Chapter 2

Equivalence-Scale Measurement

2.1 Expert Approaches

The expert approach to equivalence scales refers to the evaluation of household needs by experts, usually on a low or minimum income level. In this method experts define specific minimum needs for households of different demographic composition. It is frequently applied in practical social policy. The OECD scale and the scale implicit in the German social assistance system are examples.

Historically, the expert approach has its roots in nutritional science, where scales were based on minimal calorie needs. The first attempt to calculate an expert based equivalence scale dates back to the seminal works of Ernst Engel (1895), who measured so called centimetre weights (height times weight of a person) to extrapolate the needs of an infant to the needs of older persons. This is a mixed approach, where a basic need is defined by expert opinion, while the extension to persons with other demographic properties is done by statistic measurement of some needs index. The focus of this approach lies on needs depending on age and gender, therefore these scales are individual equivalence scales rather than household equivalence scales. They assess the different needs of individuals, but not economies of scale arising from living together.

In a second strain of research, experts define a basket of goods and services, which contains essential amounts of food, clothing, energy and other goods. This approach was first developed by Rowntree (1901), to measure poverty. The ratios of the cash equivalents of the goods baskets for different household types can be interpreted as equivalence scales. Minimal needs can of course be extended to a socio-cultural minimum, on which the German social assistance levels were based until 1990. The goods basket approach has been criticized, because its definition entails explicit *judgements* on minimum needs, while approaches that are based on the *measurement* of some needs index (such as

Engel's centimetre weights or the economic methods described later) are seen as more "objective". It is arguable, however, if an explicit judgement is not preferable over the implicit judgements that are contained in all measurement approaches (Coulter et al., 1992).

In German social policy the goods basket was abolished in favour of a so called "statistical model", which is based on the expenditures of low income households.[1] This is also an expert approach, basically a refined goods basket model, that combines judgements with measurements. Survey data are used to define the goods basket of low income working households, who do not receive any social assistance. To find the level of social assistance, all non-necessary expenditures are taken out of the basket, again based on expert opinion. It is an advantage of this approach, that the resulting minimum needs are based on actual survey data, and actual expenditure of low income households, possibly facilitating the adjustment of benefit levels to changing needs. On the downside, the decision about which expenditures are not necessary seems as arbitrary as the inclusion of goods in the basic goods basket approach, while at the same time making public discussion over the appropriate level of social benefits more difficult, because expert judgements are less clear. As applied, the approach does not prove to be particularly useful for the estimation of equivalence scales, because the statistical basket is only used to determine the needs of the single reference household. The needs of larger households are then determined by multiplying the basic needs with a factor (an equivalence scale) that is not based on the same method, but rather arbitrary.

Related to the German statistical model is a third type of expert approach deployed in the U.S. by Orshansky (1965). In this method the total needs of a household are calculated by dividing the minimum need of a certain good, as defined by experts, through the empirically determined budget share of this good. Not unlike the statistical model of the German social assistance, this method extends expert based determination of needs with elements of demand analysis. The U.S. poverty line is based on this method.

All mentioned expert methods are concerned with the determination of poverty lines and the budget of households at a minimum expenditure level. The calculation of equivalence scales is merely an interesting side effect of these methods. This is a severe limitation, when equivalence scales and child costs change with income. In applications that are concerned with a wide range of income levels, such as distribution issues and taxation, a set of scales is needed for all income levels. In these cases expert based equivalence scales that are determined at the poverty line cannot be applied, when the "true" equivalence scales change with higher incomes.

---

[1] §28 SGB, §40 SGB and Regelsatzverordnung (regulation of the basic benefit level) as of 06/03/2004.

2.2 Economic Approaches

2.2.1 Demand Theory

There is a link between observed household behaviour (such as consumption choices, labour supply and fertility decisions) and the level of household well-being. Thus, economic approaches can derive equivalence scales from the observed consumption behaviour of households. Their foundations rest firmly in neoclassical microeconomic household theory. Household preferences can be represented by a utility function $u = U(\boldsymbol{q}, s)$ where $\boldsymbol{q}$ is the vector of the n consumption goods $q_1, ..., q_n$ and s is the index of a household type.[2] The starting point for the definition of the utility function is that households behave rationally. The hypothesis of rationality is embodied in the *completeness* and *transitivity* of the preference relation. Completeness means that the household has a well defined preference relationship between any pair of two consumption bundles. Transitivity rules out cycling preferences, i.e. if bundle $\boldsymbol{q}_A$ is preferred to bundle $\boldsymbol{q}_B$ and $\boldsymbol{q}_B$ is preferred to $\boldsymbol{q}_C$ then $\boldsymbol{q}_C$ cannot be preferred to $\boldsymbol{q}_A$. Complete and transitive preferences can be represented by a utility function $U(\boldsymbol{q}, s)$ that is increasing and quasiconcave in $\boldsymbol{q}$ and assigns a numeric value to each consumption bundle. This value has only ordinal meaning: A bundle with a higher utility value is preferred to a bundle with a lower value; the difference between the utility values is not relevant. It is important to note that the utility function so defined is not unique. Preference relations are preserved by any monotonic transformation of the utility function. If the preference relations are known, the utility function can be recovered only up to a monotonic transformation.[3]

Given preferences represented by the utility function U, disposable income μ and the price vector $\boldsymbol{p}$, the rational household maximizes its utility subject to the budget constraint $\boldsymbol{p}'\boldsymbol{q} \leq \mu$. The solution to the maximization problem yields the vector of *Marshallian demand functions* $\boldsymbol{q} = \boldsymbol{g}(\mu, \boldsymbol{p}, s)$, which specify the optimal consumption bundle $\boldsymbol{q}$ depending on prices, income and household demography. Substitution of Marshallian demands into the utility function $U(\boldsymbol{q}, s)$ gives the indirect utility function $u = V(\mu, \boldsymbol{p}, s)$.

The problem of minimizing the cost $x = \boldsymbol{p}'\boldsymbol{q}$ of attaining a given utility level u is dual to the utility maximization problem. Its solution leads to *Hicksian demands* $\boldsymbol{q} = \boldsymbol{h}(u, \boldsymbol{p}, s)$ which depend on the utility level u, prices and household characteristics. Substitution into $x = \boldsymbol{p}'\boldsymbol{q}$ yields the corresponding

---

[2]The household type can be represented by a vector of characteristics $\boldsymbol{s}$, which contains at the minimum the number of persons in the household. It can also contain the number of adults and children, ages of children and other parameters. If all parameters take discrete values, then the household type can be conveniently indexed by a suitably constructed household type index s. In what follows, an index will be used instead of the vector of characteristics, whenever possible. The same applies to the reference household index r.

[3]For a full introduction to household consumption theory, compare Deaton and Muellbauer (1980), pp. 25–52.

cost function $c = c(u, \boldsymbol{p}, s)$ which specifies the cost to attain the utility level u given prices $\boldsymbol{p}$ and household characteristics s. Inversion of the cost function again gives the indirect utility function.

The derivation of cost and indirect utility function can also be reversed. Hicksian demands can be obtained by deriving the cost function with respect to prices:

$$h_i(u, \boldsymbol{p}, s) = \partial c(u, \boldsymbol{p}, s)/\partial p_i,$$

or in budget share form:

$$w_i(u, \boldsymbol{p}, s) = \partial \ln c(u, \boldsymbol{p}, s)/\partial \ln p_i, \tag{2.1}$$

This equation is also known as *Shephard's Lemma*. Marshallian demands are derived from the indirect utility function using *Roy's Identity*:

$$g_i(\mu, \boldsymbol{p}, s) = -\frac{\partial V(\mu, \boldsymbol{p}, s)/\partial p_i}{\partial V(\mu, \boldsymbol{p}, s)/\partial \mu}$$

Figure 2.1 gives a graphic overview of all relevant analytical relationships in demand theory.

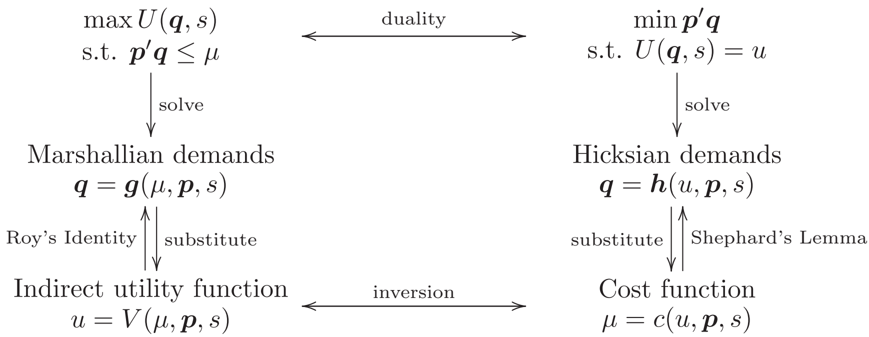

Figure 2.1: *Neoclassical demand theory: Utility maximization and cost minimization (Source: Deaton and Muellbauer, 1980, p. 38).*

Irrespective of any particular specification, the maximization of the utility function (subject to a budget constraint) imposes certain restrictions on the demand system. *Adding-up*: If household preferences are characterized by non-satiation, overall expenditures add up to household income. Intertemporal substitution of course leads to saving and consumption outside the observed time frame. This is consistent with adding-up. Because of *homogeneity*, a simultaneous and proportional increase in all prices and income leaves demands unaffected. In addition, two restrictions refer to the reaction of household demands on price changes: *negativity* means that an increase in the price of one good with utility held constant must cause the demand for that good not to increase, i.e. the derivatives of Hicksian demands with respect to own prices are negative or zero: $\partial h_i(u, \boldsymbol{p})/\partial p_i \leq 0$. *Symmetry* imposes the restriction that the cross-price derivatives of the Hicksian demands are symmetric: $\partial h_i(u, \boldsymbol{p})/\partial p_j = \partial h_j(u, \boldsymbol{p})/\partial p_i$. The properties of symmetry and negativity are tests for the consistence of a consumer's choices.

The Marshallian demands of a household can be estimated directly from observed expenditure data, while utility and cost functions cannot be estimated without an underlying theory. The functional form of a Marshallian demand function can be derived either from a direct or an indirect utility function or it can be specified without any reference to a utility function. The latter is regularly done in non-parametric demand analysis. The former allows for the identification of the respective utility and cost functions up to a monotonic transformation, if demands under different incomes and sufficiently different price regimes are observed.

2.2.2 Equivalence Scales, Demand Theory and the Identification Problem

For household type s – represented by the vector of characteristics $\boldsymbol{s}$ – relative to a reference household of type r with characteristics vector $\boldsymbol{r}$, an equivalence scale m_r^s can be calculated, if the cost function for both household types is known. In general, the scale will depend on the reference utility level u and on prices $\boldsymbol{p}$:

$$m_r^s(u, \boldsymbol{p}) = \frac{c(u, \boldsymbol{p}, \boldsymbol{s})}{c(u, \boldsymbol{p}, \boldsymbol{r})} \tag{2.2}$$

The cost function can be constructed from what Pollak and Wales (1979) call the *unconditional* utility function $U(\boldsymbol{q}, \boldsymbol{s})$ that describes a household's preferences over *both* goods $\boldsymbol{q}$ *and* household characteristics $\boldsymbol{s}$. Following Lewbel (1997), $U(\boldsymbol{q}, \boldsymbol{s})$ can be identified (up to an arbitrary monotonic transformation), if the household's choices over various combinations of $\boldsymbol{q}$ and characteristics $\boldsymbol{s}$ are observed. The unconditional equivalence scale is uniquely identified, because a monotonic transformation $t(.)$ of U equally affects all households, but not the relation between them.

Unconditional preferences account for the value that a household attributes to its characteristics $\boldsymbol{s}$. This is reflected in the equivalence scale. Pollak and Wales (1979) give the following example: if a couple chose to have two children and an income of \$12.000 when they could have had no children and an income of \$12.000, then a revealed preference argument implies that they are better off with children than without, and no compensation has to be paid.

The value of the unconditional equivalence scale would be less than one.[4]

Unconditional equivalence scales do not separate the money cost of children from their "value" to the household. For an assessment of the pure money cost of children, *conditional* equivalence scales must be estimated based on the material welfare of a household given its composition. This approach does not integrate the choice over household characteristics into the utility function. As a consequence, it suffers from a potential identification problem.

The conditional equivalence scale can be written as the relation of the two incomes μ^s and μ^r, that allow two households of different composition to attain the same level of utility u (here represented by the indirect utility function). In general, the scale will depend on the reference utility level u and prices $\boldsymbol{p}$:

$$m_r^s(u, \boldsymbol{p}) = \left. \frac{\mu^s}{\mu^r} \right|_{u=V^s(\mu^s,\boldsymbol{p})=V^r(\mu^r,\boldsymbol{p})} , \tag{2.3}$$

where $V^s(.)$ and $V^r(.)$ are the indirect utility functions of household types s and r, that were recovered using demand theory and observed expenditure data for these household types.[5]

Unfortunately, the equivalence scale cannot be identified from equation 2.3, because the utility function itself is only identified up to a monotonic transformation: Assume that $\tilde{V}^s(.)$ and $\tilde{V}^r(.)$ are indirect utility functions that are consistent with the observed demand behaviour of household types s and r. Then any monotonic transformation $t(.)$ of the two functions is consistent with the demand data as well. But the equivalence scale takes a different value if the transformation is not independent of the household type, with $t^s(\tilde{V}^s(.))$ and $t^r(\tilde{V}^r(.))$, because then the equality of utility levels depends on the transformation:

$$\tilde{V}^s(\mu^s, \boldsymbol{p}) = \tilde{V}^r(\mu^r, \boldsymbol{p}) \nLeftrightarrow t^s(\tilde{V}^s(\mu^s, \boldsymbol{p})) = t^r(\tilde{V}^r(\mu^r, \boldsymbol{p})) \ . \tag{2.4}$$

This affects the equivalence scale, and in general, the value of the equivalence scale depends on the transformation:

$$\left. \frac{\mu^s}{\mu^r} \right|_{\tilde{u}=V^s(\mu^s,\boldsymbol{p})=V^r(\mu^r,\boldsymbol{p})} \lesseqqgtr \left. \frac{\mu^s}{\mu^r} \right|_{\tilde{u}=t^s(V^s(\mu^s,\boldsymbol{p}))=t^r(V^r(\mu^r,\boldsymbol{p}))} . \tag{2.5}$$

---

[4]Of course, child benefits and earnings opportunities are influenced by the number of children, so that the choice of the couple is usually not only over the number of children but also over the varying associated income levels. The change in income that comes with an additional child will affect the choice of the marginal couple. The revealed preference argument for unconditional equivalence scales says that parents don't need to be compensated. If benefits were reduced, the birthrate would change.

[5]Following Pollak and Wales (1979), $V^s(.)$ and $V^r(.)$ are called *conditional* utility functions, because they represent the households' preferences given their characteristics s and r.

In fact, any value of an equivalence scales can be reconciled with observed behaviour, given the appropriate monotonic transformation $t(.)$ (Pollak and Wales, 1979).[6] In reaction to this fundamental identification problem, some researchers dismiss demand based equivalence scales altogether. Blundell and Lewbel (1991) write that "it is standard practice in demand analysis to propose a functional form for utility, estimate parameters from demand data, and report the resulting equivalence scales implied by the model. Our results show that this standard practice is inherently dishonest or at least uninformative, since in a given price regime, any value of equivalence scales can be rationalized by any demand system."

This view is one-sided. Nelson (1993) replies that "if one has a *theory* of intra-household allocation and how those allocations translate into demands, one may be able to justify a preference for one specific utility or cost function. Suppose for example, that the world were such that all agents were identical and treated symmetrically within households, and that there were no economies of scale in household formation. Then, if $C(u, \boldsymbol{p})$ is the cost function for a single individual, $J \cdot C(u, \boldsymbol{p})$ is the cost function for a J person household. The equivalence scale for a five person household is five." A monotonic transformation of u that depends on the number of individuals makes no sense in this model, because there is a theory of distribution and utility in a household. This theory is based on the individual utility of household members. The example "per capita theory" is certainly too simple and restrictive, but other models of household equivalence scales develop more elaborate models of intra household allocation and household economies of scale.

Lewbel (1997) points into the same direction. While emphasizing the fundamental identification problem, Lewbel concedes that identification can be possible if it is based on the individual, not the household. If the household technology can be identified, then equivalence scales can be estimated based on a *situation comparison* (Pollak and Wales, 1992), where the same type of individual is set in different household situations and its individual welfare from consumption is evaluated.

Not all equivalence scale models follow the route to identification suggested above. Some achieve a solution of the identification problem by using a proxy for household welfare as in the Engel method, or by making assumptions about the functional form of the equivalence scale as in the IB approach, but many indeed impose some structure on the intra-household decision process

---

[6] This can be illustrated by a simple example: assume there is only one good q with price p and let the direct utility function of an n person household be $U(q, n) = q/n$. Derive the indirect utility function with $V(\mu, p) = \mu/(np)$. Now compare a one person reference household r with an n person household s. It is easy to see that the equivalence scale is $m_r^s = n$. Now apply the monotonic transformation $t(x, n) = nx$ to the indirect utility function. This leads to exactly the same demand equation $q = \mu/p$, but the equivalence scale is 1. The example can be extended to any utility-function specification.

as the Rothbarth model, the Barten model and – most explicit – the collective model. The following sections will give an overview.

2.2.3 Engel Scales

One early and fundamental finding in empirical demand analysis is "Engel's law": the observation that the expenditure share of food decreases with income. This observation has been replicated in all cultures and at all times. Therefore it is reasonable to use the observed budget share of food as an indicator of a household's welfare. Another observation is that an increase in household size – income held constant – increases the budget share for food. So, a change of demographic characteristics acts in the same way as a reduction of income. Engel equivalence scales are based on the assumption that the budget share of food is an indicator of welfare that is valid also in the comparison of households of different composition. According to the assumption, two households of different composition are as well off if they spend the same budget share on food.

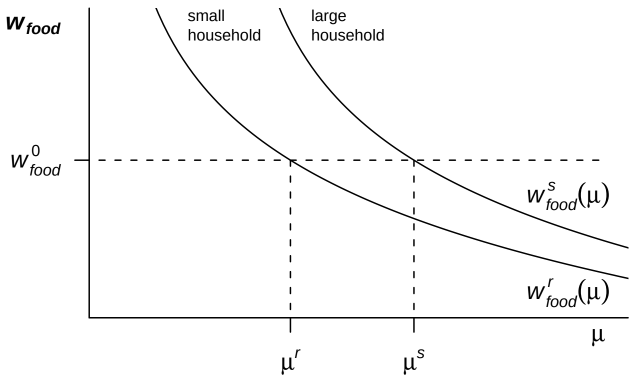

Figure 2.2: *Determination of Engel scales: a larger household of type s is as well off as a smaller household of type r at a reference income μ^r if its budget share for food is the same. This is the case at income μ^s.*

A graphic representation of the method is shown in Figure 2.2. The budget share of food is shown on the vertical axis, while the horizontal axis shows total income. The budget shares of food for two household types, a larger household of type s and a smaller household of type r, are drawn as curves w^s_{food} and w^r_{food}. As the budget share is falling in income both curves have a negative slope. Household s is as well off as household r at a reference income μ^r if its budget share for food is the same. This is the case at income μ^s. The equivalence scale at reference income level μ^r is then

$$m_r^s(\mu^r) = \left.\frac{\mu^s}{\mu^r}\right|_{w_{food}^s(\mu^s)=w_{food}^r(\mu^r)} . \tag{2.6}$$

Even though the identifying assumption of the Engel model is superficially plausible, it is an ad hoc assumption and not testable without a more complete model of the households demand pattern. A comparison with results from different methods does not help, because different methods also use different identifying assumptions that cannot be tested against each other. An argument why Engel scales are biased is given by Nicholson (1976). The argument goes as follows: A previously childless couple gets a child. If the child's consumption is more tilted towards food, which is plausible, then the household's food share will be increased. If the household were now correctly compensated for all child costs, then the household's food share were still increased, because the children's food share is higher than the parents', and Engel scales would indicate a reduction in welfare. Thus Engel scales are biased upwards. Unfortunately it is not possible to assess the size of the bias. This is a strong argument against the method, and consequentially Deaton (1997) writes "The method is unsound and should not be used."

Nevertheless, because of its ease of use there have been many practical applications in the literature. Gozalo (1997) uses non-parametric estimation methods to derive Engel scales, albeit his focus lies in the econometric aspects of the application of bootstrap analysis, not in the resulting scales themselves. Despite severe reservations, Deaton and Muellbauer (1986) estimate Engel scales in comparison with Rothbarth scales (see section 2.2.4) with a focus on the availability of data and feasibility of the estimation process. Tsakloglou (1991) estimates Engel and Rothbarth scales for a sample from Cyprus. Results are similar to those of Deaton and Muellbauer, in particular the result is replicated that Rothbarth scales are much lower than Engel scales. Bradbury (1994) estimates Engel scales for Australia. Engel scales that were estimated by Merz and Faik (1995) for Germany are reported in Table 2.1.

The Engel method can be rationalized in an economic model. Following Deaton and Muellbauer (1980) the Engel cost function and direct utility functions are:

$$c(u, \boldsymbol{p}, s) = m(s)\tilde{c}(u, \boldsymbol{p}), \tag{2.7}$$

$$u = U(q_i/m(s), \ldots, q_n/m(s)) \tag{2.8}$$

where the scale factor $m(s)$ is a function of the household type index s, which can be normalized to a value of one for the reference household r.[7]

Using Shephard's lemma in share form (equation 2.1), the food share w_f

[7] The normalization of one is arbitrary. It could also be two for a two person reference household or any other suitable value.

can be calculated:

$$w_f = \frac{\partial \ln c(u, \boldsymbol{p}, s)}{\partial \ln p_f} = \frac{\partial \ln m(s)}{\partial \ln p_f} + \frac{\partial \ln \tilde{c}(u, \boldsymbol{p})}{\partial \ln p_f} = \frac{\partial \ln \tilde{c}(u, \boldsymbol{p})}{\partial \ln p_f} \tag{2.9}$$

The food share depends only on prices and the level of utility but not on demographic characteristics. Therefore, any two households that have the same food share have the same level of utility. The scaling function $m(s, u)$ is easily identified from the Marshallian demand functions:

$$q_f = m(s) g_f(\mu/m(s), \boldsymbol{p}) \tag{2.10}$$

In budget share form this becomes

$$w_f = \frac{p_f q_f}{\mu} = \frac{p_f g_f(\mu/m(s), \boldsymbol{p})}{\mu/m(s)} \tag{2.11}$$

Now w_f is a function of $\mu/m(s)$ but not of μ or $m(s)$ separately. When the expenditure levels μ^s and μ^r at which households s and r spend the same share on food are found, the equivalence scale can be calculated:

$$\frac{\mu^r}{m(r)} = \frac{\mu^s}{m(s)} \text{ and, with } m(r) = 1 \ : \ m(s) = \frac{\mu^s}{\mu^r} \ . \tag{2.12}$$

The model can be generalized with a scale factor and an equivalence scale that depend on the level of utility by replacing $m(s)$ with $m(s, u)$.

Identification of equality of the utility level is not restricted to the food share. Indeed, according to the theory, not only the food share, but all shares will be equal given these preferences. This is certainly very restrictive, and a test is straightforward. Tsakloglou (1991) tests the restriction and rejects it forcefully.

Given the restrictive nature of the Engel model, it is a natural extension of the method *not* to require the scale factors for all commodities to be the same. This generalization is known as the Prais and Houthakker (1955) model, where the Engel curve is generalized to

$$\frac{q_i}{m_i} = g_i\left(\frac{\mu}{m_0}\right) \ . \tag{2.13}$$

m_0 is a general equivalence scale and is defined as a weighted average of the individual commodity scales m_i. Unfortunately, the Prais-Houthakker scale is mathematically not identified (Muellbauer, 1980): There are n goods and therefore n unknown individual commodity scales m_i. But there are only $n-1$ independent Engel curves: due to the budget constraint and the adding up restriction, the n^{th} Engel curve is not independent of the others. With n unknowns and only $n-1$ equations an identification of the system is not possible. In an empirical application Muellbauer confirms the underidentification result. Even though not suited for equivalence scale estimation, the model still can be applied to the integration of demographic characteristics into demand systems, as in Pollak and Wales (1981).

2.2.4 Rothbarth-Scales

Rothbarth equivalence scales are specifically designed for the estimation of child costs. They permit the comparison of households with the same number of adults and a differing number of children, i.e. none, one or more children. The method dates back to Rothbarth (1943), who suggested that adult equivalence scales can be derived from the observation of "adult goods". Typical adult goods are tobacco, alcohol and adult clothing.

According to this model, a couple with children is as well off as a couple without children if the parents can spend the same amount on their own consumption. Even though in most expenditure surveys it is not reported who in the household consumes which good, there are a few goods that are consumed exclusively by adults, either by definition or by custom. Under some additional assumptions, these goods can be used for the identification of total adult consumption. Then two households with the same number of adults are equally well off if their expenditure on adult goods is the same.

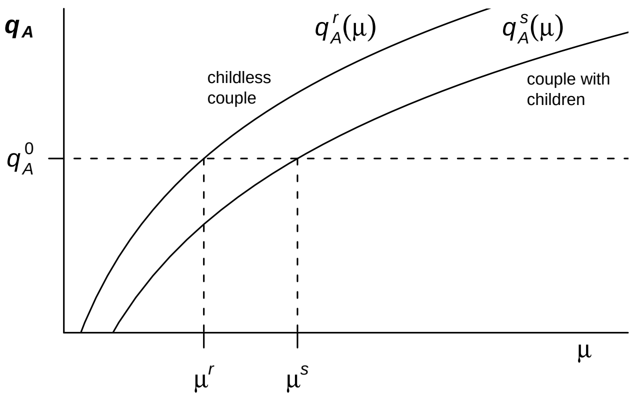

Figure 2.3: *Demands for adult goods in households S and R.*

The method is shown in Figure 2.3. The vertical axis shows the amount q_A of the adult good purchased, while the horizontal axis shows total outlay μ; $q^r_A(\mu)$ and $q^s_A(\mu)$ are the Engel curves of the adult good of a childless couple and a couple with children, respectively. It is assumed that both households have the same level of welfare if expenditure for the adult good is the same. Therefore the larger household s has the same welfare at expenditure level μ^s as the reference household r at expenditure level μ^r and the equivalence scale is: $m(\tilde{u}) = \mu^s/\mu^r$, where $\tilde{u} = V(\mu^r, r) = V(\mu^s, s)$ is the utility level both couples derive from their respective incomes. In general, the equivalence scale depends on the reference utility level $\tilde{u}$.

As with other equivalence scales, the focus of the method clearly lies on parents' utility. However, no other method makes this more explicit. It is apparent that the method focuses only on the material well-being of the parents, taking a clear position on the discussion about which is the proper definition of welfare that equivalence scales should address.

Formally, the model assumes a cost function that is additively separable in the costs of parents and children:

$$c(u, \boldsymbol{p}_A, \boldsymbol{p}_B, \boldsymbol{s}^c) = c^a(u, \boldsymbol{p}_A, \boldsymbol{p}_B) + c^c(u, \boldsymbol{p}_B, \boldsymbol{s}^c) \tag{2.14}$$

where $\boldsymbol{p}_A$ and $\boldsymbol{p}_B$ are the price vectors for adult and other goods, and $\boldsymbol{s}^c$ is the vector of demographic characteristics that are only connected with children. The first component of the cost function c^a can be seen as the cost of adults and the second c^c as cost of children. Given the additive cost function 2.14 the Hicksian demand for any adult good q_{Ai} is given by:

$$q_{Ai} = \frac{\partial c^a(u, \boldsymbol{p}_A, \boldsymbol{p}_B)}{\partial p_{Ai}} = h_{Ai}(u, \boldsymbol{p}_A, \boldsymbol{p}_B). \tag{2.15}$$

A change in the children's demographic characteristics will have only an income effect on the demand of adult consumption. Assuming that prices are the same for all households and normality of the adult good, u and q_{Ai} are correlated: an increase in q_{Ai} is equivalent to an increase in u, and q_{Ai} is an indicator of adult welfare. The cost of children $c^c(u, \boldsymbol{p}_B, \boldsymbol{s}^c)$ of a childless couple equals zero. The adult cost function $c^a(u, \boldsymbol{p}_A, \boldsymbol{p}_B)$ can be identified from the observation of childless couples and the equivalence scale with childless couples as the reference household is given by

$$m_0(u, \boldsymbol{p}_A, \boldsymbol{p}_B, \boldsymbol{s}^c) = \frac{c^a(u, \boldsymbol{p}_A, \boldsymbol{p}_B) + c^c(u, \boldsymbol{p}_B, \boldsymbol{s}^c)}{c^a(u, \boldsymbol{p}_A, \boldsymbol{p}_B)} \tag{2.16}$$

One reason for the ongoing popularity of the Rothbarth method is that it is easy to apply if adult goods are observed. The identification of adult goods in expenditure surveys might be difficult, though. Expenditures on tobacco and alcohol are often observed, but a limited income elasticity of tobacco expenditure together with a strong reaction of tobacco consumption on the presence of children makes estimation infeasible. Standard errors become very large; in the worst case estimation becomes impossible because parents with children never reach the same level of tobacco consumption as childless couples (see Figure 2.4). Engel curves for tobacco and alcohol consumption are also often not monotonic. When several expenditure levels generate the same level of consumption of the adult good, the equivalence scale is ambiguous.

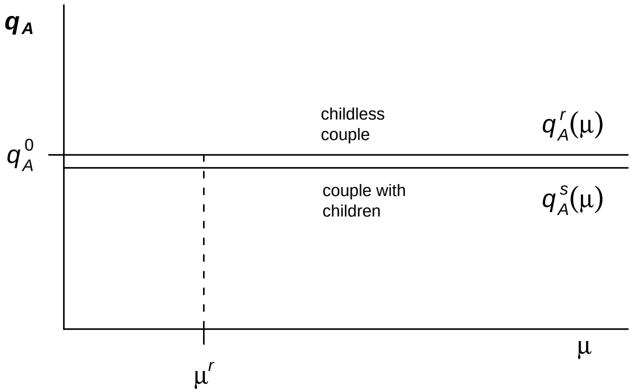

Figure 2.4: *Rothbarth scales cannot be identified, if the demand for the adult good is not income elastic.*

Low income shares of alcohol and tobacco lead to additional measurement problems.[8] Due to a high income elasticity and a high expenditure share, adult clothing does not suffer from these problems and is therefore the best adult good candidate for the estimation of Rothbarth equivalence scales. Unfortunately, more than one good is needed for a test of the model assumptions: if the assumptions are correct, equivalence scales estimated from expenditures on each good must be identical.[9]

Despite these practical problems, the method is almost as easy to apply as the Engel method and certainly preferable over that method on the ground that the theoretical assumptions of the Rothbarth model are more plausible than those of the Engel method. Therefore it finds regular application, see for example Lazear and Michael (1988). Gronau (1988, 1991) uses a Rothbarth like model to assess the intra-household distribution of income between parents and children, but he abstains from calculating equivalence scales because of the above mentioned identification problem. Gronau argues that even though demands of parents and demands of children can be separated by the method, wants remain indistinguishable from needs, thus leaving equivalence scales, which are based on needs, unidentified. Deaton et al. (1989) discuss extensively the assumption of an additive cost function on Spanish data. They call this *demographic separability*. They find mixed evidence on the validity of the separability assumption. Although rejected on conventional statistical measures, they argue that the large sample size of more than 20.000 households makes acceptance on usual measures unlikely even if there is only slight disagreement between data and theory. The assumption is accepted

[8]Sometimes it is argued that tobacco and alcohol are not appropriate for measuring adult welfare, because they are not goods but "bads" (compare Atkinson, 1983, p. 51)

[9]For a detailed discussion of tests of the model see Chapter 3, p. 58.

on a broader measure of a large sample Bayesian criterion for families with children between 8 and 13 years of age.

Deaton and Muellbauer (1986) establish a relationship between Rothbarth and Engel scales. To obtain clear results, they assume that all non-foods are adult goods. The Rothbarth scale is no larger than the Engel scale if the nonadult goods in the Rothbarth model correspond to food in the Engel model and if the nonadult goods (*food*) are necessities. If, when a child is born, the parents are compensated according to Rothbarth, their nonfood expenditure remains unchanged. But as their total expenditure has increased, the share of food has risen. Hence, according to Engel, they have been undercompensated. Therefore, Engel scales are higher than Rothbarth scales. Deaton and Muellbauer find that this result holds also in an empirical application where pure adult goods are used for identification of Rothbarth scales. The result is replicated by Tsakloglou (1991) on data from Cyprus.

To my knowledge Faik (1995) is the only recent study that calculates Rothbarth scales for Germany. Using data of the 1969 Income and Expenditure Survey of expenditure on alcohol and tobacco, Faik estimates absolute scales. He gets sensible results only for expenditures on alcohol (reported in Table 2.1, while the resulting scales for tobacco are low or even negative (and not significantly different from one in any case). This comes as no surprise. As discussed above, expenditures for tobacco are highly income inelastic, which can render estimation impossible.

Rothbarth scales are only suited to derive the distribution between parents and children, because of the focus on adult goods. The cost of characteristics of parents cannot be assessed by Rothbarth equivalence scales, because parents' characteristics are missing in the children's cost function c^c. The needs of, say employed and unemployed persons cannot be compared with the method if adult goods are not separable from employment status. However, possibly the method can be adapted by finding other goods that are separable. This would cause the same problems as the standard Rothbarth method, because separability is a strong restriction and should be tested.

Two aspects of the method can be criticized: the reliability of the observation of adult goods and the validity of the assumption of demographic separability. The problem of adult goods has been discussed already above. Provided a long enough survey period, expenditures on adult clothing can be measured with sufficient precision. For short survey periods, data on clothing suffer from censoring because of bulk purchases, but econometric methods are available to correct for possible bias (e.g. Heckman, 1980).

More serious is the bias that is introduced by a violation of the separability assumption. Children can either have a direct effect on parents' preferences for certain adult goods[10] but the presence of children can also have an indirect

[10] Prais and Houthakker (1955) note that "the father who is driven to drink by his wife and children is the textbook example".

effect on demands by influencing implicit prices. Prices can be influenced in two ways. First, and most of all children influence their parents' value of time. Apart from effects on labour supply (which is not modelled here), all time intensive activities and the accompanying consumption become much more expensive. For example a visit to the theatre is more expensive when parents have to pay a baby sitter. Second, relative opportunity costs can change when public goods are shared within a larger household, while private goods become relatively more expensive (Nelson, 1992). The omission of possible substitution effects in the method leads to a bias of the estimated equivalence scale. The direction and size of the resulting bias is not clear, though. It depends on the direction of the price change of the identifying adult good and on the size of substitution elasticities of the good. See Chapter 3, p. 55, for an extended discussion and a simple model of the resulting bias.

Despite these reservations, in Chapter 3 the Rothbarth method is used as a starting point for equivalence scale estimation in this work. There, the implications of the model and possible tests will be discussed in more detail. The size and direction of possible substitution effects will be assessed in a comparison with results from the Barten method (see below) that are estimated in Chapter 4.

2.2.5 Demand System Based Approaches

The great popularity of Engel and Rothbarth equivalence scales is mostly founded in their ease of estimation and the relatively low data requirements. Methods that estimate a complete demand system are much more data demanding and complex in their application. However, they have some theoretical advantages that make them attractive in application.

The Barten Method

In the Barten (1964) approach, demographics enter the utility function via good specific scale factors m_i^s, quite similar to the Prais-Houthakker model. Scale factors cover different needs connected with different demographic characteristics as well as different economies of scale arising from living together and sharing goods. A larger household has to buy more of every good that is not perfectly public, but the multiplier m_i^s for each good i can be different. For instance when the father wants to have an ice cream cone, he has to buy one for his wife and each of his children, too, while the fast internet connection can be shared by all members of the family. The effective quantity consumed per adult is q_i/m_i^s.[11] To compare utility between households, quantities of

[11] To be exact, q_i/m_i^s is the effective quantity per reference unit. This can be a single adult as well as a couple or any other household type. For the sake of clarity, a single adult is taken as reference unit in what follows.

all goods are scaled down by their respective scale factor:

$$U = U(q_1/m_1^s, ..., q_n/m_n^s) \ , \tag{2.17}$$

where U is the utility level of household type s and $U(.)$ is the household utility function of the reference household. The scale factors m_i^s are assumed to be exogenous, i.e. they are independent of quantities consumed, prices and income.[12] Scale factors m_i^r for the reference household r are normalized to one.

Scaling acts directly on prices because it changes the effective quantities consumed. A household with demographic composition s faces different scaled prices $p_i m_i^s$ than a reference household r. To continue the example, the price of the ice cream cone increases threefold for the father who has a wife and one child, while the price of the internet connection stays the same as compared to a single reference household.

Scaled prices can be plugged into the cost function:

$$c = c\,(u,\ p_1 m_1^s, p_2 m_2^s, ..., p_n m_n^s) \tag{2.18}$$

The equivalence scale m_r^s is the relation of the cost functions of household s and the reference household. In general, the scale depends on the reference level of utility u^r:

$$m_r^s(u^r, s) = \frac{c\,(u^r,\ p_1 m_1^s, p_2 m_2^s, ..., p_n m_n^s)}{c\,(u^r,\ p_1, p_2, ..., p_n)} \tag{2.19}$$

Hicksian demands are derived using Shephard's lemma, where the total household demand is the adult equivalent demand per head h_i times the scale factor m_i^s.

$$q_i = m_i^s h_i\,(u,\ p_1 m_1^s, p_2 m_2^s, ..., p_n m_n^s) \tag{2.20}$$

To determine the scales, a system of demand equations can be estimated together with the scale factors. It is a great practical advantage of the model that demography enters the demand system exclusively via prices. Provided data with sufficient price variation are available, Barten scales are readily estimable with any demand system specification that fits the data. For a practical application see Chapter 4.

There are three concerns about the assumption that demographic changes affect demands exclusively through price changes. The demand for goods that are not consumed by the reference household can only be explained by excessive substitution effects; the model can predict excessive substitution towards goods that are private to the reference group; and it could be that a change in demographics affects the price elasticities of the demands for some

[12] See Muellbauer (1974).

goods. All three points in question are most relevant to the comparison of families with children, when the reference household is a household without children.

1) The model cannot account for any goods that are only consumed by families with children. If the reference household does not consume a certain good, consumption of this good must be explained by substitution effects. A childless couple does not use diapers or baby foods while households with small children do. In the Barten Model this can only be explained by very far-fetched assumptions about substitution effects. Let q_b be the consumed quantity of baby specific goods and $h_b(.)$ the respective Hicksian demand function. Then the Hicksian demand of the family with children for baby goods is $q_b = m_b^s h_b(.)$. But the reference household's demand for the good is zero, so no multiplicative factor can lead to a positive demand of the good, except if the reference household is at a corner solution and a shift in prices eventually leads the household to demand some quantity of the good. The problem is more of theoretical concern, because the model is usually applied to larger groups of goods, where children's goods are always combined with their adults' counterparts. Parents do not feed their baby with the same food, but they will probably feed the baby as well as themselves, and parents have to add to their clothing bill an item for baby jumpers proportional to the expenditures on their own clothing. All things considered, the problem may be less serious than it appears at first sight, but it has to be taken into account when constructing the commodity groups of the estimated demand system.

2) A change in relative prices leads to substitution effects, away from goods with high scale factors and high scaled prices towards goods with lower scale factors. In application to families this means that there is a possible substitution away from child intensive goods, like milk towards less child intensive goods like whisky. It seems a bit odd that parents should shift demand from milk to whisky because the demands of their child make milk so expensive. Again the problem is alleviated when whisky and milk are grouped together as drinks, but some information is lost in the process. Also one might think that it is not sensible to group vices such as tobacco and alcohol together with food. One might, however, take comfort in the fact that such adult only goods typically constitute only a small part of the budget and therefore their influence on the overall result is small.

3) The last concern is about the constancy of price elasticities across household types. The demand for milk might be highly price elastic in a childless household, while it is rather inelastic in a household with children. This can even be the case for broad commodity groups with a large share of expenditures, such as housing: While childless couples might react to a change in housing prices by moving to a cheaper neighbourhood or a smaller apartment, parents could be more restricted in their choice because of circumstances such as school and neighbourhood quality. If this were the case, a general price

increase for housing would cause childless households to adjust their housing expenditures more compared to households with children, effectively leading to a lower price elasticity of housing consumption.

The Gorman (1976) extension to the model solves at least part of the problems of overstretched substitution effects by adding fixed consumption terms β_i^s to the demands of the household. The cost and Marshallian demand functions become:

$$c^s = c\,(u,\ p_1 m_1^s, p_2 m_2^s, ..., p_n m_n^s) + \sum_{j=1}^{n} p_j \beta_j^s \tag{2.21}$$

$$q_i^s = m_i^s g_i \left(\mu - \sum_{j=1}^{n} p_j \beta_j^s,\ p_1 m_1^s, p_2 m_2^s, ..., p_n m_n^s \right) + \beta_i^s \tag{2.22}$$

Demographic effects on demands now consist of two parts: some necessary expenditures, e.g. diapers, some milk, and other things for children, and a scaling effect that is proportional to the consumption of the reference household. For example after buying necessary basics, higher income parents spend additional money on their children's clothes – proportional to their own increased expenditures.

Demographic translating is a special case of the Gorman model, where all scaling factors are unity. With translating, the cost of children are fixed and do not depend on income. Marshallian demands become:

$$q_i^s = g_i \left(\mu - \sum_{j=1}^{n} p_j \beta_j^s,\ p_1, p_2, ..., p_n \right) + \beta_i^s. \tag{2.23}$$

Translating is even more restrictive than Barten scaling, because it allows only for fixed costs for children: no matter how high the income of parents, the costs of children are always the same. In a test of scaling, translating and the Gorman model, Pollak and Wales (1981) reject only the translating model against an unpooled specification. Translating fares somewhat better when it is applied to household characteristics other than the number of children, such as age, work status or housing situation. Consequentially, the model is rarely used in the estimation of equivalence scales.[13]

A further criticism of the Barten method points to the exogeneity assumption of the scale factors. This applies to both the Barten and the Gorman method. Nelson (1992) remarks that the exogeneity of the goods specific scale factors m_i^s can be tested for clothing, because in most surveys this item is

[13] Kohn and Missong (2003) contains a recent application of translating equivalence scales, but the authors are careful to interpret the translation parameters β_j^s only in terms of minimum subsistence expenditures.

reported separately for parents and children. If the scale factors were exogenous, then the income elasticities of clothing expenditures for adults and children would be equal. Instead she finds income elasticities for children's clothing that are half the size of those for adults. In the Gorman approach, elasticities can differ due to the addition of a fixed cost term. Here the ratio of children's to adults' consumption should be equal at the margin. This is rejected, as this ratio decreases significantly with rising income.

Nelson concludes that "This implies that at least one of the Barten model m_i parameters is not exogenous. Once it is allowed that these can change with income, the m_i parameters estimated with the Barten-form demand equations must be interpreted as being determined by the *outcome* of intra-household allocation decisions, not, as the Barten model holds, as the factors which determine household allocation." (emphasis in original)

Both the Barten and the Gorman method will be discussed in more detail and applied in chapter 4. The above-mentioned criticisms are considered in the application of the Barten method.

The Extended Linear Expenditure System

Merz and Faik (1995) use the Barten approach in an extended linear expenditure system (ELES) to estimate scales for Germany, following earlier work by Lluch (1973) and Kakwani (1977). Not all parameters of the linear expenditure system (LES) can be identified without the observation of some price variation. For n commodity groups, the demand equations of the LES are:

$$x_i = p_i b_i + a_i \left(\mu - \sum_{j=1}^{n} p_j b_j \right), \tag{2.24}$$

where x_i is the expenditure on good i, b_i can be interpreted as some subsistence expenditure and a_i is the marginal budget share with $\sum_{i=1}^{n} a_i = 1$. Without price variation, linear Engel curves can be estimated directly with:

$$x_i = \gamma_i + \alpha_i \mu \tag{2.25}$$

Setting prices to unity and comparing with equation 2.24 shows that $a_i = \alpha_i$ is readily identified, while $b_i = \gamma_i + \alpha_i \sum_{j=1}^{n} b_j$ cannot be identified without further restrictions. There are n b_i but only $n-1$ independent equations because of the adding up restriction: $\sum_{i=1}^{n} \gamma_i = 0$.

To identify the expenditure system without the need to observe behaviour under different price regimes, it is assumed in the ELES that the subsistence level for saving is zero. Saving is included into the demand system and the respective b_{saving} is set to zero, allowing for the identification of all other b_i.

Demographics can be included into the demand system by adding a dummy variable for each household type except for the reference household. A different set of b_i^s can be calculated for each household. With b_i the parameter of the reference household, the b_i^s can be interpreted either in terms of the Barten model ($b_i^s = m_i^s b_i$) or in terms of a translating model ($b_i^s = b_i + \beta_i^s$). The resulting equivalence scales are identical at the subsistence level $\sum_{j=1}^n p_j b_j$ but different at any higher income. This gives the results of the method some ambiguity.

At low income levels the method is equivalent to both Engel and Rothbarth scales with saving as the key commodity (Binh and Whiteford, 1990; Bradbury, 1994). The method is easy to apply and produces plausible equivalence scales, but it is probably more informative about saving behaviour than about relative household needs.

The method has been applied to German data by Scheffter (1991), Faik (1995), Stryck (1997) and Missong and Stryck (1998).

2.2.6 Cost Scaling and Independent of Base Equivalence Scales

Ray (1983) suggests an approach called *cost scaling*. This approach is similar to the theoretical model of the Engel method (Equation 2.7), in postulating a cost function that is separable in a scaling function $m_0(.)$ and a reference cost function $\tilde{c}(.)$ which is independent of the household type:

$$c(u, \boldsymbol{p}, s) = m_0(u, \boldsymbol{p}, s)\tilde{c}(u, \boldsymbol{p}) \tag{2.26}$$

Estimation is possible, given a suitable specification of both the scaling function $m_0(.)$ and the cost function $\tilde{c}(.)$. If m_0 is suitably normalized to one for the reference household, the interpretation as an equivalence scale is straightforward.

This approach solves the problem of extreme substitution present in the Barten method, because the impact of children on their parents' demands is not modelled as a quasi-price effect. To estimate equivalence scales, however, a structural model of the household is necessary. In this model, the household is treated as a "black box". The model gives no guidance on the choice of a functional form for the equivalence scale function $m_0(.)$ apart from the restriction that m_0 has to be homogeneous of degree zero in prices. Equivalence scales can be estimated, but the method suffers from the fundamental identification problem discussed in section 2.2.2: Estimated scales will vary systematically with the a-priori specification, because a monotonic transformation of the utility function can affect $m_0(.)$ as well as c and $\tilde{c}$.

To solve the identification problem in the cost scaling approach, Blackorby and Donaldson (1991) and Lewbel (1989) independently suggested the use of constant equivalence scales, which do not depend on the reference level of utility. They termed this *independence of base* (IB) or *equivalence scale exactness* (ESE), respectively. Under the IB property, the cost function is separable in characteristics and utility level. It can be written in the form:

$$c(u, \boldsymbol{p}, s) = m_0(s, \boldsymbol{p})\tilde{c}(u, \boldsymbol{p}) \tag{2.27}$$

With the equivalence scale not depending on utility, a monotonic transformation of the utility function affects $\tilde{c}$, but not m_0, and the equivalence scale is uniquely identified.

With $\tilde{V}(.)$ being the indirect utility function of the reference household, the base independent (IB) dual indirect utility function of household s can be written as:

$$V(\boldsymbol{q}, \mu, s) = \tilde{V}\left(\boldsymbol{p}, \frac{\mu}{m_0(\boldsymbol{p}, s)}\right) \tag{2.28}$$

Using Roy's identity, Pendakur (1999) derives Marshallian demand equations in terms of the demand equations of the reference household $\tilde{g}_i(\boldsymbol{p}, \mu)$:

$$g_i(\boldsymbol{p}, \mu, s) = m_0(\boldsymbol{p}, s)\tilde{g}_i\left(\boldsymbol{p}, \frac{\mu}{m_0(\boldsymbol{p}, s)}\right) + \frac{\mu}{m_0(\boldsymbol{p}, s)}\frac{\partial m_0(\boldsymbol{p}, s)}{\partial p_i} \tag{2.29}$$

Multiplication of Equation 2.29 with p_i/μ gives Marshallian share equations. Defining $\psi_i(\boldsymbol{p}, s)$ as the elasticity of $m_0(\boldsymbol{p}, s)$ with respect to p_i, one can write:

$$w_i(\boldsymbol{p}, \mu, s) = \tilde{w}_i\left(\boldsymbol{p}, \frac{\mu}{m_0(\boldsymbol{p}, s)}\right) + \psi_i(\boldsymbol{p}, s) \tag{2.30}$$

Under the IB assumption, Marshallian demands of a household s are equal to the Marshallian shares of the reference household plus the elasticity of the equivalence scale with respect to the price. The form of Engel curves of the reference household is not restricted by the model, but the shape of Engel curves is closely linked across household types. In share/log income ($\{w_i, \log \mu\}$) space, share functions must be related by a horizontal and a vertical shift, parametrized by m_0 and ψ, respectively; the shape of the Engel curves does not change. Pendakur calls this *shape invariance.*

This property can be used to estimate equivalence scales and to test the IB assumption. Pendakur uses semi-parametric estimation methods to test shape invariance and estimate scales. Shape invariance is not rejected for a comparison of childless single households with childless couples, nor for couples with one and two children. It is rejected, however, for a comparison between couples with and without children. This is an interesting result. It implies that one should be careful when estimating equivalence scales for households with and without children at the same time. It can also be interpreted as children having some fixed cost that makes children more "expensive" relative to an adult at low incomes than at high incomes. At the same time, having a second child is proportionally not more expensive for a poor household than for a rich one.

The shape invariance property also allows for the estimation of equivalence scales from data without price variation. This is particularly useful in the German case, due to the limited availability of time series data. Missong (2004) uses the method to estimate equivalence scales for Germany (reported in table 2.1), but at the same time he rejects the IB assumption for most household-type comparisons, when tested against unpooled Engel curve estimates.

The IB approach generates testable restrictions on the model other than shape invariance as well. Unfortunately the restrictions are usually rejected. Blundell and Lewbel (1991) question the empirical reasonableness of the independence of base assumption. E.g. quasi-homothetic or translog cost functions imply IB only if certain restrictions on the parameters hold. They reject the IB separability restrictions using data from the UK Family Expenditure Survey. Pashardes (1995) also tests the homogeneity conditions imposed by the IB assumption on the equivalence scale function and finds them rejected by the data.

In response to these findings Donaldson and Pendakur (2004) suggest an extension to the IB approach, where they define equivalence scales that depend on income in a specific way, called generalized equivalence scale exactness (GESE): When two different households are equally well off, independence of base (or equivalence scale exactness as it is called by the authors) implies that a one-percent increase in expenditure by one household must be matched by a one percent increase in expenditure by the other household to preserve equality of well-being. GESE allows the matching increase to be different from one percent, but it requires the increase to be the same for all expenditure levels. Donaldson and Pendakur show that the equivalent expenditure function can be identified if GESE is a maintained hypothesis and the form of the reference expenditure functions is not too simple.[14] In an empirical comparison, IB is rejected in favour of GESE. The authors note, however, that the choice of the functional form of the equivalence scale cannot be based on demand data alone (due to the identification problem), but should be based on other evaluation methods, survey methods for example. When the form of the scale is known in advance, then its level can be determined.

2.2.7 Collective Equivalence Scales

A solution of the identification problem can be achieved by either using a proxy for household welfare as in the Engel method, by making assumptions about the functional form of the equivalence scale as in the IB approach, or by imposing some structure on the intra-household decision process as in the Rothbarth and the Barten models. The collective model of equivalence scales suggested by Browning et al. (2004) falls in the last category. In contrast

[14] It is required that the expenditure function E^r is not of the PIGLOG type, which can be written as $\ln E^r(u, \boldsymbol{p}) = C(\boldsymbol{p})f(u) + \ln D(\boldsymbol{p})$.

to all other approaches, the collective model is based on individual utility functions that may differ between household members.

For simplicity of notation the model is described here for a mixed couple that consists of a man m and a woman f. The preferences of the household can be represented by a social welfare function or a bargaining function $\tilde{U}$ which is increasing in U^f and U^m and which can depend itself on prices, income and other demographic variables $\boldsymbol{s}^D$ that influence the intra-household distribution of resources:

$$\tilde{U}(U^f(\boldsymbol{q}^f), U^m(\boldsymbol{q}^m), \boldsymbol{p}, \mu, \boldsymbol{s}^D) \tag{2.31}$$

$\tilde{U}$ can arise from personal utility functions $\tilde{U}^f(U^f(\boldsymbol{q}^f),\ U^m(\boldsymbol{q}^m))$ and $\tilde{U}^m(U^f(\boldsymbol{q}^f), U^m(\boldsymbol{q}^m))$ that are increasing and separable in U^f and U^m. This means that each member of the household receives utility from the consumption of other persons in the household only in so far as they receive themselves utility from their own consumption. This implies for example that the husband does not care if his wife buys lingerie or shoes, as long as she likes it.[15] The utility functions $\tilde{U}^f$ and $\tilde{U}^m$ also embody intangible contributions to the couple's well-being such as love and companionship. Identification is limited to the personal consumption utility functions U^f and U^m.

$\tilde{U}$ is maximized subject to the budget constraint

$$\boldsymbol{p}'(\boldsymbol{q}^f + \boldsymbol{q}^m) = \mu. \tag{2.32}$$

In this version of the budget constraint there are only private goods and neither public goods nor economies of scale in consumption. A sensible model would include economies of scale. With economies of scale, the budget constraint can be generalized to

$$\boldsymbol{p}'\boldsymbol{F}(\boldsymbol{q}^f, \boldsymbol{q}^m) = \mu \tag{2.33}$$

where $\boldsymbol{q} = \boldsymbol{F}(\boldsymbol{q}^f, \boldsymbol{q}^m)$ is the bundle of goods that the household is observed purchasing, and $\boldsymbol{q}^f$ and $\boldsymbol{q}^m$ are the private consumption equivalents of the joint consumption of $\boldsymbol{q}$. Sharing $\boldsymbol{q}$ is equivalent to buying $\bar{\boldsymbol{q}} = \boldsymbol{q}^f + \boldsymbol{q}^m$ without sharing. As a start, a Barten or Gorman style joint consumption technology would be feasible.

Utility functions of men and women can be identified from single male and female households up to a monotonic transformation. From this knowledge and the observation of couples, the distribution within a couple and individual consumption of partners as well as the parameters of the joint consumption technology can be identified.[16] From the maximization program the following

[15] This also means that there are no externalities, e.g. from smoking.

[16] For a proof see Chapter 5, Section 5.4, p. 157.

equivalence scales can be calculated:

$$
\begin{aligned}
m^f &= \min_{\boldsymbol{q}^*} \left\{ (\boldsymbol{p}'\boldsymbol{q}^*)/\mu \mid U^f(\boldsymbol{q}^*) = U^f(\boldsymbol{q}^f) \right\} \qquad \text{and} \\
m^m &= \min_{\boldsymbol{q}^*} \left\{ (\boldsymbol{p}'\boldsymbol{q}^*)/\mu \mid U^m(\boldsymbol{q}^*) = U^m(\boldsymbol{q}^m) \right\},
\end{aligned} \tag{2.34}
$$

where $\boldsymbol{q}^*$ are the quantities consumed by an equivalent single household and μ is the total outlay of the reference couple. Equivalence scales are uniquely identified in the collective model, because a monotonic transformation $t(.)$ affects $U^f(\boldsymbol{q}^*)$ and $U^f(\boldsymbol{q}^f)$ in the same way: $U^f(\boldsymbol{q}^*) = U^f(\boldsymbol{q}^f) \Leftrightarrow t(U^f(\boldsymbol{q}^*)) = t(U^f(\boldsymbol{q}^f))$. The same applies to $U^m(\boldsymbol{q}^*)$ and $U^m(\boldsymbol{q}^m)$.

The framework is similar to a Becker (1965) type household production model, except that instead of using market goods to produce commodities that contribute to utility, the household produces the equivalent of a greater quantity of goods by sharing. This is essentially the technology behind the motivation of the Barten (1964) and Gorman (1976) approach to equivalence scales. However, while in those two approaches household decisions are made by a benevolent dictator, or – in the case of families with children – dictatorial parents, in the collective model, intra-household distribution can be modelled explicitly.

Browning et al. argue that it is a great advantage of the collective model that the usual argument of non-identification does not apply. As the model applies to the same person, living in different situations, a monotonic transformation of the utility function would apply to the utility function in both situations, hence the relationship between utilities in both situations is unique.

The model is not well suited to estimate equivalence scales for families with children. Even though Lewbel (2002) suggests that the model can be extended by comparing single mothers and single men with couples with children, this would stretch the model a bit far: if the model were to be applied to families with children, it would be consequent to assume a three way sharing rule that incorporates a share for the child. In addition, this approach would not allow for the estimation of an equivalence scale between a couple and a couple with a child: only an equivalence scale for the situation comparison between a single mothers and a mother living with the father can be determined. To estimate equivalence scales for families I prefer the Rothbarth model (chapter 3) and the Barten model (chapter 4). However, the model is perfectly suited for the estimation of equivalence scales between single men and women and couples. Therefore the model is developed in detail and estimated in chapter 5.

2.3 Survey Approaches

While economic approaches identify the different levels of welfare indirectly from observed demand data, survey methods attempt to measure welfare and its link to income and demography directly by asking respondents about needs or welfare evaluations. Survey methods have found many applications, e.g. in the estimation of poverty thresholds, in the assessment of the value of certain characteristics like being with a partner, being employed or unemployed or being disabled. The approach can also be used to estimate equivalence scales.

Two main survey approaches can be distinguished: the subjective and the consensual approach. In the subjective approach respondents are asked to consider which levels of income correspond to different standards of living, given their own actual household situation. The question can also be reversed to which standard of living corresponds to a certain level of income. In the consensual approach, interviewees are asked about hypothetical household situations, not about their own.

There are different varieties of the subjective approach. The first line of research is the so called Leyden School. In this approach, so called individual welfare of income functions are estimated. Respondents have to state those levels of income they consider correspond in their personal circumstances to a standard of living that is "bad", "sufficient", "good", etc. Respondents have to state income levels for hypothetical welfare levels as well as for their own welfare level which coincides with one of the response categories.

In a second line of research, the income satisfaction approach, respondents are asked to evaluate their own income or their own level of well-being on either a numerical or a verbal scale from very bad to very good. The advantage of the method is that respondents are asked exclusively about their own actual standard of living and not about hypothetical situations as in the Leyden approach.

The consensual approach is related to subjective methods, but relies on *opinions* (the "consensus") about living standards. Respondents are asked to evaluate the living standard in different demographic situations, e.g. for a single household, a couple and a couple with two children. Again, respondents can be questioned either on income levels necessary for a certain living standard or on the living standard that corresponds to certain income levels. In this method, both living standard and demographic situation are hypothetical situations. The consensual approach does not ask the subjective evaluation of the respondents own situation but their evaluation of some given situation.

The following section will give a general discussion of the mentioned survey methods.

2.3.1 Leyden School

The fundamental assumption of the Leyden approach is that each person has an individual welfare of income function which summarizes evaluations of different income levels on a bounded numerical scale (van Praag, 1968). The income levels that correspond to points on the scale are determined using surveys containing the so called Income Evaluation Question (IEQ). A typical form of the IEQ reads: *'What income would you, in your circumstances, consider to be "very bad", "bad", "insufficient", "sufficient", "good", "very good" (by income we mean after tax household income)'*[17] From the answers to this question individual welfare-of-income functions can be estimated that relate labels to points on the welfare scale.

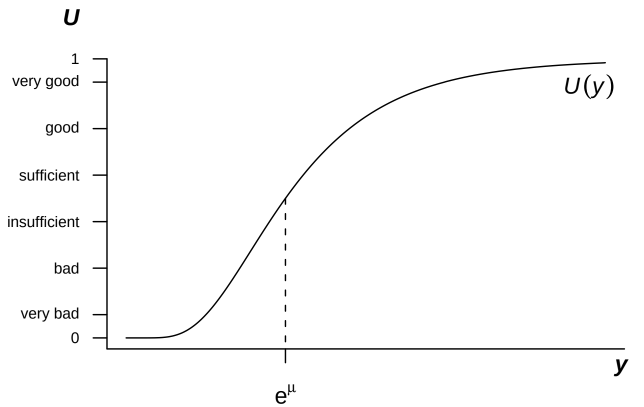

Figure 2.5: *Individual welfare function of income (Source: after van Praag and Kapteyn (1973))*

Figure 2.5 gives an illustration of an individual welfare function. Numerical welfare increases with income. Van Praag and Kapteyn (1973) justify the assumption of a bounded utility function as "reflecting the psychological reality that every individual is evaluating his income by comparing it with a worst position and a position of complete satiation although an infinite amount of money may not suffice to create such a situation." With bounded utility, the lognormal distribution function is a natural choice to be assumed as the functional form of the welfare function. The function can be conveniently normalized on a range from zero to one: zero represents the lowest possible level of well- (or in this case ill-) being, while one stands for perfect bliss. The welfare labels of the IEQ are evenly spaced on the midpoints of an equal number of intervals along the range. This "equal quantile assumption" can be

[17]See for example van Praag and Kapteyn (1973).

justified by the assumption that individuals try to maximize the information given by their answers.

A severe complication for the assessment of equivalence scales with the IEQ is that evaluations vary systematically with income. People with a higher actual income require a higher income to reach a certain welfare level, while respondents with a lower actual income also require a lower income to be on the same level of welfare. For example, when a household with a monthly income of 4000 € would consider an income of 3000 € as "good", an otherwise identical household with an income of only 1000 € would consider 2000 € as "good". Kapteyn and van Praag (1976) call this effect *preference drift*. Due to preference drift, there is a wide range of individual estimates of income equivalent to a given utility level u at each family size, depending on income. As a consequence equivalence scales cannot be identified without further assumptions.[18] One possibility to solve the problem is to select only that welfare-income relationship for which the equivalent income and the households own income coincide. It is assumed that there is no preference drift in this case. However, this solution would mean to give up a large part of the information contained in the data. In fact, this solution is close to the method followed in the income satisfaction approach. In order to preserve all information, the Leyden approach follows a different route.

A lognormal distribution function is assumed to represent the functional form of the individual welfare function:

$$U(y) = \mathscr{N}((log(y) - \mu)/\sigma; 0, 1), \tag{2.35}$$

where y is the hypothetical household income necessary to reach a certain utility level and the parameters μ and σ depend on income and other characteristics. The dependence of μ on actual income z is the preference drift. For expositional reasons σ is assumed to be constant.[19] To map responses

[18] To see this, consider the following example: Assume that households with two and four persons are to be compared, and the "true" value of the equivalence scale is 1.5. Further assume that any two person household with an income of €1500 considers an income of €2000 to be "good", while with an income of €1000 it would require only €1800. These values have to be multiplied by the true equivalence scale for four person households: With an equivalence scale of 1.5, €1500 in a larger household are equivalent to only €1000 in a smaller household. Therefore the larger household will consider $1.5 \cdot$ €1800 = €2700 (the equivalence scale times the equivalent income considered to be "good" by a €1000 equivalent household) to be "good"; with an actual income of €2250 it had to be €3000. Now assume that all surveyed households, small and large, earn the same income of €1500. The calculated naïve equivalence scale between the two household types would be the relation of the incomes considered to be "good" by both household types (m_s^r = €2700/€2000 = 1.35), which is biased due to the preference drift. The true equivalence scale can be estimated only if answers of equivalent households are compared, for example the €1000 two person and the €1500 four person household.

[19] In general σ can vary with household characteristics in the same as μ and is also estimated in the same way.

on utility levels, the Leyden approach makes an 'equal quantile' assumption, so that all response categories are equally spaced over the range from zero to one, as shown in Figure 2.5. The six descriptions given in the IEQ above correspond to the midpoints of six equal intervals between 0 and 1.

To correct for preference drift, $\hat{\mu}$ can be determined from a regression of individual μ's on the log of household size h^s and actual household income z^s:

$$\hat{\mu} = \beta_0 + \beta_1 \log h^s + \beta_2 \log z^s \tag{2.36}$$

$\beta_2 \neq 0$ implies the presence of preference drift.

At some reference utility level u for household s and reference household r, $\log y^s - \mu^s$ and $\log y^r - \mu^r$ have to be equal for both households to have the same welfare level. Substitution of $\hat{\mu}$ for μ leads to:

$$\begin{aligned} \log y^s - \beta_0 - \beta_1 \log h^s - \beta_2 \log z^s &= v \\ \text{and} \log y^r - \beta_0 - \beta_1 \log h^r - \beta_2 \log z^r &= v \end{aligned} \tag{2.37}$$

Apparently, there is no unique relationship between equivalent income and family size when preference drift is present, because the income y^s necessary to reach utility level u depends on actual respondents' income z^s. The identification problem is resolved by substituting y^s for z^s and y^r for z^r in equations (2.37) under the assumption that there is no preference drift at the actual income level. Solving for the equivalence scale then leads to:

$$m_s^r = \frac{y^s}{y^r} = \left(\frac{h^s}{h^r}\right)^{\frac{\beta_1}{1-\beta_2}} \tag{2.38}$$

By construction, the scale does not depend on the welfare-level. However, the method can be extended to depend on the base level of utility, if the σ parameter is allowed to depend on the household type.

A variant of the Leyden approach is the use of the minimum income question (MIQ) instead of the income evaluation question (IEQ). The minimum income question typically reads as: *"We would like to know which net family income would, in your circumstances, be the absolute minimum for you. That is to say, that you would not be able to make ends meet if you earned less."* Answers to the MIQ represent one point on the individual welfare of income function. The problems related to the MIQ are very similar to those of the IEQ. Preference drift can be corrected in the same way. In contrast to the IEQ, the MIQ allows only equivalence scales at some subsistence level to be estimated.

The main criticisms of the Leyden approach of the measurement of welfare are the cardinality assumption of the welfare function as well as some properties the functional form: the boundedness on the range from zero to one, the equal spacing of labels and an increasing marginal utility at low income levels (Seidl, 1994). Rejection of the cardinality assumption and doubts about the functional form can lead to problems in applications where the *difference* of utility levels is relevant. Van Praag and van der Sar (1988) show that it is not crucial in the context of equivalence scale estimation, where only households of the *same* utility level are compared. Here, full ordinal comparability is sufficient. To be able to make this comparison one has to accept the (untestable) assumption that different types of households understand the IEQ in the same way, just as one has to accept the (equally untestable) assumption that households have the same preferences when economic equivalence scales are estimated.

However, it is arguable whether different households do understand the IEQ in the same way. It might well be that households of different types do not understand the given response categories in an absolute sense but in a relative sense, comparing their own situation to households of some reference group that might consist of households of the same type. Bradbury (1989) claims that if such reference group effects exist, i.e. if a household evaluates how it is doing financially relative to other households in the same situation, instead of assessing its income on an absolute scale, then the Leyden approach to equivalence scales is not a sensible one. In this case scales simply represent the relation of average income levels of the reference and the compared households. If average incomes do not depend on household composition, equivalence scales would equal unity.

Reference group effects can be responsible for the fact that – compared to other methods – scales derived from the Leyden approach are often rather flat: the income needs of households do not increase much when family sizes increase. For example Melenberg and van Soest (1996) find implausibly flat scales using the IEQ approach on Dutch data, while scales that are estimated using the income satisfaction approach (see below) are much higher.[20]

Flat scales can also be explained by preference and life-style changes. Goedhart et al. (1977) argue that when family structure changes, families with children get accustomed to the new situation. E.g., while a formerly childless couple was used to relatively high vacation expenditures, after having children, it spends its holidays at home. Savings from lower spending on vacation can be used to compensate for the additional expenditures associated with children, so that the family gives its income almost the same evaluation as before – the respective equivalence scale would be close to one. This interpretation supports the plausibility of flat equivalence scales that are based on the IEQ. It is questionable though, how far such a compensation can go. To test the argument, it would be necessary to combine IEQ and expenditure data.

As reported in Table 2.1, scales for Germany that were generated with the traditional Leyden approach and the individual utility of income function are indeed flatter than the average scale. On the other hand, IEQ scales

[20] See Schröder (2004) for an exhaustive list of studies that estimate IEQ scales.

that relax the strict assumptions are not very different from other scales at a medium income level.

Scales reported for Western Germany by Plug et al. (1997) show some income elasticity when results from the MIQ and the IEQ are compared, assuming that MIQ results apply to low incomes while IEQ results reflect some average equivalence scale. In so far the problem of implausibly flat equivalence scales does not seem to apply at least to German data. This is counterevidence to the presence of reference-group effects, because otherwise scales would have to be much flatter. Results of the method are quite sensitive to the specification of the estimated equations: for instance van Praag and van der Sar (1988) and van Praag et al. (1982) find quite different results while using exactly the same data set.

2.3.2 Income Satisfaction Approach

The Income Satisfaction Approach is another subjective method that can be used for the estimation of equivalence scales. Its basic principle of relating income to a verbal or numerical representation of a welfare level is akin to the Leyden approach. Here, however, the question is reversed and respondents are asked about their satisfaction with current income (IS). They have to rate their income on a pre-specified welfare scale with a number of response categories. The question in the German Socio Economic Panel reads, for example: *How satisfied are you today with the following areas of your life? Please answer by using the following scale, in which 0 means totally unhappy, and 10 means totally happy. If you are partly happy and partly not, select a number in between. How satisfied are you... with your household income?*

In contrast to the Leyden approach and to economic approaches, no underlying model of the utility function is needed in the IS approach. The relationship between satisfaction with income and demographic characteristics of the household is specified and the parameters in this relationship are estimated using an ordered response model. The equivalence scale is the ratio of those incomes that create the same feeling of satisfaction about the household's situation given different household characteristics. Melenberg and van Soest (1996) find that the results based on satisfaction data lead to much higher equivalence scales than results obtained from the IEQ. They suggest that a reason for this finding might be that the satisfaction question refers to the household's actual situation, whereas the IEQ asks for information in some virtual situations that is subject to bias from preference drift. However, reference group effects that can create a bias in the Leyden approach, act here too: if respondents tend to compare their situation to households from their own group, estimated equivalence scales could become meaningless because of reference group effects.

As an extreme example, assume that satisfaction with income were only determined by a household's position in the income distribution of its type.

Then equivalence scales would merely reflect the relative incomes of different household types. The median earner of type A would be as satisfied as the median earner of type B and the equivalence scale between types A and B would be the relation of the two median incomes. A meaningful equivalence scale would require respondents to answer on the basis of what they can afford or how they manage with their income compared to some common reference level.

Even with the same reference group, every respondent may have a different reference point from which she evaluates satisfaction with income. Maybe some people are by nature happier and more satisfied than others. If these happier people were more likely to found a family and have children, then equivalence scales would suffer from a selection bias. Scales were biased downwards, because respondents in larger households families tended to be more satisfied on average than singles.

Using panel data, this problem can be solved in a fixed effects model. When the same person is observed over time in different demographic situations, the "happiness predisposition" of this person can be extracted as a fixed effect and the bias eliminated. Charlier (2002) and Schwarze (2003) apply this idea to data of the German Socio Economic Panel (GSOEP).[21] They find *lower* scales for fixed effects models than for pooled models. This result runs exactly counter to the explanation above: unhappy people tend to have more children than happy people. Unfortunately, both studies use a specification that allows only for base independent equivalence scales, so that the effect of income on equivalence scales cannot be extracted.

Charlier also estimates equivalence scales for two types of satisfaction questions: one for satisfaction with income and one for satisfaction with life. He finds scales of a size similar to other methods[22] for satisfaction with income and scales close to one for satisfaction with life. The explanation for this result could be that the latter scales measure unconditional welfare: when parents decide to have children, they are indeed happy with them, given the expected income situation with children. In comparison, when they evaluate their income they do see the actual costs of their children and their own reduced material welfare. This is reflected in the answers to the satisfaction with income question. Another, less positive explanation for flat equivalence scales would be a change in the response behaviour. As soon as they get used to a new situation, people tend to evaluate themselves at some average level of satisfaction: An increase in income would lead to a higher satisfaction at first, but this effect would fade over time and after a while the person would report the same satisfaction level as before.[23] However, this effect does not

[21] Results are reported in Table 2.1.

[22] See Table 2.1.

[23] This would mean for example that, if there are no recent positive or negative events, a person would always respond with a "6" on the 0 to 10 scale of income or life satisfaction.

explain the difference between the rather high IS scales and extremely flat satisfaction with life scales.

There are still some open questions regarding the IS approach, especially regarding the cause of the bias that is detected by the fixed effects model. It would also be interesting to estimate equivalence scales that are not independent of base by construction, to see if this affects the results. Concerns about reference group effects are not confirmed in practical applications, and problems with changing response behaviour can be answered by using time series data that reveal the immediate responses to situation changes. All in all, the method generates some quite promising results, if used with time series data.

Of course it is certainly not good practice to judge the method only from the size of the scales that it generates. That would mean to endorse or dismiss a method only on the grounds of its results coinciding with some preconceived opinion. One could as well set the scales arbitrarily. Or poll a sample of people on their opinion about equivalence scales. That is what the next method does.

2.3.3 Consensual Approach

The consensual approach is directed at scrutinizing the public opinion (the consensus) about incomes and living standards of different household types in the community. Respondents are asked to state the amount that a given hypothetical household needs in order to reach a certain specified living standard. The answers can be used to define poverty lines, for example. Equivalence scales can be calculated easily, when the survey includes necessary amounts for different household types.

Surveys differ mainly with regard to the stimulus material (here the situation description of different household types), the type of questions and the response categories. E.g. respondents are asked about the amounts necessary for different household types to attain certain pre-defined living standards: poor, getting along, being comfortable, prosperous, substantial and rich (*Boston Social Standards Survey*, see Rainwater, 1974). In an alternative set of questions, respondents are asked to rate given incomes of different family types on a poor/prosperous scale (Dubnoff, 1985).

Schröder (2004) uses a new method that is tailored to the estimation of equivalence scales, where stimulus material and response are money metric. Respondents are asked which income would be necessary for households of a given demographic composition (a childless couple, a couple with one to three children and a single with one to three children) to reach the same welfare as a reference household (a childless single) with a given income. The advantage of the method is that – in contrast to a reference situation that is described as “poor”, “prosperous” or “sufficient” – people cannot differ about the interpretation of the amount of money that is given as reference living standard.

Scales estimated using the consensual approach are correct if respondents have a correct perception of the cost of living in different situations. Then the method can be seen as some kind of "pooled expert approach" where many "experts" are asked about their opinion on the matter. Errors in the experts' judgements can be corrected, if a large number of experts is queried and if errors cancel out on average.

It is possible to test for a possible bias by scrutinizing the correlation of respondents' personal characteristics and their answers. If the average father of three gives the same answers as the average bachelor, then scales are presumably not biased: the father has some experience of the cost of living in a larger family and should be able to adjust his conception of these costs. If he gives the same answers as the bachelor, then the bachelor can evaluate the cost of living of a hypothetical household even without further experience. The same argument applies with respect to income.

Schröder (2004) carries out this test and finds some significant correlation between household characteristics and scales as well as between income and scales, but the effect is numerically small. Still, this sheds some doubt on the method. Even if there were no such correlation, it would not rule out that the consensual approach measures the normative judgement of what a society deems adequate rather than equivalence scales that reflect actual needs. This is very difficult to test without additional data. It would be interesting if equivalence scales are correlated with scales that are defined through the level social benefits in a sample of different countries. A high correlation would suggest that scales are determined by norms rather than needs, while a low correlation would suggest that households look to their own experience for an answer.

The consensual method has its problems, and some more research on the validity of scales is certainly necessary. A great advantage is the ease of data collection and computation. The method also allows for the estimation of cost of characteristics like employment status, which is impossible with most other methods and economic methods in particular. It is also less difficult to inquire into the relationship between income and the size of equivalence scales. Koulovatianos et al. (2004, in press) and Schröder (2004) find that consensual scales do depend substantially on income. They show this for a wide range of different countries.

2.4 Some Equivalence Scale Estimates for Germany

In this section, results of previous research on equivalence scales for Germany is reported, as well as the scales that will be calculated throughout the remainder of this work. The number of results (shown in Table 2.1) is fairly limited, but all approaches that have been discussed are present. Some studies report only household size without reference to the number of adults and children, but most studies explicitly report scales for families with children. For these, a number of heads between 3 and 6 implies a couple with one to four children, respectively. A childless couple is the reference household. The table is organized with respect to the different approaches, where results from the expert approach come first, survey methods second and economic equivalence scales last. In what follows, numbers in square brackets indicate the respective line in Table 2.1.

The OECD scales [1] and the scales implicit in the German social assistance system [2] are expert scales. Both scales are more a rule of thumb and contain some simplifications. The old OECD scales [1b] state that for calculating a needs adjusted household income, the first adult should be assigned a value of one, subsequent adults are assigned a value of 0.7 and children 0.5.[24] This scale was suggested by the OECD (1982) for possible use in "countries which have not established their own equivalence scale". After having used the old OECD scale in the 1980s and the earlier 1990s, the Statistical Office of the European Union (EUROSTAT) adopted in the late 1990s the so-called modified OECD equivalence scale [1b]. This scale, first proposed by Hagenaars et al. (1994), assigns a value of 1 to the household head, of 0.5 to each additional adult member and of 0.3 to each child. The German social assistance system works in a similar fashion. A single adult has a value of one. Two adults living together have a value of 0.9 each, children are differentiated by age: children of age 0–6 have a weight of 0.5, ages 7–13 get 0.65 and adolescents of age 14–17 have weight 0.9. Additional adjustments are made for single parents. Scales for households with children aged 7–13 are reported in Table 2.1. The social assistance scales do not contain housing cost. As there are substantial economies of scale in housing, the reported scales are much higher than any other scales reported in Table 2.1 and cannot be used without further adjustments. For the social assistance scales the year of introduction is reported.

Charlier (2002) and Schwarze (2003) use an income satisfaction model with panel data from the German Socio Economic Panel (GSOEP). Both estimate a fixed effects model. Charlier reports a general scale where there is no distinction made between children and adults [3a] as well as scales for families with one, two and four children aged 12, 12 and 6 and 18, 12, 6 and

[24] Division by 1.7, the needs of the two-adult reference household gives the values indicated in Table 2.1.

	Author	Method	Year	Part	Income Level	Number of Heads 1	3	4	5	6
1a	OECD (old)	Expert	(1982)	–	IB	0.59	1.29	1.59	1.88	2.18
1b	OECD (modif.)	Expert	(1994)	–	IB	0.67	1.20	1.40	1.60	1.80
2	Social Assistance	Expert	2004	G	L	0.56	1.36	1.72	2.08	2.44
3a	Charlier	IS/FE	1984-91	W	IB	0.70	1.22	1.40	1.56	1.70
3b	Charlier	IS/FE	1984-91	W	IB	0.70	1.20	1.30	1.42	1.54
3c	Charlier	IS/RE	1984-91	W	IB	0.67	1.26	1.44	1.61	1.76
3d	Charlier	IS	1984-91	W	IB	0.67	1.23	1.38	1.57	1.76
4a	Schwarze	IS/FE	1992-99	G	IB	0.81	1.13	1.23	1.31	
4b	Schwarze	IS/FE	1992-99	G	IB	0.78	1.10	1.15	1.16	
4c	Schwarze	IS	1992-99	G	IB	0.79	1.15	1.26	1.37	
4d	Schwarze	IS	1992-99	G	IB	0.75	1.13	1.19	1.19	
5a	Plug et al.	IEQ/Leyden	1992	W	IB		1.09	1.16	1.21	1.26
5b	Plug et al.	MIQ	1992	W	L		1.12	1.21	1.29	1.36
6	van Praag et al.	IEQ/Leyden	1979	W	L	0.83	1.12	1.21	1.29	1.35
7a	v.Praag&v.d.Sar	IEQ	1979	W	L/M		1.16	1.29	1.39	1.49
7b	v.Praag&v.d.Sar	IEQ	1979	W	H		1.13	1.23	1.31	1.38
8	Riffault&Rabier	Consensual	1976	W	L	0.68		1.40		
9a	Schröder	Consensual	1999	W	L	0.57	1.29	1.55	1.81	
9b	Schröder	Consensual	1999	W	M	0.67	1.15	1.28	1.41	
9c	Schröder	Consensual	1999	W	H	0.72	1.07	1.14	1.21	
10	Merz and Faik	Engel	1983	W	IB	0.55	1.21	1.35	1.53	1.68
11	Merz and Faik	ELES/Barten	1983	W	L	0.68	1.17	1.28	1.34	1.30
12a	Faik	Rothbarth	1969	W	M		1.09	1.18	1.27	
12b	Faik	Rothbarth	1969	W	L		1.19	1.39	1.58	
13	Missong	IB	1993	W	IB	0.68	1.20	1.33	1.31	
14	Missong	Translating	1988/93	W	L	0.60	1.28	1.43	1.54	
15	Chapter 3 [1]	Rothbarth	1993	W	IB		1.21	1.38	1.53	1.67
16	Chapter 4 [2]	Barten	1993	W	M		1.13	1.22	1.40	
17	Chapter 5 [3]	Collective	1993	W	M	0.70/0.72				

Table 2.1: *Equivalence Scales for Germany. A childless couple is the reference household with an equivalence scale of 1.00. Part indicates if data are from East and West Germany (G) or West Germany only (W). Income level are: L: Low Income; M: Medium Income; H: High Income; IB: Independent of Base. Year indicates the year of the data set, with the exception of OECD scales, where it refers to the year of publication. For further explanation see text.* [1] *Table 3.1, p. 65, "adult goods".* [2] *Table 4.8, p. 111, model BaC.* [3] *Table 5.3, p. 168, model C.*

1, respectively. Scales for a three person household are my own calculations based on Charlier's results for a household with children aged 15, 9 and 3.[25] Child age dependent scales are estimated with fixed effects [3b], with random effects [3c] and pooled [3d].[26] Scales depend strongly on the age of children. In combination with older children in larger households, this leads to reported scales that seem to increase more for a third and fourth child than for a second child. This is only an effect of the higher average age of children in larger households: In the fixed effects model [3b], scales for households with one to four children, where all children are 9 years old, would be 1.16, 1.29, 1.40 and 1.50, respectively. It is remarkable, that scales are *lower* in the fixed effects model. This means, that people living in a larger household are inherently less happy (or have a natural tendency to report lower scores) than those living in a smaller household. Otherwise Charlier's scales are rather high compared to other survey and economic scales. Only expert scales and Schröder's (2004) result for low incomes are higher.

The result that fixed effects estimates [4a/b] are lower than pooled estimates [4c/d], is reproduced by Schwarze on a later wave of the same data set. He uses a rather peculiar specification for his model: equivalence scales are estimated indirectly by assuming a constant equivalence scale elasticity, (models [4a] and [4c]) or an elasticity that is decreasing in the number of heads in the household (models [4b] and [4d]).[27] This specification is certainly too restrictive, in particular, because there is no distinction made between a partner (the second person) and children. As a result Schwarze's results are very low, in particular for larger families.

Plug et al. (1997) estimate equivalence scales for East and West Germany separately (only scales for West Germany are reported in Table 2.1). Scales are estimated using the income evaluation question [5a] as well as the minimum income question [5b]. Both approaches lead to scales that are independent of the base level of utility. Nevertheless, scales from the MIQ are listed here as applying to a low income, because the question refers explicitly to the *minimum* income necessary to make ends meet. Indeed, MIQ scales are somewhat higher than IEQ scales, indicating lower economies of scale at lower incomes. However, results for East Germany (not reported) do not show this difference, possibly because of the lower average income level in Eastern Germany.

Van Praag et al. (1982) do not calculate equivalence scales, but they

[25] Ages were chosen for an age structure in between the 2 and 4 person household.

[26] In the random and fixed effects models inter-individual differences in the scaling and anchoring of the satisfaction responses are captured. The random effect is independent of the covariates, while the fixed effect is not. The pooled model is estimated as a reference.

[27] With a constant elasticity, the equivalence scale takes the form $m_s^r = h^e$, where h is the number of heads in the household and e is the constant equivalence scale elasticity, that is to be estimated. The decreasing elasticity takes the form $e = a - bk$, where a and b are parameters and k is the number of children.

do estimate household type specific poverty lines using the IEQ for eight European countries, including Germany. Poverty lines can then be used for the calculation of equivalence scales (Scales shown in line [6] are my own calculations). They calculate a welfare of income function and then report the incomes that correspond to a welfare level of 0.4 and 0.5. The former number corresponds to a level between "insufficient" and "bad", while the latter corresponds to a level between "insufficient" and "sufficient". Reported equivalence scales correspond to the lower level, but the difference between scales is small.

The same data set is used by van Praag and van der Sar (1988) to estimate equivalence scales [7]. The aim of the work is to show that it is possible to estimate equivalence scales using the IEQ without some of the most disputed assumptions of the Leyden approach. Therefore van Praag and van der Sar relax the cardinality assumption inherent in the approach. They do not estimate the individual utility of welfare function but calculate equivalence scales separately for each of the six levels of the IEQ in the same way as scales are calculated using the MIQ. Due to the different assumptions, resulting scales are much higher than those of van Praag et al. (1982) and Plug et al. (1997). Because scales are evaluated separately for each response category, they can depend on the reference welfare level. German scales are identical for the levels "very bad" and "sufficient" [7a] and only slightly different for the levels in between, while they are markedly lower for the levels "good" and "very good" [7b].

In a comparative study on poverty in the then nine members of the European Community, Riffault and Rabier (1977) estimated consensual equivalence scales at a minimum income level. Respondents were asked to specify the income that seems to them the minimum to make ends meet for a single person of 30 to 50 years, for a couple of 30 to 50 years and for a family of four: the man, the woman and two children of 10 to 15 years. Scales are calculated on the basis of median answers. Scales for Germany are reported in line [8]. The range of scales for all member countries (Belgium, Denmark, Germany, France, Ireland, Italy, Luxembourg, the Netherlands and the U.K.) is quite wide with 0.64–0.76 for single households and 1.24–1.43 for a family of four. The lowest cost of children are found in the Netherlands, while the highest are reported for France and Italy.

The main concern of the work by Schröder (2004) [9] is not the absolute level but the income dependence of equivalence scales. Schröder uses the consensual approach to evaluate the opinion of respondents on the income necessary for different household types to reach the same welfare level as a single adult with an income of DM 1,000, DM 2,500, DM 4,000, DM 5,500 and DM 7,000, respectively. The approach is tailored to the assessment of equivalence scales at different welfare levels. He finds scales that are rather high for low income households at the DM 1,000 reference level [9a], but scales fall dramatically with increasing income: Scales at the DM 2,500 level are already much smaller [9b] and scales continue to fall up to the highest reference level [9c]. Only scales for the lowest income level are clearly higher than other reported (non-expert) scales. This might be due to the fact that other methods are less sensitive to the income dependence of equivalence scales. IB scales do not depend on income by construction and therefore reflect some average equivalence scale, but other scales show a rather low sensitivity as well, because the estimation is best around the centre of the data, which is far removed from these low income levels.

Merz and Faik (1995) estimate Engel equivalence scales [10] and scales using the extended linear expenditure system (ELES, [11]). The estimated Engel scales are among the highest of all income independent equivalence scales in this collection of surveys. This reiterates the common result that Engel scales are rather high. The shown scales are calculated using total food expenditures. Merz and Faik also calculate Engel scales for basic foods and a basic basket of goods, which comprises the categories food, clothing & shoes, housing & energy and body & health care. Scales for basic food are substantially higher: with a range of 1.31 to 2.17 for families of three to six persons, they are very close to the old OECD scales. Scales for the basic basket of goods are much lower with a range from 1.18–1.55 for the same households. They are closer to the ELES scales. This result supports the claim that Engel scales shouldn't be used. First, it is not clear, which good is the correct good for estimation. Different goods give different results, so that the choice of the scale is arbitrary. Second, to be theoretically valid, all goods should give the same results, but that is not the case.[28]

The situation is different for the ELES scales. The specification used by Merz and Faik is of the Barten type. Again, they also estimate a detailed demographic decomposition. With a value of 0.97 at the subsistence level the scale for men relative to women is now not significantly different from one. It is falling with rising income down to a value of 0.92 at 1.5 times the median income. In principle, the specification allows for scales to depend on income, but in fact they are almost constant. This might be due to the very restrictive linear Engel curves, but my own somewhat less restrictive Barten model gives almost constant equivalence scales as well.

Faik (1995) tries out the full range of possible economic equivalence scales. He estimates Engel and Rothbarth scales, as well as ELES scales in a translating, Barten and Gorman form. He even estimates Prais-Houthakker scales by fixing one good specific scale factor a priori, taking this information from the Barten estimates. Engel and Barten scales have been published in the above-mentioned paper together with Merz. Rothbarth scales are shown in

[28]Engel scales rest on a very weak basis, but Merz and Faik clearly overstretch the explanational abilities of the model by calculating highly disaggregated scales, where they compare even single men and women. They find an equivalence scale of 1.26 for a single man relative to a single woman for food, but 0.80 and 0.87 for basic foods and the basic basket, respectively. This is certainly not very informative.

Table 2.1 [12]. Unfortunately the only survey available to Faik that included adult goods (tobacco and alcohol) was the German Income and Expenditure Survey (EVS) of 1969. Faik reports scales for both goods, but individual children's weights for tobacco are very low or even negative. Faik estimates scales where the absolute costs of a child do not depend on income. As a consequence, scales are falling in income by construction. Scales for families with children 7–11 years of age that were calculated from alcohol consumption, are shown for two incomes: the mean income minus one standard deviation, which is 8219 DM/year [12a] and for the mean income: 17686 DM/year [12b]. Younger *and* older children have lower scales with 1.06 and 1.08 , respectively, at mean income.

Missong (2004) estimates IB scales [13] and scales based on demographic translating [14]. IB scales are estimated using log-quadratic Engel curves from the 1993 Income and Expenditure Survey for four commodity groups: education, clothing, mobility and other. Scales are estimated separately for each commodity group (not reported here) as well as in a system of equations with a restriction of equal scales for each group imposed. In a test, the IB assumption is rejected for the system.

Missong reports a wide variety of scales for different specifications. Some of his IB Scales are almost not increasing with the number of persons in the household and some even deviate from the principle of monotonicity: scales for a couple with one child are actually lower than those for a childless couple[29]. A model with differentiated adult age groups violates monotonicity for young parents, but not so much for older parents: scales for households with adults between the age of 40 and 60 are therefore reported in table 2.1, line [13][30]. Missong also experiments with a differentiation of age groups for children and adults and finds a significant influence of age on equivalence scales. A model with differentiated children's age groups still violates monotonicity between household sizes, but not between age groups: according to the model older children are more expensive than younger children.

Missong (2004) also calculates translating scales [14] within a quadratic expenditure system (QES) using the 1988 and 1993 cross sections of the German Income and Expenditure Survey and a system of seven commodity groups: food, clothing, personal care, education, mobility, housing and other. Due to the translating approach which assumes a constant absolute cost for any additional member of the household, reported scales at the subsistence level are rather high compared to other economic methods, but they are falling with increasing income. The estimated annual subsistence level for a childless couple is about DM 26,000.

Finally, my own results are reported. Rothbarth scales [15] are estimated

[29]Scales that violate monotonicity were estimated at values of 0.63, 0.97, 1.01 and 1.11 for a household of one, three, four, five and six persons, respectively.

[30]Missong (2004), table on p. 166.

in chapter 3. These scales are very high and almost identical in value to the food Engel scales of Merz and Faik. This is surprising as it runs counter to the conventional wisdom that Rothbarth scales are generally lower than Engel scales. In chapter 4 a Barten model is estimated in a quadratic expenditure system [16]. To identify the demand system, some of the same scale factors are fixed that are also used in the Rothbarth method, but because the Barten method allows for substitution effects, estimated scales are significantly lower than Rothbarth scales. In my opinion, both methods do not sensibly allow for the estimation of equivalence scales between single men and women and couples. Therefore, I estimate a collective model [17] in chapter 5. Because in this model needs are assessed separately for partners in a couple, two scales are given for one person households: one for women (0.70) and one for men (0.72). The scales are evaluated with reference to a couple with an equal distribution of income between partners. More detailed results are presented in the respective chapters.

So many Different Results?

Table 2.1 shows a wide range of results: Scales vary from 1.07 [9c] to 1.36 [2] for a family of three and from 1.26 [5a] to 2.44 [2] for a family of six. This is certainly not a satisfactory answer to the questions of policy makers and researchers. So what are the true equivalence scales then? In fact the range of results is not as wide as it first appears. First, if the *ad hoc* OECD and social assistance scales are left aside, the upper bound is reduced to 1.29 and 1.76, respectively. Then, there are several methods that are not independent of base and therefore generate a range of results depending on the reference income level. To find the "average income" equivalence scale, one should look only at the middle income results for these methods. Engel scales should be disregarded, because they tend to overstate equivalence scales and it is not clear, which goods basket should be used for estimation. Faik's Rothbarth scales are based on very old data, and are therefore not included. The result by Merz and Faik for the ELES and a family of six should be dismissed for the same reason. Schwarze's results should also be omitted, because the underlying model is too restrictive. Charlier presents a range of results; the most detailed result for families with children is picked [3b]. What remains are the results listed in Table 2.2

The band of scales for single person households is now very small. However, the list still shows a considerable range of equivalence scales for families with children, with the highest cost of a child about twice the estimate of the lowest. Notably the Leyden approach shows very low numbers, while the income satisfaction and the Rothbarth approach of chapter 3 show rather high scales. Maybe the average of all selected scales is the best approximation of the size of the "true" equivalence scale.

As for the income dependence of equivalence scales, no definite answer can

	Author	Method	Number of Heads				
			1	3	4	5	6
3b	Charlier	IS/FE	0.70	1.20	1.30	1.42	1.54
5a	Plug et al.	IEQ/Leyden		1.09	1.16	1.21	1.26
7a	v.Praag&v.d.Sar	IEQ		1.16	1.29	1.39	1.49
9b	Schröder	Consensual	0.67	1.15	1.28	1.41	
13	Missong	IB	0.68	1.20	1.33	1.31	
15	Chapter 3	Rothbarth		1.21	1.38	1.53	1.67
16	Chapter 4	Barten		1.13	1.22	1.40	
17	Chapter 5	Collective	0.70/0.72				
	Mean		0.69	1.16	1.28	1.38	1.49

Table 2.2: *"Best" equivalence scale estimates for Germany.*

be given. A large number of scales is not income dependent by construction, and a good part of the other scales is not sensitive to any income dependence. Schröder (2004) is the only reported study that is directly concerned with income dependence.