the review of income and wealth

Review of Income and Wealth
Series 60, Number 4, December 2014
DOI: 10.1111/roiw.12047

PARAMETRIC LORENZ CURVES AND THE MODALITY OF THE INCOME DENSITY FUNCTION

by Melanie Krause*

Goethe University Frankfurt

Similar looking Lorenz curves can imply very different income density functions and potentially lead to wrong policy implications regarding inequality. This paper derives a relation between a Lorenz curve and the modality of its underlying income density: given a parametric Lorenz curve, it is the sign of its third derivative which indicates whether the density is unimodal or zeromodal (i.e., downward-sloping). The density modality of several important Lorenz curves such as the Pareto, Weibull, Singh–Maddala parametrizations and hierarchical families of Lorenz curves are discussed. A Lorenz curve performance comparison with Monte Carlo simulations and data from the UNU–WIDER World Income Inequality Database underlines the relevance of the theoretical result: curve-fitting based on criteria such as mean squared error or the Gini difference might lead to a Lorenz curve implying an incorrectly-shaped density function. It is therefore important to take into account the modality when selecting a parametric Lorenz curve.

JEL Codes: C13, C16, D31

Keywords: goodness of fit, income distribution, inequality, Lorenz curve, modality

1. Introduction

In the macroeconomic field of growth and development it is of interest to analyze not only a country's total GDP but also how it is distributed across the population. Inequality is often expressed with a Lorenz curve (LC) $L(\pi)$, developed by Lorenz (1905), which links the cumulative income share L to the cumulative share π of people in the population earning an income up to that level. Very similar looking LCs can, however, imply starkly different shapes of the underlying income density $f(x)$. This insight is the main motivation for this paper.

Consider the two LCs in Figure 1. They are of the Singh–Maddala biparametric form, with different parameters, but they have the same Gini coefficient and indeed look very similar. Now turn to the underlying income density functions in Figure 2: their shapes are very different. It is particularly striking that LC1 implies a downward-sloping, zeromodal density, whereas LC2 has a unimodal one. A zeromodal density shows a higher proportion of relatively poor earners and no real middle class, unlike a unimodal one. This can have important policy implications.

Note: The author is grateful to the editor and to two anonymous referees for their helpful comments. In addition, this paper has benefited from seminar participants' suggestions at Goethe University, the RGSE conference in Duisburg, the IAES conference in Istanbul, and the meeting of the Verein für Socialpolitik in Göttingen.

*Correspondence to: Melanie Krause, Goethe University Frankfurt, Faculty of Economics and Business Administration, Chair for International Macroeconomics & Macroeconometrics, Grueneburgplatz 1, House of Finance, 60323 Frankfurt am Main, Germany (melanie.krause@wiwi.uni-frankfurt.de).

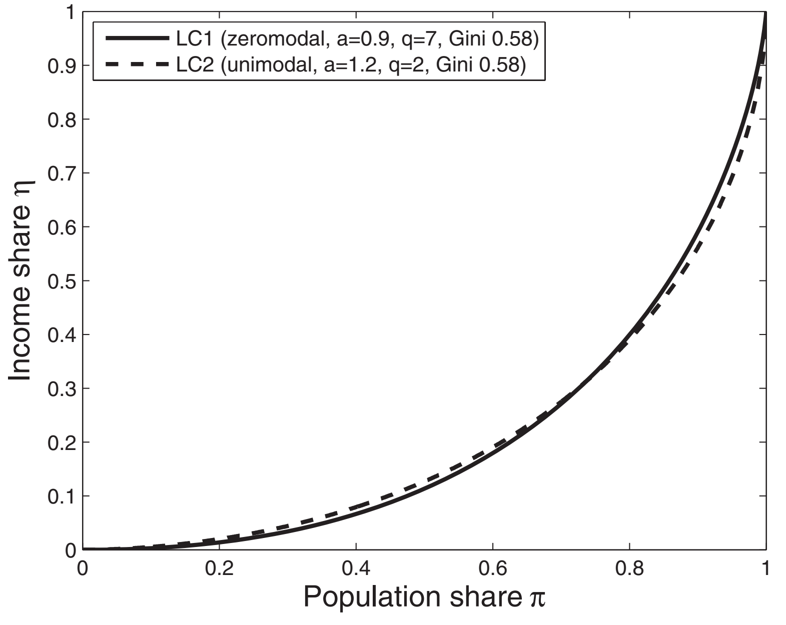

Figure 1. Two Examples of Singh–Maddala LCs with the Same Gini Coefficient of 0.58

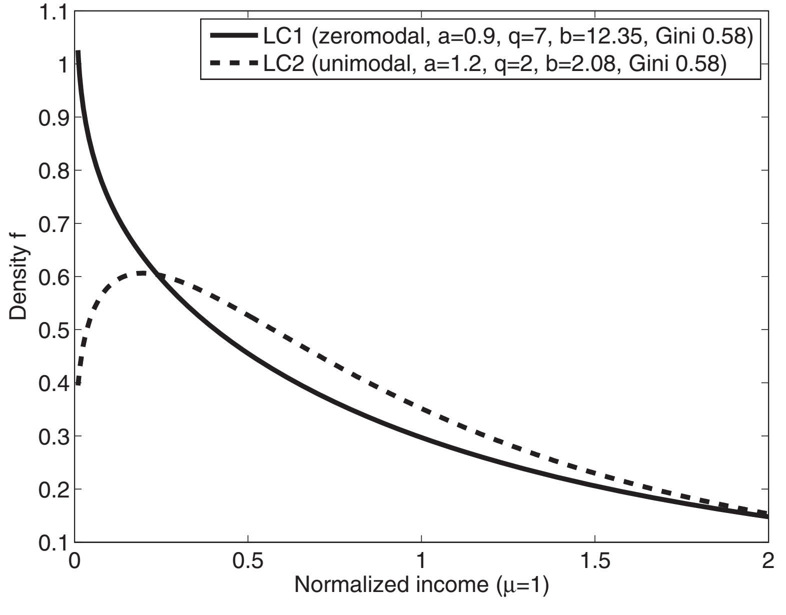

Figure 2. Densities for the Two Singh–Maddala LCs from Figure 1 with the Same Gini Coefficient of 0.58

Also it was pointed out by Dagum (1999) and Kleiber (2008) that wealth densities tend to be zeromodal, whereas income densities are mostly unimodal unless a country is very poor and overpopulated. So how can a researcher working with the two similar LCs from Figure 1 and the same Gini coefficient tell which of them has a unimodal and which a zeromodal density?

This paper presents a link between an LC and the shape of its density function: it argues that the third derivative of an LC indicates the modality of its underlying density. So it is not necessary to derive the density explicitly (which can be tedious). The theoretical result is derived in Section 2. Sections 3 and 4 discuss the density modality of a number of important uniparametric and multiparametric LCs. In Section 5 a Monte Carlo simulation illustrates the practical relevance of the density modality for a researcher fitting LCs to data points: curve-fitting based on mean squared error or Gini difference often leads to an LC with the wrong density modality. Section 6 presents a brief empirical application with a worldwide dataset and Section 7 concludes. Some technical derivations have been relegated to the Appendix.

2. Relation between the Lorenz Curve and the Modality of its Density

Following Pakes (1981) an LC $L(\pi)$ can be defined formally by:

Definition 1. *A function* $L(\pi)$*, continuous on* $[0,1]$ *and with second derivative* $L''(\pi)$*, is a Lorenz curve (LC) if and only if*

$$L(0)=0,\quad L(1)=1,\quad L'(0^+)\geq 0,\quad L''(\pi)\geq 0\quad in\,(0,1). \tag{1}$$

A density $f(x)$ with mean μ, cumulative distribution $\pi = F(x)$ and its inverse function $x = F^{-1}(\pi)$ is related to the LC $L(\pi)$ by the following formula (see Gastwirth, 1971; Kakwani, 1980):

$$F^{-1}(\pi) = L'(\pi)\mu. \tag{2}$$

Hence, starting with a given income density with $f(x)$ and $\pi = F(x)$, one can obtain the LC as follows: find the inverse cumulative distribution function $x = F^{-1}(\pi)$ and divide it by μ to obtain the slope $L'(\pi)$ of the LC. Integrating with respect to π and making sure that the properties from Definition (1) are fulfilled gives $L(\pi)$.

Let us now formally define the modality of the density function $f(x)$ in terms of sign changes in its first derivative $f'(x)$:

Definition 2. *The modality of a continuously differentiable density* $f(x)$ *on* $[x_L, x_U]$ *is the number of local maxima* $\tilde{x}\in(x_L,x_U)$ *where*

$$\begin{cases} f'(\tilde{x})=0, \\ f'(x)>0 \quad \forall x\in\tilde{X}_N \wedge x<\tilde{x}, \\ f'(x)<0 \quad \forall x\in\tilde{X}_N \wedge x>\tilde{x}, \end{cases} \tag{3}$$

with $\tilde{X}_N$ denoting the neighborhood of the point $\tilde{x}$.

In the following we will make use of a result by Arnold (1987), who differentiates (2) with respect to x and summarizes the relation in his theorem:

Theorem 1. *If for the Lorenz curve $L(\pi)$ the second derivative $L''(\pi)$ exists and is positive in an interval (x_1, x_2), then $F(x)$ has a finite positive density in the interval $(x_L, x_U) = (\mu L'(F(x_1^+)), \mu L'(F(x_2^-)))$ which is given by*

(4) $$f(x) = \frac{1}{\mu L''(F(x))}.$$

Proof. See Arnold (1987). ■

A key contribution of this paper is the following theorem on the relation between the LC and the modality of its density:

Theorem 2. *If the LC $L(\pi)$ has a third derivative $L'''(\pi)$ and the cumulative distribution $F(x)$ has a finite positive and differentiable density $f(x)$ in the interval (x_L, x_U), it holds:*

- *If and only if $L'''(\pi) > 0\ \forall \pi \in (0, 1)$, then $f'(x) < 0\ \forall x \in (x_L, x_U)$. This means that $f(x)$ is zeromodal and downward-sloping.*
- *If and only if $L'''(\pi) < 0\ \forall \pi \in (0,1)$, then $f'(x) > 0\ \forall x \in (x_L, x_U)$. This means that $f(x)$ is zeromodal and upward-sloping.*
- *If and only if $L'''(\pi) = 0\ \forall \pi \in (0,1)$, then $f'(x) = 0\ \forall x \in (x_L, x_U)$. This means that $f(x)$ is constant; the distribution is uniform.*
- *If and only if $L'''(\pi) < 0\ \forall \pi < \tilde{\pi} = F(\tilde{x})$ and $L'''(\pi) > 0\ \forall \pi > \tilde{\pi} = F(\tilde{x})$, then $f'(x) > 0\ \forall x < \tilde{x}$ and $f'(x) < 0\ \forall x > \tilde{x}$. This means that $f(x)$ is unimodal with mode $\tilde{x}$.*
- *In general: if and only if $L'''(\pi)$ has $n \geq 1$ sign changes from $L'''(\pi) < 0$ to $L'''(\pi) > 0$ occurring at n points $\tilde{\pi}_i$ (with $i = 1, \ldots, n$), then $f'(x)$ shows the corresponding sign changes from $f'(x) > 0$ to $f'(x) < 0$ occurring at the n points $\tilde{x}_i$ (with $i = 1, \ldots, n$). This means that $f(x)$ is n-modal with modes at $\tilde{x}_i$ (with $i = 1, \ldots, n$).*

Proof. Differentiating (4) with respect to x, one arrives at

(5) $$f'(x) = -\frac{f(x)}{\mu[L''(F(x))]^2} L'''(F(x)).$$

Note that $f(x)$, μ and $[L''(F(x))]^2$ are positive, so there is a negative relative between $f'(x)$ and $L'''(F(x))$. From Definition 2 one can express the modality of a density in terms of $f'(x)$, with a sign change in $f'(x)$ occurring at a mode $\tilde{x}$. The combination of these two results is the relation between $L'''(F(x))$ and the density modality as stated in the theorem, with a sign change in $L'''(F(x))$ corresponding to a mode $\tilde{x}$. ■

Some of the cases from Theorem 2 are more practically relevant than others: the second and third case, upward-sloping zeromodal and constant income densities have been included here for mathematical rigor, but are of limited practical use and will be neglected henceforth. Because of their empirical importance for income and wealth distributions, the focus of the remainder of this paper will be on unimodality and (downward-sloping) zeromodality, although according to Theorem 2, the theoretical result applies to distributions of a higher modality as well.

Summarizing, one can say that it is the sign of the third derivative of the LC which determines the modality of the underlying density function. By Definition 1, the first and second derivatives of an LC are positive. If the third derivative is positive as well, the density is zeromodal (and downward-sloping), but if its sign changes n-times from negative to positive, the density is n-modal. This means that given any parametric LC, one can infer the shape of the underlying density just from looking at its third derivative, without explicitly deriving the density (which can be tedious). The following sections will discuss the implications of this insight for LC fitting in practice.

3. The Density Modality of Some Uniparametric LCs

A number of parametric LCs have been proposed to fit empirical data and they can have one or more parameters; for an overview see Ryu and Solttje (1999) and Sarabia (2008).[1] To start the illustration of Theorem 2, this section will focus on single-parameter LCs, which are known for their simplicity, ease of parameter interpretation, and feasibility in the presence of scarce data.

3.1. *The Pareto, Chotikapanich, and Rohde LCs*

As examples, let us consider the well-known Pareto LC (see Arnold, 1983) and the forms proposed by Chotikapanich (1993) and more recently by Rohde (2009).[2] Their parametric forms $L(\pi)$ are given in the first three rows of Table 1. There one can also see their third derivatives $L'''(\pi)$, which, intriguingly, are all positive on the whole domain. According to Theorem 2, this means that the income density functions associated with these three LCs are zeromodal.[3] One should bear in mind that these LC parametrizations are used in empirical studies with income data—which is often unimodal. One could argue that, given the positivity restriction of $L'(\pi)$ and $L''(\pi)$, it is more obvious to find single-parameter forms whose third derivative is also positive rather than sign-changing. Indeed, expressions for the

[1]For the merits and drawbacks of parametric LCs compared to their non-/semiparametric counterparts, see Ryu and Slottje (1999). Research is advancing in both strands; consider, for example, Sarabia (2008) and Rohde (2009) on parametric LCs and Hasegawa (2003) and Cowell and Victoria-Feser (2007) on non-/semiparametric LCs. One should be aware that the structural assumptions underlying parametric LCs might be a limitation, but they offer the only feasible estimation approach in the presence of grouped data.

[2]As Sarabia *et al*. (2010) have pointed out, this form can be considered a reformulation of the parametrization proposed by Aggarwal (1984).

[3]Because the density functions of these parametric LCs are known, one could also look at them directly to check their zeromodality. For instance, Rohde (2009) has derived the income density of his LC as $f(x)=\frac{1}{2}\sqrt{\frac{\beta(\beta-1)\mu}{x^3}}\quad \forall x\in\left[\frac{(\beta-1)\mu}{\beta};\frac{\beta\mu}{\beta-1}\right]$. However, Theorem 2 allows us to infer its zeromodality without having to derive the functional form of the density explicitly.

TABLE 1

Third Derivatives and Density Modalities of the Pareto, Rohde, Chotikapanich, Lognormal, Weibull LCs

LC	Parameter	$L(\pi)$	$L'''(\pi)$	Density Modality	Gini Coefficient
Pareto	$\alpha > 1$	$1-(1-\pi)^{1-\frac{1}{\alpha}}$	$\frac{\alpha^2-1}{\alpha^3}\cdot(1-\pi)^{-\frac{1}{\alpha}-2}$	zero	$\frac{1}{2\alpha-1}$
Rohde	$\beta > 1$	$\pi\frac{\beta-1}{\beta-\pi}$	$\frac{6\beta(\beta-1)}{(\beta-\pi)^4}$	zero	$2\beta\cdot\left[(\beta-1)\cdot log\left(\frac{\beta-1}{\beta}\right)+1\right]-1$
Chotikapanich	$k > 0$	$\frac{e^{k\pi}-1}{e^k-1}$	$\frac{k^3e^{k\pi}}{e^k-1}$	zero	$\frac{e^k+1}{e^k-1}-\frac{2}{k}$
Lognormal	$\sigma > 0$	$\Phi(\Phi^{-1}(\pi)-\sigma)$	see (8) in the text	uni	$2\Phi\left(\frac{\sigma}{\sqrt{2}}\right)-1$
Weibull	$b > 0$	$1-\frac{\Gamma\left(-log(1-\pi),1+\frac{1}{b}\right)}{\Gamma\left(1+\frac{1}{b}\right)}$	$\frac{(-log(1-\pi))^{\frac{1}{b}-2}}{\Gamma\left(1+\frac{1}{b}\right)b^2(1-\pi)^2}\cdot[1-b+b(-log(1-\pi))]$	zero/uni	$1-2^{-\frac{1}{b}}$

latter tend to be algebraically more involved, as we will see with the Lognormal and Weibull forms, as well as in the next section on multi-parameter LCs.

3.2. *The Lognormal LC*

A widely-used example of a unimodal income distribution is the Lognormal, whose density is given by

(6) $$f(x)=\frac{1}{x\sqrt{2\pi\sigma^2}}e^{-\frac{(log(x)-\bar{\mu})^2}{2\sigma^2}},$$

with shape parameter σ and scaling parameter $\bar{\mu}$.[4] The one mode is located at $\tilde{x}=e^{\bar{\mu}-\sigma^2}$, below the mean of $\mu=e^{\bar{\mu}+\frac{\sigma^2}{2}}$ (see also Aitchison and Brown, 1957). Row 4 of Table 1 shows the Lognormal LC as presented by Sarabia (2008)

(7) $$L(\pi)=\Phi(\Phi^{-1}(\pi)-\sigma),$$

where $\Phi(x)$ denotes the cumulative distribution function of the standard normal distribution of x. Note that the LC depends only on σ because the scaling parameter $\bar{\mu}$ drops out in the step of dividing by the distributional mean. Crucial for our analysis is again the third derivative, because according to Theorem 2 the unimodality of the density implies that it must be sign-changing. In fact, one can show that the Lognormal LC's third derivative

(8) $$L'''(\pi)=\frac{\phi''(\Phi^{-1}(\pi)-\sigma)\cdot\phi(\Phi^{-1}(\pi))-\phi(\Phi^{-1}(\pi)-\sigma)\cdot\phi''(\Phi^{-1}(\pi))}{[\phi(\Phi^{-1}(\pi))]^4}$$
$$-\frac{3\left[\phi'(\Phi^{-1}(\pi)-\sigma)\cdot\phi'(\Phi^{-1}(\pi))-\frac{\phi(\Phi^{-1}(\pi)-\sigma)}{\phi(\Phi^{-1}(\pi))}\cdot(\phi')^2(\Phi^{-1}(\pi))\right]}{[\phi(\Phi^{-1}(\pi))]^4}$$

is sign-changing from negative to positive, with the change occurring at $\tilde{\pi}=F(\tilde{x})=\Phi(-\sigma)$.

3.3. *The Weibull LC*

As this paper focuses on the relation between the LC and the modality of the underlying income density, one functional form deserves particular attention: the Weibull density and distribution function given by

(9) $$f(x)=\frac{b}{a}\left(\frac{x}{a}\right)^{b-1}e^{-\left(\frac{x}{a}\right)^b}\quad\text{and}$$

(10) $$F(x)=1-e^{-\left(\frac{x}{a}\right)^b}$$

[4]In the notation of this paper μ stands for the distributional mean, which is why I resort to labeling $\bar{\mu}$ the parameter of the Lognormal distribution otherwise known as μ.

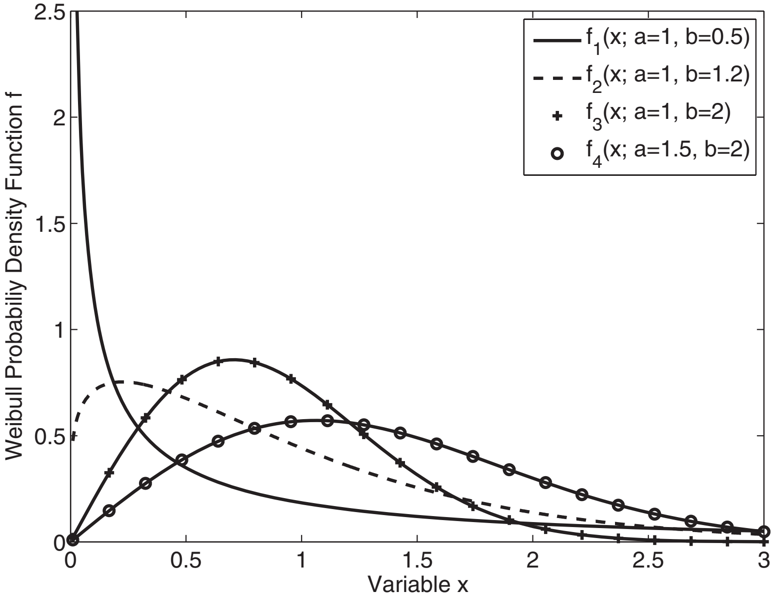

Figure 3. Weibull Density Functions with Different *a* and *b* Parameters

for $x > 0$, with scaling parameter $a > 0$ and shape parameter $b > 0$. The Weibull distribution is a special case of the Generalized Beta distribution. In this context, its ability to model income densities has been analyzed by McDonald (1984) and McDonald and Ransom (2008).

The reason why this density is so appealing to our analysis lies in the flexibility of b: for $b \leq 1$, the density is zeromodal, for $b > 1$, it is unimodal, as shown in Figure 3. The scale parameter a determines the location of the mode and how far the density is spread out, but the shape is entirely captured by b. In the derivation of the associated LC, a drops out so that we have a single-parameter LC depending only on b. Its functional form is

(11)
$$L(\pi) = 1 - \frac{\Gamma\left(-log(1-\pi), 1 + \frac{1}{b}\right)}{\Gamma\left(1 + \frac{1}{b}\right)},$$

where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt$ is the Gamma function and $\Gamma(x, \alpha) = \int_x^\infty t^{\alpha-1} e^{-t}\, dt$ is the upper incomplete Gamma function.

The LCs of the four densities from Figure 3 are plotted in Figure 4, showing that higher values of b unambiguously imply less inequality.

Row 5 of Table 1 presents the Weibull LC and its third derivative

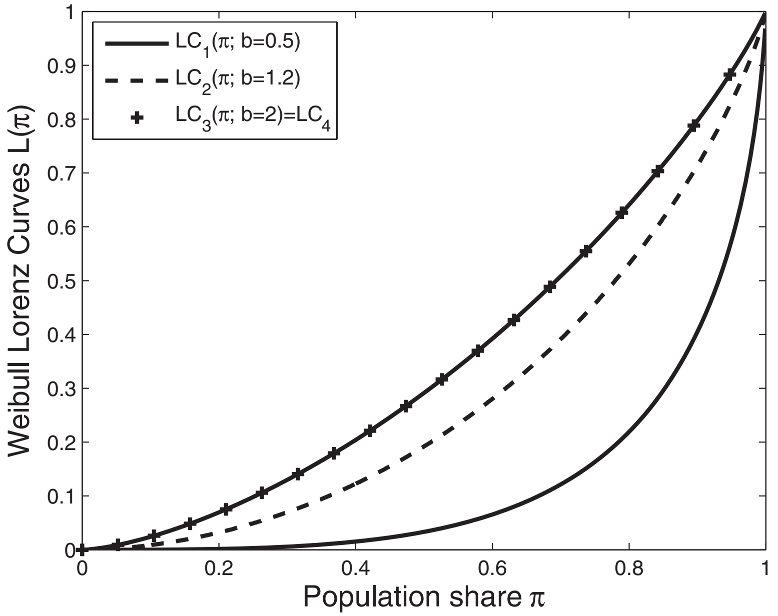

Figure 4. Weibull LCs with Different *b* Parameters

$$L'''(\pi) = \frac{(-log(1-\pi))^{\frac{1}{b}-2}}{\Gamma\left(1+\frac{1}{b}\right)b^2(1-\pi)^2}[1-b+b(-log(1-\pi))]. \tag{12}$$

According to Theorem 2, the sign of this expression determines the modality of the density. As the fraction is unambiguously positive, we only need to consider the term in square brackets: for $0 < b \leq 1$, $[1-b + b(-log(1-\pi))]$ is positive as well, hence the third derivative is positive and the associated density zeromodal. Conversely, for $b > 1$, $[1 - b + b(-log(1 - \pi))]$ is negative for small π and positive for larger π, with the sign change occurring at $\tilde{\pi} = 1 - e^{\frac{1}{b}-1}$. This point refers to the mode $\tilde{x} = a\left(1-\frac{1}{b}\right)^{\frac{1}{b}}$ of the unimodal Weibull density for $b > 1$.

4. The Density Modality of Some Multiparametric LCs

After discussing the modality of some single-parameter LCs, let us now apply Theorem 2 to LCs with more than one parameter. These are very important in practice but do entail an increased complexity in estimating and interpreting their parameters as compared to their single-parameter counterparts. This section starts by discussing the modality of two widely-used multi-parameter LCs, the

Singh–Maddala and the Elliptical LC, and focuses then on two hierarchical LC families. A general result for the modality of multi-parameter LCs of these families in terms of their single-parameter baseline functions proves very useful.

4.1. *The Biparametric Singh–Maddala LC*

The distribution proposed by Singh and Maddala (1976) has been shown to be very accurate in fitting income and wealth data (see Kleiber and Kotz, 2003). Its density and distribution functions

(13) $$f(x) = \frac{aqx^{a-1}}{b^a\left(1+\left(\frac{x}{b}\right)^a\right)^{q+1}} \quad \text{and}$$

(14) $$F(x) = 1 - \frac{1}{\left[1+\left(\frac{x}{b}\right)^a\right]^q}$$

involve the three parameters a, b, $q > 0$ with $aq > 1$. b is only a scale parameter and does not appear in the biparametric LC presented by Sarabia (2008)

(15) $$L(\pi, a, q) = I_{\left[1-(1-\pi)^{\frac{1}{q}}\right]}\left(1+\frac{1}{a}, q-\frac{1}{a}\right) = \frac{\int_0^{1-(1-\pi)^{\frac{1}{q}}} t^{\frac{1}{a}}(1-t)^{q-\frac{1}{a}-1}\,dt}{\int_0^1 t^{\frac{1}{a}}(1-t)^{q-\frac{1}{a}-1}\,dt},$$

where $I_x(m, n)$ denotes the incomplete beta function ratio

(16) $$I_x(m,n) = \frac{\int_0^x t^{m-1}(1-t)^{n-1}\,dt}{\int_0^1 t^{m-1}(1-t)^{n-1}\,dt} = \frac{B(x,m,n)}{B(m,n)}.$$

For the analysis on the third derivative and the density modality, one needs to differentiate (15) three times, yielding

(17) $$L'''(\pi, a, q) = \frac{\left(1-(1-\pi)^{\frac{1}{q}}\right)^{\frac{1}{a}-2}(1-\pi)^{-\frac{1}{aq}-2}}{aq^2\cdot B\left(1+\frac{1}{a}, q-\frac{1}{a}\right)}\cdot\left[\frac{1}{aq}+1-\left(\frac{1}{q}+1\right)(1-\pi)^{\frac{1}{q}}\right].$$

Determining the sign of this expression is easier than it might appear because all the terms in the fraction are positive and only the term in square brackets needs to be considered. For $a \le 1$ it holds

$$\frac{1}{aq}+1-\left(\frac{1}{q}+1\right)(1-\pi)^{\frac{1}{q}} \geq \frac{1}{q}+1-\left(\frac{1}{q}+1\right)(1-\pi)^{\frac{1}{q}}$$
$$=\left(\frac{1}{q}+1\right)\left(1-(1-\pi)^{\frac{1}{q}}\right)>0 \text{ for } \pi \in (0,1).$$

So for $a \leq 1$, $L'''(\pi, a, q)$ is always positive and the density thus zeromodal.

With $a > 1$, however, the term in square brackets can become negative for low π and changes to positivity at the point $\tilde{\pi}$, which refers to the mode

(18)
$$\tilde{\pi}=1-\left(\frac{\frac{1}{aq}+1}{\frac{1}{q}+1}\right)^{q}.$$

Summarizing, the density associated with the Singh–Maddala LC is zeromodal for $a \leq 1$ and unimodal for $a > 1$. This is a remarkable result: although the LC depends on two shape parameters a and q, only one of them influences the modality of the density. When conducting curve-fitting with the Singh–Maddala LC, the researcher should thus pay attention to the value of a and make sure it is in line with the modality of the underlying data.

Going back to Figures 1 and 2 from the motivating Section 1, we have the example of the two similar looking Singh–Maddala LCs with ($a = 0.9$, $q = 7$) and ($a = 1.2$, $q = 2$), both yielding the same Gini coefficient of 0.58.[5] After the previous analysis, it is directly clear that LC1, with $a = 0.9 \leq 1$, belongs to the zeromodal density from Figure 2 and LC2, with $a = 1.2 > 1$, to the unimodal density.

4.2. *Villaseñor and Arnold's Triparametric Elliptical LC*

The Elliptical LC proposed by Villaseñor and Arnold (1989) is known for its good data fit and has been incorporated into the World Bank's POVCAL software. Its triparametric form is

(19)
$$L(\pi, a, b, d)=\frac{1}{2}\left[-(b\pi+e)-\sqrt{\alpha\pi^{2}+\beta\pi+e^{2}}\right],$$

where $e = -(a + b + d + 1)$, $\alpha = b^2 - 4a < 0$ (so that the curve is an ellipse rather than a parabola or hyperbola), and $\beta = 2be - 4d$. The four necessary and sufficient

[5]The Gini coefficient of the Singh–Maddala LC can be calculated with the formula given by Sarabia (2008): $Gini=1-\frac{\Gamma(q)\Gamma\left(2q-\frac{1}{a}\right)}{\Gamma\left(q-\frac{1}{a}\right)\Gamma(2q)}$. Also note that according to (13), a third parameter b is needed for the density in Figure 2. The choice of $b = 12.35$ for LC1 and $b = 2.08$ for LC2 ensures that the median of both densities is equal to 1.

conditions for (19) to be an LC are $\alpha = b^2 - 4a < 0$, $a + b + d + 1 = -e > 0$, $d \geq 0$, and $a + d - 1 \geq 0$.[6]

In order to analyze the modality we consider the third derivative

$$L'''(\pi, a, b, d) = \frac{\frac{3}{16}(4\alpha e^2 - \beta^2)(2\alpha\pi + \beta)}{\sqrt{(\alpha\pi^2 + \beta\pi + e^2)^5}}. \tag{20}$$

In this expression, the denominator is always positive and the factor $4\alpha e^2 - \beta^2$ is always negative. So the sign of L''' is the opposite sign of $2\alpha\pi + \beta = 2(b^2 - 4a)\pi + 2be - 4d$. An analysis of this term—the details of which are in the Appendix—yields the result:

- A necessary and sufficient condition for the unimodality of the density is

$$\frac{-4a - 2d}{a + d + 1} < b < -\frac{a + d + 1}{2} + \sqrt{\left(\frac{a + d + 1}{2}\right)^2 - 2d} \tag{21}$$

  because then the third derivative is sign-changing and the mode occurs at $\tilde{\pi} = \frac{2d - be}{b^2 - 4a}$.
- For b smaller than or equal to the lower boundary in (21), the density is upward-sloping zeromodal (to be neglected in practice).
- For b greater than or equal to the upper boundary in (21), the density is (downward-sloping) zeromodal.

So the three parameters a, b, and d together determine the modality of the underlying density.

Figure 5 presents the LCs of five parameter combinations with different modality and Gini coefficients.[7] One can see that both uni- and zeromodal densities cover a vast range of inequality levels. This example underlines that from looking only at the LCs and Gini coefficients, it is impossible to infer the density modality. Compare for instance the unimodal LC1 ($a = 2$, $b = -1.5$, $d = 1$, Gini coefficient 0.21) and the zeromodal LC2 ($a = 5$, $b = -0.5$, $d = 3$, Gini coefficient 0.23). Only the third derivative and the parameter condition (21) derived from it can tell the researcher what the modality is.

4.3. *Sarabia et al.'s First Hierarchical Family* $L_1(\pi, \alpha)$

The third derivative of multiparametric LCs can become algebraically involved. If, however, the LC belongs to a hierarchical family (as defined by Sarabia *et al.*, 1999, 2001), its third derivative can be expressed using the derivatives of its uniparametric baseline function.

[6]Note that in Villaseñor and Arnold (1989) this condition is incorrectly stated as $a + d - 1 \leq 0$ (see Krause, 2013).

[7]The Gini coefficient can be calculated with the formula given by Villaseñor and Arnold (1989):

$$Gini = 1 + \frac{b + 2e}{2} + \frac{(2\alpha + \beta)(a + d - 1) + \beta e}{4\alpha} + \frac{\beta^2 - 4\alpha e^2}{8\alpha\sqrt{-\alpha}}\left[sin^{-1}\left(\frac{2\alpha + \beta}{\sqrt{\beta^2 - 4\alpha e^2}}\right) - sin^{-1}\left(\frac{\beta}{\sqrt{\beta^2 - 4\alpha e^2}}\right)\right].$$

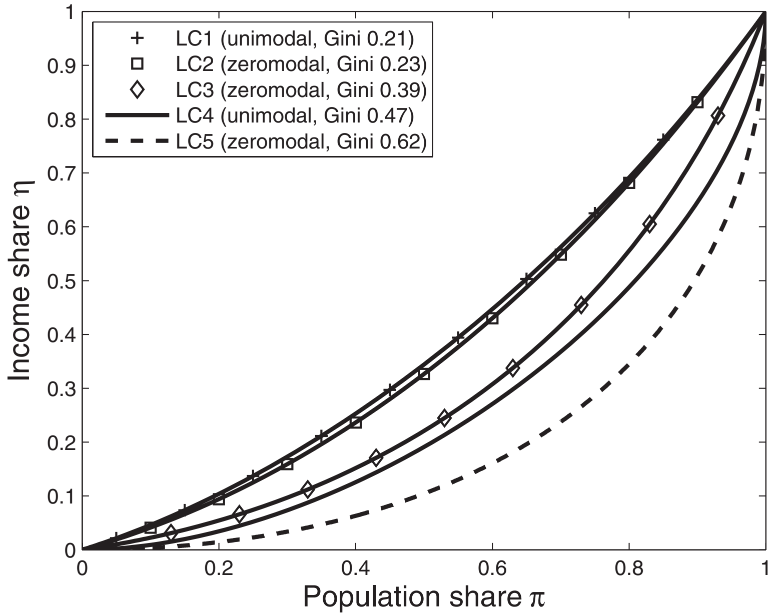

Figure 5. Five Elliptical LCs with the Following Parameters: LC1: $a = 2$, $b = -1.5$, $d = 1$; LC2: $a = 5$, $b = -0.5$, $d = 3$; LC3: $a = 2$, $b = 1.5$, $d = 1$; LC4: $a = 1$, $b = -1$, $d = 0$; LC5: $a = 1$, $b = 1$, $d = 0$

Consider any uniparametric LC $L_0(\pi)$. Sarabia *et al.* (1999) show that the first hierarchical family

$$L_1(\pi, \alpha) = \pi^{\alpha} L_0(\pi) \quad \text{with } (\alpha \geq 1) \vee (0 \leq \alpha < 1, L_0'''(\pi) \geq 0) \tag{22}$$

is also an LC. In this way uniparametric LCs can be extended to include another parameter.

When analyzing the modality of the biparametric LC $L_1(\pi, \alpha)$, one can write its third derivative as

$$L_1'''(\pi, \alpha) = \pi^{\alpha} \left[\alpha(\alpha - 1) \left((\alpha - 2) \frac{L_0(\pi)}{\pi^3} + 3 \frac{L_0'(\pi)}{\pi^2} \right) + 3\alpha \frac{L_0''(\pi)}{\pi} + L_0'''(\pi) \right]. \tag{23}$$

This expression only involves the derivatives of the baseline uniparametric LC $L_0(\pi)$ and the parameter α, which facilitates the analysis for a given family.

A general analysis of the sign of $L_1'''(\pi, \alpha)$ is only possible in a limited way, though. One can see that $L_0'''(\pi) > 0$ and $\alpha \geq 2$ will lead to a positive $L_1'''(\pi, \alpha)$. So any LC adapted from a baseline LC with a zeromodal density and family parameter $\alpha \geq 2$ will have a zeromodal density modality itself, which is a remarkable result. But this condition is only sufficient and not necessary: the first term in (23) is negative if $0 < \alpha < 2$ but the whole expression might still be positive, depending on the magnitudes of $L_0'(\pi)$, $L_0''(\pi)$ and $L_0'''(\pi)$.

TABLE 2

Density Modality for Different Combinations of the Pareto Baseline Parameter K and, Respectively the First Family Parameter α (Ortega LC, see (24)) or Second Family Parameter γ (Rasche LC, see (31))

Family Parameter α (Family Parameter γ)	$0 < k < 1$	$k = 1$
$\alpha = 0$ ($\gamma = 1$)	zeromodal	egalitarian (all incomes = μ)
$0 < \alpha < 1$ ($1 < \gamma < 2$)	unimodal	zeromodal upward-sloping
$\alpha = 1$ ($\gamma = 2$)	zeromodal	constant uniform density
$\alpha > 1$ ($\gamma > 2$)	zeromodal	zeromodal

In order to discuss (23) in more detail, let us apply it to one particular family, namely the biparametric LC by Ortega *et al.* (1991), which is based on the uniparametric Pareto LC $L_0(\pi, k) = 1 - (1 - \pi)^k$:[8]

$$L_1(\pi,\alpha,k) = \pi^\alpha(1-(1-\pi)^k) \quad \text{with } (\alpha \ge 0; 0 < k \le 1). \tag{24}$$

Plugging the first three derivatives of the Pareto LC $L_0(\pi, k) = 1 - (1 - \pi)^k$

$$L_0'(\pi,k) = k(1-\pi)^{k-1}, \ge 0, \tag{25}$$

$$L_0''(\pi,k) = -k(k-1)(1-\pi)^{k-2}, \ge 0 \quad \text{and} \tag{26}$$

$$L_0'''(\pi,k) = k(k-1)(k-2)(1-\pi)^{k-3} \ge 0 \tag{27}$$

into (23) gives the third derivative of the Ortega LC as

$$\begin{aligned} L_1'''(\pi,\alpha,k) = \pi^\alpha[\alpha(\alpha-1)[(\alpha-2)\pi^{-3}(1-(1-\pi)^k) + 3\pi^{-2}k(1-\pi)^{k-1}] \\ - 3\alpha\pi^{-1}k(k-1)(1-\pi)^{k-2} + k(k-1)(k-2)(1-\pi)^{k-3}]. \end{aligned} \tag{28}$$

The sign and thus the density modality of $L_1'''(\pi,\alpha,k)$ depend on both α and k. The main result is that for $0 < k < 1$, a family parameter $0 < \alpha < 1$ leads to a unimodal density and $\alpha > 1$ to a zeromodal one. If α and k take the boundary values 0 or 1, special cases may arise as shown in Table 2.

4.4. *Sarabia et al.'s Second Hierarchical Family $L_2(\pi, \gamma)$*

The second hierarchical family of LCs described in Sarabia *et al.* (1999)

$$L_2(\pi,\gamma) = (L_0(\pi))^\gamma \quad \text{with } \gamma \ge 1 \tag{29}$$

[8]The Pareto parameter is here called k and not α as in the previous section (Table 1), in order to avoid confusion with the family parameter α. Also note that $\alpha \ge 0$ combines the parameter spaces $0 \le \alpha < 1 \vee \alpha \ge 1$ from (22) because the Pareto LC satisfies $L_0'''(\pi) \ge 0$.

shows a different way of extending a baseline uniparametric LC $L_0(\pi)$ in a multiparametric way. Like in the case of the first family, the third derivative of $L_2(\pi, \gamma)$ can be written using only the derivatives of $L_0(\pi)$:

$$(30)\quad L_2'''(\pi,\gamma) = \gamma(L_0(\pi))^{\gamma-1}\left[(\gamma-1)\left((\gamma-2)\frac{(L_0'(\pi))^3}{(L_0(\pi))^2} + 3\frac{L_0'(\pi)L_0''(\pi)}{L_0(\pi)}\right) + L_0'''(\pi)\right].$$

Note the similarity to (23): $L_0'''(\pi)$ and $\gamma \geq 2$ are sufficient but not necessary conditions for zeromodality of $L_2(\pi, \gamma)$.

Let us consider an example of the second hierarchical family, again using the Pareto LC as the baseline function. The biparametric LC

$$(31)\qquad L_2(\pi,k,\gamma) = (1-(1-\pi)^k)^\gamma \quad \text{with } (0 < k \leq 1; \gamma \geq 1)$$

was proposed by Rasche *et al.* (1980). Using the derivatives of the Pareto LC from (25), (26), and (27), (30) becomes

$$(32)\quad L_2'''(\pi,k,\gamma) = \gamma k(1-(1-\pi)^k)^{\gamma-1}(1-\pi)^{k-3} \cdot\left[(\gamma-1)\frac{(\gamma-2)k^2(1-\pi)^{2k}}{(1-(1-\pi)^k)^2} - \frac{3k(k-1)(1-\pi)^k}{1-(1-\pi)^k} + (k-1)(k-2)\right].$$

One can show that for $0 < k < 1$ and $1 < \gamma < 2$, this expression is sign-changing and the density unimodal, whereas $0 < k < 1$ and $\gamma \geq 2$ lead to a zeromodal density. Other parameter combinations, involving boundary cases, are displayed in Table 2. The correspondence of the parameters α and γ of the first and second family is worth noticing. Together, these cases illustrate how the modality of multiparametric LCs can be influenced by the modality of their baseline functions and the choice of the family parameters.

5. Monte Carlo Simulation

The two previous sections have analyzed the density modalities of many important LCs and shown some examples of similar looking LCs with differently-shaped densities. But how important is this result in practice? If a researcher chooses an LC to fit their data in an error-minimizing way, won't the LC have the correct density modality? The answer is clearly no, as the following Monte Carlo simulation shows.

The simulation is done with the five uniparametric LCs from Section 3; repeating it with multiparametric LCs does not add further insight. As goodness of fit criteria for the LCs, the Mean Squared Error (MSE) and the Gini difference are used. While the MSE indicates how well the parametric LC fits the given data points, the Gini difference (see Chotikapanich, 1993) captures the overall fit of the shape of the LC. It exploits the following relation between an LC and its implied Gini coefficient:

$$(33)\qquad Gini = 1 - 2\int_0^1 L(\pi)\,d\pi.$$

The Gini difference is then obtained as the absolute difference between the Gini implied by (33) and the actual one (given by the data). The Gini coefficients associated with the five LCs are given in the last column of Table 1.[9]

The MC Simulation is carried out as follows:

1. Take 10,000 draws from a zeromodal or unimodal income density with mean 1.
2. Aggregate the drawn income data to decile data points (to mimic grouped data as available in many applications).
3. For each of the five LCs, solve for the parameter which best fits the decile data points by minimizing the MSE.[10] This procedure is carried out using the fminunc-routine in MATLAB.
4. Calculate and store the MSEs and the Gini differences for each of the five LCs.
5. Repeat Steps 1 to 4 10,000 times.
6. Look which of the five parametric forms has the lowest MSE and/or Gini difference overall: does the density implied by this LC have the correct shape, that is, does it look similar to the actual one?

The results for a selection of underlying densities and inequality levels are shown in Table 3; more detailed results are available from the author upon request. Of course, fitting an LC of the same parametrization as the underlying density always leads to the lowest MSE and Gini difference (in italics), but the focus lies on which of the other LC forms does best (in bold).

For instance, for an underlying Weibull density (quite egalitarian with $b = 3$), a fitted Chotikapanich LC leads to the lowest MSE at the decile data points. If, based on this criterion, the researcher chose the Chotikapanich LC, the implied density would be zeromodal and very different from the true Weibull density. Figures 6 and 7 show the LCs and densities for this case: the very similar looking LCs give rise to considerably varying income densities. A researcher aware of the unimodality of his data should rather decide on the Lognormal LC, which, despite its slightly higher MSE at the decile points, captures the unimodal shape of the Weibull income density. A diverging shape might lead to incorrect conclusions about the distribution of income, because the downward-sloping Chotikapanich density misses out on the dominance of the middle class and implies a larger number of poor earners. Incidentally, the Gini difference as goodness of fit criterion would (appropriately) point toward the Lognormal LC in this setting.[11]

[9]These formulas can be found, respectively, in Arnold (1983) (for the Pareto LC), Rohde (2009), Chotikapanich (1993) (this paper uses a simplified expression of her formula), and McDonald (1984) (with the Weibull and Lognormal distributions as special cases of the Generalized Beta).

[10]As a referee has pointed out, the LC least squares estimator is consistent but there is a lot of research going on about more efficient estimators, given heteroscedastic and autocorrelated residuals. Kakwani and Podder (1973) were among the first to suggest an alternative procedure. An exploratory analysis shows that in the given application the results would hardly change.

[11]Table 3 also gives two measures of distance from the implied density $g(x)$ to the true one $f(x)$ in order to underline the importance of the correct modality: the Kullback–Leibler Distance (see Kullback and Leibler, 1951) $\int_{-\infty}^{\infty} f(x)log\left(\frac{f(x)}{g(x)}\right)dx$ and the intuitive Integral Difference $\int_{-\infty}^{\infty} |f(x)-g(x)|dx$. In the case with the underlying Weibull density discussed, the two measures support our analysis by confirming that the unimodal Lognormal density has the smallest distance from the true unimodal Weibull.

TABLE 3

MC Simulation for Different Underlying Densities

Parametric LC Density Modality	Pareto zero	Rohde zero	Chotikapanich zero	Lognormal uni	Weibull zero/uni
Parameter	$\alpha > 1$	$\beta > 1$	$k > 0$	$b > 0$	$\sigma > 0$
Underlying WEIBULL with $b = 3$ (unimodal, Gini 0.2063)					
Est. mean parameter	3.0976	2.2010	1.2372	0.3653	*3.0007*
(S.D.)	(0.0187)	(0.0120)	(0.0096)	(0.0027)	*(0.0249)*
MSE	0.000995	0.000142	**0.000073**	0.000110	*0.000000*
(S.D.)	(0.000024)	(0.000007)	(0.000004)	(0.000007)	*(0.000000)*
Gini difference	0.013807	0.006809	0.005177	**0.002543**	*0.001218*
(S.D.)	(0.001389)	(0.001455)	(0.001485)	(0.001375)	*(0.000918)*
K-L distance	7.5213	4.5157	4.0850	**0.3215**	*0.0002*
Integral diff.	0.8367	0.4829	0.4249	**0.2510**	*0.0002*
Underlying LOGNORMAL with $\sigma = 0.8$ (unimodal, Gini 0.4284)					
Est. mean parameter	1.7162	1.3574	2.8746	*0.8000*	1.2180
(S.D.)	(0.0103)	(0.0059)	(0.0300)	*(0.0072)*	(0.0134)
MSE	0.001532	**0.000098**	0.000449	*0.000001*	0.000337
(S.D.)	(0.000048)	(0.000007)	(0.000036)	*(0.000001)*	(0.000027)
Gini difference	0.017261	0.008374	**0.004837**	*0.002781*	0.005762
(S.D.)	(0.003466)	(0.003387)	(0.003005)	*(0.002094)*	0.003255
K-L distance	9.1313	4.4648	2.6921	*0.0001*	**0.2440**
Integral diff.	0.7534	0.4185	0.4078	*0.0001*	**0.2632**
Underlying CHOTIKAPANICH with $k = 1.5$ (zeromodal, Gini 0.2411)					
Est. mean parameter	2.6629	1.9289	*1.4997*	0.4393	2.4466
(S.D.)	(0.0100)	(0.0063)	*0.0076*	(0.0021)	0.0132
MSE	0.001058	**0.000039**	*0.000000*	0.000085	0.000084
(S.D.)	(0.000024)	(0.000005)	*(0.000000)*	(0.000006)	0.000005
Gini difference	0.009927	**0.001834**	*0.000905*	0.002820	0.005621
(S.D.)	(0.001073)	(0.001028)	*(0.000682)*	(0.001120)	(0.001154)
K-L distance	9.5512	2.8658	*0.0078*	**0.3666**	0.3795
Integral diff.	0.8664	**0.2753**	*0.0004*	0.3645	0.4133
Underlying ROHDE with $\beta = 1.1$ (zeromodal, Gini 0.6725)					
Est. mean parameter	1.2428	*1.1000*	6.4609	1.4276	0.5814
(S.D.)	(0.0036)	*(0.0017)*	(0.0729)	(0.0091)	0.0049
MSE	0.001730	*0.000001*	0.001475	**0.000084**	0.000503
(S.D.)	(0.000080)	*(0.000001)*	(0.000071)	(0.000005)	(0.000034)
Gini difference	0.002646	*0.002377*	0.021085	**0.014771**	0.024003
(S.D.)	(0.002013)	*(0.001815)*	(0.003267)	(0.003077)	(0.003052)
K-L distance	13.4991	*0.0008*	1.6146	**0.4092**	0.5860
Integral diff.	0.9072	*0.0002*	0.8571	**0.4850**	0.7034
Underlying PARETO with $\alpha = 2.5$ (zeromodal, Gini 0.2500)					
Est. mean parameter	*2.5015*	1.8778	1.5277	0.4574	2.3918
(S.D.)	*(0.0439)*	(0.0284)	(0.0357)	(0.0105)	(0.0607)
MSE	*0.000000*	0.000790	0.001175	**0.000692**	0.001500
(S.D.)	*(0.000001)*	(0.000065)	(0.000097)	(0.000057)	(0.000100)
Gini difference	*0.004301*	**0.004429**	0.005958	0.005108	0.004448
(S.D.)	*(0.003536)*	(0.003492)	(0.003942)	(0.004284)	0.003799
K-L distance	*0.0402*	1.9287	2.3998	**0.6723**	1.1494
Integral diff.	*0.0021*	0.7803	0.8790	**0.6811**	0.8566

Note: Printed in italics are the results for the LC form associated with the underlying density; printed in bold are the results for the best-performing LC apart from this one.

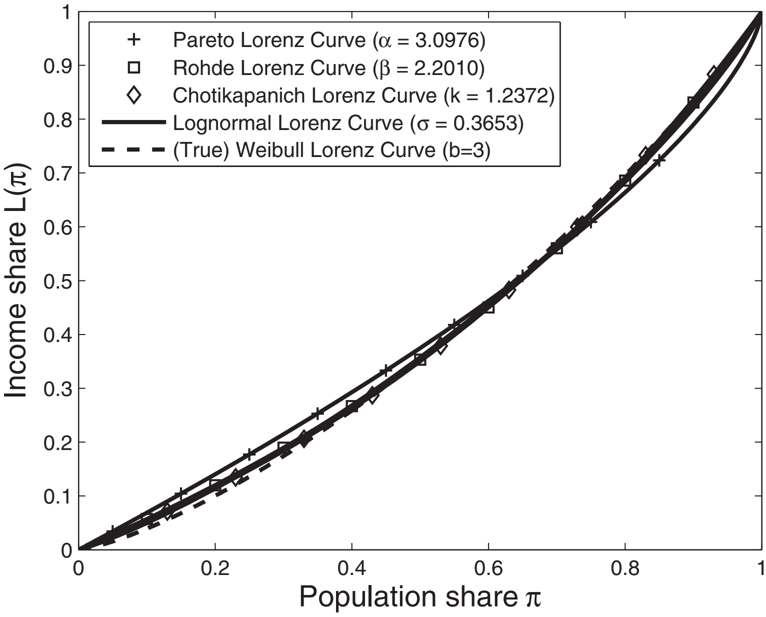

Figure 6. Parametric LCs Fitting Decile Data from a Simulated Underlying Weibull Density ($b = 3$, Gini = 0.2063)

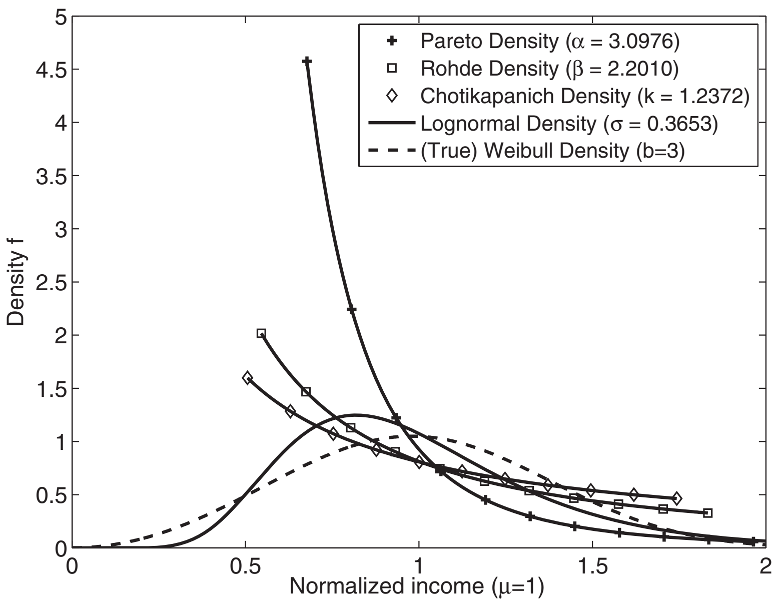

Figure 7. Normalized Densities Associated with the LCs from Figure 6

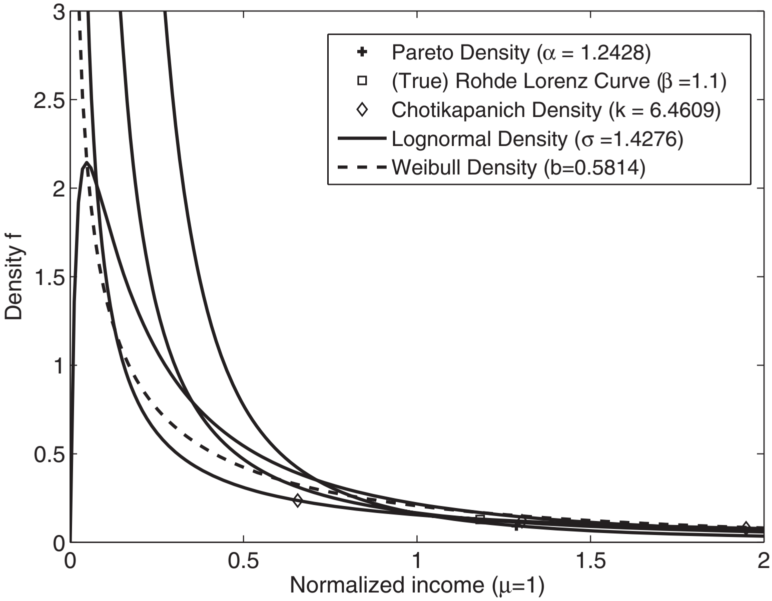

Figure 8. Normalized Densities Associated with LCs Resulting from Rohde Density with $\beta = 1.1$ (Gini = 0.6725)

However, in general one cannot count on the Gini difference reliably leading to the LC with the correct modality, as Table 3 shows.

One setting from Table 3 deserves particular attention: when the true density is a very inegalitarian zeromdodal of the Rohde form ($\beta = 1.1$), MSE and Gini difference are lowest for the Lognormal LC, the only one to imply a unimodal density. Surprisingly, the density difference measures are in accordance with this choice and suggest that the unimodal Lognormal density comes nearer to the true one than the zeromodal densities.

The reason lies at the very high inequality level: Figure 8 shows that the mode of the Lognormal density is near zero which makes it effectively downward-sloping on most of the domain.[12]

One can conclude that the relevance of choosing an LC with the correctly implied modality diminishes in extremely inegalitarian settings. This case, with a Gini coefficient of 0.6725, is however of limited practical importance. For

[12]In a Lognormal density with mean 1, the mode is located at $\tilde{x} = e^{\bar{\mu}-\sigma^2} = e^{-\frac{3}{2}\sigma^2}$. Using the Lognormal Gini coefficient $2\Phi\left(\frac{\sigma}{\sqrt{2}}\right)-1$, one can express the location of the mode as a direct function of the Gini coefficient:

$$\tilde{x} = e^{-3\left[\Phi^{-1}\left(\frac{Gini+1}{2}\right)\right]^2}. \tag{34}$$

For a Gini of 0.5, the mode is at 0.2554, while in the above case of a Gini at 0.6725, the mode is at very low 0.0563—keeping in mind that mean income is still 1.

TABLE 4
Descriptive Overview of the Dataset

Region	Europe (West)	Europe (East)	Africa	Asia	The Americas
# Countries	18	24	25	25	28
# Observations	481	261	53	284	549
Mean reported Gini coeff. (S.D.)	0.3117	0.2921	0.5647	0.3862	0.5040
	(0.0563)	(0.0726)	(0.1118)	(0.0836)	(0.0702)
# Weibull-implied unimodals	479	257	10	242	213
# Weibull-implied zeromodals	2	4	43	42	336

egalitarian and moderately inegalitarian income distributions, the researcher should definitely take the modality into account upfront because goodness of fit criteria like MSE and Gini difference can lead to an LC with an incorrect density shape.

6. An Empirical Analysis

In this final section an exploratory empirical study is undertaken to illustrate the relation between LCs from different countries and their density modality. The data source is United Nations University–World Institute for Development Economics Research (2008) because of its wide coverage, including decile data from many countries and time periods. From this database, 1,628 LC decile data rows refer to income (rather than consumption or expenditure) and have a quality rating of 1, 2, or 3 (excluding memorandum items and those from particularly unreliable old sources).[13] The decile data on LCs in the database are accompanied by a reported Gini coefficient, which has either been reported directly by the countries or calculated by the World Bank with the POVCAL software. The reported Gini coefficients range from below 0.2 (e.g., in Finland in the 1980s and 1990s) to above 0.7 from Mali, Mauritania, Zambia, and Zimbabwe in the 1980s and 1990s.

One should keep in mind that direct cross-country comparisons are beset by some caveats, for instance the slightly varying income definitions across countries and data quality ratings. But for our goal of analyzing LC performance and its implications for income density modality in a variety of settings and inequality levels, the 1,628 LCs from 120 different countries offer a good opportunity.

Table 4 shows some aggregate statistics on the 120 countries, dividing them into five regional groups: Europe (West), Europe (East, referring to the formerly centrally-planned economies), Africa, the Americas (including the Caribbean), and Asia (which also comprises the Middle East and Pacific countries). According to the mean reported Gini coefficients, inequality in formerly Socialist Europe is slightly lower than in Western Europe but its inter-country dispersion is larger.

[13]It is known that consumption is typically more equally distributed than income. Note that wealth would be even more unequally distributed than income (see, e.g., Castañeda *et al.*, 2003). Excluding all data other than income is a basis for ensuring at least a broad comparability across countries.

Income distributions in Asia are more egalitarian than in the Americas, while Africa is the continent with the highest mean Gini coefficient and the highest inter-country dispersion.

Now what is the density modality of the 1,628 settings? Ideally, researchers should obtain additional data in order to infer the modality. This paper solely relies on the decile data given and instead exploits the flexibility of the Weibull LC discussed in Section 3: it uses the modality implied by the Weibull LC as a proxy for the true one. A Weibull LC is fitted to the data and, when its fitted parameter b is larger than 1, one can conclude that the density is unimodal, otherwise it is assumed to be zeromodal. This Weibull-implied modality is also given in Table 4: according to this proxy, income is predominantly unimodal in Europe and Asia, mostly zeromodal in Africa, and sometimes unimodal, sometimes zeromodal on the American continent. This observation is in line with the consensus in the literature (see Dagum, 1999; Kleiber, 2008) that income distributions in poorer, overpopulated countries tend to be zeromodal and those in more developed countries unimodal. In terms of the Gini coefficient, here we see that countries with high Gini coefficients (in particular higher than 0.45–0.48) tend to have Weibull-implied zeromodality, those with low Gini coefficients implied unimodality.

The question is now if the Weibull-implied modality coincides with the density modality implied by the best-fitting LC. Given the decile data, it does not make sense to fit LCs with more than one parameter, so let us restrict the selection to the five uniparametric LCs from Section 3: Pareto, Rohde, Chotikapanich, Lognormal, and Weibull. Each of the 1,628 LCs is estimated separately: for each of the five parametric forms, the parameter which best fits the decile data points by minimizing the MSE is calculated. MSE and Gini difference are then calculated accordingly, using the reported Gini coefficient as a proxy for the actual one.

Detailed results for every LC and every country can be obtained from the author upon request, but the main points become clear from Tables 5 and 6. In Table 5, the best-fitting LCs are shown for the worldwide sample of Weibull-implied zeromodal and unimodal settings: with the MSE as criterion, the Lognormal LC outperforms across both the unimodal and zeromodal settings, thereby erroneously suggesting a unimodal density shape in the latter case. In fact, for 408 out of the 427 zeromodals, the Lognormal LC has the lowest MSE. When restricting the choice to the LCs which imply the correct zeromodality, Rohde's form obtains the lowest MSE for 398 of them. The Gini difference as a criterion leads to a wider range of outperformers: among the unimodals, the Weibull LC achieves the best fit in more settings than the Lognormal LC. But the Gini difference cannot guarantee the right density modality either: consider the more than 50 LCs of each of the Pareto, Rohde, and Chotikapanich forms it proposes in the unimodal setting. Among the zeromodals, each of the five forms has the lowest Gini difference in many settings and the inappropriate Lognormal LC does not dominate as starkly.

Further insights can be gained from Table 6, where the results are broken down by region.[14] In (unimodal) Western and Eastern European countries the

[14]In Europe (West) and Europe (East) only the results for the unimodals are reported, because they make up, respectively, 479 out of 481 and 257 out of 261 LCs.

TABLE 5

Results from Fitting LCs from the Worldwide Sample to Five Parametric Forms (settings with Weibull-implied uni- and zeromodality)

Parametric LC	Pareto	Rohde	Chotikapanich	Lognormal	Weibull
WEIBULL-IMPLIED UNIMODALS (1201 LCs, Mean Gini 0.3387)					
Est. min. parameter	1.5426	1.2601	0.8672	0.2612	1.0002
Est. mean parameter	2.1319	1.6152	2.1815	0.6222	1.7810
Est. max. parameter	3.9807	2.8444	3.4764	0.9417	4.4134
MSE	0.001034	0.000180	0.000417	**0.000073**	0.000372
(S.D.)	(0.000610)	(0.000202)	(0.000462)	(0.000176)	(0.000386)
MSE/Lowest MSE	14.1644	2.4658	5.7123	1.0000	5.0959
Gini difference	0.015550	0.010238	0.009169	**0.006672**	0.007345
(S.D.)	(0.009513)	(0.008969)	(0.009237)	(0.009631)	(0.010148)
Gini diff./lowest Gini diff.	2.3306	1.5345	1.3743	1.0000	1.1009
# Lowest MSE	28	4	5	**1067**	97
# Lowest MSE (uni-mod)	–	–	–	**1100**	101
# Lowest Gini diff.	127	61	77	365	**571**
# Lowest Gini diff. (uni-mod)	–	–	–	**629**	572
WEIBULL-IMPLIED ZEROMODALS (427 LCs, Mean Gini 0.5520)					
Est. min. parameter	1.1459	1.0556	3.4624	0.9394	0.4463
Est. mean parameter	1.4306	1.1969	4.2871	1.0917	0.8388
Est. max. parameter	1.5534	1.2631	9.5550	1.7400	0.9994
MSE	0.001455	0.000260	0.001493	**0.000118**	0.000785
(S.D.)	(0.000571)	(0.000253)	(0.000694)	(0.000240)	(0.000515)
MSE/Lowest MSE	12.3305	2.2034	12.6525	1.0000	6.6525
Gini difference	0.016602	0.012412	0.010441	**0.010037**	0.015237
(S.D.)	(0.010941)	(0.011227)	(0.013997)	(0.014272)	(0.015081)
Gini diff./lowest Gini diff.	1.6541	1.2366	1.0403	1.0000	1.5181
# Lowest MSE	12	3	0	**408**	4
# Lowest MSE (zero-mod)	17	**398**	0	–	12
# Lowest Gini diff.	73	57	87	**154**	56
# Lowest Gini diff. (zero-mod)	74	57	**217**	–	79

Lognormal LC obtains the best fit in terms of MSE and the Weibull in terms of the Gini difference. What is striking about the African estimation is that even among the unimodal settings, the Pareto LC achieves the lowest Gini difference. Only when the unimodality restriction is imposed, does the Lognormal LC outperform. It is also noteworthy that the Weibull LC does poorly in the African sample, in both the zeromodal and unimodal groups: it has the third-highest MSE and the highest Gini difference of all five forms. And it does not fare much better in Asia and in the Americas. In Asia, unimodal settings are best captured by the Lognormal LC, while the Pareto and Rohde forms do well for the zeromodal observations. In the Americas, MSE and Gini difference point to the Lognormal LC for both groups; after considering modality, Rohde's and Chotikapanich's forms outperform the others in the zeromodal settings. So outside Europe, the Weibull LC tends to do worse than the other forms. Obviously, its flexibility to encompass unimodal and zeromodal densities—and its ability to give a hint at the underlying modality—comes at a cost. In many circumstances the other, more specialized, forms obtain a better fit in terms of MSE or Gini difference.

It is noteworthy that the goodness of fit criteria MSE and Gini difference often point to different LCs. In fact, for only 554 out of 1,628 LCs the same

TABLE 6

Results from Fitting LCs from Different Regions to Five Parametric Forms (settings with Weibull-implied uni- and zeromodality)

Parametric LC	Pareto	Rohde	Chotikapanich	Lognormal	Weibull
UNIMODAL EUROPE (WEST) (479 out of 481 LCs, mean reported Gini 0.3109)					
MSE	0.001034	0.000152	0.000251	**0.000054**	0.000247
(S.D.)	(0.000614)	(0.000136)	(0.000249)	(0.000107)	(0.000215)
Gini difference	0.015578	0.009506	0.008211	0.004645	**0.004128**
(S.D.)	(0.006301)	(0.005427)	(0.005124)	(0.005543)	(0.005659)
UNIMODAL EUROPE (EAST) (257 out of 261 LCs, mean reported Gini 0.2882))					
MSE	0.000771	0.000145	0.000269	**0.000051**	0.000284
(S.D.)	(0.000558)	(0.000130)	(0.000260)	(0.000105)	(0.000209)
Gini difference	0.018299	0.013849	0.013511	0.009788	**0.009144**
(S.D.)	(0.013201)	(0.013177)	(0.013554)	(0.013397)	(0.013579)
UNIMODAL AFRICA (10 out of 53 LCs, mean reported Gini 0.4142))					
MSE	0.001017	0.000329	0.001091	**0.000162**	0.000846
(S.D.)	(0.000475)	(0.000148)	(0.000498)	(0.000136)	(0.000448)
Gini difference	**0.016574**	0.020897	0.022684	0.028610	0.033753
(S.D.)	(0.027227)	(0.029384)	(0.030704)	(0.030209)	(0.030922)
ZEROMODAL AFRICA (43 out of 53 LCs, mean reported Gini 0.5997))					
MSE	0.001340	0.000292	0.001861	**0.000179**	0.000869
(S.D.)	(0.000541)	(0.000457)	(0.000998)	(0.000432)	(0.000827)
Gini difference	**0.027235**	0.027647	0.031804	0.031391	0.035979
(S.D.)	(0.020746)	(0.020721)	(0.025877)	(0.025259)	(0.027548)
UNIMODAL ASIA (242 out of 284 LCs, mean reported Gini 0.3628))					
MSE	0.001005	0.000225	0.000583	**0.000109**	0.000533
(S.D.)	(0.000611)	(0.000176)	(0.000463)	(0.000151)	(0.000399)
Gini difference	0.013105	0.008290	0.007480	**0.006572**	0.008477
(S.D.)	(0.008409)	(0.007327)	(0.007190)	(0.008260)	(0.008866)
ZEROMODAL ASIA (42 out of 284 LCs, mean reported Gini 0.5208))					
MSE	0.001499	0.000328	0.001395	**0.000174**	0.000832
(S.D.)	(0.000915)	(0.000349)	(0.000793)	(0.000309)	(0.000624)
Gini difference	0.015077	**0.013929**	0.015378	0.015582	0.020701
(S.D.)	(0.012817)	(0.012605)	(0.013749)	(0.015226)	(0.016278)
UNIMODAL THE AMERICAS (213 out of 549 LCs, mean reported Gini 0.4313))					
MSE	0.001384	0.000228	0.000748	**0.000098**	0.000554
(S.D.)	(0.000487)	(0.000350)	(0.000694)	(0.000327)	(0.000616)
Gini difference	0.014899	0.009238	0.007366	**0.006553**	0.009884
(S.D.)	(0.008793)	(0.007342)	(0.007941)	(0.008887)	(0.009485)
ZEROMODAL THE AMERICAS (336 out of 549 LCs, mean reported Gini 0.5501))					
MSE	0.001463	0.000245	0.001459	**0.000102**	0.000768
(S.D.)	(0.000513)	(0.000193)	(0.000616)	(0.000189)	(0.000441)
Gini difference	0.015441	0.010322	0.007105	**0.006654**	0.011989
(S.D.)	(0.007768)	(0.007146)	(0.008286)	(0.008733)	(0.009530)

parametric form achieves both the lowest MSE and Gini difference. This is evidence that these two measures capture different aspects of the goodness of fit of an LC: a parametric LC may obtain a low MSE at the given data points; its overall form, however, may not be in line with the reported Gini coefficient. So both criteria have their virtues. However, neither unambiguously chooses LCs with the correct density modality, thus the researcher should restrict the choice of LCs upfront.

7. Conclusion

This paper has derived a relation between the LC and the modality of its density function: the third derivative of a parametric LC gives an indication of how the underlying density looks without having to derive it. The density modalities of several important uniparametric and multiparametric LCs are discussed.

Even LCs whose graphs appear similar and are associated with the same Gini coefficient can have very different density modalities. Both the Monte Carlo simulation and the empirical analysis show that LC fitting based on MSE or Gini difference can lead to an LC whose density has an incorrect modality. The resulting implications about the relative numbers of rich, middle-class and poor earners can thus be highly misleading. This paper therefore argues that researchers should limit their choice of LCs to those forms associated with the appropriate density modality (e.g., by checking the third derivative of the LC). In case the shape of the income density is unknown, more information should be gathered, or the parameter estimate of the Weibull LC can give a hint. Indeed, from our empirical analysis one may conclude that one main asset of the Weibull LC is indicating the modality, while other LCs often achieve a better fit in most unimodal or zeromodal settings.

References

Aggarwal, V., "On Optimal Aggregation of Income Distribution Data," *Sankhya*, B46, 343–55, 1984.

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

Arnold, B. C., *Pareto Distributions*, International Cooperative Publishing House, Fairland, MD, 1983.

———, *Majorization and the Lorenz Curve: A Brief Introduction*, Springer, New York, 1987.

Castañeda, A., J. Díaz-Gimémez, and J.-V. Ríos-Rull, "Accounting for the U.S. Earnings and Wealth Inequality," *Journal of Political Economy*, 111, 818–57, 2003.

Chotikapanich, D., "A Comparison of Alternative Functional Forms for the Lorenz Curve," *Economics Letters*, 41, 129–38, 1993.

Cowell, F. and M.-P. Victoria-Feser, "Robust Stochastic Dominance: A Semi-Parametric Approach," *Journal of Economic Inequality*, 5, 21–37, 2007.

Dagum, C., "A Study on the Distributions of Income, Wealth and Human Capital," *Revue Européenne des Sciences Sociales*, 113, 231–68, 1999.

Gastwirth, J. L., "A General Definition of the Lorenz Curve," *Econometrica*, 39, 1037–39, 1971.

Hasegawa, H. and H. Kozumi, "Estimation of Lorenz Curves: A Bayesian Nonparametric Approach," *Journal of Econometrics*, 115, 277–91, 2003.

Kakwani, N. C., *Income Inequality and Poverty: Methods of Estimation and Policy Applications*, Oxford University Press, Oxford, 1980.

Kakwani, N. and N. Podder, "On the Estimation of the Lorenz Curve from Grouped Observations," *International Economic Review*, 14, 278–92, 1973.

Kleiber, C., "A Guide to the Dagum Distribution," in D. Chotikapanich (ed.), *Modeling Income Distributions and Lorenz Curves (Economic Studies in Inequality)*, Springer, New York, 97–117, 2008.

Kleiber, C. and S. Kotz, *Statistical Size Distributions in Economics and Actuarial Sciences*, John Wiley & Sons, Hoboken, NJ, 2003.

Krause, M., "Corrigendum to 'Elliptical Lorenz Curves' [J. Econom. 40 (1989) 327–338]," *Journal of Econometrics*, 174, 44, 2013.

Kullback, S. and R. Leibler, "On Information and Sufficiency," *Annals of Mathematical Statistics*, 22, 79–86, 1951.

Lorenz, M. O., "Methods of Measuring the Concentration of Wealth," *Publications of the American Statistical Association*, 9, 209–19, 1905.

McDonald, J. B., "Some Generalized Functions for the Size Distribution of Income," *Econometrica*, 52, 647–63, 1984.

McDonald, J. B. and M. Ransom, "The Generalized Beta Distribution as a Model for the Distribution of Income: Estimation of Related Measures of Inequality," in D. Chotikapanich (ed.), *Modeling Income Distributions and Lorenz Curves (Economic Studies in Inequality)*, Springer, New York, 147–66, 2008.
Ortega, P., M. Fernández, M. Ladoux, and A. García, "A New Functional Form for Estimating Lorenz Curves," *Review of Income and Wealth*, 37, 447–52, 1991.
Pakes, A. G., *On Income Distributions and their Lorenz Curves*, Technical Report, Department of Mathematics, University of Western Australia, 1981.
Rasche, R., J. Gaffney, A. Koo, and N. Obst, "Functional Forms for Estimating the Lorenz Curve," *Econometrica*, 48, 1061–62, 1980.
Rohde, N., "An Alternative Functional Form for Estimating the Lorenz Curve," *Economics Letters*, 100, 61–63, 2009.
Ryu, H. K. and D. J. Slottje, "Parametric Approximations of the Lorenz Curve," in J. Silber (ed.), *Handbook of Income Inequality Measurement*, Springer, New York, 291–314, 1999.
Sarabia, J. M., "Parametric Lorenz Curves: Models and Applications," in D. Chotikapanich (ed.), *Modeling Income Distributions and Lorenz Curves (Economic Studies in Inequality)*, Springer, New York, 167–90, 2008.
Sarabia, J. M., E. Castillo, and D. Slottje, "An Ordered Family of Lorenz Curves," *Journal of Econometrics*, 91, 43–60, 1999.
———, "An Exponential Family of Lorenz Curve," *Southern Economic Journal*, 67, 748–56, 2001.
Sarabia, J. M., F. Prieto, and M. Sarabia, "Revisiting a Functional Form for the Lorenz Curve," *Economics Letters*, 107, 249–52, 2010.
Singh, S. and G. Maddala, "A Function for the Size Distribution of Incomes," *Econometrica*, 44, 963–70, 1976.
United Nations University–World Institute for Development Economics Research, *UNU-WIDER World Income Inequality Database*, Version 2.0c, 2008.
Villaseñor, J. and B. Arnold, "Elliptical Lorenz Curves," *Journal of Econometrics*, 40, 327–38, 1989.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix: Parameter combinations determining the Density Modality of the Elliptical LC (Section 4.3)