

Clusters and Equivalence Scales

Gustavo De Santis and Mauro Maltagliati

Abstract Equivalence scales have always attracted the attention of social scientists and researchers. Their estimation is difficult, however, and even apparently solid microeconomic foundations do not necessarily lead to consistent results. We contend that this depends on "style effects": households with the same economic resources and identical "needs" (i.e. observed household characteristics) may spend differently, following unobservable inclinations, that we call "style" or "taste" effects. We submit that these style effects must be controlled for in order to obtain unbiased estimates of the equivalence scale. One way of doing this is to create clusters of households, with different structure (e.g., number of members) but similar "economic profile", in terms of both standard of living and "style". The equivalence scales that derive from the comparison of the incomes of households belonging to the same cluster prove defensible on theoretical grounds, and are empirically reasonable and consistent.

Keywords: Equivalence scales, measuring poverty and inequality, cluster analysis

1 Equivalence scales: a dead end?

As Muellbauer and van de Ven (2004: 1) put it, "despite a considerable research effort ..., almost every aspect of equivalence scale specification remains controversial". Indeed, the difficulties seem to be of both micro-economic and statistical nature.

Micro-economic theory suggests that equivalence scales cannot possibly be derived from the observation of empirical consumption patterns, because of their implicit compensation nature. Consider in particular the case when the size of the family increases because of the birth of a child that the parents did want. How can they possibly be worse off than before? And how can they need extra money to be

¹

Gustavo De Santis, Dept. of Statistics, Un. of Florence, email: desantis@ds.unifi.it
Mauro Maltagliati, Dept. of Statistics, Un. of Florence, email: mmaltag@ds.unifi.it
Financial support from the Italian MIUR is gratefully acknowledged.

"compensated" for the economic loss that the presence of their child causes them? With varying emphasis, this objection can be found in virtually all authors, since when Pollak and Wales (1979) first raised the issue. See also Lewbel and Pendakur (2006) for a recent review.

Statistically speaking, estimation methods are either mono- or multi-equational. In the former case, one variable (e.g. food share, as in the Engel method) is assumed to vary monotonically with (and thus represent) the standard of living of the household. Two households, say H_1 and H_2 , where this indicator reaches the same value are thus assumed to enjoy the same standard of living, and the ratio of their incomes (Y_2/Y_1) provides the desired equivalence scale. Mono-equational models are easy to estimate. Among them, in particular, Engel's method generally leads to reasonable results: this is probably the main reason why the "Carbonaro" scale (based on the Engel method) is generally used in Italy, by both Istat (2009), and the Government (with the ISEE scale, <https://servizi.inps.it/servizi/isee/default.htm>, a variant of Carbonaro's one).

Mono-equational models, however, are scarcely defensible on a theoretical plan: whatever indicator one chooses (e.g. food share) there is always the possibility that two otherwise identical households (i.e. with the same level of economic wellbeing) will behave differently, that is, will spend a different amount of money on food, and thus incorrectly appear to be at different standards of living, according to the selected indicator. This may happen in part because of price effects (consumers tend to purchase less of what becomes comparatively more expensive), and in part because of a set of other reasons which we will here call "taste" or "style", for the sake of simplicity. If all households are observed in the same period (cross section), prices are given and identical for everybody: their impact should therefore be negligible, if not totally absent. Style effects, however, may still have an impact.

In order to see more clearly how this could happen, let us assume that the object of the research is the estimate of child costs (or, which is the same, the estimate of the equivalence scales for households with and without children) and that we try to do this using food share as an (inverse) indicator of the standard of living (Engel's method). The result may be biased basically for three, non alternative reasons:

- a) households with preference for (and therefore greater presence of) children are selected in many ways, including some that directly affect our indicator of economic well being. Imagine for instance that these households are more indoor-oriented and tend to consume more food at home. They will (incorrectly) appear poorer than they actually are, and this apparent lack of economic resources will normally be associated with (if not attributed to) the presence of children;
- b) after the birth of a child, parents may change their lifestyle and, for instance, spend more time at home, and thus also consume proportionally more in food. Once again, parents will (wrongly) appear poorer than they actually are;
- c) a child differs from his or her parents: he/she may consume proportionally more food than other items (e.g. travel, culture, apparel, etc.). Overall, the household will shift its consumption pattern towards food, which, here too, will make households with children appear poorer than they actually are.

With multi-equational models (e.g. those that derive from the so-called "complete demand systems"), the problem becomes less visible, but is still present, and emerges clearly whenever one looks a bit deeper into the data: equivalence scales that abruptly change from one year to the next without any apparent reason, or that prove too sensitive to patently minor details (e.g. number or type of expenditure items

considered), suggest that the estimates are not robust. And this happens even with the adoption of methods specifically intended to limit or altogether eliminate this bias (see. e.g. De Santis and Maltagliati 2001, 2003).

We believe that this depends on two main causes. One is that there is an intrinsic multicollinearity between the object of interest (the equivalence scale) and the style effects, whenever one tries to model them separately (De Santis and Seghieri 2006). The other is that individual (household) consumption data are collected in so short a lapse of time and are therefore so variable, and so full of zeros, that even the best models can explain only a tiny fraction of the observed variance (see. e.g. Bollino, Perali, and Rossi, 2000). Aggregating observations, or expenditure shares, or both, is an option, of course, but then other questions arise: how to proceed? When to stop? Besides, as mentioned before, aggregation is not neutral: it affects the results very deeply (De Santis and Seghieri 2006), and, to the best of our knowledge, this (major) difficulty has not been overcome yet. Are equivalence scales at a dead end?

2 A possible solution: clusters

Both issues (multicollinearity and "proper" aggregation of expenditure items) may find a solution in clustering observations in such a way that each cluster contains households that "behave similarly" in economic terms, i.e. standard of living and lifestyle. Let us imagine that each household can be characterized by the following set of variables: E (Economic indicators), Y (income) and N (Number of members), so that

$$1) \quad H = (E_h, Y_h, N_h)$$

The problem is that, for any given household, each economic indicator in E_h will reflect both its level of affluence (some households are richer than others) and its lifestyle (some households prefer the car, and some the motorbike, if they can't afford to buy both).

Indicators of lifestyle tend to "sum to zero" for any given level of affluence. This is mathematically true if one refers to budget shares, where styles are, by definition, deviations from the mean (De Santis and Maltagliati 2003). With other indicators, this is no longer true in mathematical terms but the notion tends to preserve its validity in a more general sense. In the example above, the idea is that two households at a comparable level of affluence may possess different means of transport (one may have a car and the other may have a motorbike): in this case, their economic profile will differ, and, in a clustering process based on the dichotomous variables "Car owner" and "motorbike owner", they will belong to different branches.

Let us now imagine that we consider only 1- and 2- person households, and that we take several households of both types. On the basis of our two dichotomous indicators we can form four groups, depending on whether they own both car and motorbike, only car, only motorbike, or none of these. In each of these four groups we will find households of both types, with one and two members. What we can say is that, within each cluster, households tend to share a common economic profile which means that they are comparable in terms of both affluence and lifestyle. Now let $Y_{1,\alpha}$ and $Y_{2,\alpha}$ be the average income of households belonging to cluster α who have, respectively, 1 and 2 members. We expect that $Y_{2,\alpha} > Y_{1,\alpha}$ (because two persons need more resources than

one to reach any given level of affluence) and, since we are assuming that their level of affluence is the same, we can say that $Y_{2,\alpha}/Y_{1,\alpha}$ is the equivalence factor specific for that cluster. If we now do the same also for the remaining clusters ($\beta, \gamma, \dots$) and we average the results, what we obtain is an estimate of the "true", "style-free" equivalence factor Y_2/Y_1 . Of course, if the procedure is correct, each estimate should give more or less the same value: indeed, we can compute the confidence interval for each estimate, and, with a sufficient number of clusters, we can see whether estimates tend to heap around a specific value (hopefully, the desired one), or if they vary more or less at random; we can see if there are outliers (that should preferably be discarded), etc.

Note that this procedure does not require that we distinguish between indicators of affluence and style, of that we can tell in what proportion affluence and style affect each item. We simply contend that within each cluster, all households share a common economic profile, or, in other words, they are similar in terms of the two dimensions that interest us, affluence and style.

So much for the general logic of our procedure: let us now see what we can do and obtain in practice.

3 An empirical application to Italian data on consumption

We work on data that derive from Istat's consumption survey for the years 2003 to 2006 (four years), after inflating monetary values so as to make them comparable.

We consider households made up of 1 to 5 persons. We drop larger households, because they are rare and peculiar, and because at this preliminary stage we want to make sure that our method works reasonably well on more standard cases. Next, we "trim" our data set by discarding outliers, i.e. households that, given their dimension and their overall spending, spend proportionally too much or too little on each expenditure item (the acceptable range is centered on the median, and spans 4 interquartile intervals, more and less - that is 8 interquartile intervals, overall). This has two purposes: firstly, there may be errors in these cases, and it is better to get rid of them. Secondly, we look for households, of various dimensions, that are comparable in economic terms - and these outliers, by definition, are different from virtually every other household in the survey. Obviously, the equivalence scale that we will derive will hold only for "average", not for "extreme", households. This leaves us with nearly 87 thousand households. Notice that we do not have information on their income Y : we will instead use their total spending X .

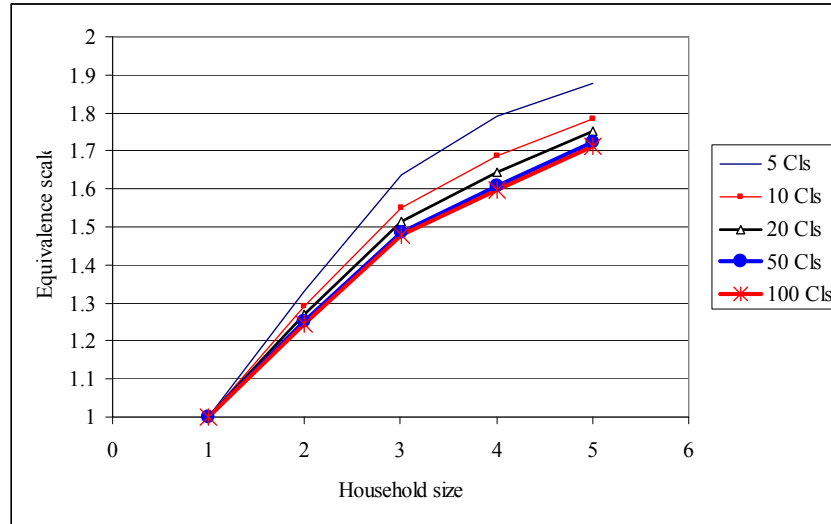
For each household we calculate several indicators that we deem in principle apt to describe its affluence, either directly (the indicator increases with total spending X - e.g. probability of having heating in one's home) or inversely (the indicator decreases with X - e.g. food share). We drop several inconsistent indicators, i.e. indicators that do not vary monotonically with X - for each household dimension, in the relevant range of X .

After this, we are left with 10 indicators, two continuous (1. expenditure share in necessities; 2. expenditure share in luxuries) and eight dichotomous variables (3. kitchen as a separate room; 4. garage; 5. motorbike; 6. washer; 7. dishwasher; 8. air conditioner; 9. telephone; 10. electric devices for house cleaning).

On the basis of these 10 indicators E_{h_i} , we form clusters. In a clustering process, several potentially important choices are arbitrary: for instance what criterion to use (here: Ward), whether the process should be hierarchical or non-hierarchical (here: hierarchical) and how many groups to form (and we formed several: cf. Figure 1). In all cases, the ratio $(X_{n,g}/X_{1,g})$, for each group g , gives the equivalence scale for that group. The weighted average of these group-specific equivalent scales gives the overall equivalence scale, which is what we are looking for (see Figures 1 and 2).

In several ways, results are as expected. Firstly, equivalence scales increase monotonically with N , which was not granted, since this is a non parametric model, and each equivalence factor (for each household dimension) forms independently of each other. Secondly, our equivalence scales are scarcely sensitive to the number of clusters that one forms. Indeed, as this number increases, the scale seems converge towards a unique and apparently stable value, which it is tempting to interpret as the "true" value of the scale.

Figure 1: Equivalence scales obtained with clustering, and a varying number of clusters (Households with 1 to 5 members)

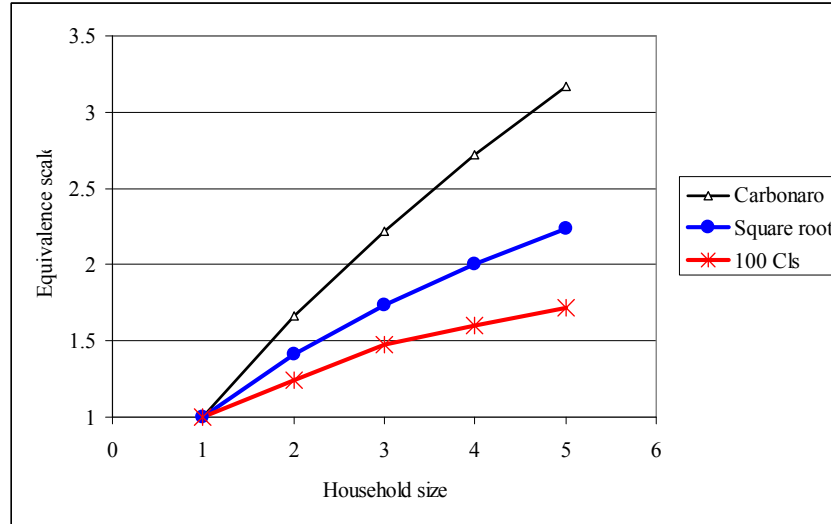


Source: Authors' elaborations on Istat data,

Besides, the scale does not seem to change too much if one forms clusters with different variables and/or different criteria, so long as clusters are numerous and the classification variables are "well behaved", i.e. vary monotonically with the total resources of the household (results not shown here).

What is perhaps less convincing is that our equivalence scale is considerably flatter than those customarily in use (Figure 2): not only flatter than Carbonaro's scale (frequently considered to be too steep, as it is generally the case with Engel-based scales), but also flatter than the scale given by the square root of the number of household members, which is generally thought to be too flat.

Figure 2: Equivalence scales with Carbonaro's method, square-root of members (OECD) and our own (100 clusters; Households with 1 to 5 members)



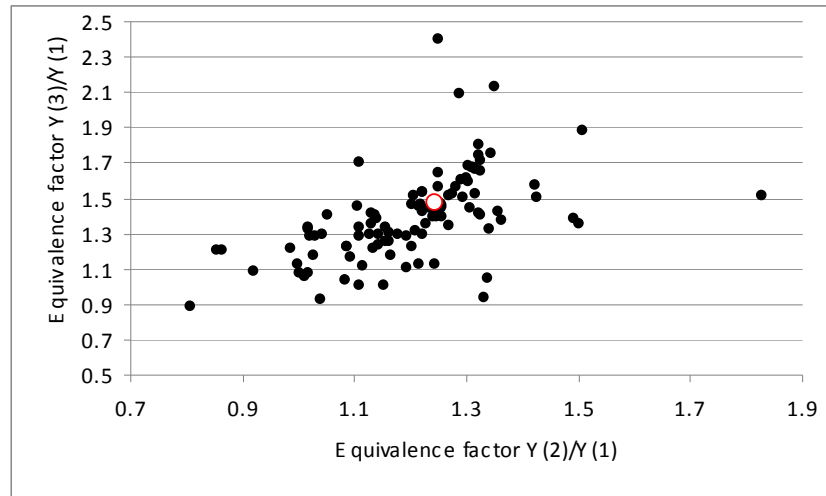
Source: Authors' elaborations on Istat data and Istat (2009).

4 Discussion

The method that we are proposing for the construction of equivalence scales has, in our view, a few merits. In the first place, to the best of our knowledge, it is completely original. Besides, the results that we have obtained up to now, although admittedly few in number, are consistent with a priori expectations and, as far as we could test, scarcely dependent on the number of clusters or the type and number of variables used for clustering. Remember that this is a non parametric method, so that, for instance, the equivalence scale for a 5-member household does not depend on what has been estimated for a 4-member household, or for any other household, for that matter. Therefore, there was no a priori guarantee that the equivalence scale would result in the shape of figures 1 and 2, that is increasing but progressively less so as the number of household members augments. Since this is what it is generally thought to happen with equivalence scales, the result is encouraging.

Besides, one can measure how reliable the estimate is: each cluster provides an estimate of the scale, whose reliability depends, among other things, on the number of households within that cluster. And all the clusters together produce a set of estimates, for which confidence intervals can be computed (see, e.g., Figure 3).

Figure 3: Estimates of the equivalence factor Y_2/Y_1 (average=1.245; std.err=0.015) and of the equivalence factor Y_3/Y_1 (average=1.477; std.err=0.025) with 100 clusters



Source: Authors' elaborations on Istat data,

Theoretically speaking, the method is more solid than it may appear at first sight: it is a conscious attempt at circumventing the thorny, but underestimated problems that derive from style effects - i.e. households of comparable standards of living who, nonetheless, choose to spend their money in different ways.

But this method has its limits, too. For instance, it relies on the implicit assumption that certain indicators (e.g. having a motorbike) have the same economic meaning for small and large households. It may become difficult to handle when households differ not only because of their size but also because of other characteristics (e.g. the sex and age of their members). It mixes flow data (expenditure) with stock data (ownership of certain durables), which may not be fully justifiable.

Obviously, the method is still in its very early phase: it needs to be tested on several more databases, for several years, with different indicators of affluence, before it can qualify as a viable alternative for the estimation of equivalence scales. Notice, however, that one of its advantages is that it does not necessarily need consumption data to be applied: income and saving data, such as those that are available in the Bank of Italy's SHIW (Survey on Household Income and Wealth), can work just as well, provided, of course, that a sufficient number of reliable (i.e. affluence-dependent) indicators is available. Our preliminary analyses for Italy, not shown here, are very encouraging in this respect: we could indeed derive equivalence scales also from income and saving data (Source: Bank of Italy SHIW), and these proved very close to those shown above, obtained from consumption data (Source: Istat Survey on Households' Budgets). Once again, to the best of our knowledge, this type of comparison is an absolute novelty in the history of equivalence scales.

References

1. Bollino, C.A., Perali, F., Rossi, N.: Linear household technologies, *J. of Appl. Econom.*, 15, pp. 275-287, (2000)
2. De Santis, G., Maltagliati, M.: De gustibus non est disputandum? A new approach to the estimation of equivalence scales, *Statistical Methods and Applications, J. of the Ital. Stat. Soc.*, 10(1-3), pp. 211-236, (2001)
3. De Santis, G., Maltagliati, M.: Equivalence scales: a fresh look at an old problem. Theory and empirical evidence, in C. Dagum, G. Ferrari (eds.) *Household behavior, equivalence scales and well-being*, pp. 29-53, New York/Berlin, Physica Verlag, (2003)
4. De Santis, G., Seghieri, C.: How robust are equivalence scales?, *Proceedings of the 43rd SIS Sci. Meet. (Torino, 14-16 June)*, pp. 133-136, Padova, Cleup, (2006)
5. Istat: La povertà in Italia nel 2008, *Statistiche in breve*, (2009), http://www.istat.it/salastampa/comunicati/non_calendario/20090730_00/
6. Lewbel, A., Pendakur, K., Equivalence Scales. Entry for *The New Palgrave Dictionary of Economics*, 2nd edition, Boston College and Simon Fraser University, (2006), <http://www.sfu.ca/~pendakur/palequiv.pdf>
7. Muellbauer, J., van de Ven, J.: Estimating Equivalence Scales for Tax and Benefits Systems, *NIESR (Natl. Inst. of Econ. and Soc. Res.) discussion papers*, 229, (2004), <http://www.niesr.ac.uk/pubs/dps/dp229.pdf>
8. Perali F.: *The Behavioral and Welfare Analysis of Consumption*, Kluwer Academic Publishers, Boston, (2001)