PARAMETRIC REPRESENTATION OF THE TOP OF INCOME DISTRIBUTIONS: OPTIONS, HISTORICAL EVIDENCE AND MODEL SELECTION

Vladimir Hlasny

COMMITMENT TO EQUITY

CEQ INSTITUTE
COMMITMENT TO EQUITY
Tulane University

The CEQ Working Paper Series

The CEQ Institute at Tulane University works to reduce inequality and poverty through rigorous tax and benefit incidence analysis and active engagement with the policy community. The studies published in the CEQ Working Paper series are pre-publication versions of peer-reviewed or scholarly articles, book chapters, and reports produced by the Institute. The papers mainly include empirical studies based on the CEQ methodology and theoretical analysis of the impact of fiscal policy on poverty and inequality. The content of the papers published in this series is entirely the responsibility of the author or authors. Although all the results of empirical studies are reviewed according to the protocol of quality control established by the CEQ Institute, the papers are not subject to a formal arbitration process. Moreover, national and international agencies often update their data series, the information included here may be subject to change. For updates, the reader is referred to the CEQ Standard Indicators available online in the CEQ Institute's website www.commitmentoequity.org/datacenter. The CEQ Working Paper series is possible thanks to the generous support of the Bill & Melinda Gates Foundation. For more information, visit www.commitmentoequity.org.

The CEQ logo is a stylized graphical representation of a Lorenz curve for a fairly unequal distribution of income (the bottom part of the C, below the diagonal) and a concentration curve for a very progressive transfer (the top part of the C).

COMMITMENT TO EQUITY

PARAMETRIC REPRESENTATION OF THE TOP OF INCOME DISTRIBUTIONS: OPTIONS, HISTORICAL EVIDENCE AND MODEL SELECTION

Vladimir Hlasny[*]

CEQ Working Paper 90

MARCH 2020

ABSTRACT

Approximating the top of income distributions with smooth parametric forms is valuable for descriptive purposes, as well as for correcting income distributions for various top-income measurement and sampling problems. The proliferation of distinct branches of modeling literature over the past decades has given rise to the need to survey the alternative modeling options and develop systematic tools to discriminate among them. This paper reviews the state of methodological and empirical knowledge regarding the adoptable distribution functions, and lists references and statistical programs allowing practitioners to apply these parametric models to microdata in household income surveys, administrative registers, or grouped-records data from national accounts statistics. Implications for modeling the distribution of other economic outcomes including consumption and wealth are drawn. For incomes, recent consensus shows that among the many candidate distribution functions, only a handful have proved to be consistently successful, namely the generalized Pareto, and the 3–4 parameter distributions in the generalized beta family, including the Singh-Maddala and the GB2 distributions. Understanding these functions in relation to other known alternatives is one contribution of this review.

JEL Codes: C1, D31, D63.

Key words: statistical size distribution of incomes; top incomes measurement; parametric estimation; extreme value theory; Pareto; inequality.

Parametric Representation of the Top of Income Distributions: Options, Historical Evidence and Model Selection

Vladimir Hlasny*
March 5, 2020

Abstract

Approximating the top of income distributions with smooth parametric forms is valuable for descriptive purposes, as well as for correcting income distributions for various top-income measurement and sampling problems. The proliferation of distinct branches of modeling literature over the past decades has given rise to the need to survey the alternative modeling options and develop systematic tools to discriminate among them. This paper reviews the state of methodological and empirical knowledge regarding the adoptable distribution functions, and lists references and statistical programs allowing practitioners to apply these parametric models to microdata in household income surveys, administrative registers, or grouped-records data from national accounts statistics. Implications for modeling the distribution of other economic outcomes including consumption and wealth are drawn. For incomes, recent consensus shows that among the many candidate distribution functions, only a handful have proved to be consistently successful, namely the generalized Pareto, and the 3–4 parameter distributions in the generalized beta family, including the Singh-Maddala and the GB2 distributions. Understanding these functions in relation to other known alternatives is one contribution of this review.

I. Introduction

Since Vilfredo Pareto's (1895, 1896ab) seminal work, welfare economists have been noting regularities in the shape of the top of various income distributions and the value of using parametric approximations in place of actual income distributions. The distributions tend to be smooth and monotonically declining with the level of incomes – systematically a higher share of households have lower incomes while only a few households have higher incomes. The density functions of the distributions are also known to be heavy, meaning that the dispersion of top incomes is high, and the richest households are significantly richer than those that follow.

It has been widely accepted that these distributions can fruitfully be described by a handful of parameters. These parameters vary modestly across countries, years, income concepts, or exact income ranges. The broadly similar shape of the distribution of top incomes in various settings mean that estimates from one empirical distribution could be used to approximate or validate incomes from other distributions.

Recent academic literature has recognized the value of this tendency in light of measurement problems associated with top incomes. It is now well known that top ends of empirical income distributions suffer from various representation problems including data contamination (Cowell and Flachaire 2002, 2007), top coding (Hlasny and Verme 2018a), or sampling issues such as sampling errors, exclusion from sampling frame, and systematic nonresponse (Hlasny and Verme 2018c). Evidence from the up-to-date stock of modeling literature has shown that problematic or missing incomes can satisfactorily be imputed using parametric approaches (Hlasny and Verme 2015; Lustig 2019).

The implicit assumption motivating these analyses is that the observed top incomes are realizations from a stochastic distribution well approximated by one of the known families of parametric distribution functions. Vilfredo Pareto proposed a one-parameter distribution function (referred to as *Pareto 1*[1]), but subsequently generalized it to a two-parameter form (*generalized Pareto*, or *Pareto 2*) providing better fit across income distributions.

Benini (1897) provided an early confirmation of the relevance of the Pareto distribution for incomes, but also proposed 2- and 3-parameter variants as providing a better fit. Literature that followed in the first half of the 20th century called for higher-parametric distributions (Winkler 1950). As McDonald and Xu (1995:139) point out, a 1-parameter distribution can model well one independent characteristic in a distribution, such as the mean among top incomes. To model multiple distribution characteristics, more complex forms are needed. Over the 20th century, alternative distribution functions have been proposed as providing a closer fit to the shape of income distributions, or more flexibility across different national distributions.[2] The improvement in computing power in recent decades has meant that more complicated functional forms could be fitted, using more robust estimators and using finer-grain data.

Distribution functions nested in the *Pareto*, *beta*, and *gamma* distribution families, either as generalized or limiting forms of the baseline specifications, are some of the promulgated variations. These parametric distributions differ in the rates of decay of their densities with income levels (e.g., Pham-Gia and Duong 1989:1616–1617 illustrate for 3-parametric forms), in the existence of closed-form solutions for their moments and measures of inequality, and in the availability of statistical programs allowing their estimation and evaluation. Empirically, these

---

[1] Hereafter, for clarity of presentation, Arabic numbers will be used instead of Roman numerals, without the word 'type' or 'kind': e.g., Pareto 2 rather than Pareto type II, GB2 rather than GB of the second kind.
[2] Persky (1992) and Maccabelli (2009) offer historical perspectives on the mixed acceptance of Pareto's theory.

distribution functions have also been found to differ in their fit across different income concepts, income brackets, data types, and survey samples.

With the proliferation of alternative options for modeling top incomes, a need arose to be able to discriminate among them as well as among alternative analytical approaches. The approaches relate to the choice of the appropriate estimator, lower and upper cutoffs, and the approach for combining the parametric distribution for the top incomes under analysis with the rest of the nonparametric incomes or with statistics on their dispersion. The aim of this paper is to review the state of existing methodological and empirical knowledge, and contribute to clearing the confusion over the large number of distribution functions proposed, their names and relations, and their empirical performance. The survey provides references and statistical programs allowing practitioners to assess these distribution functions in relation to microdata in household income surveys, administrative registers, or grouped-records data generated from national accounts statistics. This survey draws heavily on McDonald and Xu (1995), Kleiber and Kotz (2003), Cowell and Flachaire (2015:13–18), and other studies cited below.[3]

The rest of the paper is organized as follows. Section II provides an overview of the families of distributions and individual functions considered in the existing literature. Section III then summarizes the evidence of their fit to empirical income distributions. Section IV reports the key lessons and implications for practitioners. Finally, three appendices review some of the methodological options for estimating and diagnosing the alternative parametric models: Appendix 1 lists a number of relevant user-written programs for Stata software, and Appendices 2 and 3 briefly review some alternative estimators and diagnostic tests for alternative models.

II. Distribution functions and their relations

Several families of smoothly decaying distribution functions have historically been considered for modeling top incomes. This includes the generalized Pareto, beta (aka, Burr; see Burr 1942), gamma (Ammon 1895), and other distribution families. These families overlap, as some of their members can be derived or approximated as special cases of distribution functions in other families (Cowell 2011). Moreover, new distribution functions resembling the standard families continue to be proposed as providing better or more flexible fit in relation to top incomes.

These distribution functions are generally well understood, and their moments and degrees of inequality represented by them can be computed analytically. (For the exponential, Pareto, lognomal, Burr and Dagum distributions, see references in Giorgi 1999:253–254.) A large arsenal of statistical programs has grown to estimate and diagnose them (see Appendix 1). This section provides a brief overview of a number of the promulgated distribution functions, some of their

---

[3] We consider only stand-alone parametric distribution functions, not their mixtures or kernel estimators (Cowell and Flachaire 2015:ch.3.2–3.3; Lubrano 2017:ch.5) – which are important modeling classes in themselves, with unique properties and challenges, and unique diagnostic tests. One reason for excluding them is that these data-driven methods are appropriate for modeling observations but not for prediction. Using them on contaminated data, or on one income bracket may not allow consistent imputation of the misreported or out-of-sample values. Kernel density estimation may also perform poorly in the tails. Relatedly, performance of these methods depends crucially on sufficient sample sizes. Single-distribution methods are more robust to data issues, because each distribution is estimated on an effectively larger sample, and the methods allow imputation and extrapolation. Finally, mixture models often produce multi-modal composite distributions, and can yield unrealistic predictions in between modes.

The finite mixture model can be estimated automatically in Stata, starting in Stata 15. For instance, `fmm 2: streg, distribution(gamma)` estimates a fixed mixture model of 2 gamma distributions. The count and the type of parametric distributions are user-selected and can be evaluated using AIC, BIC and other model-selection criteria.

properties, and the relationships among them. Sarabia and Castillo (2005) study the stability of multiple distribution functions with respect to the maximum order statistics, and their use in the analysis of incomes. More technical detail for many of the distribution functions can be found in McLaughlin (1999:App. A), Kleiber and Kotz (2003), Cowell (2011:ch.4 and A.3), or individual references cited below.

(Generalized) Pareto family

The Pareto family of distribution functions, including those introduced by Pareto himself in the 1890s and those proposed over the following century (Arnold 2015) remain among the most commonly applied functions and the first candidates to turn to for top-income modeling. This is because of their theoretical justification (Jones 2015), good empirical fit, ease of fitting, and ease of computing moments and inequality indexes. Pareto distributions exhibit slow decay based on the power law, which makes them appropriate for modeling of the heavy upper-most tail of income distributions (Gabaix 2009).

There are several commonly used specifications of the Pareto distribution functions. They differ according to the rate of decay or shape in different income ranges, and their income scale. Pareto 1 has one shape parameter, while Pareto 2 and Pareto 3 are 2-parameter models allowing for shape and scale parameters, and Pareto 4 has an additional shape parameter (Arnold 2015). Pareto 2 and 3 can be viewed as special cases of the Pareto 4 and, in turn, Pareto 1 is a special case of theirs. Pareto 3 was characterized by Arnold et al. (1986), who showed that the logistic and Zipf distributions are mere transformations of the Pareto 3.

Several less commonly used forms are also worth mentioning. Benini (1905) proposed 2- and 3-parameter generalizations of the Pareto akin to the lognormal and log-Weibull, with a sub-exponential rate of decay (Kleiber 2013). Stoppa's (1990ab) two generalized Pareto 2 distributions, obtained by exponentiating the Pareto distribution, relied on up to 4 parameters. Generalized Pareto 3 (Arnold and Laguna 1977; reviewed by Cowell 2011:ch. A.3.3) has links to other Pareto distributions as well as the exponential, Fisk's sech-squared, logistic, reverse Pareto, and Weibull (for observations with a bounded support) distribution functions as special cases. Beirlant et al. (2009), and Papastathopoulos and Tawn (2013) proposed several re-parameterized variants of 3-parameter extended generalized Pareto distributions providing better fit to heavy upper tails. These were applied to incomes by Charpentier and Flachaire (2019).

A number of distribution functions have top tails decaying at a similar rate as the Pareto, such as the closely related Champernowne (1937, 1952, 1953, 1973) family of distributions. Colombi (1990) proposed a Pareto-lognormal distribution as an alternative distribution for fitting incomes beside the commonly used Singh-Maddala and Dagum models (described under generalized beta distributions below). Johnson et al. (1994) evaluate the similarly behaving generalized exponential, logistic and Gompertz distributions. Zandonatti (2001) proposed a generalized Pareto 1 (aka, generalized Fisk) distribution, which also serves as the generalization of the Lomax (aka, shifted Pareto) distribution.[4] Lomax is a special case of the generalized Pareto with one parameter being fixed. Hamed et al. (2018) also introduced a class of transformed T-Pareto distributions, with Pareto distributions and those approximating them as special cases.

---

[4] Simon (1955) has referred to the class of Pareto-Yule-Zipf distributions, as a class of functional forms with similar properties and rates of decay that enable them to serve as approximations for one another. Pareto distributions also belong in the class of Pareto-Lévy distributions, which satisfy the weak Pareto law on the rate of decay of top incomes, and have the property of 'stable distributions' meaning that if two income sources follow the same law (up to location and scale parameters), their sum will too (Mandelbrot 1960, 1961; Zolotarev 1986; Dagsvik et al. 2013).

Among these related distributions, the Champernowne family has been accepted most widely for modeling incomes, particularly the 5-parameter Champernowne 1 and the 4-parameter Champernowne 2 (Ord 1975). Other special cases of the Champernowne distribution include Fisk's (1961) sech-squared distribution (aka, sech2, square-of-hyperbolic-sequant, log-logistic, Fisk, Champernowne-Fisk, Burr distribution with d = 1), and the Yule distribution (Yule 1925; Kleiber and Kotz 2003:242–247; Duan 2016). Yule distribution has also been found to share properties with the lognormal distribution (Aitchison and Brown 1957; Simon 1955; Simon and Bonini 1958). Fisk (1961) compares the sech-squared distribution to the general Champernowne (4-parameter Champernowne 2) and the Pareto distribution. Finally, 3-parameter conic distributions have been proposed for incomes, with Pareto 1 distribution as a special case (Houthakker 1992; Taylor 2010:472; Kleiber and Kotz 2003:97–99).

(Generalized) Gamma family

The generalized gamma family of distributions exhibit exponential decay that is faster than the power-function decay of the Pareto. They encompass all gamma distributions (Stacy 1962; Salem and Mount 1974), as well as the chi-square, exponential, lognormal, Rayleigh, and Weibull (Bartels and van Metelen 1975) distributions as special cases. Exponential distribution is a limiting case (shape parameter=1) of the Weibull distribution. Amoroso (1925) proposed a 3–4 parameter generalized gamma distribution with gamma as a special case (Lee 1984; Esteban 1986).

As with the Pareto family, a number of distribution functions have been proposed that approximate the behavior of the gamma distributions. Gibrat (1931) proposed a 2-parameter lognormal (aka, Galton, McAlister, Cobb-Douglas, or Gibrat) distribution as providing a superior fit to high incomes that fall short of the topmost tail (see also Kalecki 1945; Aitchison and Brown 1954), and Metcalf (1969) enhanced this distribution with a third shifting parameter. The Johnson S_B distribution function (aka, 4-parameter lognormal), appropriate for outcomes with spread-out lower incomes and thin (platykurtic) tails, requires minimum and maximum limits on income (Johnson 1949). Dubey (1970) proposed a compound-gamma generalization that can be re-parameterized as the beta 1 and 2 described below and the F distributions. Benktander (1970) proposed two distributions, one approximating the lognormal (aka, Benktander 1), and one approximating the Weibull distribution (aka, Benktander 2) (Kleiber and Kotz 2003:247–252). Finally worth noting here, Marshall and Olkin (1997) proposed a method for expanding the lognormal and Weibull distributions by another parameter with some useful properties.

These various distributions differ in their rate of decay, with lognormal and Weibull distributions decaying more gradually than gamma or exponential, a desirable property for modeling incomes. Lee (1984), and Chakraborti and Patriarca (2008) review the properties and relations among the lognormal, exponential, gamma and Weibull distributions.

(Generalized) Beta family

The beta family of distributions has been used for incomes with bounded support, including a maximum limit on income (such as on an interval 0 to 1; Thurow 1970). Beta 1 and 2 distributions embrace as special cases the chi-squared, exponential, F, gamma, Lomax, power, uniform, and unit gamma distributions (Tadikamalla 1980; McDonald and Richards 1987; Johnson et al. 1994;

McDonald and Xu 1995). Moreover, both the beta and gamma families can be said to belong to the generalized beta (henceforth GB) family of distribution functions.[5]

The GB distribution allows for up to 5 parameters, with GB1 and GB2 providing the most commonly used 4-parameter specifications.[6] Their flexibility allows for thin or thick upper tails. Following the exposition by McDonald and Butler (1990), most recent reviews of distributions for top-income modeling use the flexible 5-parameter GB distribution as their starting point (McDonald and Xu 1995).[7]

GB nests all the distributions in the beta and gamma families, as well as the Cauchy, chi-square, exponential, F, Fisk's sech-squared, lognormal, Lomax, power, uniform and Weibull distributions as special or limiting parametric cases (McDonald 1984; Patil et al. 1984; Cummins et al. 1990; Johnson et al. 1994; McDonald and Xu 1995). Figure 1 illustrates the relations among the more common types of distribution functions taking the GB2 as the starting point. (Other illustrations of the relations among distribution functions appear in Song 2005; or Leemis and McQueston 2008.)

Among the family of GB functions, GB2 is the most widely accepted today for modeling incomes, and various representations – and classifications – of its density function have been proposed, including as the generalized Feller-Pareto (Arnold and Laguna 1977; Arnold 2015), generalized beta prime (Patil et al. 1984), transformed beta (Venter 1984), and generalized F (in different parameterization; Kalbfleisch and Prentice 1980) distributions. The GB2 distribution as well as its 3-parameter restricted versions decay like a heavy-tailed power function. By contrast, the lognormal distribution decays exponentially with a mid-heavy upper tail, and the Pareto class of distributions has a polynomial decay.

Among the 3-parameter special cases of the GB2, the Singh-Maddala (aka, Burr, Burr 12, beta-P, Pareto IV, or generalized log-logistic distribution) and Dagum (aka, Burr 3, inverse Burr, 3-parameter kappa, beta-K) distributions have been adopted most widely. The Singh-Maddala distribution can be said to be a generalization of the Pareto, Weibull and sech-squared distributions, and behaves as Pareto among the highest incomes (Schluter and Trede 2002). The Dagum distribution itself has several commonly-used variants. Dagum 1 is the standard 3-parameter type that was promoted by Dagum (1999). Generalized 4-parameter Dagum 2 (aka, generalized logistic-Burr; evaluated by Jenkins and Jäntti 2005) and Dagum 3 distributions were also promoted by Dagum (1980, 1983).

As with the Pareto and gamma distributions, the family of GB distribution functions shares properties and even overlaps with other families. For instance, if a variable is distributed as under the power law function, then its inverse is distributed as Pareto. The Lomax (1954) distribution is a 2-parameter distribution that is a special case of the generalized Pareto distribution (special

---

[5] Another representation of the beta 1 and 2 distributions is as the 3-parameter generalized beta (G3B) distribution, or the generalized F (G3F) (Libby and Novic 1982; Pham-Gia and Duong 1989; Warsono et al. 2017). These can be viewed as special cases of the GB2 (Cowell 2011:ch.A.3.3).

[6] In fact, GB family can be further generalized by a 6-parameter 'compound confluent hypergeometric' distribution (Gordy 1998), which also covers the Gauss hypergeometric distribution (Armero and Bayarri 1994) as a special case.

[7] Figure A.1 in Cowell (2011) starts with GB2 and has the Singh-Maddala, a 3-parameter special case, at its center. The 5-parameter GB distribution has a probability density function

$$GB(y; a, b, c, p, q) = \frac{|a| y^{ap-1} [1-(1-c)(y/b)^a]^{q-1}}{b^{ap} B(p,q) [1+c(y/b)^a]^{p+q}} \text{ for } 0 < y^a < \frac{b^a}{1-c}, 0 \le c \le 1, b, p, q > 0$$

where a is a 'peakedness of the density' parameter, b and c are 'scale' parameters, p and q are 'shape and skewness' parameters, and $B(p,q)$ is the beta function (McDonald and Xu 1995). The GB2 distribution results from setting c=1; the Dagum distribution results from setting c=1 and q=1; the Singh-Maddala distribution results from setting c=1 and p=1; and the generalized gamma is the limit of the GB as $q \to \infty$ when $b = q^{1/a}\beta$, c=1 (McDonald 1984).

parameterization; aka, Pareto 2), the beta-prime distribution (scale parameter 1), the F distribution (shape parameter 1 and scale parameter 1), the log-logistic (special parameterization; shape parameter 1), the *q*-exponential distribution (special parameterization), or a mixture of exponential distributions (special parameterization) (Kleiber and Kotz 2003; Van Hauwermeiren et al. 2012). The Hall (1982) class of distributions is also said to including the Singh-Maddala, student, Fréchet (for unbounded variables with a lower limit and a heavy tail) and Cauchy distributions, or a mixture of two strict Pareto-I distributions.

Exponential Generalized Beta family

A related family of distributions derived from the GB family is the exponential generalized beta (EGB) distribution. If a random variable *Y* is distributed as GB, then $\log(Y)$ is said to be distributed as EGB.

For income distributions with extremely heavy tails, EGB can provide a viable way of parametric modeling in these tails. The EGB family covers the generalized forms of the exponential (EGB1), logistic (EGB2), Gompertz and Gumbel distributions; the 'exponential' versions of the Singh-Maddala, Fisk, power and Weibull distributions; and the standard versions of the Burr 2 (aka, generalized logistic, log-Burr 3), exponential, logistic and normal distributions as special cases (Ragab et al. 1991; Johnson et al. 1994; McDonald and Xu 1995). Generalized Gumbel (for variables with an infinite support but light tails, like the exponential distribution) is a limiting case of the EGB (with p=q) (McDonald and Xu 1995). The 2-parameter (log-)Gompertz distribution is another limiting case (refer to figure 3 in McDonald and Xu 1995:142).
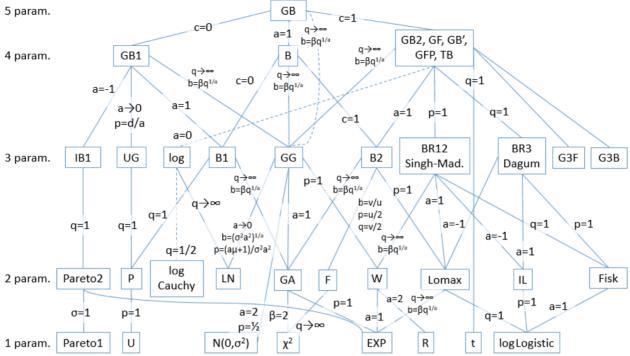
Pearson family

Many distribution functions in the generalized Pareto, beta and gamma families belong to the class of Pearson distributions (Lee 1984), even though generalized gamma, Singh-Maddala and Dagum do themselves not (McDonald and Xu 1995:138). "Pearson system purports to account for all possible shapes of observed distributions in any field [...], a general purpose system" (Dagum 2012:3).[8]

Extreme value distributions

Another category of distributions that has been cited in the top-incomes literature is the extreme value (EV; aka, von Mises, or von Mises-Jenkinson type) distributions, based on the EV theory (Kotz and Nadarajah 2000; Coles 2001). EV distribution. There are 3 types commonly used: EV1 (aka, doubly exponential) covers the exponential Weibull (aka, log-Weibull), Gumbel and Gompertz distributions; EV2 is a Fréchet-type distribution; and EV3 is the Weibull distribution (either the ordinary type, or reversed Weibull type). The generalized Pareto distribution can be represented as an EV1 distribution related to the exponential Weibull.

Figure 1. Relations among univariate distribution functions in the generalized beta family

---

[8] There are 7 types of Pearson distributions, with the generalized Pareto and beta distributions covered by types 1 and 2, the inverse Gaussian and gamma covered in type 5, the beta prime covered in type 6, and the *t* and Cauchy distributions covered in type 7. Beta 1 and 2 can be referred to as Pearson 1, and the 2-parameter gamma as Pearson 3. Amoroso's (1925) 3-parameter generalized gamma distribution has the Pearson 5 distribution (aka, inverse gamma, inverted gamma, reciprocal gamma, or Vinci) as its special case. Cauchy distribution is itself a special case of Pearson 7. Finally worth noting, transforming the distributions in the Pearson family can also lead to the so-called *Gompertz-Verhulst* distributions, with favorable properties for the fitting of extreme values (Ahuja and Nash 1967). Aiuppa (1988) provided calculations to estimate a number of distributions in the Pearson family.

5 param.   GB   c=1

c=0   a=1   q→∞ b=βq^{1/a}   GB2, GF, GB', GFP, TB

q→∞ b=βq^{1/a}

4 param.   GB1   B

q→∞ b=βq^{1/a}   c=0   q→∞ b=βq^{1/a}   q=1

a=−1

a→0 p=d/a   a=1   a=1   p=1   q=1

a=0   c=1

3 param.   IB1   UG   log   B1   GG   B2   BR12 Singh-Mad.   BR3 Dagum   G3F   G3B

q→∞ b=βq^{1/a}   p=1   q→∞ b=βq^{1/a}   p=1

q→∞   b=v/u p=u/2 q=v/2   a=1

q=1   q=1   q=1   a→0 b=(σ²a²)^{1/a} p=(aμ+1)/σ²a²   a=1   a=−1   q→∞ b=βq^{1/a}   a=−1   q=1   p=1

q=1/2   a=1

2 param.   Pareto2   P   log Cauchy   LN   GA   F   W   Lomax   IL   Fisk

σ=1   p=1   a=2 p=½   β=2   p=1   a=2 q→∞ b=βq^{1/a}   p=1   a=1

a=1   q=1   q→∞

1 param.   Pareto1   U   N(0,σ²)   χ²   EXP   R   t   log Logistic

Notes: Own analysis based on McDonald and Butler (1990), McDonald and Xu (1995). Only primary cross-row relations shown. Other relations ignored.

Key (by row, alphabetically):

B = beta; GB' = generalized beta prime; GF = generalized F; GFP = generalized Feller-Pareto; TB = transformed beta.
BR = Burr; GG = generalized gamma; IB = inverse beta; Log = logistic; UG = unit gamma.
F = variance ratio (F); Fisk = Fisk's sech-squared; GA = gamma; IL = inverse logistic; LN = lognormal; P = power; W = Weibull
EXP = exponential; N = normal; R = Rayleigh; t = Student's t; U = uniform

Other distributions

Beside the above oft-used families of distributions and their reparameterizations, various other models have been considered to represent the top income distribution, although they have not been adopted widely or have not been retained in recent literature. Among these, we could mention D'Addario's (1949) generating system of income distributions that includes the Amoroso generalized gamma and the Davis (1941ab) distributions (including Vinci distribution) as special cases (Kleiber and Kotz 2003:238–242; Dagum 1990a, 1996, 2012). Topp-Leone (1955) class of distributions has been adopted by Van Dorp and Kotz (2006) in proposing a 2-parameter generalization for fitting lower-range incomes. Also, Barndorff-Nielsen (1978) proposes a class of generalized hyperbolic distributions for fitting log incomes.

III. Empirical properties

Some distributions fit better at the top tail than above the middle of the income distribution; some distributions are more robust to data irregularities and can provide better out-of-sample fit. Parametric selection can be done through empirical testing, or using prior evidence or conceptual fit of each distribution. One important consideration is the nature of the data at hand. The incidence of income measurement errors (particularly at the uppermost tail), negative or zero incomes, and whether income data are available for individual units or as population densities in various income brackets (unit-record vs. grouped-records data) affects the estimation. Alternative distribution

functions differ in how easily their parameters can be estimated on various income distributions, and in the stability and precision of the estimated parameters (Metcalf 1972:ch.2; Merkies 1987).[9]

A "largely settled" modern consensus is that the vast bulk of incomes are best represented by lognormal, Singh-Maddala, Dagum, or GB2 distributions (Jäntti et al. 2015:310), while the uppermost 0.1–10 percent are distributed according to the extreme-values or Pareto distributions. This has to do with a "transition that occurs in the shape of an empirical income distribution from the middle, which decays exponentially, to the upper tail, which decays with respect to power" (Nirei and Souma 2007:441). One advantage of relying on distributions in the GB family is that they have known properties, and inequality indexes under these distributions can be derived analytically. For instance, the Gini coefficient can be computed using the estimated coefficients for all GB class distributions, except for the 5-parameter GB itself (Dagum 1977; McDonald 1984; Boundarian et al. 2003).

This section reviews existing evidence of the empirical fit of alternative parametric models, and draws lessons for practitioners' future work. One aim is to show how the modeling literature has evolved from using simple distribution functions to more complex ones, how some classes of distributions have been sidelined in favor of better performing ones, and how the nature of data at hand affect the modeling choice.

Pareto 1

Pareto (1896) concluded that his one-parameter model provided a good fit for historical incomes in various places including England, Italian and Peruvian cities, German states, and Paris. On the general nature of the Pareto coefficient, Davis (1941b) noted that the coefficient has been estimated at similar levels for years as far apart as 1471 and 1894. "There is no reason to believe that a significant change has occurred in [the Pareto] parameter during the depression or afterwards, although there has been a tendency for it to increase as the tax burden has grown since 1933" (Davis (1941b:31).

More recent literature has restricted the use of the Pareto to a narrower and narrower range of top incomes (Jenkins 2017; Charpentier and Flachaire 2019) and considered other parametric functions for lower incomes. Aitchison and Brown (1954, 1957) found that for lower earnings in homogeneous occupational groups – 1950 UK data for nine agricultural occupations – a lognormal distribution provides a good fit, while for higher earnings a power law distribution, such as Pareto, fits better. Clementi and Gallegati (2005), using the US Panel Study of Income Dynamics (PSID) for 1980–2001, the British Household Panel Survey (BHPS) for 1991–2001, and the German Socio-Economic Panel (GSOEP) for 1990–2002, propose using a two-parameter lognormal (or an exponential) distribution for the low and middle income population (up to the 97th–99th percentile), and the Pareto distribution for the highest income group (1%–3% of the population).

The parsimonious Pareto 1 model fares well even compared to more-parametric alternatives. Gabaix (2009) concluded that the 2-parameter lognormal distribution does not necessarily outperform the Pareto 1 on the topmost incomes, on account of the degrees of freedom lost. Hlasny and Verme (2018ab) similarly argued that Pareto 1 provides more consistent results than the Pareto 2 for data in the US Current Population Survey (CPS), and than the GB2 for data in the EU Surveys on Income and Living Conditions and the US CPS. Hlasny and Verme (2018c), Ibragimov and Ibragimov (2018), and Hlasny (2019) found that the Pareto 1 fit well the distribution of incomes

---

[9] In prior decades, an important property of distribution functions was the ease of their estimation (Campano and Salvatore 2006). Each additional estimable parameter represented a challenge for estimation.

in transitional economies – Egypt, Russia and Mexico, respectively – and that the results were surprisingly similar to those obtained in high-income countries. Hlasny and Intini (2015) evaluated the fit of the Pareto 1 model to the distributions of household consumption in 11 Arab-region surveys, using lower cutoff points between the 80th and the 99.9th percentiles. They concluded that the tops of the Egyptian, Palestinian and Tunisian distributions are adequately described by Pareto 1, but Jordan and Sudan suffer from influential observations, which affects the estimation of the Pareto coefficients.

With grouped-records data, an important consideration is the assumed distribution of incomes in each income range. Taking advantage of the ease of working with the Pareto distribution, and taking Pareto's (1896) lead, densities in each income bracket can be modeled as Pareto-distributed (Feenberg and Poterba 1993; Piketty 2001). Gastwirth (1972) considers the extremes of the least versus most inequality in each income range to compute bounds on the composite inequality indexes. Alfons et al. (2013) derive the parameters of the Pareto 1 distribution in the presence of sampling weights, and numerically show their desirable properties.

Atkinson et al. (2011) summarize the literature contending that the distribution of top incomes is best approximated by the Pareto distribution.[10] Atkinson (2017) promotes the Pareto 1 distribution as a valuable point of departure for modeling top incomes, but he favors more complex distributions including Pareto 2 for specific income distributions. Harrison (1981), using the British New Earnings Survey, concludes that the Pareto distribution fits the dispersion of top 15–20 percent of incomes well, provided that an appropriate criterion of fit is used (i.e., Gastwirth-Smith rather than chi-squared test). Disaggregating data by occupational group leads to better estimation results than using all data jointly, but the Pareto model outperforms the lognormal model both in the aggregated and the disaggregated data. The estimated Pareto coefficients differ across occupational groups, and even across choices for the lower cutoff point for a single occupational group – while the coefficient appears more stable in the aggregate sample. The poor performance of the lognormal distribution appears to be due to the income differences across occupational groups.

As a final note, a number of studies take advantage of the power-law property, also known as van der Wijk's law, that the mean of a Pareto-distributed variable can be approximated as $\bar{x} = E(x|x \geq x_0) = x_0 \, a/(a-1)$, where a is the Pareto coefficient. This property makes it easy to approximate the mass of top incomes above a specific threshold. Bernstein and Mishel (1997) offered an early application. This is followed in recent studies by the World Inequality Lab, which have adopted the Pareto approximation and van der Wijk's law extensively.[11] Blanchet et al. (2018) estimated different Pareto distributions across income fractiles, and used a cubic spline to restrict the fitting and provide smoothness between fractiles ranges.

Pareto 2

---

[10] Beside incomes and consumption, Pareto distribution has also been advocated for modeling top wealth (Eckerstorfer et al. 2016; Vermeulen 2016; Credit Suisse 2019; Kennickell 2019). A convex combination of the Pareto and lognormal has been used by Hruschka et al. (2015).

In fact, the Pareto class of distributions may be used even in the bottom tail of incomes. Van Kerm (2007) fitted an inversed Pareto distribution to negative incomes in the EU-SILC surveys, and Hlasny et al. (2020) did so for Mediterranean-region surveys in the Luxembourg Income Study database. Toda (2012) evaluated positively the double Pareto and the double Pareto-lognormal distributions on US CPS and PSID data. However, newer evidence suggests that exponential distribution may fit negative incomes better (Jenkins and Jäntti 2005; Jäntti et al. 2015).

[11] https://wid.world/world-inequality-lab/ , e.g., Piketty and Saez (2003); Blanchet et al. (2017,2018b).

In recent years, understanding has shifted toward advocating the more flexible Pareto 2 distribution, particularly when a larger range of incomes are analyzed (Atkinson 2017). This shift in preference is based on the Pickands–Balkema–de Haan theorem (Balkema and de Haan 1974; Pickands 1975), also called the second theorem in the EV theory, giving the asymptotic tail distribution of values of a random variable above a certain threshold. For a large class of true distributions, as the lower threshold increases, the distribution is said to converge to the Pareto 2.[12]

Burkhauser et al. (2009) used one of Stoppa's (1990a) exponentiated Pareto distributions to correct for top coding in US-CPS income data, because the distribution function has a heavy tail and can be made to mimic the mean as well as the variance of the raw top incomes.

Departing from a baseline Pareto 1 specification, Jenkins (2017) found strong support for the Pareto 2 model for UK incomes, except in the complete top of the distribution (top 0.5–1%, or even fewer topmost incomes depending on year) where the Pareto 1 performs as well as the Pareto 2 or even better. Based on Jenkins (2017), the most appropriate left-truncation threshold varies greatly across survey waves, from below the 85th percentile to as high as the 99.9th percentile for Pareto 1, even though it is more consistently between the 95th and the 99th percentile for Pareto 2.

Charpentier and Flachaire (2019) compared Pareto 1, Pareto 2, and Beirlant et al.'s (2009) 3-parameter extended Pareto distribution on the US CPS and the South African National Income Dynamics Study surveys. They found that the one-parameter Pareto 1 is clearly dominated by the two/three-parametric alternatives, but the two-parameter Pareto 2 and the three-parameter extended Pareto cannot be ranked easily across various modeling specifications. Only among the topmost 1 percent or fewer incomes the Pareto 1 and 2 perform similarly, and the extended Pareto distribution appears to outrank them in the precision of fitting the empirical distribution. Finally, Charpentier and Flachaire (2019) report that the extended Pareto distribution also outperforms the Pareto 4 model (Arnold 2015) and a variant of an extended generalized Pareto model (EGPD3 in Papastathopoulos and Tawn, 2013).[13]

Lognormal

Aitchison and Brown (1954, 1957) promoted the lognormal distribution as offering good fit to a wide range of incomes. Incomes between the 10th and the 80th percentiles (Fournier 2015), or even between the 5th and the 95th percentiles (Montroll 1978) were found to follow the lognormal distribution well.

Metcalf (1969, 1972) found that the lognormal distribution overcorrects for the positive skew of the distribution of US-CPS incomes in 1949–1965. He proposed a generalized 3-parameter shifted version (i.e., displaced lognormal) that fits better than the lognormal and the exponential distributions, particularly lower down in the distribution. Metcalf (1972) described a method for

[12] We can also point out the related Fisher–Tippett–Gnedenko theorem, or the first EV theorem (Fisher 1930), pertaining to the distribution of extreme order statistics. To the extent that the distribution converges, it is said to converge to only one of 3 possible models, all in the class of generalized EV distributions: the Gumbel, the Fréchet, or the Weibull.

[13] Charpentier and Flachaire (2019) also discuss three potential sources of bias in the estimation of a tail index (such as the Pareto coefficient): misspecification bias, estimation bias and sampling bias. Misspecification bias can be evaluated - with some success - using graphical methods such as the quantile-quantile (QQ) plot or the Pareto diagram. Increasing sample size does not mitigate the misspecification bias; only a more appropriate parametric form or a different (i.e., higher) threshold can mitigate it. Estimation bias arises because the observed empirical upper tail differs from the underlying population upper tail, and the divergence can be systematic, on account of oversampled (rare outliers) or undersampled units (underreporting, top coding). Having a larger sample size may partially mitigate these problems. Sampling bias comes from survey nonresponse or incidental inclusion of out of sampling frame units. This, once again, is unlikely to be mitigated by increasing the sample size.

estimating the displaced lognormal distribution by anchoring it at three symmetrically-spaced quantiles in the empirical data. Hart (1983) estimated the moments of the bivariate lognormal distribution for British 1971–1978 male earnings accounting for household size and composition.

While Gibrat's and Metcalf's 2 and 3-parameter lognormal distributions have been shown to provide inferior fit at the complete top, they continue to be used for incomes lower down in the distribution. In rural China, the lognormal distribution provided 'perfect' fit to income data (Kmietowicz and Ding 1993). Harrison (1981) concludes that the lognormal distribution fits well up to the 80th percentile of incomes. Applying the lognormal distribution to individual occupational groups, and adding the groups to one another, he fails to confirm the summation property of lognormal distributions, suggesting that not all occupational groups exhibit incomes distributed as lognormal. McDaniel et al. (1988) successfully fitted a tri-variate form of the Johnson S_B (4-parameter lognormal) distribution to monthly income, financial assets, and a physical disability indicator. Finally, Cowell (2015:15) concludes that the lognormal model fits the distribution of wages, but not of broader income concepts.

Other Pareto-like models
Winkler (1950; as reported by Kleiber 2013) fitted a 3-parameter Benini distribution to US incomes of 1919. Thatcher (1968) used the Champernowne distribution to model UK earnings, and Campano (1987) used the 5-parameter Champernowne on 1969 US incomes. Colombi (1990) estimated that a Pareto-lognormal distribution fit the Italian 1984–1986 family incomes as closely as the Singh-Maddala and Dagum models. Chung and Cox (1994) modeled the dispersion of superstar performers' earnings using the Yule distribution, and found that it fits very well. Fisk distribution was successfully fitted to incomes in homogeneous occupations (Fisk 1961), in 17 Peruvian cities during 1971–1972 (Arnold and Laguna 1977), and to the 1972 gross weekly earnings for full-time male workers in seven occupational groups in the UK, where it outperformed the lognormal or the Pareto distributions (Harrison 1979, 1981). Fisk distribution also fared well on the UK Family Expenditure Survey (Henniger and Schmitz 1989). Linear and quadratic special cases of conic distributions (evaluated positively by Kleiber and Kotz 2003) provided better fit to US individual earnings than the Singh-Maddala (Houthakker 1992; Taylor 2010).

More recently, Dagsvik et al. (2013) found that a 3-parameter Pareto-Lévy distribution outperforms the 4-parameter GB2 distribution on grouped-records income data from 5 out of 8 evaluated OECD countries. Fournier (2015) concluded that the Pareto 3 and 4, Champernowne, Fisk and Singh-Maddala distributions fit well the observed distribution of top incomes in the French fiscal authority data for 2006.

Estimation specifications: Lower and upper cutoff points for Pareto models
The discussion in the preceding paragraphs suggests that the choice of functional form and the quality of fit of Pareto models depend on the income bracket evaluated. Various lower cutoff points have been used in existing studies, from the 80th (Piketty et al. 2017), 90th (Chancel and Piketty 2017), 99th (Burkhauser et al. 2016), or as high as the 99.5th percentile of survey observations (Jenkins 2017). Coles (2001), and Gilleland and Katz (2005) explore the determination of the lower cutoff for the Pareto distribution. Cowell and Flachaire (2007), and Davidson and Flachaire

(2007) choose the 90th percentile as the lower point based on the graphical Hill-plot approach (Drees et al. 2000).14

Hlasny and Verme (2018abc) compare different combinations of lower and upper truncation points in survey data, and argue for a choice that maintains a large sample size for fitting and at the same time avoids using contaminated uppermost values for estimation. Jenkins (2017) evaluates various left-truncation points for the Pareto 1 and 2 distributions analytically, using alternative diagnostic tests. Using British administrative data, he identifies the optimal cutoff point at between the 90th and the 99th income percentile in various years, and typically at the 95th or higher percentile. Charpentier and Flachaire (2019) put the left cutoff point for the Pareto 1 model even higher, at the 99th or higher percentile.

(Generalized) Gamma

Gamma distribution was promoted as recently as the 1970s as offering good fit among 2-parameter models, and was evaluated positively on incomes in the US (Salem and Mount 1974) and the Netherlands (Bartels and van Metelen 1975). Salem and Mount (1974) showed that the gamma fitted better than the lognormal for the middle-income range of US family incomes during 1960–1969. In earlier decades, the 3–4 parameter generalized gamma distributions by Amoroso (1925) were successfully used for German and Italian incomes, even though the fit was weaker for US incomes (Davis 1941a:405–406).

More recent evidence suggests that distributions with more parameters, or other functions relying on the same number of parameters provide a closer fit to income data. Kloek and Van Dijk (1978) found that the lognormal and gamma functions showed poor fit to Dutch earnings, instead favoring the log-*t* and the log Pearson 4 models, and in some applications also a 3-parameter generalized gamma and a 4-parameter Champernowne function. Hippel et al. (2012) found that the extended generalized gamma, power normal and newly-proposed power logistic distributions compared favorably to Dagum, GB2 and logspline distributions on grouped-records household or family data in the US. Most recently, however, studies have used functions in other distribution families, including the 2-parameter Pareto 2, or the 3–5 generalized beta 2 functions.

(Generalized) Beta

Beta distributions have been at the forefront of empirical modeling of the overall income distributions. A large body of research has focused on comparing individual members of the beta family. Among the early research, a 3-parameter beta distribution was successfully fitted to the 1949–1966 and 1952–1980 US income data (Thurow 1970; Slottje 1984).

Recent decades have seen a quest for using more parametric representations that offer superior flexibility. One conceptual argument for using distribution functions with 3+ parameters is that the Lorenz (1905) functions of any two such distributions can show any relationship with one another. By contrast, Lorenz curves of 1–2 parameter distributions (such as F, lognormal, Lomax, Pareto, Weibull etc.) cannot intersect. To allow two estimated distributions to have intersecting Lorenz curves – as is common with real-world distributions – distribution functions with 3+ parameters should be applied.

Higher-parametric distributions also appear to offer better fit at the top. Majumder and Chakravarty (1990) proposed a four-parameter generalization of the Dagum and Singh-Maddala

---

14 The Hill plot is a line showing the Pareto 1 coefficients estimated with various lower cutoff points. Under the Pareto law, the line should be horizontal. Hence, departures from that shape are an indication that the original data are not Pareto-distributed, and 'Hill's horror plot' is an indication of a specific systematic type of deviation.

distributions that exhibited better fit for personal income than the Dagum, Singh-Maddala and Champernowne distributions. However, McDonald and Mantrala (1993, 1995) showed that Majumder and Chakravarty's model was a mere re-parameterization of the GB2 model. McDonald and Mantrala favored Dagum and Singh-Maddala models for 1970–1990 US family incomes. Only the four-parameter GB2 and the five-parameter GB distribution outperformed the Dagum 1 (McDonald and Xu 1995). Gordy (1998) proposed a 6-parameter 'compound confluent hypergeometric' distribution to household liquid assets in the US, and found that the added parameters were significant and useful, relative to the more restrictive GB and gamma models.

Across the majority of empirical studies of income, GB2 has emerged as the preferred functional form representing well not only the top but also much of the support of the income distribution (Nadarajah and Kotz 2006; Jenkins et al. 2011; Graf and Nedyalkova 2013; Graf and Tille 2013; Chotikapanich et al. 2018). A number of empirical studies – Singh and Maddala (1976), Dagum (1977, 1980, 1983), McDonald (1984), Butler and McDonald (1989), Majumder and Chakravarty (1990), McDonald and Xu (1995), Bantilan et al. (1995), Victoria-Feser (1995, 2000), Brachmann et al. (1996), Bordley et al. (1996), Tachibanaki et al. (1997) and Bandourian et al. (2003) – found that the GB2, the Singh-Maddala and the Dagum distributions perform better than distributions with three or fewer parameters.

For 1970–1980 US family incomes, McDonald (1984) preferred the GB1, GB2, Singh-Maddala and generalized gamma distributions to the gamma and Weibull distributions, which themselves outperformed the Fisk distribution. For grouped-records Japanese income data for 1963–1971, Singh-Maddala slightly outperformed the Fisk, which outperformed beta, gamma, lognormal and Pareto 2 (Suruga 1982). Atoda et al. (1988) and Tachibanaki et al. (1997) found that more-parametric models were needed for Japanese incomes. Chotikapanich et al. (2007) applied the GB2 successfully to grouped-records data in China.

Bandourian et al. (2003) found that the Weibull, Dagum and GB2 distributions provide the best fit for group-records pre-tax incomes among ten GB-class distributions with 2, 3, and 4 parameters, respectively (compared to gamma, lognormal, beta 1, beta 2, generalized gamma, Singh-Maddala, GB1, GB). They used 23 high and upper-middle income countries, and income data grouped into 20 5-percentage point groups. McDonald et al. (2013) found that the GB2 approximates well the skewness and kurtosis in a set of 78 income distributions, compared to the generalized gamma, Dagum, Singh-Maddala, beta 1 and 2, and GB1 distributions. Okamoto (2013) evaluated a newly proposed k-generalized beta distribution favorably compared to GB2, on 60 income distributions in the Luxembourg Income Study database.

Brzezinski (2013) found that the GB2 model outperforms the Dagum and Singh-Maddala models on most distributions of disposable household income per adult equivalent in Central and East Europe during 1991–2010, with a few exceptions where Dagum and the GB2 perform similarly. The GB and GB2 models have been found to provide the best fit for US family income (grouped-records for 11 or 21 quantiles), while the lognormal distribution provide the worst fit (the other evaluated distributions being GB1, gamma, generalized gamma, beta, beta 1, beta 2, Singh-Maddala, and Dagum). The GB2, Dagum, and gamma are the best-fitting four-, three-, and two-parameter distributions.

The conventional wisdom from recent research is that the GB2 distribution can be used among middle and upper-middle incomes, but in the upper-most tail one might prefer a power-law distribution such as the Pareto 1 instead. Because the right tail of incomes may also be contaminated by nonresponse or income under-reporting, right-truncation to the GB2 distribution may be applied at the 95–99th percentile (Hlasny and Verme 2018ab; Hlasny 2019). Finally,

because bottom incomes may not be approximated well by the GB2 distribution, Jenkins et al. (2011:69) propose estimating the GB2 distribution on income data left-truncated at the 30th percentile.

Singh-Maddala

In response to Salem and Mount's (1974) advocating of the gamma distribution, Singh and Maddala (1976) compared the gamma, the lognormal and the Singh-Maddala distributions, and found that the latter provided a better fit to the 1960–1972 US family incomes. In regard to the degrees of freedom used up in estimation, particularly on grouped-records data, Cramer (1978) called for the development of more appropriate model-selection criteria recognizing the nonlinear nature of parameter estimation. Commenting on Singh and Maddala's (1976) assumptions and model-fit statistics, he questioned whether their distribution really outperformed the gamma distribution.

Nevertheless, subsequent research has largely confirmed the superior fit of the Singh-Maddala model to income distributions worldwide, compared to lower-parametric specifications. McDonald and Ransom (1979ab) compared the fit of the lognormal, gamma, beta and Singh-Maddala distributions to decile-group data for 1960–1975 family incomes. The Singh-Maddala outperformed the beta distribution across most years, the beta in turn outperformed the gamma, and the gamma outperformed the lognormal. This ranking held when the same estimation technique was used across the competing distribution functions, but not necessarily otherwise. Brachmann et al. (1996) concluded that the GB1 and B1 distributions did not provide a good fit to German net household incomes, while the Singh-Maddala and gamma fared well among the classes of 3- and 2-parameter distributions.

Comparing the lognormal, gamma, log-logistic, Weibull, generalized gamma and Singh Maddala distributions, Tachibanaki et al. (1997) concluded that the generalized gamma fit better than the lognormal, the gamma, or the Weibull. The Singh-Maddala outperformed the Weibull and the log-logistic. Different estimations yielded vastly different parameters and inequality indexes, on account of 1) different estimators, and 2) data differences such as individual/grouped-records data. Tachibanaki et al. (1997) also compared income distributions estimated using individual-level data (maximum likelihood method) versus grouped-records data (minimum chi-square or maximum likelihood method). They found that grouped-records data allowed estimating a well-fitting distribution even when the values of individual incomes did not allow the estimation on individual-level data (i.e., the beta and the Johnson S_B distributions).

Biewen and Jenkins (2005) evaluated the fit of the Singh-Maddala and Dagum distributions for income distributions in the US, Great Britain and Germany, and concluded that the Singh-Maddala distribution performed better. Similarly, Jäntti et al. (2015) found that the Singh-Maddala distribution offered a good fit to the distributions of positive incomes in the US, Germany, Italy, Luxembourg and Spain.

Dagum and log-logistic models

The natural alternative to the Singh-Maddala model is the other 3-parameter Burr-family specification, the Dagum function. As the studies in the previous and this section show, the evidence of their relative performance is mixed. Dagum (1983) found that his distributions (Dagum 1, 2 and 3) outperformed the lognormal, gamma and Singh-Maddala models on US family

incomes. Using the UK data, the Dagum 1 was also shown to outperform the gamma distribution.[15] Dagum also found that a 2-parameter log-logistic distribution, as a special case of the Dagum model, fit Canadian data well. Suruga (1982) and Atoda et al. (1988) also evaluated the log-logistic model positively for Japanese incomes, relative to the lognormal, gamma, and beta distributions, but found it to be comparable to the 3-parameter Singh-Maddala distribution.

Using French incomes, Espinguet and Terraza (1983) confirmed that the Dagum 2 distribution provided a superior fit compared to the Weibull, the Singh-Maddala, the Box-Cox-transformed logistic, the 3-parameter lognormal, and the 4-parameter beta 1 models. Dagum 1–3 showed good fit to Italian incomes (Dagum and Lemmi 1989), and incomes in the Buenos Aires region (Botargues and Petrecolla 1997, 1999ab), where they outperformed the lognormal and the Singh-Maddala. Bantilan et al. (1995) found that the Dagum 1 provided a good fit for family incomes in the Philippines. Campano and Salvatore (2006) reported that the United Nations had concluded that the 4-parameter Dagum distribution "was the model of choice for over 60 countries" (p.51). The Dagum fit as well as or better than the 5-parameter Champernowne distribution (Campano 1987). In fact, estimating "one serves as a check on [estimating] the other" (Campano and Salvatore 2006:53). Jäntti et al. (2015) also favored the Dagum distribution for positive incomes (and the exponential distribution for negatives).

In relation to other GB-family distributions, Bordley et al. (1996) concluded that the Dagum 1 worked better than other 3- and 4-parameter models, including the gamma and the GB1, but the GB2 remained as the most preferred form. Kleiber (1996, 2008:110) listed all the known empirical applications of the Dagum distribution, and summarized them by noting that the Dagum outperformed other models with 3 or fewer parameters, but concurred that the 4-parameter GB2 could perform better.

Beyond beta: other distribution functions

A small number of studies have assessed the empirical fit of models outside the GB family, but their results are difficult to compare because of the lack of consistent benchmark specifications and test statistics. Here is the limited evidence. Horsky (1990) used a 2-parameter EV distribution (Gumbel distribution in the EGB class) to model wages and total incomes in the US Census data. The model fit wages better than the lognormal distribution, and as well as the best of the evaluated 3- and 4-parameter models. Chumacero and Paredes (2005) found that the EV distribution outperformed the generalized Cauchy, logistic and other functional forms in fitting the Chilean income distribution, with the lognormal distribution also having desirable properties.

Van Dorp and Kotz (2006; also Kotz and van Dorp 2004) applied the 'reflected' generalized Topp–Leone distribution to 2001 US household incomes, and found that it performed better at lower ranges of incomes than the beta family of distributions. Atoda et al. (1988) found that a 4-parameter *J*-shaped beta distribution, related to the Topp–Leone distributions, and a 2-parameter Weibull distribution provided a good fit to narrow income categories such as farmers' primary incomes. The Weibull performed as well as the 4-parameter Singh-Maddala. Barndorff-Nielsen (1978) successfully fitted the generalized hyperbolic distribution to the 1962/1963 personal log-income data for Australian residents.

IV. Summary and implications for practitioners

---

[15] Dagum (1999) also proposed using the Dagum 1 distribution for the estimation of the joint distribution of income and wealth. Dagum 1 was used to model positive incomes, an exponential distribution for negative data, and a point-mass for zero incomes.

Fitting smooth parametric curves to the top of empirical income distributions is valuable for describing the general shape of the income density functions, as well as for correcting income distributions for various top-income measurement problems. Thanks to the extensive research undertaken in the past quarter-century, we now have access to a large body of empirical evidence and tools for estimating and diagnosing alternative models, and for replacing actual top incomes with model-generated ones. Alternative estimators have been developed, and a number of statistics exist to inform our model and estimator selection. This body of knowledge helps us understand the nature of the process underlying the distribution of incomes, and to some degree even the mechanisms driving the process. To the extent that observed incomes suffer from various top-tail measurement problems, the existing knowledge enables us to overcome them.

The proliferation of distinct branches of modeling literature over the past decades gave rise to the need to survey the alternative modeling approaches and be able to discriminate among them based on their fit to empirical data. To this end, we have reviewed the state of methodological and empirical knowledge regarding the adoptable distribution functions, and provided references to original sources and statistical programs (refer to Appendix 1) allowing practitioners to apply these parametric functions to microdata in household income surveys, administrative registers, or grouped-records data generated from national accounts statistics.

Many of the distribution functions reviewed in this study have theoretical or empirical merit. With arbitrary income data at hand, it is a priori unclear which distribution function will provide the best fit for the incomes used in estimation, as well as for those outside of the estimation sample. A typical confirmatory research path is to choose a set of theoretically or empirically comparable parametric forms, fit each of them, and then evaluate their degree of fit.[16] The alternative approaches to fitting each distribution function are beyond the scope of this survey, but a brief, non-exhaustive review is provided in Appendix 2. Similarly, some guidelines how to diagnose the alternative distribution functions and rank them are provided in Appendix 3.

Our review suggests that among the large number of candidate families of distribution functions, only a handful have proved to be consistently successful in terms of feasibility of fitting and their fit to the original data, and have been used over the years and to this day. The choice depends in part on the range of incomes that needs to be represented parametrically. Across many countries and years, 3–4 parameter distributions in the generalized beta family – namely the Singh-Maddala, the Dagum and the GB2 distributions – have proved successful at representing the vast mass of income distributions, approximately from the 30_{th} to the 95_{th} percentile. The lognormal distribution, itself a 2-parameter member of the generalized beta family, offers a weaker fit for that range of incomes, but may also be estimated due to degrees-of-freedom considerations, say on grouped-records data. For top incomes, flexible power function distributions, including 2–3 parameter members of the EV and Pareto families exhibit a better-fitting rate of decay of income densities. Finally, in the uppermost tail of the top 0.5% or up to top 5% of income observations, due to data sparseness and heavy rate of decay, the Pareto 1 serves as an adequate parsimonious model.

Model selection clearly also depends on the properties of the data at hand. Unit-records data call for functional forms and estimators that are less sensitive to individual income values.

---

[16] Esteban (1986) propose an alternative search path: formulating a set of hypotheses, or stylized facts, about the empirical distribution, and then identifying the (only) parametric distribution that satisfies them. For instance, hypothesizing that the income share elasticities have a constant rate of decline (plus two other assumptions), he shows that the 3-parameter generalized gamma is the only density function within its class that satisfies them.

Grouped-records data, on the other hand, call for functional forms and estimators that are less demanding with regard to the degrees of freedom, such as distributions with fewer parameters. The narrower the income concept (say, wages or primary income), and the more homogeneous the sample (say, farmers, or workers in the same occupation, or residents in one city), the better the fit of simpler distribution functions such as the lognormal, the Weibull, or the Pareto 1 models. Finally, to the extent that multiple types of income must be added together (say, income sources) or juxtaposed (say, those of two subgroups), parametric functions that are additive or that allow the Lorenz curves to cross (e.g., 3+ parametric functions), respectively, would facilitate the most relevant comparisons. In this respect it should be noted that there is presently limited record of stability of distribution functions across disparate components of total incomes.

To summarize the evidence presented in this review, a large number of distribution functions have been considered over the past 120 years for modeling economic outcomes. We have introduced several classes of these models, including their theoretical properties and empirical record of fit. Among the dozens of models, only a handful have proved to be consistently successful, namely the generalized Pareto, and the 3–4 parameter distributions in the generalized beta family, including the Singh-Maddala, Dagum and GB2 distributions. Recent literature has focused on ascertaining the settings under which each model is appropriate, which estimator should be used in fitting each model, and how to compare or replace the observed values with the model-estimated ones. It should be re-emphasized that the distribution functions presented here can perform only as well as the income observations they are based on. In the presence of top-income measurement biases, the models should be estimated on right-truncated samples or on samples where the measurement errors have been allayed.

The quest for modeling best practices does not stop here. Future academic research should focus on bridging the disparate branches of modeling literature and developing a single-stop tool for model selection, estimation and evaluation. Among practitioners, a consensus should be sought on how to embrace the findings most effectively to, say, track measurement issues such as tax evasion, report on developmental policies, and improve fiscal targeting and redistribution.

References

Ahuja, J.C., and Stanley W. Nash (1967) The generalized Gompertz-Verhulst family of distributions, Sankhyā: The Indian Journal of Statistics, Series A 29(2):141–156.

Aitchison, John, and James A.C. Brown (1954) On criteria for descriptions of income distribution, Metroeconomica 6:88–98.

Aitchison, John, and James A.C. Brown (1957) The Lognormal Distribution. Cambridge: Cambridge University Press.

Aiuppa, Thomas A. (1988) Evaluation of Pearson Curves as an Approximation of the Maximum Probable Annual Aggregate Loss, Journal of Risk & Insurance 55(3):425–441.

Alfons, Andreas, Matthias Templ, and Peter Filzmoser (2013) Robust estimation of economic indicators from survey samples based on Pareto tail modelling, Journal of the Royal Statistical Society: Series C (Applied Statistics) 62:271–286.

Amoroso, Luigi (1925) Ricerche intorno alla Curva dei Rediti, Matematica Pura ed Applicata 2(1):123–159.

An, Di, and Roderick J.A. Little (2007) Multiple imputation: an alternative to top coding for statistical disclosure control, Journal of the Royal Statistical Society A 170:923–940.

Armour, Philip, Richard V. Burkhauser, and Jeff Larrimore (2016) Using the Pareto Distribution to Improve Estimates of Topcoded Earning, Economic Inquiry 54(2):1263–1273.

Arnold, Barry C. (2008) Pareto and Generalized Pareto Distributions. In Duangkamon Chotikapanich (ed), Modeling Income Distributions and Lorenz Curves, New York: Springer Verlag.

Arnold, Barry C. (2015) Pareto Distributions (2nd ed.), Boca Raton, FL: CRC Press.

Arnold, Barry C., and Leonor Laguna (1977) On Generalized Pareto Distributions with Applications to Income Data. International Studies in Economics, Monograph #10, Department of Economics, Iowa State University.

Arnold, Barry C., Christopher A. Robertson, and Hsiaw-Chan Yeh (1986) Some properties of a Pareto-type distribution. Sankhyā: Indian Journal of Statistics – Series A 48(3):404–408.

Atkinson, Anthony B. (2017) Pareto and the upper tail of the income distribution in the UK: 1799 to the present, Economica 84(334):129–156.

Atkinson, Anthony, Thomas Piketty, and Emmanuel Saez (2011) Top incomes in the long run of history, Journal of Economic Literature 49:3–71.

Atoda, Naosumi, Terukazu Suruga, and Toshiaki Tachibanaki (1988) Statistical inference of functional forms for income distribution, Economic Studies Quarterly 39(1):14–40.

Balkema, August A., and Laurens de Haan (1974) Residual life time at great age, Annals of Probability 2:792–804.

Bandourian, Ripsy, James B. McDonald, and Robert S. Turley (2003) A comparison of parametric models of income distribution across countries and over time, Estadística 55:135–152.

Bantilan, Maria C.S., N.B. Bernal, M.M. de Castro, and Joanne M. Pattugalan (1995) Income distribution in the Philippines, 1957–1988: An application of the Dagum model to the family income and expenditure survey (FIES) data. Pp.11–43 in C. Dagum and A. Lemmi (eds.) Research on Economic Inequality, Vol. 6: Income Distribution, Social Welfare, Inequality and Poverty. Greenwich, CT: JAI Press.

Barndorff-Nielsen, Ole (1978) Hyperbolic distributions and distributions on hyperbolae, Scandinavian Journal of Statistics 5(3):151–157.

Beirlant, Jan, Elisabeth Joossens, and Johan Segers (2009) Second-order refined peaks-over-threshold modelling for heavy-tailed distributions, Journal of Statistical Planning and Inference 139:2800–2815.

Benini, Rodolfo (1897) Di alcune curve descritte da fenomeni economici aventi relazione colla curva del reddito o con quella del patrimonio, Giornale degli Economisti 14:177–214. English translation in Rivista di Politica Economica 87:701–744, 1997.

Benini, Rodolfo (1905) I diagrammi a scala logaritmica (a proposito della graduazione per valore delle successioni ereditarie in Italia, Francia e Inghilterra), Giornale degli Economisti 16:222–231.

Benktander, Gunnar (1970) Schadenverteilungen nach Grösse in der Nicht-Lebensversicherung [Loss distributions by size in non-life insurance]. Bulletin of the Swiss Association of Actuaries (in German):263–283.

Bernstein, Jared, and Lawrence Mishel (1997) Has wage inequality stopped growing? Monthly Labor Review 120(12):3–16.

Biewen, Martin, and Stephen P. Jenkins (2005) A framework for the decomposition of poverty differences with an application to poverty differences between countries, Empirical Economics 30:331–358.

Blanchet, Thomas, Juliette Fournier, Thomas Piketty (2017) Generalized Pareto curves: theory and applications, WID.world working paper 2017/3.

Blanchet, Thomas, Ignacio Flores, and Marc Morgan (2018a) The Weight of the Rich: Improving Surveys Using Tax Data, WID.world working paper 2018/12.

Blanchet, Thomas, Bertrand Garbinti, Jonathan Goupille-Lebret, and Clara Martínez-Toledano (2018b) Applying generalized Pareto curves to inequality analysis, AEA papers and proceedings 108:114–118.

Bordley, Robert F., James B. McDonald, and Anand Mantrala (1996) Something new, something old: parametric models for the size distribution of income, Journal of Income Distribution 6:91–103.

Brachmann, Klaus, Andreas Stich, and Mark Trede (1996) Evaluating parametric income distribution models, Allgemeines Statistisches Archiv 80:285–298.

Brzezinski, Michal (2013) Parametric modelling of income distribution in Central and Eastern Europe, University of Warsaw Working Paper 31(116).

Burkhauser, Richard V., Shuaizhang Feng, and Jeff Larrimore (2009) Improving imputations of top incomes in the public-use current population survey by using both cell-means and variances, Economics Letters 108:69–72.

Burr, Irving W. (1942) Cumulative frequency functions, Annals of Mathematical Statistics 13(2):215–232.

Butler, Richard J., and James B. McDonald (1989) Using incomplete moments to measure inequality, Journal of Econometrics 42:109–119.

Campano, Fred (1987) A fresh look at Champernowne's five-parameter formula, Economie Appliquee 40:162–175.

Campano, Fred, and Dominick Salvatore (2006) Income distribution, Oxford: Oxford University Press.

Chakraborti, Anirban, and Marco Patriarca (2008) Gamma-distribution and wealth inequality, Pramana Journal of Physics 71(2):233–243.

Champernowne, David G. (1937) The theory of income distribution, Econometrica 5:379–381.

Champernowne, David G. (1952) The graduation of income distributions, Econometrica 20(4):591–615.

Champernowne, David G. (1953) A model of income distribution, Economic Journal 63(250):318–351.

Champernowne, David G. (1973) The distribution of income, Cambridge: Cambridge University Press.

Charpentier, Arthur, and Emmanuel Flachaire (2019) Pareto models for top incomes, Aix-Marseille Universite working paper, May.

Chotikapanich, Duangkamon (1994) Techniques for measuring income inequality: an application to Thailand, Brookfield, VT: Avebury, Ashgate Publishing.

Chotikapanich, Duangkamon, ed. (2008) Modeling Income Distributions and Lorenz Curves, New York: Springer Verlag.

Chotikapanich, Duangkamon, William E. Griffiths, Gholamreza Hajargasht, Wasana Karunarathne, and D.S. Prasada Rao (2018) Using the GB2 income distribution, Econometrics 6(2):21.

Chotikapanich, Duangkamon, D.S. Prasada Rao, and Kam K. Tang (2007) Estimating income inequality in China using grouped data and the generalized beta distribution, Review of Income and Wealth 53:127–147.

Chumacero, Rómulo A., and Ricardo Paredes M. (2005) Characterizing income distribution: policy implications for poverty and inequality, Estudios de Economía 32(1):97–117.

Chung, Kee H., and Raymond A.K. Cox (1994) A Stochastic Model of Superstardom: An Application of the Yule Distribution, Review of Economics and Statistics 76(4):771–775.

Clementi, Fabio, and Mauro Gallegati (2005) Pareto's Law of Income Distribution: Evidence for Germany, the United Kingdom and the United States, In: Chatterjee A., Yarlagadda S., Chakrabarti B.K. (eds) Econophysics of Wealth Distributions. New Economic Windows. Milano: Springer.

Clementi, Fabio, T. DiMatteo, Mauro Gallegati and G. Kaniadakis (2008) The k-Generalised Distribution: a New Descriptive Model for the Size Distribution of Incomes, Physica A 387:3201–3208.

Coles, Stuart (2001) An introduction to statistical modeling of extreme values. New York: Springer.

Colombi, Roberto (1990) A new model of income distribution: the Pareto-lognormal distribution, ch.2 in C. Dagum and M. Zenga (eds.) Studies in Contemporary Economics: Income and Wealth Distribution, Inequality and Poverty, Berlin: Springer.

Cox, Christopher, Haitao Chu, Michael F.E. Schneider, and Álvaro Muñoz (2007) Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution, Statistics in medicine 26(23):4352–4374.

Cowell, Frank A. (2011) Measuring Inequality, (3rd ed.). Oxford: Oxford University Press.

Cowell, Frank A., and Emmanuel Flachaire (2002) Sensitivity of inequality measures to extreme values, LSE STICERD Research Paper, DARP 60.

Cowell, Frank A., and Emmanuel Flachaire (2007) Income distribution and inequality measurement: The problem of extreme values, Journal of Econometrics 141(2):1044–1072.

Cowell, Frank A., and Emmanuel Flachaire (2015) Statistical Methods for Distributional Analysis, AMSE Working Papers 1507.

Cowell, Frank A., and Philippe Van Kerm (2015) Wealth inequality: a survey, Journal of Economic Surveys 29(4):671–710.

Cramer, Jan S. (1978) A function for size distribution of incomes: comment, note, Econometrica 46:459–460.

Credit Suisse (2019) Global wealth report 2019: Report and databook, October 2019.

Cronin, David C. (1979) A Function for Size Distribution of Incomes: A Further Comment, Econometrica 47(3):773–774.

Cummins, J. David, Georges Dionne, James B. McDonald, B. Michael Pritchett (1990) Applications of the GB2 family of distributions in modeling insurance loss processes, Insurance Mathematical Economics 9:257–272.

Dagsvik, John K., Weizhen Zhu, Bjørn H. Vatne, Zhiyang Jia (2013) Is the Pareto-Lévy Law a Good Representation of Income Distributions? Empirical Economics 44(2):719–737.

Dagum, Camilo (1975) A model of income distribution and the conditions of existence of moments of finite order, Proceedings of the 40th session of the International Statistical Institute, pp.196–202, Warsaw.

Dagum, Camilo (1977) A new model of personal income distribution: specification and Estimation, Economie Appliquée 30:413–437.

Dagum, Camilo (1980) The generation and distribution of income, the Lorenz curve and the Gini ratio, Economie Appliquée 33:327–367.

Dagum, Camilo (1983) Income Distribution Models, In: S. Kotz, N.L. Johnson, and C. Read (eds.): Encyclopedia of Statistical Sciences, Vol. 4., New York: John Wiley, p.27–34.

Dagum, Camilo (1990) Generation and properties of income distribution functions, ch.1 in C. Dagum and M. Zenga (eds.) Studies in Contemporary Economics: Income and Wealth Distribution, Inequality and Poverty, Berlin: Springer.

Dagum, Camilo (1999) A study on the Distributions of Income, Wealth and Human Capital, Revue europeenne des sciences sociales 37(133):231–268.

Davidson, Russell, and Emmanuel Flachaire (2007) Asymptotic and bootstrap inference for inequality and poverty measures, Journal of Econometrics 141(1):141–166.

Davis, Harold T. (1941a) The Analysis of Economic Time Series. Bloomington, IN: Principia Press.

Davis, Harold T. (1941b) The theory of econometrics. Bloomington, IN: Principia Press.

Duan, Yutong (2016) Implementing the Multimodel Generalized Beta Estimator in Stata and Its Application, Masters thesis, Department of Statistics, University of Texas at Austin.

Dubey, Satya D. (1970) Compound gamma, beta and F distributions, Metrika 16:27–31.

Eckerstorfer, Paul, Johannes Halak, Jakob Kapeller, Bernhard Schutz, Florian Springholz, and Rafael Wildauer (2016) Correcting for the missing rich: an application to wealth survey data, Review of Income and Wealth 62(4):605–627.

Esteban, Joan (1986) Income share elasticity and the size distribution of income, International Economic Review 27:439–444.

Feng, Shuaizhang, Richard V. Burkhauser, and John S. Butler (2006) Levels and long-term trends in earnings inequality: overcoming Current Population Survey censoring problems using the GB2 distribution, Journal of Business & Economic Statistics 24:51–62.

Ferrari, Silvia L.P., and Francisco Cribari-Neto (2004) Beta regression for modelling rates and proportions, Journal of Applied Statistics 31(7):799–815.

Fisher, Ronald A. (1930) The genetical theory of natural selection, New York: Oxford University Press.

Fisk, Peter R. (1961) The Graduation of Income Distributions, Econometrica 29:171–185.

Foster, James, Joel Greer, and Erik Thorbecke (1984) A Class of Decomposable Poverty Measures, *Econometrica* 52:761–766.

Fournier, Juliette (2015) Generalized Pareto curves: Theory and application using income and inheritance tabulations for France 1901–2012, Paris School of Economics Masters thesis.

Fréchet, Maurice (1927) Sur la loi de probabilité de l'écart maximum, Annales de la Société Polonaise de Mathematique, Cracovie 6:93–116.

Gabaix, Xavier (2009) Power Laws in Economics and Finance, Annual Review of Economics 1:255–294.

Gastwirth, Joseph L. (1972) The Estimation of the Lorenz Curve and Gini Index, Review of Economics and Statistics 54:306–316.

Gastwirth, Joseph L., and Marcia Glauberman (1976) The interpolation of the Lorenz curve and Gini index from grouped data, Econometrica 44(3):479–483.

Gibrat, Robert (1931) Les Inégalités économiques: applications aux inégalités des richesses, à la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc., d'une loi nouvelle: la loi de l'effet proportionnel, Paris: Librairie du Recueil Sirey.

Giorgi, Giovanni M. (1999) Income inequality measurement: the statistical approach, ch.8 in J. Silber (ed.) Handbook of Income Inequality Measurement, Norwell MA: Kluwer.

Gordy, Michael B. (1998) A generalization of generalized beta distributions, Federal Reserve Board, FEDS paper 98-18.

Graf, Eric, and Yves Tille (2013) Imputation of income data with generalized calibration procedure and GB2 law: illustration with SILC data, pp.5509–5514, Proceedings of the 59th session of the International Statistical Institute, Hong Kong.

Graf, Monique, and Desislava Nedyalkova (2014) Modeling of income and indicators of poverty and social exclusion using the generalized B distribution of the second kind, Review of Income and Wealth 60(4):821–842.

Hamed, Duha, Felix Famoye, and Carl Lee (2018) On families of generalized Pareto distributions: properties and applications, Journal of Data Science 4(8):377-396.

Harrison, Alan J. (1979) The Upper Tail of the Earnings Distribution: Pareto or Lognormal?, Economics Letters 2:191–195.

Harrison, Alan J. (1981) Earnings by size: a tale of two distributions, Review of Economic Studies 48(4):621–631.

Hart, Peter E. (1983) The size mobility of earnings, Oxford Bulletin of Economics and Statistics 45(2):181–193.

Hill, Bruce M. (1975) A Simple General Approach to Inference About the Tail of a Distribution, Annals of Statistics 3(5):1163–1174.

Hippel, Paul T. von, Igor Holas, and Samuel V. Scarpino (2012) Estimation with binned data, Cornell University working paper.

Hlasny, Vladimir (2019) Redistributive impacts of fiscal policies in Mexico: corrections for top income measurement problems, CEQ Working Paper 84.

Hlasny, Vladimir, and Vito Intini (2015) Representativeness of top expenditures in Arab region household surveys. UN ESCWA/EDID working paper 11, December, Beirut: UN ESCWA.

Hlasny, Vladimir, and Paolo Verme (2015) Top incomes and the measurement of inequality: a comparative analysis of correction methods using Egyptian, EU and US survey data, Paper presented at the 6th meeting of the Society for the Study of Economic Inequality (ECINEQ), Luxembourg, July 13-15, 2015. www.ecineq.org/ecineq_lux15/FILESx2015/CR2/p145.pdf

Hlasny, Vladimir, and Paolo Verme (2018a) The Impact of Top Incomes Biases on the Measurement of Inequality in the United States, Ecineq working paper 452.

Hlasny, Vladimir, and Paolo Verme (2018b) Top incomes and inequality measurement: a comparative analysis of correction methods using the EU SILC data, Econometrics 6(2):30.

Hlasny, Vladimir, and Paolo Verme (2018c) Top incomes and the measurement of inequality in Egypt, World Bank Economic Review 32(2):428–455.

Hlasny, Vladimir, Lidia Ceriani, and Paolo Verme (2020) Bottom incomes and the measurement of poverty and inequality, Economic Research Forum and Luxembourg Income Study Working Paper.

Horsky, Dan (1990) A diffusion model incorporating product benefits, price, income and information, Marketing Science 9(4):342–365.

Houthakker, Hendrik S. (1992) Conic distributions of earnings and income, Harvard Institute of Economic Research Working Papers 1598.

Hruschka, Daniel J., Drew Gerkey, and Craig Hadley (2015) Estimating the absolute wealth of households, Bulletin of the World Health Organization 93:483-490.

Ibragimov, Marat, and Rustam Ibragimov (2018) Heavy tails and upper-tail inequality: the case of Russia, Empirical Economics 54:823–837.

Jäntti, Markus, Eva M. Sierminska, and Philippe Van Kerm (2015) Modeling the joint distribution of income and wealth, Research on Economic Inequality 23:301–327.

Jenkins, Stephen P. (2007) Inequality and the GB2 income distribution. IZA Discussion Papers 2831, Institute for the Study of Labor (IZA), Bonn, Germany.

Jenkins, Stephen P. (2009) Distributionally-sensitive inequality indices and the GB2 income distribution, Review of Income and Wealth 55:392–398.

Jenkins, Stephen P. (2017) Pareto distributions, top incomes, and recent trends in UK income inequality, Economica 84(334):261–289.

Jenkins, Stephen P., Richard V. Burkhauser, Shuaizhang Feng, and Jeff Larrimore (2011) Measuring inequality using censored data: a multiple-imputation approach to estimation and inference, Journal of the Royal Statistical Society 174(1):63–81.

Jenkins, Stephen P., and Markus Jäntti (2005) Methods for summarizing and comparing wealth distributions, ISER Working Paper 2005-05.

Jenkins, Stephen P., and Philippe Van Kerm (2007) Paretofit: Stata module fitting a Pareto (Type I) distribution by ML to unit record data, April 2007.

Johnson, Norman L., Samuel Kotz, and Narayanaswamy Balakrishnan (1994) Distributions in statistics: continuous univariate distributions, Vols. 1 and 2 (2nd Ed.), New York: Wiley.

Johnson, Norman L. (1949) Systems of frequency curves generated by methods of translation, Biometrika 36(1/2):149–176.

Jones, Charles I. (2015) Pareto and Piketty: The macroeconomics of top income and wealth inequality, Journal of Economic Perspectives 29(1):29–46.

Kalecki, Michał (1945) On the Gibrat distribution, Econometrica 13(2):161–170.

Kennickell, Arthur B. (2019) The tail that wags: differences in effective right tail coverage and estimates of wealth inequality, Journal of Economic Inequality 17:443-459.

Kleiber, Christian (1996) Dagum vs. Singh-Maddala income distributions, Economics Letters 53:265–268.

Kleiber, Christian (2008) A guide to the Dagum distributions, Ch.6 in Duangkamon Chotikapanich (ed.) Modeling Income Distributions and Lorenz Curves, pp.97–118, New York: Springer.

Kleiber, Christian (2013) On moment indeterminacy of the Benini income distribution, Statistical Papers 54:1121–1130.

Kleiber, Christian, and Samuel Kotz (2003) Statistical Size Distributions in Economics and Actuarial Sciences, Hoboken, NJ: John Wiley & Sons.

Kloek, Teun, and Herman K. Van Dijk (1978) Efficient estimation of income distribution parameters, Journal of Econometrics 8:61–74.

Kmietowicz, Zbigniew W., and H. Ding (1993) Statistical analysis of income distribution in the Jiangsu Province of China, Journal of the Royal Statistical Society 42(2):107–121.

Kotz, Samuel, and Saralees Nadarajah (2000) Extreme Value Distributions: Theory and Applications, London: World Scientific.

Kotz, Samuel, and Johan R. van Dorp (2006) Beyond Beta: Other Continuous Families of Distributions with Bounded Support and Applications, Singapore: World Scientific.

Lee, Lung-Fei (1984) Maximum likelihood estimation and a specification test for non-normal distributional assumption for the accelerated failure time models, Journal of Econometrics 24(1):159–179.

Leemis, Lawrence M., and Jacquelyn T. McQueston (2008) Univariate distribution relationships, Teachers' corner, American Statistician 62(1):45–53.

Lomax, Kenneth S. (1954) Business failures: another example of the analysis of failure data, Journal of the American Statistical Association 49:847–852.

Lorenz, Max O. (1905) Methods of measuring the concentration of wealth, Publications of the American Statistical Association 9(70):209–219.

Lubrano, Michel (2017) The econometrics of inequality and poverty, unpublished manuscript, AIX Marseille School of Economics.

Lustig, Nora (2019) The 'missing rich' in household surveys: causes and correction approaches, CEQ Working Paper 75.

Maccabelli, Terenzio (2009) Measuring inequality: Pareto's ambiguous contribution. History of Political Economy 41:183–208.

Majumder, Amita, and Satya R. Chakravarty (1990) Distribution of personal income: development of a new model and its application to US income data, Journal of Applied Econometrics 5(2):189–196.

Mandelbrot, Benoit (1960) The Pareto-Lévy law and the distribution of income, International Economic Review 1:79–106.

Mandelbrot, Benoit (1961) Stable Paretian random functions and the multiplicative variation of income, Econometrica 29:517–543.

Marshall, A.W. and I. Olkin (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. Biometrika, 84, 641–652.

McDaniel, Reuben R., Robert S. Sullivan, and James R. Wilson (1988) A simulation model for welfare policy analysis, Socio-Economic Planning Sciences 22(4):157–165.

McDonald, James B. (1984) Some generalized functions for the size distribution of income, Econometrica 52:647–663.

McDonald, James B., and Richard J. Butler (1990) Regression models for positive random variables, Journal of Econometrics 43:227–251.

McDonald, James B., and Anand Mantrala (1993) Apples, oranges and distribution trees, Discussion paper, Brigham Young University.

McDonald, James B., and Anand Mantrala (1995) The distribution of personal income: revisited, Journal of Applied Econometrics 10(2):201–204.

McDonald, James B., and Michael R. Ransom (1979a) Alternative parameter estimators based upon grouped data, Communications in Statistics 8(9):899–917.

McDonald, James B., and Michael R. Ransom (1979b) Functional forms, estimation techniques and the distribution of income, Econometrica 47(6):1513–1525.

McDonald, James B., and Dale O. Richards (1987) Hazard rates and generalized beta distributions, IEEE Transactions on Reliability 36:463–466.

McDonald, James B., Patrick A. Turley, Jeff Sorensen (2013) Skewness and kurtosis properties of income distribution models, Review of Income and Wealth 59(2):360–374.

McDonald, James B. and Yexiao J. Xu (1995) A generalization of the beta distribution with applications, Journal of Econometrics 66:133–152.

McLaughlin, Michael P. (1999) A Compendium of Common Probability Distributions, Version 2.3, Appendix A of Regress+ documetation, www.ub.edu/stat/docencia/Diplomatura/Compendium.pdf (accessed 24 September 2018).

Medeiros, Marcelo, Juliana de Castro Galvao, and Luisa de Azevedo Nazareno (2018) Correcting the underestimation of top incomes: combining data from income tax reports and the Brazilian 2010 census, Social Indicators Research 135(1):233–244.

Merkies, Arnold H.Q.M. (1987) Inductive analysis from empirical income distributions, In: Heijmans R., Neudecker H. (eds.) The Practice of Econometrics: Preference Functions of Dutch political parties, pp. 151–164, Dordrecht, Netherlands: Springer.

Merkies, Arnold H.Q.M. (1994) The scope of inequality coefficients, In: Eichhorn W. (eds) Models and Measurement of Welfare and Inequality, pp. 141–161, Berlin: Springer.

Metcalf, Charles E. (1969) The size distribution of personal income during the business cycle, American Economic Review 59:657–669.

Metcalf, Charles E. (1972) An econometric model of the income distribution, Institute for Research on Poverty monograph, Chicago: Markham Publishers.

Montroll, Elliott W. (1978) On some mathematical models of social phenomena, pp.161–216 in Proceedings of an International Symposium on Nonlinear Equations in Abstract Spaces, University of Texas at Arlington, 8–10 June 1977.

Nirei, Makoto, and Wataru Souma (2007) A two-factor model of income distribution dynamics, Review of Income and Wealth 53(3):440–459.

Okamoto, Masato (2013) Extension of the κ-generalized distribution: new four-parameter models for the size distribution of income and consumption, Luxembourg Income Study working paper 600, December.

Ord, Keith (1975) Statistical models for personal income distributions. In Statistical distributions in scientific work (ed. G. P. Patil, S. Kotz and J. K. Ord), Vol. 2, pp. 151-158. D. Reidel, Publ. Co., Dordrecht.

Papastathopoulos, Ioannis, and Jonathan A. Tawn (2013) Extended generalised Pareto models for tail estimation, Journal of Statistical Planning and Inference 143:131–142.

Pareto, Vilfredo F.D. (1895) La legge della domanda, Giornale degli Economisti 10:59–68.

Pareto, Vilfredo F.D. (1896a) La courbe de la repartition de la richesse, Ecrits sur la courbe de la repartition de la richesse, (writings by Pareto collected by G. Busino, Librairie Droz, 1965), pp.1–15.

Pareto, Vilfredo F.D. (1896b), On the distribution of wealth and income, translation in Frank A. Cowell (ed.), The economics of poverty and inequality, Vol. 2, Elgar Reference Collection, Cheltenham, 2003.

Persky J. (1992) Retrospectives: Pareto's law, Journal of Economic Perspectives 6:181–192.

Pham-Gia, Thu, and Quang P. Duong (1989) The generalized beta and F distributions, statistical modeling, Mathematical Computational Modeling 12(12):1613–1625.

Pickands, James (1975) Statistical inference using extreme order statistics, Annals of Statistics 3:119–131.

Piketty, Thomas (2001) Les Hauts revenus en France au 20e siècle: inégalités et redistribution, 1901–1998. Paris: Grasset.

Piketty, Thomas, and Emmanuel Saez (2003) Income inequality in the United States 1913–1998, Quarterly Journal of Economics 118(1):1–39.

Ragab, Aisha, Jack R. Green, and Maurice C.K. Tweedie (1991) The Burr type II distribution: properties, order statistics, Microelectronics Reliability 31(6):1181-1191.

Rasche, Robert H., Janice Gaffney, Anthony Y.C. Koo and Norman Obst (1980) Functional forms for estimating the Lorenz curve, Econometrica 48:1061–1062.

Reed, William J. and Wu Fan (2008) New four- and five-parameter models for income distributions. In D. Chotikapanich (ed.), Modeling Income Distributions and Lorenz Curves, Chapter 13, New York: Springer-Verlag, pp. 211–224.

Roodman, David M. (2015) EXTREME: Stata module to fit models used in univariate extreme value theory. Statistical Software Components archive S457953, Boston College Department of Economics, revised 18 Dec 2017. https://ideas.repec.org/c/boc/bocode/s457953.html.

Salem, Ali B.Z., and Timothy D. Mount (1974) A Convenient Descriptive Model of Income Distribution: The Gamma Density, Econometrica 42:1115–1127.

Sarabia, José M., and Enrique Castillo (2005) About a class of max-stable families with applications to income distributions, Metron-International Journal of Statistics 63(3):505–527.

Sarabia, José M., Enrique Castillo, and Daniel J. Slottje (2002) Lorenz ordering between McDonald's generalized functions of the income size distribution, Economics Letters 75:265–270.

Schluter, Christian, and Mark Trede (2002) Tails of Lorenz curves, Journal of Econometrics 109:151–166.

Simon, Herbert A. (1955) On a Class of Skew Distribution Function, Biometrika 42(3/4):425–440.

Simon, Herbert A., and Charles P. Bonini (1958) The Size Distribution of Business Firms, American Economic Review 48:607–617.

Singh, Vijendra P., and Huizhen Guo (1995) Parameter estimation for 3-parameter generalized Pareto distribution by the principle of maximum entropy (POME), Hydrological Sciences Journal 40(2):165–181.

Singh, Surendra K., and Gary S. Maddala (1976) A function for the size distribution of income, Econometrica 44:963–970.

Song, Wheyming T. (2005) Relationships among some univariate distributions, IIE Transactions 37:651–656.

Stacy, E. Webb (1962) A generalization of the gamma distribution, Annals of Mathematical Statistics 33:1187–1192.

Stoppa, Gabriele (1990a) A new model for income size distributions, ch.3 in C. Dagum and M. Zenga (eds.) Studies in Contemporary Economics: Income and Wealth Distribution, Inequality and Poverty, Berlin: Springer.

Stoppa, Gabrielle (1990b) Proprieta campionarie di un nuovo modello Pareto generalizzato. Atti XXXV Riunione Scientifica della Societa Italiana di Statistica, Padova: Cedam, pp.137–144.

Suruga, Terukazu (1982) Functional forms of income distribution: the case of yearly income groups in the 'Annual report on the family income and expenditure survey', Economic Studies Quarterly 361–395.

Tachibanaki, Toshiaki, Terukazu Suruga, and Naosumi Atoda (1997) Estimations of Income Distribution Parameters for Individual Observations by Maximum Likelihood Method, Journal of the Japan Statistical Society 27(2):191–203.

Tadikamalla, Pandu R. (1980) A Look at the Burr and Related Distributions, International Statistical Review 48(3):337–344.

Taylor Lester D. (2010) Conic Distributions of Earned Incomes, In: L.D. Taylor and H.S. Houthakker, Consumer Demand in the United States: Prices, Incomes and Consumption Behavior, New York: Springer.

Thurow, Lester C. (1970) Analyzing the American Income Distribution, American Economic Review 60(2):261–269.

Toda, Alexis A. (2012) The double power law in income distribution: explanations and evidence, Journal of Economic Behavior and Organization 84:364–381.

Topp, Chester W., and Fred C. Leone (1955) A family of J-shaped frequency functions, Journal of the American Statistical Association 50(269):209–219.

Van Dorp, Johan R., and Samuel Kotz (2006) Modeling Income Distributions Using Elevated Distributions, Ch. 1 in Herreiras Pleguezuelo, Rafael, Jose Callejon Cespedes, and Jose M. Herrerias Velasco (eds.) Distribution Models Theory, p.1–25, Singapore: World Scientific Publishing.

Van Hauwermeiren, Michael, David Vose, and Stijn Vanden Bossche (2012) A Compendium of Distributions (2nd ed.). [e-book], Ghent: Vose Software.

Van Kerm, Philippe (2007) Extreme incomes and the estimation of poverty and inequality indicators from EU-SILC, IRISS working paper 2007-01.

Vermeulen, Philip (2016) Estimating the top tail of the wealth distribution, American Economic Review 106(5):646–650.

Victoria-Feser, Maria-Pia (1995) Robust Methods for Personal Income Distribution Models, with Application to Dagum's Model, pp.225–239 in C. Dagum and A. Lemmi (Eds.) Research On Economic Inequality: Income Distribution, Social Welfare, Inequality and Poverty, Greenwich:JAI Press.

Victoria-Feser, Maria-Pia (2000) Robust methods for the analysis of income distribution, inequality and poverty, International Statistical Review 68:277–293.

Warsono, Dian Kurniasari, Widiarti, Mustofa Usman and Faiz A.M. Elfaki (2017) The moment generating function of the four-parameter generalized F distribution and related generalized distributions, Global Journal of Pure and Applied Mathematics 13(3):941-956.

Weibull, Waloddi (1951) A statistical distribution function of wide applicability Journal of Applied Mechanics 18(3):293–297.

Wifling, Berndt, and Walter Kraemer (1993) The Lorenz-ordering of Singh-Maddala income distributions, Economics Letters 43:53–57.

Winkler, W. (1950) The corrected Pareto law and its economic meaning, Bulletin of International Statistical Institute 32:441–449.

Xekalaki, Evdokia (1983) A Property of the Yule Distribution and its Applications, Communication in Statistics-Theory and Methods 12(10):1181–1189.

Yule, George U. (1924) A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis F.R.S., Philosophical Transactions B, Royal Society of London 213:21–87.

Zolotarev, Vladimir M. (1986) One-dimensional stable distributions, (translated from Russian by H.H. McFaden), Providence, RI: American Mathematical Society.

Appendix 1. Statistical programs (Stata)

The stock of statistical programs is continuously expanding – and so this section (non-exhaustive as it is) will eventually become obsolete. The following programs are presently available in *Stata* software version 13 (most of them installable through `ssc install ZZZ, update`), but even these programs are being updated to allow additional options, such as left or right truncation, weighting, testing, graphical analysis and post-estimation. Unless noted, the following commands are designed for unit-record data.

Estimators of selected parametric distributions, and some basic diagnostics

`paretofit` (Jenkins and Van Kerm 2007): fits by ML the 1-parameter Pareto 1 distribution. (`paretowml` (Brzezinski 2014a): weighted maximum likelihood estimator (WMLE) for the shape parameter of Pareto 1 distribution)

`extreme gpd` (Roodman 2015): fits by ML the 2-parameter generalized Pareto distribution (Pareto 2).

`extreme gev`: fits by ML the 3-parameter generalized extreme value distribution.

`lognfit` (Jenkins 2007b): fits by ML the 2-parameter lognormal distribution.

`betafit` (Buis et al. 2003): fits by ML the 2-parameter beta distribution.
    (`qbeta` (Cox 1998d): quantile-quantile plot for data vs fitted beta distribution.)
    (`pbeta` (Cox 1998a): probability plot for data vs fitted beta distribution.)

`gammafit` (Cox and Jenkins 2003a): fits by ML the 2-parameter gamma distribution.
    `streg, distribution(gamma)`: fits the generalized gamma
    (`qgamma` (Cox 1998f): quantile-quantile plot for data vs fitted gamma distribution.)
    (`pgamma` (Cox 1998c): probability plot for data vs fitted gamma distribution.)

`gumbelfit` (Cox and Jenkins 2003b): fits by ML the 2-parameter Gumbel distribution.
    (or `extreme gev, gumbel`: restrict location parameter to 0, estimate 2-parameter Gumbel distribution)
    (`qgumbel`: quantile-quantile plot for data vs fitted Gumbel distribution.)
    (`pgumbel`: probability plot for data vs fitted Gumbel distribution.)

`fiskfit` (Buis and Jenkins 2008): fits by ML the 2-parameter Fisk (log-logistic; sech-squared) distribution.

`invgammafit` (Cox and Jenkins 2007a): fits by ML the 2-parameter inverse-gamma distribution.
    (`qinvgamma`: quantile-quantile plot for data vs fitted inverse-gamma distribution.)
    (`pinvgamma`: probability plot for data vs fitted inverse-gamma distribution.)

`streg, distribution(exponential)`: fits the exponential distribution
    (or `extreme gpd, gumbel`: restrict location parameter to 0, estimate 1-parameter exponential distribution)
    (`pexp` (Cox 1998b): probability plot for data vs fitted exponential distribution.)
    (`qexp` (Cox 1998e): quantile-quantile plot for data vs fitted exponential distribution.)

`streg, distribution(gompertz)`: fits the Gompertz

`mgbe` (Duan and von Hippel 2016): fits 10 distributions from the GB2 family on grouped-records data (Hippel et al. 2016), including: GB2, Dagum, Singh-Maddala, beta 2, log-logistic, generalized gamma, gamma, Weibull, Pareto 2, lognormal

`gb2fit` (Jenkins 2007a): fits by ML the 4-parameter GB2 distribution
`gb2lfit` (Jenkins 2014): fits by ML the 4-parameter GB2 distribution. equivalent to `gb2fit`, but estimating parameters in the logarithmic metric, and allowing additional options including right-censoring or left-truncation of data, and estimation of inequality measures.

(gbgfit (Nichols 2010): GB2 distribution fitted to grouped-records data.)
(gb2dist (Brzezinski 2014c): computes poverty and inequality indices for the GB2 distribution, following gb2fit/gbgfit estimation.)

dagumfit (Jenkins 1999a): fits by ML the 3-parameter Dagum distribution.
 (dagumpred: derivation of additional statistics for the Dagum distribution.)
 (qdagum: quantile-quantile plot for data vs fitted Dagum distribution.)
 (pdagum: probability plot for data vs fitted Dagum distribution.)

smfit: (Jenkins 1999b): fits by ML the 3-parameter Singh-Maddala distribution.
 (qsm: quantile-quantile plot for data vs fitted Singh-Maddala distribution.)
 (psm: probability plot for data vs fitted Singh-Maddala distribution.)

weibullfit (Cox and Jenkins 2007b): fits by ML the 2-parameter Weibull distribution. Restricting $\gamma=1$ gives the exponential distribution; $\gamma=2$ gives the Rayleigh distribution.
 (qweibull: quantile-quantile plot for data vs fitted Weibull distribution.)
 (pweibull: probability plot for data vs fitted Weibull distribution.)

dirifit (Buis et al. 2006): fits by ML the 2 parameter Dirichlet distribution.

Graphical diagnostic tests

qqplot3 (Linden 2014): quantile-quantile plot comparing an empirical distribution, particularly its tails, against quantiles of a reference distribution, allowing weights and by-variables.
 (qqplot: basic command, without weights and by-variables.
 qnorm: quantile-quantile plot comparing a distribution, particularly its tails, to a reference normal distribution (aka, *probit plotting*))

ppplot (Cox 2004): probability-probability plot comparing an empirical cumulative distribution, particularly its center, to a reference distribution.
 (pnorm: probability plot comparing a distribution, particularly its center, to a reference normal distribution)

extreme plot, pp qq return density: PP & QQ plots following extreme estimation

hillp (Scotto 2001): creates the Hill plot, with estimates of the Pareto 1 coefficient on the vertical axis, and the share of top income observations on which the coefficient was estimated on the horizontal axis (with confidence interval).

hangroot (Buis 2007): hanging rootogram comparing an empirical distribution to the best-fitting reference parametric distribution.

dpplot (Cox 2002): plots density probability plots for an empirical distribution given a reference distribution.

`ggtax` (Gonzalez Rangel 2013): creates a graph of the estimated generalized-gamma shape and scale parameters, against other generalized-gamma family models (including lognormal, exponential and Weibull).

`glcurve` (Jenkins and Van Kerm 2004): draw the Generalised Lorenz curve, and generate GL ordinates

Other diagnostic tests

`powerlaw` (Brzezinski 2014b): detecting power law or Pareto-like behavior in empirical data using model fitting, goodness of fit and model selection tests (Clauset et al. 2009).
available at: http://coin.wne.uw.edu.pl/mbrzezinski/software/

`rpec` (Brzezinski 2014c): robust Pareto error criterion for choosing the scale parameter (x_0) of Pareto 1 distribution.

`nrpec` (Brzezinski 2014c): non-robust Pareto error criterion for choosing the scale parameter (x_0) of Pareto 1 distribution.

`qpfit` (Cox 2007): quantile and probability plot commands for assessing the fit of specific distributions, namely the Dagum, generalised beta (second kind), lognormal and Singh-Maddala

`ksmirnov`: one- or two-sample Kolmogorov-Smirnov test of the equality of distributions.

`lrtest GEV`: likelihood ratio test of Gumbel vs. unrestricted extreme-value model following `extreme` estimation

`compsta` (Brzezinski 2014c): test of the computational stability of estimation using perturbation tests

Appendix 2. An illustration of alternative estimators for distribution functions

One major reason for considering the parametric modeling of the top tail of income distributions is that the empirically observed distributions suffer from various measurement problems or incomparability across samples or survey waves. The highest incomes are to a large extent excluded from household surveys, because the households are not identified on the sampling frame – and therefore are not selectable – or have a probability of being selected, but for one reason or another end up excluded. This could be simply due to a low probability of selection, accounting for their rareness and the inadequate oversampling of the relevant demographic group. Secondly, unit and item nonresponse tend to be high in this group, as is underreporting and post-survey topcoding by statistical agencies.

The parametric models are applied to imperfect, empirical income distributions, and the specific functional forms and parameter values must be estimated and assessed. To this end, a number of estimators have been proposed to fit the distribution functions presented in the main text. The aim of this short and non-exhaustive review is to make readers aware of the variety of estimation approaches, and provide some key references.

Hill (1975) proposed a maximum-likelihood estimator of the Pareto 1 coefficient (see also Quandt 1966; Schluter and Trede 2002:155). For fixed k, this estimator $H_{k,n}$ converges in distribution to a gamma distribution as $n \to \infty$. Huisman et al. (2001) proposed a weighted version

of Hill's estimator correcting it for its small sample bias. A robust estimator of the Pareto coefficient was proposed by Cowell and Victoria-Feser (1996ab, 2007). ML 'Optimal b-robust estimator' (ML-OBRE) places an upper limit on the score function to make the estimation robust to the presence of high values (Victoria-Feser and Ronchetti 1994; Doğru and Arslan 2015). Brzezinski (2013, 2016) and Jenkins (2017) review alternative estimators for the Pareto coefficient.

Mahdi and Cenac (2006) compared three methods for estimating the logistic and Rayleigh distributions – the maximum likelihood, the moment and the probability weighted moment methods. Hippel et al. (2012) proposed a 'best-of-breed' model selection criterion, which chooses the best fitting model among the extended generalized gamma, power normal and power logistic families.

McDonald and Ransom (1979ab, 2008) identified and evaluated three estimators for the estimation of GB and gamma models on grouped-records data. Pearson minimum chi-squared estimators (asymptotically equivalent to the method-of-scoring maximum-likelihood estimators, and generalizations of estimators minimizing the sum of squared differences between the predicted and observed probabilities. On decile-grouped data, the scoring and minimum chi-squared estimators were found to perform similarly, and outperformed the least-squares estimators in terms of sample bias, variance, and mean squared error.

Appendix 3. Diagnostic tests

There are three types of diagnostic tests available to distinguish parametric forms for the modeling of the top tail of income distributions. One, standard confidence-interval based tests can help assess goodness of fit of stand-alone models or estimation approaches. These tests can also assess the relative fit of multiple models nested in one another, as well as the relative fit of models estimated on different subsamples. Two, graphical tests can help to evaluate visually any deviations in model fit from known yardsticks. Three, simulations on well understood (synthetic or uncontaminated) data can help evaluate the relative performance of alternative models or estimators, and their robustness to various experimental designs of data contamination. The following sections present some of the commonly used tests and their results.

Confidence-interval based tests

To test whether empirically observed data are consistent with a particular smooth parametric function, we must assess the deviations of the empirical data from the predicted pattern of dispersion. The available model selection criteria include sum of squared errors, sum of absolute errors, log-likelihood value, likelihood ratios, the Pearson chi-squared, Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling statistics, and others. These are briefly introduced below. The likelihood ratio, equal to twice the difference between the unrestricted and restricted model likelihoods, is distributed as chi-squared with degrees of freedom equal to the parameter restrictions.

The Pearson chi-squared can be computed as $\chi^2 = \sum_{i=1...m} \frac{(O_i - E_i)^2}{E_i}$, where O and E are the observed and expected shares of observations in each quantile *i=1…m*, respectively. This statistic provides a consistent rejection criterion in large samples, but results in false non-rejections in small samples, and is sensitive to the delineations of quantiles *i* (Stephens 1986). Model selection criteria such as the Akaike information criterion ($AIC = 2k - 2\log(L)$) can also be used for pairs of nested models. The Kolmogorov-Smirnov statistic compares the observed cumulative density of all order

statistics with that estimated under a parametric form, the Cramér-von Mises statistic evaluates the squared deviations between these cumulative densities, and the Anderson-Darling statistic assigns greater weight to the tails.

Regression-based tests rely on coefficient standard errors (*t* and *F* tests) and measures of model fit (R-squared, AIC, BIC, Wald F, likelihood ratio, and Lagrange multiplier tests).[17] When the parametric distributions are fitted by maximum likelihood, the required test statistics are available. In other cases, such as when parameters are on a boundary of the allowed parameter space, testing can be done using standard errors obtained by bootstrapping the estimation procedures (Cowell and Flachaire 2015:50–51; McDonald and Xu 1995:144).

To discriminate among nested models, including between the Singh-Maddala, the Dagum and the GB2, the parameter restrictions can be tested simply by looking at the corresponding parameter standard errors. However, confidence-interval tests based on asymptotic behavior perform poorly in small samples or in distributions with heavy tails as is common with income distributions (Cowell and Flachaire 2015:52–56).

Clauset et al. (2009) proposed a statistical framework for testing whether the data are consistent with the Pareto behavior, relying on maximum-likelihood fitting methods and goodness-of-fit tests based on the Kolmogorov–Smirnov statistic and likelihood ratios.

Graphical tests

One problem with data-reduction approaches such as relying on logarithmic transformations of distributions, or replacing full distributions with distributional statistics is that they lose sight of the true dispersion of original observations, particularly those in the extreme tails, where data are sparse and the fit is likely weakest. These extreme observations may also influence the estimates obtained under any fitted curves. As far back as a century ago, Lorenz (1905:217) criticized tests of Pareto's power-law property and complained that "logarithmic curves are more or less treacherous," suppressing any deviations in the original variables. Cirillo (2013) raised similar criticisms in regard to various graphs relying on logarithmic transformations, because they downplay the effect of outliers. These criticisms also apply to using distributional statistics in place of the full Lorenz curve. This section reviews the existing methods for testing the shape of income distributions. Like with distributional statistics, the selection depends on one's needs and on the part of income distribution of special interest.

Dombos (1982) provided an early review of available graphical diagnostic tools. A good start is to review the *density histograms* or *kernel density plots* compared to reference parametric distributions. Chauvel (2016) proposed what he called *isographs* showing the diversity of local inequality along the income scale. He writes: "The isograph presents the slope of the 'Fisk Graph' (Fisk, 1961:176) that is indeed a *logit-log transformation of the Pen's parade* (Pen, 1971:49–59), a transformation of the cumulative distribution function graph." Chauvel also reviews the shape of the density functions of distributions (strobiloids), and the evolution in them over time.

Another set of graphical tools is based on the so-called Zipf law (Gabaix 1999), proposing that one should find a linear relationship between log(rank order) and log(frequency) among the relevant range of incomes. Under the Pareto distribution, we should find a linear relationship

---

[17] These tests are valid for comparing nested models, except for nested distributions with parameters on the boundary of the allowed parameter space, which violate the regularity conditions necessary to invoke the Wald and likelihood-ratio properties even asymptotically. McDonald and Xu (1995:139–140) discuss the appropriate estimation approaches for the alternative cases, including maximum likelihood, method of moments, and maximum product spacing.

between the logarithm of the proportion of individuals with income greater than a threshold and the logarithm of the threshold itself. This suggest a graphical test of the validity of this property.

The plot of the log of incomes on the x-axis and the log of the survival function (log(1-CDF)) on the y-axis, known as the *Pareto diagram* (Cowell 2011) or *Zipf plot* (Cirillo 2013) is among the most commonly used diagnostics. *Pareto QQ-plot*, {-log(1-F(xi)) ; log xi}, has also been commonly used, but this plot performs poorly at distinguishing related distributions such as the lognormal and the Pareto (Cirillo 2013). The *PP-plot*, also referred to as the *probability plot*, shows the predicted versus actual values of the cumulative densities under various lower thresholds. The tightness of the estimated curve around the 45° line indicates goodness of fit.

Relative density plots juxtapose a best-fitting reference distribution function against the observed distribution. If the empirical distribution exactly follows the reference distribution, the plot will be a uniform density.

Jenkins (2017) used various graphical (and confidence-interval based) tools – the Zipf plots, *mean excess plots*, *Zenga* (1984) curves, and *Hill plots* (Drees et al. 2000) – to test Pareto 1 against Pareto 2 on an empirical income distribution. Mean-excess plots show the mean income above a threshold for various thresholds. For Pareto, the graph is a positively-sloped straight line above a lower threshold. Ghosh and Resnick (2010) used the mean-excess plots to diagnose several candidate functions in the generalized Pareto family. Zenga curves show a transformation of the Lorenz curve above a certain threshold. For Pareto distributions the Zenga curve is positively-sloped (and the more heavy-tailed the distribution, the higher the curve), while for a lognormal distribution, the curve is flat. Hill plots show the (Pareto) top-income index estimated across different delineations of top incomes, with various lower thresholds.

Similarly, *generalized Pareto curves* (Fournier 2015) show how the inverted Pareto coefficients (the ratio between the average income above a threshold, and the threshold income) evolve with varying thresholds in terms of income percentiles. Pareto distributions, the Champernowne distribution, the Sech2 distribution and the Singh Maddala distribution have a generalized Pareto curve with b>1 as one approaches the top end. On the contrary, lognormal, the three-parameter lognormal distributions, the gamma and the generalized gamma distributions and the Weibull distribution have a generalized Pareto curve with b=1 as one approaches the top end.

The *Cullen and Frey* (1999) *graph*, showing the degree of kurtosis on one axis, and the square of skewness on the other, indicates the departure of observations from a confidence interval.[18]

Cirillo (2013) and Jenkins (2017) illustrate a number of graphical (and confidence-interval based) tests of the Pareto 1 versus the Pareto 2 models. Araar and Duclos (2013) review various graphical methods including *differences between Lorenz curves*, and *differences between the so-called Foster-Greer-Thorbecke* (FGT) *cumulative-density curves* plotted against income level expressed relative to the poverty line (Foster et al. 1984).

Appendix references

Araar, Abdelkrim, and Jean-Yves Duclos (2013) User Manual for Stata Package DASP: Version 2.3, PEP, World Bank, UNDP and Université Laval.

Brzezinski, Michal (2013) Parametric modelling of income distribution in Central and Eastern Europe, University of Warsaw Working Paper 31(116).

Brzezinski, Michal (2014a) Paretowml: Stata program for weighted maximum likelihood estimator (WMLE) for the shape parameter of Pareto distribution.

[18] https://stats.stackexchange.com/questions/243973/is-my-data-gamma-distributed?rq=1

Brzezinski, Michal (2014b) Powerlaw: Stata programs to detect power law (or Pareto) behavior in empirical data using model fitting, goodness of fit and model selection tests.

Brzezinski, Michal (2014c) Wealth distributions and power laws: evidence from "rich lists," Physica A 406:155–162. http://coin.wne.uw.edu.pl/mbrzezinski/software/ (Accessed 4-March 2020).

Brzezinski, Michal (2016) Robust estimation of the Pareto tail index: a Monte Carlo analysis, Empirical Economics 51(1):1–30.

Buis, Maarten L. (2007) Hangroot: Stata module creating a hanging rootogram comparing an empirical distribution to the best fitting theoretical distribution, Statistical Software Components S456886, Boston College Department of Economics, revised 18 May 2011.

Buis, Maarten L., Nicholas J. Cox, and Stephen P. Jenkins (2003) Betafit: Stata module to fit a two-parameter beta distribution, Statistical Software Components S435303, Boston College Department of Economics, revised 03 Feb 2012.

Buis, Maarten L., Nicholas J. Cox, and Stephen P. Jenkins (2006) Dirifit: Stata module to fit a Dirichlet distribution, Statistical Software Components S456725, Boston College Department of Economics, revised 19 Apr 2010.

Buis, Maarten L., and Stephen P. Jenkins (2008) Fiskfit: Stata module to fit a Fisk distribution by ML to unit record data, Statistical Software Components S456900, Boston College Department of Economics.

Chauvel, Louis (2016) Intensity and shape of inequalities: the ABG method for the analysis of distributions, Review of Income and Wealth 62(1):52–68.

Cirillo, Pasquale (2013) Are your data really Pareto distributed?, Physica A 392(23):5947–5962.

Clauset, Aaron, Cosma R. Shalizi, and Mark E.J. Newman (2009) Power-law distributions in empirical data, SIAM Review 51(4):661–703.

Cowell, Frank A., and Emmanuel Flachaire (2015) Statistical Methods for Distributional Analysis, AMSE Working Papers 1507.

Cowell, Frank A., and Maria-Pia Victoria-Feser (1996a) Poverty measurement with contaminated data: A robust approach, European Economic Review 40:1761–1771.

Cowell, Frank A., and Maria-Pia Victoria-Feser (1996b) Robustness properties of inequality measures, Econometrica 64:77–101.

Cowell, Frank A., and Maria-Pia Victoria-Feser (2007) Robust Lorenz curves: a semiparametric approach, Journal of Economic Inequality 5:21–35.

Cox, Nicholas J. (1998a) Pbeta: Stata module to generate probability plot for data compared with fitted beta distribution, Statistical Software Components S350702, Boston College Department of Economics, revised 19 Nov 2003.

Cox, Nicholas J. (1998b) Pexp: Stata module to produce probability plot for data compared with fitted exponential distribution, Statistical Software Components S357701, Boston College Department of Economics, revised 30 Oct 2003.

Cox, Nicholas J. (1998c) Pgamma: Stata module to generate probability plot for data vs fitted gamma distribution, Statistical Software Components S336202, Boston College Department of Economics, revised 17 May 2011.

Cox, Nicholas J. (1998d) Qbeta: Stata module to generate quantile-quantile plot for data vs fitted beta distribution, Statistical Software Components S350703, Boston College Department of Economics, revised 19 Nov 2003.

Cox, Nicholas J. (1998e) Qexp: Stata module to produce quantile-quantile plot for data vs fitted exponential distribution, Statistical Software Components S357702, Boston College Department of Economics, revised 30 Oct 2003.

Cox, Nicholas J. (1998f) Qgamma: Stata module to generate quantile-quantile plot for data vs fitted gamma distribution," Statistical Software Components S336203, Boston College Department of Economics, revised 17 May 2011.

Cox, Nicholas J. (2002) Dpplot: Stata module for density probability plots, Statistical Software Components S426201, Boston College Department of Economics, revised 04 Aug 2004.

Cox, Nicholas J. (2004) Ppplot: Stata module for P-P plots, Statistical Software Components S438002, Boston College Department of Economics.

Cox, Nicholas J. (2007) Qpfit: Stata module containing various quantile and probability plots for assessing distribution fit, Statistical Software Components S456827, Boston College Department of Economics.

Cox, Nicholas J., and Stephen P. Jenkins (2003a) Gammafit: Stata module to fit a two-parameter gamma distribution, Statistical Software Components S435301, Boston College Department of Economics, revised 17 May 2011.

Cox, Nicholas J., and Stephen P. Jenkins (2003b) Gumbelfit: Stata module to fit a two-parameter Gumbel distribution, Statistical Software Components S435302, Boston College Department of Economics, revised 07 Dec 2012.

Cox, Nicholas J., and Stephen P. Jenkins (2007a) Invgammafit: Stata module to fit a two-parameter inverse gamma distribution, Statistical Software Components S456889, Boston College Department of Economics, revised 21 Aug 2011.

Cox, Nicholas J., and Stephen P. Jenkins (2007b) Weibullfit: Stata module to fit a two-parameter Weibull distribution, Statistical Software Components S456891, Boston College Department of Economics, revised 25 Sep 2014.

Cullen, Alison C., and H. Christopher Frey (1999) Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs, New York: Springer.

Doğru, Fatma Z., and Olcay Arslan (2015) Optimal B-robust estimators for the parameters of the Burr XII distribution, Journal of Statistical Computation and Simulation 86(6):1133–1149.

Dombos, Paul (1982) Some ways of examining dimensional distributions, Acta Sociologica 25(1):49–64.

Drees, Holger, Laurens de Haan, and Sidney I. Resnick (2000) How to make a Hill plot, Annals of Statistics 28(1):254–274.

Duan, Yutong, and Paul T. von Hippel (2016) Mgbe: Stata module to compute multimodel generalized beta estimator, Statistical Software Components S458189, Boston College Department of Economics, revised 14 Sep 2019.

Fournier, Juliette (2015) Generalized Pareto curves: Theory and application using income and inheritance tabulations for France 1901–2012, Paris School of Economics Masters thesis.

Gabaix, Xavier (1999) Zipf's Law for cities: an explanation, Quarterly Journal of Economics 114(3):739–767.

Ghosh, Souvik, and Sidney I. Resnick (2010) A discussion on mean excess plots, Stochastic Processes and their Applications 120:1492–1517.

Gonzalez Rangel, Andres L. (2013) Ggtax: Stata module to identify the most suitable GG family model, Statistical Software Components S457649, Boston College Department of Economics.

Hill, Bruce M. (1975) A Simple General Approach to Inference About the Tail of a Distribution, Annals of Statistics 3(5):1163–1174.

Hippel, Paul T. von, Igor Holas, and Samuel V. Scarpino (2012) Estimation with binned data, Cornell University working paper.

Hippel, Paul T. von, Samuel V. Scarpino, and Igor Holas (2016) Robust estimation of inequality from binned incomes. Sociological Methodology 46(1):212–251.

Huisman, Ronald, Kees G. Koedijk, Franz Palm, and Clemens J.M. Kool (2001) Tail index estimates in small samples, Journal of Business and Economic Statistics 19:208–216.

Jenkins, Stephen P. (1999a) Dagumfit: Stata module to fit a Dagum distribution by maximum likelihood, Statistical Software Components S366101, Boston College Department of Economics, revised 15 Mar 2014.

Jenkins, Stephen P. (1999b) Smfit: Stata module to fit a Singh-Maddala distribution by maximum likelihood, Statistical Software Components S366102, Boston College Department of Economics, revised 16 Mar 2007.

Jenkins, Stephen P. (2007a) Gb2fit: Stata module to fit Generalized Beta of the Second Kind distribution by maximum likelihood, Statistical Software Components S456823, Boston College Department of Economics, revised 17 Jul 2012.

Jenkins, Stephen P. (2007b) Lognfit: Stata module to fit lognormal distribution by maximum likelihood, Statistical Software Components S456824, Boston College Department of Economics, revised 01 Jun 2013.

Jenkins, Stephen P. (2014) Gb2lit: Stata module to fit Generalized Beta of the Second Kind distribution by maximum likelihood (log parameter metric), Statistical Software Components S457897, Boston College Department of Economics.

Jenkins, Stephen P. (2017) Pareto distributions, top incomes, and recent trends in UK income inequality, Economica 84(334):261–289.

Jenkins, Stephen P., and Philippe Van Kerm (2004) Glcurve: Stata module to derive generalised Lorenz curve ordinates, Statistical Software Components S366302, Boston College Department of Economics, revised 24 Jun 2008.

Jenkins, Stephen P., and Philippe Van Kerm (2007) Paretofit: Stata module to fit a Type 1 Pareto distribution, Statistical Software Components S456832, Boston College Department of Economics, revised 11 Nov 2015.

Linden, Ariel (2014) Qqplot3: Stata module to plot the unweighted and weighted quantiles of one variable against the corresponding quantiles of another, Statistical Software Components S457856, Boston College Department of Economics.

Lorenz, Max O. (1905) Methods of measuring the concentration of wealth, Publications of the American Statistical Association 9(70):209–219.

Mahdi, Smail, and Myrtene Cenac (2006) Estimating and Assessing the Parameters of the Logistic and Rayleigh Distributions from Three Methods of Estimation, Caribbean Journal of Mathematical Computer Science 13:25–34.

McDonald, James B., and Michael R. Ransom (1979a) Alternative parameter estimators based upon grouped data, Communications in Statistics 8(9):899–917.

McDonald, James B., and Michael R. Ransom (1979b) Functional Forms, Estimation Techniques and the Distribution of Income, Econometrica 47(6):1513–1525.

McDonald, James B., and Michael R. Ransom (2008) The generalized beta distribution as a model for the distribution of income: estimation of related measures of inequality, Ch.8 in

Duangkamon Chotikapanich (ed), Modeling Income Distributions and Lorenz Curves, New York: Springer Verlag, pp.147–166.

Nichols, Austin (2010) Gbgfit: Stata module to fit a Generalized Beta (Type 2) distribution to grouped data via ML, Statistical Software Components S457132, Boston College Department of Economics.

Pen, Jan (1971) Distribution: Facts, Theories, Policies, New York: Praeger.

Quandt, Richard E. (1966) Old and new methods of estimation and the Pareto distribution, Metrika 10(1):55–82.

Roodman, David M. (2015) Extreme: Stata module to fit models used in univariate extreme value theory, Statistical Software Components S457953, Boston College Department of Economics, revised 18 Dec 2017.

Schluter, Christian, and Mark Trede (2002) Tails of Lorenz curves, Journal of Econometrics 109:151–166.

Scotto, Manuel G. (2001) Hill estimator for the index of regular variation, Stata Technical Bulletin 10(56).

Stephens, Michael A. (1986) Tests based on EDF statistics. In R.B. d'Agostino and M.A. Stephens (Eds.), Goodness-of-fit techniques, pp.97–193. New York: Marcel Dekker.

Victoria-Feser, Maria-Pia., and Elvezio Ronchetti (1994) Robust methods for personal income distribution models, Canadian Journal of Statistics 22:247–258.

Zenga, Michele (1984) Proposta per un indice di concentrazione basato sui rapporti fra quantili di popolazione e quantili di reddito, Giornale degli Economisti e Annali di Economia 48:301–326.