MULTIVERSAL MODEL OF MEASUREMENT OF A COMPOSITE INDEX OF WELL-BEING

Giulio Giacomo Cantone

Abstract. The aggregation of many indicators into a unique index may hide the latent uncertainty in the rank orders of the statistical units due to the freedom in the choice of the aggregating methods. Multiversal modelling enhances this flexibility in the operative definition to derive a posterior distribution of estimates for the latent composite index. This study applies this principle to map the quality of life in 107 provinces in Italy and assess them into robust clusters.

1. Introduction

Gross Domestic Production (GDP) is the fundamental benchmark for measuring the impact of socio-economic policies of well-being and development. Nevertheless, not always a GDP goes along with a consensus on a positive state of happiness or quality of life in a population. In fact, GDP ignores two ideal features of sustainable development: the impact of inequalities in the access to opportunities of development, and the environmental balance between the needs of economic activities and the ecological stability (Patrizii *et al.*, 2017).

The dominance of GDP has been progressively juxtaposed to more granular systems, made of many indicators of socio-economic development. These systems allow the monitoring of indicators for abstract concepts, often articulated in multiple intertwined semantic dimensions. For example, the famous composite Human Development Index (HDI) has the ambition to represent an underlying unity of human activities as biological, technological, and economic facts. In 2014 the Italian National Institute of Statistics of Italy (ISTAT) developed the equally ambitious project for a monitoring system of the Equitable and Sustainable Well-Being of the territories (BES, standing for ``Benessere Equo e Sostenibile"). Originally composed as a system of few indicators to represent 12 dimensions (*dominii*, domains in the Italian) related to the Sustainable Development Goals (SDG) of the United Nations Agenda 2030, over time the number of indicators has grown to exceed a hundred, achieving a remarkable level of granularity even for small territorial areas.

A monitoring system with a hundred variables that, as a collective, define the unitary yet polysemous concept of equitable and sustainable well-being poses more than one challenge for the operative definition of a composite synthetic index, considering that BES is already hierarchically structured along well-distinct domains, too. First of how, abstract concepts as “quality of life” or “sustainability” does not fit well conceptually into continuous scales, where the numeric difference between two units is uninterpretable. Even adopting the composite to establish a ranking of the units, a method that meets the demand of applications more oriented towards discrete decision-making than the estimation of latent features, the compression of many variables into a singular number could result in a decision more driven from the model of aggregation than from evidence (Permanyer, 2012; Greco *et al*., 2019; Alaimo and Maggino, 2020; Munda and Matarazzo, 2020). A solution for such lack of transparency of uncertainty involved in the aggregation is to represent an interval of ranks instead of a single rank (Permanyier, 2012; Paruolo *et al*., 2013; Munda, 2022).

This study extends the application of the framework of Multiverse Analysis (MA, Cantone and Tomaselli, 2024) for the elicitation of a set of 20 rank estimates each of the BES in 107 Provinces. The assumption of MA is that by fitting a sufficiently large and well-specified set of combinations of elements of a statistical model, it is possible to approximate the inherent component of variability of results due to the differences in admissible options in modelling the relations between variables. By achieving such result, it would be possible to isolate inferences mainly driven by difference in evidence from inferences mainly driven by the analytical choices in the model. This study extends this principle to models of aggregation of variables, showcasing the benefits of MA over typical alternatives for sensitivity and robustness analysis (see Leibel and Bornmann, 2024). The latter involve random perturbations of the weighting schemes, while MA, running a finite and regular set of combinations, is deterministic. So it is possible to run paired t-tests for a more robust clustering of the Italian provinces into tiers of BES, from the less concerning to the most.

2. The challenges of synthetising many indicators

In a model of a composite index, the relationship between the indicators and the latent variables can either be *reflective* or *formative* (Coltman *et al*., 2008).

- In reflective relations the latent construct is a ‘factor’ of the indicators, e.g. “the quality of the health systems is *reflected* in a higher life expectancy”. The factor is reconstructed through the tools of Confirmatory Factor

Analysis (CFA) with the hypothesis of a singular factor as communal cause for all the indicators grouped within the domain:

$$\mathrm{X_i} = \mu(X_i) + \lambda_i \cdot \Xi + \epsilon_\mathrm{i} \tag{1}$$

In Eq. 1 Ξ is the factor, also working later as composite index. $\mathrm{X_i}$ is the generic variable within the domain, $\mu(X)$ is its center, generally omitted for standardised X. λ_i is the linear coefficient that best fits the equation (factor loading). This model allows a ϵ_i residual, due to fixed nature of Ξ.
The most immediate test for the hypothesis of a singular fixed Ξ regards the capacity of a singular component of the matrix of the grouped indicators is in the context of the assumptions of the Principal Component Analysis: there must be a singular component that reproduces most of the total variance of the indicators. However, given the confirmatory nature of the model for Ξ, its estimation involves the maximisation of the likelihood for the system of equations of the $X_1, \dots, X_i, \dots, X_k$ indicators in the form of Eq. 1.

- Formative model: in this case the indicators are the causes of the latent dimension (e.g. "an increase of participation in associations and clubs increases the social cohesion of a territory"). Often these processes of formation are designed explicitly through a weighted aggregation of the indicators:

$$\Xi := f(X_1 \cdot w_1, X_2 \cdot w_2, \dots, X_i \cdot w_i, \dots, X_k \cdot w_k) \tag{2}$$

where *W*: $\{w\}$ is the set of weights (*weighting scheme*) and *f(**X**,W)* is the aggregative function.

These models work well when the analysts can identify and access the sufficient set of variables necessary for a correct specification of the construct. In reality, in many cases the analysts do not design the elicitation of the variables to observe, and as in the case of BES, are provided with sets of semantically tied variables collected by a central institution, without a fully specified causal structure. While in these cases formative models should be preferred, the inclusion of a variable in a system like BES should not be regarded as an event independent from the inclusion of the others, for example because the central institution has more ease to observe some indicators than others; as a consequence, an excess of positive correlations could signal a redundance of information, which should be calibrated with the weighting scheme. in the elicitation of the weighting scheme.

Finally, a linearly aggregative function (e.g. arithmetic average, sum, etc.) implies the principle of compensability of variables, i.e. that it is materially possible (and socially desirable) to trade a reduction in one dimension for a larger gain in another; the option of a non-linear function suffers of being a relatively arbitrary choice, instead.

In other words, compared to methodologies based on the selection of one singular key indicator (e.g. GDP), methodologies involving multiple indicators suffer from the risk of being sensitive to the analytical choices of synthesis, leading to logically incoherent results across operational definitions of the target construct. The uncertainty resulting from the freedom of choices in formative models justifies the adoption of a whole *posterior* distribution of multiversal estimates for Ξ, where each specification of the formative model combines a weighting scheme and a aggregative function.

3. The dimensions of equitable and sustainable well-being in Italy

This study adopts the Territorial BES for year 2019 (TBES19), consisting in 58 indicators grouped in 11 domains (dimensions), across 107 Italian provinces[1]. Indicators are rescaled in their standard z-values. and polarity aligned to the semantic of the domain (Mazziotta and Pareto, 2013), e.g. if the Domain is "Security" and the indicator is "Violent crimes" then the indicator is multiplied for -1:

$$\mathrm{z}_*(x) = \frac{x_* - \bar{x}}{s(x)} \tag{3}$$

where x is the element of variable X for a generic province, $\bar{x}$ is the average x across the 107 province, and $s(x)$ is the standard deviation across the 107 provinces; the asterisk reminds the possibility of an multiplication to invert polarity.

Since the scope of the multiversal method is not to estimate a singular index, but to reproduce the variability in rankings due to the analytical choices of a formative model on a system of indicators, instead of synthetising directly the 58 variables, 11 intermediate variables, one for each domain are synthetised through a simple reduction. The scope of this reduction is to not offer intermediate indexes, but to simplify computation and to avoid to bring an implicit weighting due the different availability of indictors for the different domain in the BES of Provinces. For each domain is conducted a preliminary Principal Component Analysis to check if the reported indicators constitute a reflection of a unitary latent factor (Eq. 1). The method of aggregation of the grouped indicators depends on such test:

[1] The dimension "Subjective well-being" is not represented, since it is based on surveys which are not representative at the granularity of provinces. There are 49 missing values in this dataset, less than 0.8% of it. These has been imputed with the predictive algorithm 'missForest'.

- Whereas the principal component with the higher eigenvalue ("first factor") is capable to reproduce more than 50% of the total variance of the variables grouped under the domain, then Ξ is estimated through the Maximum Likelihood Estimator of it (see Eq. 1).
- Where this test fails, the intermediate variables are the arithmetic averages of the indicators with uniform weights, instead.
- The domain "Social participation" groups only one indicator in the TBES19, so that indicator is considered alone as a proxy of the dimension. The domain "Political participation" groups only two indicators, so their average is considered.

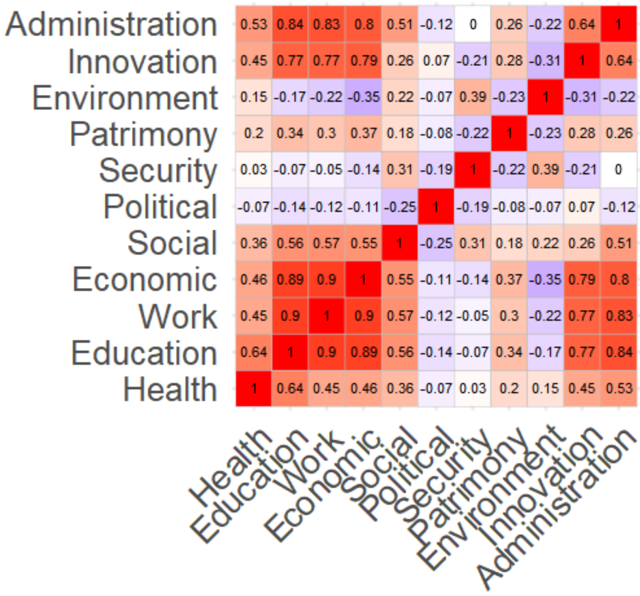
Results of this reduction of variables are summarised in Table 1. For domains Health, and Environment, an eigenvalue of the second factor definitely higher than 1 suggests that hypothetically the grouped indicators may conjointly convey information about two independent latent dimensions, for each domain. For example, in the case of indicators of Health, these include both indicators of health in the population and indicators on the performance of the local health systems. However, it is difficult to assume that the health in a population is independent from the performance of hospitals, etc. For this reason, a *naive* formative model that roughly captures both latent sub-dimensions serves as the alternative to a cross-loaded estimation. Similar considerations are applied for Security, Environment, etc.

Table 1 – *Summary of the aggregation of indicators in intermediate constructs.*

Dimensions	N. of indicators	Variance reprod. by 1st Factor	Eigenvalue of 2nd Factor	Method of synthesis
Health	10	28.7%	2.17	Arith. Average
Education	9	62.4%	.87	CFA
Conditions of work	6	77.93%	.88	CFA
Economic activities	4	78.87%	.48	CFA
Social participation	1	---	---	Direct
Political participation	2	---	---	Arith. Average
Security	6	36.37%	1.19	Arith. Average
Territorial patrimony	3	40.27%	1.01	Arith. Average
Environment	9	26.2%	1.88	Arith. Average
Innovative activities	3	52.36%	.83	CFA
Quality of Administration	3	52.6%	.92	CFA

The matrix of correlations of the 11 intermediate variables (Fig 1) allows to hypothesise a very strong relational dependence among the block of Education, condition of Work and Economic activities; meanwhile Security, Environment and Political participation do not positively correlate with the other Domains.

Figure 1 – *Correlation matrix of the 11 intermediate variables.*



4. Multiversal estimates for a composite index

Once the first step of estimation of 11 intermediate variable is completed, in the second step these are aggregated with a formative model (see Eq. 2) crossing 5 weighting schemes (W) and 4 aggregating functions (f), for a total of 20 multiversal estimates for a unique index Ξ_0:

$$\Xi_0 = f(\Xi_{Health} \cdot w_{Health}, \ldots, X_{QoA} \cdot w_{QoA}) \tag{4}$$

All the considered weighting schemes penalise those intermediate variables highly positively correlated with the other intermediate variables. This principle is adopted for the reason to reduce the potential compensability: in this model of evaluation policymakers cannot not ignore one (or few) dimensions of BES to enhance others, because if they do so, these will become positively correlated to each other and be penalised. In addition, weights must be defined as positive quantities (Munda and Nardo, 2009). The weighting schemes are specified as follow:

- For each pair of constructs for the domains, the similarity of their vectors is measured of through 5 distances: Euclidean, Mahalanobis, Canberra, Soergel (that is the complement of Tanimoto proximity), and Cosine. A summary of these five distances is provided in Appendix A.
- For each Ξ_{i} intermediate construct the distances with other domains, in the form of $d(\Xi_i, \Xi_j)$, are summed. The weight $\mathrm{w_i}$ is equal to

$$\mathrm{w_i} = \frac{\sum_{j}^{QoA} d(\Xi_i,\Xi_j)}{\sum d(\Xi_i,\Xi_j)} \tag{5}$$

The weighted intermediate variables are aggregated into the final index through four specification of the same function:

$$m_q[z_* - \min(z_*)] = \lim_{q' \to q} \left[\sum_{x \neq 0}^{k} \left(\left[z_*^{q'} - \min\left(z_*^{q'}\right) \right] \cdot w \right)^{\frac{1}{q'}} \right] \tag{6}$$

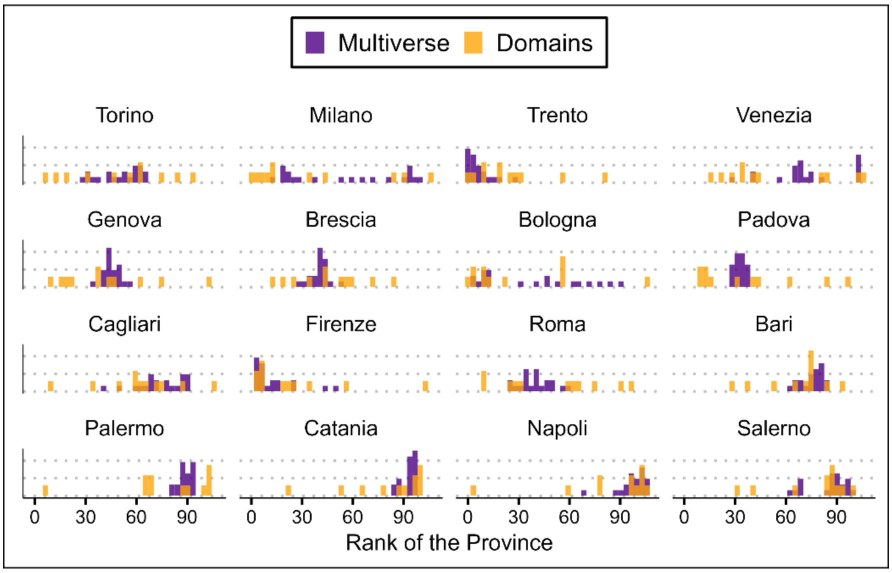
which is a generalised form of the normalised mean (de Carvalho, 2016) converging to the Harmonic Mean, Geometric Mean, Arithmetic Mean and Quadratic Mean for the *q* integer shifting from -1 to 2. Being normalised by substraction of the sample minimum, Eq. 6 is never negative, and being a limit, it ignores the minimum unit in the aggregation, ergo its argument is always positive.
The 20 estimates can be represented as $\hat{\Xi}_{(d,q)}$ where *d* represent which formula for distance is adopted to calibrate the weighting scheme, and *q* the value of the parameter for Eq. 6. Given a finite population of 107 provinces, these estimates can are evaluated through a rank statistic: the higher is $\hat{\Xi}_{(d,q)}$ of a province, the lower is the the rank of the province in the specification (*d*,*q*).

5. Results

The adoption of the 20 multiversal estimates, compared to the ranks for the 11 intermediate variables, result in a shrinkage in the dispersion in the ranks of provinces, as expected[2].

[2] The average Median Absolute Deviation (MedAD) of ranks of the scores of the Italian provinces is equal 14.35 across the 11 dimensions, but it only to 8.58 in the 20 estimates of the composite. This result is visually checked at Figure 3.

Figure 2 – *Relative frequency of ranks for 20 large Italian provinces. 20 multiverse estimates of the composite index are less dispersed than 11 distinct intermediate Domains.*

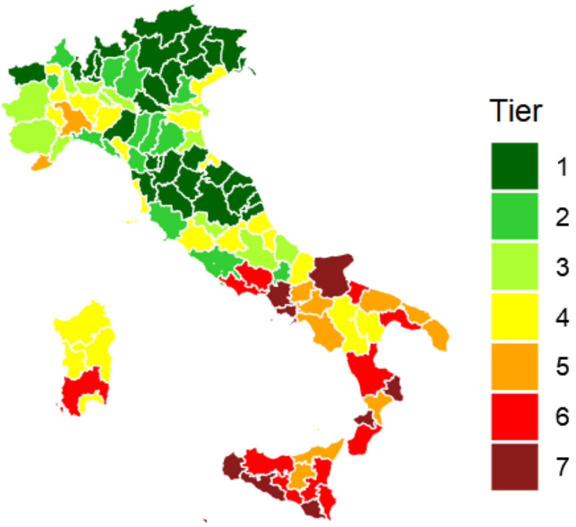


In the multiversal posterior, Pordenone and Trento, the two provinces with the lowest median rank (i.e. the distribution is shifted to the left side of the x-axis in Fig. 2) are characterised not only by a high-performing outlook across the 11 sub-dimensions but also by relatively low variability in ranks. The worst performers Caserta, Agrigento, Crotone and Trapani share concerning states across all domains of the BES system.

For a full visualisation of Italy, a clustering algorithm is run on the multiversal estimates with the sake to establish levels of approximate equivalence in sustainable quality of life among provinces. The core principle of the algorithm is that each member of a higher cluster must be significantly at a higher ranks (more shifted to the right, in Fig 2) of the best performer (lowest ranks) of all the lower clusters. The algorithm is detailed in Appendix B. Given the combinatorial nature of the specifications of the posterior estimates in the multiverse, the Wilcoxon test is ideal to establish this degree of separation. It is important to remark that such procedure does not just split the provinces in percentiles of their outcomes, but optimises am ideal division of territories accounting for the uncertainty in the measurement, so, for example, if a province has an outstanding outperformance of the others, it could still potentially form a cluster in its own, etc.

The algorithm identified 14 clusters of performance. For the sake of a more synthetic representation, these have been manually agglomerated into only 7 tiers associated to seven colours, from green (more desirable state) to red (worst). This result is reported in Fig 3.

Figure 3 – *Clustering of the Italian provinces in tiers of quality of life.*



The procedure clearly identified clusters of excellence (Tier 1) in Alps, North-East and North of Center of Italy, while North-West is significantly lower scoring than these areas. The North-West is penalised by its associated with lower Security and Environment, which in the specific scheme of Eq. 5 are weighted more for being less correlated with the other domains (see Fig. 1). Among the Southern provinces, the algorithm identifies Sicily and Calabria as the two regions more in need of structural interventions to equate social development. The little cluster of provinces between Rome and Naples (plus Foggia) is another concerning area, too.

Appendix A

Distances and weighting schemes

Consider i as the index for the province.

Euclidean distance:

$$d_E(X,Y) = \sqrt{[\sum_i (x_i - y_i)^2]} \tag{A1}$$

Mahalanobis distance:

$$d_M(X,Y) = \sqrt{(\boldsymbol{x}-\boldsymbol{y}) \cdot S_{(X,Y)}^{-1} \cdot (\boldsymbol{x}-\boldsymbol{y})^T} \tag{A2}$$

where **x** and **y** are the vectors of X and Y, and S is their covariance matrix.

Canberra distance:

$$d_C(X,Y) = \sum_i \frac{|x_i - y_i|}{|x_i + y_i|} \tag{A3}$$

Soergel distance:

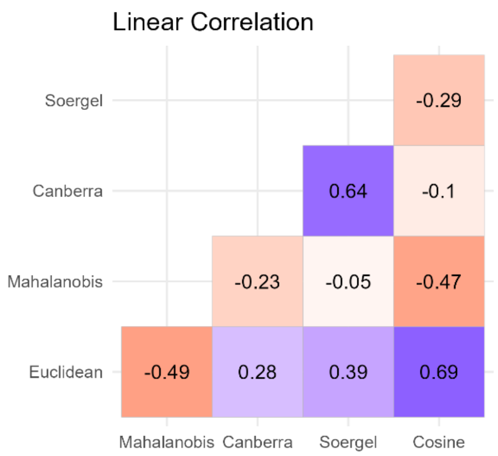
$$d_S(X,Y) = \frac{\sum_i |x_i - y_i|}{\sum_i \max(x_i, y_i)} \tag{A4}$$

Cosine distance:

$$d_{COS}(X,Y) = 1 - \frac{\sum_i [x_i - \min(x_i)] \cdot [y_i - \min(y_i)]}{\sqrt{\sum_i [x_i - \min(x_i)]^2} \cdot \sqrt{\sum_i [y_i - \min(y_i)]^2}} \tag{A5}$$

Figure A1 represents the linear correlations among estimates combinations of 107 provinces and 4 values of the q parameter, for a total of 428 pairings.

Figure A1 – *Linear correlations among estimates of the formative index weighting schemes.*



Appendix B

Algorithm for Ordinal Clustering

1. Order the provinces from the lowest median rank to the highest, and assign them a *i* for their position.
   Set i=1, and the i=1st province to cluster = 1.

2. Increase *i* by 1. Look at the cluster of province with i-1, then look at the province with the lowest median rank within that cluster.
   Call this "benchmark".

3. Run a Wilcoxon test between the estimates the i-th province and the benchmark; match the estimates through their (d,q).

4. If p < $\frac{\alpha=.05}{106}$, then assign the i-th province to new cluster,
   If not, assign *i* to the same cluster of the benchmark.

5. If i $\neq 107$, then reiterate from point 2.
   If not, stop.

Acknowledgements

This is a product of research of project *GRINS- Growing Resilient, INclusive and Sustainable* project (GRINS PE00000018 – CUP E63C22002120006). The views and opinions expressed are solely those of the author.

TBES19 dataset is a courtesy of Prof. Venera Tomaselli.

References

ALAIMO L.S., MAGGINO F. 2020. Sustainable Development Goals Indicators at Territorial Level: Conceptual and Methodological Issues—The Italian Perspective, *Social Indicators Research*, Vol. 147, No. 2, pp. 383–419.

CANTONE G.G., TOMASELLI V. 2024. Characterisation and Calibration of Multiversal Methods, *Advances in Data Analysis and Classification*, 10.1007/s11634-024-00610-9.

COLTMAN T., DEVINNEY T.M., MIDGLEY D.F., VENAIK S. 2008. Formative versus reflective measurement models: Two applications of formative measurement, *Journal of Business Research*, Vol. 61, No. 12, pp. 1250–1262.

DE CARVALHO M. 2016. Mean, What do You Mean? *The American Statistician*, Vol. 70, No. 3, pp. 270–274.

GRECO S., ISHIZAKA A., TASIOU M., TORRISI G. 2019. On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness, *Social Indicators Research*, Vol. 141, No. 1, pp. 61–94.

LEIBEL C., BORNMANN L. 2024. Specification uncertainty: What the disruption index tells us about the (hidden) multiverse of bibliometric indicators. *Scientometrics*. https://doi.org/10.1007/s11192-024-05201-1.

MAZZIOTTA M., PARETO A. 2013. Methods for constructing composite indices: one for all or all for one? *Italian Journal of Economic, Demographic and Statistical Studies - Rivista Italiana Di Economia Demografia e Statistica*, Vol. LXVII, No 2.

MUNDA G. 2022. Qualitative reasoning or quantitative aggregation rules for impact assessment of policy options? A multiple criteria framework, *Quality & Quantity*, Vol. 56, No. 5, pp. 3259–3277.

MUNDA G., MATARAZZO A. 2020. On the impossibility of using “the correct” cost–benefit aggregation rule, *Journal of Economic Studies*, Vol. 47, No. 5, pp. 1119–1136.

MUNDA G., NARDO M. 2009. Noncompensatory/nonlinear composite indicators for ranking countries: A defensible setting, *Applied Economics*, Vol. 41, No. 12, pp. 1513–1523.

PARUOLO P., SAISANA M., SALTELLI A. 2013. Ratings and Rankings: Voodoo or Science? *Journal of the Royal Statistical Society Series A: Statistics in Society*, Vol. 176, No. 3, pp. 609–634.

PATRIZII V., PETTINI A., RESCE G. 2017. The Cost of Well-Being, *Social Indicators Research*, Vol. 133, No. 3, pp. 985–1010.

PERMANYER, I. 2012. Uncertainty and robustness in composite indices rankings. *Oxford Economic Papers*, Vol. 64, No. 1, pp. 57–79.

Giulio Giacomo CANTONE, University of Catania / University of Sussex, ggc23@sussex.ac.uk