Thailand and The World Economy
P-ISSN: 2630-0931
E-ISSN: 2651-0529
Homepage : *https://www.tci-thaijo.org/index.php/TER*

Vol. 41, No.1, January – April 2023 Page [106-125]

An Alternative Functional Form for the Lorenz Curve with Empirical Applications

*Thitithep Sitthiyot**
Faculty of Commerce and Accountancy, Chulalongkorn University, Thailand

Kanyarat Holasut
Faculty of Engineering, Khon Kaen University, Thailand

Received 11 July 2021, Received in revised form 22 November 2021,
Accepted 8 December 2021, Available online 6 January 2023

Abstract

Given that the Lorenz curve is widely used for analyzing income distribution and inequality, this study introduces an alternative functional form for the Lorenz curve that is constructed based on the weighted average of the exponential function and the functional form implied by Pareto distribution. Using the data on the Gini index and the decile income shares of Thailand and other 4 countries with different income inequality, socioeconomic, and regional backgrounds, this study shows that the alternative functional form meets required criteria for a good functional form suggested by Dagum (1977). Moreover, this study compares the performance of the alternative functional form to that of Kakwani (1980). The results show that the performance of the alternative functional form is comparable to that of Kakwani (1980). However, the alternative functional form has an advantage in that the Gini index can be conveniently computed since it has an explicit mathematical solution whereas, for the Kakwani (1980)'s functional form, the Gini index is computed by using the numerical integration since its closed-form expression does not exist. Furthermore, this study finds that when the values of cumulative normalized rank of income are low, the Kakwani (1980)'s functional form does not always satisfy the monotonic increasing condition for the Lorenz curve. Thus, when applying any functional form for the Lorenz curve to analyze and formulate policy at the lower tail of income distribution, the shape of the estimated Lorenz curve should be considered together with the values of goodness-of-fit statistics and the estimated Gini index.

Keywords: Lorenz Curve; Parametric Functional Form; Income Distribution; Gini Index; Inequality; Income Share; Size Distribution
JEL Classifications: C80; D31

* **Corresponding author:** Mahitaladhibesra Bld., 10th Fl., Phayathai Rd., Pathumwan, Bangkok 10330, E-mail address: thitithep@cbs.chula.ac.th

1. Introduction

The Lorenz curve is a graphical representation of the relationship between the cumulative normalized rank of income of population from the lowest to the highest and the cumulative normalized income of this population from the lowest to the highest. It was originally developed by an American economist named Max O. Lorenz in 1905 as a method for measuring the concentration of wealth. For the past century, the Lorenz curve has been widely used as a method for illustrating size distribution of income and wealth and as a device for examining inequality in the distribution (Chotikapanich, 1993). The use of the Lorenz curve is not limited to economics, however. The application of the Lorenz curve has grown beyond economics and reached various disciplines of science (Eliazar and Sokolov, 2012). Examples include astrophysics – the analysis of galaxy morphology (Abraham, van den Bergh, & Nair, 2003); econophysics – scale invariance in the distribution of executive compensation (Sitthiyot, Budsaratragoon, & Holasut, 2020); engineering – the analysis of load feature in heating, ventilation, and air conditioning systems (Zhou, Yan, & Jiang, 2015); finance – the analysis of fluctuations in time intervals of financial data (Sazuka & Inoue, 2007); human geography – measuring differential accessibility to facilities between various segments of population (Cromley, 2019); informetrics – the analysis of citation (Bertoli-Barsotti & Lando, 2019); medical chemistry – the analysis of kinase inhibitors (Graczyk, 2007); physics – inequality indices (Eliazar, 2018); renewable and sustainable energy – the analysis of irregularity of photovoltaic power output (Das, 2014); transport geography – equity in accessing public transport (Delbosc & Currie, 2011); selection of tram links for priority treatments (Pavkova, Currie, Delbosc, & Sarvi, 2016). Focusing on the use of the Lorenz curve to analyze income distribution and its inequality, Jordá et al. (2021) pointed out that income inequality would be relatively simple to estimate if individual records on personal or household income data were available. Unfortunately, much of the existing research on economic inequality has been plagued by a lack of individual data. Nevertheless, periodic report of certain summary statistics on income distribution has become quite common. The United Nations University-World Income Inequality Database (UNU-WIID), the World Bank's PovcalNet, and the World Income Database (WID) are the largest cross-country databases that provide grouped income data, typically including the information on income and population shares. This type of grouped data depicts sparse points of the Lorenz curve which makes defining a method to connect those points an essential requisite for estimating inequality measures.

The Lorenz curve could be estimated 1) by using interpolation techniques 2) by estimating a specified functional form for income distribution and deriving the respective Lorenz curve and 3) by specifying a parametric functional form for the Lorenz curve. Given that the interpolation techniques underestimate inequality unless the individual data are available and no existing statistical distribution has proved to be adequate for representing the entire income distribution (Chotikapanich, 1993), various studies have proposed a variety of parametric functional forms in order to directly estimate the Lorenz curve. Examples are Kakwani and Podder (1973; 1976), Kakwani (1980), Rasche, Gaffney, Koo, and Obst, (1980), Aggarwal (1984), Gupta (1984), Rao and Tam (1987), Basmann, Hayes, Slottje, and Johnson (1990), Ortega, Martín, Fernández, Ladoux, and García (1991), Chotikapanich (1993), Ogwang and Rao (1996; 2000), Ryu and Slottje (1996), Sarabia (1997), Sarabia, Castillo, and Slottje (1999; 2001), Sarabia and Pascual (2002), Rohde (2009), Helene (2010), Sarabia, Prieto, and Sarabia, (2010), Sarabia, Prieto, and Jordá (2015), Wang and Smyth (2015), Sarabia, Jordá, and Trueba (2017), Paul and Shankar (2020).

According to Dagum (1977), a good parametric functional form for the Lorenz curve has to be able to describe the income distributions of different countries, regions, socioeconomic groups, and in different time periods, as well as from different sources of income through the changes in parameter values. The specified functional form should also provide a good fit of the entire range of income distribution since all observations are relevant for an accurate measurement of income inequality, supporting social and income policies, as well as determining taxation structure. Many functional forms often cited in the literature do not have a closed-form expression for the Gini index, making it computationally inconvenient to calculate since they require the valuation of the beta function such as Kakwani and Podder (1976), Kakwani (1980), Rasche, Gaffney, Koo, and Obst (1980), and Ortega, Martín, Fernández, Ladoux, and García (1991) or the confluent hyper-geometric function such as Rao and Tam (1987). A good functional form, therefore, should have an explicit mathematical solution for the Gini index. In addition, Dagum (1977) suggested that the specified functional form should use the smallest possible number of parameters for adequate and meaningful representation with well-defined economic meanings. While three- or four- parameter functional form implies a loss in simplicity, a functional form that fits the empirical observations well with an associated measure of income inequality such as the Gini index usually requires more than two parameters. Furthermore, from a viewpoint of computational cost and the acceptance of the specified functional form in applied economics, a simple method of parameter estimation is always an advantage.

Provided that earlier studies have shown that no parametric functional form for the Lorenz curve is always optimal, different attempts are therefore still worth studying (Fellman, 2018). This study introduces an alternative functional form for estimating the Lorenz curve with a closed-form expression for the Gini index. Our specified functional form has 3 parameters. It is constructed based on the weighted average of 2 well-known functional forms for the Lorenz curve which are the exponential function and the functional form implied by Pareto distribution. Using the data on the Gini index and the decile income shares of Thailand and other 4 countries, namely, Austria, Taiwan, Mexico, and Namibia, this study demonstrates that the alternative functional form meets a number of required criteria for a good functional form as noted by Dagum (1977). We also compare the performance of the alternative functional form to that of Kakwani (1980) which, based on Cheong (2002) and Tanak, Mohtashami Borzadaran, and Ahmadi (2018), has the best overall performance among different functional forms employed in approximating the Lorenz curve. The overall results indicate that on the criteria of the coefficient of determination (R^2), the mean-squared error (MSE), the mean absolute error (MAE), the maximum absolute error (MAS), the information inaccuracy measure (IIM), and the Kolmogorov-Smirnov test (K-S test), the performance of our alternative functional form is comparable to that of Kakwani (1980). However, the estimated values of the Gini index that is calculated based on the alternative functional form are closer to the actual observations than those computed according to the functional form proposed by Kakwani (1980). This is because our alternative functional form has an explicit mathematical solution for the Gini index which can be conveniently computed whereas, for the Kakwani (1980)'s functional form, the Gini index has to be calculated by using the numerical integration since its closed-form expression for the Gini index does not exist. Even though the Kakwani (1980)'s functional form fits the actual observations quite well on the basis of R^2, MSE, MAE, MAS, IIM, and the K-S test, this study shows that for a certain range of values of cumulative normalized rank of income that are close to zero which could vary from country to country, the Kakwani (1980)'s functional form does not always satisfy the monotonic increasing condition which is considered a required mathematical condition for the Lorenz curve.

This study is divided into 4 sections. Following the Introduction, Section 2 describes the method and data used in this study. Section 3 presents and discusses the results. Section 4 concludes and provides policy implications.

2. Method and Data

To demonstrate our method, let x be the cumulative normalized rank of income of population from the lowest to the highest, y be the cumulative normalized income of population from the lowest to the highest, and α, β, and P denote parameters. According to Ogwang and Rao (2000), a convex linear combination is a way to overcome an important shortcoming of functional forms for estimating the Lorenz curve which is the lack of satisfactory fit over the entire range of a given income distribution. While there are a variety of existing and already known functional forms for the Lorenz curve, as discussed in Section 1, which could be used in combination by assigning each functional form a weight between 0 and 1, such that all weights are added up to 1, our specified functional form is constructed based on the weighted average of 2 well-known functional forms, which are the exponential function ($y(x) = x^P$) and the functional form implied by Pareto distribution ($y(x) = 1 - (1-x)^{\frac{1}{P}}$) as shown in equation (1).

$$y(x) = [1 - (\alpha + \beta x)] * x^P + (\alpha + \beta x) * \left[1 - (1-x)^{\frac{1}{P}}\right], \quad (1)$$

$$0 \leq x, y \leq 1,$$

$$0 \leq \alpha, \beta \leq 1,$$

$$0 \leq \alpha + \beta \leq 1,$$

$$\alpha = \frac{a^2}{1+a^2},$$

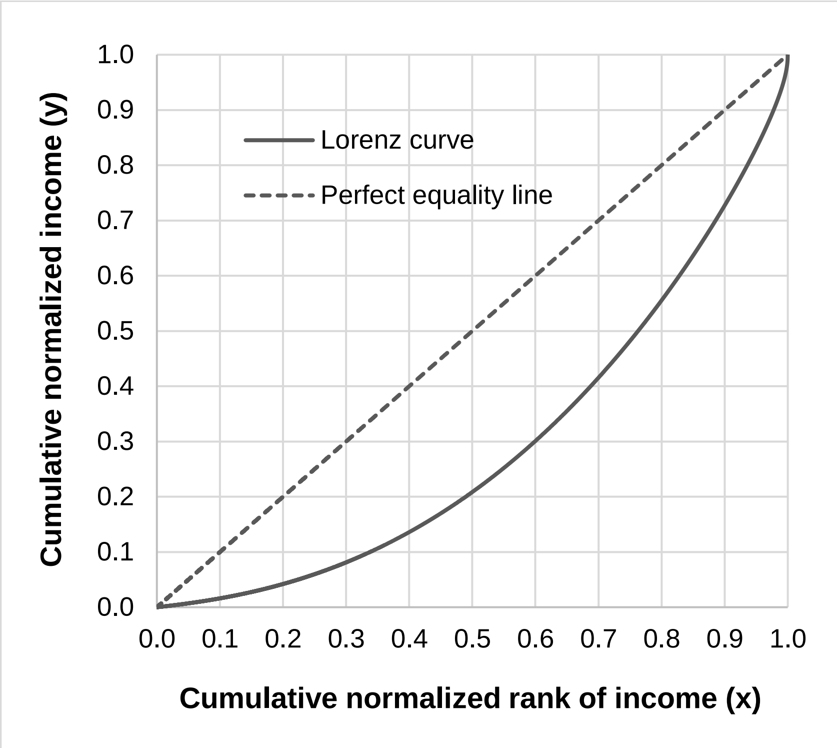
$$\beta = \left(1 - \frac{a^2}{1+a^2}\right) * \left(\frac{b^2}{1+b^2}\right),$$

$$a, b \in \mathbb{R},$$

$$1 \leq P$$

This alternative functional form satisfies all required mathematical conditions for the Lorenz curve, namely, $y\,(0) = 0$, $y\,(1) = 1$, $y(x) \leq x$, $y(x)\ is\ convex$, $\frac{dy}{dx} \geq 0$, and $\frac{d^2y}{dx^2} \geq 0$. The condition $\frac{d^2y}{dx^2} \geq 0$ also takes into account the case where everyone has the same income. The main reason that we choose the exponential function and the functional form implied by Pareto distribution, besides their simplicity, is because our specified functional form that is based on the weighted average of these 2 well-known functions could be conveniently used to derive an explicit mathematical solution for the Gini index. To our knowledge, no study has employed a functional form for estimating the Lorenz curve that is based on the combination of the exponential function and the functional form implied by Pareto distribution in this way before. The parameter P represents the degree of inequality in income distribution while the term $(\alpha + \beta x)$ is the weight that controls the convexity of the Lorenz curve. It should be noted that the parameter α is set to be equal to $\left(\frac{a^2}{1+a^2}\right)$ so that the values of α would lie between 0 and 1 whereas the parameter β is set to be equal to $\left(1-\frac{a^2}{1+a^2}\right)*\left(\frac{b^2}{1+b^2}\right)$ so that the values of $(\alpha+\beta)$ would also lie between 0 and 1. Figure 1 illustrates the Lorenz curve.

Figure 1: The Lorenz Curve.



Source: Authors' production.

Based on the alternative functional form for the Lorenz curve as shown in equation (1), we can calculate the area under the Lorenz curve by integrating equation (1) from 0 to 1 as follows:

$$\int_0^1 y(x)dx = \frac{1}{P+1} - \left[\frac{\beta*P*(P-1)}{2*(P+1)*(P+2)*(2P+1)}\right] \quad (2)$$

Given the area under the Lorenz curve as shown in equation (2), the closed-form expression for the Gini index can be computed as shown in equation (3).

$$Gini\ index = 1 - 2*\int_0^1 y(x)dx \ = \ 1 - 2*\left[\frac{1}{P+1} - \left[\frac{\beta*P*(P-1)}{2*(P+1)*(P+2)*(2P+1)}\right]\right], \quad (3)$$

$$0 \leq Gini\ index \leq 1$$

The Gini index takes the values in the unit interval. The closer the index is to 0, the more equal the income distribution whereas the closer the index is to 1, the more unequal the distribution of income. According to Sitthiyot and Holasut (2020), the advantage of the Gini index as a measure of income inequality is that the inequality of the entire income distribution could be summarized by using a single statistic that is relatively easy to interpret since its values are between 0 and 1. This allows for comparison among countries with different population sizes. Furthermore, the data on the Gini index is easy to access, regularly updated and published by countries and/or international organizations. For these reasons, the Gini index has arguably been the most popular measure of inequality despite the fact that there are over 50 inequality indices (Coulter, 1989).

To demonstrate our method, this study uses the latest data on the Gini index and the income shares by decile of Thailand and other 4 countries, namely, Austria, Taiwan, Mexico, and Namibia. The data of Thailand are from the Office of the National Economic and Social Development Council (NESDC), Thailand, whereas the data from the other 4 countries come from the United Nations University-World Income Inequality Database (UNU-WIID). These countries are chosen mainly to reflect the difference in level of income inequality, socioeconomic, and regional backgrounds. By using the specified functional form for the Lorenz curve as shown in equation (1), the parameters a, b, and P can be estimated by employing the curve fitting technique based on minimizing error sum of squares. The estimated parameters a and b are then used to compute the values of parameters α and β.

To determine how well the estimated Lorenz curve fits the actual observations, five goodness-of-fit statistics are employed. They are R^2, MSE, MAE, MAS, and IIM. The closer the value of R^2 is to 1 as well as the closer the values of MSE, MAE, and MAS are to zero, the better the estimated functional form. For the IIM criterion, the estimated functional specification that has a smaller value of IIM is better than those with larger values of IIM. In addition, the K-S test is performed in order to compare whether or not the estimated income shares by decile are statistically different from the actual observations with the null hypothesis being no difference between the two. According to Sitthiyot, Budsaratragoon, and Holasut (2020), the K-S test is commonly used to determine whether two datasets differ statistically. It has the advantage of making no assumption regarding the distribution of the data. In this study, the Microsoft Excel Solver program and the Microsoft Excel Spreadsheet program, which are available in most, if not all, computers, are employed for estimating the parameters and calculating the estimated Gini index. From a viewpoint of computational cost and the acceptance of the specified functional form for practical purposes, Dagum (1977) suggested that a simple method of parameter estimation is always an advantage.

3. Results and Discussion

3.1 The performance of the alternative functional form

Table 1 reports the estimated values of parameters a, b, α, β, and P for the Lorenz curves of Thailand and other 4 countries, namely, Austria, Taiwan, Mexico, and Namibia. The results indicate that the estimated Lorenz curves of these 5 countries fit the empirical data practically well with the values of R^2 ranging between 0.9998 and 1.0000.

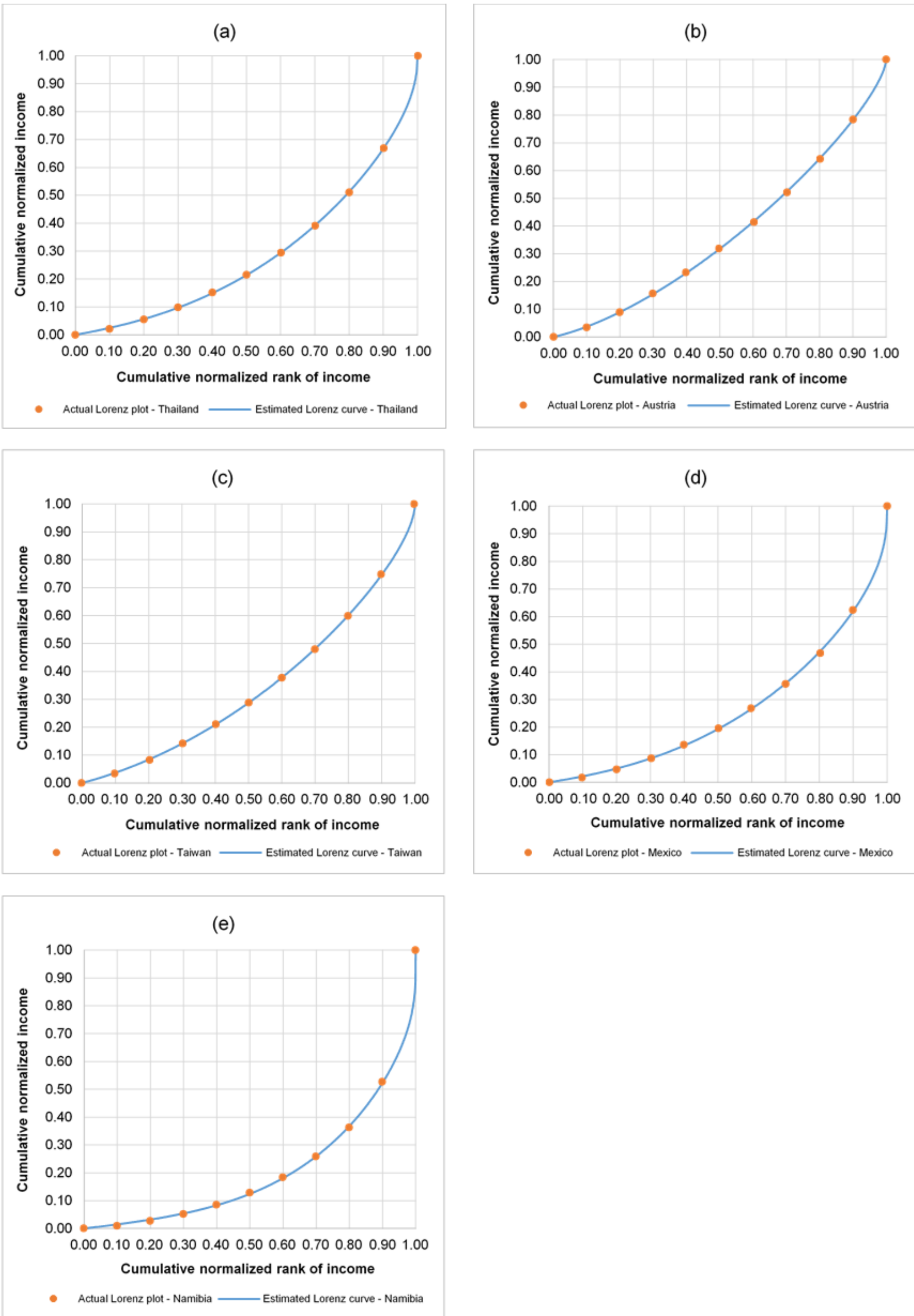
Table 1**:** The values of parameters a, b, α, β, and P for the estimated Lorenz curves of Thailand, Austria, Taiwan, Mexico, and Namibia using the alternative functional form**.**

Year	**Country**	a	b	α	β	P	R^2
2019	Thailand	1.15	0.31	0.57	0.04	2.51	1.0000
2018	Austria	0.86	0.54	0.42	0.13	1.72	1.0000
2016	Taiwan	1.21	0.06	0.59	0.002	1.92	1.0000
2018	Mexico	1.12	0.69	0.56	0.14	2.78	0.9999
2016	Namibia	1.08	0.58	0.54	0.12	3.89	0.9998

Source: Authors' calculation.

The actual Lorenz plots and the estimated Lorenz curves of Thailand and other 4 countries, namely, Austria, Taiwan, Mexico, and Namibia based on the alternative functional form that is constructed based on the weighted average of the exponential function and the functional form implied by Pareto distribution are illustrated in Figure 2**.**

Figure 2**:** The actual Lorenz plots and the estimated Lorenz curves based on the alternative functional form**.** (a) Thailand**.** (b) Austria**.** (c) Taiwan**.** (d) Mexico**.** (e) Namibia**.**



Source: Authors' production.

This study then uses the estimated parameters to calculate the income shares by decile of 5 countries. The results are reported in Table 2. All 4 goodness-of-fit statistical measures, namely, MSE, MAE, MAS, and IIM, indicate that the estimated decile income shares are close to the actual observations in all 5 countries. The results from the K-S test also confirm that the estimated income shares by decile are statistically not significant from the actual observations with *p*-value = 1.000 in all cases.

Table 2: The estimated income shares by decile calculated using the estimated Lorenz curves based on the alternative functional form.

Decile	Thailand		Austria		Taiwan		Mexico		Namibia	
	Actual	**Estimate**	**Actual**	**Estimate**	**Actual**	**Estimate**	**Actual**	**Estimate**	**Actual**	**Estimate**
D1	0.0208	0.0249	0.0340	0.0367	0.0336	0.0366	0.0170	0.0220	0.0096	0.0147
D2	0.0341	0.0316	0.0550	0.0526	0.0491	0.0470	0.0300	0.0280	0.0179	0.0174
D3	0.0434	0.0408	0.0670	0.0652	0.0590	0.0572	0.0400	0.0365	0.0250	0.0220
D4	0.0528	0.0519	0.0760	0.0763	0.0684	0.0676	0.0490	0.0470	0.0331	0.0295
D5	0.0641	0.0649	0.0860	0.0867	0.0779	0.0785	0.0600	0.0595	0.0428	0.0408
D6	0.0785	0.0800	0.0960	0.0969	0.0890	0.0903	0.0720	0.0740	0.0556	0.0567
D7	0.0966	0.0978	0.1070	0.1077	0.1022	0.1036	0.0880	0.0913	0.0742	0.0786
D8	0.1197	0.1203	0.1200	0.1205	0.1199	0.1201	0.1120	0.1136	0.1043	0.1088
D9	0.1582	0.1542	0.1430	0.1396	0.1493	0.1449	0.1560	0.1488	0.1649	0.1556
D10	0.3317	0.3336	0.2160	0.2178	0.2517	0.2540	0.3760	0.3792	0.4725	0.4759
MSE	0.000005		0.000003		0.000005		0.00001		0.00002	
MAE	0.0020		0.0015		0.0018		0.0030		0.0037	
MAS	0.0041		0.0034		0.0044		0.0072		0.0093	
IIM	0.0006		0.0002		0.0004		0.0012		0.0019	
K-S test	*D*-statistic = 0.1000 *p*-value = 1.000		*D*-statistic = 0.1000 *p*-value = 1.000		*D*-statistic = 0.1000 *p*-value = 1.000		*D*-statistic = 0.1000 *p*-value = 1.000		*D*-statistic = 0.1000 *p*-value = 1.000	
Gini	**Actual**	**Estimate**	**Actual**	**Estimate**	**Actual**	**Estimate**	**Actual**	**Estimate**	**Actual**	**Estimate**
Index	0.430	0.432	0.268	0.268	0.315	0.316	0.475	0.476	0.591	0.596

Source: Authors' calculation.

Next, we use the parameters β and P to calculate the values of the Gini index according to equation (3). We then compare the estimated values of the Gini index of each country to those from the actual observations. The results shown in Table 2 indicate that the estimated values of the Gini index of each country are virtually similar to the observed values of the Gini index as published in the NESDC and the UNU-WIID.

3.2 The performance comparison between the alternative functional form and the Kakwani (1980)'s functional form

In addition to the statistical measures of goodness-of-fit used to gauge the performance of the alternative functional form that is constructed based on the weighted average of the exponential function and the functional form implied by Pareto distribution, we compare the performance the alternative functional form to the functional form proposed by Kakwani (1980). The reason that we choose the Kakwani (1980)'s functional form is mainly because, based on Cheong (2002) and Tanak et al. (2018), it is overall superior to other functional forms used for estimating the Lorenz curves, most of which are often cited in the literature. Using the same notations for the cumulative normalized rank of income (x) and the cumulative normalized income (y) as denoted in Section 2, the Kakwani (1980)'s functional form for approximating the Lorenz curve is as follows:

$$y(x) = x - a * x^{\gamma} * (1 - x)^{\delta}, \tag{4}$$

$$a > 0,$$

$$0 < \gamma \leq 1,$$

$$0 < \delta \leq 1$$

Given the functional form as shown in equation (4), the Gini index could be approximated as $2a\ B\ (\gamma+1,\ \delta+1)$, where B is the beta function (Cheong, 2002) since Kakwani (1980)'s functional form does not have an explicit mathematical solution for the Gini index.

To estimate the Lorenz curve, this study employs the same dataset of Thailand and other 4 countries which are Austria, Taiwan, Mexico, and Namibia and uses the Kakwani (1980)'s functional form along with the curve fitting technique based on minimizing sum of squared errors. The estimated parameters a, γ, and δ including the values of R^2 are reported in Table 3. On the basis of the value of R^2, the Kakwani (1980)'s functional form performs just slightly better than our alternative functional form as reported in Table 1.

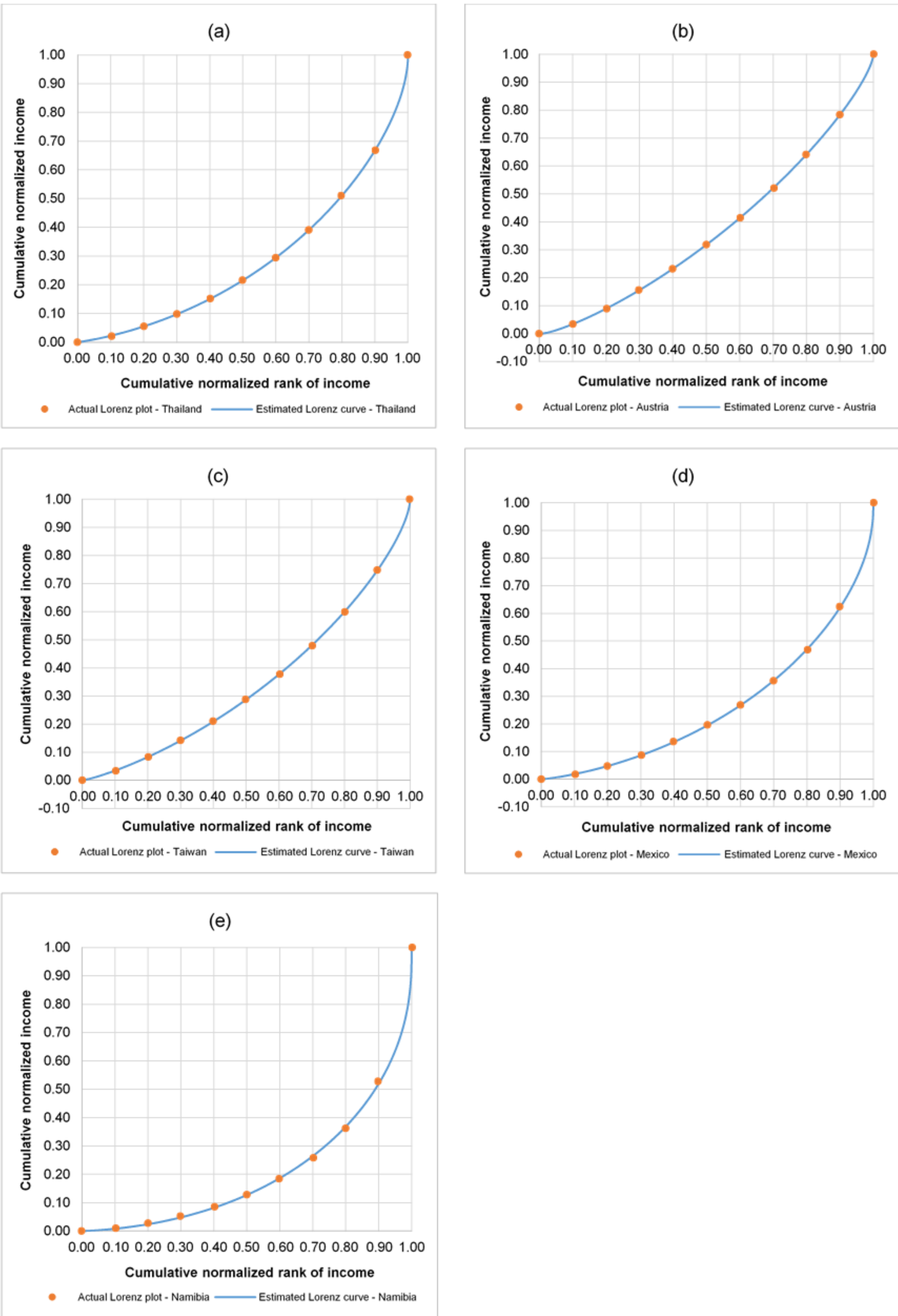
Table 3: The estimated parameters a, γ, and δ for the Lorenz curves of Thailand, Austria, Taiwan, Mexico, and Namibia using the Kakwani (1980)'s functional form.

Year	**Country**	$\boldsymbol{a}$	$\boldsymbol{\gamma}$	$\boldsymbol{\delta}$	$\mathbf{R^2}$
2019	Thailand	0.79	0.98	0.49	1.0000
2018	Austria	0.49	0.84	0.59	1.0000
2016	Taiwan	0.59	0.93	0.54	1.0000
2018	Mexico	0.80	0.97	0.41	1.0000
2016	Namibia	0.95	1.00	0.35	0.9998

Source: Authors' calculation.

Figure 3 illustrates the actual Lorenz plots and the estimated Lorenz curves of Thailand and other 4 countries which are Austria, Taiwan, Mexico, and Namibia based on the Kakwani (1980)'s functional form.

Figure 3: The actual Lorenz plots and the estimated Lorenz curves based on the Kakwani (1980)'s functional form. (a) Thailand. (b) Austria. (c) Taiwan. (d) Mexico. (e) Namibia.



Source: Authors' production.

Next, the estimated parameters based on the Kakwani (1980)'s functional form are used to calculate the income shares by decile of 5 countries. This study then compares the income shares by decile estimated using the Kakwani (1980)'s functional form to those estimated using the alternative functional form that is constructed based on the weighted average of the exponential function and the functional form implied by Pareto distribution. The results are reported in Table 4.

The overall results show that there are no significant differences among the values of income shares by decile estimated using the alternative functional form and the Kakwani (1980)'s functional form. Table 5 reports the comparison of the values of goodness-of-fit statistics between the alternative functional form and the functional form proposed by Kakwani (1980).

The values of goodness-of-fit statistics shown in Table 5 indicate that, with the exception of the K-S test which gives identical results in that the differences between the estimated income shares by decile and the actual observations are not statistically significant with *p*-value equal to 1.000 in all cases, the performance of the Kakwani (1980)'s functional form is slightly better than that of the alternative functional form on the basis of the values of MSE, MAE, MAS, and IIM.

When using both functional forms to calculate the Gini index, the estimated values of the Gini index computed using the alternative functional form are closer to the actual observations than those calculated using the functional form proposed by Kakwani (1980). It should be noted that, for the alternative functional form that we introduce, the value of the Gini index can be conveniently computed since it has an explicit mathematical solution for the Gini index as shown in equation (3) in Section 2 whereas, for the Kakwani (1980)'s functional form, the value of the Gini index has to be calculated by using the numerical integration since its closed-form expression for the Gini does not exist. The comparison of the estimated values of the Gini index between the alternative functional form and the functional form proposed by Kakwani (1980) is reported in Table 6.

Table 4**:** The comparison of the estimated income shares by decile between the alternative functional form and the Kakwani **(**1980**)’**s functional form**.**

Decile	**Thailand**			**Austria**			**Taiwan**			**Mexico**			**Namibia**		
	Actual	**Alternative**	**Kakwani (1980)**	**Actual**	**Alternative**	**Kakwani (1980)**	**Actual**	**Alternative**	**Kakwani (1980)**	**Actual**	**Alternative**	**Kakwani (1980)**	**Actual**	**Alternative**	**Kakwani (1980)**
D1	0.0208	0.0249	0.0223	0.0340	0.0367	0.0339	0.0336	0.0366	0.0345	0.0170	0.0220	0.0184	0.0096	0.0147	0.0083
D2	0.0341	0.0316	0.0326	0.0550	0.0526	0.0554	0.0491	0.0470	0.0484	0.0300	0.0280	0.0293	0.0179	0.0174	0.0158
D3	0.0434	0.0408	0.0424	0.0670	0.0652	0.0666	0.0590	0.0572	0.0583	0.0400	0.0365	0.0387	0.0250	0.0220	0.0242
D4	0.0528	0.0519	0.0531	0.0760	0.0763	0.0763	0.0684	0.0676	0.0680	0.0490	0.0470	0.0486	0.0331	0.0295	0.0338
D5	0.0641	0.0649	0.0650	0.0860	0.0867	0.0858	0.0779	0.0785	0.0782	0.0600	0.0595	0.0598	0.0428	0.0408	0.0453
D6	0.0785	0.0800	0.0790	0.0960	0.0969	0.0958	0.0890	0.0903	0.0895	0.0720	0.0740	0.0730	0.0556	0.0567	0.0593
D7	0.0966	0.0978	0.0963	0.1070	0.1077	0.1070	0.1022	0.1036	0.1029	0.0880	0.0913	0.0896	0.0742	0.0786	0.0776
D8	0.1197	0.1203	0.1196	0.1200	0.1205	0.1211	0.1199	0.1201	0.1202	0.1120	0.1136	0.1127	0.1043	0.1088	0.1041
D9	0.1582	0.1542	0.1573	0.1430	0.1396	0.1418	0.1493	0.1449	0.1469	0.1560	0.1488	0.1521	0.1649	0.1556	0.1513
D10	0.3317	0.3336	0.3323	0.2160	0.2178	0.2164	0.2517	0.2540	0.2529	0.3760	0.3792	0.3778	0.4725	0.4759	0.4804

Source: Authors’ calculation.

Table 5**:** The comparison of the values of goodness-of-fit statistical measures between the alternative functional form and the Kakwani (1980)'s functional form.

Goodness-of-Fit Measures	Thailand		Austria		Taiwan		Mexico		Namibia	
	Alternative	**Kakwani (1980)**	**Alternative**	**Kakwani (1980)**	**Alternative**	**Kakwani (1980)**	**Alternative**	**Kakwani (1980)**	**Alternative**	**Kakwani (1980)**
MSE	0.000005	0.0000008	0.000003	0.0000003	0.000005	0.000001	0.00001	0.000003	0.00002	0.00003
MAE	0.0020	0.0007	0.0015	0.0004	0.0018	0.0008	0.0030	0.0013	0.0037	0.0036
MAS	0.0041	0.0015	0.0034	0.0012	0.0044	0.0024	0.0072	0.0039	0.0093	0.0136
IIM	0.0006	0.00011	0.0002	0.00001	0.0004	0.0001	0.0012	0.0002	0.0019	0.0011
K-S test										
- *D*-statistic	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000
- *p*-value	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Source: Authors' calculation.

Table 6**:** The comparison of the estimated values of the Gini index between the alternative functional form and the Kakwani (1980)'s functional form.

Gini Index	Thailand			Austria			Taiwan			Mexico			Namibia		
	Actual	**Alternative**	**Kakwani (1980)**	**Actual**	**Alternative**	**Kakwani (1980)**	**Actual**	**Alternative**	**Kakwani (1980)**	**Actual**	**Alternative**	**Kakwani (1980)**	**Actual**	**Alternative**	**Kakwani (1980)**
	0.430	0.432	0.419	0.268	0.268	0.262	0.315	0.316	0.308	0.475	0.476	0.460	0.591	0.596	0.574

Source: Authors' calculation.

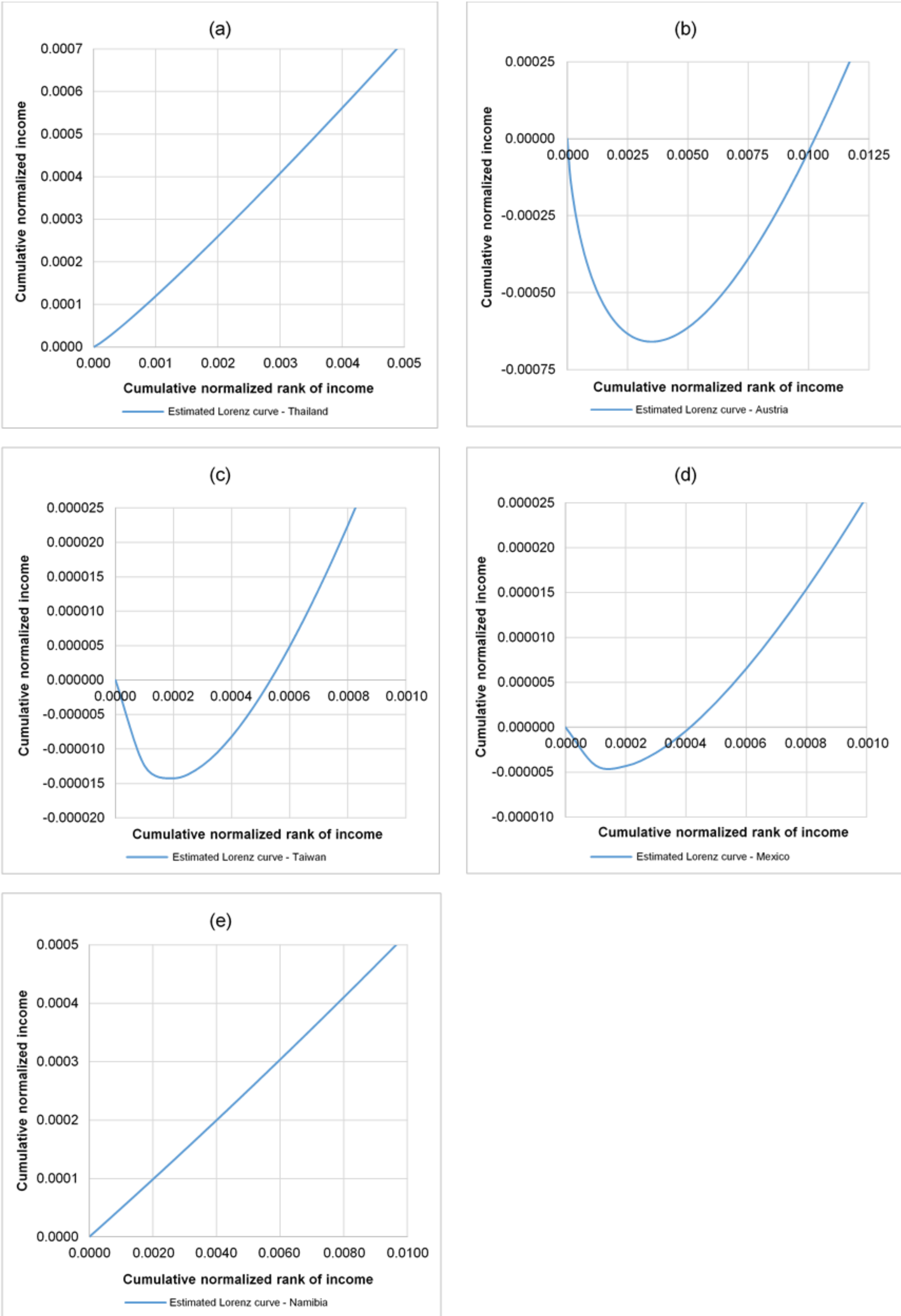
It is important to note that the estimated values of the Gini index of Austria, Taiwan, and Mexico based on the Kakwani (1980)'s functional form are calculated by rescaling the originally estimated values of cumulative normalized income (y) so that all values would lie between 0 and 1 since there are some negative values of cumulative normalized income (y) for a given range of values of cumulative normalized rank of income (x) that are close to zero. The Lorenz curves estimated using the Kakwani (1980)'s functional form which result in the negative values of cumulative normalized income (y) are illustrated and discussed in details in the next Section.

3.3 The analysis of the shape of the estimated Lorenz curves based on the Kakwani (1980)'s functional form when the value of cumulative normalized rank of income is close to zero

For a parametric functional form to be appropriately used to approximate the Lorenz curve, it must satisfy all required mathematical conditions, namely, $y\ (0) = 0$, $y\ (1) = 1$, $y(x) \leq x$, $y(x)\ is\ convex$, $\frac{dy}{dx} \geq 0$, and $\frac{d^2y}{dx^2} \geq 0$. Provided that the Lorenz curve has to start from the origin, if the data on income contain some negative values, all income data are typically rescaled and then normalized so that their values would lie between 0 and 1.

Using the same dataset on the Gini index and the income shares by decile of Thailand and other 4 countries which are Austria, Taiwan, Mexico, and Namibia, this study examines the shape of the Lorenz curves of these 5 countries estimated using the Kakwani (1980)'s functional form. We find that when the values of the cumulative normalized rank of income (x) are close to zero, the Kakwani (1980)'s functional form does not always satisfy one of the monotonic increasing conditions which is $\frac{dy}{dx} \geq 0$. It depends upon the values of the estimated parameters a, γ, and δ as well as the range of values of cumulative normalized rank of income (x), all of which vary from country to country. Our findings are shown in Figure 4.

Figure 4**:** The shape of the estimated Lorenz curves based on the Kakwani **(**1980**)'**s functional form when the values of cumulative normalized rank of income are close to zero**.** **(**a**)** Thailand**.** **(**b**)** Austria**.** **(**c**)** Taiwan**.** **(**d**)** Mexico**.** **(**e**)** Namibia**.**



Source: Authors' production.

These findings indicate that, with the exception of Thailand and Namibia, when the values of the cumulative normalized rank of income (x) are close to zero, the Lorenz curves of the other 3 countries, namely, Austria, Taiwan, and Mexico, estimated using the Kakwani (1980)'s functional form do not satisfy one of the monotonic increasing conditions since $\frac{dy}{dx} < 0$. This implies that when evaluating a functional form for the Lorenz curve, the shape of the estimated Lorenz curve should be taken into consideration in addition to the values of the goodness-of-fit statistical measures and the estimated value of the Gini index since a good functional form should satisfy all required mathematical conditions for the Lorenz curve.

4. Conclusions and Policy Implications

Finding a good functional form for the Lorenz curve is a theoretical and practical challenge. This study introduces an alternative parametric functional form for estimating the Lorenz curve that is constructed based on the weighted average of the 2 well-known functional forms, namely, the exponential function and the functional form implied by Pareto Distribution. Even though there are a variety of existing and already known parametric functional forms for estimating the Lorenz curve, which could be used in combination by assigning each functional form a weight between 0 and 1, such that all weights are added up to 1, to our knowledge, no study has employed a parametric functional form for approximating the Lorenz curve by combining the exponential function and the functional form implied by Pareto distribution in this way before. This alternative functional form satisfies several of the required criteria for a good functional form as suggested by Dagum (1977) in that all parameters have well-defined economic meanings. In addition, the Gini index can be computationally convenient to calculate since our functional form has a closed-form expression for the Gini index. Moreover, using the data on the Gini index and the income shares by decile of Thailand and other 4 countries which are Austria, Taiwan, Mexico, and Namibia from the NESDC and the UNU-WIID, this study shows that, on the criteria of R^2, MSE, MAE, MAS, IIM, and the K-S test, the alternative functional form fits the whole range of actual income distributions of countries that have different degree of income inequality, socioeconomic, and regional backgrounds reasonably well. The estimated values of the Gini index of all 5 countries are also virtually similar to the actual observations. Furthermore, the computational cost is relatively low since all parameters and the Gini index could be estimated and computed by using the Microsoft Excel Solver program and the Microsoft Excel Spreadsheet program which are available in most, if not all, computers.

In addition, this study compares the performance of the alternative functional form to the functional form proposed by Kakwani (1980) which, according to Cheong (2002) and Tanak et al. (2018), has the best overall performance among different existing functional forms for approximating the Lorenz curve often cited in the literatures. While the overall results indicate that, on the basis of R^2, MSE, MAE, MAS, IIM, and the K-S test, the performance of our alternative functional form is comparable to that of Kakwani (1980), the estimated values of the Gini index of all 5 countries calculated according to our alternative functional form are practically closer to the actual observations than that of the Kakwani (1980). Moreover, our specified functional form has an advantage over the Kakwani (1980)'s functional form in that it has an explicit mathematical solution for the Gini index that is computationally convenient to calculate whereas the estimated values of the Gini index have to be calculated by using the numerical integration since its explicit mathematical solution does not exist for the Kakwani (1980)'s functional form. Furthermore, this study finds that when the values of cumulative normalized rank of income are close to zero, the functional form proposed by Kakwani (1980) does not always satisfy the monotonic increasing condition $\left(\frac{dy}{dx} \geq 0\right)$ which is considered the required mathematical condition for the Lorenz curve. Thus, the shape of the Lorenz curve that satisfies all mathematical conditions should be considered as one of the criteria for evaluating and comparing the performance of different parametric functional forms for estimating the Lorenz curve in addition to the values of goodness-of-fit statistical measures and the estimated value of the Gini index. As shown in this study, applying the Lorenz curve that is estimated based on the parametric functional form which could possibly yield negative values for the cumulative normalized income for a certain range of values of the cumulative normalized rank of income at the lower tail of distribution could have effects on the measurement of income inequality, social and income policy formulation, as well as tax structure determination.

Finally, given the extensive use of the Lorenz curve and the Gini index in various scientific disciplines besides economics (Eliazar & Sokolov, 2012), we hope that the alternative parametric functional form that is constructed based on the weighted average of the exponential function and the functional form implied by Pareto distribution and a closed-form expression for the Gini index could be useful for measuring size distributions of non-negative quantities and inequality.

Acknowledgements

T.S. is grateful to Dr. Suradit Holasut for guidance and comments.

References

Abraham, R. G., van den Bergh, S., & Nair, P. (2003). A new approach to galaxy morphology. I. Analysis of the Sloan Digital Sky Survey early data release. *The Astrophysical Journal*, *588*(1), 218-229.

Aggarwal, V. (1984). On optimum aggregation of income distribution data. *Sankhyā: The Indian Journal of Statistics, Series B*, *46*(3), 343-355.

Basmann, R. L., Hayes, K., Slottje, D., & Johnson, J. (1990). A general functional form for approximating the Lorenz curve. *Journal of Econometrics*, *92*, 727-744.

Bertoli-Barsotti, L., & Lando, T. (2019). How mean rank and mean size may determine the generalised Lorenz curve: With application to citation analysis. *Journal of Informetrics*, *13*, 387-396.

Chakrabarti, B. K., Chakraborti, A., Chakravarty, S. R., & Chatterjee, A. (2013). *Econophysics of income and wealth distributions*. New York: Cambridge University Press.

Cheong, K. S. (2002). An empirical comparison of alternative functional forms for the Lorenz curve. *Applied Economics Letters*, *9*, 171-176.

Chotikapanich, D. (1993). A comparison of alternative functional forms for the Lorenz curve. *Economics Letters*, *41*, 129-138.

Coulter, P. B. (1989). *Measuring inequality: A methodological handbook*. Boulder: Westview Press.

Cromley, G. A. (2019). Measuring differential access to facilities between population groups using spatial Lorenz curves and related indices. *Transactions in GIS*, *23*(6), 1332-1351.

Dagum, C. (1977). A new model of personal income distribution: Specification and estimation. In D. Chotikapanich (Ed.), *Modeling income distributions and Lorenz curves. Economic studies in equality, social exclusion and well-being, vol 5*. (pp. 3-25). New York: Springer.

Das, A. K. (2014). Quantifying photovoltaic power variability using Lorenz curve. *Journal of Renewable and Sustainable Energy*, *6*, 033124.

Delbosc, A., & Currie, G. (2011). Using Lorenz curves to access public transport equity. *Journal of Transport Geography*, *19*, 1252-1259.

Eliazar, I. I. (2018). A tour of inequality. *Annals of Physics*, *389*, 306-332.

Eliazar, I. I., & Sokolov, I. M. (2012). Measuring statistical evenness: A panoramic overview. *Physica A: Statistical Mechanics and its Applications*, *391*, 1323-1353.

Fellman, J. (2018). Income inequality measures. *Theoretical Economics Letters*, *8*, 557-574.

Graczyk, P. P. (2007). Gini coefficient: A new way to express selectivity of kinase inhibitors against a family of kinases. *Journal of Medicinal Chemistry*, *50*(23), 5773-5779.

Gupta, M. R. (1984). Functional form for estimating the Lorenz curve. *Econometrica*, *52*, 1313-1314.

Helene, O. (2010). Fitting Lorenz curves. *Economics Letters*, *108*, 153-155.

Jordá V., Sarabia J. M., & Jäntti, M. (2021). Inequality measurement with grouped data: Parametric and non-parametric methods. *Journal of the Royal Statistical Society: Statistics in Society, Series A*, *184*, 964-984.

Kakwani, N. C. (1980). On a class of poverty measures. *Econometrica*, *48*, 437-446.

Kakwani N. C., & Podder N. (1973). On the estimation of Lorenz curves from grouped observations. *International Economic Review*, *14*, 278-292.

Kakwani, N. C., & Podder, N. (1976). Efficient estimation of the Lorenz curve and associated inequality measures from grouped observations. *Econometrica*, *44*, 137-148.

Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, *9*(70), 209-219.

Ogwang, T., & Rao, U.L.G. (1996). A new functional form for approximating the Lorenz curve. *Economics Letters*, *52*, 21-29.

Ogwang, T., & Rao, U.L.G. (2000). Hybrid models of the Lorenz curve. *Economics Letters*, *69*, 39-44.

Ortega, P., Martín, G., Fernández, A., Ladoux, M., & García, A. (1991). A new functional form for estimating Lorenz curves. *Review of Income and Wealth*, *37*, 47-452.

Paul, S., & Shankar, S. (2020). An alternative single parameter functional form for Lorenz curve. *Empirical Economics*, *59*, 1393-1402.

Pavkova, K., Currie, G., Delbosc, A., & Sarvi, M. (2016). Selecting tram links for priority treatments – The Lorenz curve approach. *Journal of Transport Geography*, *55*, 101-109.

Rao, U. L. G., & Tam, A. Y-P. (1987). An empirical study of selection and estimation of alternative models of the Lorenz curve. *Journal of Applied Statistics*, *14*, 275-280.

Rasche, R. H., Gaffney, J. M., Koo, A. Y. C., & Obst, N. (1980). Functional forms for estimating the Lorenz curve. *Econometrica*, *48*, 1061-1062.

Rohde, N. (2009) An alternative functional form for estimating the Lorenz curve. *Economics Letters*, *105*, 61-63.

Ryu, H., & Slottje, D. (1996). Two flexible functional forms for approximating the Lorenz curve. *Journal of Econometrics*, *72*, 251-274.

Sarabia, J. M. (1997). A hierarchy of Lorenz curves based on the generalized Tukey's lambda distribution. *Econometric Reviews*, *16*, 305-320.

Sarabia, J. M., Castillo, E., & Slottje, D. (1999). An ordered family of Lorenz curves. *Journal of Econometrics*, *91*, 43-60.

Sarabia, J. M., Castillo, E., & Slottje, D. (2001). An exponential family of Lorenz curves. *Southern Economic Journal*, *67*, 748-756.

Sarabia, J. M., & Pascual, M. (2002). A class of Lorenz curves based on linear exponential loss functions. *Communications in Statistics – Theory and Methods*, *31*, 925-942.

Sarabia, J. M., Prieto, F., & Sarabia, M. (2010). Revisiting a functional form for the Lorenz curve. *Economics Letters*, *107*, 249-252.

Sarabia, J. M., Prieto, F., & Jordá, V. (2015). About the hyperbolic Lorenz curve. *Economics Letters*, *136*, 42-45.

Sarabia, J. M., Jordá, V., & Trueba, C. (2017). The lame class of Lorenz curves. *Communications in Statistics – Theory and Methods*, *46*, 5311-5326.

Sazuka N., & Inoue, J. (2007). Fluctuations in time intervals of financial data from the view point of the Gini index. *Physica A: Statistical Mechanics and its Applications*, *383*, 49-53.

Sitthiyot, T., Budsaratragoon, P., & Holasut, K. (2020). A scaling perspective on the distribution of executive compensation. *Physica A: Statistical Mechanics and its Applications*, *543*, 123556.

Sitthiyot, T., & Holasut, K. (2020). A simple method for measuring inequality. *Palgrave Communications*, *6*, 112.

Tanak, A. K., Mohtashami Borzadaran, G. R., & Ahmadi, J. (2018). New functional forms of Lorenz curves by maximizing Tsallis entropy of income share function

under the constraint on generalized Gini index. *Physica A: Statistical Mechanics and its Applications*, *511*, 280-288.
The Office of the National Economic and Social Development Council. (2021). Poverty and Income Distribution Statistics [dataset]. Retrieved from https://www.nesdc.go.th/main.php?filename=social
The United Nations University World Institute for Development Economics Research (UNU-WIDER). (2020). World Income Inequality Database (WIID), Version date: 6 May 2020 [dataset]. Retrieved from https://www4.wider.unu.edu/
Wang, Z., & Smyth, R. (2015). A hybrid method for creating Lorenz curves. *Economics Letters*, *133*, 59-63.
Zhou, X., Yan, D., & Jiang, Y. (2015). Application of Lorenz curve and Gini index in the analysis of load feature in HVAC systems. *Procedia Engineering*, *121*, 11-18.