Summary

This paper illustrates a method for estimating the equivalence scales based on the results of a survey carried out by Banca d’Italia on the household budgets. The method was proposed by Kot in 2002. Here we formulate a different procedure on the base of the quantiles. Then we estimate the scale through maximum likelihood method. Finally we propose a confidence interval for the elasticity of the equivalence scale.

Key words: equivalence scales, Dagum distribution, maximum likelihood estimator.

0. Introduction

For many purposes household income has to be adjusted to household composition A way to do it is to utilize a household equivalence scale which divides the income by an index which reflects the relative housing expenditures needs. The method chosen for estimating the equivalence scales may influence comparisons of income distributions.

Kot (2002, 2002b) proposed to analyse social welfare treating the behaviour of a complex system as a whole. He showed how the unobservable distribution of welfare could be inferred from the distribution of income. He defined the income evaluation social function (IES) as the tool to summarize society attitude towards the distribution of income. Kot considered also the situation in which there were two demographic variables. Here we will consider only one variable.
The aim of this paper is to estimate the equivalence scales in Italy in order to compare the income of households having different demographic attributes. We propose a different procedure which leads to the same finale formulation of the one proposed by Kot. Here we start by considering the equality between the quantiles of the distribution in the group considered and the quantiles of the distribution considered as reference by the researcher.

In section 1 we summarize briefly the most important classification of equivalence scales. In the following section it is explained the method of estimation. We propose a confidence interval for the scale value.

In Section 3 we describe the results obtained for the Italian situation.

1. Definition of equivalence scale

An equivalence scale is a devise to convert the income of units having different demographic profiles into a common base measuring income-unit purchasing power. First of all it is necessary to calculate the income conversion coefficient and then to divide the unit income by this coefficient. There are several methods proposed to estimates the equivalence scales. The choice of the kind of scale does affect the shape of the household income distribution for research purposes and it does have implications for policy purposes involving the income distribution.

Equivalent scales are commonly used in the studies about the poverty of a region (see, f.i. Carbonaro, 1985).

The original idea on which is based Engel’s method is that the share of budget devoted to food decreases as the number of family components increases. Many authors suggested that the share adopted as welfare indicator should include a wider number of goods.

A rigorous way to estimate household equivalence scale is to use an econometric method based on utility theory (see f.i. Michelini, 2002). It is also interesting to consider the cost that families meet in maintain their children (see Vernizzi, Siletti, 2004) for both food and other goods.

2. Description of the method

Let income $X$ be the r.v. with Dagum\(^{(1)}\) type I distribution (see, f.i., Dancelli (1986), Dagum (1990)).

The model proposed by Dagum fulfils many properties considered relevant to a model of income distribution: model specifications explain the economic reality, the convergence to the Pareto law, economic significance of the parameters. In the present paper, the choice of Dagum model is also supported by the fact that it provides a good fitting to both the low income part and the high income part of the distribution in Italy (Latorre, 1989).

Let $M$ be a socio-demographic variable which takes values in a set and has known probability function $p(\cdot)$. Here $M$ indicates the number of household members.

\(^{(1)}\) Kot (2002) uses Burr III type distribution. Dagum I type distribution used in the present paper may be reconducted to Burr distribution.
Let $X_m$ denote household income in the group in which the demographic variable takes value $m$. The cumulative distribution function (c.d.f.) of the r.v. $X_m$, that is the conditional income distribution given the size $m$ of the family, is:

$$F(x | m) = \frac{1}{\left[ 1 + \frac{\beta_0 m^{-\beta_1}}{\alpha - 1} x^{-(\alpha - 1)} \right]^\gamma}$$

Let $X_0$ be the income of the family taken as reference. In the present paper we chose the one-member reference household group, that is $m_o = 1$.

If we put $m = m_0$ in (1), we obtain the c.d.f. of the r.v. $X_0$

$$F(x | m_0) = \frac{1}{\left[ 1 + \frac{\beta_0 m_0^{-\beta_1}}{\alpha - 1} x^{-(\alpha - 1)} \right]^\gamma}$$

Equalizing (1) and (2), we obtain

$$\left[ 1 + \frac{\beta_0 m^{-\beta_1}}{\alpha - 1} x^{-(\alpha - 1)} \right] = \left[ 1 + \frac{\beta_0 m_0^{-\beta_1}}{\alpha - 1} x_0^{-(\alpha - 1)} \right]^\gamma$$

it follows that

$$m^{-\beta_1} x_m^{-(\alpha - 1)} = m_0^{-\beta_1} x_0^{-(\alpha - 1)}$$

Equation (3) allows to find the relation between the quantiles of a fixed order in the group of $m$ components and in the one with $m_0$ components.

From (4) it is easy to find

$$\frac{x_m}{x_0} = \left( \frac{m}{m_0} \right)^{\frac{\beta_1}{\alpha - 1}}$$

Let

$$\varepsilon = -\frac{\beta_1}{\alpha - 1}$$

where $\varepsilon$, the elasticity of the equivalence scale, is positive if $\beta_1 < 0$.

When we put $m_o = 1$, it is possible to transform the income of a household in which $M$ assumes value $m$ (for instance the size is equal to $m$) into the income which allows the same welfare to a family composed by one member through the simple transformation
\[ X_m \]

\[ m^\varepsilon. \]

It is possible to order the income of the households of the population considered in terms of non-decreasing welfare after having depurated them from the influence of the variable \( M \).

As a consequence of the previous considerations, \( Y = \frac{X_m}{m^\varepsilon} \) has the following cumulative distribution function (c.d.f.):

\[
V(y|m) = \frac{1}{\left[ 1 + \frac{\beta_0 m^{-\beta_1 - \varepsilon (a-1)}}{\alpha - 1} y^{-(a-1)} \right]^\gamma}. \tag{5}
\]

Given a simple random sample of size \( n \), we obtain \((x_i, m_i) \ (i=1,2,\ldots,n)\), where \( x_i \) represents the global expenditure and \( m \) the size of the household considered, the likelihood function is:

\[
L \left( \theta; x, m \right) = \prod_{i=1}^{n} f(x_i / m_i, \theta) c(m_i) \tag{6}
\]

where \( \theta = (\alpha, \beta_0, \beta_1, \gamma), \bar{x} = (x_1, \ldots, x_i, \ldots, x_n), \ m = (m_1, \ldots, m_i, \ldots, m_n), c(m_i) \) is the probability function of the household size and \( f(x_i / m_i, \theta) \) is the density function obtained differentiating (2) with respect to \( x_i \), that is

\[
f(x_i | m_i, \theta) = \frac{\gamma \beta_0 m_i^{-\beta_1} x_i^{-\alpha}}{\left[ 1 + \frac{\beta_0 m_i^{-\beta_1}}{\alpha - 1} x_i^{-(a-1)} \right]^{\gamma+1}}. \tag{7}
\]

The likelihood function (5) may be maximized in order to estimate the parameters of the vector \( \theta \):

\[
L \left( \hat{\theta}; x, m \right) = \max_{\theta \in \Theta} L \left( \theta; x, m \right)
\]

The logarithm of the likelihood function attains its maximum for the same values of \( \theta \):

\[
\log L \left( \hat{\theta}; x, m \right) = \max_{\theta \in \Theta} \log L \left( \theta; x, m \right)
\]

If we put

\[
u = \frac{\beta_0 m_i^{-\beta_1}}{\alpha - 1} x_i^{-(a-1)}
\]

\[
u = \frac{\beta_0 m_i^{-\beta_1}}{\alpha - 1} x_i^{-(a-1)}
\]
we can write

\[
\log L = n \log (\gamma) + n \log(\beta_0) - \beta_1 \sum_{i=1}^{n} \log(m_i) - \alpha \sum_{i=1}^{n} \log(x_i) - (\gamma + 1) \sum_{i=1}^{n} \log(1 + u_i) + \sum_{i=1}^{n} \log(f(m_i)) \tag{8}
\]

In order to find the maximum of \( \log L(\cdot) \), we have to solve the system:

\[
\begin{align*}
\frac{\partial \log L}{\partial \gamma} &= \frac{n}{\gamma} - \sum_{i=1}^{n} \log(1 + u_i) = 0 \\
\frac{\partial \log L}{\partial \alpha} &= \sum_{i=1}^{n} \log(x_i) - (\gamma + 1) \sum_{i=1}^{n} u_i \left[ \frac{1}{\alpha - 1} - \log(x_i) \right] = 0 \\
\frac{\partial \log L}{\partial \beta_0} &= \frac{n}{\beta_0} - (\gamma + 1) \sum_{i=1}^{n} m_i e^{-\beta_1 x_i (\alpha - 1)} = 0 \\
\frac{\partial \log L}{\partial \beta_1} &= \sum_{i=1}^{n} \log(m_i) - (\gamma + 1) \sum_{i=1}^{n} u_i \log(m_i) = 0 
\end{align*}
\tag{9}
\]

Maximum likelihood estimators are asymptotically Normal, that is:

\[
\sqrt{n} \left( \hat{\theta}_n - \theta \right) \xrightarrow{d} \mathcal{N} \left( \theta ; \Sigma_\theta \right)
\]

Where \( \mathcal{N}(\cdot) \) indicates a multivariate Normal Distribution and \( \Sigma_\theta = nI_\theta^{-1} \)

The information matrix \( I_\theta \) is a \((4 \times 4)\) matrix whose elements are:

\[
I_{ij} = -E \left( \frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j} \right) \quad i, j = 1, \ldots, 4
\]

It is possible to find the matrix \( I_{\hat{\theta}} \) extending the results reported in Latorre (1998), but this way presents many difficulties.

Variances and covariances of the estimators of \( \beta_1 \) and of \( (\alpha - 1) \) are estimated on the base of 100 bootstrap resamplings. We have found the estimates

\[
\hat{\text{Var}}\left( \hat{\beta}_1 \right), \quad \hat{\text{Var}}\left( \hat{\alpha} - 1 \right), \quad \hat{\text{Cov}}\left( \hat{\beta}_1, (\hat{\alpha} - 1) \right)
\]

If we put

\[
Y_1 = \frac{\hat{\beta}_1 - \bar{\beta}_1}{\sqrt{\hat{\text{Var}}(\hat{\beta}_1)}} \quad \text{and} \quad Y_2 = \frac{\hat{\alpha} - \bar{\alpha}}{\sqrt{\hat{\text{Var}}(\hat{\alpha})}}
\]

The r.v. \((Y_1, Y_2)\) has asymptotic Standardized Bivariate Normal Distribution.

The Distribution of the r.v. \( T = \text{Max}\left\{ Y_1, |Y_2| \right\} \) is a mixture of two arctangent densities with parameters:
\[ a_1 = \frac{1 + \rho}{\sqrt{1 - \rho}} \quad \text{and} \quad a_2 = \frac{1 - \rho}{\sqrt{1 + \rho}} \]

and with proportion
\[ \pi_1 = \frac{2}{\pi} \arctan(a_1) \quad \text{and} \quad \pi_2 = \frac{2}{\pi} \arctan(a_2) \]

respectively (see Pollastri, Tornaghi, 2004).

The distribution of \( T \) is useful to construct the confidence region for the two parameters \( \beta_i \) and \( \alpha \) at a fixed probability \((1 - \alpha^*)\).

The same region allows the construction of the confidence interval for \( \varepsilon = -\frac{\beta_1}{\alpha - 1} \).

Actually the lower bound of the confidence region for \( \varepsilon \) corresponds to ratio of the upper bound of the confidence region for \( \beta_i \) and the lower bound of the confidence region for \((\alpha - 1)\). The upper bound of the confidence interval for \( \varepsilon \) is the ratio of the lower bound of the region for \( \beta_i \), and the upper bound for \((\alpha - 1)\). Certainly, the probability is always equal to \((1 - \alpha^*)\).

3. Empirical results

A very important source of microdata about expenditure, income and wealth in Italy is represented by Banca d'Italia survey which is compiled of a process of interviews. The sampling unit is the household and the survey population is the whole set of households who live in Italy. In 2002, the sample size was 8011. Each family is randomly drawn through a two stage sample.

We use expenditures as the statistical measure of income.

The distribution of the household size is presented in Tab. 1.

<table>
<thead>
<tr>
<th>m</th>
<th>F(m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2212</td>
</tr>
<tr>
<td>2</td>
<td>0.2861</td>
</tr>
<tr>
<td>3</td>
<td>0.2185</td>
</tr>
<tr>
<td>4</td>
<td>0.1996</td>
</tr>
<tr>
<td>5</td>
<td>0.0587</td>
</tr>
<tr>
<td>6 or more</td>
<td>0.0159</td>
</tr>
</tbody>
</table>

The maximum likelihood estimates of the parameters of the model found through numerical approximations are:
\[ \hat{\gamma} = 4.75 \quad \hat{\beta}_0 = 9 \times 10^{-7} \quad \hat{\beta}_1 = -0.96 \quad \hat{\alpha} = 3.08 \]
The estimation of equivalence scale is $\hat{\epsilon} = 0.462$. It indicates that the income of a household having $m$ components divided by $m^{0.462}$ allows the same welfare of a family having only one component.

For instance, let $X_2$ the income of a household of size $m=2$. The household in question has the same welfare of a household constituted by a single person having income:

$$Y = \frac{X_2}{2^{0.462}} \cdot \frac{X_2}{1.377}$$

Kot estimated the elasticity $\hat{\epsilon}$ of the power of the equivalence scale using the data of the Polish household budget survey. He obtained $\hat{\epsilon} = 0.538$ in 1993 and $\hat{\epsilon} = 0.463$ in 1999.

In Pollastri (2003) the parameter of the equivalence scale for the region Lombardia was estimated equal to $\hat{\epsilon} = 0.6449$. The elasticity of the equivalence scales means that a new member of a family corresponds to an increment of the amount of the expenses greater in Lombardia than in Italy. If we compare the elasticity in Italy and in Poland, we do not find appreciable differences.

Then we have estimated on the base of $n=100$ resamplings

$$\hat{\text{Var}}(\hat{\beta_1}) = 0.00011 \quad \hat{\text{Var}}(\hat{\alpha - 1}) = 0.000144 \quad \hat{\text{Cov}}(\hat{\beta_1}, (\hat{\alpha - 1})) = 0.0000085$$

Having fixed $(1 - \alpha^*) = 0.95$, we can find the simultaneous region for $\beta_1$ and $(\alpha - 1)$.

In Pollastri (1979) it is possible to find that the percentile of the r.v. $T$ is $c=2.235$.

The rectangular region in which $(\beta_1, (\alpha - 1))$ assumes values is limited by the intervals $(-0.9834; -0.9366)$ for $\beta_1$ and $(2.0735; 2.0865)$ for $(\alpha - 1)$. For the considerations of the previous section, we can say that the true value $\epsilon$ lies in the interval $(0.4517; 0.4744)$ at the fixed confidence level.

**4. Concluding remarks**

The parameter of equivalence scales has been estimated for the Italian situation. It has been found through numerical maximization of the likelihood function. The value found is smaller than the one estimated for the region Lombardia. It means that the elasticity of the scale is greater in a region of the North of Italy than in the whole country.

The elasticity of the equivalence scale estimated in Italy is similar to the one found by Kot in 1999 in Poland.

Then it is proposed a method in order to give confidence interval for the unknown parameter of the equivalence scales. The method, proposed in the present paper, is based on the distribution of the maximum of the components of a Correlated Bivariate Normal.
References


