Information Theoretic Approaches to Income Density Estimation with an Application to the U.S. Income Data*

Sung Y. Park † Anil K. Bera ‡

Abstract

The size distribution of income is the basis of income inequality measures which in turn are needed for evaluation of social welfare. Therefore, proper specification of the income density function is of special importance. In this paper, using information theoretic approach, first, we provide a maximum entropy (ME) characterization of some well-known income distributions. Then, we suggest a class of *flexible* parametric densities which satisfy certain economic constraints and stylized facts of personal income data such as the weak Pareto law and a decline of the income-share elasticities. Our empirical results using the U.S. family income data show that the ME principle provides economically meaningful and a very parsimonious and, at the same time, flexible specification of the income density function.

JEL code: C16; D31

Key words: Income density estimation; Information theoretic approach; Maximum entropy; Weak Pareto law.

*We are grateful to the editor and three anonymous referees for many pertinent comments and helpful suggestions. We would also like to thank Duncan Foley, Essie Maasoumi, Roger Koenker, Zhongjun Qu and Ximing Wu for their comments on an earlier version of the paper. However, we retain the responsibility for any remaining errors. Financial support from the Research Board, University of Illinois at Urbana-Champaign is gratefully acknowledged.

†School of Economics, Chung-Ang University, 84 Heuksok-Ro, Dongjak-Gu, Seoul, Korea. Tel: +82-2-820-5622. Fax: +82-2-812-9718. Email: sungpark@cau.ac.kr.

‡Corresponding author. Department of Economics, University of Illinois, 1407 W. Gregory, Urbana, Illinois 61801, USA. Tel.: +1-217-333-4596; fax: +1-217-244-6678; Email: abera@illinois.edu.

1 Introduction

Measurement of the income inequality to evaluate social welfare is of particular interest to economists and policy makers. Since the size distribution of income is the basis of such inequality measures, correct specification of the income density function is of utmost importance. There is a very long history of research on models for the size distribution of income: Pareto (1895, 1896, 1897) formulated the "laws" of personal income after observing that a plot of the logarithm of the number of people above a certain level of income against the logarithm of that level has almost a *linear* representation with a negative slope. Since then there had been numerous developments in formulating various income density functions and measuring income inequality. Aitchison and Brown (1957) advocated the use of the log-normal density to describe the income distribution, whereas Salem and Mount (1974) suggested using the gamma distribution. Empirical evidences accumulated over a long period of time show that the log-normal and gamma distributions fit the data relatively well in the middle range of income but tend to exaggerate the skewness and fit poorly toward the tails, see for instance, Gastwirth (1972), Kloek and van Dijk (1977, 1978), McDonald and Ransom (1979), Dagum (1977), Ransom and Cramer (1983), and McDonald (1984).

Subsequently, many more models for income distributions have been proposed. Among *parametric* densities, along with the log-normal and gamma, Singh and Maddala (SM) (1976), generalized gamma (GG) (Kloeck and van Dijk, 1978) and generalized beta of the first (GB1) and -second kind (GB2) (McDonald, 1984) are most popular. Many well-known density functions such as exponential, Weibull, Fisk, the beta of the first and -second kinds, log-normal, gamma and SM, are special cases of GG, GB1 and GB2. Kleiber and Kotz (2003) provides an excellent overview of size distributions and estimation methods. The empirical results of various studies show that GB2 outperforms other two- to four-parameter distributions in the goodness-of-fit sense (see, McDonald, 1984; McDonald and Xu, 1995; Bordley, McDonald and Mantrala, 1996; Dastrup, Hartshorn and McDonald, 2007). Although the density GB2 is flexible enough to take care of various types of income data, following the tradition of

Haavelmo (1944), it is necessary to carry out analysis for any possible misspecification. Boccanfuso, Decaluwé and Savard (2008) found that selecting inappropriate income distribution can yield biased results in terms of poverty analysis. If a density function is correctly specified, then, the maximum likelihood estimation preserves consistency and efficiency. However, due to lack of complete information, the true density is rarely achieved; therefore, it is necessary to subject any proposed income density function to a battery of specification tests. In this paper, we consider a class of flexible income densities derived from optimizing a well-defined distance measure, namely, the maximum entropy (ME) subject to certain moment constraints. Careful choice of moment-constraints based on economic theory, past empirical evidence and specification tests leads to an well-specified density function that is able to extract the essential information from the data. By so doing we are, as our empirical application demonstrates, able to arrive close to the "true" model for income distribution.

Density estimation based on minimum divergence methods, i.e., minimizing some appropriate distance norm between the assumed and the true densities, has been studied quite extensively, see for instance, Kullback and Leibler (1951), Renyi (1960), Cressie and Read (1984), and Lindsay (1994). By minimizing such distance measures subject to certain moment constraints possibly involving some unknown parameters, one can obtain a very probable distribution. We consider the Kullback-Leibler information criteria with uniform reference density as the distance norm, which is nothing but the negative of Shannon's (1948) entropy measure (SEM). A ME density (MED) is obtained by maximizing SEM subject to certain moment constraints which can be regarded as "prior" information. By choosing different sequences of moment constraint functionals, we have various flexible MED functions belonging to a generalized exponential family. MED is known to be the least biased distribution given known moment constraints (Kapur and Kesavan (1992)).

Although we consider only parametric family, one can also adapt *non-parametric* approaches to estimate income distribution. For the non-parametric density, however, the estimated tail-behavior may not be satisfactory due to the scarcity of data in the tail parts. Minoiu and Reddy (2014) showed that the performance of the estimated non-parametric kernel density has nontrivial biases in estimated poverty levels and suggested to use a parametric income density. Furthermore, in our approach to MED estimation, we can ensure that the Pareto law is not violated by selecting appropriate moment functions.

Although parametric approaches have such advantages, we should be very careful in selecting specific density that takes care of some kind of stable structure of personal income data. One can almost always construct a flexible density having higher goodness-of-fit measure than that of all previously considered models. Finding a parsimonious density that obeys the Pareto law and other stylized facts is always a challenge. Esteban (1986) showed that the GG is the only density function that satisfies three stylized facts, i.e., (i) the weak Pareto law; (ii) possessing at least one interior mode; (iii) a constant rate of decline of the income-share elasticities. Majumder and Chakravarty (1990) proposed a 4-parameter income density function satisfying conditions (i) and (iii) above. McDonald and Mantrala (1995) showed that Majumder and Chakravarty (1990)'s density function is a reparameterized version of GB2. We demonstrate that by choosing appropriate moment functions, stylized facts and constraints implied by economic theory can be incorporated into the estimated density in a parsimonious way.

Standard MED models used in the literature are based on the power series or some orthonormal series as the moment functions. One exception is Leipnik (1990) that considered general moment functions based on utility functions, and derived many well-known income distributions, such as log-normal, Pareto, SM and GB2. However, since the distribution function, $F(x)$ itself appears in the moment function, deriving an explicit form of the MED is quite difficult, and this methodology is hard to implement in practice. Ryu and Slottje (1997) and Wu (2003) considered the Legendre polynomials up to 4-th order and arithmetic moments up to 12-th order, respectively. Wu and Perloff (2005) used $\ln(1+x)^i$, $i = 1, 2, 3, 4$ as moment functions and reported that these provided the best overall fit for Chinese household income data. Such approaches may lack economic interpretation in the sense that some of the moment functions are used without considering stylized facts and economic implications of the resulting density function. Furthermore, existence of higher-order arithmetic moments cannot always be ensured. Wu and Perloff (2007) also considered a generalized method of moment (GMM) estimator for the distribution of a variable where summary statistics are available only for certain intervals. However, they do not consider the characterization of various income densities that satisfy empirical stylized facts of personal income data. We consider moment functions that represent distributional characteristics more directly.[1]

The rest of the paper is organized as follows. Section 2 introduces MED under general moment conditions and gives examples of ME income densities. In Section 3, we present some basic characteristics of the MED in the context of modeling income distribution and discuss which moment functions are appropriate to capture the stylized facts. In Section 4, we discuss estimation and suggest a moment selection criteria based on Rao's score (RS) test principle. Section 5 provides an empirical application to the U.S. income data with specific moment functions. The paper is concluded in Section 6.

2 ME density under general moment conditions

When prior information are available in the form of moments of unknown distribution, ME principle recovers the distribution without using any further information but only those moments. This principle has been used in the construction of the prior density functions in the Bayesian literature (Zellner (1977) and Berger (1985, pp. 90-94)). Simple moment conditions usually take form of $E[\phi(x)] = \int \phi(x) f(x) dx = \mu$, where $\phi(\cdot)$ and μ are, respectively, a vector valued function and a given constant vector. We propose using generalized moments where the function $\phi(\cdot)$ also involves an additional unknown parameter vector γ, i.e., $E[\phi(x, \gamma)] = C(\gamma)$. The generalized MED is obtained by maximizing Shannon's (1948) entropy measure

$$H(f) = -\int f(x) \ln f(x) dx, \tag{1}$$

[1] Park and Bera (2009) demonstrated the usefulness of MED approach in finding suitable density functions for autoregressive conditional heteroskedasticity (ARCH)-type models for financial time series data.

subject to the constraints

$$\int \phi_j(x,\gamma) f(x) dx = C_j(\gamma), \quad j = 0, 1, 2, \cdots, q. \tag{2}$$

The normalization constraint corresponds to $j = 0$ by setting $\phi_0(x,\gamma)$ and $C_0(\gamma)$ to 1. The solution to (1)-(2), obtained by applying the Lagrangian procedure, is the generalized exponential density

$$f(x;\theta) = \frac{1}{\Omega(\theta)} \exp\left[-\sum_{j=1}^{q} \lambda_j \phi_j(x,\gamma)\right], \tag{3}$$

where $\Omega(\theta) = \int \exp\left[-\sum_{j=1}^{q} \lambda_j \phi_j(x,\gamma)\right] dx$, and $\theta = (\lambda', \gamma')'$ with $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_q)'$. When the parameter vector γ is absent from $\phi_j(x,\gamma)$ in (2), the solution (3) belongs to the exponential family, where λ_j is the Lagrange multiplier corresponding to the j-th constraint in (2) for maximizing $H(f)$, $j = 1, 2, \cdots, q$. For example, if the moment function is given by $\phi_1(x,\gamma) = x$ with $\phi_0(x,\gamma) = 1$, the resulting MED is the exponential density, i.e., $f(x;\lambda_1) = \lambda_1 e^{-\lambda_1 x}$, $x \geq 0$. With an additional moment function, $\phi_2(x,\gamma) = x^2$, the MED is a $N(\mu,\sigma^2)$ density with $\mu = -\lambda_1/(2\lambda_2)$ and $\sigma^2 = 1/(2\lambda_2)$. In the context of general exponential family, λ_j can be viewed as a natural parameter and $\sum_i^N \phi_j(x_i,\gamma)$ the corresponding sufficient statistic for a given γ, $j = 1, \cdots, q$, where N is the sample size. The statistical characteristics of exponential family are well-known, and therefore, it is convenient for estimation and inference. However, in some cases, a large number of moment functions might be needed to fit the unknown density if the *simple* moment functions $\phi_j(x)$'s are not flexible enough. For example, in order to fit a fat tailed density we need high orders of the arithmetic moment functions. Having parameter vector γ in $\phi_j(\cdot,\cdot)$, makes the (generalized) moment function more flexible, and the resulting MED quite general.

Since the solution form of prescribed constrained maximization problem is exponential, we can regard the MED estimation procedure as a non-linear "regression" model in which the log-density and moments functions are dependent and independent variables, respectively, which is seen by writing (3) as

$$\ln f(x;\theta) = \text{Const} - \sum_{j=1}^{q} \lambda_j \phi_j(x,\gamma). \tag{4}$$

Fitting a log-density might be easier than estimating a density function directly since log-transformation generally reduces the degree of curvature of an underlying function. Moreover, the log-density can take any value in the real line, while a proper density should take only non-negative values.

To provide an early illustration of our approach, in Figure 1 we plot three estimated log-densities along with the empirical log-probability. The first model, $f1$, is the MED based on two moment functions, $\ln x$ and $\ln(1+(x/b)^a)$, where a and b are unknown parameters, while only $\ln(1+(x/b)^a)$ is considered as the moment function for $f2$. Finally, two simple moment functions, $\ln x$ and $(\ln x)^2$ are used in $f3$. Clearly, $f1$ fits better than the other two models. Note that the major difference between estimated $f1$ and $f2$ comes from the lower income intervals. $f2$ is flat in the lower income levels but achieves quite similar goodness-of-fit to that of $f1$ in the middle and higher income levels. $f3$ though not flat for the lower income level, it strays away from the histogram throughout all income levels. Therefore, we can say that the moment functions $\ln x$ and $\ln(1+(x/b)^a)$ are informative for the higher and lower income intervals, respectively. In fact, $f1$ and $f3$ are re-parameterized versions of GB2 and log-normal distributions, respectively. Therefore, selection of appropriate and well-behaved moment functions is critical to the construction of MED, as we elaborate in the next Section.

[Figure 1]

3 Characterization of ME income distribution

Maximum entropy distribution has a very flexible functional form. By choosing a sequence of moment functions $\phi_j(x)$, $j = 1, 2, \cdots, q$, we can generate a sequence of various flexible MED functions. Many well-known families of distributions can be obtained as special cases of MED function. Kagan, Linnik and Rao (1973) provided characterization of many distributions, such as, the beta, gamma, exponential and Laplace distributions as ME densities. Gokhale (1975) presented characterization of the univariate normal, double exponential and Cauchy, and the multivariate Dirichlet and Wishart distributions. Ord, Patil and Taillie (1981) provided characterization of three typical income distributions, gamma, Pareto and log-normal. Cobb, Koppstein and Chen (1983) obtained a general class of multimodal density functions within a unified framework of stochastic catastrophe models. Similar to ME principle, the system in stochastic catastrophe models behaves as if it moves towards the points of lowest potential.

Table 1 shows the characterization of some well-known income distributions. These distributions can be interpreted in an information theoretic way that they can be obtained by imposing moment constraints which are inherent in the data and thus the relationships among various distributions can be seen from the underlying moment restrictions. For example, the log-normal (LN) distribution is the resulting MED if the prescribed moment conditions are $E[\ln x] = \mu$ and $E[(\ln x)^2] = \mu^2 + \sigma^2$. Weibull and generalized gamma (GG) are characterized by the same two moment functions $\phi_1(x) = \ln x$ and $\phi_2(x, a) = x^a$; the resulting densities are, however, not the same. This is due to the flexibility of the expected values of the moment functions of GG can take. For Weibull case, stronger restrictions of the constraints, $E[x^a] = 1$ and $E[\ln x] = -\gamma/a$ lead to an *one*-parameter distribution, while the flexible ranges of values that $E[x^a]$ and $E[\ln x]$ can take for the GG case are much wider and that results in a *three*-parameter density. Similar arguments can be made for the Fisk, SM, Dagum and GB2 distributions. We can check that the MED associated with two moment functions, $E[\ln x]$ and $E[\ln(1 + (x/b)^a)]$ is a re-parameterized version of GB2 distribution. Clementi, Gallegati and Kaniadakis (2010) proposed κ-generalized distribution derived from the ME principle with 3 moment functions: $\ln x$, $\ln(1+\kappa^2(x/\beta)^{2\alpha})$ and $\sinh^{-1}(-\kappa(x/\beta)^{\alpha})$, where α, β and κ are parameters of the distribution. As McDonald (1984) noted and can be easily seen from the expression of densities given in Table 1, Fisk is a special case of SM with $q = 1$, and SM is obtained from GB2 by setting $p = 1$. Finally, note also that Dagum distribution is a special case of GB2 when $q = 1$.

[Table 1]

Using Figure 1 we earlier noted $\ln x$ and $\ln(1 + (x/b)^a)$ can capture the distributional patterns of the lower and upper income levels, respectively. It is, therefore, worthwhile to consider some well-behaved functions having similar behavior to the moment functions corresponding to GB2. There are, of course, many such functions. From these, we can select those functions that satisfy the weak Pareto law (WPL). In many empirical applications, the Pareto distribution has been found to fit well toward the upper tail region. Let us define the share of total income earned by individuals with income in the interval $[x, x + h]$ as

$$\xi(x, x + h) = \frac{1}{\mu} \int_x^{x+h} z f(x) dz,$$

where $f(\cdot)$ is the density function, and $\mu = \int_0^\infty z f(z) dz$ the mean income. Esteban (1986) showed that an income distribution can be *uniquely* characterized by its income share elasticity function, defined by

$$\eta(x, f) = \lim_{h \to 0} \frac{d \ln \xi(x, x + h)}{d \ln x} = 1 + \frac{x f'(x)}{f(x)}.$$

Then, the density $f(\cdot)$ satisfies the WPL if

$$\lim_{x \to \infty} \eta(x, f) = -\alpha, \tag{5}$$

for some $\alpha > 0$. Some commonly used densities, such as lognormal and gamma, do not satisfy WPL. The income share elasticity associated with MED in (3) is given by

$$\eta(x, f) = 1 - \left[\sum_{j=1}^{q} \lambda_j \phi'_j(x, \gamma) \right] x. \tag{6}$$

Therefore, we can easily verify whether WPL is satisfied or not by checking the boundedness of $\phi'_j(x, \gamma)x$ with respect to x, $j = 1, 2, \cdots, q$.

As presented in Table 1, the well-known income densities can be characterized by a few moment functions, for example, x, $\ln x$, $\ln(1-x)$, x^a and $\ln(1+(x/b)^a)$. More general income density functions that have the similar behavior with those of the well-known income densities, can be obtained by considering the flexible moment functions. For example, the shapes of $\tan^{-1}(x)$ and $\sinh^{-1}(x)$ are quite close to those of $\ln x$ and $\ln(1+x)$. Moreover, the behavior of $\sinh^{-1}(x^a)$ is very similar to that of $\ln(1+x^a)$ for $a > 0$. Thus one can construct a generalized income density function by incorporating more flexible moment functions.

To get more insight of various moment functions, in Figure 2 we plot four common functions: $\ln x$, $\ln(1+x)$, $\tan^{-1}(x)$ and $\sinh^{-1}(x)$ used in the literature to generate a wide variety of densities; the positive and negative functions are in the left and right panels, respectively. The effect of $\phi_j(x, \gamma)$ to the corresponding log-density depends on the Lagrange multiplier λ_j, $j = 1, 2, \cdots, q$ [see equation (4)]. If $\lambda_j < 0$ then the associated moment function can take care of the distribution at the lower income levels, where, usually, the number of individuals increases sharply as x increases from zero. The rate of increase in $\ln x$ is higher than that of the other three functions, and therefore, when there is sharp increases of the number of individuals at the lower income levels, it fits this part of the log-density well. It is not surprising that most, if not all, well-known income distributions are characterized by the moment condition $E[\ln x] = c$ for some constant c [see Table 1]. When $\lambda_j > 0$, all corresponding moment function plays a different role and explains the middle and upper tail of the distribution. For instance, in the restricted income range, $x > x_0 > 0$, $E[\ln x] = 1/\alpha + \ln x_0$ characterizes the Pareto density, $f(x : \alpha) = \alpha x_0^\alpha / x^{\alpha+1}$, where $\alpha > 0$, which in turn can be written as

$$\alpha x_0^\alpha \exp[-(\alpha+1)\ln x] = \exp[-\lambda \ln x]/\Omega(\alpha) \quad \text{(say)},$$

where since $\alpha > 0$, we have $\lambda > 0$. Therefore, it is not surprising that the Pareto distribution fits the upper tail part well but not the rest of the distribution.

[Figure 2]

Three moment functions $\ln(1+x^2)$, $\tan^{-1}(x^2)$ and $\sinh^{-1}(x^2)$ are plotted in the lower panel of Figure 2. Comparing to plots in the upper panel of Figure 2, these moment functions do not increase sharply for the smaller values of x. Thus these are not appropriate to represent distributions that have sharp increases at the lower income level. Instead, if the Lagrange multiplier is positive (negative moment functions), these functions can take care of the middle and upper tail part of the distribution. However, distribution generated by these functions, except $-\tan^{-1}(x^2)$, may not have enough thick tails compared to $-\ln x$, $-\tan^{-1}(x)$ and $-\sinh^{-1}(x)$. We should, however, note that $\tan^{-1}(x)$ and $\tan^{-1}(x^2)$ cannot be used as the moment function to explain right tail part of income distribution; $\tan^{-1}(\cdot)$ function being bounded, the associated MED cannot be defined over the whole range $0 < x < \infty$. We can consider more flexible moment functions that involve additional parameters. In Figure 3, we plot $\ln(1 + x^a)$ and $\sinh^{-1}(x^a)$ for five different values of a, and these two moment functions have very similar behavior.

[Figure 3]

Overall, most of the densities used in the literature [see, Table 1] can be characterized by just *two* moment functions that take care of lower, middle and upper income levels at the same time. From this point of view, the fitted log-density is nothing but a linear combination of two moment functions $\phi_1(x, \gamma)$ and $\phi_2(x, \gamma)$ with corresponding Lagrange multipliers, λ_1 and λ_2, as the weights plus a constant [see equation (4)]. When these two moment functions are not enough to take care of underlying income data, we can add one more moment function $\phi_3(x, \gamma)$ to the ME problem to extract further relevant information from the data. There is an added safeguard in our approach. Using the RS test principle, we check the relevancy of our moment conditions for the data in hand, as discussed in the next section.

4 Estimation and moment selection test

4.1 Estimation of MED

Many of the papers that considered estimation of distribution using ME approach dealt with only pure discrete or continuous data, not grouped data. Income data are, however, sometimes available for a fixed number of intervals with respective frequencies. There are some studies that used the grouped income data. For example, Singh and Maddala (1976) and McDonald (1984) used the grouped income data to estimate the income densities and Wu and Perloff (2007) considered the GMM estimation of a MED with interval data. In the case of grouped income data the standard maximum likelihood method is not directly applicable.

Using the multinomial distribution one can construct the likelihood function. Assume that the income range I can be divided into K intervals I_k, $k = 1, 2, \cdots, K$. Let n_k be the frequency (number of individuals) of the k-th interval with $\sum_{k=1}^{K} n_k = N$, N being the total number of individuals. The likelihood function associated with the observed frequencies $n_1, n_2, \cdots, n_K$ can be written as

$$L(\theta) = N! \prod_{k=1}^{K} \frac{[P_k(\theta)]^{n_k}}{n_k!},$$

where $P_k(\theta) = \int_{I_k} f(x;\theta)dx$ with $f(x;\theta)$ having the form of (3). Thus the log-likelihood function $l(\theta)$ that we use for our estimation of parameter vector θ, is given by

$$l(\theta) = \ln L(\theta) = d + \sum_{k=1}^{K} n_k \ln P_k(\theta), \tag{7}$$

where $d = \ln N! - \sum_{k=1}^{K} \ln n_k$. The maximum likelihood estimator which maximizes the above log-likelihood function (7) is known to be asymptotically efficient relative to other estimators based on the interval data [see Aigner and Goldberger (1970)]. Since $f(x;\theta)$ in $P_k(\theta)$ is selected from a variety of moment functions, one cannot guarantee $P_k(\theta)$ to have analytic forms. Therefore, in practical applications these are computed numerically using numerical integration. We find Gauss-Legendre quadrature approach to solve above numerical integration problem works very well.

For the unit record income data the same moment functions can be used to construct a flexible income density function. The only difference in the estimation procedure between the grouped data and the unit record data is the log-likelihood function. In the case of the unit record income data the log-likelihood function, $l(\theta)$, can be expressed by

$$l(\theta) = -N \ln \Omega(\theta) - \sum_{i=1}^{N} \sum_{j=1}^{q} \lambda_j \phi_j(x_i, \gamma),$$

where $\Omega(\theta) = \int \exp\left[-\sum_{j=1}^{q} \lambda_j \phi_j(x, \gamma)\right] dx$ and x_i represents actual unit income observation, $i = 1, 2, \cdots, N$. It is easy to see that the above log-likelihood function results naturally from the MED given in (3). The standard maximum likelihood estimation method can be directly applied to estimate the unknown parameter vector θ. In this case, generally, $\ln \Omega(\theta)$ does not have an analytic expression so that it should be computed by numerical integration.

4.2 Moment selection test

In order to construct a flexible ME income density, it is of importance to select appropriate moment functions which determine the shape and characteristics of the ME income density function. As discussed in Section 3, there are many candidate moment functions such as $\tan^{-1}(x/b)$, $\sinh^{-1}(x/b)$, $\ln(1 + (x/b)^2)$, $\sinh^{-1}(x/b)^2$, $\ln(1 + (x/b)^a)$ and $\sinh^{-1}((x/b)^a)$. In this subsection, we propose a test statistics by which the suitability of the moment functions can be decided.

As we discussed earlier the Lagrange multipliers ($\lambda_j, j = 1, 2, \cdots, q$) provide rates of change of the maximum attainable value of $H(f)$ in (1) with respect to the change in the constraints, (2). Therefore, empirically λ_j should be very close to zero if the j-th associated moment function $\phi_j(x, \gamma)$ does not provide useful information. Our proposed test statistic for moment selection, i.e., for testing $H_{0j} : \lambda_j = 0$, is based on the RS principle. This can be easily extended to a joint test for multiple moment conditions.

By substituting $\int_{I_k} f(x;\theta)dx$ for $P_k(\theta)$ in (7) and then using (3), we obtain

$$\begin{aligned} l(\theta) &= d + \sum_{k=1}^{K} n_k \ln\left[\int_{I_k} f(x,\theta)dx\right] \\ &= d + \sum_{k=1}^{K} n_k \ln \int_{I_k} \frac{1}{\Omega(\theta)} \exp\left[-\sum_{j=1}^{q} \lambda_j\phi_j(x,\gamma)\right]dx \\ &= d - N\ln\Omega(\theta) + \sum_{k=1}^{K} n_k \ln \int_{I_k} \exp\left[-\sum_{j=1}^{q} \lambda_j\phi_j(x,\gamma)\right]dx. \end{aligned}$$

The first derivatives of $l(\theta)$ with respect to λ_l, $l = 1, 2, \cdots, q$, is given by [here temporarily we use 'l' to avoid confusion with 'j', used in the above expression of $l(\theta)$]

$$d_l(\theta) = \frac{\partial l(\theta)}{\partial \lambda_l} = -N\frac{\partial \ln\Omega(\theta)}{\partial \lambda_l} + \sum_{k=1}^{K} n_k \int_{I_k} -\phi_l(x,\gamma)\Omega_{I_k}^{-1}(\theta) \exp\left[-\sum_{j=1}^{q} \lambda_j\phi_j(x,\gamma)\right]dx, \tag{8}$$

where $\Omega_{I_k}(\theta) = \int_{I_k} \exp\left[-\sum_{j=1}^{q} \lambda_j\phi_j(x,\gamma)\right]dx$ and

$$\frac{\partial \ln\Omega(\theta)}{\partial \lambda_l} = \frac{\int_x -\phi_l(x,\gamma)\exp\left[-\sum_{j=1}^{q} \lambda_j\phi_j(x,\gamma)\right]dx}{\Omega(\theta)}.$$

The score function (8), under the null hypothesis $H_{0l} : \lambda_l = 0$, reduces to

$$d_l^0 \equiv d_l(\theta)|_{\lambda_l=0} = N\delta_l - \Delta_l, \tag{9}$$

where

$$\delta_l = E_{\tilde{f}}[\phi_l(x,\gamma)] = \int \phi_l(x,\gamma)\tilde{\Omega}^{-1}(\theta)\exp\left[-\sum_{\{j=1,2,\cdots,q\}\backslash\{l\}} \lambda_j\phi_j(x,\gamma)\right]dx,$$

$$\text{and} \quad \Delta_l = \sum_{k=1}^{K} n_k \int_{I_k} \phi_l(x,\gamma)\tilde{\Omega}_{I_k}^{-1}(\theta)\exp\left[-\sum_{\{j=1,2,\cdots,q\}\backslash\{l\}} \lambda_j\phi_j(x,\gamma)\right]dx.$$

Here, $\tilde{\Omega}(\theta) = \int \exp\left[-\sum_{\{j=1,2,\cdots,q\}\backslash\{l\}}^{q} \lambda_j\phi_j(x,\gamma)\right]dx$ and $\sum_{i=\{1,2,\cdots,q\}\backslash\{j\}}$ means summation over $i = 1, 2, \cdots, j-1, j+1, \cdots, q$. Hence, the RS test for $H_{0j} : \lambda_j = 0$ will be based on

$$R_j(\theta) = d_j^0 = \delta_j - \frac{1}{N}\Delta_j, \tag{10}$$

which can be regarded as the difference between population mean relating to the expected value of the j-th moment function, i.e., $E[\phi_j(x,\gamma)]$, and its sample counterpart Δ_j, all evaluated under the null hypothesis. An operational form of RS statistic is

$$RS_j = N \cdot \frac{R_j^2(\hat{\theta})}{\hat{V}_j}, \tag{11}$$

where $\hat{\theta}$ is the maximum likelihood estimates of θ, and $\hat{V}$ is a consistent estimator of asymptotic variance of $\sqrt{N}R_j(\hat{\theta})$.[2] Under the null hypothesis, RS_j will be distributed asymptotically as χ_1^2. Since analytical expression of V_j is quite complicated, we obtain variance of $\sqrt{N}R_j(\hat{\theta})$ is by the bootstrap method. Using the estimated bootstrap variance $\hat{V}_{j,B}$, an operational form of the RS test statistic is given by $RS_{j,B} = N \cdot R_j^2(\hat{\theta})/\hat{V}_{j,B}$, where B denotes bootstrap sample size. Under the null hypothesis, as $B \to \infty$, $RS_{j,B}$ is asymptotically distributed as χ_1^2. For finite B, $RS_{j,B}$ is asymptotically distributed as $F_{1,B-1}$.[3] In our application in the next section, we set $B = 200$, i.e., 200 bootstrap samples are drawn from a multinomial distribution with parameters $(n_1/N, n_2/N, \cdots, n_k/N)'$ to calculate $\hat{V}_{j,B}$.[4]

5 Empirical application to U.S. income data

To illustrate the suitability of our methodology, we consider various types of ME income distribution using the U.S. family income data for the years 1970, 1980, 1990, 2000 and 2005.[5] Although family income measure has some drawback in some cases, for example, it disregards persons living in nonfamily household, it is appropriate in some other cases to exclude nonfamily households, for example, housing affordability. The Census Bureau conducts a survey from which it derives annual estimates of the distribution of income across households, families, and individuals with income.[6] The official measure of income of the Census Bureau is money income. It includes earnings, dividends, pensions, interest,

[2]The moment selection test for the unit record income data can be derived similarly based on the log-likelihood function for the unit record income data. We refer the readers to Park and Bera (2009) for a more detailed derivation of moment selection test for the unit record data.

[3]Dhaene and Hoorelbeke (2004) shows that $RS_{j,B}$ is asymptotically distributed as Hotelling's T^2 with $(1, B-1)$ degrees of freedom, in short $T^2_{1,B-1}$ which can be also represented by $F_{1,B-1}$.

[4]See Jhun and Jeong (2000) and Morales, Pardo and Santamaría (2004) for detailed expositions of the bootstrap method for categorical data.

[5] In the Current Population Survey (CPS), a household is defined as all of the individuals who occupy a housing unit as their usual place of residence. A family is defined as a group of two or more individuals who reside together and who are related by birth, marriage, or adoption.

[6]The data are publicly available online at the web pages of U.S. Bureau of the Census (series P-60): `https://www.census.gov/prod/www/population.html`

and government non means-tested income, for example, unemployment compensation, social security benefits, and veterans' payments. Money income is calculated on a pre-tax basis so that it does not include the value of noncash benefits, such as food stamps, medicare, medicaid, public or subsidized housing, and employment-based fringe benefits. The data are in a grouped format with 11 groups for 1970 and 1980 and 21 groups for 1990, 2000 and 2005. The total money income data are given in Table A1 (Appendix), and were taken from the Census Population Report. To conserve space we report our test and estimation results only for the years 2000 and 2005. The results for the remaining years are similar, and reported in an earlier version of the paper.[7]

We consider 3-, 4- and 5-parameter ME income densities (MEIDs) with four common income distributions: Log-normal (LN), GG (generalized gamma), SM (Singh-Maddala), Dagum and GB2 (generalized beta of the second kind). For 5-parameter MEID, appropriate density functions are chosen by proposed moment selection test. In Table 2, we list all the two-moment-functions, used in 3 (Three)- and 4 (Four)-parameter MEIDs, respectively, denoted by T1, T2,$\cdots$, T6 and F1,$\cdots$, F4. Here the first and second functions take cares of fitting lower and upper levels of income, respectively, and therefore, the expected signs of respective Lagrange multipliers should be negative and positive. Income share elasticity expression, $\eta(x, f)$ in (6), along with their limit values as $x \to \infty$, for all the ten listed models, are also given in Table 2. The parameter "a" makes the difference between 3- and 4-parameter MEIDs, as we have seen in Figure 3. Three-parameter MEIDs are special cases of 4-parameter models with $a = 2$; for example, T1 and T2 are special cases of F1 and F2, respectively. F1 is of special interest since GB2 is derived as MEID with these two moment functions (see Table 2), of which the first and second explain the lower and upper region of income, respectively, as we have seen in Figure 1.

[Table 2]

[7] We refer the reader to an earlier version of the paper, Park and Bera (2013) for a more detailed set of results on estimation, see `www.sungpark.net/MEID_ParkBera_2013.pdf`

The estimates of 3-parameter MEIDs along with a sum of squared errors (SSE) and absolute errors (SAE), chi-squared (CSQ) and cross entropy (CE) are reported in Tables 3 and 4 for the years 2000 and 2005, respectively. SSE, SAE, CSQ and CE are calculated as

$$\mathrm{SSE} = \sum_{k=1}^{21} \left(\frac{n_k}{N} - P_k(\hat{\theta}) \right)^2,$$

$$\mathrm{SAE} = \sum_{k=1}^{21} \left| \frac{n_k}{N} - P_k(\hat{\theta}) \right|,$$

$$\mathrm{CSQ} = N \sum_{k=1}^{21} \left(\frac{n_k}{N} - P_k(\hat{\theta}) \right)^2 / P_k(\hat{\theta}),$$

$$\text{and} \quad \mathrm{CE} = \sum_{k=1}^{21} P_k(\hat{\theta}) \ln \left(P_k(\hat{\theta}) / \frac{n_k}{N} \right).$$

$\ln L$, as in (7), provides the log-likelihood values. Since $\Omega(\theta)$ and $P_k(\theta)$ do not have analytical forms, these were computed using numerical integration. Since our estimation involves nonlinear optimization technique, there would be some approximation errors.[8] In order to check the validity of estimated models we calculate the mean value and Gini coefficient for each model, and reported as mean and Gini in Tables 3 and 4. Following McDonald (1984), the Gini coefficient can be expressed by

$$\mathrm{Gini} = E(|y - x|)/2\mu = (1/\mu)(I^*(1, 0) - I^*(0, 1)),$$

where $\mu = E(y)$ and $I^*(i, j) = \int_0^\infty x^i f(x) \int_0^x y^j f(y) dy dx$. The Gini coefficient can be calculated using numerical integration.[9] As expected, λ_1 and λ_2, respectively, have negative and positive signs, for *all* the cases. This confirms that the first and second moment functions take care of left and right parts of income distribution, respectively. The mean values and Gini coefficients are found to be close to the census estimates. For 2000 (Table 3), SM provides better fit than GG judged by all five goodness-of-fit criterion, and SM and GB2 have almost identical goodness-of-fit. SM fits even better than GB2 in terms of SSE and SAE,

[8]The `CML` procedure in the `GAUSS 9.0` program was used to maximize the multinomial log-likelihood function. For the numerical integration we use the `intqud1` function in the `GAUSS` program. The `intquad1` function uses Gauss-Legendre quadrature to integrate a function. Program codes are available from the authors upon request.

[9]We use the `intgrat2` function in the `GAUSS` program to calculate the double integral.

however, its log-likelihood value is small since it is a special case of GB2 ($p = 1$). There is no 3-parameter MEID that outperforms SM, however, T1 and T2 outperforms GG. It is interesting to see that the performance of T1 is better than that of GB2 in terms of all goodness-of-fit measures. In Table 4 (2005 data), T1 and T2 are better than GG and SM based on all five goodness-of-fit criteria and, interestingly, T2 gives lower values of SSE, SAE and CSQ than GB2. These results are of interest since a 3-parameter model is parsimonious.

Overall, T1 and T2 fit the underlying U.S. income data very well among 3-parameter MEIDs and dominate GG and Dagum for 2000 and SM and GG for 2005. Unfortunately, MEID for T2 does not have analytic normalizing constant so the estimation of such an MEID involves numerical integration. However, MEID for T1, of course, has an analytical form for $\Omega(\theta)$, given by $2b^{\lambda_1-1}/B(1/2-\lambda_2/2, -1/2+\lambda_1/2+\lambda_2)$ if $\lambda_1+2\lambda_2>1$, $\lambda_1<1$ and $b>0$.

[Table 3]

[Table 4]

For the 5-parameter MEIDs (G), moment selection tests are performed starting with four 4-parameter MEIDs, F1 to F4 as distribution under the respective null hypothesis. The results of our tests based on the RS statistic given in (11) with bootstrap sample size 200 are reported in Table 5. For the estimated distributions F1 to F4 we test whether each of the five listed additional moment functions is informative enough to capture the shape of income density. None of the additional moment functions are informative for 2000 and 2005 income data when the null density is based on F1. As noted earlier, F1 is a reparameterized version of GB2. Therefore, we can say that the functions in GB2 act like "sufficient" moment functions in the sense that once we start with GB2 specification, any additional moment function does not add any further information. In other words, moments constraints of GB2 exhaust "all" the information regarding the density that is available from the data. When F3 and F4 models are the null models, it is clear that we need to add the moment function $\ln x$. An additional third moment function can explain lower or upper (and middle) parts of income distribution when the first two miss some valuable information in the data. Assuming that signs of current Lagrange multipliers do not change, statistically significant negative and positive Lagrange multiplier associated with an additional moment constraint indicates deficiency of the density function under the null hypothesis over lower and upper income regions, respectively. It is, of course, possible that such moment function can alter the signs of current Lagrange multipliers so that new moment function takes care of one region alone.

[Table 5]

In Tables 6 and 7, we present results from models under several combinations of moment functions for which the Lagrange multipliers were significant (in Table 5) along with four 4-parameter MEIDs, F1 to F4, for the years 2000 and 2005. Among 4-parameter MEIDs, F4 fits better than any other models for 2000 and 2005, and therefore, we suggest the density based on F4 as an alternative 4-parameter model to GB2. For 5-parameter (Generalized) MEID, six models are considered by adding one extra moment function to 4-parameter MEIDs from the results of moment selection test in Table 5: (i) F4+$\ln x$; (ii) F2+$\ln(1+x/b)$; (iii) F2+$\tan^{-1}(x/b)$; (iv) F3+$\ln(1+x/b)$; (v) F3+$(x/b)/(1+(x/b)^2)$; (vi) F4+$(x/b)/(1+(x/b)^2)$, and we denote them by G1, G2, G3, G4, G5 and G6, respectively. In most cases of 5-parameter MEIDs, the sign of the first two Lagrange multipliers are unchanged, and that associated with the additional moment constraint is either negative or positive depending on the null model. The exception is G3 for 2005. Due to the addition of the moment function, $\tan^{-1}(\cdot)$, to F2, the sign of the first moment function, $\ln x$, is changed from (-) to (+). Thus we can say that $\ln x$ turns to put more weights on lower income region than others. Among 5-parameter MEIDs, in terms of five all goodness-of-fit criterion, G1, G5 and G6 outperform GB2 for 2000, and all considered models dominate GB2 for 2005. We should note that the performance of G5 is superior for both years. The SSE, CSQ and CE of G5 for 2005 are, respectively, 0.000034, 68.6 and 0.000361, which are quite small compared with those of GB2, 0.000176, 382.97 and 0.001868. Therefore, G5 is our preferred model for the recent U.S. income data. The significance of the Lagrange multipliers in Table 5 can be corroborated by the likelihood ratio (LR) statistics for the null hypothesis of $H_0 : \lambda_3 = 0$, $LR = 2[\ln L(\hat{\theta}) - \ln L(\hat{\theta}_0)]$, where $\hat{\theta}$ and $\hat{\theta}_0$ are the maximum likelihood estimates for the 5- and 4-parameter models, and using the $lnL(\cdot)$ values from Tables 6 and 7.

[Table 6]

[Table 7]

The income share elasticities for Gi, $i = 1, 2, \cdots, 6$ can be calculated using (6) and are given by

$$
\begin{aligned}
\eta(x, G1) &= 1 - \lambda_3 - \frac{\lambda_1 x}{b\sqrt{1 + (x/b)^2}} - \frac{a\lambda_2 (x/b)^a}{1 + (x/b)^a}, \\
\eta(x, G2) &= 1 - \lambda_1 - \frac{\lambda_3 x}{b + x} - \frac{a\lambda_2 (x/b)^a}{\sqrt{1 + (x/b)^{2a}}}, \\
\eta(x, G3) &= 1 - \lambda_1 - \frac{b\lambda_3 x}{b^2 + x^2} - \frac{a\lambda_2 (x/b)^a}{\sqrt{1 + (x/b)^{2a}}}, \\
\eta(x, G4) &= 1 - \frac{\lambda_3 x}{b + x} - \frac{b\lambda_1 x}{b^2 + x^2} - \frac{a\lambda_2 (x/b)^a}{\sqrt{1 + (x/b)^{2a}}}, \\
\eta(x, G5) &= 1 - \frac{bx((\lambda_1 + \lambda_3)b^2 + (\lambda_1 - \lambda_3)x^2)}{(b^2 + x^2)^2} - \frac{a\lambda_2 (x/b)^a}{\sqrt{1 + (x/b)^{2a}}}, \\
\eta(x, G6) &= 1 - \frac{a\lambda_2 (x/b)^a}{1 + (x/b)^a} - \frac{bx((\lambda_3 + \lambda_1\sqrt{1 + (x/b)^2})b^2 + (-\lambda_3 - \lambda_1\sqrt{1 + (x/b)^2})x^2)}{(b^2 + x^2)^2},
\end{aligned}
$$

where λ_1 and λ_2 are Lagrange multipliers corresponding to the first and second moment functions (for the 4-parameter MEIDs in Table 2), and λ_3 is associated with additional moment function.

The values of $\lim_{x\to\infty} \eta(x, Gi)$ are given by $1 - \lambda_1 - a\lambda_2 - \lambda_3$, $1 - \lambda_1 - a\lambda_2$, $1 - a\lambda_2 - \lambda_3$ and $1 - a\lambda_2$ for $i = \{1, 2\}$, $i = \{3, 6\}$, $i = \{4\}$ and $i = \{5\}$, respectively. Thus if the values of $\lim_{x\to\infty} \eta(x, Gi)$ are negative, the WPL is readily satisfied. In the last rows of Tables 6 and 7, values of the α's are reported, and all the above models satisfy the WPL. Lastly, since signs of the Lagrange multipliers, particulary, for MEID having three moment functions, are not known before estimation process the direct check of decreasing income share elasticity through above equations is difficult. However, it can also be easily checked by calculating the first order derivative of the income share elasticity evaluated at maximum likelihood estimates:

$$-\sum_{j=1}^{k}\left[\hat{\lambda}_j\left(\frac{\partial\phi_j(x,\hat{\gamma})}{\partial x}+\frac{\partial\phi_j^2(x,\hat{\gamma})}{\partial^2 x}x\right)\right]<0.$$

We calculate the first order derivatives of six models at given income horizon and find that they are all negative.

In Figure 5 we plot the estimated densities for GB2, F4 and G5 in which graphs in the right panel represent the magnified version of the estimated densities in a limited range. F4 and G5 are chosen since they have the lowest values of SSE, SAE, CSQ and CE for both the years. We can observe that model F4 with the same number of parameters as in GB2 performs much better. However, for 2000 and 2005, GB2 and F4 are not flexible enough to explain the behavior of the middle (peaked) income region. As we can see clearly in Figure 5, by considering *one* more parameter, G5 can take care of the peaked behavior. Basically, the moment functions in F4, $\tan^{-1}(x/b)$ and $\sinh^{-1}(x/b)$ capture the shape of lower and upper regions of the household income distribution. The additional moment function $(x/b)/(1+(x/b)^2)$ can extract further useful information of the shape of lower household income distribution. Thus, our maximum entropy density construction approach based on an appropriate moment selection search is useful in modeling household income data.[10]

[Figure 5]

[10]Note that the specification of G5 may not always yield the best income density with other income data set, for example, the unit record income data. However, we should stress that one can follow our proposed MEID methodology to obtain a flexible density that would be very close to the "true" income density function given a set of data.

6 Concluding remarks

In this paper using maximum entropy principle, we have provided characterization of some well-known income density functions and suggested a class of flexible parametric densities that satisfy weak Pareto law and other stylized facts. We introduced some moment functions that would be appropriate to capture lower and upper income regions of distribution. To select suitable moment functions, a test based on the Rao score principle is proposed. One can consider our proposed score test as a pretest. It is quite well known that a pretest may affect the inference of selected model, for example, Guggenberger (2009). However, in this paper, we do not consider the inference after our moment selection test. Our empirical results demonstrate that our suggested maximum entropy income densities are quite useful in capturing the behavior of the underlying features of income data. Many other moment functions could be chosen to generate more flexible density. Since income inequality measures can be calculated accurately from parametric form of income distribution, finding appropriate density is quite important, and therefore, our models would be useful to obtain good estimates of income inequality measure and understand the true behavior of the income data.

Appendix

Table A1: U.S. Family Total Money Income Share

Income range (in thousands)	1970	1980	Income range (in thousands)	1990	2000	2005
0 - 2.49	6.6	2.1	0 - 4.99	3.57	2.17	2.70
2.5 - 4.99	12.5	4.1	5 - 9.99	5.84	2.85	2.61
5 - 7.49	15.2	6.2	10 - 14.99	7.50	4.53	3.73
7.5 - 9.99	16.6	6.5	15 - 19.99	7.89	5.64	4.79
10 - 12.49	15.8	7.3	20 - 24.99	8.47	5.83	5.24
12.5 - 14.99	11.0	6.9	25 - 29.99	8.00	6.08	5.31
15 - 19.99	13.1	14.0	30 - 34.99	8.15	5.95	5.40
20 - 24.99	4.6	13.7	35 - 39.99	7.57	5.60	4.93
25 - 34.99	3.0	19.8	40 - 44.99	6.61	5.32	4.97
35 - 49.99	1.1	12.8	45 - 49.99	5.87	4.99	4.71
50 - ∞	0.5	6.7	50 - 54.99	5.13	5.00	4.58
			55 - 59.99	4.14	4.31	4.07
			60 - 64.99	3.62	4.63	4.32
			65 - 69.99	2.95	3.85	3.78
			70 - 74.99	2.38	3.68	3.62
			75 - 79.99	2.09	3.24	3.22
			80 - 84.99	1.64	2.91	3.05
			85 - 89.99	1.33	2.45	2.58
			90 - 94.99	1.02	2.28	2.56
			95 - 99.99	0.79	1.71	2.11
			100 - ∞	5.44	16.97	21.81
Obs (in thousands)	52227	60309		66322	72388	77418
Mean (in thousands)	11.106	23.974		43.652	65.570	73.300
Gini	0.354	0.365		0.395	0.415	0.414

Notes: Source: Current Population Survey. Obs denotes the number of family as of March of the following year. Mean and Gini represent Census estimates for mean and Gini index. The data are obtained by multiplying the reported figures by (100/obs), i.e., n_k/N for each k-th income group.

References

Aigner, D. J., and A. S. Goldberger (1970): "Estimation of Pareto's law from grouped Observation," *Journal of the American Statistical Association*, 65, 712–723.

Aitchison, J., and J. Brown (1957): *The Lognormal Distribution*. Cambridge University Press, Cambridge.

Berger, J. (1985): *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.

Boccanfuso, D., B. Decaluwé, and L. Savard (2008): "Poverty, income distribution and CGE micro-simulation modeling: Does the functional form of distribution matter?," *Journal of Economic Inequality*, 6, 149–184.

Bordley, R. F., J. B. McDonald, and A. Mantrala (1996): "Something new, something old: Parametric models for the size distribution of income," *Journal of Income Distribution*, 6, 91–103.

Clementi, F., M. Gallegati, and G. Kaniadakis (2010): "A model of personal income distribution with application to Italian data," *Empirical Economics*, 39, 559–591.

Cobb, L., P. Koppstein, and N. H. Chen (1983): "Estimation and moment recursion relations for multimodal distributions of exponential family," *Journal of the American Statistical Association*, 78, 124–130.

Cressie, N., and T. Read (1984): "Multinomial goodness-of-fit tests," *Journal of the Royal Statistical Society: Ser B*, 46, 440–464.

Dagum, C. (1977): "A new model of personal income distribution: Specification and estimation," *Economie Appliquée*, 30, 413–437.

Dastrup, S. R., R. Hartshorn, and J. B. McDonald (2007): "The impact of taxes and transfer payments on the distribution of income: A parametric comparison," *Journal of Economic Inequality*, 5, 353–369.

Dhaene, G., and D. Hoorelbeke (2004): "The information matrix test with bootstrap-based covariance matrix estimation," *Economics Letters*, 82, 341–147.

Esteban, J. M. (1986): "Income-share elaticity and the size distribution of income," *International Economic Review*, 27, 439–444.

Gastwirth, J. (1972): "The estimation of the Lorenz curve and Gini index," *Review of Economics and Statistics*, 54, 306–316.

Gokhale, D. (1975): "Maximum entropy characterization of some distributions," *Statistical Distribution in Scientific Work*, 3, 299–304.

Guggenberger, P. (2009): "The impact of a Hausman pretest on the asymptotic size of a hypothesis test," *Econometric Theory*, 26, 369–382.

Haavelmo, T. (1944): "The probability approach in Econometrics," *Econometrica*, 12, Supplement, iii–vi+1–115.

Jhun, M., and H. Jeong (2000): "Applications of bootstrap methods for categorical data analysis," *Computational Statistics & Data Analysis*, 35, 83–91.

Kagan, A., Y. Linik, and C. Rao (1973): *Characterization Problems in Mathematical Statistics*. Wiley, New York.

Kapur, J., and H. Kesavan (1992): *Entropy Optimization Principles with Applications*. Academic Press.

Kleiber, C., and S. Kotz (2003): *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley, Hoboken, NJ.

Kloek, T., and H. van Dijk (1977): "Further results on efficient estimation of income distribution parameters," *Economie Appliquée*, 30, 439–459.

Kloek, T., and H. Van Dijk (1978): "Efficient estimation of income distribution parameters," *Journal of Econometrics*, 8, 74–81.

Kullback, S., and R. Leibler (1951): "On information and sufficiency," *Annals of Mathematical Statistics*, 22, 79–86.

Leipnik, R. B. (1990): "A maximum relative entropy principle for distribution of personal income with derivations of several known income distributions," *Communications in Statistics-Theory and Methods*, 19, 1003–1036.

Lindsay, B. (1994): "Efficiency versus robustness: The case for minimum Hellinger distance and related methods," *Annals of Statistics*, 22, 1081–1104.

Majunmder, A., and S. Chakravarty (1990): "Distribution of personal income: Development of a new model and its application to U.S. income data," *Journal of Applied Econometrics*, 5, 189–196.

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

McDonald, J., and A. Mantrala (1995): "The distribution of income: Revisited," *Journal of Applied Econometrics*, 10, 201–204.

McDonald, J., and M. Ransom (1979): "Functional forms, estimation techniques, and the distribution of income," *Econometrica*, 47, 1513–1526.

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

Minoiu, C., and S. G. Reddy (2014): "Kernel density estimation on grouped data: the case of poverty assessment," *Journal of Economic Inequality*, 12(2), 163–189.

Morales, D., L. Pardo, and L. Santamaría (2004): “Bootstrap confidence regions in multinomial sampling,” *Applied Mathematics and Computation*, 155, 295–315.

Ord, J. K., G. P. Patil, and C. Taillie (1981): “Choice of a distribution to decribe personal income,” *Statistical Distribution in Scientific Work*, 6, 193–201.

Pareto, V. (1895): “La legge della domanda,” *Giornale degli Economisti*, 10, 59–68, English translation in *Rivista di Politica Economica*, 87 (1997), 691-700.

——— (1896): “La courbe de la répartition de la richesse,” Reprinted in G.Busoni(ed.): *Eeuvres complètes de Vilfredo Pareto , Tom 3: Écrits sur la courbe de la répartition de la richesse*, Geneva: Librairie Doroz. English translation in *Rivista di Politica Economica*, 87 (1997), 647-700.

——— (1897): “Aggiunta allo studio della curva della entrate,” *Giornale degli Economisti*, 14, 15–26, English translation in *Rivista di Politica Economica*, 87 (1997), 645-700.

Park, S. Y., and A. K. Bera (2009): “Maximum entropy autoregressive conditional heteroskedasticity model,” *Journal of Econometrics*, 150, 219–230.

——— (2013): “Information theoretic approaches to density estimation with an application to the U.S. personal income data,” *Working Paper*, University of Illinois.

Ransom, M., and J. Cramer (1983): “Income distribution with disturbances,” *European Economic Review*, 22, 363–372.

Renyi, A. (1960): “On measures of entropy and information,” *Proceeding of the fourth berkeley Symposium on Mathematics*, 1, 547–561.

Ryu, H., and D. Slottje (1997): “Another perspective on recent changes in the U.S. income distribution: An index space representation,” *Advances in Econometrics*, 12, 319–340.

Salem, A., and T. Mount (1974): "A convenient descriptive model of income distribution: The Gamma density," *Econometrica*, 42, 1115–1127.

Shannon, C. (1948): *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, Illinois.

Singh, S., and G. Maddala (1976): "A function for the size distribution of income," *Econometrica*, 44, 963–970.

Wu, X. (2003): "Calculation of maximum entropy densities with application to income distribution," *Journal of Econometrics*, 115, 347–354.

Wu, X., and J. M. Perloff (2005): "China's income distribution, 1985-2001," *Review of Economics and Statistics*, 87, 763–775.

Wu, X., and J. M. Perloff (2007): "GMM estimation of a maximum entropy distribution with interval data," *Journal of Econometrics*, 138, 532–546.

Zellner, A. (1977): "Maximal data information prior distribution," *In: New Methods in the Applications of Bayesian Methods, North-Holland*, pp. 117–132.

Table 1: Maximum entropy characterization of some common income densities

Type	Side constraints	Resulting density, $f(x) =$	Common form, $f(x) =$	
LN	$\int \ln x f(x) dx = \mu$ $\int (\ln x - \mu)^2 f(x) dx = \sigma^2$	$\exp\left[-\lambda_0 - \lambda_1 \ln x - \lambda_2 (\ln x)^2\right]$	$\frac{1}{\sigma\sqrt{2\pi x}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right]$,	$0 < x < \infty$ $\mu > 0, \quad \sigma > 0$
Gamma	$\int x f(x) dx = a$ $\int \ln x f(x) dx = \psi(a)$	$\exp\left[-\lambda_0 - \lambda_1 x - \lambda_2 \ln x\right]$	$\frac{1}{\Gamma(a)} \exp^{-x} x^{a-1}$,	$0 < x < \infty$ $a > 0$
Beta	$\int \ln x f(x) dx = \psi(a) - \psi(a+b)$ $\int \ln(1-x) f(x) dx = \psi(b) - \psi(a+b)$	$\exp\left[-\lambda_0 - \lambda_1 \ln x - \lambda_2 \ln(1-x)\right]$	$\frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$,	$0 < x < 1$ $a > 0, \quad b > 0$
Weibull	$\int \ln x f(x) dx = -\gamma/a$ $\int x^a f(x) dx = 1$	$\exp\left[-\lambda_0 - \lambda_1 x^a - \lambda_2 \ln x\right]$	$a x^{a-1} \exp\left[-x^a\right]$,	$0 < x < \infty$ $a > 0$
Fisk	$\int \ln x f(x) dx = -\frac{\ln(b^{-a})}{a}$ $\int \ln(1 + (x/b)^a) f(x) dx = 1$	$\exp\left[-\lambda_0 - \lambda_1 \ln x - \lambda_2 \ln(1 + (x/b)^a)\right]$	$\frac{a x^{a-1}}{b^a (1 + (x/b)^a)^2}$,	$0 < x < \infty$ $a > 0, \quad b > 0$
GG	$\int \ln x f(x) dx = -\frac{\ln(b^{-a}) - \psi(p)}{a}$ $\int x^a f(x) dx = \frac{\Gamma(1+p)}{\Gamma(p)} b^a$	$\exp\left[-\lambda_0 - \lambda_1 \ln x - \lambda_2 x^a\right]$	$\frac{a x^{ap-1} \exp[-(x/b)^a]}{b^{ap}\Gamma(p)}$,	$0 < x < \infty$ $a, b, p > 0$
SM	$\int \ln x f(x) dx = -\frac{\gamma + \ln(b^{-a}) + \psi(q)}{a}$ $\int \ln(1 + (x/b)^a) f(x) dx = \frac{1}{q}$	$\exp\left[-\lambda_0 - \lambda_1 \ln x - \lambda_2 \ln(1 + (x/b)^a)\right]$	$\frac{a q x^{a-1}}{b^a (1 + (x/b)^a)^{1+q}}$,	$0 < x < \infty$ $a, b, q > 0$
Dagum	$\int \ln x f(x) dx = -\frac{-\gamma + \ln(b^{-a}) - \psi(p)}{a}$ $\int \ln(1 + (x/b)^a) f(x) dx = \gamma + \frac{1}{p} + \psi(p)$	$\exp\left[-\lambda_0 - \lambda_1 \ln x - \lambda_2 \ln(1 + (x/b)^a)\right]$	$\frac{a p x^{ap-1}}{b^{ap} (1 + (x/b)^a)^{p+1}}$,	$0 < x < \infty$ $a, b, p > 0$
GB2	$\int \ln x f(x) dx = -\frac{\ln(b^{-a}) - \psi(p) + \psi(q)}{a}$ $\int \ln(1 + (x/b)^a) f(x) dx = \psi(p+q) - \psi(q)$	$\exp\left[-\lambda_0 - \lambda_1 \ln x - \lambda_2 \ln(1 + (x/b)^a)\right]$	$\frac{a x^{ap-1}}{b^{ap} B(p,q) (1 + (x/b)^a)^{p+q}}$,	$0 < x < \infty$ $a, b, p, q > 0$

Notes: $\psi(a) = \Gamma'(a)/\Gamma(a)$ denotes digamma function, and γ denotes Euler's constant.

Table 2: Combination of moment functions corresponding to different models along with the corresponding income share elasticity functions and their limits

Model	First moment function	Second moment function	Income share elasticity function ($\eta(x, f)$)	$\lim_{x\to\infty} \eta(x, f)$
		Three-parameter MEIDs		
T1	$\ln x$	$\ln(1 + (x/b)^2)$	$1 - \lambda_1 \frac{2\lambda_2 x^2}{b^2+x^2}$	$1 - \lambda_1 - 2\lambda_2$
T2	$\ln x$	$\sinh^{-1}((x/b)^2)$	$1 - \lambda_1 - \frac{2\lambda_2 x^2}{b^2\sqrt{1+(x/b)^4}}$	$1 - \lambda_1 - 2\lambda_2$
T3	$\ln(1 + x/b)$	$\ln(1 + (x/b)^2)$	$1 - \frac{\lambda_1 x}{b+x} - \frac{2\lambda_2 x^2}{b^2+x^2}$	$1 - \lambda_1 - 2\lambda_2$
T4	$\ln(1 + x/b)$	$\sinh^{-1}((x/b)^2)$	$1 - \frac{\lambda_1 x}{b+x} - \frac{2\lambda_2 x^2}{b^2\sqrt{1+(x/b)^4}}$	$1 - \lambda_1 - 2\lambda_2$
T5	$\tan^{-1}(x/b)$	$\ln(1 + (x/b)^2)$	$\frac{b^2 - b\lambda_1 x + x^2 - 2\lambda_2 x^2}{b^2+x^2}$	$1 - 2\lambda_2$
T6	$\sinh^{-1}(x/b)$	$\sinh^{-1}((x/b)^2)$	$1 - \frac{\lambda_1 x}{b\sqrt{1+(x/b)^2}} - \frac{2\lambda_2 x^2}{b^2\sqrt{1+(x/b)^4}}$	$1 - \lambda_1 - 2\lambda_2$
		Four-parameter MEIDs		
F1	$\ln x$	$\ln(1 + (x/b)^a)$	$1 - \lambda_1 - \frac{a\lambda_2 (x/b)^a}{1+(x/b)^a}$	$1 - \lambda_1 - a\lambda_2$
F2	$\ln x$	$\sinh^{-1}((x/b)^a)$	$1 - \lambda_1 - \frac{a\lambda_2 (x/b)^a}{\sqrt{1+(x/b)^{2a}}}$	$1 - \lambda_1 - a\lambda_2$
F3	$\tan^{-1}(x/b)$	$\sinh^{-1}((x/b)^a)$	$1 - \frac{b\lambda_1 x}{b^2+x^2} - \frac{a\lambda_2 (x/b)^a}{\sqrt{1+(x/b)^{2a}}}$	$1 - a\lambda_2$
F4	$\sinh^{-1}(x/b)$	$\ln(1 + (x/b)^a)$	$1 - \frac{\lambda_1 x}{b\sqrt{1+(x/b)^2}} - \frac{a\lambda_2 (x/b)^a}{1+(x/b)^a}$	$1 - \lambda_1 - a\lambda_2$

Table 3: Estimated density functions (common and 3-parameter MEIDs) : Year 2000

	LN	GG	SM	Dagum	GB2	T1	T2	T3	T4	T5	T6
a (λ_1)	3.8570	1.1221	1.6071	2.7497	1.9285	-0.5344	-0.4931	-4.7904	-3.9845	-3.2408	-3.3963
b (λ_2)	0.7870	43.2993	139.3646	73.1575	101.6993	2.7649	2.0055	3.8483	2.9239	2.1813	3.0888
p (b)		1.5185		0.5266	0.7990	95.4776	92.9525	65.9555	78.8410	45.3819	74.3686
q (a)			3.8524		2.2441						
SSE	0.002177	0.000170	0.000142	0.000186	0.000144	0.000145	0.000164	0.000254	0.000393	0.000247	0.000424
SAE	0.168034	0.045510	0.042151	0.049731	0.042545	0.042345	0.045757	0.063058	0.076721	0.062715	0.079821
CSQ	5524.01	364.51	299.23	312.47	279.98	278.65	288.46	414.46	657.87	400.90	710.61
CE	0.023040	0.002785	0.002260	0.001956	0.002107	0.002019	0.001978	0.003063	0.004698	0.002646	0.005185
lnL	-211561.37	-209690.00	-209655.36	-209690.32	-209648.39	-209644.99	-209650.01	-209708.97	-209832.42	-209701.78	-209859.09
Mean	69.7673	61.0268	62.4039	65.3483	63.1558	63.4936	65.8757	66.6061	70.8343	65.4424	69.4452
Gini	0.4670	0.3855	0.3982	0.4225	0.4046	0.4075	0.4268	0.4330	0.4641	0.4236	0.4543
α						3.9954	2.5179	1.9062	0.8633	3.3626	1.7813

Notes: Census estimates for mean and Gini are 65.594 and 0.415, respectively. λ_1 and λ_2 are Lagrange multipliers correspond to first and second moment constraints. Parameters in the parenthesis, λ_1, λ_2, b and a in order, denote those for models Ts. α denotes the constant for WPL in (5).

Table 4: Estimated density functions (common and 3-parameter MEIDs) : Year 2005

	LN	GG	SM	Dagum	GB2	T1	T2	T3	T4	T5	T6
a (λ_1)	3.9575	1.3454	1.4610	2.7953	2.1586	-0.3863	-0.3674	-3.9803	-3.3578	-2.7716	-2.8763
b (λ_2)	0.9103	69.2895	319.6971	86.9740	112.1601	2.8887	1.9369	3.4810	2.6270	2.1326	2.7710
p (b)		1.0727		0.4711	0.6315	121.2684	109.0262	76.5951	87.1563	55.5404	82.4953
q (a)			9.0846		1.8520						
SSE	0.002513	0.000221	0.000197	0.000186	0.000176	0.000174	0.000174	0.000149	0.000219	0.000146	0.000235
SAE	0.185251	0.049661	0.043920	0.048173	0.044313	0.042475	0.043280	0.044935	0.053188	0.044731	0.055720
CSQ	7277.19	492.98	472.29	374.86	383.97	394.95	383.03	285.98	418.07	279.98	448.36
CE	0.026877	0.002702	0.002759	0.001386	0.001868	0.001999	0.002247	0.001297	0.002052	0.001159	0.002037
lnL	-222515.86	-220024.27	-220009.82	-220003.47	-219974.63	-219973.02	-219967.46	-219915.31	-219981.04	-219912.11	-219995.99
Mean	81.1471	67.3306	68.3355	72.5946	70.5013	70.3297	73.5717	74.7726	80.3687	73.5611	79.0753
Gini	0.4928	0.3901	0.3985	0.4302	0.4151	0.4140	0.4373	0.4442	0.4788	0.4358	0.4711
α						4.3911	2.5064	1.9817	0.8962	3.4936	1.6657

Notes: Census estimates for mean and Gini are 73.304 and 0.414, respectively. λ_1 and λ_2 are Lagrange multipliers correspond to first and second moment constraints. Parameters in the parenthesis, λ_1, λ_2, b and a in order, denote those for models Ts. α denotes the constant for WPL in (5).

Table 5: Moment function selection test results

Null model	Moment function	2000		2005	
(i) F1	$(x/b)/(1+(x/b)^2)$	0.484	(0.487)	1.345	(0.246)
	$\tan^{-1}(x/b)$	0.000	(0.988)	0.324	(0.569)
	$\sinh^{-1}(x/b)$	0.430	(0.512)	0.000	(0.998)
	$\tan^{-1}((x/b)^2)$	0.834	(0.361)	1.664	(0.197)
	$\sinh^{-1}((x/b)^2)$	0.002	(0.965)	0.011	(0.918)
(ii) F2	$\ln(1+x/b)$	4.563*	(0.033)	4.104*	(0.043)
	$(x/b)/(1+(x/b)^2)$	0.045	(0.833)	0.788	(0.375)
	$\tan^{-1}(x/b)$	9.025**	(0.003)	9.763**	(0.002)
	$\ln(1+(x/b)^2)$	0.013	(0.908)	0.197	(0.657)
	$\tan^{-1}((x/b)^2)$	0.012	(0.914)	0.000	(0.996)
(iii) F3	$\ln(x)$	487.164**	(0.000)	276.836**	(0.000)
	$\ln(1+x/b)$	5.916*	(0.015)	1.525	(0.217)
	$(x/b)/(1+(x/b)^2)$	8.630**	(0.003)	4.366*	(0.037)
	$\ln(1+(x/b)^2)$	1.473	(0.225)	0.758	(0.384)
	$\tan^{-1}((x/b)^2)$	0.288	(0.592)	0.108	(0.743)
(iv) F4	$\ln(x)$	39.762**	(0.000)	144.389**	(0.000)
	$(x/b)/(1+(x/b)^2)$	3.814	(0.051)	3.403	(0.065)
	$\tan^{-1}(x/b)$	2.207	(0.137)	1.330	(0.249)
	$\tan^{-1}((x/b)^2)$	3.049	(0.081)	2.922	(0.087)
	$\sinh^{-1}((x/b)^2)$	0.678	(0.411)	0.645	(0.422)

Notes: (i)-(iv) denote null density corresponds to the moment functions in Table 3. The test statistics are calculated with the bootstrap sample size 200. P-values are calculated using asymptotic χ^2_1 distribution and given in the parentheses. 5% and 1% critical values of Hotelling's $T^2_{1,99}$ are 3.937 and 6.898, respectively. * and ** indicate statistical significance at the 5% and 1% levels, respectively.

Table 6: Estimated 4- and 5-parameter MEIDs : year 2000

	F1	F2	F3	F4
λ_1	-0.5237	-0.5387	-5.8426	-10.3641
λ_2	2.5498	2.5074	3.9445	11.3087
b	91.2540	104.0125	43.3102	46.6547
a	2.0761	1.7535	1.2149	1.3966
SSE	0.000147	0.000149	0.000229	0.000139
SAE	0.042484	0.042847	0.060171	0.042748
CSQ	278.20	279.14	373.32	239.79
CE	0.002227	0.002041	0.002751	0.001883
lnL	-209644.74	-209645.38	-209688.19	-209623.32
Mean	63.7159	64.8072	65.0011	63.6833
Gini	0.4094	0.4186	0.4200	0.4088
α	3.7699	2.8580	3.7922	4.4296

	G1	G2	G3	G4	G5	G6
λ_1	-9.8981	-0.3861	-0.4124	-6.1732	-7.8673	-11.5359
λ_2	11.4364	11.4093	5.0201	10.7979	1.5439	9.6802
λ_3	-0.1157	-11.2629	-4.5249	-7.9702	7.5917	4.2798
b	53.5454	127.8902	74.5703	61.0381	19.9022	45.1606
a	1.3878	1.1261	1.2037	1.0861	3.3913	1.7861
SSE	0.000129	0.000146	0.000149	0.000215	0.000077	0.000102
SAE	0.040998	0.042809	0.043167	0.057563	0.032440	0.036884
CSQ	232.50	263.76	271.04	351.48	138.00	191.47
CE	0.001860	0.002002	0.002154	0.002635	0.000959	0.001449
lnL	-209620.29	-209636.89	-209640.67	-209677.83	-209571.26	-209599.64
Mean	63.3096	65.2761	63.6397	65.4972	64.1612	63.1142
Gini	0.4056	0.4225	0.4084	0.4240	0.4134	0.4040
α	4.8576	0.1990	4.6303	2.7574	4.2358	4.7539

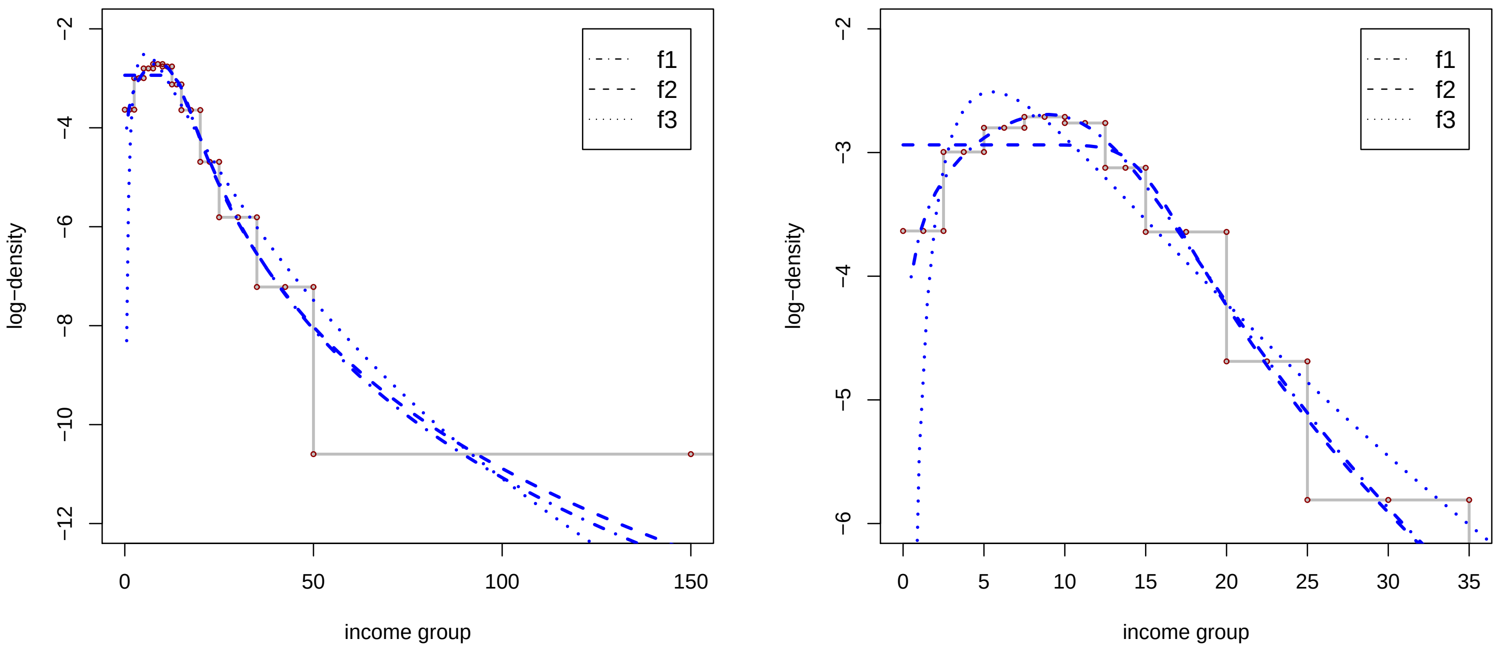
Notes: G1: F4+$\ln(x)$; G2: F2+$\ln(1+x/b)$; G3: F2+$\tan^{-1}(x/b)$; G4: F3+$\ln(1+x/b)$; G5: F3+$(x/b)/(1+(x/b)^2)$; G6: F4+$(x/b)/(1+(x/b)^2)$. Census estimates for mean and Gini are 65.594 and 0.415, respectively. λ_1, λ_2 and λ_3 are Lagrange multipliers correspond to first, second and additional moment constraints selected from the test, respectively. α denotes the constant for WPL in (5).

Table 7: Estimated 4- and 5-parameter MEIDs : year 2005

	F1	F2	F3	F4
λ_1	-0.3371	-0.3450	-5.4562	-7.0794
λ_2	1.5553	1.6257	3.9739	8.1248
b	86.8065	99.8908	52.3811	70.2124
a	2.6478	2.2165	1.1925	1.5192
SSE	0.000172	0.000177	0.000140	0.000115
SAE	0.044647	0.044672	0.044128	0.039814
CSQ	367.26	377.42	272.61	238.32
CE	0.001747	0.001812	0.001130	0.000900
lnL	-219959.58	-219964.94	-219908.65	-219891.80
Mean	73.1060	74.8709	72.9577	71.4000
Gini	0.4337	0.4457	0.4316	0.4206
α	2.7810	2.2584	3.7388	4.2637
	G1	G3	G5	
λ_1	-9.8760	0.0277	-7.6171	
λ_2	9.8924	3.8856	1.4376	
λ_3	0.2544	-5.4944	8.0585	
b	47.6501	50.8447	22.5469	
a	1.4504	1.1980	3.6497	
SSE	0.000108	0.000142	0.000034	
SAE	0.038692	0.044140	0.020928	
CSQ	206.18	272.62	68.60	
CE	0.000659	0.001221	0.000361	
lnL	-219874.42	-219908.36	-219804.79	
Mean	72.3757	73.1101	71.6525	
Gini	0.4275	0.4327	0.4226	
α	3.7263	3.6826	4.2468	

Notes: G1: F4+$\ln(x)$; G3: F2+$\tan^{-1}(x/b)$; G5: F3+$(x/b)/(1+(x/b)^2)$. Census estimates for mean and Gini are 73.304 and 0.414, respectively. λ_1, λ_2 and λ_3 are Lagrange multipliers correspond to first, second and additional moment constraints selected from the test, respectively. α denotes the constant for WPL in (5).

Figure 1: Log-empirical probability and log-density estimates



Notes: Left: Log-empirical probability (solid step function) and three log-density estimates using 1970 U.S. grouped income data which are in 11-group format. $f1$ is the MED with two moment functions, namely, $\log(x)$ and $\log(1+(x/b)^a)$; $f2$ uses only one moment function $\log(1+(x/b)^a)$, while $f3$ is based $\log(x)$ and $\log(x)^2$. Right: Enlarged version of left graph for a shorter, (0,35) income range.

Figure 2: Moment functions related to income distribution

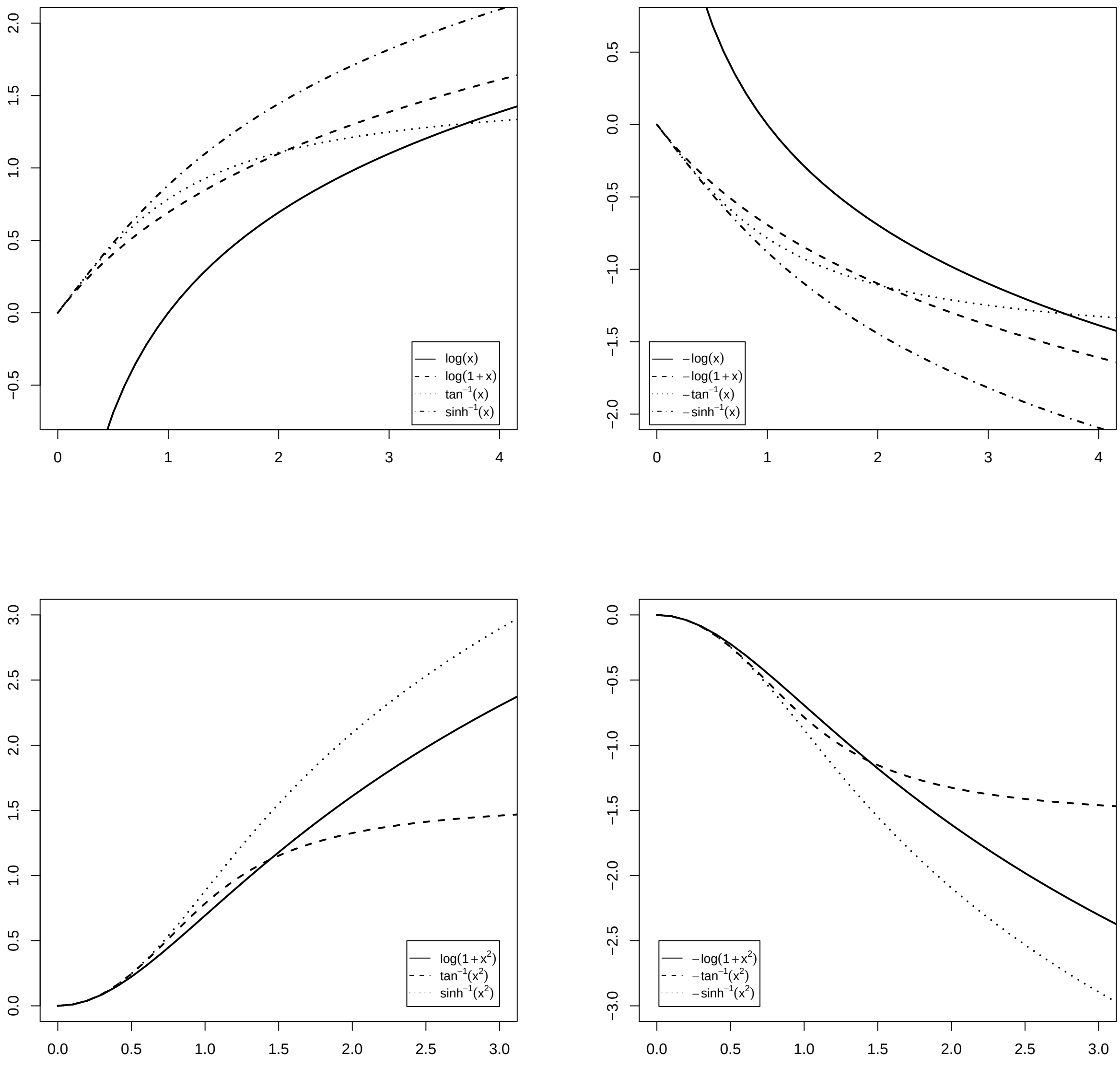


Figure 3: Moment functions: $\ln(1 + x^a)$ and $\sinh^{-1}(x^a)$

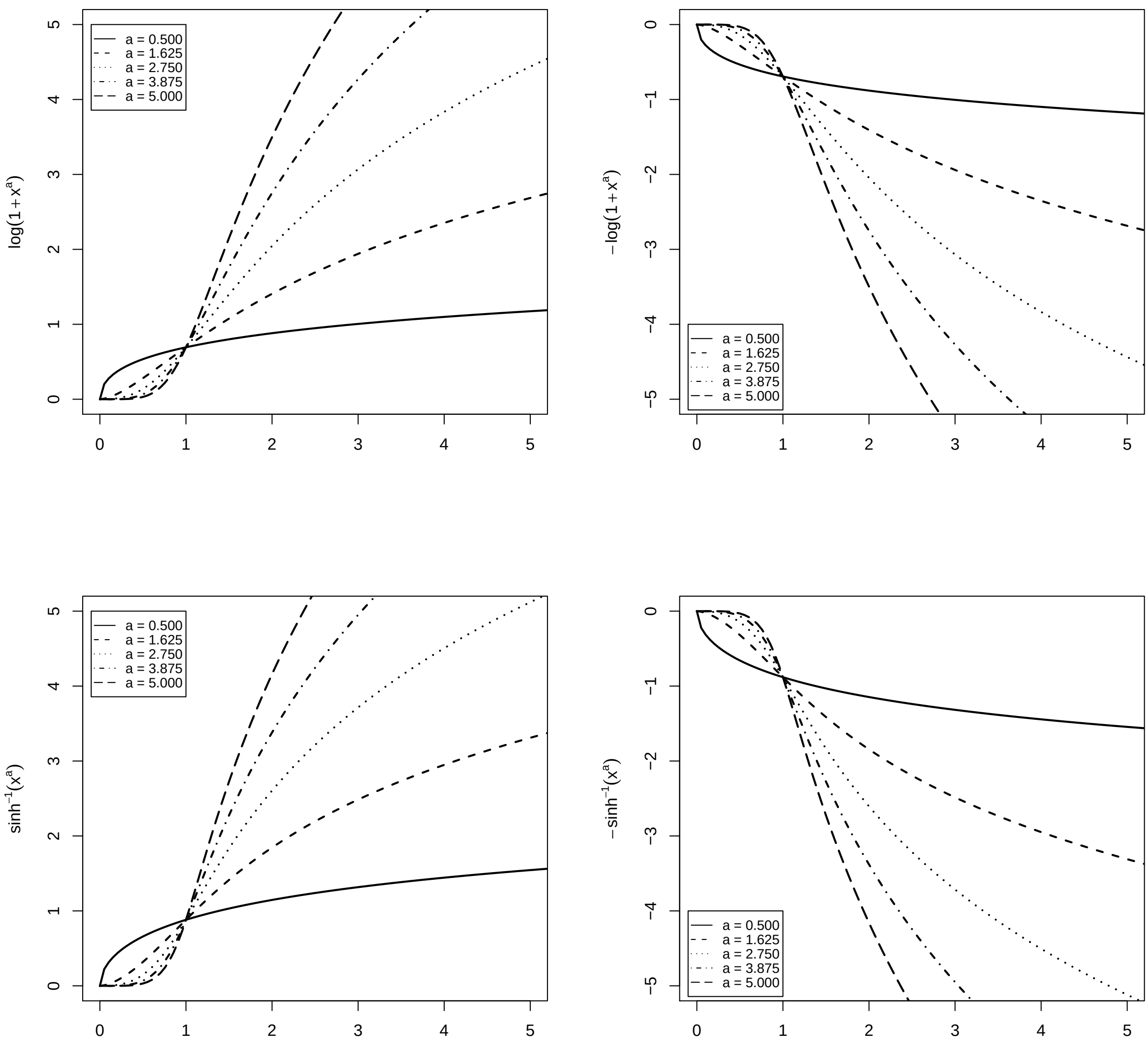


Figure 4: Density estimates for the US personal income for 2000 and 2005