Article

Not peer-reviewed version

Energy Efficiency Forecast as an Inverse Stochastic Problem: A Cross-Entropy Econometrics Approach

Second Bwanakare *

Posted Date: 4 October 2023

doi: 10.20944/preprints202310.0230.v1

Keywords: overall energy efficiency score; energetic intensity; non-extensive cross-entropy econometrics; stochastic inverse problems; regional innovation doi:10.20944/preprints202310.0230.v1

Article

Energy Efficiency Forecast as an Inverse Stochastic Problem: A Cross-Entropy Econometrics Approach

Second Bwanakare

Cardinal Stephan Wyszynski, Institute of Statistics in Rzeszów University in Warsaw, s.bwanakare@uksw.edu.pl

Abstract: This paper proposes a non-extensive entropy econometric technique to predict energy efficiency at province (NUT-2) level based on imperfect knowledge of the national overall efficiency in the sectors of industry, transport, households and services. The model is applied to the polish case. As acknowledged in recent literature, non-extensive entropy model should remain a valuable device for econometric modelling even in the case of low frequency series since outputs provided by the Gibbs-Shannon entropy approach correspond to the Tsallis entropy limiting case of the Gaussian law when the Tsallis q-parameter converges to unity. Therefore, we set up a q-Tsallis-Kullback-Leibler entropy criterion function with a priori consistency moment and model data constraints, including province energy intensity (known with uncertainty), regional climate differentiation and regular conditions. The model outputs continue to conform to empirical expectations. In spite of the close to unity q-Tsallis parameter, this Tsallis related approach reflects higher stability for parameter computation in comparison with the Shannon-Gibbs entropy econometrics technique. The proposed technique can be applied in different EU countries and elsewhere for example in the context of experimental official statistics.

Keywords: overall energy efficiency score; energetic intensity; non-extensive cross-entropy econometrics; stochastic inverse problems; regional innovation

PACS: 89.65.Gh; 89.70.cf

---

1. Introduction

The oil shocks and market supply instabilities and uncertainties that have been occurring since 1973 until today have led to an increase in energy costs with negative consequences on the growth of different economies in different parts of the world. Next, the 1997 Kyoto Protocol operationalized the United Nations Framework Convention on Climate Change by committing industrialized countries and economies in transition to limiting and reducing greenhouse gas (GHG) emissions in accordance with agreed individual goals. It is therefore the occurrence of these two joint goals of the end of the last century that forced public decision-makers to initiate energy and carbon saving policies. Consequently, Europe defined the goals for 2030 [1] of which reduction of 40% of greenhouse gas emissions with respect to 1990 values and an increment of energy saving of at least 32.5%.

Following IRENA [2], energy efficiency and renewable energy sources may provide more than 80% of the emission savings required [3]. These saving-oriented objectives requires a management system enabling the evaluation of progresses with respect to these defined goals at different time and space scales. In fact, many authors are aware of the complexity of this problem, which has also been highlighted in recent publications (e.g., [4,5]). One of the issues is the lack of statistical data on a disaggregated scale which contrasts with the pressing needs for information of the parallel level for a such strategic sector of the economy.

Therefore, this paper tries to answer to one of these issues. We will present and apply a recent statistical approach to recover statistical information in conditions when other mathematical or statistical techniques are unable to pass muster. The approach is related to the maximum entropy principle known to tackle with ill-posed inverse problems as the one to be solved in this paper. In this document, the question will arise when trying to recover the estimates from a dashboard of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2023 doi:10.20944/preprints202310.0230.v1

energy efficiency scores at a disaggregated national level where the statistical data does not exist. The case of polish provinces will be methodologically illustrative for many similar countries, particularly those within the European Union. Many developed countries publish annually (e.g. [6]) sectoral energy efficiency scores aggregated at the national level. However, such an aggregated information makes it very difficult to assess and plan the energy policies whose the scope of operation happens to be at the sub-regional levels. To overcome this problem, statistical institutes instead calculate energy intensity scores, which are easier to estimate. The first question is whether or not the production of statistics on energy efficiency coefficients at a disaggregated level (in our case at NUT 2 level) deserve to be calculated alongside energy intensity coefficients. The counter argument is that the energy intensity coefficients available should be sufficient to provide with information on its efficiency. Nevertheless, the correlation between energy intensity and energy efficiency is generally far from perfect. For instance, a small service-based economy in a mild climate region will be characterized by a lower intensity than a large industry-based economy in a colder climate region, even though the latter uses energy more intensively and efficiently. In addition, other elements will also play a role in defining intensity levels and trends. Among these, the regional economic structure (share of large energy-consuming industries), geographical characteristics (e.g. longer distances leading to higher demand for transport), climatic and weather conditions (changing demand heating or cooling) (e.g. [6].

Based on what has just been presented above, let us now concretize the problem targeted by this paper as follows: we need to estimate the coefficients of energy efficiency at the sixteen province level of Poland. Energy efficiency into question concerns the following 4 sectors: Industry, Transport, Households and Services. The ODYSSEE-MURE project (e.g. [6] publishes national statistics on the sectoral averages of these coefficients. Next, some institutes such as the Polish Institute of Statistics [7]publish energy intensity coefficients at the disaggregated level, e.g. at the province level. The question posed in this article is therefore how to reconcile these two sources of information to forecast the sector ratios of energy efficiency at the province level. Mathematically, the problem to be presented in a rectangular table(see tab. 1) is ill-posed. On the one hand, we only have information on the averages of the energy intensity coefficients for each of the 16 provinces while at the level of the 4 sectors we have the averages of energy efficiency without, in both cases, any details at the individual interprovincial sectoral level. On the other hand, it follows that the total of the rows cannot correspond to the total of the columns. Finally, if we take additionally into consideration the fact that the systems generating these two sources of information are not known with precision, the problem to be solved turns out to be an ill-posed stochastic inverse problem whose solution goes beyond traditional mathematical techniques.

The remaining part of this paper is organized as it follows. The Section II introduces concepts and definitions related to the energy optimal use and saving through a recent literature on the subject. Section III present mathematical and statistical insights of model in the context of the problem to be solved. In this context, the concept of *inverse problem* is defined and this will be followed by a detailed presentation of the non-extensive cross-entropy econometrics as a main technique applied by the model. Section IV presents the model outputs and comments. The final Section draws conclusions and highlights potential outcomes related to the application of the presented method while suggesting perspectives for future research.

2. Energy Efficiency and Its Measurement

This Section provides basic concepts to enable a good understanding and interpretation of the computed outputs. There exists a vast literature on the energy efficiency definition and measurement (e.g., [8,9]. The reason may particularly reside in the complexity of the energy efficiency technical evaluation [10]. Issues are the choice of the right aggregation level, the appropriate variables to construct a reference energy consumption trend, the energy units to be applied and interaction between various effects. Uncertainty margins for results lack in most presentations as well. Among a vast literature, we underscore some of these works which disserve a higher attention in the context of this paper. In particular, the authors [11] present two approaches to measuring energy efficiency.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2023 doi:10.20944/preprints202310.0230.v1

The bottom-up approach to which the 'Odex' index is linked. It has been developed under the EU Odyssee-Mure programme. It will be developed in this article. The second approach is that of Top-down which brings together the 'Decomposition' methodologies as used by e.g. Netherlands, Canada and New Zealand. Likewise, the authors [12] propose the calculation of total factor energy efficiency (TFEE) using the concept of global production instead of domestic product which excludes intermediate consumption. A next paper worthy to cite comes from the authors [13] . In that work, a literature review was carried out and authors found that the currently in use definitions of total-factor energy efficiency and total factor carbon emissions efficiency are confusing and misleading enough.

Regarding institutional research, the Environmental and Energy Study Institute [14] uses a definition of energy efficiency emphasizing using less energy to perform the same task, i.e. eliminating energy waste. The European Parliamentary Research Service (EPRS) definition [15] highlights the fact that energy efficiency should refer, in general terms, to the amount of output that can be produced with a given energy input. Most commonly, energy efficiency is measured as the amount of energy output for a given energy input . However, other kinds of output can also be used. The EU Energy Efficiency Directive uses a very broad definition: 'energy efficiency' means the ratio of output of performance, service, goods or energy, to input of energy. Following the International Energy Agency [16], the energy efficiency ratio is a ratio between energy consumption (measured in energy units) and activity data (measured in physical units).

Finally, the UNDP, as an organ of the United Nations, has published a summary work which takes up different approaches to calculating energy efficiency. In fact, this institution in the third chapter [17] begins by identifying the methodological challenges associated with defining and measuring energy efficiency. It then proposes a framework for understanding energy efficiency trends, integrating the current UNDP approach to energy efficiency developed by various international agencies and national institutions, and establishing a methodology to identify a starting point in relation at which future improvements in energy efficiency can be measured globally and at national levels.

It is worthy to pay attention to the related three next concepts to be used in the next part of the paper. For an economy-wide measure, GDP is often compared to energy use, to give *the energy intensity* (measured for example in kilowatt-hours per euro). Next, *energy savings* are the reduction of energy use, without reference to output produced. Finally, as far as energy efficiency assessment is concerned, this can be done at different levels and according to different techniques. These levels may start from economy-wide and sectoral energy intensity to individual units of activity.

We now present some details regarding the ODEX energy efficiency index from which the data to be used in our model are extracted as a priori. This index is published by the Odyssee-Mure project [18] to measure the energy efficiency progress by main sector (industry, transport, households, services) and for the whole economy (all final consumers). The ODEX composite indicator is calculated as a weighted average of sectoral indices.

For each sector, the index is calculated as a weighted average of sub-sectoral indices of energy efficiency progress. The sub-sectors stand for industrial branches, service sector branches, end-uses for households or transport modes.

This project calculates the scores by including the following energy efficiency components:

- the energy efficiency level,
- the energy efficiency trends,
- the energy efficiency policies,
- the overall energy efficiency.

These three criteria are scored between 0 and 1 on the basis of a variety of indicators (extracted from the Odyssee Database) and of energy policies (extracted from the Mure Database). The overall energy efficiency score is obtained as an average of the three scores obtained for "energy efficiency level", "energy efficiency progress" and "energy efficiency policies" (i.e. one third weighting). This work will use data representing the overall energy efficiency. Following [18], the energy efficiency scoring technique is based on the OECD Composite Indicator methodology . This method allows the countries or regions to be compared in a relevant range where minimum and maximum values

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2023 doi:10.20944/preprints202310.0230.v1

indicators define the best and worst scores and countries or regions are ranked between these two extrema. The indicators are calculated and normalized so that they range between 0 and 1 following this formula:

$$Normalized\ score = [\ Indicator - Min\ indicator \backslash (\ (Max\ indicator - Min\ indicator) * direction)] + 0.5 * (1 - direction)$$

where:

Indicator: The indicator value of the country\region.

Min indicator: The minimum indicator value across all countries\regions.

Max indicator: The maximum indicator value across all countries\regions.

Direction: The favored direction in the level of indicator; -1 if the decline is favored, 1 if the incline is favored.

In spite of a "one third weighting", the most influential score is the one of the energy efficiency level to which the two remaining scores should be related. Its scoring, according to Odyssee-Mure practice, is done as follows:

- Scoring is done separately for the four considered sectors (households, transport, industry and services) and for all sectors together.
- The score by sector is based on scores computed for statistically selected indicators of end-uses in buildings or modes in transport. For industry sector an aggregate score is obtained from various industrial branch scores that account for the energy efficiency characteristics of each of them.
- The score by sector is calculated as a weighted score of each indicator. The weights correspond to the average shares over the last 3 years of each end-use or transport mode in the sector consumption.

Finally, for comparative reasons, sector score values are normalized into interval 0-1 according to to the next formula:

$$Normalized\ score = (\ Indicator - Min\ indicator) / \ (Max\ indicator - Min\ indicator)$$

where:

Indicator: The indicator value of the sector

Min indicator: The minimum indicator value across all sectors.

Max indicator: The maximum indicator value across all sectors.

To close this Section, it is worthy to briefly comment on energy efficiency estimation in the presence of recent energy system known as *integrated energy system (IES)* .

In the context of the energy crisis and environmental degradation, an integrated energy system (IES) based on the complementarity of multiple energy sources and the cascading use of energy is considered an effective way to mitigate these problems. Due to the different forms of energy and the different characteristics of IESs, the interrelationships between different forms of energy are complicated, which increases the difficulty of assessing the energy efficiency of IESs. A limited number of technique exists. We can send the interested authors to the authors [19]. These authors have proposed a technique mixing Energy Use Efficiency (EUE) and Exergy Efficiency (EXE) based on the first/second law of thermodynamics [4].

1. Mathematical problem setting

a) Inverse problem and the maximum entropy principal

In many real-world situations, theorists and empiricists observe at a given time two or more quantifiable multivariate stochastic systems and want to infer on an unknown cross-correlation between their random elements. To illustrate that, we implement a cross-entropy formalism to forecast an interprovincial sectoral energy efficiency score matrix based on imperfect and contradictory information from province or sector aggregates.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2023 doi:10.20944/preprints202310.0230.v1

The basic model for[1] dealing with poorly posed inverse problems is to solve the integral equation of the first type. We formulate this – in the context of a model that will be developed later – as follows:

$$G(\zeta) = \int_D f(\beta) h(\beta, \zeta) d\beta + b(\zeta) \quad (1)$$

– G the amounts observed in rows or columns;
– f is the unknown regional cross-sectoral greenhouse contamination matrix;
– D defines the Hilbert support space of the model,
– h is the transformation kernel linking measures G and f,
– b explains random errors.

The literature on various methodologies devoted to the recovery of ill-posed inverse problem is expansive. In addition to the well-known Tychonoff regularization theory [20], the Gibbs–Shannon–Jaynes principle of maximal (minimum) entropy, [21,22] and recent extensions, [23,24] remain the most commonly used techniques to solve this class of problems. The general rule that applies to both approaches is to associate the linear or nonlinear problem of the least squares with the regularization rule (a *priori* or additional information) to *get to a well-posed problem*. Moreover, the Gibbs–Shannon–Jaynes principle of maximum (minimum) entropy formalism tries to search for global regularity — related to the second law of thermodynamics — while producing the smoothest reconstructions consistent with the data available in the Bayesian spirit. In this research, as it is often the case in many empirical applications, the discrete form of equation 1 was implemented.

As far as the social sciences are concerned, a number of other techniques have been tried for this class of inverse problems. Examples include the pseudo-inverse Moore–Penrose problem approach or the bi-proportional RAS approach and its extension [25]. Although the latter technique requires an initial transaction matrix, it offers a less good solution when the model studied is s*tochastic*. The authors [24] showed poorer performance of the Markov chain model compared to the generalized Gibbs–Shannon entropy for this class of inverse problems. Subsequently, the Bayesian approach showed its relative superiority, particularly that associated with the principle of maximal entropy. A neural class model can also be proposed. However, it is not based on a compact theory, its application takes time and the results are not always guaranteed. The entropy model has been successfully applied to update and balance social accounting matrices [25] . However, on theoretical grounds, this assumes that entropy is a positive linear function of the number of possible states and then ignores the possibility of interdependencies between states and their influence. In a recent paper, the authors [26] demonstrated the convergence of two standard regularization techniques to two special power law (PL) related q Tsallis values. For q = 2, a Tychonoff regularization is obtained, and for q = 1, the classical formulation of the Boltzmann–Gibbs–Shannon entropy is obtained. The central point is that, in addition to the well-known law of scale, PL exhibits a series of interesting characteristics related to its aggregation properties, in that it is preserved under addition, multiplication, polynomial transformation, minimum and maximum [27].

Since we are dealing with a poorly conditioned inverse problem, we must meet all three conditions of regularity (existence, uniqueness, and stability of the solution) at the same time.

If, in general, the conditions of existence *and uniqueness* remain available due to regular a priori constraints, the *stability* of the optimal solution, by random or systematic errors, is much more difficult to find. In short, the problem is, among an infinite number of distributions that meet all the restrictions imposed, to find the one that best replicates the data generation system (DGS). As for the formalism of maximum entropy, thanks to Jaynes' contribution [5], the reasonable candidate should

1 For example, by limiting itself to ordinary cases of signal or imaging, the basic equation (equation 1) can be extended to, for example, the impulse response of a measurement system.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2023 doi:10.20944/preprints202310.0230.v1

be the one which reduces uncertainty about the system the most. Similarly, according to Kullback-Leibler, information divergence metrics, [21,28] the best candidate should be those posteriors who meet all the binding conditions and deviate the least from the priors.

a) Non extensive cross-entropy energy model and confidence interval area

Kullback-Leibler information divergence metrics are best known when they apply to Gaussian phenomena. In the case of PL phenomena - of which the Gaussian law is a particular case, the divergence metric applied is the q-Generalized Kullback-Leibler relative entropy. This metric- in its essence the best fitted to nonlinear systems, works even in the case of scale invariance, long-memory correlation, among others. These properties give it practical advantages over many other laws(e.g., [29,30]). The model related to this metric has been extensively presented in different publications [e.g. [30]] in the context of macroeconomic analysis. Since this work deals with energy management, it is worthy to reformulate the model in the energy management context to enable interpretation of the outputs.

We implement[2] the usual[3] discrete form of the q-Generalized Kullback-Leibler relative entropy (Eq. 5). With the cross-entropy new constraining data, the model updates the initial information(priors in tab.1) and provides new outputs(posteriors).

It is necessary to redefine the parameterization of the generalized linear model (equation 2), which plays the role of constraints. The inside table elements to be forecasted can be meaningfully presented by columns as the discrete Bayesian joint probabilities explaining each region's average cross-sector weight or probabilities corresponding to individual energy efficiency ratio. The ratio total per column will sum up to unit. We recall that energy efficiency coefficients to be forecasted are in the form of normalized indicators, thus belonging within the space zero and one. In this case, the parameter processing space coincides with the probability space. Under these conditions, the accuracy of the estimated parameters is greater, since a priori there is no loss of information from these data [30]. In any case, let us briefly present the general procedure of reparametrization in the case of a general linear inverse model:

$$Y = X^{'} \cdot \beta + \varepsilon \qquad (2)$$

where the values of unknown parameters β are not necessarily bounded between 0 and 1 indicating the need for reparameterization. This ε term is an unobservable random term for perturbations, plausibly with finite variance exhibiting observation errors from empirical measurement or random shocks that may be driven by PL. The variable Y consists of data observed – with errors – of an unknown data generating system of energy efficiency coefficients by sector and X may represent known average regional energy intensity coefficients – with uncertainty – through the relational parameter matrix β and the unobservable disturbance ε to be estimated through the observable error components e. Unlike classical econometric models, no binding assumptions are required – for example, about the distribution of random errors. In particular, as we deal with an ill-behaved inverse problem, the number of parameters to be estimated should be greater than the observed data points, and the quality of the information data collected should be low. The process of the true system recovery requires the entropy objective function to include all of the interacting constraining consistency moments. Thus, referring to the properties of the relative entropy principle, each new piece of constraining information will reduce the entropy level of the system in accordance with the degree of data consistency with the system. For this multi- dimensional space inverse problem , among an unlimited number of model solution candidates, the best solution will result from identifying the one that—in terms of probability—best simulates the data generating system. By taking each of β_{kl} $\left(k = 1 \ldots K, l = 1 \ldots L\right)$ as a discrete

2 Generalized Bregman Kullback-Leibler may be an alternative version of this model.

3 The generalized Bregman Kullback-Leibler may be the alternative version of this model.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2023 doi:10.20944/preprints202310.0230.v1

and random variable with a compact support (Golan et al., 1996) and $2 < M < \infty$ possible outcomes, it can be estimated using , B_{kl} i.e.:

$$B_{kl} = \sum_{m=1}^{M} p_{klm} v_{klm} \tag{3}$$

where p_{klm} is the probability of the outcome v_{klm}, and the probabilities must be non-negative and added to one. Similarly, by treating each element $e_{\bullet l}$ (which affects the total uncertainty of the sector efficiency) as a e finite, discrete random variable with a compact support and $2 < J < \infty$ possible results centred at zero, we can express $e_{\bullet l}$ as:

$$e_{\bullet l} = \sum_{j=1..J} r_{\bullet lj} z_{\bullet lj} \tag{4}$$

As mentioned, it can be assumed that each previous entire row of errors has been evaluated, $\omega_{k\bullet}$ and a similar support space should be constructed as follows:

$$\omega_{k\bullet} = \sum_{s=1}^{S} \mu_{k\bullet s} \upsilon_{k\bullet s}$$

where $r_{\bullet lj}$ and $\mu_{k\bullet s}$ are the outcome probabilities in the support spaces respectively. z_{klj} v_{ls} $j = 1...J$ $s = 1...S$. Therefore, $k = 1...K$ and $l = 1...L$ represent respectively the indexes of the number of row and column whose coefficient sum were estimated with errors. Moreover, e the term error ω is empirically set around the empirical standard error of the stated variables and a priori represent the Bayesian hypothesis. The choice of error limits, of course, depends on their own properties. In this study, their sets were determined by Chebyshev's inequality [31] with the boundaries of the support space ranging from -3 to +3. Notice that in spite of this Gaussian property of the priors, posterior probabilities in the support space may represent a class of a non-Gaussian distribution, in particular a PL.

The element v_{klm} constitutes an a priori information provided by the researcher while p_{klm} is an unknown probability generating the true parameter β_{kl} the value of which must be determined by solving a non-extensive cross-entropy econometrics problem. In matrix notation, let us rewrite $\beta = V \cdot P$ with $p_{klm} \geq 0$ and $\sum_{k=1}^{K} \sum_{j>2...J} p_{klm} = 1$. Also, let $e = r \cdot z$ with $r_{\bullet lj} \geq 0$ and $\sum_{j>2...J} r_{\bullet lj} = 1$ for K and L the number of rows and columns and J the number of data points over the support space for the error terms inside the regional cross-sector matrix. The same conditions of normality can be easily formulated for any vector of column sums. Next, the cross-entropy econometric estimator of the Tsallis entropy can be presented as follows:

$$MinH_q(p // p^0, r // r^0, \mu // \mu^0) \equiv \alpha \sum p_{klm} \frac{\left[p_{klm} / p^o{}_{klm}\right]^{q-1} - 1}{q-1} + \beta \sum r_{\bullet lj} \frac{\left[r_{\bullet lj} / r^o{}_{\bullet lj}\right]^{q-1} - 1}{q-1} + ... + \delta \sum \mu_{k\bullet s} \frac{\left[\mu_{k\bullet s} / \mu^o{}_{k\bullet s}\right]^{q-1} - 1}{q-1} \tag{5}$$

Subject to

$$Y_{\bullet l} = ccj.l \sum_{k} ([Y_{\bullet l} P_{kl}]' + e_{\bullet l})$$
$$= \left[ccj.l \sum_{k}^{K} \left[\left(\sum_{m>2}^{M} Y_{\bullet l}{}' v_{klm} \left(p_{klm}{}^{q}\right) \right]' + \sum_{j=1..J} r^{q}{}_{\bullet lj} z_{\bullet lj} \right) \right] \tag{6}$$

$$H_{k\bullet} = C_{k\bullet}(X_{k\bullet} + \omega_{k\bullet}) = C_{k\bullet}\left(X_{k\bullet} + \sum_{s=1}^{S} \mu^{q}{}_{k\bullet s} \upsilon_{k\bullet s}\right) \tag{7}$$

$$\sum_{k..K} H_{k\bullet} = \sum_{l..L} Y_{\bullet l} \quad (8)$$

$$\sum_{k=1}^{K} \sum_{j>2...J} p_{klj} = 1 \quad (9)$$

$$\sum_{j>2...J}^{J} r_{\bullet lj} = 1 \quad (10)$$

$$\sum_{s>2...S}^{S} \mu_{k\bullet s} = 1$$

where:

$Y_{\bullet l}$: indicates each sum per column l (values observed by energy sector l , including unknown errors),

$H_{k\bullet}$: each row total (observed values per province k) corrected for $C_{k\bullet}$,

$X_{k\bullet}$: total energy intensity indicators by region affected by unknown errors;

p_{kl} : probabilistic structure of energy efficiency ratios by sector and region,

$C_{k\bullet}$ is a positive scaling factor related to Province climatic factor(average annual temperature by province) to match the totals of energy intensity and energy efficiency indicator levels,

$CC_{\bullet j}$ is the arbitrary additional scaling factor representing other random variables but $C_{k\bullet}$ to balance row and column sums

" $\bullet$ ": refers to a variable bound to a row or column total, depending on the context.

Non-extensive statistics use a number of binding forms in which expectations can be set. The above model uses Curado-Tsallis (C-T) constraints [32,33], the general form of which is as follows:

$$\langle y_q \rangle = \sum_i p_i^q y_i$$

The parameter q , as already mentioned, represents the Tsallis parameter . As Table 1 suggests, the $C_{k\bullet}$ remains a scaling factor related to province climatic factor contributing to match the totals of energy intensity and energy efficiency indicator levels that are known with uncertainty. Unlike the factor $C_{k\bullet}$ whose role has just been expressed above, the factor $CC_{\bullet j}$ is the arbitrary additional scaling factor representing other random variables but $C_{k\bullet}$ to complementarily contribute balancing row and column sums. This is because the $C_{k\bullet}$ which explains climatic differences within provinces could not alone play the balancing role since other factors may exist in explaining the difference between energy intensity and energy efficiency. This point has been alluded to in the introduction Section when talking about the relationships between energy intensity and energy efficiency. Still the sum per column (i.e.per sector) of the posterior probabilities is constrained to unity, given $CC_{\bullet j}$. In the above model [equations 5 - 10], their values will play the role of new Bayesian data discriminating in favor of new inferential evidence. We must recall that when no new data are included in the cross-entropy model, the results correspond to those from formulating a maximum entropy principle, without additional conditions, except those of normality.

Above $H_q\left(p // p^0, r // r^0, \mu // \mu^0\right)$ is nonlinear and measures entropy in the model. The relative entropy of the three independent terms (respectively the three posteriors p , r , μ and corresponding priors p^0 r^0 μ^0) are then added together using the weights . These are the real positives that amount to unity within the mentioned constraints. The first term known as the one of "parameter precision" takes into account the discrepancies between the estimated parameters α, β, δ and the prior parameters (usually defined in the support space). The second and third terms "ex-post predictions" include the empirical error term as the difference between the predicted and observed data values(see the last row and column in tab.1) in the model. The first component of the criterion function can relate to the structure of table parameters, the second component to errors in row totals, and the last component to errors in column totals.

We must find a minimum discrepancy between priors and posteriors, while respecting the restrictions imposed by the system.

It should be noted that the estimates of the model and their variances should be influenced not only by the length of the support space, but also by the spatial scale effect, i.e. the number of affected point values [24]. The greater the number of these points, the better the prior information- i.e. nonlinear starting points, about the system.

Second, ω_i random errors (see Equation 7) explain errors in data collection and processing and are not necessarily related to the Gaussian distribution being itself a particular case of a PL. Traditionally, when it comes to Bayesian formulations and relative entropy, it is worth noting that both models will lead to similar results if and only if the real expected error associated with the data generation system is zero in the symmetric support space around zero(see e.g. [30]). Similarly, the results of Gibbs-Shannon cross-entropy and Tsallis' non-extended cross-entropy will match when errors included in the model are not correlated, and the system distribution will then evolve towards the Gaussian attractor e.g., [34–36].

With respect to the confidence interval of the parameters, equation 11 shows the non-additivity of the Tsallis entropy for two - probable - independent systems, one related to the probability distribution of the parameters and the other to the probability distribution of error perturbations:

$$S(\hat{\Pr}) = [S(\hat{p}+\hat{r})] = \{[S(\hat{p})+S(\hat{r})] + (1-q)\cdot S(\hat{p})\cdot S(\hat{r})\} \quad (11)$$

Where:

$$S(\hat{p}) = -\left[1-\sum_k\sum_m (p_{klm})^q\right] / \left[K\cdot\left(M^{1-q}-1\right)\right]$$

$$S(\hat{r}) = -\left[\left(1-\sum_l\sum_j r^q_{\bullet lj}\right)\right] / \left[L\cdot\left(J^{1-q}-1\right)\right]$$

$S(\hat{\Pr})$ is then the sum of the normalized entropy associated with the parameters of the $S(\hat{p})$ model and the term of perturbation $S(\hat{r})$. Similarly, the last value is $S(\hat{r})$ obtained for all observations l, with J the number of data points above the support of the estimated probabilities r related to the time length of errors.

The values of these normalized entropy indices range from zero to one. The values, close to one, indicate a weak information variable, while lower values indicate a parameter that is estimated to be more informative $\hat{\beta}_{kl}$ in the model.

1. **Outputs and comment**

We will illustrate the empirical basis of the theoretical model developed in the above sections by applying Polish energy efficiency to the case. Table 1 - whose row and column symbols correspond to equation Eq.6 - is presented to illustrate the extent to which a researcher can have limited information about quantity and quality before solving an inverse and ill-posed inverse problem such as the one in this article.

Table 1. Sector and province energy efficiency ratios(2021) .

Polish	**Energy demand sectors**						
province	**industry**	**Transport**	**..**	**Households**	**..**	**Services**	**Average energy intensity coefficients (GWh\mln zl value added)**
Lodz	$Y_{\bullet 1}P_{11}$	$Y_{\bullet 2}P_{12}$	..	$Y_{\bullet l}P_{1l}$	..	$Y_{\bullet L}P_{1L}$	0.08
Mazowieckie	$Y_{\bullet 1}P_{21}$	..	..	$Y_{\bullet l}P_{2l}$	..	$Y_{\bullet l}P_{2L}$	0.05
Malopolskie	..	..	..	..	..	..	0.06
Silesian	..	..	..	..	..	..	0.09
Lublin	..	..	..	..	..	..	0.07
Podkarpackie	..	..	..	..	..	..	0.06
Podlaskie	..	..	..	..	..	..	0.06
Swietokrzyskie	..	..	..	..	..	..	0.09
Lubuskie	$Y_{\bullet 1}P_{k1}$	..	..	$Y_{\bullet l}P_{kl}$	..	.. $Y_{\bullet L}P_{kL}$	0.07
Wielkopolska	..	..	..	..	..	..	0.05
Zachodnia pomors	..	..	..	..	..	..	0.07
Dolnoslaskie	..	..	..	..	..	..	0.07
Opole	..	..	..	..	..	..	0.11
Kujawska-pomorska	..	..	..	..	..	..	0.07
Pomorska	..	..	..	..	..	..	0.06
Warminsko-mazurskie	$Y_{\bullet 1}P_{K1}$	..	..	$Y_{\bullet l}P_{Kl}$	..	.. $Y_{\bullet L}P_{KL}$	0.06
Average energy efficiency ratio	0.372	0.612		0.621		0.689	

Sources: own based on Odyssey-Mure and Polish Institute of Statistics(GUS).

We have two aggregated energy indicators each from the Polish energy demand sectors per province. The problem in hands stands as [21] a discrete problem in a multidimensional space, leading to (k -1)x(l -1) degree of freedom that would illustrate the case of the standard inverse problem. In this problem, as illustrated in the Table 1 [37,38] we have 3x15=45 degrees of freedom related to the number of energy efficiency indicators per sector and province. Moreover , as alluded to before, the row total and the column total are different as probably the result of two separate data sources with different nature and scale.

As the above pages made us aware of, this kind of problem corresponds the best to the philosophy naturally contained in the principle of maximum entropy that we have implemented to forecast the energy efficiency ratios displayed in Table 2.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2023 doi:10.20944/preprints202310.0230.v1

Table 2. Post-entropic 2020 energy efficiency ratio forecasts and efficiency progress in 2020\2021(%).

	Industry	Transport	Households	Services	Average efficiency ratio per province	Efficiency ratio Change_2021\2020
Lodz	0.4	0.667	0.677	0.754	0.625	**-0.12**
Slaskie	0.429	0.729	0.741	0.83	0.682	**0.81**
Mazowieckie	0.302	0.471	0.477	0.521	0.443	**-3.16**
Wielkopolska	0.333	0.526	0.533	0.583	0.494	**7.49**
Dolnoslaskie	0.369	0.602	0.61	0.676	0.564	**-1.29**
Opole	0.482	0.847	0.861	0.974	0.791	**2.43**
Malopolskie	0.336	0.536	0.543	0.596	0.503	**-2.44**
Swietokrzyskie	0.445	0.764	0.777	0.873	0.715	**5.32**
Zachodnia pomors	0.373	0.61	0.618	0.685	0.572	**0.04**
Kujawska-pomorska	0.37	0.604	0.613	0.678	0.566	**-0.88**
Pomorska	0.336	0.536	0.543	0.597	0.503	**-2.44**
Lublin	0.369	0.602	0.61	0.675	0.564	**-1.29**
Podkarpackie	0.336	0.536	0.543	0.597	0.503	**-2.33**
Lubuskie	0.386	0.636	0.645	0.716	0.596	**4.07**
Podlaskie	0.336	0.536	0.543	0.597	0.503	**-2.39**
Warminsko-mazurskie	0.339	0.541	0.548	0.602	0.507	**-1.53**
Sector average ratio	0.371	0.609	0.618	0.685		
Progress_2021/2020(in %) ratio	**0.067**	**0.369**	**0.385**	**0.548**		

Let us now comment on the model outputs presenting the final model solution about the cross "energy efficiency ratios " (column 5 of Table 2) by sector and province. As presented in the theoretical model(equations 5-10), in addition to the classical normality conditions and moment consistency, the earlier explained random factors $CC_{\bullet j}$ and $C_{k\bullet}$ allowed the system for balancing. The first factor is related to various variables differentiating energy intensity from energy efficiency except the climatical factor represented by $C_{k\bullet}$ As already said, these calculated ratios in the above table are not deterministic since the priors are not known with certainty. Therefore, we treat them as random variables, from which the posteriors result from the optimization of the model. It is worthy to recall that the presented energy efficiency scores stand for the overall energy efficiency score, i.e. a combination of the three already presented components: the energy efficiency level, the energy efficiency progress (i.e. energy efficiency trends) and the energy efficiency policies. The not presented in this paper prior matrix of energy efficiency ratio has been initiated on the basis of knowledge of province energy intensity index and sector energy efficiency ratio averages. Next, using proportions of energy intensity index(last column of tab.1) of different provinces, we computed the initial energy efficiency ratio by sector, so as to sum up to the total of energy efficiency ratios-known with error, available from the last row of the same Table 1. In probabilistic terms, we have assumed the energy efficiency ratio to be a uniform distribution across provinces, thus the last column of energy intensity as a marginal probability. The assumption behind this procedure is that the province with lower or higher average energy intensity index will have lower or higher average energy efficiency ratio respectively, irrespective of the considered sector. Doing so, we enabled the nonlinear mathematical system to start the search of the model global optimal solution from the best starting points leading to a quick convergence. The post-entropy posterior ratios in Table 2 are normalised and the higher the value the lower the energy efficiency level for a given sector and\or province. The obtained forecasts are empirically close to real world expectations of energy efficiency ranking within polish provinces. We notice that in the industry sector the provinces Mazowieckie (including Warsaw) and Wielkopolski (including Poznan) have the lowest energy efficiency ratios, respectively around 0,320 and 0,333. The highest ratio is shown in the provinces Swietokrzyskie and Opole with ratios around to 0,445 and 0,482 respectively. Globally, we notice that Mazowieckie displays the highest efficiency in all sectors while Opole displays the lowest efficiency. We notice too that industry sector globally remains the most efficient among all sectors.

Table 3 presents the level at which the model has discriminated the prior in favor of the post cross-entropy solution given the new model consistency moments and normality conditions. Consequently, we notice in that table the highest post cross-entropy discriminating values through all sectors in the case of Opole in comparison with the remaining sectors. It is worthy to recall that, as described in the precedent paragraphs, the prior values presented the same structure for different sectors. The cross-entropy formalism has discounted non valuable information to just retain information fitting to corresponding sectors and province given initial information on the energy intensity indicator and the climatic factor.

Table 3. Model cross matrix increment (in %) between priors and posteriors.

	Industry	Transport	Households	Services
lod	-4.901	-4.569	-4.566	-4.558
sla	-9.481	-8.137	-8.098	-7.812
maz	14.89	11.26	11.13	10.104
wiel	14.89	11.26	11.13	10.104
dol	0.405	-0.418	-0.455	-0.76
opo	-17.153	-14.151	-14.051	-13.297
mal	6.77	4.634	4.552	3.9
swi	-9.481	-8.137	-8.098	-7.812
zac	0.405	-0.418	-0.455	-0.76
kuj	0.405	-0.418	-0.455	-0.76
pom	6.77	4.634	4.552	3.9
lub	0.405	-0.418	-0.455	-0.76
pod	6.77	4.634	4.552	3.9
lubu	0.405	-0.418	-0.455	-0.76
podl	6.77	4.634	4.552	3.9
war	6.77	4.634	4.552	3.9
AEE ratio	-0.289	-0.908	-0.937	-1.178

Source: own calculations.

Following the formulation of the equation 11, the confidence interval of the model is a normalized values ranging between zero to one. The values, close to one, indicate weak informative variables, while lower values indicate a model whose estimates are more informative. In the present case, the equivalent to the classical coefficient of determination $\widehat{R^2}$ is equal to 0.011, then much closer to zero than unit. Then, the proposed model variables have enough discriminated far from the priors, given all constraints encountering the energy efficiency system defined in this paper.

1. Concluding remarks

The objective of this article was to carry out a forecast of energy efficiency ratios in conditions where statistical data are not only missing but also from contradictory sources. In empirical research, this case is rather a rule than an exception. In most situations, this will lead to a nonlinear ill-posed inverse problem for which traditional statistics fail to find an analytical solution unless commodity additional hypotheses are made as is often the case in empirical econometrics.

In the present paper, we have analysed this problem at the level of the Polish 16 regions for four sectors of energy consumption. We started with sparse information concerning the annual sectoral energy efficiency ratio averages aggregated at the national level. We disposed as well as the annual averages of energy intensity aggregated at the provincial level. The problem therefore consisted of bringing together these two aggregates of different nature to produce reliable information on interprovincial sectoral energy efficiency coefficients.

The forecasts obtained on these ratios seem to be in line with the expectations of those who know realities behind the Polish energy sector.

This work more or less described in details the generalized cross-entropy technique and illustrated its applicability in the field of energy experimental statistics. Furthermore, thanks to this approach, this problem could also be solved in the case of further disaggregation, for example to the NUTS 3 level. This is of capital importance because it is at the more disaggregated levels that economic actors act.

In this work, the explanatory variable used was the climatic factor in the form of average annual temperature by province. Depending on the countries studied, this variable or few other variables presented in this paper, like the structure and nature of industry in different regions, could much better influence the difference between energy intensity and energy efficiency. It would be interesting to simulate future values of these efficiency ratios by linking energy intensity with these influential variables. Comparing such an information on energy efficiency across sectors and regions will be of significant benefit to target gray areas and redefine more appropriate energy policies fitting to social values inside a disaggregated region. In particular, the outputs of the model could allow public energy institutions for monitoring the relative progress of each region or province in their endeavour to optimize the energy use given their own specifical conditions. Finally, an extensive forecast at European or larger scale of these indicators through experimental official statistics could allow for a better management of energy saving as targeted by the European or worldwide climate change mitigation institutions or projects.

References

1. S. Tsemekidi Tzeiranaki, P. Bertoldi, D. Paci, L. Castellazzi, T. Ribeiro Serrenho, M. Economidou and P. Zangheri, "Energy Consumption and Energy Efficiency trends in the EU-28, 2000–2018," Office of the European Union, Luxembourg, 2020.
2. IRENA, Global Energy Transformation: A Roadmap to 2050, Abu Dhabi, 2021
3. IRENA, Reaching Zero with Renewables.2020, Abu Dhabi: IRENA, 2021.
4. Jakimowicz, "The Material Entropy and the Fourth Law of Thermodynamics in the Evaluation of Energy Technologies of the Future," *Energies*, vol. 16 , no. 9, 2023.
5. Jakimowicz and D. Rzeczkowski, "The Future of the Energy Sector and the Global Economy: Prosumer Capitalism and What Comes Next," *Energies*, vol. 15 , no. 23, 2022.
6. E. Union. [Online]. Available: https://www.odyssee-mure.eu/.
7. P. I. o. Statistics, "Polish Institute of Statistics(GUS)," 2020-2021. [Online]. Available: https://bdl.stat.gov.pl/bdl/pomoc/stanzasilenia?active=0#.
8. L. &. K. A. Proskuryakova, "Measuring energy efficiency: Is energy intensity a good evidence base?," *Elsevier*, vol. vol. 138(C), pp. 450-459., 2015.
9. E. E. &. R. Energy, "Energy Efficiency vs. Energy Intensity Analysis," Waszyngton DC, 2023.
10. P. Paolo Renna and M. S., "An Energy Efficiency Index Formation and Analysis of Integrated Energy System Based on Exergy Efficiency," *Applied Sciences*, vol. 11, no. 16, July 2021.

11. Househam, "ENERGY EFFICIENCY METRICS," IIEC-Africa, South-Africa.
12. Shuanjie Li, "Energy Efficiency Measurement: A VO TFEE Approach and Its Application," *Sustainability,* vol. 13, no. 4, p. 1605, 2021.
13. Shuangjie Li, "Shuangjie Li; Wei Wang; Hongyu Diao; Liming Wang. Measuring the Efficiency of Energy and Carbon Emissions: A Review of Definitions, Models, and Input-Output Variables," *Energies,* vol. 15, p. 962, 2022.
14. [Online]. Available: https://www.eesi.org/.
15. E. European Union, "EU," [Online]. Available: https://www.europarl.europa.eu/at-your-service/en/stay-informed/factsheets-on-the-european-union.
16. IEA, "Energy Efficiency Indicators: Fundamentals on Statistics," IEA, Paris, 2014.
17. UNDP, "CHAPTER 3: energy efficiency," in *Global tracking framework*, United Nations.
18. Odyssee-Mure, "METHODOLOGY EUROPEAN ENERGY EFFICIENCY SCOREBOARD," Luxemburg, 26 November 2020.
19. H. Su, Q. Huang and Z. Wang, "An Energy Efficiency Index Formation and Analysis of Integrated Energy System Based on Exergy Efficiency," *Front. Energy Res.*, 20 August 2021.
20. T. A.N. and A. V.I, "Solutions to poorly conditioned problems," *John Wiley & Sons, ,* 1977.
21. J. E. T, Probability Theory: The Logic Of Science, Washington: Washington University, 1944.
22. S. C. E., "A Mathematical Theory of Communication,, , , 1948.," *Bell System Technical Journal,* pp. s.27, s. 379-423 & 623-656, 1948.
23. S., "Econometrics of non-extended entropy: new statistical characteristics of constant elasticity of substitution-related models," *Entropy,* vol. 16, pp. 2713-2728, 2014.
24. G. M. D. Golan A., Maximum Entropy Econometrics: Robust Estimation with Limited Data, Chichester, Angleterre, : Wiley , 1996.
25. E.-S. M. Robinson S., Updating and Estimating a Social Accounting Matrix Using Cross Enropy Methods, vol. 13, Economic Systems Research , 2001, pp. 47-64.
26. F. H. d. C. VELHO, Elcio H. SHIGUEMORI and F. M. R. i. J. C. CARVALHO, "A Theory of Unified Regularization: The Principle of Non-Extensive Maximum Entropiy. Differential equation," Vols. 25, n. 2-3, Appl. Math., 2006, pp. 307-330.
27. G. X., The Laws of Power in Economics and Finance, Washyngton: BER, 2008.
28. K. S., Information Theory and Statistics, NY: John Wiley and Sons, 1959.
29. Jakimowicz, "The Role of Entropy in the Development of Economics," *Entropy,* vol. 22, no. 4, p. 452, 2020.
30. S. Bwanakare, Non-Extensive Entropy Econometrics for Low Frequency Series: National Accounts-Based Inverse Problems, Warsaw: De Gruyter Open Poland, 2018.
31. Z. Shen and P. J. M., Maximum enropy and Bayesian approaches to the ratio problem,, vol. 104, Journal of Econometrics , 2001, pp. 289-313.
32. P. F., The Three Sigma Rule, vol. 48, American Statistical Association, 1994, pp. 88-91.
33. V. R. C. and P. A., "Deformed Statistics Kullback-Leibler Divergence Minimization within a Scaled Bregman Framework," 2011.
34. S. and B. G. B., Constraints and Relative Entropies in Nonextensive Statistical Mechanics, 2004.
35. Tsallis, Introduction to Non-extensive Statistical Mechanics: Approaching a Complex World, Berlin: Springer, 2009.
36. S. Bwanakare, "Solving a generalized constant elasticity of substitution function of production: A non ergodic maximum enropy principle," *Acta Phys. Pol.,* vol. 123, p. 502, 3/2013.
37. S. Bwanakare, "Non-extensiveeconometric model of entropy (NO): a case of labour demand in Podkarpackie Voivodeship," *Acta Phys. Pol.,* vol. 117, p. 647, 4/2010.
38. EU, "Eurostat," 2021. [Online]. Available: http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/themes.
39. GUS, 2021. [Online]. Available: http://old.stat.gov.pl/gus/publikacje_a_z_PLK_HTML.htm.