CHAPTER 5

Modelling Income Distributions with Limited Data

Duangkamon Chotikapanich, William Griffiths, and Gholamreza Hajargasht

INTRODUCTION

It is generally recognized that poverty and excessive inequality are socially undesirable. Reducing global poverty so that fewer individuals are deprived of basic needs is a major objective of international agencies. While what constitutes too much inequality is debatable, there is concern about the negative effects of rising inequality on health, crime and other aspects of society. Also, in extreme cases, inequality has led to the overthrow of governments and changes in the international order.

D. Chotikapanich
Monash University, Melbourne, VIC, Australia

W. Griffiths (✉)
University of Melbourne, Melbourne, VIC, Australia
e-mail: wegrif@unimelb.edu.au

G. Hajargasht
University of Queensland, Brisbane, QLD, Australia

233

It is important, therefore, be able to monitor changes in inequality and poverty using suitable measurement techniques. For this purpose, modelling and estimation of income distributions and Lorenz curves play an important role. Data available for modelling and estimation can be available in many forms. They may come from taxation data or from a variety of surveys. We focus on modelling and estimation when the data are limited in the sense that they come in grouped form, typically as the proportion of total income allocated to each of a number of groups, ordered according to increasing income, and with a specified proportion of the population within each group. These so-called income and population shares form the basis for estimating inequality through the Lorenz curve.[1] When share data are combined with data on mean incomes, income distributions can also be estimated, and their relationship with Lorenz curves can be exploited.

Data in grouped form are often utilized for large scale projects where inequality and poverty on a regional or global scale are being measured, and where compilation and dissemination of data in a more disaggregated form would be overly resource intensive. An example of such a study is Chotikapanich et al. (2012). Examples of locations where grouped share data are available for researchers are the World Bank's PovcalNet website[2] and that of the World Institute for Development Economic Research.[3]

Our objective is to summarize methods for estimating parametric income distributions using grouped data, to specify the functions needed for estimation for a number of popular parametric forms, and to provide formulae that can be used to compute inequality and poverty measures from the parameters of each of the distributions. In section Concepts, we introduce notation and concepts to be utilized later in the paper. The density, distribution and moment distribution functions that play an important role are introduced, along with poverty and inequality measures whose values can be calculated from estimates of the parameters of income distributions. We also describe the nature of the data that we assume are available. Section Estimation is devoted to estimation. Choice of estimation technique is influenced by whether or not

---

[1] We will continue to refer to *income* distributions and *income* shares, but recognize that data are often for *expenditure* that can be treated in the same way.

[2] http://iresearch.worldbank.org/PovcalNet/povOnDemand.aspx.

[3] https://www.wider.unu.edu/database/wiid.

group bounds are provided in the available data and on how the data are grouped: fixed group bounds and random population proportions or fixed population proportions and random group bounds. Both minimum distance (MD) and maximum likelihood (ML) estimators are considered, and results are provided for variants of the MD estimators which depend on which "distance" is being minimized. In section Specification of Distributions, Inequality and Poverty Measures, we tabulate the common parametric distributions that have been used to model income distributions; their density, distribution and moment distribution functions, and moments, are provided. Expressions that can be used to calculate inequality measures from the parameters of the different distributions are also tabulated. Expressions for some poverty measures are given in section Concepts; those for the Watts poverty index are tabulated in section Specification of Distributions, Inequality and Poverty Measures. In large projects, involving many countries and many years, MD and ML estimation can be daunting tasks. In section Simple Recipes for Two Distributions, we describe two relatively simple estimators for two specific distributions: the lognormal and the Pareto-lognormal. Some concluding remarks follow in section Concluding Remarks.

Concepts

We assume a population of incomes y, with $y > 0$, can be represented by a probability density function (pdf) $f(y; \theta)$ where θ is a vector of unknown parameters. Our objective is to review several alternative functional forms that have been suggested for $f(y; \theta)$, to describe methods for estimating θ from grouped data, and to provide expressions that can be used to compute estimates of inequality and poverty measures from estimates for θ.

We further assume y has a finite mean $\mu = \int\limits_{0}^{\infty} y\, f(y; \theta)\, dy$. Its cumulative distribution function (cdf) will be denoted by

$$\lambda = F(y; \theta) = \int\limits_{0}^{y} f(t; \theta)dt \tag{5.1}$$

and its first moment distribution function (fmdf) by

$$\eta = F^{(1)}(y; \theta) = \frac{1}{\mu} \int_0^y t\, f(t; \theta)\, dt \tag{5.2}$$

We will also utilize the second moment distribution function (smdf)

$$\psi = F^{(2)}(y; \theta) = \frac{1}{\mu^{(2)}} \int_0^\infty t^2 f(t; \theta) dt \tag{5.3}$$

where $\mu^{(2)}$ is the second moment $\mu^{(2)} = \int_0^\infty y^2 f(y; \theta)\, dy$. The Lorenz curve, relating the cumulative proportion of income to the cumulative proportion of population, is given by[4]

$$\eta = L(\lambda; \theta) = F^{(1)}\left(F^{-1}(\lambda; \theta); \theta\right) \tag{5.4}$$

When modelling begins with the specification of a Lorenz curve, the quantile function $y = F^{-1}(\lambda; \theta)$ can be found from it via differentiation,

$$y = F^{-1}(\lambda; \theta) = \mu \frac{dL(\lambda; \theta)}{d\lambda} \tag{5.5}$$

Inequality Measures

The most commonly cited inequality measure is the Gini coefficient g which is given by twice the area between the Lorenz curve and the line of equality where $\eta = \lambda$. That is,

$$g = 1 - 2 \int_0^1 L(\lambda; \theta)\, d\lambda$$

$$= -1 + \frac{2}{\mu} \int_0^\infty y F(y; \theta) f(y; \theta) dy \tag{5.6}$$

---

[4] See Gastwirth (1971).

Two further inequality measures that we consider are the Theil indices which are special cases of a generalized entropy class of measures. Unlike the Gini coefficient, members of this class have the advantage of being additively decomposable into population subgroups. The general class is given by

$$\text{GE}(v) = \frac{1}{v^2 - v} \left[\int_0^\infty \left(\frac{y}{\mu} \right)^v f(y; \theta) dy - 1 \right] \qquad v \neq 0, 1 \qquad (5.7)$$

The parameter v controls the sensitivity of the index to income differences in different parts of the income distribution; larger positive values imply greater sensitivity to income differences in the upper part of the distribution and more negative values imply greater sensitivity to differences in the lower part of the distribution. The Theil special cases are those for $v \to 0$ and $v \to 1$. They are given by

$$T_0 = \text{GE}(0) = \int_0^\infty \ln \left(\frac{\mu}{y} \right) f(y; \theta) \, dy \qquad (5.8)$$

$$T_1 = GE(1) = \int_0^\infty \left(\frac{y}{\mu} \right) \ln \left(\frac{y}{\mu} \right) f(y; \theta) dy \qquad (5.9)$$

The last inequality measure that we consider is the Pietra index which is equal to the maximum distance between the Lorenz curve and the equality line $\eta = \lambda$. It can be written as the difference between the cdf and the fmdf, evaluated at μ.

$$P = F(\mu; \theta) - F^{(1)}(\mu; \theta) \qquad (5.10)$$

Poverty Measures

Modelling and estimating income distributions are also useful for evaluating poverty. We consider four poverty measures, the headcount ratio HC, the poverty gap PG, the FGT index with the inequality aversion parameter set at 2 and the Watts index, WI. For convenience, we express HC, PG and FGT in terms of distribution and moment distribution

functions, and moments, which are tabulated for specific distributions in section Specification of Distributions, Inequality and Poverty Measures. The Watts index requires more work, however; we defer specific parametric expressions for it until section Specification of Distributions, Inequality and Poverty Measures. Given a specific poverty line z, we have

$$H = F(z; \theta) \tag{5.11}$$

$$\text{PG} = \int_0^z \left(\frac{z - y}{z} \right) f(y; \theta) \, dy = F(z; \theta) - \frac{\mu}{z} F^{(1)}(z; \theta) \tag{5.12}$$

$$\text{FGT}(2) = \int_0^z \left(\frac{z - y}{z} \right)^2 f(y; \theta) \, dy$$

$$= F(z; \theta) - 2\frac{\mu}{z} F^{(1)}(z; \theta) + \frac{\mu^{(2)}}{z^2} F^{(2)}(z; \theta) \tag{5.13}$$

$$\text{WI} = \int_0^z [\ln(z) - \ln(y)] f(y; \theta) \, dy \tag{5.14}$$

Data Setup

For estimating the various inequality and poverty measures, we assume we have a sample $\mathbf{y'} = (y_1, y_2, \ldots, y_T)$ randomly drawn from $f(y; \theta)$, and grouped into N income classes $(x_0, x_1), (x_1, x_2), \ldots, (x_{N-1}, x_N)$ with $x_0 = 0$ and $x_N = \infty$. We denote the proportion of observations in the i-th group as c_i, mean income in the i-th group as $\overline{y}_i$, and mean income for the whole sample as $\overline{y}$. The income share for the i-th group is $s_i = c_i \overline{y}_i / \overline{y}$. Sometimes observations $\mathbf{c'} = (c_1, c_2, \ldots, c_N)$ and $\mathbf{s'} = (s_1, s_2, \ldots s_N)$ are available from one source and $\overline{y}$ is available from another source, in which case group mean incomes can be found from $\overline{y}_i = s_i \overline{y} / c_i$. In the next section, we describe various methods for estimating θ, given the observations $(c_i, s_i, \overline{y})$.

Estimation

The estimation methods that we review can be categorized according to the way in which the data are generated, and whether the group bounds $x' = (x_0, x_1, \ldots, x_N)$ are known in addition to the observations on $(c_i, s_i, \overline{y})$. There are two ways in which the data can be generated. The group bounds x can be specified a priori, making the proportions of observations which fall into each group c_i, and the group means $\overline{y}_i$, the random variables. Alternatively, the c_i can be specified a priori, in which case the group bounds x are random variables, along with the group means $\overline{y}_i$. We consider estimation techniques for each of these cases in turn, noting the implications of known and unknown values for the group boundaries.

Estimation with Fixed x, Random c, Random $\overline{y}_i$

One approach for estimating θ when the group bounds x are known and the c_i are random is to maximize the likelihood function for the multinomial distribution. This approach uses information on x and c, but does not utilize the information contained in s and $\overline{y}$. The log of the likelihood function is given by

$$L(\theta) \ \propto \ K + \sum_{i=1}^{N} c_i \ln \big[F(x_i; \theta) - F(x_{i-1}; \theta) \big] \qquad (5.15)$$

where K is a constant.

In a series of papers (Griffiths & Hajargasht, 2015; Hajargasht & Griffiths, 2020; Hajargasht et al., 2012), three minimum distance (MD) estimators suitable for random c_i and $\overline{y}_i$ were introduced.[5] These estimators utilize information on c, s and $\overline{y}$, and can be applied with or without knowledge of x. When x is unknown it can be treated as a set of unknown parameters and estimated along with θ. The three estimators all have the same limiting distribution, but do not yield identical estimates. They are more efficient than the ML estimator from the multinomial likelihood

---

[5] The estimator in the first of these papers was described as a generalized method of moments estimator. Here, we use the term minimum distance estimator because it includes not only estimators that minimize the squared distance between sample and population moments, but also those that minimize the squared distance between sample quantities and their probability limits.

function where only information from c_i is utilized. To introduce the three estimators, we begin by noting the following:

$$\text{plim}\, c_i \;=\; F(x_i; \theta) - F(x_{i-1}; \theta) \;=\; \lambda_i(\phi) - \lambda_{i-1}(\phi) \tag{5.16}$$

$$\text{plim}\, s_i \;=\; F^{(1)}(x_i; \theta) - F^{(1)}(x_{i-1}; \theta) \;=\; \eta_i(\phi) - \eta_{i-1}(\phi) \tag{5.17}$$

where we write $\phi = (x, \theta)$ to accommodate the case where x is unobserved, making the unknown parameter vector equal to ϕ. If x is observed, we can proceed in the same way, utilizing the known x and treating θ as the unknown parameter vector.

MD Estimator 1
For the first MD estimator, we define

$$\tilde{y}_i = s_i \overline{y} = c_i \overline{y}_i \tag{5.18}$$

Since $\sum_{i=1}^{N} \tilde{y}_i = \sum_{i=1}^{N} c_i \overline{y}_i = \overline{y}$, we interpret $\tilde{y}_i$ as that part of mean income $\overline{y}$ that comes from the i-th group. Then, from (5.17) and (5.18),

$$\begin{aligned}
\text{plim}\, \tilde{y}_i &= \text{plim}\, \overline{y}\, \text{plim}\, s_i \\
&= \mu \Big[F^{(1)}(x_i; \theta) - F^{(1)}(x_{i-1}; \theta) \Big] \\
&= \mu \big[\eta_i(\phi) - \eta_{i-1}(\phi) \big]
\end{aligned} \tag{5.19}$$

From (5.16) and (5.19), we can set up the MD estimator

$$\hat{\phi}_1 = \arg\min_\phi H_1(\phi)' W H_1(\phi) \tag{5.20}$$

where

$$H_1(\phi) = \begin{bmatrix} c_1 - [\lambda_1(\phi) - \lambda_0(\phi)] \\ \vdots \\ c_{N-1} - \big[\lambda_{N-1}(\phi) - \lambda_{N-2}(\phi)\big] \\ \tilde{y}_1 - \mu[\eta_1(\phi) - \eta_0(\phi)] \\ \vdots \\ \tilde{y}_N - \mu\big[\eta_N(\phi) - \eta_{N-1}(\phi)\big] \end{bmatrix} \tag{5.21}$$

and W is a weight matrix. Note that μ will also depend on ϕ, the exact function depending on the parametric pdf chosen for the income distribution. Also, $c_N - \left[\lambda_N(\phi) - \lambda_{N-1}(\phi)\right]$ has been omitted since having $\sum_{i=1}^{N} c_i = 1$ makes one of the c_i entries redundant.

A possible weight matrix, one suggested by Chotikapanich et al. (2007), is to set the diagonal elements of W as $w_i = 1/c_i^2$ for $i = 1, 2, \ldots, N-1$ and $w_{N-1+i} = 1/\tilde{y}_i^2$ for $i = 1, 2, \ldots, N$, and the off-diagonal elements to zero. With this setting $\hat{\phi}_1$ minimizes the sum of squares of percentage errors. This weight matrix, call it W_{CGR}, is a simple one, and it works well in practice, but it is not optimal; it does not lead to the most efficient estimator for ϕ. Hajargasht et al. (2012) show that the inverse of the optimal weight matrix is given by

$$W_1^{-1}(\phi) = \begin{bmatrix} D_1 & \begin{bmatrix} D_2 & 0_{N-1} \end{bmatrix} \\ \begin{bmatrix} D_2 \\ 0'_{N-1} \end{bmatrix} & D_3 \end{bmatrix} - \begin{bmatrix} A_1 & A_2 \\ A_2' & A_3 \end{bmatrix} \qquad (5.22)$$

where 0_{N-1} is an $(N-1)-$ dimensional vector of zeros, and D_1, D_2 and D_3 are diagonal matrices. Their elements, and those of A_1, A_2 and A_3, are as follows.

$$[D_1]_{ii} = \lambda_i - \lambda_{i-1}, \qquad i = 1, 2, \ldots, N-1$$

$$[D_2]_{ii} = \mu(\eta_i - \eta_{i-1}), \qquad i = 1, 2, \ldots, N-1$$

$$[D_3]_{ii} = \mu^{(2)}(\psi_i - \psi_{i-1}), \qquad i = 1, 2, \ldots, N$$

$$[A_1]_{ij} = (\lambda_i - \lambda_{i-1})(\lambda_j - \lambda_{j-1}), \qquad i, j = 1, 2, \ldots, N-1$$

$$[A_2]_{ij} = (\lambda_i - \lambda_{i-1})(\eta_j - \eta_{j-1}) \qquad i = 1, 2, \ldots, N-1; \ j = 1, 2, \ldots, N$$

$$[A_3]_{ij} = (\eta_i - \eta_{i-1})(\eta_j - \eta_{j-1}), \qquad i, j = 1, 2, \ldots, N$$

All these quantities depend on the unknown parameter vector ϕ. To ease the notation, we have not made this dependence explicit. Note also that, through D_3, W will depend on the second moment $\mu^{(2)}$ and the second moment distribution function $\psi_i = F^{(2)}(x_i; \theta)$.

After inverting W_1^{-1} to find W_1, and simplifying, the objective function in (5.20) can be shown to be equal to

$$H_1(\phi)' W_1(\phi) H_1(\phi) = \sum_{i=1}^{N} w_{1i} \big[c_i - (\lambda_i - \lambda_{i-1})\big]^2$$
$$+ \sum_{i=1}^{N} w_{2i} \big[\tilde{y}_i - \mu(\eta_i - \eta_{i-1})\big]^2$$
$$- 2 \sum_{i=1}^{N} w_{3i} \big[c_i - (\lambda_i - \lambda_{i-1})\big]\big[\tilde{y}_i - \mu(\eta_i - \eta_{i-1})\big]$$

$$(5.23)$$

where

$$w_{1i} = \frac{\mu^{(2)}(\psi_i - \psi_{i-1})}{v_i} \tag{5.24}$$

$$w_{2i} = \frac{(\lambda_i - \lambda_{i-1})}{v_i} \tag{5.25}$$

$$w_{3i} = \frac{\mu(\eta_i - \eta_{i-1})}{v_i} \tag{5.26}$$

and

$$v_i = \mu^{(2)}(\lambda_i - \lambda_{i-1})(\psi_i - \psi_{i-1}) - \mu^2(\eta_i - \eta_{i-1})^2$$

There are three possible ways to approach the problem of finding an estimate $\hat{\phi}$ that minimizes $H_1(\phi)' W_1(\phi) H_1(\phi)$:

1. A two-step estimator where first an estimate $\hat{\phi}_{CGR}$ is obtained using the weight matrix W_{CGR}, and then a second estimate $\hat{\phi}_{2-STEP}$ is obtained by minimizing $H_1(\phi)' W_1(\hat{\phi}_{CGR}) H_1(\phi)$.
2. An iterative estimator obtained by iterating the 2-step estimator until convergence is achieved.
3. A "continuous updating estimator" where the whole function in (5.23) is minimized with respect to ϕ.

These three estimators all have the same limiting distribution but can produce different estimates. Their asymptotic covariance matrix is

$$\mathrm{var}\left(\hat{\phi}_1\right) = \frac{1}{T}\left[\left(\frac{\partial H_1^{*'}}{\partial \phi}\right)W_1^*\left(\frac{\partial H_1^*}{\partial \phi'}\right)\right]^{-1} \tag{5.27}$$

where H_1^* is a $(2N \times 1)$ vector obtained from H_1 by including $c_N - (\lambda_N - \lambda_{N-1})$ in the N-th position, and W_1^* is a $(2N \times 2N)$ matrix with 4 $(N \times N)$ diagonal blocks D_{11}, D_{12}, $D_{21} = D_{12}$ and D_{22}. The i-th diagonal elements of these matrices are w_{1i} for D_{11}, w_{2i} for D_{22} and $-w_{3i}$ for D_{12}. See Eqs. (5.24) to (5.26).

MD Estimator 2
The second MD estimator is that considered by Griffiths and Hajargasht (2015). It follows the same principles as the previous one, but it replaces $\tilde{y}_i$ by $\bar{y}_i$. To accommodate this replacement, we note that, from (5.16)–(5.18),

$$\mathrm{plim}\,\bar{y}_i = \frac{\mathrm{plim}\,\bar{y}\,\mathrm{plim}\,s_i}{\mathrm{plim}\,c_i}$$
$$= \frac{\mu(\eta_i - \eta_{i-1})}{\lambda_i - \lambda_{i-1}}$$

In this case, the MD estimator can be written as

$$\hat{\phi}_2 = \arg\min_\phi H_2(\phi)' W_2 H_2(\phi) \tag{5.28}$$

where

$$H_2(\phi) = \begin{bmatrix} c_1 - [\lambda_1 - \lambda_0] \\ \vdots \\ c_{N-1} - \left[\lambda_{N-1} - \lambda_{N-2}\right] \\ \bar{y}_1 - \frac{\mu[\eta_1 - \eta_0]}{\lambda_1 - \lambda_0} \\ \vdots \\ \bar{y}_N - \frac{\mu\left[\eta_N - \eta_{N-1}\right]}{\lambda_N - \lambda_{N-1}} \end{bmatrix} \tag{5.29}$$

and W_2 is a specified weight matrix. The weight matrix that is analogous to W_{CGR}, suggested for the previous estimator as a simple choice, or as a starting point for estimators that use an optimal weight matrix, is a diagonal matrix with elements $w_i = 1/c_i^2$ for $i = 1, 2, \ldots, N-1$, and $w_{i+N-1} = 1/\bar{y}_i^2$ for $i = 1, 2, \ldots, N$. Griffiths and Hajargasht (2015) show that the optimal weight matrix, for use with a 2-step, iterative or continuous updating estimator, is given by

$$W_2(\phi) = \begin{bmatrix} E_1 & 0 \\ 0 & E_2 \end{bmatrix} \qquad (5.30)$$

where

$$[E_1]_{ij} = \frac{\delta_{ij}}{\lambda_i - \lambda_{i-1}} + \frac{1}{\lambda_N - \lambda_{N-1}} \qquad i, j = 1, 2, \ldots, N-1 \qquad (5.31)$$

$$[E_2]_{ij} = \frac{\delta_{ij}(\lambda_i - \lambda_{i-1})^3}{\mu^{(2)}(\lambda_i - \lambda_{i-1})(\psi_i - \psi_{i-1}) - \mu^2(\eta_i - \eta_{i-1})^2} \qquad i, j = 1, 2, \ldots, N \qquad (5.32)$$

and $\delta_{ij} = 1$ when $i = j$ and $\delta_{ij} = 0$ when $i \neq j$. Using these results, the objective function can be simplified to

$$H_2(\phi)' W_2(\phi) H_2(\phi) = \sum_{i=1}^{N} \frac{\left[c_i - (\lambda_i - \lambda_{i-1}) \right]^2}{\lambda_i - \lambda_{i-1}}$$
$$+ \sum_{i=1}^{N} [E_2]_{ii} \left(\bar{y}_i - \frac{\mu(\eta_i - \eta_{i-1})}{\lambda_i - \lambda_{i-1}} \right)^2 \qquad (5.33)$$

As before, $H_2(\phi)' W_2(\phi) H_2(\phi)$ can be minimized using a 2-step estimator, an iterative estimator or a continuous updating estimator. The weights are $1/(\lambda_i - \lambda_{i-1})$ for the first terms in (5.33) and $[E_2]_{ii}$ for the second. In contrast to the earlier formulation in (5.23), there are no cross product terms, making the minimization problem simpler and convergence easier to obtain. The large sample covariance matrix of an estimator $\hat{\phi}_2$ using an optimal weight matrix is

$$\text{var}\left(\hat{\phi}_1 \right) = \frac{1}{T} \left[\left(\frac{\partial H_1^{*'}}{\partial \phi} \right) W_1^* \left(\frac{\partial H_1^*}{\partial \phi'} \right) \right]^{-1} \qquad (5.34)$$

where H_2^* is a $(2N \times 1)$ vector obtained from H_2 by including $c_N - (\lambda_N - \lambda_{N-1})$ in the N-th position, and W_2^* is a $(2N \times 2N)$ block-diagonal matrix with elements $1/(\lambda_i - \lambda_{i-1})$ in the first diagonal block and elements $[E_2]_{ii}$ in the second diagonal block.

MD Estimator 3
The third MD estimator that we describe is that considered by Hajargasht and Griffiths (2020). Its essential difference is that it considers cumulative population and income shares. To develop it, we begin by defining.

$$\hat{\lambda}_i = \sum_{j=1}^{i} c_j \text{ and } \hat{\eta}_i = \sum_{j=1}^{i} s_j \tag{5.35}$$

.

and recognizing that

$$\text{plim } \hat{\lambda}_i = F(x_i; \theta) = \lambda_i(\phi) \tag{5.36}$$

$$\text{plim } \bar{y} \hat{\eta}_i = \mu F^{(1)}(x_i; \theta) = \mu \eta_i(\phi) \tag{5.37}$$

Using (5.36) and (5.37), we can construct the MD estimator as

$$\hat{\phi}_3 = \arg\min_{\phi} H_3(\phi)' W_3 H_3(\phi) \tag{5.38}$$

where

$$H_3(\phi) = \begin{bmatrix} \hat{\lambda}_1 - \lambda_1 \\ \vdots \\ \hat{\lambda}_{N-1} - \lambda_{N-1} \\ \bar{y} \hat{\eta}_1 - \mu \eta_1 \\ \vdots \\ \bar{y} \hat{\eta}_{N-1} - \mu \eta_{N-1} \\ \bar{y} - \mu \end{bmatrix} \tag{5.39}$$

and W_3 is a pre-specified weight matrix. A simple weight matrix that can be used to simplify calculations or as a starting point for estimators that use an optimal weight matrix is a diagonal matrix with elements $w_i =$

$1\big/\hat{\lambda}_i^2$ for $i = 1, 2, \ldots, N-1$ and $w_{N-1+i} = 1\big/\left(\overline{y}\,\hat{\eta}_i\right)^2$ for $i = 1, 2, \ldots, N$. Hajargasht and Griffiths (2020) show that the optimal weight matrix is given by

$$W_3(\phi) = \begin{bmatrix} L_{11} & L_{12} \\ L'_{12} & L_{22} \end{bmatrix} \tag{5.40}$$

where

1. L_{11} is a $[(N-1) \times (N-1)]$ tri-diagonal matrix with the following nonzero elements:

$$[L_{11}]_{ii} = \frac{\mu^{(2)}(\psi_{i+1} - \psi_i)}{v_{i+1}} + \frac{\mu^{(2)}(\psi_i - \psi_{i-1})}{v_i} \quad i = 1, 2, \ldots, N-1$$

$$[L_{11}]_{ij} = \begin{cases} -\dfrac{\mu^{(2)}(\psi_i - \psi_{i-1})}{v_i} & i = 2, 3, \ldots, N-1; \ j = i-1 \\[3mm] -\dfrac{\mu^{(2)}\left(\psi_j - \psi_{j-1}\right)}{v_j} & j = 2, 3, \ldots, N-1; \ i = j-1 \end{cases}$$

$$\tag{5.41}$$

2. L_{12} is a $[(N-1) \times N]$ matrix with the following nonzero elements:

$$[L_{12}]_{ii} = -\frac{\mu(\eta_{i+1} - \eta_i)}{v_{i+1}} - \frac{\mu(\eta_i - \eta_{i-1})}{v_i} \quad i = 1, 2, \ldots, N-1$$

$$[L_{12}]_{ij} = \begin{cases} \dfrac{\mu(\eta_i - \eta_{i-1})}{v_i} & i = 2, 3, \ldots, N-1; \ j = i-1 \\[3mm] \dfrac{\mu\left(\eta_j - \eta_{j-1}\right)}{v_j} & j = 2, 3, \ldots, N; \ i = j-1 \end{cases}$$

$$\tag{5.42}$$

3. L_{22} is a $[N \times N]$ tri-diagonal matrix with the following nonzero elements

$$[L_{22}]_{ii} = \frac{(\lambda_{i+1} - \lambda_i)}{v_{i+1}} + \frac{(\lambda_i - \lambda_{i-1})}{v_i} \quad i = 1, 2, \ldots, N-1$$

$$[L_{22}]_{ij} = \begin{cases} \dfrac{-(\lambda_i - \lambda_{i-1})}{v_i} & i = 2, 3, \ldots, N; \ j = i-1 \\[2mm] \dfrac{-(\lambda_j - \lambda_{j-1})}{v_j} & j = 2, 3, \ldots, N; \ i = j-1 \end{cases}$$

$$[L_{22}]_{NN} = \frac{-(\lambda_N - \lambda_{N-1})}{v_N}$$

$$(5.43)$$

As in the previous two cases, the objective function can be minimized using a two-step estimator, an iterative estimator or a continuous updating estimator. The asymptotic covariance matrix for $\hat{\phi}_3$, when using an optimal covariance matrix, is

$$\text{var}(\hat{\phi}_3) = \frac{1}{T}\left[\left(\frac{\partial H_3'}{\partial \phi}\right) W_3 \left(\frac{\partial H_3}{\partial \phi'}\right)\right]^{-1} \tag{5.44}$$

A Quasi ML Estimator
Building on the work of Hilomi et al. (2008), Eckernkemper and Gribisch (2021), propose a quasi ML estimator. They combine the multinomial likelihood in Eq. (5.15) with a Gaussian approximation for the group means $\bar{y}_i$. Including the extra information means that estimation can proceed with or without knowledge of the group bounds, with these bounds being treated as parameters to be estimated when they are unknown. Let $T_i = c_i T$ be the number of observations in groups i. Each $\bar{y}_i$ is assumed to be $N\left(\tilde{\mu}_i, \tilde{\sigma}_i^2 / T_i\right)$ where the $\tilde{\mu}_i$ and the $\tilde{\sigma}_i^2$ are the means and variances of y from truncations $(x_{i-1} < y_i < x_i)$ of the originally specified distribution. That is,

$$\tilde{\mu}_i = E(y|x_{i-1} < y < x_i) = \frac{\mu[\eta_i(\phi) - \eta_{i-1}(\phi)]}{\lambda_i(\phi) - \lambda_{i-1}(\phi)} \tag{5.45}$$

and

$$\tilde{\sigma}_i^2 = \text{var}(y|x_{i-1} < y < x_i) = \frac{\mu^{(2)}\big[\psi_i(\phi) - \psi_{i-1}(\phi)\big]}{\lambda_i(\phi) - \lambda_{i-1}(\phi)} - \tilde{\mu}_i^2 \qquad (5.46)$$

Using these results, the log of the likelihood function can be written as

$$L(\phi) \;\propto\; K_1 + \sum_{i=1}^{N} \left\{ c_i \ln\big[\lambda_i(\phi) - \lambda_{i-1}(\phi)\big] - \ln\tilde{\sigma}_i - \frac{c_i}{2\tilde{\sigma}_i^2}(\overline{y} - \tilde{\mu}_i)^2 \right\}$$
$$(5.47)$$

Eckernkemper and Gribisch (2021) show that the estimator for ϕ that maximizes $L(\phi)$ is consistent and that the covariance matrix of its limiting distribution is the same as that for MD estimators 1 and 2.

Estimation with Fixed c, Random x, Random $\overline{y}_i$

In this case, the observations are grouped such that the proportion of observations in each group is pre-specified. Examples are 10 groups with 10% of the observations in each group or 20 groups with 5% of the observations in each group. This setup implies the proportion c_i are fixed (non-random) and the sample group boundaries x as well as the average cumulative incomes $\overline{y}_i$ are random variables. Let $y_{[1]}, y_{[2]}, \ldots, y_{[T]}$ be the order statistics obtained by arranging the original observations y in ascending order. An estimate for a group bound x_i is the largest order statistic in the i-th group, $\hat{x}_i = y_{[\hat{\lambda}_i T]}$. If the $\hat{x}_i$ are observed, estimation can use both the $\hat{x}_i$ and the $\overline{y}_i$; if the $\hat{x}_i$ are unobserved, then only the information in $\overline{y}_i$ can be utilized. We consider MD and ML estimation for both these cases. MD estimation with unobserved $\hat{x}_i$ corresponds to Lorenz curve estimation which has attracted a great deal of attention in the literature. See, for example, Chotikapanich (2008). A Lorenz curve implied by a specific income distribution is defined by Eq. (5.4). An alternative is to start with a specific parametric Lorenz curve in which case the corresponding income distribution is defined via the quantile function in (5.5). A problem with the latter approach is that the income distributions corresponding to some Lorenz curves are not defined for all values of y.

MD Estimation

The MD estimators that we consider are those proposed by Hajar-gasht and Griffiths (2020). Suppose, in the first instance, that the $\hat{x}_i$ are observed. To use this information in an MD estimator, we recognize that[6]

$$\text{plim}\,\hat{x}_i \;=\; F^{-1}\Big(\hat{\lambda}_i; \theta\Big) \tag{5.48}$$

To use information on the income shares, we use the cumulative shares $\hat{\eta}_i$ multiplied by mean income $\overline{y}$, in line with MD estimator 3 for the random c case. One difference, however, is that we express its probability limit in terms of the non-random c, instead of x, which is now a random variable. That is,

$$\text{plim}\,\overline{y}\,\hat{\eta}_i \;=\; \mu\,F^{(1)}\Big(F^{-1}(\hat{\lambda}; \theta); \theta\Big) \tag{5.49}$$

To set up the MD estimator, it is convenient to define notation for a generalized Lorenz curve which can be written as

$$\mu\,\eta = G(\lambda; \theta) = \mu\,L(\lambda; \theta) = \mu\,F^{(1)}\Big(F^{-1}(\lambda; \theta); \theta\Big) \tag{5.50}$$

Then, from (5.48)–(5.50), we can set up the following MD estimator,

$$\hat{\theta}_4 = \arg\min_\theta H_4'(\theta)\,W_4\,H_4(\theta) \tag{5.51}$$

---

[6] To avoid introducing more notation to what is already a very substantial amount, we will continue to use $\hat{\lambda}_i$ to denote the observed cumulative proportion of population, despite the fact that, in the current context, it is a non-random fixed quantity.

where

$$H_4(\theta) = \begin{bmatrix} \hat{x}_1 - F^{-1}(\hat{\lambda}_1; \theta) \\ \hat{x}_2 - F^{-1}(\hat{\lambda}_2; \theta) \\ \vdots \\ \hat{x}_{N-1} - F^{-1}(\hat{\lambda}_{N-1}; \theta) \\ \overline{y}\,\hat{\eta}_1 - G(\hat{\lambda}_1; \theta) \\ \vdots \\ \overline{y}\,\hat{\eta}_{N-1} - G(\hat{\lambda}_{N-1}; \theta) \\ \overline{y} - \mu \end{bmatrix} \tag{5.52}$$

and W_4 is a suitable chosen weight matrix. It can be shown that the optimal weight matrix is given by

$$W_4(\theta) = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega'_{12} & \Omega_{22} \end{bmatrix}^{-1} \tag{5.53}$$

where

$$\underset{[(N-1)\times(N-1)]}{[\Omega_{11}]_{ij}} = \begin{cases} \dfrac{\hat{\lambda}_i(1 - \hat{\lambda}_j)}{f(\hat{x}_i)f(\hat{x}_j)} & i \leq j \\[3mm] \dfrac{\hat{\lambda}_j(1 - \hat{\lambda}_i)}{f(\hat{x}_i)f(\hat{x}_j)} & j \leq i \end{cases} \tag{5.54}$$

$$\underset{[N\times N]}{[\Omega_{22}]_{ij}} = \begin{cases} \mu^{(2)}\psi_i + \left[\hat{\lambda}_i\hat{x}_i - G(\hat{\lambda}_i)\right]\left[\hat{x}_j - \hat{\lambda}_j\hat{x}_j + G(\hat{\lambda}_j)\right] \\ \qquad\qquad\qquad - \hat{x}_i G(\hat{\lambda}_i) \quad i \leq j \\ \mu^{(2)}\psi_j + \left[\hat{\lambda}_j\hat{x}_j - G(\hat{\lambda}_j)\right]\left[\hat{x}_i - \hat{\lambda}_i\hat{x}_i + G(\hat{\lambda}_i)\right] \\ \qquad\qquad\qquad - \hat{x}_j G(\hat{\lambda}_j) \quad j \leq i \end{cases} \tag{5.55}$$

$$\underset{[(N-1)\times N]}{[\Omega_{12}]_{ij}} = \begin{cases} \dfrac{\hat{\lambda}_i[G(\hat{\lambda}_j) - \hat{x}_j\hat{\lambda}_j + \hat{x}_j] - G(\hat{\lambda}_i)}{f(\hat{x}_i)} & i \leq j \\[3mm] \dfrac{[\hat{\lambda}_i - 1][G(\hat{\lambda}_j) - \hat{x}_j\hat{\lambda}_j]}{f(\hat{x}_i)} & j \leq i \end{cases} \tag{5.56}$$

The covariance matrix for the limiting distribution of $\hat{\theta}_4$ is

$$\text{var}(\hat{\theta}_4) = \frac{1}{T}\left(\frac{\partial H_4'}{\partial \theta} W_4 \frac{\partial H_4}{\partial \theta'}\right)^{-1} \tag{5.57}$$

When there are a large number of groups, the matrix inversion in (5.53) can be computationally demanding. Hajargasht and Griffiths (2020) show how W_4^{-1} can be derived from W_3^{-1} which has computationally convenient tri-diagonal blocks. They also demonstrate that, if the groupings for this set up are equivalent to those for the MD3 setup in the sense that, a priori, $x_i = F^{-1}(\lambda_i; \theta)$, then the asymptotic covariance matrices for $\hat{\theta}_3$ and $\hat{\theta}_4$ are identical.

Minimizing (5.51) to find an estimate $\hat{\theta}_4$ can proceed using one of the three algorithms described in section Estimation with Fixed $\boldsymbol{x}$, Random $\boldsymbol{c}$, Random $\overline{\boldsymbol{y}}_i$. However, there are two requirements which will not always be met: estimates of the bounds $\hat{x}_i = y_{[\hat{\lambda}_i T]}$ must be observed and the cdf must be invertible, either algebraically or computationally, so that quantiles $F^{-1}(\hat{\lambda}_i; \theta)$ can be found. Note that $F^{-1}(\hat{\lambda}_i; \theta)$ appears not only in the first $(N-1)$ elements of H_4 but also in the next $(N-1)$ elements that involve the generalized Lorenz curve $G(\lambda; \theta) = \mu F^{(1)}\big(F^{-1}(\lambda; \theta); \theta\big)$. One way to overcome non-invertibility of the cdf is to replace the assumption of a parametric income distribution with an assumption of a parametric Lorenz curve. Doing so overcomes the problem for the second set of elements in H_4, and relationships between the generalized Lorenz curve and the quantile function—see Hajargasht and Griffiths (2020)—can be exploited to obtain the first set of elements in H_4.

When the $\hat{x}_i$ are unobserved, estimation can proceed using the last N elements in H_4, with calculations made from an assumed income distribution if the cdf is invertible, or from an assumed Lorenz curve if the cdf is not invertible. This last approach is that most closely aligned with suggestions for Lorenz curve estimation which have appeared in the literature.[7] Earlier suggestions are sub-optimal in the sense that they do not use the best weighting matrix. Details can be found in Hajargasht and Griffiths (2020).

---

[7] See Chotikapanich (2008) for access to this literature.

ML Estimation

ML estimation of θ for fixed c_i, and random x_i and $\bar{y}_i$ was considered by Eckernkemper and Gribisch (2021). Recognizing that the joint density for the group bounds and group means can be written as

$$f(\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_N, \hat{\boldsymbol{x}}) = f(\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_N | \hat{\boldsymbol{x}}) f(\hat{x}_1)$$
$$f(\hat{x}_2 | \hat{x}_1) \ldots f(\hat{x}_{N-1} | \hat{x}_{N-2}) \qquad (5.58)$$

they set up a likelihood function that uses distribution theory for order statistics for $f(\hat{\boldsymbol{x}})$ and a Gaussian approximation for $f(\bar{y}_i | \hat{x}_i, \hat{x}_{i-1})$. Using results in David and Nagaraji (2003), the conditional means and variances for the $\bar{y}_i$ can be written as

$$\mu_i = E(\bar{y}_i | \hat{x}_i, \hat{x}_{i-1}) = \frac{T_i - 1}{T_i} \tilde{\mu}_i + \frac{\hat{x}_i}{T_i} \qquad (5.59)$$

and

$$\sigma_i^2 = \text{var}(\bar{y}_i | \hat{x}_i, \hat{x}_{i-1}) = \frac{T_i - 1}{T_i^2} \tilde{\sigma}_i^2 \qquad (5.60)$$

The log-likelihood is

$$L(\theta) = K_2 - \frac{1}{2} \left[\ln \tilde{\sigma}_N^2 + \frac{T_N}{\tilde{\sigma}_N^2} (\bar{y}_N - \tilde{\mu}_N)^2 \right] + T_N \ln \left[1 - F(\hat{x}_{N-1}; \theta) \right]$$
$$+ \sum_{i=1}^{N-1} \left\{ \begin{array}{l} -\frac{1}{2} \left[\ln \sigma_i^2 + \left(\dfrac{\bar{y}_i - \mu_i}{\sigma_i} \right)^2 \right] \\ + (T_i - 1) \ln \left[F(\hat{x}_i; \theta) - F(\hat{x}_{i-1}; \theta) \right] + \ln f(\hat{x}_i; \theta) \end{array} \right\}$$
$$(5.61)$$

where K_2 is a constant. The estimator $\hat{\theta}_5$ that minimizes (5.61) can be interpreted as a quasi ML estimator. Eckernkemper and Gribisch (2021) establish its asymptotic covariance matrix as

$$\text{var}(\hat{\theta}_5) = \frac{1}{T} \left[\sum_{i=1}^{N} \left(\frac{\partial \mu_i}{\partial \theta} \frac{\partial \mu_i}{\partial \theta'} \right) \left(\frac{c_i}{\tilde{\sigma}_i^2} \right) \right]^{-1} \qquad (5.62)$$

One difference between the estimator $\hat{\theta}_5$ and those estimators considered in the earlier sections is that it requires knowledge of the sample size T,

from which the number of observations in each group can be found from $T_i = c_i T$. All estimators require knowledge of T to compute standard errors, but knowledge of the proportions c_i, without knowledge of T, is sufficient for the earlier estimators for θ and ϕ to be employed.

For ML estimation of θ when the $\hat{x}_i$ are not observed, Eckernkemper and Gribisch (2021) integrate out the $\hat{x}_i$ from the likelihood in (5.61) to obtain the following log-likelihood

$$L(\theta) = K_3 - \frac{1}{2}\Big[\log|\Xi| + T\big(\overline{\mathbf{y}} - \boldsymbol{\mu}^*\big)'\Xi^{-1}\big(\overline{\mathbf{y}} - \boldsymbol{\mu}^*\big)\Big] \tag{5.63}$$

where K_3 is a constant, $\overline{\mathbf{y}}' = (\overline{y}_1, \overline{y}_2, \ldots, \overline{y}_N)$, $\boldsymbol{\mu}^*$ is an $(N \times 1)$ vector with i-th element equal to

$$\mu_i^* = \frac{1}{c_i}\Big[G\big(\hat{\lambda}; \theta\big) - G\big(\hat{\lambda}_{i-1}; \theta\big)\Big] \tag{5.64}$$

and $\Xi = DB\,\Omega_{22}^*B'D$ where $D = \mathrm{diag}\big(c_1^{-1}, c_2^{-1}, \ldots, c_N^{-1}\big)$, $[B]_{ii} = 1$, $[B]_{ij} = -1$ for $i = j + 1$, $j = 1, 2, \ldots, N - 1$, and zero elsewhere, and Ω_{22}^* is equal to Ω_{22} defined in (5.55), but with $\hat{x}_i$ and $\hat{x}_j$ replaced by $F^{-1}\big(\hat{\lambda}_i; \theta\big)$ and $F^{-1}\big(\hat{\lambda}_j; \theta\big)$, respectively. The asymptotic covariance matrix for the estimator $\hat{\theta}_6$ obtained by maximizing (5.61) is

$$\mathrm{var}\,(\hat{\theta}_6) = \frac{1}{T}\left[\partial\mu^{*'}\,\Xi^{-1}\frac{\partial\mu^*}{\partial\theta'}\right] \tag{5.65}$$

Specification of Distributions, Inequality and Poverty Measures

To implement the estimation methods described in section Estimation, a specific parametric distribution has to be specified and we need its moments, its pdf, cdf, fmdf and smdf. This information is provided in Table 5.1 for several popular income distributions. Once the parameters of a chosen distribution have been estimated, estimates for inequality and poverty incidence are frequently of interest. In Table 5.2, we provide expressions that can be used to compute inequality estimates from the estimates of the parameters. Expressions for the poverty estimates were

Table 5.1 Probability distributions, distribution functions and moments

Distribution	pdf	cdf/fmdf/smdf[a]	Moments
Pareto	$f(y;\alpha,y_0) = \dfrac{\alpha y_0^\alpha}{y^{\alpha+1}}$ $y \geq y_0 > 0$	$F^{(k)}(y;\alpha,y_0) = 1 - \left(\dfrac{y_0}{y}\right)^{\alpha-k}$ $\alpha > k$	$\mu^{(k)} = \dfrac{\alpha y_0^k}{\alpha-k}$ $\alpha > k$
Lognormal[b]	$f(y;\beta,\sigma) = \dfrac{1}{\sigma y\sqrt{2\pi}} \exp\left\{-\dfrac{1}{2\sigma^2}(\ln y - \mu)^2\right\}$	$F^{(k)}(y;\beta,\sigma) = \Phi\left(\dfrac{\ln(y)-\beta-k\sigma^2}{\sigma}\right)$	$\mu^{(k)} =$ $\exp\left\{k\beta + \dfrac{k^2\sigma^2}{2}\right\}$
Pareto-lognormal	$f(y;\alpha,\beta,\sigma) =$ $\dfrac{\alpha}{y^{\alpha+1}} \exp\left\{\dfrac{\alpha^2\sigma^2}{2} + \alpha\beta\right\} \Phi\left(\dfrac{\ln y - \beta - \alpha\sigma^2}{\sigma}\right)$	$F^{(k)}(y;\alpha,\beta,\sigma) = \Phi\left(\dfrac{\ln y - \beta - k\sigma^2}{\sigma}\right) -$ $y^{k-\alpha} \exp\left\{\dfrac{\sigma^2}{2}(\alpha^2 - k^2) + \beta(\alpha - k)\right\} \Phi\left(\dfrac{\ln y - \beta - \alpha\sigma^2}{\sigma}\right)$ $\alpha > k$	$\mu^{(k)} =$ $\dfrac{\alpha}{\alpha-k} \exp\left\{k\beta + \dfrac{k^2\sigma^2}{2}\right\}$ $\alpha > k$
GB2[c]	$f(y;a,b,p,q) =$ $\dfrac{ay^{ap-1}}{b^{ap} B(p,q)\left[1+\left(\frac{y}{b}\right)^a\right]^{p+q}}$	$F^{(k)}(y;a,b,p,q) = B_u\left(p + \dfrac{k}{a}, q - \dfrac{k}{a}\right)$ $k < qa$ $u = \dfrac{(y/b)^a}{1+(y/b)^a}$	$\mu^{(k)} =$ $\dfrac{b^k B\left(p+\frac{k}{a},q-\frac{k}{a}\right)}{B(p,q)}$ $k < qa$
Singh-Maddala[d] (p = 1)	$f(y;a,b,q) = \dfrac{aqy^{a-1}}{b^a\left[1+\left(\frac{y}{b}\right)^a\right]^{1+q}}$	$F(y;a,b,q) = 1 - \left[1 + \left(\dfrac{y}{b}\right)^a\right]^{-q}$ $F^{(k)}(y;a,b,q) = B_u\left(1 + \dfrac{k}{a}, q - \dfrac{k}{a}\right)$ $k < aq$ $u = \dfrac{(y/b)^a}{1+(y/b)^a}$	$\mu^{(k)} =$ $\dfrac{b^k \Gamma\left(1+\frac{k}{a}\right)\Gamma\left(q-\frac{k}{a}\right)}{\Gamma(q)}$ $k < aq$
Beta-2 (a = 1)	$f(y;b,p,q) = \dfrac{y^{p-1}}{b^p B(p,q)\left[1+\frac{y}{b}\right]^{p+q}}$	$F^{(k)}(y;b,p,q) = B_u(p+k, q-k)$ $k < q$ $u = \dfrac{y/b}{1+y/b}$	$\mu = \dfrac{bp}{q-1}$ $\mu^{(2)} = \dfrac{bp(p+1)}{(q-1)(q-2)}$

Distribution	pdf	cdf/fmdf/smdf[a]	Moments
Dagum ($q=1$)	$f(y;a,b,p) = \dfrac{apy^{ap-1}}{b^{ap}\left[1+\left(\frac{y}{b}\right)^a\right]^{p+1}}$	$F(y;a,b,p) = \left[1+\left(\frac{y}{b}\right)^{-a}\right]^{-p}$ $F^{(k)}(y;a,b,p) = B_u\left(p+\frac{k}{a},1-\frac{k}{a}\right)$ $u = \dfrac{(y/b)^a}{1+(y/b)^a}\quad k<a$	$\mu^{(k)} =$ $\dfrac{b^k\Gamma\left(p+\frac{k}{a}\right)\Gamma\left(1-\frac{k}{a}\right)}{\Gamma(p)}$ $k<a$
Generalized gamma(5)	$f(y;a,b,p) = \dfrac{ay^{ap-1}\exp\left[-\left(\frac{y}{b}\right)^a\right]}{b^{ap}\Gamma(p)}$	$F^{(k)}(y;a,b,p) = \Gamma_u\left(p+\frac{k}{a}\right)$ $u = \left(\frac{y}{b}\right)^a$	$\mu^{(k)} = \dfrac{b^k\Gamma\left(p+\frac{k}{a}\right)}{\Gamma(p)}$
Gamma ($a=1$)	$f(y;b,p) = \dfrac{y^{p-1}\exp\left[-\frac{y}{b}\right]}{b^p\Gamma(p)}$	$F^{(k)}(y;b,p) = \Gamma_u(p+k)$ $u = \dfrac{y}{b}$	$\mu = bp$ $\mu^{(2)} = b^2p(p+1)$

Notes Kleiber and Kotz (2003) is an excellent source for many of the results in this table

[a]Of interest are $k=0$ (cdf = cumulative distribution function), $k=1$ (fmdf = first moment distribution function), and $k=2$ (smdf = second moment distribution function

[b] $\Phi(x)$ is the standard normal cdf evaluated at x

[c]GB2 is the generalized beta distribution of the 2nd kind.$B(p,q)$ is the beta function evaluated at (p,q). $B_u(p,q)$ is the standard beta cdf, with parameters (p,q), evaluated at u

[d]$\Gamma(p)$ is the gamma function evaluation at p

[e]$\Gamma_u(p)$ is the standard gamma cdf, with parameter p, evaluated at u

given in section Poverty Measures, with the exception of the Watts Index whose expressions we have tabulated in Table 5.3.

SIMPLE RECIPES FOR TWO DISTRIBUTIONS

In some instances, where large scale projects involving many countries and many time periods are being undertaken, it may be prudent to use estimation techniques which are relatively simple. In this section, we consider two estimation techniques that fall into this category—one for the lognormal distribution and one for the Pareto-lognormal distribution.

Lognormal Distribution

In the previous section, we indicated that the Gini coefficient for the lognormal distribution is $g = 2\Phi\left(\sigma / \sqrt{2}\right) - 1$ and its mean is $\mu = \exp\{\beta + \sigma^2 / 2\}$. Using grouped data the Gini coefficient can be estimated from

$$\hat{g} = \sum_{i=1}^{N-1} \hat{\eta}_{i+1} \hat{\lambda}_i - \sum_{i=1}^{N-1} \hat{\eta}_i \hat{\lambda}_{i+1} \qquad (5.66)$$

and the mean can be estimated using $\overline{y}$,

$$\hat{\mu} = \overline{y} = \exp\left\{\hat{\beta} + \frac{\hat{\sigma}^2}{2}\right\} \qquad (5.67)$$

Utilizing these two equations and the expression for the Gini coefficient yields the parameter estimates.

$$\hat{\sigma} = \sqrt{2}\Phi^{-1}\left(\frac{g+1}{2}\right) \qquad (5.68)$$

$$\hat{\beta} = \ln(\overline{y}) - \frac{\hat{\sigma}^2}{2} \qquad (5.69)$$

This approach was adopted by Chotikapanich et al. (1997).

Pareto-Lognormal Distribution

For the Pareto-lognormal distribution, we can estimate the Theil inequality measures from the grouped data, and then use these estimates,

Table 5.2 Inequality measures[a]

Distribution	Gini coefficient	Theil index (0)
Pareto	$g = \frac{1}{2\alpha-1}$	$T_0 = -\frac{1}{\alpha} + \ln\left(\frac{\alpha}{\alpha-1}\right)$
Lognormal	$g = 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1$	$T_0 = \frac{\sigma^2}{2}$
Pareto-lognormal	$g = \frac{2\exp\left[\alpha(\alpha-1)\sigma^2\right]}{2\alpha-1}\Phi\left(\frac{(1-2\alpha)\sigma}{\sqrt{2}}\right) + 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1$	$T_0 = \ln\left(\frac{\alpha}{\alpha-1}\right) - \frac{1}{\alpha} + \frac{\sigma^2}{2}$
GB2[b]	See Kleiber and Kotz (2003, p. 193)[c]	$T_0 = \ln\left(\frac{\mu}{b}\right) - \frac{\Psi(p)-\Psi(q)}{a}$
Beta-2 $(a=1)$	$g = \frac{2B(2p,2q-1)}{pB^2(p,q)}$	$T_0 = \ln\left(\frac{p}{q-1}\right) - [\Psi(p) - \Psi(q)]$
Singh-Maddala $(p=1)$	$g = 1 - \frac{\Gamma(q)\Gamma\left(2q-\frac{1}{a}\right)}{\Gamma\left(q-\frac{1}{a}\right)\Gamma(2q)}$	$T_0 = \ln\left(\frac{\mu}{b}\right) - \frac{\Psi(1)-\Psi(q)}{a}$
Dagum $(q=1)$	$g = \frac{\Gamma(p)\Gamma\left(2p+\frac{1}{a}\right)}{\Gamma\left(p+\frac{1}{a}\right)\Gamma(2p)} - 1$	$T_0 = \ln\left(\frac{\mu}{b}\right) - \frac{\Psi(p)-\Psi(1)}{a}$
Generalized Gamma	See Kleiber and Kotz (2003, p.155)[d]	$T_0 = \ln\left(\frac{\Gamma\left(p+\frac{1}{a}\right)}{\Gamma(p)}\right) - \frac{\Psi(p)}{p}$

(continued)

Table 5.2 (continued)

Distribution	Gini coefficient	Theil index (0)
Gamma $(a = 1)$	$g = \dfrac{\Gamma\left(p + \frac{1}{2}\right)}{\sqrt{\pi}\,\Gamma(p+1)}$	$T_0 = \ln p - \dfrac{\Psi(p)}{p}$

Distribution	Theil index (1)	Pietra index
Pareto	$T_1 = \dfrac{1}{\alpha - 1} - \ln\left(\dfrac{\alpha}{\alpha - 1}\right)$	$P = \dfrac{(\alpha - 1)^{\alpha - 1}}{\alpha^\alpha}$
Lognormal	$T_1 = \dfrac{\sigma^2}{2}$	$P = 2\Phi\left(\dfrac{\sigma}{2}\right) - 1$
Pareto-lognormal	$T_1 = \ln\left(\dfrac{\alpha - 1}{\alpha}\right) + \dfrac{1}{\alpha - 1} + \dfrac{\sigma^2}{2}$	$P = \Phi\left(u + \dfrac{\sigma}{2}\right) - \Phi\left(u - \dfrac{\sigma}{2}\right)$ $\quad + \left\{\dfrac{(\alpha - 1)^{\alpha - 1}}{\alpha^\alpha} - \exp\left\{\dfrac{\alpha\sigma^2}{2}(\alpha - 1)\right\}\right\}\Phi\left(u + \sigma\left(\dfrac{1}{2} - \alpha\right)\right)$ $\quad u = \dfrac{1}{\sigma}\ln\left(\dfrac{\alpha}{\alpha - 1}\right)$
GB2[b]	$T_1 = \dfrac{\Psi\left(p + \frac{1}{a}\right) - \Psi\left(q - \frac{1}{a}\right)}{a} + \ln\left(\dfrac{b}{\mu}\right)$	$P = B_u(p,q) - B_u\left(p + \dfrac{1}{a}, q - \dfrac{1}{a}\right)$ $\quad u = \dfrac{(\mu/b)^a}{1 + (\mu/b)^a}$
Beta-2 $(a = 1)$	$T_1 = \Psi(p+1) - \Psi(q-1) + \ln\left(\dfrac{q-1}{p}\right)$	$P = \dfrac{u^p(1-u)^{q-1}}{pB(p,q)}$ $\quad u = \dfrac{\mu/b}{1 + \mu/b}$

Distribution	Theil index (1)	Pietra index
Singh-Maddala $(p=1)$	$T_1 = \dfrac{\Psi\left(1+\frac{1}{a}\right) - \Psi\left(q-\frac{1}{a}\right)}{a} + \ln\left(\frac{b}{\mu}\right)$	$P = 1 - \left[1 + \left(\frac{\mu}{b}\right)^a\right]^{-q} - B_u\left(1+\frac{1}{a}, q-\frac{1}{a}\right)$ $u = \dfrac{(\mu/b)^a}{1+(\mu/b)^a}$
Dagum $(q=1)$	$T_1 = \dfrac{\Psi\left(p+\frac{1}{a}\right) - \Psi\left(1-\frac{1}{a}\right)}{a} + \ln\left(\frac{b}{\mu}\right)$	$P = \left[1 + \left(\frac{\mu}{b}\right)^{-a}\right]^{-p} - B_u\left(p+\frac{1}{a}, 1-\frac{1}{a}\right)$ $u = \dfrac{(\mu/b)^a}{1+(\mu/b)^a}$
Generalized Gamma	$T_1 = \dfrac{\Psi\left(p+\frac{1}{a}\right)}{a} - \ln\left(\dfrac{\Gamma\left(p+\frac{1}{a}\right)}{\Gamma(p)}\right)$	$P = \Gamma_u(p) - \Gamma_u\left(p+\frac{1}{a}\right)$ $u = \left(\frac{y}{b}\right)^a$
Gamma $(a=1)$	$T_1 = \dfrac{1}{p} + \Psi(p) - \ln p$	$P = \dfrac{p^p \exp(-p)}{\Gamma(p+1)}$

Notes [a]Many of these measures can be found in Kleiber and Kotz (2003). Sarabia et al (2017) has a comprehensive list of the Theil measures. Sarabia and Jordá (2014) consider the Pietra coefficient in detail

[b]$\Psi(p)$ is the digamma function $\Psi(p) = d\ln\Gamma(p)/dp$.

[c]The expression in Kleiber and Kotz involves lengthly hypergeometric functions. It also contains a typographical error; the first p in the denominator should be deleted. Computationally, we have found that evaluating the Gini coefficient via numerical integration is easier than computing values for the hypergeometric functions

[d]Like the GB2, this Gini expression involves hypergeometric functions. It may be preferable to evaluate the Gini integral numerically

Table 5.3 Watts poverty indices for selected distributions

Distribution	*Index*
Pareto	$WI = \ln\left(\frac{z}{y_0}\right) + \left(\frac{\alpha}{\alpha-1}\right)\left(\frac{1}{y_0}\right)\left[1 - \left(\frac{y_0}{z}\right)^{\alpha+1}\right]$
Lognormal	$WI = \sigma[u\Phi(u) + \phi(u)]$ $u = \frac{\ln z - \beta}{\sigma}$
Pareto-lognormal	$WI = -\frac{1}{\alpha}F(z; \alpha, \beta, \sigma) + \sigma[u\Phi(u) + \phi(u)]$ $u = \frac{\ln z - \beta}{\sigma}$
GB2[a]	$WI = \ln\left(\frac{z}{b}\right)B_u(p,q) -$ $\frac{1}{a}\{D_p B_u(p,q) - D_q B_u(p,q) + B_u(p,q)[\Psi(p) - \Psi(q)]\}$ $u = \frac{(z/b)^a}{1+(z/b)^a}$
Beta-2 $(a = 1)$	$WI = \ln\left(\frac{z}{b}\right)B_u(p,q) + D_q B_u(p,q) - D_p B_u(p,q) -$ $B_u(p,q)[\Psi(p) - \Psi(q)]$ $u = \frac{z/b}{1+z/b}$
Singh-Maddala $(p = 1)$	$WI = \ln\left(\frac{z}{b}\right)F(z; a, b, q) +$ $\frac{1}{a}\left\{\left[1 + \left(\frac{z}{b}\right)^a\right]^{-q}\ln\left[1 + \left(\frac{z}{b}\right)^a\right] - F(z; a, b, q)[\Psi(1) - \Psi(q)]\right\}$
Dagum $(q = 1)$	$WI = \ln\left(\frac{z}{b}\right)F(z; a, b, p) +$ $\frac{1}{a}\left\{\left[1 + \left(\frac{z}{b}\right)^{-a}\right]^{-p}\ln\left[1 + \left(\frac{z}{b}\right)^{-a}\right] - F(z; a, b, p)[\Psi(p) - \Psi(1)]\right\}$
Generalized Gamma[b]	$WI = \ln\left(\frac{z}{b}\right)\Gamma_u(p) - \frac{1}{a}\left[D_p\Gamma_u(p) + \Gamma_u(p)\Psi(p)\right]$ $u = \left(\frac{z}{b}\right)^a$
Gamma $(a = 1)$	$WI = \ln(u)\Gamma_u(p) - D_p\Gamma_u(p) + \Gamma_u(p)\Psi(p)$ $u = \frac{z}{b}$

Notes [a]$D_p B_u(p,q)$ and $D_q B_u(p,q)$ are the derivatives of the beta cdf $B_u(p,q)$ with respect to p and q respectively. These derivatives are available in some software such as EViews. A derivation of the expression can be found in Chotikapanich et al. (2013)
[b]$D_p\Gamma_u(p)$ is the derivative of the gamma cdf $\Gamma_u(p)$ with respect to p. It too is available in software such as EViews. Derivation of the result uses a similar approach to that for the GB2

along with sample mean income to estimate the parameters. Working in this direction, the grouped-data sample estimates are

$$\hat{T}_1 = \sum_{i=1}^{N} c_i\left(\frac{\overline{y}_i}{\overline{y}}\right)\ln\left(\frac{\overline{y}_i}{\overline{y}}\right) \tag{5.70}$$

$$\hat{T}_0 = \sum_{i=1}^{N} c_i \ln\left(\frac{\overline{y}}{y_i}\right) \tag{5.71}$$

$$\hat{\mu} = \overline{y} \tag{5.72}$$

The corresponding quantities in terms of the parameters of the Pareto-lognormal distribution are

$$T_1 = \frac{1}{\alpha - 1} + \frac{\sigma^2}{2} + \ln\left(\frac{\alpha - 1}{\alpha}\right) \tag{5.73}$$

$$T_0 = -\frac{1}{\alpha} + \frac{\sigma^2}{2} - \ln\left(\frac{\alpha - 1}{\alpha}\right) \tag{5.74}$$

$$\mu = \frac{\alpha}{\alpha - 1} \exp\left\{\beta + \frac{\sigma^2}{2}\right\} \tag{5.75}$$

Assuming the mean exists ($\alpha > 1$), from (5.70)–(5.75) we can retrieve parameter estimates using the following three steps:

1. Find $\hat{\alpha}$ as the Solution to the Equation

$$\frac{2\hat{\alpha} - 1}{\hat{\alpha}(\hat{\alpha} - 1)} + 2\ln\left(\frac{\hat{\alpha} - 1}{\hat{\alpha}}\right) = \hat{T}_1 - \hat{T}_0 \tag{5.76}$$

2. Find $\hat{\sigma}^2$ from

$$\hat{\sigma}^2 = \hat{T}_1 + \hat{T}_0 - \frac{1}{\hat{\alpha}(\hat{\alpha} - 1)} \tag{5.77}$$

3. Find $\hat{\beta}$ from

$$\hat{\beta} = \ln(\overline{y}) + \ln\left(\frac{\hat{\alpha} - 1}{\hat{\alpha}}\right) - \frac{\hat{\sigma}^2}{2} \tag{5.78}$$

Concluding Remarks

Inequality and poverty, both nationally and globally, continue to be two of the most pressing issues facing today's society. Accurate measurement

of inequality and poverty involves a multitude of non-trivial considerations including reliable data collection, specification of purchasing power parities and definition of a suitable poverty line. We have focused on a further consideration, how to model and estimate income distributions, and how to estimate inequality and poverty from the parameters of those income distributions, when using grouped data. Single observations are becoming increasingly available and their use is preferred to the use of grouped data if resources are adequate for doing so. However, countries and time periods for which only grouped data are available are still prevalent, and it can be advantageous to use grouped data for large scale regional and global projects. Our objective has been to summarize available techniques in a convenient form for researchers working along these lines.

References

Chotikapanich, D., Valenzuela, M. R., & Rao, D. S. P. (1997). Global and regional inequality in the distribution of income: Estimation with limited/incomplete data. *Empirical Economics, 20*, 533–546.

Chotikapanich, D., Griffiths, W. E., & Rao, D. S. P. (2007). Estimating and combining national income distributions using limited data. *Journal of Business and Economic Statistics, 25*, 97–109.

Chotikapanich, D. (Ed.). (2008). *Modeling income distributions and Lorenz curves*. Springer.

Chotikapanich, D., Griffiths, W. E., Rao, D. S. P., & Valencia, V. (2012). Global income distributions and inequality, 1993 and 2000: Incorporating country-level inequality modelled with beta sistributions. *The Review of Economics and Statistics, 94*, 52–73.

Chotikapanich, D., Griffiths, W. E., Karunarathne, W., & Rao, D. S. P. (2013). Calculating poverty measures from the generalized beta income distribution. *Economic Record, 89*(S1), 48–66.

David, H. A., & Nagaraja, H. N. (2003). *Order statistics*. Wiley.

Eckernkemper, T., & Gribisch, B. (2021). Classical and Bayesian inference for income distributions using grouped data. *Oxford Bulletin of Economics and Statistics, 83*, 32–65.

Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica, 39*, 1037–1039.

Griffiths, W. E., & Hajargasht, G. (2015). On GMM estimation of distributions from grouped data. *Economics Letters, 126*, 122–126.

Hajargasht, G., Griffiths, W. E., Brice, J., Rao, D. S. P., & Chotikapanich, D. (2012). Inference for income distributions using grouped data. *Journal of Business and Economic Statistics, 30*(4), 563–576.

Hajargasht, G., & Griffiths, W. E. (2020). Minimum distance estimation of parametric Lorenz curves based on grouped data. *Econometric Reviews, 39*(4), 344–361.

Hilomi, K., Liu, Q.-F., Nishiyama, Y., & Sueishi, N. (2008). Efficient estimation and model selection for grouped data with local moments. *Journal of the Japan Statistical Society, 38*, 131–143.

Kleiber, C., & Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences*. Wiley.

Sarabia, J. M., & Jordá, V. (2014). Explicit expressions of the Pietra Index for the generalized function for the size distribution of income. *Physica a: Statistical Mechanics and Its Applications, 416*, 582–595.

Sarabia, J. M., Jordá, V., & Remuzgo, L. (2017). The Theil indices in parametric families of income distributions—A short review. *The Review of Income and Wealth, 63*(4), 867–880.