

## Annex D. Recovery of Pollutant Emissions by Industrial Sector and Region: an Instructional Case

### 1.1 Introduction

In this section, we are going to present a case that, unlike the case studies carried out in previous parts of the monograph, shows the limits of the entropy econometrics approach when some initial conditions are not fulfilled while dealing with an inverse problem. Thus, the example illustrates what should not be done to reach reliable model estimates. This is important for the less experienced modellers who try to recover information on the basis of this theory. In this instance, we see that an important property of respective continuity of probability measures of two hypotheses, as those were defined in part II of the book, is missing. In such a case, more statistical data should be collected to allow for an inference based on regular conditions. In this example, we use a GAMS code.

### 1.2 Recovery of Pollutant Emission by Industrial Sector and Region: the Role of the a Priori<sup>62</sup>

Suppose we dispose of information on aggregated greenhouse emissions simultaneously at the regional and sectorial level. In Table 23 below, we can depict this total information in the last column and last row, respectively. It is important to underscore here that these two pieces of information are from two different sources and, plausibly, were initially collected for different purposes. Thus, we have to join these two pieces with the hope of recovering the real distribution state of pollutant emission by industrial sector and region. In this example, there is an additional complication related to the impossibility of comparing these two totals. In fact, values in the last column constitute an unknown aggregate of standardized quantities of different pollutants with different units of measure. Inversely, in the last row, available statistical data of aggregates have been converted into tons of CO<sub>2</sub> equivalent. Suppose, in this problem, we are asked to recover the regional and sectorial level of greenhouse emissions in tons of CO<sub>2</sub> on the basis of the above availability of statistical data.

If column and row totals were explained in the same unit of measure, we would then have to deal with a direct extension of the Jaynes dice problem (see Table 1 in Part II) from one to two-dimensional discrete space. In this context, finding a unique optimal solution for such a problem requires finding an equation system with a

---

<sup>62</sup> For a complete solution of this problem, we send the reader to: *Acta Physica Polonica A* 2015, 127, A-13. [CrossRef]

closed form. Additional and consistent a priori information is required to enable an optimization entropy device moving from uniform distribution toward an optimal, global solution. For a researcher unacquainted with solving this kind of problem, the present example aims at revealing the limitations of the maximum entropy principle when minimum conditions are not fulfilled. In this example, the a priori is uniformly distributed information of pollutant emissions by industrial sector and region. Doing so, we try then to solve the problem using the well-known principle of insufficient reason already discussed in the introduction of this work.

As already observed, in the present problem, we have in column and row totals different units of measure. In this case, we suggest introducing a scaling factor (as an additional new a priori) in consistency equations of the model, theoretically allowing the table to then balance.

### The Model

According to what has been said above and then without a priori information about the system, we propose applying non-extensive maximum entropy in the criterion function under traditional restrictions such as moments and normality conditions. The model takes the following usual form:

$$\max(H_q(p)) = \left[ 1 - \sum_i \sum_j (p_{ij})^q \right] / (q-1)$$

s.t.:

$$Y_i = c(i) \sum_j X_j p_{ij}$$

$$x_{ij} = X_j \frac{p_{ij}^q}{\sum_i p_{ij}^q}$$

$$1_j = \sum_i \frac{p_{ij}^q}{\sum_i p_{ij}^q}$$

where:

$Y_i$ : means total by row  $i$ ,

$X_j$ : means total by column  $j$ ,

$p_{ij}$ : probabilistic structure of greenhouse emissions by region

$x_{ij}$ : quantity of pollutant emission inside cells.

In this problem, we dispose of up to 122 points to be estimated (then including an additional 17 scaling factors  $c(i)$ ) and only 23 observation points. Naturally, we are dealing with an inverse problem that we know how to solve from theoretical and

Table 23: Total pollutants emission by industrial sector and by region (before estimation)

Regions\	Energy	industry	construction	Transport	Industrial processing	Agriculture	Waste	Other	Total GreenhouseGas (standardized)
Lodz									3.88928103
Mazowieckie									12.2540834
Małopolskie									-0.0723657
Śląskie									8.8672366
Lubelskie									-0.82089026
Podkarpackie									-3.41721603
Podlaskie									-3.92242795
Świętokrzyskie									-4.41226141
Lubuskie									-5.9413236
Wielkopolskie									5.7919945
Zachodniopomorskie									-3.02668716
Dolnośląskie									0.67034141
Opolskie									-4.61146968
Kujawsko-Pomorskie									-1.11131699
Pomorskie									-1.83960985
Warmińsko-Mazurskie									-2.29736824
Total tons CO <sub>2</sub> equivalent	180369		32469	36443	28877	34581	9437	67787	-16.0709987

Source: own calculation on the basis of EUROSTAT information

**Table 24:** Total pollutant emission by industrial sector and region (post entropy estimation).

	indusener	indutrans	trans	procind	agr	dech	aut	Total
lod	445.943	0.00E+00	0.00E+00	0.00E+00	92.881	0.00E+00	0.00E+00	538.824
maz	25838.76	0.00E+00	15494.79	3373.983	2457.911	1027.966	0.00E+00	48193.4
mal	10463.45	1201.047	760.303	0.00E+00	993.605	4634.521	10796	28848.92
sla	7265.899	1201.05	10.309	0.00E+00	0.00E+00	0.00E+00	44955.66	53432.92
lub	26102.65	1201.046	2023.227	0.00E+00	0.00E+00	0.00E+00	0.00E+00	29326.92
pod	11110.19	1201.047	1.773	0.00E+00	2289.638	0.00E+00	0.00E+00	14602.65
podl	7888.465	2031.926	20.356	0.00E+00	753.312	0.00E+00	942.165	11636.22
swi	3001.472	1171.582	1934.917	0.00E+00	2240.56	0.00E+00	298.363	8646.894
lubu	372.657	3300.148	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	3672.805
wiel	2985.15	2914.626	0.00E+00	23710.43	15831.89	2197.649	7562.365	55202.12
zac	9952.093	691.237	1130.836	1792.583	2400.436	1352.623	250.21	17570.02
dol	21697.25	1443.422	9855.435	0.00E+00	2401.033	224.241	995.144	36616.52
opo	7448.701	827.828	3820.23	0.00E+00	317.823	0.00E+00	0.00E+00	12414.58
kuj	25001.55	0.00E+00	1.17	0.00E+00	2401.077	0.00E+00	993.112	28396.91
pom	20794.83	0.00E+00	5.611	0.00E+00	2400.832	0.00E+00	993.983	24195.26
war	0.00E+00	15284.04	1384.048	0.00E+00	0.00E+00	0.00E+00	0.00E+00	16668.09
Total	180369	32469	36443	28877	34581	9437	67787	

Source: own calculation.

empirical models encountered earlier in this work. Outputs from derivations are shown in Table 24. Only a limited part of the information has been recovered. In fact, column totals have been accurately recovered, and line totals seem, to some extent, to conform to our expectations. In fact, we see that predominant regions in greenhouse gas emissions, like Mazowieckie, Slaskie, seem to produce the highest levels of pollutants.

As already known, adding a new piece of a priori information far from uniform distribution and consistent with properties of the system would significantly improve the quality of prediction of these emission quantities. Additionally, as has been observed in the case of forecasting input-output systems (Part III), entropy estimators belong to the family of Stein estimators. As such, smaller probabilities are shrunk and higher probabilities then dominate in the space of solutions. Fortunately, adding more a priori information in the model enhances parameter precision and then allows for recovering the influence of smaller events.