.

Hamamoto, M., 2006. Environmental Regulation and the Productivity of Japanese Manufacturing Industries. Resource and Energy Economics 28, 299–312.

ICRG, 2011. International country risk guide database.

IEA (2012) IEA energy efficiency database. http://www.iea.org/policiesandmeasures/energyefficiency/, accessed: 2012-10-30

IEA (2017) Energy Prices and Taxes. https://www.iea.org/statistics/relateddatabases/energypricesandtaxes/, accessed 2017-04-30

Jaffe, A., Palmer, K., 1997. Environmental Regulation and Innovation: A Panel Data Study. Review of Economics and Statistics 79 (4), 610–619.

Jeppsen, T., Folmer, H., 2001. The Confusing Relationship between Environmental Policy and Location Behavior of Firms: A Methodological Review of Selected Case Studies. Annals of Regional Science 35, 523–546.

Kellenberg, D., 2009. An empirical investigation of the pollution haven effect with strategic environment and trade policy. Journal of International Economics 78 (2), 242–255.

Lanjouw, J., Mody, A., 1996. Innovation and the International Diffusion of Environmentally Responsive Technology. Research Policy 25 (4), 549–571.

Mair, P., Hatzinger, R., 2007. CML-based estimation of extended Rasch models with the ERM package in R. Psychology Science 49 (1), 26.

Masters, G.N., 1982. A Rasch model for partial credit scoring. Psychometrika 47 (2), 149–174.

Masters, G. N., Wright, B. D., 1997. The partial credit model. In: Handbook of modern item response theory. Springer, pp. 101–121.

MURE Mésures d'Utilisation Rationnelle de l'Energie (2012), ODYSSEE MURE Energy efficiency policy and measures database. http://www.indicators.odyssee-mure.eu/energy-efficiency-database.html. Accessed: 2012-10-30.

Murray, A. G., Mills, B. F., 2012. An application of dichotomous and polytomous Rasch models for scoring energy insecurity. Energy Policy 51 (C), 946–956.

Nicolli, F., Vona, F., 2012. The Evolution of Renewable Energy Policy in OECD Countries: Aggregate Indicators and Determinants, FEEM Nota di Lavoro 2012.51, Fondazione Eni Enrico Mattei.

OECD, 2013. Patents by main technology and by international patent classification (IPC). URL /content/data/data-00508-en

Oreja-Rodriguez, J., Yanes-Estevez, V., 2007. Perceived Environmental Uncertainty in Tourism: a New Approach Using the Rasch Model. Tourism Management 28, 1450–1463.

Perman, R., Ma, Y., Common, M., Maddison, D., Mcgilvray, J., 2011. Natural Resource and Environmental Economics, 4/E. Addison-Wesley.

Popp, D., 2003. Pollution Control Innovations and the Clean Air Act of 1990. Journal of Policy Analysis and Management 22 (4), 641660.

Popp, D., 2010. Exploring the Link between Innovation and Diffusion: Adoption of NOx Control Technologies and U.S. Coal fired Power Plants. Environmental and Resource Economics 45, 319–352.

Schaffrin, A., Sewerin, S., Seubert, S., 2015. Toward a Comparative Measure of Climate Policy Output. Policy Studies Journal 43, 257-282.

van der Linden, W.J., Hambleton, R.K. (Eds.), 1997. Handbook of Modern Item Response Theory. Springer-Verlag, New York.

Environmental policy performance and its determinants: application of a three-level random intercept model

Tables

Table 1: Estimation of Item/Policy Thresholds

Difficulty Thresholds	Coefficient	Standard Error	p-value
Regulatory Policy - Threshold 1	0.47	0.22	0.03
Regulatory Policy - Threshold 2	1.87	0.31	0.00
Voluntary Measures - Threshold 1	0.51	0.23	0.03
Voluntary Measures - Threshold 2	1.01	0.29	0.00
Voluntary Measures - Threshold 3	0.87	0.35	0.01
Financial Instruments -Threshold 1	-0.73	0.26	0.00
Financial Instruments -Threshold 2	0.35	0.26	0.19
Financial Instruments -Threshold 3	-0.17	0.28	0.53
Financial Instruments -Threshold 4	1.14	0.30	0.00
Fiscal/Tax Reductions - Threshold 1	1.29	0.23	0.00
Fiscal/Tax Reductions - Threshold 2	2.43	0.41	0.00
Information/Education - Threshold 1	0.43	0.23	0.07
Information/Education - Threshold 2	0.57	0.27	0.03
Information/Education - Threshold 3	1.10	0.33	0.00
Variances of random effects			
Second Level Variance (Time)	9.46e-21	1.953e-11	0.00
Second Level Variance (Country)	0.73	0.22	0.00

Table 2: Descriptive Statistics of Covariates of the Latent Regression

Country	EPO Energy Efficient Patent Applications	Energy Intensity	Fuel Imports, % of Merchandise Imports	Energy Intensity	GDP per capita	Law and Order	Bureacratic Quality	Length of Democracy	Growth rate of Renewable Electricity	Manufacturing in Value Added, % of GDP
Austria	25.40	1.393	11.22	0.135	42.585	6	3.917	52.08	0.0275	19.66
Belgium	20.62	1.536	12.44	0.191	41.466	5.475	4	76.53	0.148	17
Bulgaria	0.833	1.251	14.48	0.796	5.357	4.261	2	16.42	0.0692	16.29
Croatia	0.143	1.376	16.09	0.241	12.734	4.917	2.853	14.64	0.0245	17.40
Cyprus	2.269	1.320	17.94	0.187	27.01	4.315	3.591	.	.	8.780
Czech Republic	3.042	1.433	8.413	0.408	17.339	5.204	3	16.58	0.0621	24.53
Denmark	74.17	1.582	6.378	0.0975	53.45	6	4	75.92	0.0710	14.49
Estonia	1	1.331	15.13	0.492	14.445	4	2.628	14.93	0.0280	17.05
Finland	20.43	1.535	15.43	0.227	43.676	6	3.980	76.50	0.0154	23.33
France	93.20	1.563	14.02	0.155	38.891	5.148	3.620	76.67	0.0390	13.50
Germany	386.7	1.580	11.24	0.149	38.91	5.407	3.992	58.46	0.113	22.45
Greece	3.367	1.463	18.30	0.157	25.083	3.969	2.747	32.47	0.0524	9.338
Hungary	2.308	1.416	8.758	0.299	12.757	4.787	3.299	16.62	0.0928	22.41
Ireland	8.786	1.541	8.909	0.0935	51.23	5.404	3.866	76.50	0.105	26.15
Italy	57.18	1.652	14.29	0.126	34.084	4.863	2.940	30.79	0.0818	18.68
Latvia	0.400	1.247	13.74	0.346	10.513	4.865	2.372	13.53	-0.000226	12.41
Lithuania	0	1.218	23.40	0.413	10.445	4	2.372	10.23	0.0638	19.21
Luxembourg	5.333	1.326	9.213	0.145	98.952	6	3.980	76.67	0.111	8.878
Malta	0.333	1.469	14.07	0.185	18.27	4.567	2.642	31.60	0.250	16.85
Netherlands	53.75	1.653	13.85	0.157	45.547	6	4	76.50	0.0759	14.08
Norway	20.13	1.813	4.957	0.116	81.272	6	3.826	76.73	-0.0555	10.01
Poland	4.042	1.372	10.41	0.363	10.312	4.506	2.562	16.33	0.0516	18.22
Portugal	3.827	1.521	14.00	0.165	21.006	5.090	2.744	30.73	0.0549	15.23
Romania	0.667	1.277	11.79	0.453	7.249	3.682	0.769	14.83	0.0620	22.98
Slovakia (Slovak Republi	1.143	1.479	12.56	0.443	14.505	4.593	3.058	14.29	0.0685	23.25
Slovenia	0.733	1.324	11.06	0.242	21.954	4.673	3	15.33	0.0407	23.69
Spain	48.79	1.303	15.84	0.150	29.575	4.694	3.171	29.47	0.0870	16.05
Sweden	28.24	1.599	12.04	0.166	47.111	6	4	76.53	0.0322	19.25
United Kingdom	66.08	1.721	9.548	0.120	39.145	5.465	4	76.36	0.130	13.41

Table 3: Results on Second and Third Level Covariates

Variable	Coefficient	Standard Error	t ratio
Covariates of Second Level Random Effect (Time)			
Energy Price (t-1)	1.2506	0.2684	4.6595
Energy Intensity (t-1)	-1.661	0.7802	2.1289
Fuel Imports, % of Merchandise Imports (t-1)	0.0141	0.0142	0.9930
GDP per capita, (t-1)	-0.0162	0.0073	2.2192
Energy Efficient Patent Applications to EPO (t-1)	0.0021	0.0013	1.6154
Covariates of Third Level Random Effect (Country)			
Manufacturing in Value Added, % of GDP (mean)	0.0745	0.0243	3.0658
Growth rate of Renewable Electricity, (mean)	-5.1386	2.3542	2.1827
Law and Order (mean)	-0.9709	0.39	2.4895
Quality of the Bureaucracy (mean)	0.0964	0.3394	0.2840
Length of Democracy (mean)	0.0325	0.0085	3.8235

Figure(s)

Figure 1: Policy variables distributions per country and year. (a) Regulatory Policy, (b) Voluntary agreements, (c) Financial instruments, (d) Fiscal/tax reduction, (e) Information/Education