

2 Ill-posed Inverse Problem Solution and the Maximum Entropy Principle

2.1 Introduction

As explained in the introduction, many economic relationships are characterized by indeterminacy. This may be because of long-range feedback and complex correlations between source and targets, thus rendering causal relationships more difficult to investigate.

In this part of the work, the formal definition of the inverse problem will be discussed. A Moore-Penrose approach will be presented for solving this kind of problem and its limits will be stressed. The next step will be to present the concept of the maximum entropy principle in the context of the Gibbs-Shannon model. Extensions of the model by Jaynes and Kullback-Leibler will be presented and a generalisation of the model will be implemented to take into account random disturbance. The next step will concern the non-ergodic form of entropy known in the literature of thermodynamics as non-extensive entropy or non-additive statistics. There will be a focus on Tsallis entropy, and its main properties will be presented in the context of information theory. To establish a footing in the context of real world problems, non-extensive entropy will be generalized and then random disturbances will be introduced into the model. This part of the work will be concluded with the proposition of a statistical inference in the context of information theory.

2.2 The Inverse Problem and Socio-Economic Phenomena

An inverse problem, e.g., Thikonov et al., (1977), Bwanakare (2015), Golan et al., (1996) explains a situation where one tries to capture the causes of phenomena for which experimental observations represent the effect.

The essence of the inverse problem is conveyed by the expression:

$$Y = X\beta + \xi \quad (2.12)$$

or its equivalent in continuous form:

$$Y(\zeta) = \int_D g(X)B(X, \zeta)dX + b(\zeta) \quad (2.13)$$

where

X represents the state space,

Y designates the observation space,

D defines the Hilbert support space of the model,

B is the transformation kernel linking measures X and Y ,

$b(\zeta)$ displays random error process.

In classical econometrics, when given a state X , an operator B and, as happens most of time, a disturbance term (ζ), what is Y ? This is referred to as a forward problem. In social science, one must often cope with the above random (Gaussian or not) disturbance term, and this usually complicates matters in spite of significant, recent developments in econometrics, particularly concerning stochastic time-series analysis (Engle & Granger, 1987). Furthermore, the inverse question is more profound: Given y and a specific B , what is the true state X ?

If B should also be a functional of X , the problem becomes arbitrarily complex. Correlation between (ζ) and X will be at the base of such additional complexity.

Everyday, psychologists cope with such inferential problems. Patients display identical symptoms from different sicknesses. Health practitioners need more historical (*a priori*) information on patients to try to find the solution.

In economics, the same national output growth rate may result from different combinations of factors. One of the main problems encountered by practicing economists is isolating the causes of economic phenomena once they have occurred. In most cases, the economist becomes inventive in finding an appropriate hypothesis before trying to solve the problem. As an example, in the case of a recession or financial turbulence, it is usually difficult to point to principal causes and fix them. Schools of economics suggest different, even contradictory, solutions—the legacy of its inverse problem nature.

In empirical research, many techniques exist to try to solve the inverse problem. In the context of the present work, the presentation will be limited to those more applicable to matrix inversion, like the Moore-Penrose pseudo-inverse approach, and, naturally, maximum entropy based approaches. The approach better known in economics for updating national accounts on the basis of bi-proportionalities will then be added to these two techniques.

2.2.1 Moore-Penrose Pseudo-Inverse

Let us consider the discrete and determinist case and rewrite (2.12) as follows:

$$Y = XB = X\rho \quad (2.14)$$

In the right equality reflects the case where we have to deal with a ratio or probability parameter, for example, after reparametrizing B . We then have:

$$\begin{aligned} \rho &= YS \Leftrightarrow Y = XBS \\ \hat{\rho} &= BY \Leftrightarrow Y = XYV \\ Y &= X\hat{\rho} = YXV = XBX\rho, \end{aligned} \quad (2.15)$$

which means:

$XBX = X$ and V , representing the generalized inverse matrix (Golan, 1996), (Kalman, 1960).

This is a matrix with the symbol B^+ that satisfies the following requirements:

$$\begin{aligned} B B^+ B &= B, \\ B^+ B B^+ &= B^+, \\ B^+ B &\text{ is symmetric,} \\ B B^+ &\text{ is symmetric.} \end{aligned}$$

Following Theil (1967), a unique B^+ can be found for any matrix: square, non-singular or not. When the matrix B^+ happens to simultaneously be square and non-singular, then the generalized inverse will be the ordinary inverse B^{-1} . The problem that interests us is the over-determined system of equations

$$Y = XB$$

where B has n rows, $K < n$ columns and column rank equal to $R \leq K$.

If we retain the particular case when R equals K to ensure the existence of $(B'B)^{-1}$, then the generalised inverse of B is

$$B^+ = (B'B)^{-1}B'$$

as can be easily verified. A solution to the system of equations can be presented as:

$$X = B + Y.$$

Following Green (2003, pp. 833), we note in this case that the length of this vector minimizes the distance between Y and BX , according to the least squares properties method. This distance will naturally remain equal to zero if y lies in the column space of B .

If we now retain the more general case where B does not have full rank, the above solution is no longer valid and a spectral decomposition using the reciprocals of the characteristic roots is involved to compute the inverse which becomes:

$$B^+ = C_1 A_1^{-1} C_1' B'$$

where C_1 are the R characteristic vectors corresponding to the non-zero roots arrayed in the diagonal matrix A_1 .

The next and last case is the one where B is symmetric and singular, that is, with the rank $R \leq K$. In such a case, Moore-Penrose inverse is computed as in the preceding case but without pre-multiplying by B' . Thus, for such a symmetric matrix,

$$B^+ = C_1 A_1^{-1} C_1', \quad (2.16)$$

with A_1^{-1} being a diagonal matrix of the reciprocals of the non-zero roots of B .

It is important to note that only matrix B with full rank ensures a minimum distance between Y and BX . In other cases, there may exist an infinite number of combinations of elements of matrix B or $\hat{\rho}$ which satisfy (2.14).

To conclude, in spite of strong advantages of the Moore-Penrose generalised inverse, outputs will not always reflect an optimal solution.

2.2.2 The Gibbs-Shannon Maximum Entropy Principle and the Inverse Problem

Let us introduce the concept of Shannon entropy by continuing with the case of pure linear inverse problem solution discussed above. The simplest (one dimensional case) example is the Jaynes dice inverse problem.

If a dice is fair, and we throw it a large number of times n , with k different output modalities⁹ ($k = 1, \dots, K$), the expected value will be 3.5, as from a uniform distribution with probability f_k equal $1/6$. How can one infer about p_k if we have 'loaded' (unfair) dice and the expected value of the trial becomes:

$$4.5 = \sum_{k=1}^K k p_k \quad (2.17)$$

where frequencies p_k is $\frac{n_k}{n}$?

In this case, the central question is: Which estimate of the set of frequencies would most likely yield this number? The problem is underdetermined since there are many sets of f_k that can be found to fit the single datum of equation (2.17). Here we have to deal with a multinomial distribution where the multinomial coefficient W is given by:

$$W = \frac{N!}{N p_1! N p_2! \dots N p_k!} = \frac{N!}{N k_1! N k_2! \dots N k_k!}$$

Deriving and using the Stirling approximation $\ln x! \cong x \ln x - x$ for a large number of N , we get the Shannon entropy formulation:

$$\text{Max}_p H(p) = - \sum_{k=1}^K p_k \ln p_k \quad (2.18)^{10}$$

In the case of a die, parameter K equals 6, and W is the multinomial coefficient, i.e., the number yielding a particular set of frequencies among 6^N possible outcomes.

⁹ Generally, if the number of trials is equal to n , we will have n_k possible outputs corresponding to each modality k with $n = \sum_k n_k$. Thus, the frequency $p_k = \frac{n_k}{n}$ is related to each modality k .

¹⁰ Note that the generalized form of Shannon entropy in the continuous case has the form: $\text{Max}_{f(y)} H(f(y)) = - \int f(y) \log f(y) dy$.

We need only find the set of frequencies maximizing W in order to find the set that can be realized in the greatest number of ways. This is the most plausible combination in the case of fair dice.

This turns out to convey the same logic as maximizing Shannon Gibbs entropy. Thus, starting from two pieces of information that is, the number k equal to six and N a large number of trials we are able to derive six probabilities related to a die distribution.

Next, Jaynes (1994) maximized the Shannon function through the restriction of consistent information at hand. This opened entropy theory application to many scientific fields, including the social sciences.

Thus, if we add to the formulation (2.18) the moment-consistency and the adding up-normalization constraints, we then get:

$$\text{Max}_p H(p) = -\sum_{k=1}^K p_k \ln p_k \quad (2.19)$$

subject to:

$$\sum_{k=1}^K p_k f_t(x_k) = y_t, 1 \leq t \leq T \quad (2.20)$$

$$\sum_{k=1}^K p_k = 1 \quad (2.21)$$

where $\{y_1, y_2, \dots, y_T\}$ denotes a set of observations (e.g., aggregate accounts or their averages) being consistent with a function $f_t(x_k)$ of explicative variables weighted by a corresponding distribution of probabilities $\{p_1, p_2, \dots, p_K\}$. As usually happens, T is less than K , and the problem is ill-posed (underdetermined).

Two main results emerge from the above formulation. First, if all events are independent or quasi-independent (locally dependent) and equally probable, then the above entropy is a *linear function of the number of the possible system states* and then is *extensive*¹¹.

A second fundamental result is connected with information theory and suggests that a Gaussian variable has the largest entropy among all random variables of equal variance (see Papoulis, 1991 for proof). In the next chapter on non-extensive entropy, a measure to assess the divergence of a given distribution from Gaussian distribution will be presented.

¹¹ For this reason, as earlier alluded to, the Gibbs-Shannon entropy is called *extensive*. In reverse, as it will be commented on in the coming sections, the hypothesis of long-range correlation between events leads to the concept of *non-extensive* entropy (e.g., Tsallis entropy) suggesting an entropy no longer being a linear function of data.

Table 1: Recovering probability distribution of an unbalanced die through the maximum entropy principle.

P_1	P_2	P_3	P_4	P_5	P_6
0.054	0.079	0.114	0.166	0.240	0.348

Coming back to the dice case, maximization of Shannon entropy in (2.19), that is $H(P) = -P' \ln P$ under Jaynes consistency, leads to the distribution presented in Table 1. To solve this inverse problem of six unknowns, the only two pieces of information available are the expected value—from the experiments in this example—assumed to be equal to 4.5 and the information that the probability of different possibilities adds up to one. However, since we are dealing with unbalanced dice, we have no idea about the distribution.

The next chapters extend the Shannon-Gibbs-Jaynes maximum entropy principle with Kullback-Leibler relative entropy. The next to the last targeted presentation will deal with the general linear entropy model, that is, the one with a stochastic component. To conclude, Tsallis power law distribution to generalize Kullback-Leibler cross-entropy will be considered.

2.2.3 Kullback-Leibler Cross-Entropy

Kullback (1959), Good (1963) extended the Jaynes-Shannon-Gibbs model by formulating the principle of minimum (cross or relative) entropy. Using an *a priori* piece of information q about unknown parameter p , the resulting formulation is as follows:

$$\text{Min}_- H(p, q) = \sum_{k=1}^K p_k \ln(p_k / q_k) = p' \ln p - p' q \quad (2.22)$$

under restrictions:

$$Y = XP \quad (2.23)$$

$$P'1 = 1 \quad (2.24)$$

where $p = (p_1, \dots, p_K)'$, $q = (q_1, \dots, q_K)$.

These restrictions are the same as those presented earlier. In the criterion function (2.22), *a posteriori* and *a priori* vectors or matrices p and q are confronted with the purpose of measuring entropy reduction resulting from exclusive new content of data information.

One should note that when q is fully consistent with moments, then $p = q$ and the distribution becomes uniform with $q_k = 1/K$. This leads to the solution of the maximum entropy principle.

Thus, the cross-entropy principle stands for a certain form of generalization of maximum entropy. Relation (2.22) above is an illustration of the previous Kullback formulation in (2.8) as a mean information from (2.23) and (2.24) for discrimination in favour of p against q .

2.3 General Linear Entropy Econometrics

In social science, it is rare to encounter the situation described by the relation (2.14) where the random term is meaningless as is often encountered in the experimental sciences. Social phenomena are particularly affected by stochastic components. Let us rewrite it below in its generalized form:

$$y_i = \sum_{j=1}^K B_j X_j + \zeta_i \quad (2.12')$$

with the random term $\zeta_i \in e$ and

$i = (1, \dots, I)$ (I being the number of observations); K is the number of model parameters to be estimated.

2.3.1 Reparametrization of Parameters

Following Golan et al., (1996), we first reparametrize the above generalized entropy model (2.12').

We treat each B_j ($j = 1, \dots, K$) as a discrete random variable within a compact support and $2 < M < \infty$ possible outcomes. So, we can express B_j as:

$$B_k = \sum_{m=1}^M p_{km} v_{km} \quad \forall k \in K \quad (2.25)$$

where p_{km} is the probability of outcome v_{km} and the probabilities must be non-negative and sum up to one.

Similarly, let us treat each element ζ_i of e as a finite and discrete random variable with compact support and $2 < M < \infty$ possible outcomes centred on zero. We can express ζ_i as:

$$\zeta_i = \sum_{j=1}^J r_{nj} \cdot w_{nj} \quad (2.26)$$

where r_n is the probability of outcome w_n . The term ζ_i , like any prior value in the model, reflects Bayesian properties and is not a fixed value as in the case of classical econometric models. In practice, support sets with three or more points¹² are used to take into account higher moments of the distribution during the process of information recovery.

2.4 Tsallis Entropy and Main Properties

2.4.1 Definition and Shannon-Tsallis Entropy Relationships

This relatively new form of entropy is emerging over an immense area of applications in social science, including economics. One of the fields of interest is modeling and predicting markets of financial returns (Drożdż & Kwapień, 2012), (Grech & Pamula, 2013). Nevertheless, due to the high frequency nature of Big Data in Official Statistics (e.g., Braaksma & Zeelenberg, 2015), the PL-based non-extensive entropy econometrics should be seen as a potential and natural estimation device in this new statistical area. As in statistical physics, socioeconomic random events display two types of stochastic behaviour: ergodic and non-ergodic systems. Whenever isolated in a closed space, ergodic systems dynamically visit with equal probability all the allowed micro-states (Gell-Mann & Tsallis, 2004). However, it seems logical to imagine systems visiting the allowed micro-states in a much more complex way than defined by ergodicity. The financial market is a well-known example of such complex systems, as characterized by multifractal dimensions (Drożdż & Kwapień, 2012), (Grech & Pamula, 2013). Other examples include income distribution inside a given region, evolution of a given disease inside a region, size of cities, or cellular structure. These forms seem to display an organized structure owing to long-range correlation between micro-elements, heavy queues with respect to Gaussian distribution, scale-invariant structures, and criticality. Such phenomena would be better described by a stable law-based Levy process, like power law distribution.

Shannon-Kullback-Leibler Equations (2.22–2.24) are generalized by Tsallis relative entropy formulation. To emphasize consistency among the principal formulations, it is worthwhile to reiterate the statistical theory connection between the above relations and the Kullback relation presented in (2.8) or to some extent (2.9), which

¹² Golan, Judge, and Miller (1996) suggest the Chebyshev inequality as a good starting point to define the error support set: $\Pr[|x| < v\sigma] \geq v^{-2}$ where v is a positive real and x a random variable, such that $E(x) = 0$ while $\text{var}(x) = \sigma^2$. This inequality leads to three-sigma rule (Pukelsheim, 1994) for $v = 3$, i.e., to the probability $\Pr[-3\sigma < x < 3\sigma]$, which is at least 0.88 and higher when x displays a standard normal distribution. Let us remember that this inequality has the additional advantage of being independent of distribution laws.

measures the divergence between two hypotheses H_1 and H_2 . A similar concept will be introduced in the case of non-extensive entropy, which will constitute the final step of Shannon entropy extensions.

Let us generalize the Shannon Gibbs inverse problem through ordinary differential equation characterization (Tsallis, 2009). First, we need to introduce the three simplest—in terms of dynamic complexity—differential equations and their inverse functions,

$$\frac{dy}{dx} = 0 \quad (y(0) = 1). \quad (2.27)$$

Its solution is $y=1$ ($\forall x$), and its inverse function is $X=1$ ($\forall y$).

The next simplest differential equation is

$$\frac{dy}{dx} = 1 \quad (y(0) = 1). \quad (2.28)$$

Its solution is $y=(1+x)$ and its inverse $Y=(x-1)$.

The next higher step in increasing complexity is the differential equation

$$\frac{dy}{dx} = y \quad (y(0)=1). \quad (2.29)$$

Its solution is $y = e^x$, and its inverse is $y = \ln x$.

Note that the latter inverse equation satisfies the additive property:

$$\ln(x_a x_b) = \ln(x_a) + \ln(x_b). \quad (2.30)$$

Following Gell-Mann & Tsallis (2004) and trying to unify the three cases (without preserving linearity), we get:

$$\frac{dy}{dx} = y^q \quad (y(0)=1; q \in \mathbb{R}). \quad (2.31a)$$

We observe that this expression displays power-law distribution form. Its solution is

$$y = [1 + (1-q)x]^{\frac{1}{1-q}} \equiv e_q^x (e_1^x = e^x),$$

and its inverse function is

$$y = \frac{x^{(1-q)} - 1}{1-q} \equiv \ln_q x \quad (\ln_1 x = \ln x). \quad (2.31b)$$

The above represents the non-extensive (Tsallis) entropy formula. Though it will be discussed in the next section, let us immediately show here the relationship between Shannon and Tsallis entropies through the next pseudo-additive property:

$$\ln_q(x_a x_b) = \ln_q(x_a) + \ln_q(x_b) + (1-q)\ln_q(x_a)\ln_q(x_b) \quad (2.32)$$

for $q \rightarrow -\infty$, $q = 0$, $q = 1$ we obtain the three initial cases (2.27 – 2.29), respectively.

In particular for $q = 1$, we then obtain (after using l'Hôpital's rule) the solution of (2.29), the case of Shannon Gibbs entropy. The expression (2.30) states that if two systems x_a and x_b are logarithmically multiplied, the output is the additive sum of these systems in a logarithmic scale. This explains why Shannon entropy is sometimes referred to as additive entropy. This observation has been taken from (2.21) to emphasize that Shannon entropy is a direct function of data. The term q is referred to as “ q -Tsallis.” When it is equal to unity, we reach in this limiting case the Shannon entropy.

Tsallis entropy should now be described and compared with other entropy forms. This description indirectly replies to the question of why Tsallis or Shannon entropy rather than Renyi entropy or another is appropriate for a given problem.

2.4.2 Characterization of Non-Extensive Entropy

2.4.2.1 Correlations

Following Tsallis (2009), suppose we have a system composed of subsystems¹³ A (with W_A possibilities of complexities) and B (with W_B possibilities of complexity). Their joint probabilities can be presented as $\{p_{ij}^{AB}\}$ ($i = 1, 2, \dots, W_A$, $j = 1, 2, \dots, W_B$) and marginal probabilities as

$$p_i^A \equiv \sum_{j=1}^{W_B} p_{ij}^{AB} \text{ (hence } \sum_{i=1}^{W_A} p_i^A = 1) \text{ and}$$

$$p_j^B \equiv \sum_{i=1}^{W_A} p_{ij}^{AB} \text{ (hence } \sum_{j=1}^{W_B} p_j^B = 1).$$

In general,

$$p_i^A p_j^B \neq p_{ij}^{AB} \quad (2.33)$$

if they happen to be equal, then A and B are said to be probabilistically independent. Otherwise, they are *dependent* or *correlated*. Let us then define entropies:

$$S_q(p^A) \equiv S_q(\{p_i^A\})$$

and

$$S_q(p^B) \equiv S_q(\{p_j^B\}).$$

¹³ For models to be presented later, subsystem A can be considered as a data generating process and B as a subsystem of disturbances. After reparametrization, these two subsystems will be associated in terms of probabilities.

More interestingly, the conditional entropies definition, that is,

$$S_q(p_{A/B}) \text{ and } S_q(p_{B/A}),$$

may deserve closer attention, as it can intervene for the definition of estimation precision of a model if the hypothesis of independence between the model variables and its random terms has been accepted or not. If and only if A and B are independent,

$$S_q(p_{A/B}) = S_q(p^A)$$

and

$$S_q(p_{B/A}) = S_q(p^B).$$

Next ,

$$S_q(AB) \equiv S_q(\{p_{ij}^{AB}\}),$$

in general, satisfies:

$$\begin{aligned} S_q(AB) &= S_q(p^A) + S_q(p_{B/A}) + (1-q)S_q(p^A)S_q(p_{B/A}) = S_q(p^B) + S_q(p_{A/B}) + \\ &+ (1-q)S_q(p^B)S_q(p_{A/B}) \end{aligned} \quad (2.34)$$

Finally, to be more explicit than in the previous section, S_q is said to be non-extensive in the sense that given two independent random systems A and B, i.e., $P(A, B) = P(A)P(B)$, then,

$$S_q(AB) = S_q(p^A) + S_q(p^B) + (1-q)S_q(p^A)S_q(p^B) \quad (2.35)$$

In the next, empirical part of this book, for inferential purposes and for optimal simplification of numerical computations, this formula will play a key role in determining the level of entropy of a complex system under the hypothesis of independence of subsystems, i.e., between the model and the random term.

2.4.2.2 Concavity

The concept of concavity is important since, among others things, it allows us to determine whether or not a system is stable. Stability is a meaningful concept in econometrics since it implies stationarity of a process in a given system. Testing for stationarity and cointegration using entropy distribution seems thus to be an open area of further research¹⁴.

¹⁴ However, the job may be rendered difficult since optimal equilibrium responding to economic

S_q is concave (convex) for all probability distributions and all $q > 0$ ($q < 0$) (Gell-Mann & Tsallis, 2004). Let us follow the traditional mathematical definition of concavity and let $\{p_i\}$ and $\{p_i'\}$ ($i = 1, 2, \dots, W$) be two arbitrary probability distributions. The next relation of intermediate distribution follows:

$$p_i'' \equiv \lambda p_i + (1 - \lambda) p_i' \quad (0 < \lambda < 1).$$

By concavity we mean that it can be proven that for all α ,

$$S_q(\{p_i''\}) \geq \lambda S_q(\{p_i\}) + (1 - \lambda) S_q(\{p_i'\}).$$

2.4.3 Tsallis Entropy and Other Forms of Entropy

Let us first review the mathematical main forms of entropies before presenting their most important distinctive properties.

$$S_{BG}(p) \equiv -\sum_i p_i \ln p_i = S_1 = S_1^R = S_1^{LVRA} \quad (2.36)$$

$$S_q(p) \equiv \left(1 - \sum_i p_i^q / (q - 1) \right) \quad (2.37)$$

$$S_q^R(p) \equiv \left(\ln \sum_i p_i^q \right) / (1 - q) = \ln [1 + (1 - q) S_q] / (1 - q) \quad (2.38)$$

$$S_q^{LVRA}(p) \equiv (S_q) / \left(\sum_i p_i^q \right) = \left(1 - \left[\sum_i p_i^q \right]^{-1} \right) / (1 - q) = S_q / [1 + (1 - q) S_q] \quad (2.39)$$

A key element deserves attention here. We see from the first mathematical relation in (2.36) above that Shannon-Gibbs entropy may be generalized, too, by Renyi entropy (2.38) or by the normalized non-extensive form (2.39), independently introduced by Landsberg & Vedral (1998) and by Rajagopal and Abe (2000). Both forms of entropy are monotonically increasing functions of S_q . Tsallis (Gell-Mann & Tsallis, 2004, p. 11) poses and explains a relevant question concerning relationships between these forms of entropy. In fact, after pointing out that monotonicity *makes* S_q , S_q^R , and S_q^N extreme for the same probability distribution, he asks why not base thermodynamics on S_q^R or S_q^N rather than only on Tsallis entropy. The response lies in the

laws does not necessarily fit into optimal entropy equilibrium. This problem will be briefly covered, later.

Table 2: Comparison of different forms of entropy with regard to important properties

Property	Entropy			
	S_{BC}	S_q	S_q^R	S_q^{LVRA}
Extensive ($\forall q)(p_{ij}^{AB} = p_i^A p_j^B)$	yes	no	yes	no
Concave ($\forall q \succ 0$)	Yes	Yes	No	no
Stable ($\forall q \succ 0$)	Yes	Yes	No	no
Optimizing distribution ($\forall q$)	exponential Law ¹⁶	Power Law	Power Law	Power Law

Source: own, based on Tsallis (2009) and Gell-Mann & Tsallis (2004)

disadvantages of these two forms of competitive entropy. In fact, it happens that they are not concave for all positive values of q , but only for $0 < q \leq 1$. Since many physically meaningful phenomena for which q are higher than unity exist, this becomes a serious drawback of both competitive entropies. As far as economic, financial, or social phenomena are concerned, the problem does not allow for any ambiguity since, as we will see in the next section, $1 \leq q < 5/3$. For the majority of them¹⁵, extreme events are on average more frequent (with persistence) than predicted by Gaussian law and not the reverse (i.e., less frequent—with persistence—than predicted by Gaussian law). Tsallis entropy thus remains the one form that not only generalizes SG entropy but also ensures concavity (stability) inside the whole finite interval where probability distribution is defined. The reader should thus far understand why non-extensive Tsallis entropy has been recently used to generalize all other forms of entropy, at least in many fields where entropy is applied.

2.4.3.1 Characterization

In the following table, we illustrate different links between the commonly used forms of entropy with respect to the characterization in Table 2.

“yes” and “no” correspond, respectively, to what, according to recent thermodynamic literature (Gell-Mann & Tsallis, 2004), are thermodynamically allowed and forbidden violations of the Boltzmann–Gibbs (BG) entropy properties.

¹⁵ For example, for stock market returns, q is around 1.4, far enough from the unity which characterises Gaussian distribution.

¹⁶ “Exp” for exponential and “P” for power.

NB: R stands for Renyi, and $S_q^{LVRA} \equiv S_q^N$ (LVRA and N stand for Landsberg-Vedral-Rajagopal-Abe and normalized, respectively (Gell-Mann & Tsallis, 2004).

2.4.3.2 Scale of q-Tsallis Index and its Interpretation

Following the thermodynamic literature built on Lévy-like anomalous diffusion, it has been shown that $p(x) \propto e_q^{-\beta x^2}$ optimizes

$$S_q = \frac{1 - \int dx [p(x)]^q}{q-1}$$

under appropriate constraints. If one convolutes n times $p(x)$ ($n \rightarrow \infty$), we approach a Gaussian distribution if $q < 5/3$, and a Lévy $L_{\gamma L}(x)$ if $5/3 < q < 3$. The index γL of Lévy distribution is related to q as follows:

$$q = \frac{\gamma L + 3}{\gamma L + 1} \quad (5/3 < q < 3).$$

Thus, in empirical applications, the value of q should vary inside an interval from unity to $5/3$, which corresponds to cases of finite variance for phenomena dwelling within the Gaussian basin of attraction.

2.5 Kullback-Leibler-Tsallis Cross-Entropy

2.5.1 The q-Generalization of the Kullback-Leibler Relative Entropy

Kullback-Leiber-Tsallis cross-entropy is known in literature as the q -generalization of Kullback-Leibler relative entropy. The Kullback-Leiber-Shannon entropy introduced in Part II can be q -generalized (Tsallis, 2009) in a straightforward manner. The discrete version becomes:

$$I_q(p, p_o) \equiv \sum p_i \frac{[p_i / p_{oi}]^{q-1} - 1}{q-1} \quad (2.40)^{17}$$

since with any real $r > 0$, one has the following properties:

$$\frac{r^{q-1} - 1}{q-1} \geq 1 - \frac{1}{r} \quad \text{if } q > 0 \quad (2.41)$$

¹⁷ In a continuous case, we have:

$$I_q(p, p^{(0)}) \equiv - \int dx p(x) \ln_q \left[\frac{p^{(0)}(x)}{p(x)} \right] = \int dx p(x) \frac{[p(x) / p^{(0)}(x)]^{q-1} - 1}{q-1}$$

$$= 1 - \frac{1}{r} \quad \text{if } q = 0$$

$$\leq 1 - \frac{1}{r} \quad \text{if } q < 0$$

Thus, retaining the practical case of , we can write:

$$\frac{[p(x)/p_o(x)]^{q-1} - 1}{q-1} \geq 1 - \frac{p_o(x)}{p(x)}.$$

Hence¹⁸,

$$\sum p(x) \frac{\left[\frac{p(x)}{p_o(x)} \right]^{q-1} - 1}{q-1} \geq \sum p(x) \left[1 - \frac{p_o(x)}{p(x)} \right] = 1 - 1.$$

Therefore, coming back again to the generalized K-Ld cross-entropy, we have¹⁹:

$$\begin{aligned} I_q(p, p_o) &\geq 0 \text{ if } q > 0, \\ &= 0 \text{ if } q = 0, \\ &\leq 0 \text{ if } q < 0. \end{aligned} \tag{2.42}$$

Thus, as Tsallis (2009) has made us aware, the above q-Kullback-Leibler index has the same basic property as the standard Kullback-Leibler entropy and can be used for the same purpose while having the additional advantage of an adaptive q according to the system with which we are dealing.

There exist two different versions of the Kullback-Leibler divergence (K-Ld) in Tsallis statistics, the usual generalized K-Ld shown above and the generalized Bregman K-Ld. According to Venkatesan et al., (Plastino & Venkatesan, 2011), problems have been encountered in empirical thermodynamics trying to reconcile these two versions. Unfortunately—or fortunately!—the same problems seem to reappear while applying this theory in social science since every version of generalized K-Ld leads to different outputs. Let us try to synthesize what recent literature says about this problem.

¹⁸ It is straightforward to derive this property in the case of the continuous case.

¹⁹ The same conclusion is obtained by using Jensen's inequality (e.g., Gell-Mann & Tsallis, 2004).

2.5.2 Tsallis Versions of the Kullback-Leibler Divergence in Constraining Problems

This short section represents the final bridge between theory and the applications in the last parts of this work. In a recent study, Plastino & Venkatesan (2011) lay out interesting aspects of empirical research when q -generalized K-Ld cross-entropy is associated with constraining information. Since, in the social sciences, we particularly need discrete forms of these relative entropies, let us first rewrite these forms before commenting on their conditions of applicability:

$$I_q(p, p_o) \equiv \sum_i p_i \frac{[p_i / p_{oi}]^{q-1} - 1}{q-1} \quad (2.43)$$

$$I_q[p \| p_o] = \frac{1}{q-1} \sum_i p_i [(p_i)^{q-1} - (p_{oi})^{q-1}] - \sum_i (p_i - p_{oi})(p_{oi})^{q-1} \quad (2.44)$$

The form (2.43) is the one derived directly from Kullback-Leibler formalism and presented in (2.40). The second form is referred to as the generalized Bregman form of K-Ld cross-entropy, and it is more appealing than (2.43) from an information-geometric viewpoint (Plastino & Venkatesan, 2011) even if it does contain certain inherent drawbacks.

A study by Abe and Bagci (2005) has demonstrated that the generalized K-Ld defined by (2.44) is jointly convex in terms of both p_i and p_{oi} while the form defined by (2.43) is convex only in terms of p_i . A further distinction between the two forms of the generalized K-Ld concerns the property of composability. While the form defined by (2.44) is composable, the form defined by (2.43) does not exhibit this property.

The second interesting aspect for practitioners concerns the manner in which mean values are computed. Non-extensive statistics has employed a number of forms in which expectations may be defined. The first among these are the linear constraints initially used by Tsallis (2009), also known as normal averages, that is:

$$\langle y \rangle = \sum_i p_i y_i$$

The second is the Curado-Tsallis (C-T) constraints of the form:

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

and the normalized Tsallis-Mendes-Plastino (TMP) constraints (also known as q -averages or an escort distribution) of the form:

$$\langle y_q \rangle = \sum_i \frac{p_i^q}{\sum_i p_i^q} y_i$$

A fourth—less applied by practitioners—constraining procedure is the optimal Lagrange multiplier approach.

Among these four methods to describe expectations, the most commonly employed by Tsallis practitioners is TMP, referred to as escort distribution.

Recent work by Abe (2009) suggest that, in generalized statistics, expectations defined in terms of normal averages, in contrast to those defined by q-averages, seem to display higher consistency in material chaos hypotheses. Recent reformulation of the variational perturbation approximations in non-extensive statistical physics followed from these findings. To my knowledge, application in the social sciences to assess the universality of this finding has not been done yet.

Finally, there is the issue of consistency. This stems from the form of the generalized K-Ld defined by (2.43) being consistent with expectations and constraints defined by q-averages (“prominently” the TMP) while, on the other hand, the generalized Bregman K-Ld defined by (2.44) is consistent with expectations defined by normal averages.

Thus, through reformulations of an empirical inverse problem, this last point may play a key role since non-appropriated constraints should lead to a non-optimal solution in the best case or to computational problems, as is often the case.

2.6 A Generalized Linear Non-Extensive Entropy Econometric Model

2.6.1 A General Model

This section presents a generalized linear non-extensive entropy econometric approach to estimate econometric model. Following Golan et al., (1996), we first reparametrize the generalized linear model of the equation (2.12') rewritten below:

$$y_i = \sum_{k=1}^K B_k X_k + \zeta_i \quad (2.12')$$

With, once again, the random term $\zeta_i \in e$ and $i = (1, \dots, I)$ (being the number of observations); K is the number of model parameters to be estimated.

Where B values are not necessarily constrained between 0 and 1, and ζ is an unobservable disturbance term with finite variance, owing to the nature of economic data that exhibits error observation from empirical measurement or random shocks. If we treat each B_j ($k = 1 \dots K$) as a discrete random variable with compact support and $2 < M < \infty$ possible outcomes, we can express B as:

$$B_k = \sum_{m=1}^M p_{km} v_{km}, \forall k \in \{1, \dots, K\} \quad (2.45)$$

where p_{km} is the probability of the outcome v_{km} . The probabilities must be non-negative and add up to one. Similarly, by treating each element ζ_i of ζ as a finite and dis-

crete random variable with compact support and $2 < M < \infty$ possible outcomes centred around zero, we can express ζ_i as:

$$\zeta_i = \sum_{j=1}^J r_{ij} w_{ij} \quad (2.46)$$

where r_i is the probability of outcome w_i on the support space j , with $j \in \{1, \dots, J\}$ and $i \in \{1, \dots, N\}$. Note that the term e (an estimator of ζ) can be fixed as a percentage of the explained variable, as an a priori Bayesian hypothesis. Posterior probabilities within the support space may display non-Gaussian distribution. The element v_{km} constitutes a priori information provided by the researcher while p_{km} is an unknown probability whose value must be determined by solving a maximum entropy problem. In matrix notation, let us rewrite $\beta = V \cdot P$ with $p_{km} \geq 0$ and

$$\sum_{k=1}^K \sum_{m=2 \dots M} p_{km} = 1,$$

where again, K is the number of parameters to be estimated and M the number of data points in the support space. Also, let $e = r \cdot w$, with $r_{ij} \geq 0$ and

$$\sum_{i=1}^N \sum_{j=2 \dots J} r_{ij} = 1$$

for N the number of observations and J the number of data points on the support space for the error term. Then, the maximum Tsallis Entropy Econometric (MTEE) estimator can be stated as:

$$\max[H_q(p; r)] = \left[1 - \sum_k \sum_m \alpha \cdot (p_{km})^q \right] + \left[1 - \sum_i \sum_j (1 - \alpha) \cdot (r_{ij})^q \right] \cdot (q - 1)^{-1} \quad (2.47)$$

subject to

$$y_i = \sum_{k=1}^K B_k X_k + e_i = X \cdot \sum_{m=1}^M v_m \left(\frac{p_m^q}{\sum_{m=1}^M p_m^q} \right) + \sum_{j=1}^J w_j \left(\frac{r_j^q}{\sum_{j=1}^J r_j^q} \right) \quad (2.48)$$

$$\sum_{k=1}^K \sum_{m=2}^M p_{km} = 1 \quad (2.49)$$

$$\sum_{i=1}^N \sum_{j=2}^J r_{ij} = 1 \quad (2.50)$$

where the real q , as previously stated, stands for the Tsallis parameter.

Above, $H_q(p, r)$ weighted by α dual criterion function is nonlinear and measures the entropy in the model. The estimates of the parameters and residual are sensitive

to the length and position of support intervals of β parameters. When parameters of the proposed mode²⁰ concern elasticity or error correct coefficients, the values of which lie between 0 and 1, then the support space should be defined inside the interval zero and one. In other cases, the support space may be defined between minus and plus infinity, according to the intuitive evaluation of the modeller. Additionally, within the same interval support, the model estimates and their variances should be affected by the number of support values (Golan et al., 1996). Increasing the number of point values inside the support space leads to improving the a priori information about the system. A few years of modelling with the maximum entropy approach seem to show that a well-defined support space is crucial to obtaining better results. The weights α and $(1 - \alpha)$ are introduced into the above dual objective function. The first term “of precision” accounts for deviations of the estimated parameters from the prior (defined under support space). The second, “prediction ex post,” accounts for an empirical error term as a difference between predicted and observed data values of the model.

2.6.2 Parameter Confidence Interval Area

In this section, we will propose the normalized Tsallis entropy coefficient $S(\hat{a}_k)$ as an equivalent to a standard error measure in the case of classical econometrics. An equivalent of the determination coefficient R^2 will be introduced, also under the entropy symbol $S(\hat{\text{Pr}})$. The departure point is that the maximum level of entropy-uncertainty is reached when significant information-moment constraints are not enforced. This leads to a uniform distribution of probabilities over the k states of the system. As we add each piece of informative data in the form of a constraint, a departure from the uniform distribution will result, which means a lowering in uncertainty. Thus, the value of the proposed $S(\hat{\text{Pr}})$ below reflects a global departure from the maximum uncertainty for the whole model. Without giving superfluous theoretical details, we follow formulations in, e.g., Bwanakare (2014) and propose a normalized non-extensive entropy measure of $S(\hat{a}_k)$ and $S(\hat{\text{Pr}})$.

From the Tsallis entropy definition, S_q vanishes (for all q) in the case of $M = 1$; for $M > 1$, $q > 0$, whenever one of the $p_i (i = 1..M)$ occurrences equals unity, the remaining probabilities, of course, vanish. We get a global, absolute maximum of S_q (for all q) in the case of a uniform distribution, i.e., when all $p_i = 1/M$. This vanishes (for all q) in

20 As already presented, the expression $P_m = \frac{p_m^q}{\sum_{m=1}^M p_m^q}$ is referred to as *escort probabilities*, and we

have for $q=1$ (then P_m is normalized to unity), that is, in the case of Gaussian distribution (Gell-Mann i Tsallis, 2004), (Tsallis, 2009).

the case $M = 1$; and for $M > 1$, $q > 0$, whenever one of the p_i equals unity, the remaining ones vanish. A global, absolute maximum of S_q (for all q) is obtained in the case of equiprobability, i.e., when all $p_i = 1/M$. Note that we are interested, for our economic analysis, in q values lying inside the interval $(1, 5/3)$. In such an instance, we have for our two systems:

$$S_q(p) = (M^{1-q} - 1) \cdot (1 - q)^{-1} \quad (2.51)$$

and

$$S_q(r) = (N^{1-q} - 1) \cdot (1 - q)^{-1} \quad (2.52)$$

in the limit when $q = 1$, relation (2.51) or (2.52) leads to the Boltzmann-Shannon expression (Gell-Mann & Tsallis, 2004).

Below, a normalized entropy index is suggested, one in which the numerator stands for the calculated entropy of the system while the denominator displays the highest maximum entropy of the system owing to the equiprobability property:

$$S(\hat{a}_k) = - \left[1 - \sum_k \sum_m (p_{km})^q \right] / \left[k \cdot (M^{1-q} - 1) \right] \quad (2.53)$$

with k varying from 1 to K (number of parameters of the system) and m belonging to M (number of support space points), with $M > 2$. $S(\hat{a}_k)$ then reporting the accuracy on estimated parameters. Equation (2.48) reflects the non-additivity property of Tsallis entropy for two (probably) independent systems; the first, parameter probability distribution, and the second, error disturbance probability distribution (plausibly with quasi-Gaussian properties):

$$S(\hat{Pr}) = [S(\hat{p} + \hat{r})] = \{[S(\hat{p}) + S(\hat{r})] + (1 - q) \cdot S(\hat{p}) \cdot S(\hat{r})\} \quad (2.54)$$

where:

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

and

$$S(\hat{r}) = - \left[\left(1 - \sum_n \sum_j r^q \right) \right] / \left[k \cdot N \cdot (J^{1-q} - 1) \right]$$

$S(\hat{Pr})$ is then the sum of normalized entropy related to parameters of the model $S(\hat{p})$ and to the disturbance term $S(\hat{r})$. Likewise, the latter value $S(\hat{r})$ is derived for all observations n , with J the number of data points on the support space of estimated probabilities r related to the error term.

The values of these normalized entropy indexes $S(\hat{a}_k)$, $S(\hat{\text{Pr}})$ vary between zero and one. Their values, near to one, indicate a poor informative variable while lower values are an indication of better informative parameter estimate \hat{a}_k about the model.

The next part of the book will present in detail national accounts tables used for building or forecasting macroeconomic models. The statistical theory will be implemented particularly in the case of the inverse problem, while keeping in line with this work's objective.

2.7 An Application Example: a Maximum Tsallis Entropy Econometrics Model for Labour Demand

This example presents, through Monte Carlo simulations, a model for labour demand adjustment for the Polish private sector. It constitutes an extension of an initial model presented by Bwanakare (2010) for the labour demand adjustment by the private sector of Subcarpathia province in Poland. The model aims at displaying short-run and long-run relationships between labour demand determinants through an error self-correct process. Due to the relatively short period of the sample (fourteen annual data points) and the autoregressive nature of the model, we may have to deal with limited possibilities of statistical inference in the absence of convergence properties or, in the worst case, an inverse ill-behaved problem. Thus, traditional methods of parameter estimation may fail to be effective. We then propose to apply the generalized maximum Tsallis entropy econometric approach—as an extension of Jaynes-Shannon-Gibbs Information theoretic entropy formalism, already applied in econometrics (Golan, Judge & Miller, 1996). Due to an annual data frequency of the sample, the approach proves to be applicable in the case of classical econometrics when a small, lower frequency data sample is available. Such a small data sample should display tail queue Gaussian distribution. Through this application, Monte Carlo experiment outputs seem to confirm the reliability of the Tsallis entropy econometrics approach, which in this particular case performs as well as the generalized least square technique.

2.7.1 Theoretical Expectation Model

In the short run, managers decide on the number of employees to be hired (or dismissed) in accordance with the expected long-run optimal level of production. However, because of institutional or economic reasons, that optimal number is not hired (or fired) at once. First, uncertainty remains a predominant characteristic of business. For this reason, employers naturally prefer a moderate and progressive adjustment of recruited workers to the targeted optimal level. Recruitment in some economic sectors could be time-consuming as well, especially when searching for

good specialists. Second, relatively well organized trade unions could prevent employers from abrupt, large-scale layoffs, or the cost of dismissing a worker may become high, depending on prevailing labour laws at a given period. In both cases, the process of shock correction will be more or less long, depending on its origin and magnitude.

Under classical assumptions of constant returns to scale, *ex ante* and *ex post* complementarities of factors, and long-run constant rate of labour productivity, the desired level of labour demand L_t^* is a function of the output Y_t and the technical progress t^{21} :

$$L_t^* = \alpha \cdot \exp(-\beta \cdot t) \cdot Y_t \quad (2.55)$$

Assuming that labour demand adjusts to its targeted level by an error correction model:

$$\log(L_t / L_{t-1}) = \lambda \cdot \log(L_t^* / L_{t-1}^*) + \mu \cdot \log(L_{t-1}^* / L_{t-1}), \quad (2.56)$$

combining (2.55) and (2.56) leads to:

$$\log(L_t / L_{t-1}) = \lambda \cdot \log(Y_t / Y_{t-1}) + \mu \cdot \log(Y_{t-1} / L_{t-1}) + \mu \cdot \beta \cdot t + \alpha_0 \quad (2.57)$$

The parameter λ is the impact of output on labour demand, and then a short-run elasticity of labour demand with respect to output Y_t , μ being the error correction parameter. Since a relation $-1 \leq \mu \leq 0$ should prevail, the equilibrium error is only partly adjusted at each period. In other words, this parameter synthesizes employers' determinants of labour demand adjustment once a shock in sales for the coming period is expected.

2.7.2 A Generalized Non-Extensive Entropy Econometric Model

2.7.2.1 General Model

Presently we are interested in the estimation of parameters of a Podkarpacki labour demand model, applying a generalized non-extensive entropy econometric approach. Following Golan, Judge & Miller (1996) and Bwanakare (2014a, 2014b), we reparametrize, in the first step, the generalized linear model before fitting it to equation (2.48). This step allows for including in moment equations-restrictions the same probability variables as those optimized in the criterion function.

To reparametrize the model, we follow each equation in (2.45–2.46) where each β_k ($k = 1, \dots, K$) is treated as a discrete, random variable with compact support and $2 < M < \infty$ possible outcomes. Next, for the estimation of the model, we maximize the entropy criterion function in (2.47) under moment and normality condition restric-

21 This is a simplification stipulating that technical progress is a linear function of time.

tions presented in (2.48–2.50). For confidence area analysis, we need to apply Equations (2.53–2.54).

With the purpose of improving estimated parameter quality, one can add additional *a priori* restrictions to (2.48–2.50) as follows:

$$e = Y - \hat{Y} = \hat{Y} - XV\hat{p} = 0. \quad (2.58)$$

Then we constrain the error term e to sum up to zero²² which provides an additional quality of requiring an unbiased parameter estimator.

Efficiency property mainly depends upon informative quality of the prior. When it is poor, the values of the estimated \hat{p}_i from the model tends to be equal for all p_i , i.e., the case of a uniform distribution.

According to economic theory, we constrain elasticity parameters within a point support space of *zero* and *one*. As known (e.g., Golan, Judge & Miller 1996), sharper support area points of a parameter act as increasing quality of the “*a priori*” information. Furthermore, this allows computations of this nonlinear model to promptly converge to an optimum solution. This is explained as follows:

$$0 \leq \lambda = V\hat{p} \leq 1 \quad (2.59)$$

Likewise, we may add additional economic restrictions to the model (2.57) parameters; this leads to the following formulations:

$$-1 \leq \mu = V\hat{p} \leq 0 \quad (2.60)$$

$$-\infty \leq \beta = V\hat{p} \leq 0 \quad (2.61)$$

2.7.3 Estimated Confidence Area Of Parameters

In classical econometrics, we usually combine the variance of random model error with the co-linearity level of explicative variables to determine the standard error of estimated parameters and to infer their confidence area while assuming a normal distribution law of random errors. This is particularly true in the case of the Least Squares approach for a linear model.

In entropy econometrics, the approach is very different. We use the normalized entropy $S(\hat{a}_{ij})$ (equation 2.53) as an equivalent of the estimate standard error measure in classical linear model econometrics. Likewise, the equivalent of the coefficient of determination R^2 is a $S(\hat{Pr})$ (equation 2.54). Following Golan et al. (1996a, 1996b, 1996c, 2002) and Soofi (1992, 1994), in the case of maximum entropy formulation, the

²² Note that our model has a constant term, suggesting that the economic initial condition may impact the optimal solution.

maximum level of entropy-uncertainty results when the information-moment constraints (*a priori* information) are not enforced. Furthermore, this leads to uniform distribution of probabilities over the *k* states of the system. As we add each piece of informative data in the form of a constraint, a departure results from the uniform distribution, which explains an uncertainty reduction. Thus, the value of $S(\hat{P})$ reflects a global departure from the maximum uncertainty for the whole model. A similar measure, $1 - S(\hat{P})$, called *the information index*, explains the level of informative content of the model. For theoretical details, we refer you to the formulations presented above in Equations (2.51–2.54) or, e.g., in Golan et al., (1996).

2.7.4 Data and Model Outputs

In this section, the output parameter of Tsallis entropy, Shannon entropy, and least squares econometric models are presented. Next, the obtained results will be compared to those from a Monte Carlo simulation using the same data.

Data used in the model (equation 2.57) come from the Polish Office of Statistics (GUS) and concern the period 1997–2010. Parameters of the model have been computed with the GAMS (General algebraic modelling system) code with the incorporated solver *PATHNLP*. We have noticed, through different simulations, that the Shannon-Gibbs entropy model seems more sensitive to initial conditions (support space of parameters in particular) than Tsallis entropy. This is a useful property, particularly when an economic theory does not exist to prompt us as to the starting parameters with which to begin. Parameter estimation by robust standard errors least squares (LS) approach has been carried out, using freeware *Gretl* software (<http://gretl.sourceforge.net/>). Thus, the HAC estimator is used for heteroscedasticity and autocorrelation correction.

a) Parameter outputs of Tsallis entropy model:

Dependent variable: $\log\left(\frac{L_t}{L_{t-1}}\right)$

Exogenous variables	$\log\left(\frac{y_t}{y_{t-1}}\right)$	$\log\left(\frac{Y_{t-1}}{L_{t-1}}\right)$	T	a0
Estimates \hat{a}_j	0.710	0.010	-0.020	-0.266
Precision error $S(\hat{a}_j)$ on estimated parameters	0.135	0.250	0.250	0.236
Information Index $I[S(\hat{Pr})] = 1 - S(\hat{Pr}) = 0.852$				
Tsallis -q parameter (for a weight $\alpha_i=15\%$) = 2.091				

Throughout many conducted experiments, we have observed the coefficient $S(\hat{Pr})$ to be very sensitive to weighting parameters α in the objective function. Tsallis-q value being itself influenced by the above weights, its values closer or higher to 5/3 correspond to meaningless information index coefficients for which $S(\hat{Pr})$ vanishes to zero. In empirical research, the Tsallis-q coefficient may take much higher values as a consequence of model linearity attributes or in the case when the sample is small. In the present case, we have noticed a high sensitivity of this Tsallis-q parameter on the change of the weight α_i in the criterion function. The higher the weight α_i , the higher the value of the Tsallis-q parameter. We have retained the value of this weight for which $I[S(\hat{Pr})]$ is the highest.

b) Parameter outputs of Shannon-Gibbs entropy model:

Dependent variable: $\log\left(\frac{L_t}{L_{t-1}}\right)$

Exogenous variables	$\log\left(\frac{y_t}{y_{t-1}}\right)$	$\log\left(\frac{Y_{t-1}}{L_{t-1}}\right)$	T	a0
Estimates \hat{a}_j	0.709	0.010	-0.020	-0.263
Precision error $S(\hat{a}_k)$ on estimated parameters	0.297	0.518	0.518	0.421
Information Index $I[S(\hat{Pr})] = 1 - S(\hat{Pr}) = 0.829$				

c) Robust standard errors LS estimation:

Dependent variable: $\log\left(\frac{L_t}{L_{t-1}}\right)$

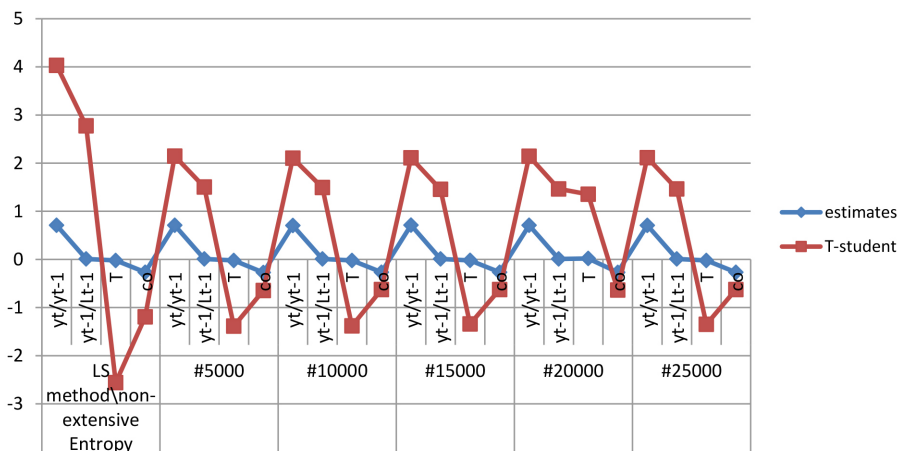
Exogenous variables	$\log\left(\frac{y_t}{y_{t-1}}\right)$	$\log\left(\frac{Y_{t-1}}{L_{t-1}}\right)$	T	a0
Estimates \hat{a}_j	0.709963	0.010417	-0.02031	-0.26578
P-values	2.73e-011***	4.34e-06	2.89e-06***	0.0302**
Corrected R2 = 0.79				
DW = 1.4832				

Three parameters are different from zero at 1%, and one on the variable a_0 significant at 10%. The above precision on the estimated parameters from such a small data sample of an autoregressive model suggests the presence of co-integrating—at the same order—variables L_t and Y_t . Such a particular situation leads to super-consistency of estimated parameters.

Table 3: Monte Carlo simulation outputs

Poland Labor demand Model and simulation outputs							
	NEE/GLS	#5000	#10000	#15000	#20000	#25000	
Estimates	y_t/y_{t-1}	0.709963106	0.7092748	0.70445994	0.7126441	0.709756	0.708636
	Y_{t-1}/L_{t-1}	0.010416764	0.0105625	0.01052379	0.0103705	0.010461	0.010458
	T	-0.020312463	-0.020624	-0.0205689	-0.020213	0.020406	-0.02038
	C_0	-0.26577817	-0.272354	-0.2652226	-0.266263	-0.26784	-0.26663
	y_t/y_{t-1}	0.176056192	0.3303391	0.33502592	0.3377129	0.331089	0.335143
	Y_{t-1}/L_{t-1}	0.003753087	0.0070234	0.00704595	0.0071221	0.007143	0.007144
Standard errors	T	0.007938838	0.0148845	0.01489532	0.0150799	0.015074	0.015107
	C_0	0.222643946	0.4216688	0.42344287	0.4260381	0.420446	0.426712

Source: own elaboration.

**Figure 1:** Monte Carlo model estimates and T-student from simulations: initial model, #5000, #10000, #15000, #20000, #25000.

For comparative purposes, the table 3 presents outputs from Monte Carlo experiments (computed with *Mathlab* 7.3.0 software).

The above outputs have been derived under the hypothesis of random normal law. The empirical standard error initially computed for random value generation is 0.02035. This constitutes 50% of the observed endogenous variable standard error. We observe that Shannon-Tsallis entropy and the least squares outputs are similar and almost reflect Monte Carlo convergence outputs. The initial t-student related to the parameters on the variables y_t/y_{t-1} and y_{t-1}/L_{t-1} decrease when we carry out the #5000 simulation and remain practically unchanged up at the #25000 simulation experiment. Nevertheless, we observe that parameter estimates of the model remain unchanged irrespective of the number of the simulations.

To conclude, we note the accuracy in the similarity of outputs from the three models. This suggests that we are dealing with a convergent case of power law to Gaussian distribution. If the Tsallis-q parameter is too high, it cannot be interpreted in a model where its nonlinearity and the small sample size (in this case 14 observation years) should have a significant impact on the value of that parameter (Grech & Pamula, 2013). The impact parameter is around 0.71. This is, on average, a 0.71% growth of labour demand when gross profits shifts up to 1%. As it has been indicated, these outputs are related to a period (1997–2010) during which Poland was undergoing structural, post-communism reforms. As such, their interpretation should be done carefully. As far as exogenous technical progress is concerned, we observe a negative sign on the value of the estimated parameter on the symptomatic variable, which indicates an expected adverse impact of technical progress on labour demand.

Annex A

The solution for the above constrained equation is obtained by forming the Lagrange function:

$$L = -\sum_{k=1}^K p_k \ln p_k + \sum_{t=1}^T \lambda_t \left[y_t - \sum_{k=1}^K p_k f_t(x_k) \right] + \mu \left(1 - \sum_{k=1}^K p_k \right) \quad (2.20)$$

After defining the first order conditions, the solution to this $(K + T + 1)$ equations and parameters is:

$$\hat{p}_k = \exp \left(-\sum_{t=1}^T \hat{\lambda}_t f_t(x_k) - 1 - \hat{\mu} \right), k = 1, 2, \dots, K,$$

$$\sum_{k=1}^K \exp \left(-\sum_{t=1}^T \hat{\lambda}_t f_t(x_k) - 1 - \hat{\mu} \right) f_t(x_k) = y_t, t = 1, 2, \dots, T,$$

$$\sum_{k=1}^K \exp\left(-\sum_{t=1}^T \hat{\lambda}_t f_t(x_k) - 1 - \hat{\mu}\right) = 1,$$

and, as a formal solution, we obtain:

$$\hat{p}_k = \frac{1}{\Omega(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_T)} \exp\left(-\sum_{t=1}^T \hat{\lambda}_t f_t(x_k)\right),$$

where

$$\Omega(\hat{\lambda}) = \sum_{k=1}^K \exp\left(-\sum_{t=1}^T \hat{\lambda}_t f_t(x_k)\right)$$

is a normalization factor.

It is easy to prove the uniqueness of the primal ME solution. In fact, given the first order conditions, elements of the Hessian matrix are as follows:

$$\frac{\partial^2 L}{\partial p_k^2} = \frac{-1' \exp(f(x_k)' \lambda)}{\exp(f(x_k)' \lambda)} = -\frac{1}{p_k} \text{ for the diagonal elements,}$$

and

$$\frac{\partial^2 L}{\partial p_k \partial p_j} = 0 \text{ for the off-diagonal elements.}$$

Thus, the Hessian is a negatively defined and sufficient condition for a unique global maximum is fulfilled. Furthermore,

$$y_t = \left(\frac{\partial}{\partial \lambda_t} \right) \ln \Omega, 1 \leq t \leq T$$

One may observe that the value of the entropy H is a function of the given data:

$$H = \ln \Omega(\hat{\lambda}) + \sum_t \hat{\lambda}_t y_t \quad (2.21),$$

Annex B: Independence of Events Within q-Generalized Kullback-Leibler Relative Entropy

Let us consider (Tsallis, 2009) the problem of independence of random variables in the case of two-dimensional random variable (x, y) , and its corresponding distribution function $p(x, y)$ with $\int dx dy p(x, y) = 1$

As expected, the marginal distribution functions are then given by

$$h_1(x) \equiv \int dy p(x, y)$$

and

$$h_2(y) \equiv \int dx p(x, y).$$

The discrimination criterion for independence concerns the comparison of $p(x, y)$ with $p_o(x, y) \equiv h_1(x)h_2(y)$. Once again, the one-dimensional random variables x and y are independent if and only if $p(x, y) = p_o(x, y)$. Therefore, the criterion becomes:

$$\int dx dy \frac{\left[\frac{p(x, y)}{h_1(x)h_2(y)} \right]^{q-1} - 1}{q-1} \geq 0, \text{ for } q > 0.$$

When $q \rightarrow 1$, we then recover the usual discrimination criterion, i.e. :

$$\int dx dy p(x, y) \ln p(x, y) - \int dx h_1(x) \ln h_1(x) - \int dy h_2(y) \ln h_2(y) \geq 0.$$

An interesting case is if $q \rightarrow 2$, then we have:

$$\int dx dy \left[\frac{p(x, y)}{h_1(x)h_2(y)} \right]^2 \geq 1,$$

The value of this quantity, useful in economics, may give a sign of independence between x and y , when it vanishes.

Finally, if we generalize to the case of many variables, the Kullback-Leibler-Tsallis index of information becomes:

$$I_q(p(x_1, x_2, \dots, x_d), p_o(x_1, x_2, \dots, x_d)) \geq 0 \text{ (for } q \geq 0)$$

or its symmetrized version:

$$\frac{1}{2} [I_q(p(x_1, x_2, \dots, x_d), p_o(x_1, x_2, \dots, x_d)) + I_q(p_o(x_1, x_2, \dots, x_d), p(x_1, x_2, \dots, x_d))] \geq 0 \text{ (} q \geq 0)$$

When equality holds for these two above relations, this means that all elements x_1, x_2, \dots, x_d are independent among them (almost everywhere).

Bibliography – Part II

- Abe, S. (2009). Generalized molecular chaos hypothesis and the H theorem: Problem of constraints and amendment of nonextensive statistical mechanics. *Phys. Rev. E* 79 (4), 041116.
- Abe, S., & Bagci, B.G. (2005). Necessity of q-expectation value in nonextensive statistical mechanics. *Phys. Rev. E* 71(1), 016139.
- Braaksma, B., & Zeelenberg, K. (2015). “Re-make/Re-model”: Should big data change the modelling paradigm in official statistics? *Statistical Journal of the IAOS* 3:193–202.

- Bwanakare, S. (2010). Non-Extensive Entropy Econometric Model (NEE): the Case of Labour Demand in the Podkarpackie Province. *Acta Physica Polonica A*, 117(4):647–651.
- Bwanakare, S. (2014). *Econometric Balancing of a Social Accounting Matrix Under a Power-law Hypothesis*. *Przegląd Statystyczny* 61(3):263–281
- Bwanakare, S. (2014). Non-Extensive Entropy Econometrics: New Statistical Features of Constant Elasticity of Substitution-Related Models. *Entropy* 16(5):2713–2728.
- Bwanakare, S. (2015). Greenhouse Emission Forecast as an Inverse Stochastic Problem: A Cross-Entropy Econometrics Approach. *Acta Physica Polonica A* 127(3A):13–20.
- Carter, T. (2011). *An introduction to information theory and entropy*. Complex Systems Summer School. Retrieved from: <http://astarte.csustan.edu/~tom/SFI-CSSS>
- Drożdż, S. & Kwapień, J. (2012). Physical approach to complex systems. *Physics Reports* 515(3–4):115–226.
- Engle, R.F., & Granger, C.W.J. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55(2):251–276.
- Gell-Mann, M., & Tsallis, C. (Eds.). (2004). *Nonextensive Entropy, Interdisciplinary Applications*. USA: Oxford University Press.
- Golan, A., Karp, S.L., & Perloff, J.M. (1996a). *Estimating a Mixed Strategy Employing Maximum Entropy*. eScholarship, California. Retrieved from <http://escholarship.org/uc/item/5rx8z14p>
- Golan, A., Judge, G.G., & Miller, D. (1996b). *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. Chichester, England: Wiley.
- Golan, A., Judge, G.G., & Perloff, M.J. (1996c). Estimating the Size Distribution of Firms Using Government Summary Statistics. *The Journal of Industrial Economics*, 3(1):69–80.
- Golan, A., & Perloff, J.M. (2002). Comparison of Maximum Entropy and Higher-Order Entropy Estimators. *Journal of Econometrics*, 107(1):195–211.
- Good, I.J. (1963). Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables. *The Annals of Mathematical Statistics* 34(3):911–934.
- Grech, D., & Pamula, G. (2013). On the multifractal effects generated by monofractal signals. *Physica A*, 392(23):5845–5864.
- Green, W.H. (2003). *Econometric Analysis* (5th Ed.). NY, USA: Prentice Hall.
- Halmos, P.R., & Savage, L.J. (1949). Application of the Radon-Nikodym Theorem to the Theory of Sufficient Statistics. *The Annals of Mathematical Statistics*, 20(2):225–241.
- Hartley, R.V.L. (1928). Transmission of Information. *Bell Labs Technical Journal*, 7(3):535–563.
- Jaynes, E.T. (1994). *Probability Theory: The Logic Of Science*. Washington DC, USA: Washington University.
- Jeffreys, H. (1946). An Invariant Form for the Prior Probability in Estimation Problems. *Proceedings of the Royal Society A*, 186(1007):453–461. JSTOR 97883.
- Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *J. Basic Eng.*, 82(D):35–45.
- Khinchin, I. (1957). *Mathematical Foundations of Information Theory*. NY: Dover.
- Kullback, S. (1959). *Information Theory and Statistics*. NY: John Wiley and Sons.
- Landsberg, T.P., & Vedral, V. (1998). Distributions and channel capacities in generalized statistical mechanics. *Physics Letters A*, 247(3):211–217.
- Loeve, M. (1955). *Probability Theory: Foundations, Random Sequences*. NY: Van Nostrand.
- Papoulis, A. (1991). *Probability, Random Variables and Stochastic Processes* (3rd Ed.). N.Y., USA: McGraw-Hill.
- Plastino, A., & Venkatesan, R.C. (2011). Deformed Statistics Kullback-Leibler Divergence Minimization within a Scaled Bregman Framework. *Physics Letters A*, 375(48):4237–4243.
- Pukelsheim, F. (1994). The Three Sigma Rule. *The American Statistician*, 48:88–91.
- Rajagopal, A.K., & Abe, S. (2000). Justification of Power Law Canonical Distributions Based on the Generalized Central Limit Theorem. *Europhysics Letters*, 52(610).

- Rényi, A. (1961). On measures of information and entropy. *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob.*, Vol. 1:547–561. (Univ. of Calif. Press, 1961)
- Savage, L.J. (1954). *The Foundations of Statistics*. NY, USA: Wiley.
- Shannon, C.E. (1948). *A Mathematical Theory of Communication*. *Bell System Technical Journal*, 27: 379–423 and 623–656.
- Soofi, E.S. (1992). A Generalizable Formulation of Conditional Logit with Diagnostics. *Journal of the American Statistical Association*, 87:812–816.
- Soofi, E.S. (1994). Capturing the Intangible Concept of Information. *Journal of the American Statistical Association*, 89:1243–1254.
- Theil, H. (1967). *Economics and Information Theory*. Chicago, USA: Rand McNally and Company.
- Tikhonov, A.N., & Arsenin, V. I. (1977). *Solutions of Ill-conditioned Problems*. New Jersey: John Wiley & Sons.
- Tsallis, C. (2009). *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*. Berlin: Springer.