Dipartimento di Statistica
"Giuseppe Parenti"

WORKING PAPER 2010/10

Comparing like with like: cluster-specific equivalence scales

Mauro Maltagliati,
Gustavo De Santis

Università degli Studi
di Firenze

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

Comparing like with like: cluster-specific equivalence scales

Mauro Maltagliati and Gustavo De Santis

Dept. of Statistics, University of Florence, Italy

Summary. On the basis of a few well-behaved indicators of economic well being, we create clusters of households, with different structure (number of members) but similar "economic profile", in terms of both standard of living and "style" (i.e. way of spending money, for any given standard of living). Since, by assumption, households are comparable within clusters, the ratios between their total monthly outlays produces cluster-specific equivalence scales. By properly averaging over clusters, we derive the general equivalence scales for Italy for the years 2003-2008. With the same logic, and a few adjustments, we also obtain measures of inflation and, separately, of purchasing power parities (PPP) for different regions within Italy

Keywords: clusters, equivalence scales, purchasing power parity, inflation.

1. Introduction

An equivalence scale [S_r, ..., S_h, ...] is a set of index numbers, each of which summarizes the economic needs of the *h*-th household relative to those of the *r*-th, or reference, household, whose index S_r is conventionally set to 1. An illustrative example is the recent OECD (2008a, b) square-root equivalence scale [1.00; 1.41; 1.73; 2.00; 2.34; ...], where households are assumed to differ only by their size, and where the needs of a household of, say, 3 members appear to be S_3=1.73 times as high as those of the reference (1-person) household. In principle, an equivalence scale can account also for other household characteristics (e.g. sex, age, and health status of members; place of residence, etc.), but we will ignore these additional dimensions in this paper.

Equivalence scales permit one to make utility comparisons of households with different characteristics (size). In order to do this, one must take a measure of the economic resources of a household of type *h* (e.g. income Y_h, or, as in this paper, total monthly expenditure X_h) and divide it by its corresponding equivalence factor S_h: this gives $X_h/S_h=E_h$ the equivalent expenditure of that household (and, similarly, $X_k/S_k=E_k$ for household *k*). While a comparison of the two monetary expenditures X_h and X_k may not tell much about the relative economic standing of the two households, *h* and *k*, comparing their equivalent counterpart E_h and E_k does, because these measures encompass their possibly differing needs.

This is why equivalent expenditures (or incomes) are, in general, more interesting than monetary ones, for instance in domains such as welfare analysis and optimal taxation (Lambert, 2004; Muellbauer and van de Ven, 2004), economic consequences of a divorce (Aassve *et al.*, 2007) or other demographic events (e.g. birth of a child), microeconomics of consumer behavior (Fernández-Villaverde and Krueger, 2007), etc. This is also why equivalence scales have attracted so much attention and have originated such a large literature, although the latter characteristics also depends on the several difficulties that their estimation poses (see, e.g., Perali, 2001; Lewbel and Pendakur, 2006).

In this paper, after recalling briefly what we believe are the cornerstones of this literature (section 2), we introduce an approach to the estimation of equivalence scales that, to the best of our knowledge, is completely new, and very promising (section 3). After introducing the database that we use (Italian consumer expenditure microdata, 2003 to 2008 - section 4), we present our results, in section 5, and we show that they go well beyond mere equivalence scales: indeed our methodology can also produce measures of inflation and of purchasing power parity for various regions, within Italy. In section 6 we discuss our main findings and suggest lines of possible future research.

2. How to estimate an equivalence scale: a glance at the literature

In the UN square-root equivalence scale introduced above [1.00; 1.41; 1.73; 2.00; 2.34; ...] household needs increase with size, but less than proportionally, so that every additional member is comparatively cheaper than the preceding one. Unfortunately, decreasing marginal costs[1] are perhaps the only universally accepted feature in a field where, "despite a considerable research effort [...], almost every aspect [...] remains controversial" (Muellbauer and van de Ven, 2004: 1).

How can one hope to estimate an equivalence scale in practice? The basic idea is to look for some empirical indicator suggesting that, on average, households with n members need to spend X_n to be "as well off as" the reference (often, one-person) household spending X_r: the ratio $X_n/X_r=S_n$ is the desired equivalent factor. Strictly speaking, there is no guarantee that this S_n factor remains constant as X_r changes. If it does, scholars speak of independence of base (IB - Lewbel, 1989), or, less evocatively, equivalence scale exactness (ESE - Blackorby and Donaldson, 1993). For the sake of simplicity, we will initially assume that the IB property holds (but our results, as shown for instance in tabel 7, indicate that this assumption may be scarcely defensible).

What empirical indicator can be used to assert that two households are equally well off? It would seem natural to look at their consumption behavior. Unfortunately, according to several micro-economists, equivalence scales cannot be derived from empirical observation of consumption patterns. Formally, the reason is that - given prices (vector $\mathbf{p}$), a certain utility level u, and certain household characteristics ($\mathbf{h}$) - beside the "basic" cost function C($\mathbf{p}$, u, $\mathbf{h}$), there is also an infinite set

[1] In the OECD (2008) example of the text, marginal costs are [1.00; 0.41; 0.32; 0.27; ...]. In the "old" OECD equivalence scales, used especially in the eighties, after the first member, whose cost is always set to 1, additional members cost 0.7, if adult (>14 years), or 0.5, if children. In the "modified" OECD equivalence scales, used especially in the nineties, these marginal costs reduce to 0.5 and 0.3, respectively.

of other cost functions C(**p**, $\phi(u; \mathbf{h})$, **h**) that derive from this and that are all undistinguishable from each other, if the transformation $\phi(u; \mathbf{h})$ is increasing in u. Pollak and Wales (1979) were probably the first to raise this point, which has become a sort of *leit-motif* ever since.

This conclusion, however, appears to be somewhat too extreme. On the practical plan, nobody can seriously contend that, with, say, 1000 Euros per month (and everything else equal), a couple with two children (n=4) can as well off as a single person (r). There *must* be a sum of money, greater than 1000 (but plausibly smaller than 4000), that grants these two households the same (or at least a comparable) standard of living. To find this amount, however difficult it may be, is to find the equivalence scale.

But there is also a theoretical objection, about the meaning of the term *utility*. In a broad sense, it is true that, given income, rational parents who choose to have children cannot be worse off than before: they may have access to fewer material resources, but what they gain, or expect to gain, in some other sphere (company, personal fulfillment, assistance in old age, ...), must at least compensate what they lose in economic terms. In a stricter sense, however, utility can be interpreted as a sort of "real" availability of resources, where "real" means "keeping into account the consumption needs of every household member". This implies that, with given resources but more members, a household may be both happier and economically poorer - and equivalence scales try to measure precisely the latter dimension.

Let us imagine, then, that a proper deflator S_h exists for each household h, so that the "real" resources of this household can be expressed as $u_h=E_h=X_h/S_h$. In this case, its cost function becomes C(**p**, u_h) = C(**p**, X_h/S_h), where, since **p** and X_h are observable, S_h can, at least in principle, be estimated. Even in this case, however, household heterogeneity remains an issue: given prices and utility (stricter sense), two different households need not have the same consumption profile: for instance, because some are vegetarians, and others are not. Now, if household size is systematically associated with non-observed components of heterogeneity, for instance, because vegetarians tend to live in larger households, this may bias the estimate of the equivalence scale. We will get back to this point shortly.

At this stage, we have already implicitly introduced a strong assumption: that the standard of living is the same for all household members. Bourguignon et al. (1994), among others, object to this idea, but the assumption seems to us to be tenable in several cases, and certainly in the countries where husband and wife are, by law, on an equal footing, where divorce exists (and is relatively widespread), and where children are highly valued - sometimes even more than parents themselves (Dalla Zuanna, 2004). Italy is surely a country that meets all the three criteria.

Several other non-neutral assumptions are frequently adopted, because they make estimation easier. Two of them we have already encountered: that the equivalence scale depends on household size only, and not on other household characteristics, and that the IB (or ESE) property holds (i.e., S_h does not vary with utility). Another assumption refers to the functional form to adopt: should it be linear, log-linear, or what? Besides, what is the best dependent variable in regressions: Q_c (quantity of item c that is consumed), X_c (expenditure for item c), X_c/S_c (equivalent expenditure for item c, where S_c is an item-specific equivalence scale that needs to be estimated), $w_c = \frac{X_c}{X}$ (share of the total expenditure devoted to good c), ...? Should all the consumption items be considered, or just a few of them? Can elementary consumption items be aggregated into broader categories (e.g. can pasta, meat and fruit be summed up, so as to form "food")? If yes, how many categories must be formed, and how?

Most of these arbitrary choices or assumptions can, at least in principle, be put to test, and the data should reveal which functional form works better, what explanatory variables one is justified to omit, etc. Unfortunately, no estimation method can simultaneously keep everything under control: in this field of research analysts normally work with micro data (see Section 4), regressions on which typically explain only a small fraction of the total variance. Therefore, virtually any "complication" of the model (e.g. more variables, squared effects, interactions, etc.) proves statistically justified (Bollino, Perali, and Rossi, 2000), and common sense, more than statistical tests, guides most of the decisions.

Another thorny issue is about the "nature" of our equivalence factors S_h. The multiplicative form that we adopted above, which leads to *scaling*, is not the only possible option: *translating* is another, by which equivalent expenditure results in $E_h = X_h - T_h$. Besides, both scaling and translating, in either order, can be envisaged, which results in either $E_h = \frac{X_h}{S_h} - T_h$ or $E_h = \frac{X_h - T_h}{S_h}$. Notice, in passing, that the introduction of translating implies the abandonment of the IB assumption: if $T_h \neq 0$, equivalence scales do depend on the level of utility.

Let us now get back to our original question: how can one derive an equivalence scale? The basic assumption is that two households (h and r) would have the same expenditure pattern if the difference in their size were properly compensated, by inflating the resources available to h. There are several ways of tackling the issue, among which the most popular consists in trying to equate budget shares: $w_{ch} = w_{cr}$.[2] In some cases, only some "relevant" budget shares matter: for instance,

[2] Some applications use absolute expenditure, instead: e.g. Rothbarth's (1943) method, based on adult's good, or Prais and Houthakker's one (1955). Note, however, that, in some instances, Rothbarth's ideas have also been applied by working on budget shares (e.g. Koohi-Kamali, 2008).

Engel's (1895) method considered only the share of a few basic necessities, among which food. In other cases, with the so-called complete demand systems, the whole expenditure profile is kept under control. Remember that w_{ch}=f_c(**p**, X_h, **h**), and that, in cross sectional observation, prices **p** are the same for every household, and can therefore be ignored. Notice, further, that it is always possible to find item-specific transformations, for instance of the type $\frac{X_h}{S_{hc}}$ or $\frac{X_h - T_{hc}}{S_{hc}}$ (h=household; c=consumption item; S=scaling factor; T=translating factor), that lead to $w_{ch}=w_{cr}$. The problem is how to guarantee that this equality hold for all goods c and households h at the same time, or, at least, that the deviations from this requirement be as small as possible[3]. Originally, the predominant idea was that of estimating item-specific equivalence factor S_{hc} (or perhaps both S_{hc} and T_{hc}), and later to summarize them in a proper synthetic measure $S_h=f(S_{hc})$ [and, possibly, $T_h=f(T_{hc})$], typically with a weighted average[4].

A more direct approach, proposed by Ray (1983), ignores item-specific transformations factors (of the type S_{hc} and T_{hc}), ignores translation, and focuses directly on the general scaling factor S_h that makes larger households poorer (same resources, but more needs, and therefore lower equivalent expenditure). This factor can be estimated by minimizing the differences in the consumption profiles of households with different size (or, more generally, different structure), through a set of regression equations of the type[5]

1) $$w_c = b_{0c} + b_{1c}\ln(X_h/S_h) + \varepsilon_c, \quad \left[\forall c; \sum b_0 = 1; \sum b_1 = 0\right]$$

All these approaches have one important feature in common: every systematic difference in the consumption profile is assumed to depend on a cost effect. Ray's (1983) approach, however, when applied empirically, brings to the fore a peculiarity that several scholars had already noted, especially in their criticism of Engel's method (see, e.g., Deaton and Muellbauer, 1986): even after "monetary compensation" (i.e., at the same level of utility), structurally different households can still differ in their consumption profile. De Santis and Maltagliati (2004) talk of "style", or "taste",

[3] In practice, both T and S can be modeled in several different ways, and this, in turn, affects the results. One of the simplest solution is to write $T=t(N\text{-}1)$, where t is a parameter to be estimated, N is the size of the household, and the reference household r has N=1. Analogously, one can write $S=s(N\text{-}1)$, in order to link the scaling effect to the household size.

[4] This is the idea that lies behind the approaches of Barten (1964), Gorman (1976). Prais and Houthakker (1955), too, had something of this type in mind, although they did not work with budget shares (cf. footnote #2).

[5] The regressions are typically more complicate than this: they normally introduce also the square of the term $\ln(X/S_h)$ and they may add a set of other regression variables. We will disregard these details, here.

effect, and offer three main, non alternative explanations for it, all linked to the notion of heterogeneity:

1) larger households (who typically also have more children at home) may be selected in several ways, some of which affect our indicators of utility. For instance, they may prefer to spend more time at home: in this case they will also consume more food indoor, watch more television (instead of going to the cinema), possibly spend less in alcohol and tobacco, etc.;
2) the birth of a child may make parents change their lifestyle and, for instance, may induce them to give up some outings, because they now value their child's company more than their friends' one. They may also decide to move to a larger apartment, possibly in a less chaotic part of the town: in Italy, for instance, this means leaving the center of towns and going towards the periphery, where rents are lower (and gardens larger);
3) a child differs from his (her) parents: he/she may consume proportionally more of some items (e.g. food and health) than the couple used to do. The general consumption pattern of the family will thus have to strike a new balance between a tighter budget constraint, on the one hand, and the changed needs of its members, on the other.

Notice that it is precisely because of these style effects that all mono-equational estimation methods (Engel's food share, Rothbarth's adult goods, etc.) may be misleading: they interpret as an income effect a change in consumption that may depend (also) on some other cause. De Santis and Maltagliati (2004) tried to apply these ideas by extending Ray's (1983) method, as follows

2) $$w_c = b_{c0} + b_{c1}(X/S_h) + \lambda_c L_h + \varepsilon_c \quad \left[\forall c; \sum b_0 = 1; \sum b_1 = \sum \lambda = 0\right]$$

where the term L_h (linked to the size of the household) was expected to capture this style effect, i.e. the variation in the consumption profile (for each consumption item c) that takes places beyond the impoverishment effect (incorporated in the S_h term).

In practice, however, both this model and all the preceding ones do not prove satisfactory: if one tries to apply them to several consecutive years (same country, same survey), results are unstable, more than "normal" year-to-year and sample variability would justify. Even worse: for any given dataset (same country, same survey, same year), for any given estimation method, the equivalence scale changes in a non trivial way by a simple re-aggregation of elementary consumption items into expenditure categories. In Italy, for instance, there are about 300 elementary consumption items in the Istat set of microdata on household expenditure (cf. section 4), but complete demand systems typically work with only 3 to 8 expenditure categories. And forming more or fewer categories - or, for any given number of categories, aggregating items in one way or another - affects estimation, and sometimes very deeply (De Santis and Seghieri, 2006).

Mono-equational estimation systems, too, are disturbed by apparently irrelevant changes in the treatment of elementary consumption items. Take Engel, for instance: originally, he worked on "necessities", which included food, apparel and housing. Currently, instead, Engel only means "food share", because the other categories do not ... "behave well". Apparel is now a luxury (its expenditure share increases with income - cf. table 1); housing is still a necessity, but its share decreases with household dimension (cf. De Santis and Maltagliati, 2004), whereas, in the case of food, the share decreases with income, but increases with household size.

In short, when scrutinized closely, almost all the work that has been done thus far on equivalence scales has eventually shown some weakness, on the theoretical or the empirical plan (or both). Besides, the research on this topic seems to have somewhat withered recently, and the words of OECD (2008b) reflect the discouragement and disenchanted view that, we think, is now prevailing: "In general, there is *no accepted method* for determining equivalence scales, and no equivalence scale is recommended by the OECD for general use" (italics in the original).

The time has probably come for researchers to try something new.

3. A different approach to equivalent scales estimation

3.1. An overview of our approach

The essentials of our approach to the estimation of equivalence scales are as follows. We form several subgroups (clusters) of households who, in our opinion, share a similar standard of living (and incidentally, also a similar life-style), at least with regard to how they spend their money. Within each cluster k, we consider households of different size (with 1, 2, 3, ... members), and calculate the average total monthly expenditure of each of them (X_{1k}, X_{2k}, X_{3k}, ...). Using X_{1k} as a standard of reference, we calculate the ratios $S_{nk}=X_{nk}/X_{1k}$. Since, by assumption, within clusters, the standard of living is (roughly) the same, these $S_{.k}$'s can be interpreted as cluster-specific equivalence scales (1, S_{2k}, S_{3k}, ...). By properly averaging over the various cluster-specific scales, we arrive at what we are looking for: the general equivalence scale (1, S_2, S_3, ...).

The details of the procedure are discussed in the following sub-sections.

3.2. Active variables

"Active" are the variables on the basis of which we form clusters. The criterion for their choice is empirical: they must be "well-behaved" indicators of economic wellbeing, or they must be "ancillary" variables that permit well-behaved variables to perform properly.

A well-behaved indicator of utility is a variable that evolves consistently with resources (in practice: with total expenditure), over the whole range of empirically relevant expenditure levels, and this separately for every household typology[6]. Take, for instance, w_f, the share of total outlays devoted to food: for every given household typology, it is always true that w_f decreases as total expenditure increases.[7] Therefore, according to our standards, w_f qualifies as a reliable indicator of economic wellbeing.

Now consider the share of "rich" proteins out of the total amount of proteins consumed (in terms of monetary expenditure), where the label "rich" refers to proteins that derive from fish or meat, excluding poultry.[8] Empirically, it works well (or, in our terminology, is "well-behaved"), because this share increases consistently with resources, for every household size. But what about the households who did not consume proteins at all in the observed period (one week - cf. Section 4)? How should we treat these "vegetarian" households? We solve the problem with the introduction of an ancillary, dummy variable (vegetarian? yes/no). This variable does not identify rich or poor households in itself, but it permits the previous variable (share of rich protein) to perform well, although only on a subset of households, the non vegetarians.

Table 1 lists the 14 active variables that we eventually retained for our analysis: they are either luxuries (if their share increases with total expenditure), necessities (if their share decreases with total expenditure),[9] or ancillary (if they form subgroups on which luxuries and necessities can be identified). Eventually, we retained 11 luxury, 1 necessity, and 2 ancillary variables.

---

[6] When the variable is a dummy ("does the household own this durable?"), it is the probability of owning a specific consumer durable that must evolve consistently with resources.

[7] This holds on average, of course. We regressed w_f=f{ln(X), [ln(X)]^2}, separately by household size, and we verified that f'<0 for every household size, and over the whole range of empirically relevant total expenditure.

[8] The others derive from poultry, milk and eggs. Needless to say, all our indicators, including this rather unconventional one, were chosen after a long series of attempts and failures.

[9] This distinction is helpful for presentation purposes only. In practice, every indicator can easily be reversed into its opposite. For instance, instead of the share of food (w_f, a necessity) we could have considered the share of non-food (1-w_f, a luxury), without affecting our results.

Table 1 - Active variables

Category	Label	#	Variable	Filters	Notes
Luxuries	%Lux	1	Share of luxuries on total expenses		(1)
	Ktchn	2	Kitchen as an independent room (Dummy: yes/no)		
	Garage	3	Garage (Dummy: yes/no)		
	Motor	4	Motorbike (Dummy: yes/no)		
	Washr	5	Washer (Dummy: yes/no)		
	Dishw	6	Dishwasher (Dummy: yes/no)		
	Hclnrs	7	Home-cleaning machines (Dummy: yes/no)		
	Cond	8	Air conditioning (Dummy: yes/no)		
	Tel	9	Telephone (Dummy: yes/no)		
	%Rchprt	10	Share of rich proteins on total proteins	var. 13	(2)
	%Olive	11	Share of olive oil out of total oil purchased	var. 14	(3)
Necessities	%Food	12	Share of food on total expenditure		
Ancillaries	Vgtrn	13	Vegetarian (Dummy: yes/no)		
	Oilcnsmr	14	Oil consumer (Dummy: yes/no)		

Notes
1) Includes a long list of expenditure items ranging from men's apparel to domestic help; from sportswear to pets; from private lessons to jewelry; etc.
2) This variable applies only to non vegetarian households (var. 13). "Rich" are the proteins that derive from the purchase of meat (except chicken) or fish. Total protein expenditure includes "rich" protein, plus chicken, milk and eggs.
3) This variable applies only to oil-consuming households (var. 14). Beside olive oil, households can also purchase the less valuable seed oil and olive pomace oil.

Note that total expenditure is *not* an active variable: we do not form clusters on the basis of this information. It is only *ex post* that we calculate the average monthly expenditure for each household typology within each cluster (X_{kn}) and, from here, also cluster-specific equivalence scales ($S_{kn}=X_{kn}/X_{kr}$).

Within each cluster, households are homogeneous in terms of standards of living: this derives from the very choice of our active variables, which are essentially well-behaved indicators of utility. However, we contend that, within each cluster, households are homogeneous also in terms of style. Let us see this with an example. Among our clusters, we frequently find situations as those depicted in Table 2. The 2-person households in these two clusters (Alpha and Beta) have virtually the same average monthly expenditure, about 1900 Euros per month: although we did not form cluster on the basis of total expenditure, we conclude, ex post, that that their standard of living is basically the same. If we now look more closely at our active variables, we notice that the two clusters do show some traits in common (e.g. they both have a dishwasher but no motorbike, they spend about 13-14% of their monthly budget in luxuries, they are not vegetarians and they spend about 65% of their protein outlays in "rich proteins", i.e. meat and fish), but they also differ in several other respects: Alpha households have kitchen as a separate room, have fixed telephone connections, and are oil consumers. Their food share is considerably higher (23% as against 15%), they more frequently own a garage (100% as against 59% in the other group) but they do not have a dishwasher, or home-cleaning machines, or air conditioning, which are instead occasionally or even frequently found in the other group.

In short the two groups, virtually undistinguishable in terms of standards of living (same structure, same total expenditure), differ in terms of their consumption profile, which is what we call style.

Table 2 - Average values (with standard deviations) of 15 variables for 2-person households in two selected clusters (Italy, 2003-2008)

			Cluster alfa			Cluster beta	
#	Label	Variable	Average	Std.err.		Average	Std.err.
1	%Lux	Share of expenses for luxury items	0.1321	0.003		0.1403	0.0047
2	Ktchn	Kitchen as an independent room (Dummy: yes/no)	1	0		0	0
3	Garage	Garage (Dummy: yes/no)	1	0		0.5889	0.0245
4	Motor	Motorbike (Dummy: yes/no)	0	0		0	0
5	Washr	Washer (Dummy: yes/no)	1	0		1	0
6	Dishw	Dishwasher (Dummy: yes/no)	0	0		0.3941	0.0243
7	Hclnrs	Home-cleaning machines (Dummy: yes/no)	0	0		0.937	0.0121
8	Cond	Air conditioning (Dummy: yes/no)	0	0		0.2299	0.0196
9	Tel	Telephone (Dummy: yes/no)	1	0		0	0
10	%Rchprt	Share of rich proteins on total proteins	0.6519	0.0069		0.6472	0.0149
11	%Olive	Share of olive oil out of total oil purchased	0.9380	0.0044		0	0
12	%Food	Share of food on total expenditure	0.2338	0.003		0.1462	0.0043
13	Vgtrn	Vegetarian (Dummy: yes/no)	0	0		0	0
14	Oilcnsmr	Oil consumer (Dummy: yes/no)	1	0		0	0
	X	*Total expenditure*	*1 845.8*	*32.237*		*1 901.0*	48.173
		No. obs	*838*			*409*	

Source: own elaborations on Istat consumption microdata. Example taken from a Ward clustering, with 50 clusters.

There may be several reasons, that need not concern us here, why households with the same resources and the same structure decide to spend their money differently: personal taste, tradition, religion, weather conditions, local constraints (but see also section 3.3), etc. What is important is that, within cluster Alpha, all households typologies, with 1 to 5 members behave similarly (table 3, upper part); and the same happens within cluster Beta (table 3, lower part), and in all other clusters (not shown here). In short, this procedure permits us to compare like with like: we only make comparisons within (homogeneous) clusters.

Table 3 - Average values (with standard deviations) of 15 variables in two selected clusters

	Cluster Alfa	1 person		2 persons		3 persons		4 persons		5 persons	
#	Label	Avrg	Std.err.	Avrg	Std.err.	Avrg	Std.err.	Avrg	Std.err.	Avrg	Std.err.
1	%Lux	0.130	0.004	0.132	0.003	0.149	0.004	0.152	0.006	0.144	0.007
2	Ktchn	1	0	1	0	1	0	1	0	1	0
3	Garage	1	0	1	0	1	0	1	0	1	0
4	Motor	0	0	0	0	0	0	0	0	0	0
5	Washr	1	0	1	0	1	0	1	0	1	0
6	Dishw	0	0	0	0	0	0	0	0	0	0
7	Hclnrs	0	0	0	0	0	0	0	0	0	0
8	Cond	0	0	0	0	0	0	0	0	0	0
9	Tel	1	0	1	0	1	0	1	0	1	0
10	%Rchprt	0.626	0.010	0.652	0.007	0.683	0.010	0.664	0.010	0.660	0.020
11	%Olive	0.957	0.004	0.938	0.004	0.924	0.006	0.908	0.008	0.897	0.014
12	%Food	0.215	0.004	0.234	0.003	0.215	0.004	0.227	0.004	0.244	0.008
13	Vgtrn	0	0	0	0	0	0	0	0	0	0
14	Oilcnsmr	0	0	0	0	0	0	0	0	0	0
	X	*1 501.9*	*31.5*	*1 845.8*	*32.2*	*2 307.3*	*52.5*	*2 464.2*	*71.4*	*2 744.1*	*129.0*
	Obs.	*590*		*838*		*419*		*323*		*101*	
	Equiv. scal	***1***		***1.229***		***1.536***		***1.641***		***1.827***	
	Cluster Beta	1 person		2 persons		3 persons		4 persons		5 persons	
#	Label	Avrg	Std.err.	Avrg	Std.err.	Avrg	Std.err.	Avrg	Std.err.	Avrg	Std.err.
1	%Lux	0.132	0.004	0.140	0.005	0.146	0.005	0.158	0.006	0.158	0.013
2	Ktchn	0	0	0	0	0	0	0	0	0	0
3	Garage	0.442	0.019	0.589	0.024	0.505	0.030	0.511	0.033	0.527	0.070
4	Motor	0	0	0	0	0	0	0	0	0	0
5	Washr	1	0	1	0	1	0	1	0	1	0
6	Dishw	0.262	0.017	0.394	0.024	0.312	0.028	0.358	0.032	0.269	0.066
7	Hclnrs	0.924	0.010	0.937	0.012	0.943	0.013	0.949	0.015	0.767	0.054
8	Cond	0.173	0.014	0.230	0.020	0.328	0.027	0.286	0.030	0.226	0.063
9	Tel	0	0	0	0	0	0	0	0	0	0
10	%Rchprt	0.579	0.013	0.647	0.015	0.609	0.016	0.601	0.017	0.599	0.033
11	%Olive	0	0	0	0	0	0	0	0	0	0
12	%Food	0.140	0.003	0.146	0.004	0.160	0.005	0.174	0.005	0.186	0.010
13	Vgtrn	0	0	0	0	0	0	0	0	0	0
14	Oilcnsmr	1	0	1	0	1	0	1	0	1	0
	X	*1 569.2*	*36.0*	*1 901.0*	*48.2*	*1 994.4*	*60.6*	*2 091.3*	*64.9*	*1 903.2*	*105.4*
	Obs.	*680*		*409*		*286*		*226*		*50*	
	Equiv. scal	***1***		***1.211***		***1.271***		***1.333***		***1.213***	

Source: own elaborations on Istat consumption data. Cf. table 2.

Of course, one may always suspect that, with a different set of active variables, results would have differed. The answer is "yes", in terms of the clusters that are formed (they are no longer the same), and of their characteristics, but it is "no" (or, at least "not much") in terms of the resulting equivalence scale. Indeed, we replicated our analysis with half our active variables, selected at random, and the resulting equivalence scale did not change appreciably (results not shown here)

Admittedly, this does not prove, but at least it does suggest that the procedure is robust, as long as our active variables are numerous enough, and not too strictly correlated to each other (cf. table 4).

Table 4 - Correlation matrix for the 14 active variables retained

	Name	%Lux	Ktchn	Garage	Motor	Washr	Dishw	Hclnrs	Cond	Tel	%Rchprt	%Olive	%Food	Vgtrn	Oilcnsmr
MEAN		0.1550	0.8282	0.5890	0.0708	0.9727	0.3921	0.7821	0.2160	0.7870	0.6356	0.4700	0.1782	0.0171	0.4321
STD		0.1086	0.3772	0.4920	0.2565	0.1630	0.4882	0.4128	0.4115	0.4094	0.2556	0.4787	0.0899	0.1297	0.4954
1	%Lux	1	0.0064	0.0928	0.0663	0.0228	0.1075	0.0758	0.0863	0.0996	-0.0086	-0.0504	-0.2791	0.0257	0.0523
2	Ktchn	0.0064	1	0.0682	0.0015	0.0634	0.0629	0.0811	0.0295	0.1121	0.0264	0.0092	-0.0471	-0.0179	-0.0079
3	Garage	0.0928	0.0682	1	0.0921	0.0896	0.2572	0.2620	0.1265	0.2081	0.0172	-0.0238	-0.2525	0.0154	0.0224
4	Motor	0.0663	0.0015	0.0921	1	0.0212	0.1117	0.0881	0.0723	0.0351	0.0055	-0.0162	-0.1160	0.0159	0.0195
5	Washr	0.0228	0.0634	0.0896	0.0212	1	0.0954	0.1983	0.0574	0.1315	0.0479	0.0121	-0.0522	-0.0436	-0.0164
6	Dishw	0.1075	0.0629	0.2572	0.1117	0.0954	1	0.3202	0.2168	0.1526	0.0409	-0.0385	-0.2830	0.0094	0.0543
7	Hclnrs	0.0758	0.0811	0.2620	0.0881	0.1983	0.3202	1	0.1705	0.1605	0.0328	-0.0191	-0.2959	0.0102	0.0274
8	Cond	0.0863	0.0295	0.1265	0.0723	0.0574	0.2168	0.1705	1	0.0614	0.0381	-0.0006	-0.1460	0.0064	0.0088
9	Tel	0.0996	0.1121	0.2081	0.0351	0.1315	0.1526	0.1605	0.0614	1	0.0530	-0.0010	-0.1123	-0.0225	0.0003
10	%Rchprt	-0.0086	0.0264	0.0172	0.0055	0.0479	0.0409	0.0328	0.0381	0.0530	1	0.1150	0.1391	-0.3281	-0.1196
11	%Olive	-0.0504	0.0092	-0.0238	-0.0162	0.0121	-0.0385	-0.0191	-0.0006	-0.0010	0.1150	1	0.2321	-0.1027	-0.8564
12	%Food	-0.2791	-0.0471	-0.2525	-0.1160	-0.0522	-0.2830	-0.2959	-0.1460	-0.1123	0.1391	0.2321	1	-0.1770	-0.2566
13	Vgtrn	0.0257	-0.0179	0.0154	0.0159	-0.0436	0.0094	0.0102	0.0064	-0.0225	-0.3281	-0.1027	-0.1770	1	0.1171
14	Oilcnsmr	0.0523	-0.0079	0.0224	0.0195	-0.0164	0.0543	0.0274	0.0088	0.0003	-0.1196	-0.8564	-0.2566	0.1171	1

Source: own elaborations on Istat consumption data. Cf. table 2

3.3. Stratification

Within clusters, we will also distinguish by geographic area, occasionally referring separately to 5 macro-regions: North-West, North-East, Center, South, Islands. Price levels differ in the various parts of Italy (Istat, 2010a) and, if the cost of living is associated with household dimension, in either way, this will bias our estimates of the equivalence scale. Imagine, for instance (as it is indeed the case - see further in the text) that in the South of Italy households are larger and the cost of living is lower. In a geographically uncontrolled estimate, large households will spend comparatively little to reach the same standard of living of the reference household, merely because they come prevalently from the South, where prices are lower: in short, the equivalence scale will appear flatter than it actually is (a downward bias). We will show shortly that the regional dimension is not to be overlooked in this type of study - at least, not in Italy. Incidentally, our procedure also permits us to arrive at a very original estimate of regional purchasing power parity, with a methodology that, at least in theory, could also be applied cross-nationally.

3.4. Other details of the clustering procedure

a. Criterion for clustering and number of clusters

There are several ways of forming clusters, even within the hierarchical, agglomerative ("bottom-up"), Euclidean-distance procedure that we decided to follow: Ward, average distance, centroid,...

(Bouroche and Saporta, 1980). We tried all of them (with SAS), and we found out that using one or another does not make any difference with a very large number of clusters (500 or more), because in this case all the methods converge to (basically) the same results (see Figure 3). With fewer clusters, however, the question of which method to choose arises. What we found out is that the Ward method is normally the best, for three main reasons. In the first place, for any given number of clusters, its results are almost always closer to the asymptotic value than those of any other method. Secondly, at least with the SAS statistical package we used here, Ward is one of the few methods that accepts an unlimited number of observations, while most other methods are limited (in our case, to 65 thousand units). Finally, Ward creates clusters of comparable weight (i.e. with roughly the same number of observations per cluster), which proves useful for some of our tests.

The best number of clusters, as mentioned, is rather high - although, of course this also depends on the number of observations. In our case, with slightly more than 128 thousand households, we find convergence with all methods with about 500 clusters. However, what we judged to be the best method (Ward) yields very satisfactory results already with 100 or even 50 clusters. With fewer clusters (or with other methods), instead, results may be biased.

b. From cluster-specific to the general equivalence scale

Our method yields an equivalence scale for each cluster. Therefore, if we work with, say, 100 clusters, we end up with 100 different equivalence scales, which we need to synthesize in one general equivalence scale. How? Consider again table 3, for instance, and imagine that we only work with two clusters (Alfa and Beta), and that we are only interested in calculating the equivalence factor S_3, starting from $S_{3\alpha}$=1.5363 and $S_{3\beta}$=1.2709. Both S_3 are ratios (=X_3/X_1), and both depend on the number of observations within each cluster, i.e. on the frequency of 3-person households (in the numerator) and of 1-person households (in the denominator). Of course, we attach more importance to information based on a greater number of observations, but the balance between numerator and denominator is important too. For instance, with 100 households in a cluster, we prefer to have 50 in the numerator (3-person households) and 50 in the denominator (1-person households) than, say, 90 and 10 respectively. This is why we eventually decided to weight each cluster (and each equivalence factor) with the geometric mean of the number of households in the numerator and in the denominator. The general equivalence scale S_n is the weighted average of cluster-specific equivalence scales.

3) $$S_n = \sum_k S_{n,k}\, w_k \quad (w_k = \text{weight}; \Sigma w_k = 1)$$

c. Chaining datasets

As discussed in Section 4, we work with household-level consumption data that derive from 6 different (independent, cross-sectional) collections, for the years 2003 to 2008. We decided to append these six dataset one after the other, so as to form only one, with more observations and therefore also more robust estimates. However, for reasons that will be explained shortly, we followed two different paths, and thus created two different databases, labeled respectively 1- and 6-period database.

The dataset that we will generally use is the 1-period database, which is obtained after inflating all the expenditures of the years 2003 to 2007, so that all the monetary variables are expressed in Euros of 2008. The inflating factors are those of the official consumer price index (table 5).

Table 5 - Inflation in Italy, 2003-2008

	Base	
	preceding year	2003
2003	--	1
2004	1.022	1.022
2005	1.019	1.042
2006	1.021	1.064
2007	1.018	1.084
2008	1.033	1.120

Source: Istat

The 6-period database too contains the 6 original databases, but this time without inflating the monetary variables: each year is thus specific and expenditure values of different years must be treated as if they were expressed in different currencies. We will use this year-specific dataset to get to an alternative estimate of inflation in the observed period (see Table 8).

e. Confidence intervals

The equivalence scales that we estimate are random variables, the variance of which can be estimated, too. Remember that, within each cluster, we obtain our equivalence scale $S_{n,k}$ for households with *n* members as a ratio between a numerator ($X_{n,k}$ - the average monthly expenditure of households of that type) and a denominator ($X_{1,k}$ - the average monthly expenditure of the reference households). Both $X_{n,k}$ and $X_{1,k}$ are sample estimates of averages, which are produced with a certain degree of uncertainty, synthesized by their variance [$Var\ (X_{n,k})$, $Var\ (X_{1,k})$]. These four pieces of information, together with the estimated value of the ratio $S_{n,k} = \frac{X_{n,k}}{X_{1,k}}$, permit us to estimate the variance of each cluster-specific equivalence scale, as follows (Kish, 1965[1995], p. 187)

4) $$Var\left(S_{n,k}\right)=\frac{1}{\left(X_{1,k}\right)^2}\left\{Var\left(X_{n,k}\right)+S_{n,k}^2\left[Var\left(X_{1,k}\right)\right]\right\}$$

Since we estimate our equivalence scale S_n as separate ratio estimate $S_n=\sum_k S_{n,k}\ w_k$ (eq. 3), the variance of S_n becomes (Kish, 1965[1995], p. 206)

5) $$Var\left(S_n\right)=\sum_k w_{n,k}^2\cdot Var\left(S_{n,k}\right)$$

4. Data source: micro data on household consumption in Italy

Our data are those of the yearly consumption survey conducted, at the household level, by the National Institute of Statistics.[10] Of the original 148 thousand odds households, we eventually retained only slightly more than 128 thousand, after exclusion of the large ones (6 or more members) and of those that we considered outliers.[11]

Of these households we know the composition, with a few details about each member: age, sex, education, employment status (yes/no), and proportion of household income secured by that specific member. Then, we know how much each household spends monthly on about 300 elementary consumption items (X_i) and how much it spends overall ($X=\Sigma X_i$). We ignore income: total expenditure X will therefore be our indicator of the resources available to the household.

Finally, but importantly: we also know if the household owns a few durables (e.g. color TV set; car, motorbike; etc) and, as explained in section 3.2, we used some of these as active variables, for clustering.

The sample is not auto-representative. All the observations come with a weight, which we take into account in our computations of average values.

[10] For more information, see http://en.istat.it/dati/microdati/ and http://en.istat.it/salastampa/comunicati/in_calendario/consfam/20100705_00/.

[11] We first formed 16 groups: by household dimensions (4 classes: the last being 4+ members) and by levels of affluence (4 classes, or quartiles, within each household dimension). Then for each group we considered 13 expenditure shares (food, beverages, housing, energy, education, transportation, leisure, health, tobacco, furniture, apparel, communication, other - summing up to 1), and for each we computed the median M and the inter-quartile interval Q. Then we excluded from the observation the households that, for any of these 13 expenditure shares, fell outside the interval ranging from (M-4Q) to (M+4Q). We verified, however, that ignoring these outliers, the equivalence scales do not change appreciable (results not shown here).

5. Results

5.1. An alternative (perhaps too flat?) equivalence scale

The most important result of our paper is the equivalence scale for Italy in the period 2003-2008: Figure 1 and table 6 show the scale that we find most convincing, together with its confidence interval. The most surprising feature of our equivalence scale is that it is flat: flatter, for instance, than Carbonaro's one, a semiofficial equivalence scale used by Istat (2010b) and, with adjustments, also by the government, in its anti-poverty policy.[12] This was expected, because Carbonaro's method is a variant of Engel's one, which is generally suspected to lead to overestimation. However, our scale is also lower than the OECD (2008b) square-root one, which, on the contrary, is generally believed to be somewhat underestimated.

Figure 1 - Equivalence scales in Italy (2003-2008): Carbonaro's method, square-root of members (OECD) and our own (Ward, 100 clusters, with 5 regional strata)

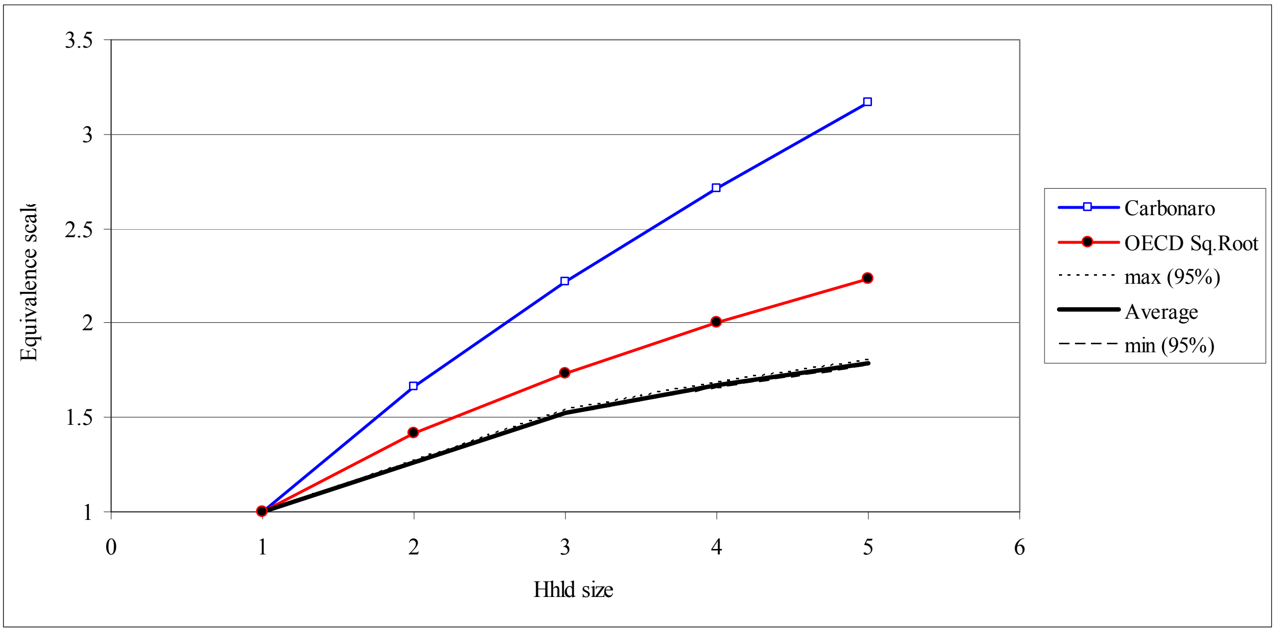


Table 6 - Estimated equivalence scale (S) and confidence intervals (95%; Italy, 2003-2008)

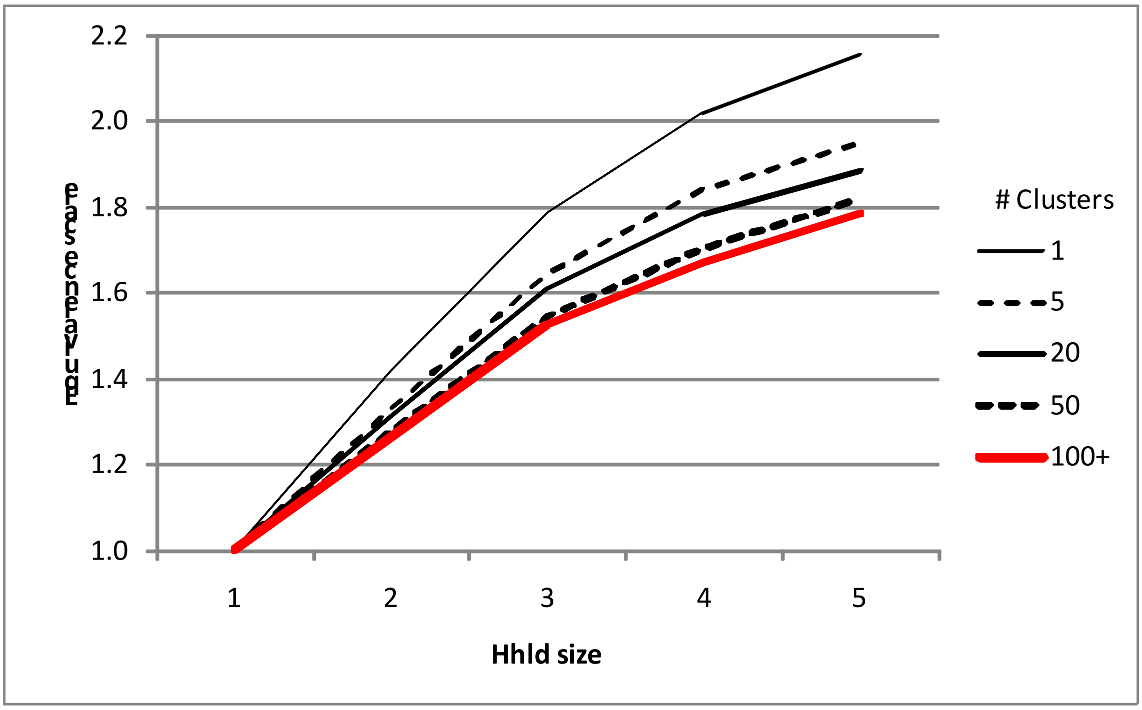
Hhld	S with 100 Clusters (Ward method)			Other, selected S's	
Size	Average	min (95%)	max (95%)	Carbonaro	OECD Sq.Root
1	1	1	1	1	1
2	1.263	1.253	1.272	1.667	1.414
3	1.526	1.515	1.536	2.222	1.732
4	1.672	1.658	1.686	2.716	2.000
5	1.786	1.770	1.803	3.169	2.236

Source: Authors' elaborations on Istat consumption data.

[12] This is the so-called ISEE scale: see, e.g., http://it.wikipedia.org/wiki/ISEE.

The 95% confidence interval around our scale is very narrow, because of the very large number of observations and, also, because of the remarkable homogeneity within each cluster. This is good news, of course, but its importance should not be overestimated, because non sampling errors and biases are probably more important than others, in this case. For instance, our estimates depend negatively on the number of clusters we form, at least up to 100 (Figure 2). With more than 100 clusters and with the Ward clustering method, changes in the equivalence scale are irrelevant (see also figure 3). The case with just 1 cluster simply measures the average monthly expenditure of households with *n* members - or, better, the ratio of this average expenditures to the standard case, the average monthly expenditure of 1-person households.

Figure 2 - Equivalence scales in Italy (2003-2008), by number of clusters (Ward, 5 regions)



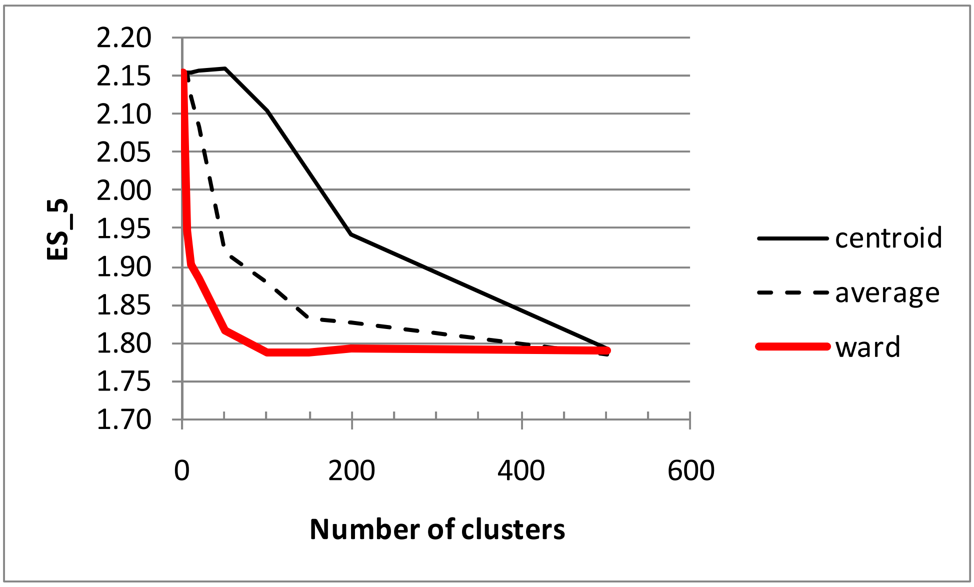
Source: Authors' elaborations on Istat consumption data.

The clustering method matters - at least with relatively few clusters. Figure 3, for instance, shows the different estimates that we obtain for equivalence factors relative to 5-person households[13]). With a very large number of clusters (here: 500 or more), the values converge towards what we interpret as the "true" equivalence scale: in this case, the clustering method is irrelevant. But, since it is unpractical to work with a very large number of clusters, it is also worthwhile to note that the

[13] The same convergence, towards smaller values of course, can also be observed for other equivalence factors (smaller household dimensions; not shown here).

Ward method is the one that converges most quickly: already with 50 (or, even better, 100) clusters, the values it yields are basically the same as those obtained with more, or even many more clusters.

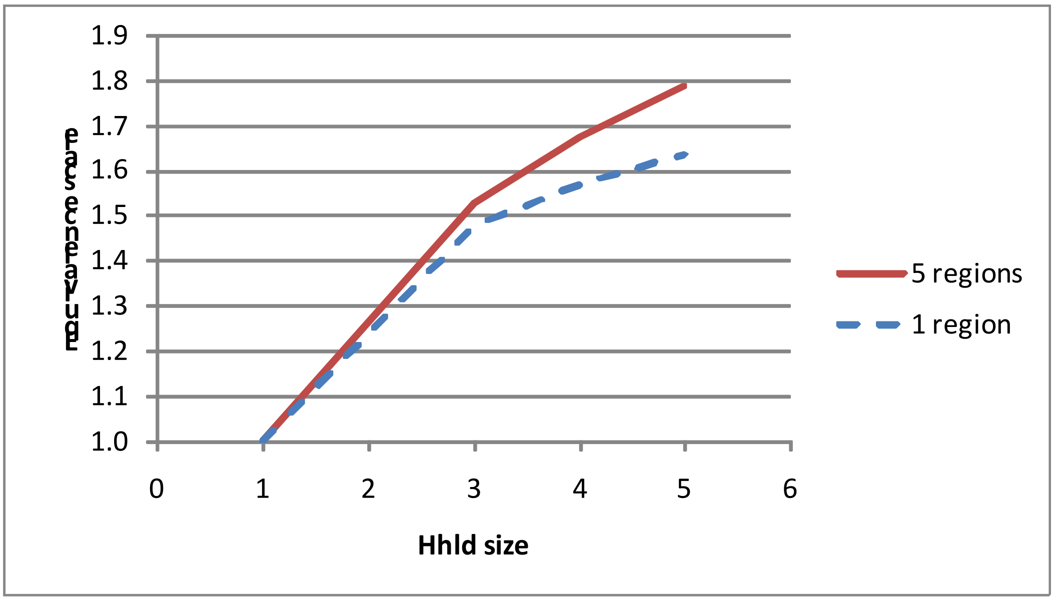
Figure 3 - Equivalence factors for a 5-person household (ES_5) in Italy (2003-2008), by number of clusters, estimated with three clustering methods: Ward, average link, centroid (5 regions)



Source: Authors' elaborations on Istat consumption data.

Finally, our estimates are also sensitive to if and how we form regional strata: ignoring regional differences systematically yields a lower equivalence scale. Figure 4 shows this for one case only (Ward, 100 clusters), but the phenomenon is always there. The reason for this lies, we believe, in the association between household size and the cost of living: households are larger in the South, where prices are lower (see further in the text), and estimating separately 5 regional equivalence scales and merging them together corrects for this bias. But, of course, there may be other biases of which we are unaware.

Figure 4 - Equivalence scale with and without consideration of regional differences in Italy (2003-2008) (Ward method, 100 clusters, 1 or 5 regions)



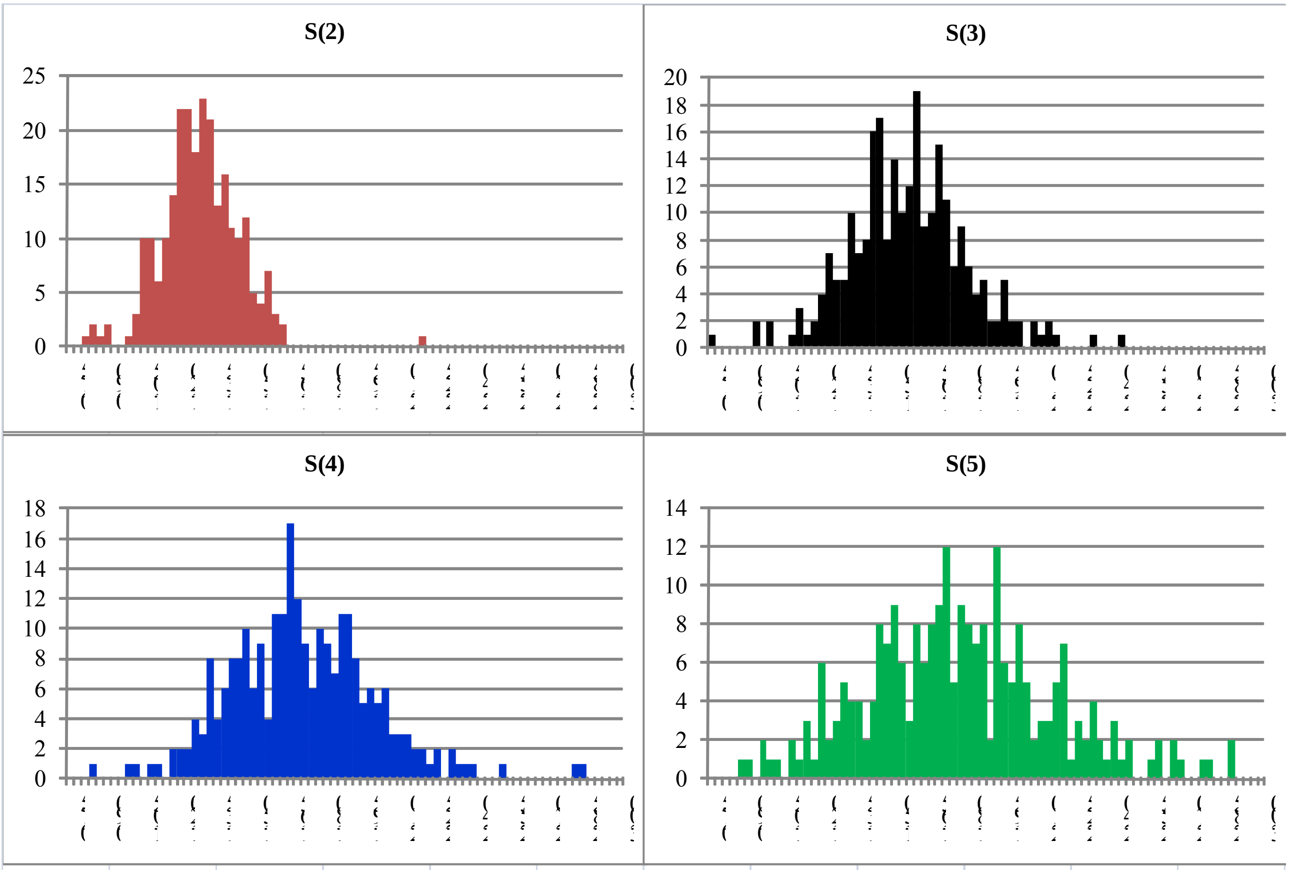
Source: Authors' elaborations on Istat consumption data.

5.2. A few tests on the results

A new method, even if theoretically convincing, needs to pass several empirical tests before others may find it worthy of consideration. Let us consider a few of them, starting with the one that our method has already passed: our results are plausible. Remember that this is a non parametric method, and that in principle *any* result could have emerged, for instance diseconomies of scale, or negative marginal costs. Instead, regardless of the clustering criterion and of the number of clusters, our equivalence factors evolve smoothly, and economies of scales emerge as expected (i.e., marginal costs are positive and decrease with n), even if, admittedly, their effect is stronger than generally believed.

The second test is whether cluster-specific equivalence scales are (at least roughly) consistent between cluster. Figure 5 provides a first answer, which we find only partially satisfactory. Although the bars of the histogram do not reveal that the outliers weigh considerably less than more centered values, dispersion is high, especially for ES_4 and ES_5. This depends also on the high number of comparisons that we make: with fewer clusters or fewer regions, of course, variability is considerably lower, but then heterogeneity increases, which may result in other types of bias.

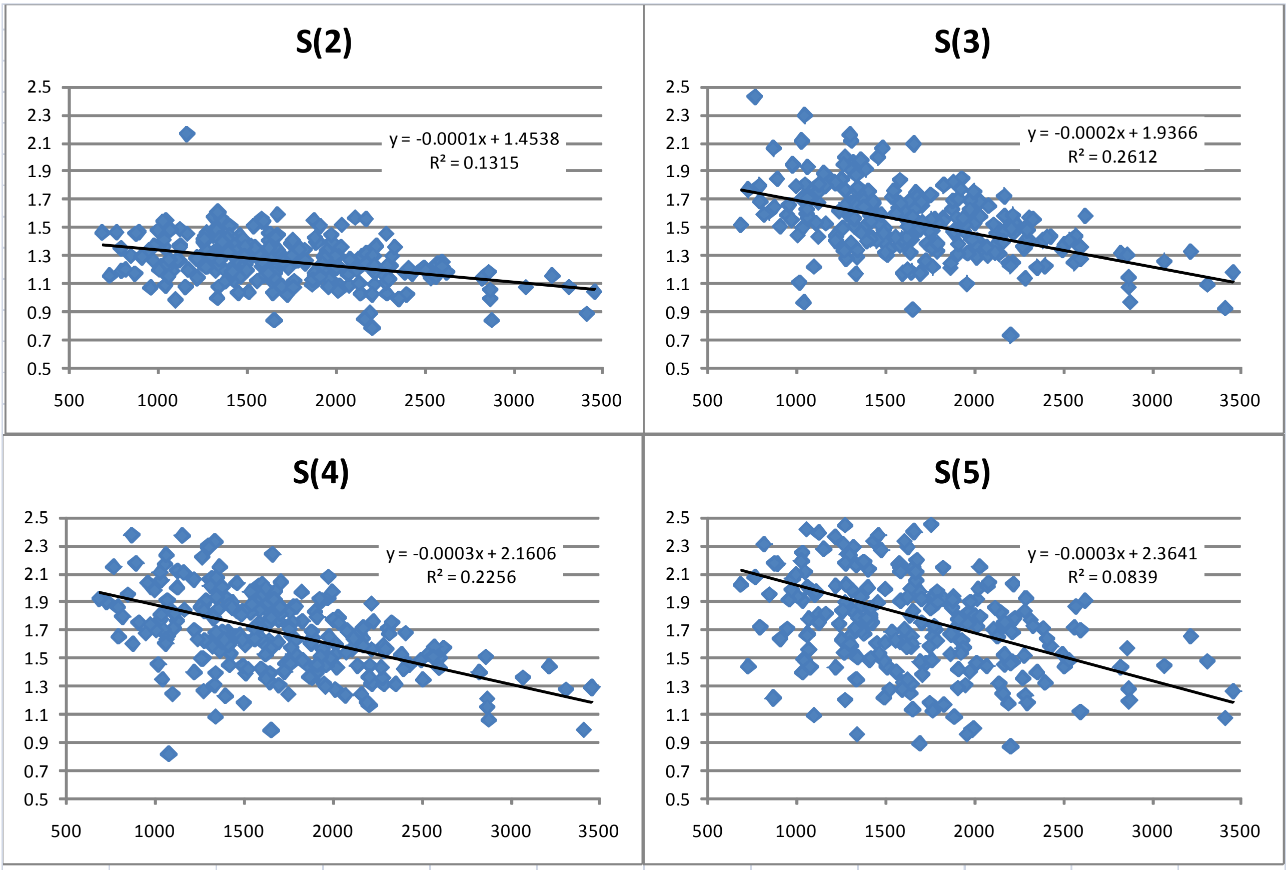
Figure 5 - Cluster specific equivalence scales (S_n) in Italy (2003-2008)



Source: Authors' elaborations on Istat consumption data (Ward, 250 cases: 50 clusters and 5 regions)

The third test qualifies this answer. Do our cluster-specific equivalence scales depend on variables whose influence is (at least, qualitatively) known? Utility is a candidate: the IB (independence of base) assumption is very useful for practical reasons, but when put to test, it invariably fails: equivalence scales are typically flatter for the rich and steeper for the poor. And this is reasonable: if some costs are more or less fixed (think of school fees, for instance), they affect more low than high family budgets. Figure 6 and table 7 show that, within clusters, equivalence factors are on average flatter for the rich than for the poor - and the impact of income is not trivial. If the standard of reference is a 1-person household spending 1000 Euros per month (therefore, relatively poor), the scale is steep: almost as much as the OECD square-root one. Conversely, for well-off, and even more so, for rich one-person households (spending 2000 and 3000 Euros/month, respectively), the equivalence scale rises only moderately with the number of members.

Figure 6 - Equivalence factors (S_n) for different utility levels in Italy (2003-2008)



Source: Authors' elaborations on Istat consumption data.

Table 7 - Estimated equivalence scales for three illustrative utility levels (Italy, 2003-2008)

	Three expenditure levels for 1-person hhlds		
	1 000	2 000	3 000
n	Equivalence scales		
1	1	1	1
2	1.341	1.229	1.117
3	1.698	1.460	1.221
4	1.878	1.596	1.314
5	2.023	1.682	1.342

Source: Authors' elaborations on Istat consumption data. Equivalence scales estimated using the regression parameters of Figure 6.

Inflation in Italy is monitored constantly, with the same (Laspeyres-based) methodology used in the other EU-27 countries, so as to obtain comparable results (cf. table 5). Our methodology, too, may be used to estimate inflation, but with a different approach and, what is more remarkable, without knowing prices! Let us now work with the 6-period database, where each year differs from each other, because prices are not the same, because of inflation. Let us form clusters, as before, and, within each cluster k, let us now consider $X_{krn(t)}$, that is the average expenditure, in year t, of households with n members in region r. Let us then calculate the ratio $I_{krn(t)} = \frac{\overline{X}_{krn(t+\lambda)}}{\overline{X}_{krn(t)}}$, which is an index number, revealing how much households need to spend in year t+λ to be as well off as in year t (λ=1, 2, ... 5). Remember that, within clusters, households are assumed to be at the same level of utility, and that they are assumed to be rational, that is to spend as little as possible (given prices, household characteristics and "style") to attain that level. In short, $I_{krn(t)}$ is an index number, (roughly) measuring inflation from year t to year t+λ. As usual, we obtain several such estimates - krn, with k clusters, r regions and n household typologies (dimensions). And their synthesis in a unique measure of inflation is not straightforward: official estimates of inflation, for instance, based on Laspeyres's method, implicitly weigh rich households more than poor ones. If we use our standard weights, on the contrary, we simply count the number of households. Table 8 shows that, despite the adoption of a totally different methodology, the results are strictly comparable: with respect to Istat, we slightly overestimate inflation in 2004 (4.2% instead of 2.2%), and slightly underestimate it in 2008 (1.7% as against 3.3%). But in the other years and more broadly speaking, over the whole period considered (2003-2008), our results are virtually undistinguishable from Istat's. Except that we do not need ad hoc (very large, very expensive) surveys: we obtain it simply as a by-product of our elaborations.

Table 8 - Estimated inflation in Italy (2003-2008)

	Base 2003=1			Base year (t-1)=1	
	Own elab.	Istat		Own elab.	Istat
2003	1	1		--	--
2004	1.042	1.022		1.042	1.022
2005	1.061	1.042		1.018	1.019
2006	1.084	1.064		1.022	1.021
2007	1.104	1.084		1.018	1.018
2008	1.122	1.120		1.017	1.033

Notes: 1250 cells (50 Clusters, 5 regions. 5 household dimensions); Clustering method: Ward.
Source: own elaborations on Istat consumption data, and Istat's official data on inflation.

5.3.Regional Purchasing Power Parity (and implications on the Equivalence scale)

The cost of living is generally believed to be lower in the South than in the rest of Italy. Until recently, however, to the best of our knowledge, this regional gap had never been reliably measured. In 2009 Istat did make an extra effort to compare price levels in the most important cities of Italy, and published its results the following year (Istat, 2010a), but these do not apply to the whole of Italy: only to selected towns.

With our methodology, instead, we can once again estimate what we are interested in (differences in regional price levels) - and without knowing prices! The procedure is the same one adopted at the end of section 5.2 - except that we now work once again with the 1-period dataset, where all the monetary values (expenditures) are expressed in Euros of 2008. We form clusters, as before, and, within each cluster *k*, we now consider $\overline{X}_{kn(r>1)}$, that is the average expenditure, in region *r*, of households with *n* members. We then calculate the ratio $I_{kn(r>1)} = \frac{\overline{X}_{kn(r>1)}}{\underline{X}_{kn(r=1)}}$, which is an index number, revealing how much households (of type *n*, in cluster *k*) need to spend in region *r*=2, 3, 4, 5 (North-East, Center, South, Islands) to be as well off as in region *r*=1 (North-West). Since, by assumption, within clusters, households are all at the same level of utility, $I_{kn(r)}$ is an index number, (roughly) measuring the difference in the cost of living between regions. Once again, we obtain *kn* such estimates - with *k* clusters and *n* household typologies (dimensions), and, once again, their synthesis in a unique measure of PPP (purchasing power parity) is not obvious. If we weigh our results on the basis of households frequencies, we get the results of table 8.

Table 8 - Purchasing Power Parities for 5 Italian macro-regions (Italy, 2003-2008)

NW	NE	Center	South	Islands
1	1.001	0.989	0.829	0.779

Note: Ward method, 500 clusters. Source: own elaborations on Istat consumption data.

In short, we estimate that the cost of living is basically the same in the Center-North of Italy, but it is about 20% lower in the Mezzogiorno (South, plus the islands of Sicily and Sardinia).[14]

7. Discussion

The key variable of our approach are the $\overline{X}_{krnt}$, i.e. the average expenditure levels $\overline{X}$ in cluster *k*, in region *r*, for household dimension *n*, in year *t*. Starting from these, and *assuming* that, within clusters, the standard of living is the same, by simply changing one of the indexes (e.g. another region *r'*, or another household dimension *n'*, or another year *t'*) we can calculate cluster-specific index numbers that can be interpreted, in turn, as measures of purchasing power parity, as equivalence factors, or as measures of inflation. The label changes, but the underlying idea is

[14] The estimates reported in Istat (2010c) are not strictly comparable, because they only refer to cities, and there is no summary measure for macro-areas (like North-West, Centre, and the like). In general, however, our results show more regional variations in prices than Istat does.

always the same: how much (more) is necessary to reach the same standard of living of the reference case, if one of the elements changes?

The crucial assumption of our approach is that the standard of living is (roughly) the same within clusters. This, of course, depends on the variables used for clustering: we selected 14 that passed a few formal tests, and we verified that using only a subset of these did not change our results appreciably. It is true, however, that the choice of these active variables is in good part arbitrary. The validity of the assumption depends also on the number of years considered: if technology, or tastes, or relative prices change over time, the meaning of certain items may change too, and some of them, that could once have been considered good indicators of the standard of living, may lose this property over time. This is probably the case of cars, or color TV sets, which, as readers may have remarked, are *not* among our clustering variables, because they do not evolve consistently with outlays, for any given household typology.

Another thorny issue concerns the number of clusters and the clustering method: once again, we have little theoretical support for our findings, which are mainly empirical and which, not surprisingly are also affected by the very large number of observations that we have (over 128 thousand). What should one do with smaller datasets? It is also worth noticing that some of our indicators of affluence are consumer durables that may reflect past more than current levels of economic well-being. Finally, cluster variably seems to be very high (especially when clusters are many), and this requires, at the very least, that the database be large enough so as to arrive at robust estimates.

But at this stage we would like to emphasize especially the merits of our approach: it is original, it can be applied to different types of datasets[15], and, perhaps most importantly, it permits to overcome at least in part the theoretical reservations that micro-economist have against equivalence scales. In our case, the indicators of economic well-being are not chosen a priori: they are based on actual household behavior, and they must evolve consistently with resources *for any given household typology*.

Our method is non parametric: while this is in part an advantage, in that it proves flexible (and this is how, for instance, we could analyze the influence of income on equivalence scales, that seem to get flatter for richer households) it may also make some analyses more difficult. For instance, investigating the impact of the age of household members, of their sex, and of their employment or health status, even if the information were available, would, at this stage require that many more

[15] E.g. Bankitalia's SHIW (Survey of Household Income and Wealth); LIS (Luxembourg Income Study); ECHP (European Consumer Household Panel); EU-SILC (Statistics on Income and Living Conditions); etc. Actually, we did try to apply our method to Bankitalia's SHIW and our preliminary results (not shown here) are very encouraging.

cells be formed: too many, given the size of most of the currently available databases, and therefore full of zeros. But Figure 6 and Table 7 also highlight a possible solution: non parametric analysis first, on the basis of which a few parameters can be estimated, thus permitting a more parsimonious use of the available information in subsequent applications.

Acknowledgments

We acknowledge financial support from the MIUR (Italian Ministry of Education, University and Research). We thank Andrea Giommi (Un. of Florence) for his advice on the estimation of the variance of equivalence scales.

Bibliography

Aassve, A., Betti, G., Mazzuco, S. and Mencarini, L. (2007) Marital disruption and economic well-being: a comparative analysis. J. R. Statist. Soc. A, **170**, 781–799.

Barten, A.P. (1964) Family Composition, Prices and Expenditure Pattern, in Econometric Analysis for National Economic Planning. Proceedings of the 16th Symposium of the Colston Research Society (eds P.E. Hart, G. Mills and J.K. Whitaker), 277-292. London: Butterworths.

Blackorby, C. and Donaldson, D. (1993) Adult-equivalence scales and the economic implementation of interpersonal comparisons of well-being. Social Choice and Welfare, **10**, 335-361

Bollino, C.A., Perali F., Rossi N. (2000) Linear household technologies. J. Appl. Econometrics, **15**, 275-287.

Browning, M., Bourguignon, F., Chiappori. P.A. and Lechène, V. (1994) Income and outcomes: a structural model of intrahousehold allocation. J. Pol. Econ., **102**, 1067-1097.

Bouroche, J.M. and Saporta, G. (1980) L'analyse des données, Paris, Presses Universitaires de France.

Dalla Zuanna, G. (2004) The banquet of Aeolus. A familistic interpretation of Italy's lowest low fertility. Demographic Research, **4**, 133-162. (doi:10.4054/DemRes.2001.4.5)

Deaton, A. and Muellbauer, J. (1986) On measuring child costs: with application to poor countries. J. Pol. Econ., **54**, 720-744.

De Santis, G. and Maltagliati, M. (2004) Equivalence scales: a fresh look at an old problem. Theory and empirical evidence, In Household Behaviour, Equivalence Scales, Welfare and Poverty. (eds C. Dagum and G. Ferrari), 29-53. New York: Physica-Verlag.

De Santis, G. and Maltagliati, M. (2010) Clusters and equivalence scales. In Proceedings of the 45th Scientific Meeting of the Italian Statistical Society, (in press).

De Santis, G. and Seghieri, C. (2006) How robust are equivalence scales? In Proceedings of the 43rd Scientific Meeting of the Italian Statistical Society, 133-136. Padova: Cleup.

Engel, E. (1895) Die Lebenskosten belgischer Arbeiterfamilien früher und jetzt. International Statistical Institute Bulletin, **9**, 1-74

Fernández-Villaverde, J. and Krueger, D. (2007) Consumption over the life cycle: facts from consumer expenditure survey data. Rev. Econ. Statist., **89**, 552-565.

Gorman, W.M. (1976) Tricks with Utility Functions. in Essays in Economic Analysis (eds M.J. Artis and A.R. Nobay), 211-243. Cambridge: Cambridge University Press.

Istat (2010a) Le differenze nel livello dei prezzi al consumo tra i capoluoghi delle regioni italiane, Rome (http://www.istat.it/salastampa/comunicati/non_calendario/20100707_00/testointegrale20100707.pdf)

Istat (2010b) La povertà in Italia nel 2009. Rome, (http://www.istat.it/salastampa/comunicati/in_calendario/povita/20100715_00/testointegrale20100715.pdf)

Kish, L. (1965[1995]) Survey Sampling. New York, Wiley.

Koohi-Kamali, F. (2008) Intrahousehold inequality and child gender bias in Ethiopia policy. World Bank Research Working Paper, **4755**. (http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2008/10/21/000158349_20081021080752/Rendered/PDF/WPS4755.pdf)

Lambert, P.J. (2004) Equivalence scales, horizontal equity and horizontal inequity. In Household Behaviour, Equivalence Scales, Welfare and Poverty. (eds C. Dagum and G. Ferrari), pp. 75–84. New York: Physica-Verlag.

Lewbel, A. (1989) Household equivalence scales and welfare comparisons. J. Pub. Econ., **39**, 377-391.

Lewbel, A. and Pendakur, K. (2006) Equivalence Scales. Entry for The New Palgrave Dictionary of Economics. 2nd edition, Boston College and Simon Fraser University. (http://www.sfu.ca/~pendakur/palequiv.pdf).

Muellbauer, J. (1975) The cost of living and taste and quality change, J. Econ. Theory, **10**, 269-283.

Muellbauer, J. and van de Ven, J. (2004) Estimating Equivalence Scales for Tax and Benefits Systems, NIESR (Natl. Inst. of Econ. and Soc. Res.) discussion papers, 229, http://www.niesr.ac.uk/pubs/dps/dp229.pdf.

OECD (2008a) Growing Unequal? Income Distribution and Poverty in OECD Countries. Paris. (http://browse.oecdbookshop.org/oecd/pdfs/browseit/8108051E.PDF)

OECD (2008b) What are equivalence scales?, mimeo. (http://www.oecd.org/dataoecd/61/52/35411111.pdf)

Perali, F. (2001) The Behavioral and Welfare Analysis of Consumption. Boston, Kluwer Academic Publishers.

Pollak, R.A. and Wales, T.J. (1979) Welfare comparisons and equivalence scales. Am. Econ. Rev. Papers and Proceedings, **69**, 216-221.

Prais, S.J. and Houthakker, H.S. (1955) The Analysis of Family Budgets. Cambridge, Cambridge Un. Press.

Ray, R. (1983) Measuring the cost of children: an alternative approach. J. Pub. Econ. **22**, 89-102.

Rothbarth, E. (1943) Note on a method of determining equivalent income for families of different composition. In War-time Pattern of Saving and Spending. (ed C. Madge), 123-130. Cambridge: Cambridge University press.