Relation Between Two Income Inequality Measures: The Gini coefficient and the Robin Hood Index

EDWARD ALLEN
Department of Mathematics and Statistics
Texas Tech University
Lubbock, Texas 79409-1042
UNITED STATES

Abstract: The objective of this investigation is to study the relation between two common measures of income inequality, the Gini coefficient and the Robin Hood index. An approximate formula for the Robin Hood index in terms of the Gini coefficient is developed from 100,000 Lorenz curves that are randomly generated based on 100 twenty-parameter families of income distributions. The approximate formula is tested against Robin Hood indexes of commonly-used one-parameter Lorenz curves, income data of several countries, and reported results of Robin Hood indexes. The approximate formula is also tested against results of a stochastic income-wealth model that is introduced in the present investigation. The formula is useful conceptually in understanding why Gini coefficients and Robin Hood indexes are correlated in distribution data and is useful practically in providing accurate estimates of Robin Hood indexes when Gini coefficients are known. The continuous piecewise-linear approximation is generally within 5% of standard one-parameter Lorenz curves and income distribution data and has the form: $R \approx 0.74G$ for $0 \leq G \leq 0.5$, $R \approx 0.37 + 0.90(G - 0.5)$ for $0.5 \leq G \leq 0.8$, and $R \approx 0.64 + 1.26(G - 0.8)$ for $0.8 \leq G \leq 0.95$, where R is the Robin Hood index and G is the Gini coefficient.

Key-Words: Robin Hood index, Gini coefficient, income inequality, Lorenz curve, Pietra index, Hoover index, Schutz index

Received: June 19, 2021. Revised: February 6, 2022. Accepted: February 21, 2022. Published: March 8, 2022.

1 Introduction

The Gini coefficient was introduced by Corrado Gini in 1912. Gini proposed that the area between two curves describing equal and actual incomes be used as a measure of inequality [1, 2]. The Gini coefficient ranges from 0 to 1 and is a measure of how far a given income distribution differs from a distribution of complete equality. The Robin Hood index, a second measure of income inequality, is known under many names, in particular, the Pietra index, the Hoover index, and the Schutz index. The Pietra index was introduced as an inequality measure in 1915 [3, 4]. Inequality measures equivalent to the Pietra index were proposed several times in later investigations, in particular, by Hoover [5, 6] in 1936, by Schutz [7, 8] in 1951, and as the Robin Hood index by Atkinson and Micklewright [9] in 1992.

The Gini coefficient is perhaps the most commonly-used measure of income inequality [10, 11]. The Robin Hood index is useful conceptually and is equal to the proportion of the population's total income that, if redistributed, would give perfect income equality. An approximate formula for the Robin Hood index in terms of the Gini coefficient is useful in achieving additional understanding of an income distribution when the Gini coefficient is known. In the present investigation, a continuous piecewise-linear approximation of the Robin Hood index in terms of the Gini coefficient is determined for Gini coefficients between 0.0 and 0.95. The form of the approximate formula is indicated from mathematical examination, improved through parameter value estimation of randomly generated income distributions, and tested against standard Lorenz curves and a stochastic income model. The resulting approximation is generally accurate to within 5% of Robin Hood indexes of standard one-parameter Lorenz curves and income distribution data.

2 Problem Formulation

Let $0 \leq x \leq 1$ be the cumulative proportion of the population ranked by income level. Let $0 \leq I(x) < \infty$ be the income distribution, i.e., the proportion of the population with income less than or equal to $I(x)$ is equal to x. It follows that $\int_0^x I(u)\,du / \int_0^1 I(u)\,du$ is the proportion of the total income earned by the fraction x of the population with the lowest income and $\int_0^1 I(u)\,du$ is the average income of the population. As an example, if income is distributed exponentially in a population with density $p(I) = \gamma \exp(-\gamma(I - I_1))$ for $I_1 \leq I < \infty$, then $I(x)$ has the form $I(x) = I_1 - \frac{1}{\gamma}\log(1-x)$ for $0 \leq x < 1$.

The Lorenz function, $L(x)$, of the population is defined as [12, 13, 14]:

$$L(x) = \frac{\int_0^x I(u)\,du}{\int_0^1 I(u)\,du}. \quad (1)$$

The Lorenz function, $L(x)$, is a continuous non-decreasing function for $0 < x < 1$ and satisfies, for example,

$$L(0) = 0,\ L(1) = 1,\ \text{and}\ L'(x) \geq 0,\ \text{for}\ 0 < x < 1.$$

Conceptually, $L(x)$ is the fraction of the total income for the proportion x of the population having the lowest incomes. For a population with complete income equality, $I(x)$ is constant for $0 \leq x \leq 1$ and so $L(x) = x$. Otherwise, $L(x) \leq x$ for $0 \leq x \leq 1$. The area between the curves $y = x$ and $y = L(x)$ is defined to be $G/2$ where $0 \leq G \leq 1$ is the Gini coefficient [13, 15]. The Gini coefficient G is a measure of the income inequality in the population with complete equality if $G = 0$.

The Robin Hood index is the proportion of the total income that can be redistributed to achieve income equality. The Robin Hood index, R, is equal to the maximum difference between the curves $y = x$ and $y = L(x)$. That is, for $x - L(x)$ achieving its maximum value at $x^* \in (0,1)$, then

$$R = x^* - L(x^*)\ \text{where}\ L'(x^*) = \frac{I(x^*)}{\int_0^1 I(x)\,dx} = 1. \quad (2)$$

By Eq. (2), x^* satisfies $I(x^*) = \int_0^1 I(u)\,du$ and so, $\left(\int_{x^*}^1 I(x)\,dx - (1-x^*)I(x^*)\right) / \int_0^1 I(x)\,dx$ is the proportion of total income to distribute so that the entire population has income $I(x^*)$. But $\int_{x^*}^1 I(x)\,dx - (1-x^*)I(x^*) = R\int_0^1 I(x)\,dx$, so the proportion of the total income to distribute from incomes above $I(x^*)$ to incomes below $I(x^*)$ is equal to the Robin Hood index[1] $R = x^* - L(x^*)$.

As $G/2 = \int_0^1 (x - L(x))\,dx \leq \int_0^1 (x^* - L(x^*))dx = R$, it is clear that G and R are closely related. Indeed, by considering the area of the triangle defined by the vertices $(0,0), (1,1)$, and $(x^*, L(x^*))$, it can also be shown [13] that $R \leq G$. Thus, the Robin Hood index satisfies the rough bounds

$$G/2 \leq R \leq G.$$

Given a value of G, there are many different Lorenz curves and, correspondingly, there are many possible values of R. In the remainder of this investigation, given a value of G, it is shown that the values of R are generally within 10% of each other. To study this, many families of Lorenz curves are examined. One-parameter families of Lorenz functions considered in the present investigation are assumed to have G as the specified parameter. That is, each curve of a one-parameter family of Lorenz functions is uniquely defined by the value of the Gini coefficient G on some interval of $[0,1]$. In addition, any n-parameter family of Lorenz functions considered yields a unique Lorenz curve when n permissible parameter values are given. A Lorenz function from a one-parameter family is written in the present investigation as $L(x,G)$ and the Robin Hood coefficient for the Lorenz function is written $R(G)$.

Consider $R(G)$ for a one-parameter Lorenz curve $L(x,G)$. From Eq. (2), the Robin Hood index is a function of G and satisfies

$$R(G) = x^* - L(x^*,G),\ \text{where}\ \frac{\partial L(x^*,G)}{\partial x} = 1. \quad (3)$$

Of interest in the present investigation is how $R(G)$ is approximately related to G. Approximations for $R(G)$ can be inferred by Taylor series expansions of $R(G)$ about the G variable. In particular, for small values of G, say $0 \leq G \leq G_1$,

$$R(G) \approx R(0) + R_G(0)\,G,\ \text{for}\ 0 \leq G \leq G_1,$$

and for larger G, say $G_1 \leq G \leq G_2$ and $G_2 \leq G \leq G_3$,

$$R(G) \approx \begin{cases} R(G_1) + R_G(G_1)(G - G_1), & G_1 \leq G \leq G_2, \\ R(G_2) + R_G(G_2)(G - G_2), & G_2 \leq G \leq G_3. \end{cases}$$

Equating by continuity the two expressions at $G = G_1$ and $G = G_2$ and setting $R(0) = 0$, the following continuous piecewise-linear approximation[2] is obtained:

$$R(G) \approx \begin{cases} \alpha G,\ \text{for}\ 0 \leq G \leq G_1, \\ \alpha G_1 + \beta_1(G - G_1),\ \text{for}\ G_1 \leq G \leq G_2, \\ \alpha G_1 + \beta_1(G_2 - G_1) + \beta_2(G - G_2), \\ \quad \text{for}\ G_2 \leq G \leq G_3, \end{cases} \quad (4)$$

where α, β_1, and β_2 are constants.

An initial estimate of α can be made by considering the Lorenz function $L(x,G)$. The values of β_1, β_2 depend on the income distribution, i.e., the family of Lorenz curves, and estimation of β_1 and β_2 is considered in the next section. To estimate α, it is assumed that $L(x,G)$ is approximately equal to x for G small and so, $L(x,G) = x + \varepsilon(x,G)$ for $0 \leq x \leq 1$ where $\varepsilon(x,G)$ is small. Assuming additionally that

[1] As Robin Hood took from the rich and gave to the poor, a second possible Robin Hood index would be $\hat{R} = (x^* - L(x^*))/(1 - L(x^*)) = R/(1 - L(x^*))$ which is the proportion of all income above the average income that, if redistributed to individuals with income below the average income, would result in complete income equality. For example, for the Pareto Lorenz function, $\hat{R}$ has the form $\hat{R} = 2G/(1+G)$.

[2] Continuous piecewise-linear approximations are common in applications [16].

$\varepsilon(x,G)$ is approximately a quadratic function of x for G small, then

$$\begin{aligned} L(x,G) &\approx x + a_1(G) + a_2(G)x + b(G)x^2 \\ &= (1-b(G))x + b(G)x^2, \end{aligned} \tag{5}$$

as $L(0,G)=0$ and $L(1,G)=1$, where the coefficient $b(G)$ depends on G. Furthermore, as $G/2 = \int_0^1 (x - L(x,G))\,dx$, $R(G) = x^* - L(x^*,G)$, and $L_x(x^*,G)=1$, then

$$b(G) \approx 3G,\ \ x^* = 0.5,\ \text{ and }\ R(G) \approx 0.25\,b(G).$$

It follows that

$$R(G) \approx 0.75G \ \text{ for } \ G \ \text{ small}$$

and, by Eq. (4), $\alpha \approx 0.75$. In the next section, based on randomly-generated Lorenz curves, $\beta_1, \beta_2, G_1, G_2$, and G_3 are estimated and the value of α is modified. The approximate formula is then tested against results from standard and derived Lorenz curves.

An example of a two-parameter family of Lorenz functions, similar to the multi-parameter families studied in the next section, is derived from a piecewise constant income probability density with parameters p_1, p_2, specifically,

$$p(I) = \begin{cases} p_1 \ \text{ for } \ I_1 \le I < I_2, \\ p_2 \ \text{ for } \ I_2 \le I < I_3, \\ p_3 \ \text{ for } \ I_3 \le I < I_4, \end{cases}$$

where $I_1=0, I_2=1, I_3=2, I_4=3$. As $\int_0^1 p(I)\,dI = 1$, then $p_1+p_2+p_3=1$ and so, $p_3 = 1-p_1-p_2$ is fixed given p_1 and p_2. The two positive parameters p_1 and p_2 satisfy $p_1+p_2 \le 1$. The function $I(x)$ for this income probability density is

$$I(x) = I_i + (x-x_i)/p_i \ \text{ for } \ x_i \le x \le x_{i+1},$$

for $i=1,2,3$ where $x_1=0$, $x_2=p_1$, $x_3=p_1+p_2$, and $x_4=1$. The Lorenz function $L(x,p_1,p_2)$ and the Gini coefficient depend on the values of the two parameters p_1, p_2, e.g.,

$$G = G(p_1,p_2) = 1 - 2\int_0^1 L(x,p_1,p_2)\,dx.$$

For example, if $p_1=1/2, p_2=1/4$, then $p_3=1/4$ and $G=2/5$. For this family of income densities, more than one set of parameters can result in the same value of G. For example, $p_1=1, p_2=0, p_3=0$ and $p_1=1/3, p_2=1/3, p_3=1/3$ both give $G=1/3$. The value of G for this family is exactly given by the formula

$$G = 1 - \frac{2p_1^2 + 6p_1p_2 + 8p_2^2 + 6p_1p_3 + 18p_2p_3 + 14p_3^2}{3p_1 + 9p_2 + 15p_3}$$

where $p_3 = 1-p_1-p_2$. The Robin Hood index can also be explicitly found for this two-parameter family of Lorenz functions and has the form:

$$R = \begin{cases} \frac{1}{2}p_1\phi, \ \text{ when } 1/2 \le \phi \le 1, \\ \left(p_1(\phi-\frac{1}{2}) + \frac{1}{2}p_2(\phi-1)^2\right)/\phi, \ 1 \le \phi \le 2, \\ \left(p_1(\phi-\frac{1}{2}) + p_2(\phi-\frac{3}{2}) + \frac{1}{2}p_3(\phi-2)^2\right)/\phi, \\ \qquad \text{when } 2 \le \phi \le 5/2, \end{cases}$$

where $\phi = p_1/2 + 3p_2/2 + 5p_3/2$. For the three examples above, $R = 1/4 = 0.75G$ when $p_1=1/3, p_2=1/3$ or when $p_1=1, p_2=0$, and $R = 49/160 = 0.766G$ when $p_1=1/2, p_2=1/4$.

3 Approximate formula for 100 families of income distributions

To estimate $\alpha, \beta_1, \beta_2, G_1, G_2$, and G_3 in Eq. (4), 100 families of income distributions are considered where each family has 20 parameters. (That is, twenty different parameter values must be specified to define a unique Lorenz curve for each of the 100 different families.) Then, $\alpha, \beta_1, \beta_2, G_1, G_2$, and G_3 are estimated by comparing how the Robin Hood indexes are related to the Gini coefficients for the 100 families of Lorenz curves. For each family, the income probability densities are piecewise constant functions defined for $I \ge 0$ where the number of income intervals, m, is equal to twenty-one. First, a general but simple family of income densities is initially studied. For this family, the income density is piecewise constant and has the form:

$$p(I) = \sum_{i=1}^{m} p_i \chi_i(I) \ \text{ for } \ i = 1,\ldots,m=21,$$

where $I_{i+1} = I_i + \Delta_i$ for $i=1,2,\ldots,m$, with $I_1=0$, and

$$\chi_i(I) = 1 \ \text{ for } \ I_i \le I < I_{i+1} \ \text{ and } \ \chi_i(I) = 0 \ \text{ otherwise.}$$

As $\int_0^1 p(I)\,dI = 1$, then $\sum_{i=1}^m p_i\Delta_i = 1$ defines one value of p_i in terms of the others. The function $I(x)$ where $0 \le x \le 1$ can readily be calculated for this probability density and has the form:

$$I(x) = I_i + (x-x_i)/p_i, \text{ for } x_i \le x \le x_{i+1},\ \ i=1,2,...,m,$$

$$\text{where } \ x_0 = 0, \ \text{ and } x_i = \sum_{j=1}^{i-1} p_j\Delta_j \ \text{ for } i = 2,\ldots,m+1.$$

For this initial density, the income intervals have equal widths and the probabilities for each interval are randomly chosen, i.e., Δ_i and p_i are, respectively,

$$\Delta_i = 1 \ \text{ and } \ p_i = \gamma_i/D \ \text{ for } \ i=1,2,\ldots,m, \ \text{ where}$$

γ_i are uniformly distributed random numbers on $[0,1]$ and $D = \sum_{i=1}^m \gamma_i\Delta_i$. This represents a large number

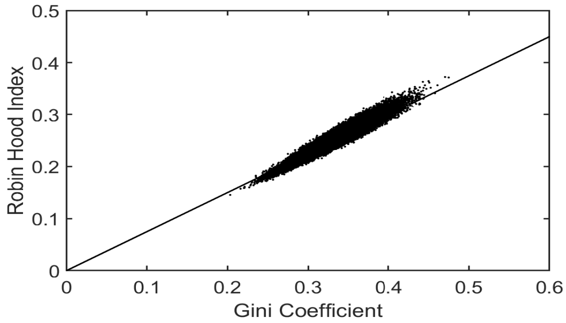

Figure 1: 100,000 points generated for the initial twenty-parameter family of income distributions plotted along with the line $R = 0.75G$.

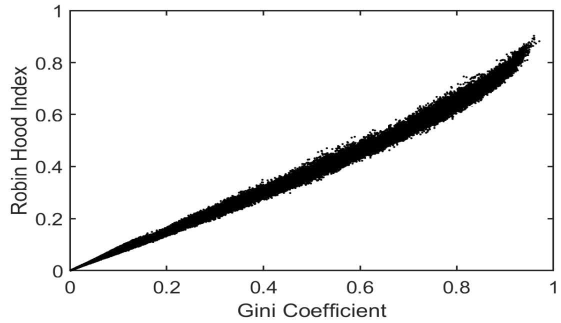

Figure 2: 100,000 Robin Hood index/Gini coefficient points generated from the 100 twenty-parameter families of income distributions.

of income probability densities, i.e., over the selections of $\{\gamma_i\}_{i=1,m}$. The Lorenz functions can be found for each income density and the Gini coefficients and Robin Hood indexes rapidly calculated. For 100,000 randomly-selected distributions from this family, the values of Gini coefficient, however, range only from about 0.2 to 0.45. A plot of the Robin Hood index with respect to the Gini coefficient is given in Figure 1 for 100,000 realizations of this family. Figure 1 also illustrates that $R(G) \approx 0.75G$ for this family.

To extend values of the Gini coefficient from about 0.0 to 0.95, this original initial twenty-parameter family of income distributions is expanded to 100 twenty-parameter families. Each family is a modification of the simple family above by generalizing the income interval widths and the interval probabilities. There are undoubtedly many ways to generalize these interval widths and probabilities. In the present investigation, the (j,k)th family is defined for $j,k = 1,2,\ldots,10$ by

$$\Delta_{i,j} = (i)^{j-3}, p_{i,j,k} = \gamma_i(k/5)^i/D_{j,k}, \text{ for } i = 1,2,\ldots,m,$$

with

$$D_{j,k} = \sum_{i=1}^{m} \gamma_i(k/7)^i \Delta_{i,j}.$$

For example, the original family of probability densities corresponds to $j = 3$ and $k = 5$. The forms for the interval widths $\Delta_{i,j}$ and the probabilities $p_{i,j,k}$ are motivated by considering exponential and power-law income probability distributions [17]. (In particular, if income is distributed as $p(I) = \alpha e^{-\alpha I}$, and income interval widths are uniformly selected, then the average probability in the ith interval is proportional to c^i for a constant c. In addition, if income is distributed such as $p(I) \propto I^\alpha$ over an interval in I and the probabilities are uniformly selected, then interval widths are approximately proportional to i^c for a constant c.) As the interval probabilities and widths can be increasing or decreasing as i increases depending on the family (j,k) and the interval probabilities are assigned random magnitudes, a diverse and large number of income probability densities are represented by these 100 families. A plot of the Robin Hood indexes and Gini coefficients for 100,000 income densities from these 100 families is given in Figure 2. Each of the calculated 100,000 points is selected randomly from the 100 families, i.e., each family is sampled approximately 1000 times. Least squares fits to the 100,000 points give the following approximation for the Robin Hood index R in terms of the Gini coefficient G:

$$R \approx \begin{cases} 0.74G, \text{ for } 0 \le G \le 0.5, \\ 0.37 + 0.90(G - 0.5), \text{ for } 0.5 \le G \le 0.8, \\ 0.64 + 1.26(G - 0.8), \text{ for } 0.8 \le G \le 0.95. \end{cases} \tag{6}$$

In comparing the approximate formula (6) with the 100,000 Robin Hood index/Gini coefficient points illustrated in Figure 2, 52% of the points are within 2.5% of curve (6), 81% are within 5% of the curve, and 97% of the points are within 10% of the curve. Also, as illustrated in Figure 3, good agreement is seen in a plot of the approximate formula (6) with 1000 Robin Hood index points obtained from sampling the 100 families of income distributions.

4 Comparison with several one-parameter Lorenz functions

For the one-parameter families of income distributions considered in this section, each Lorenz function, $L(x, G)$, in the family is unique given a value of $G \in [0, 1]$. There are many one-parameter Lorenz functions commonly used for studying income distributions. One-parameter Lorenz functions are generally considered sufficiently accurate approximations

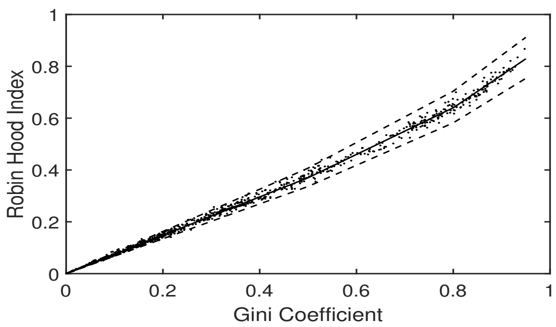

Figure 3: 1000 points generated from the 100 families of income distributions plotted with the approximate formula (6) and with curves 10% above and below formula (6).

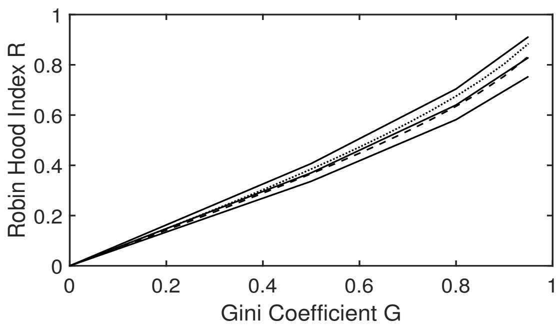

Figure 4: Pareto (dotted) and Log-normal (dashed) Robin Hood indexes compared with Eq. (6) (solid) and with curves 10% above and below Eq. (6) (solid).

for most income distributions [18]. Four commonly-used Lorenz functions are compared in this section. It is shown that the Robin Hood index in terms of the Gini coefficient G for the four one-parameter Lorenz functions is well-approximated by the proposed formula (6). The four Lorenz functions considered are the Pareto, Log-normal, Weibull, and Lamé. The Pareto Lorenz function has the form:

$$L_P(x,G) = 1-(1-x)^{\frac{1-G}{1+G}}.$$

The Log-normal Lorenz function is given by:

$$L_{LN}(x,G) = \Phi\left(\Phi^{-1}(x) - \sqrt{2}\Phi^{-1}\left(\frac{G+1}{2}\right)\right),$$

where $\Phi(x) = \frac{1}{2}(1+\mathrm{erf}(x/\sqrt{2})$ is the standard normal cumulative distribution. The Weibull function is:

$$L_W(x,G) = 1 - \frac{\Gamma\big(1-\log(1-G)/\log(2), -\log(1-x)\big)}{\Gamma\big(1-\log(1-G)/\log(2)\big)},$$

where $\Gamma(\cdot,\cdot)$ is the incomplete (upper) gamma function. (For example, in MATLAB, $L_W(x,G) = \mathrm{gammainc}\big(-\log(1-x), 1-\log(1-G)/\log(2)\big)$.) The fourth Lorenz function considered is the Lamé function given by:

$$L_L(x,G) = (1-(1-x)^{\alpha})^{1/\alpha},$$

where $G = 1 - \dfrac{(\Gamma(1/\alpha))^2}{\alpha\Gamma(2/\alpha)}$ (See, e.g, [14, 15, 18, 19, 20, 21] for more information about these Lorenz functions.) Each of these Lorenz functions is uniquely defined by the value of the Gini coefficient and the Robin Hood index depends on the Gini coefficient. Generally, given the Lorenz function, finding the Robin Hood index requires numerical solution. For the Pareto curve, however, the Robin Hood index, $R_P(G)$, is given explicitly by

$$R_P(G) = \left(\frac{1-G}{1+G}\right)^{\frac{1-G}{2G}}\left(\frac{2G}{1+G}\right).$$

The graph of the Robin Hood index $R(G)$ for each one-parameter Lorenz function has certain similarities. In particular, for each Lorenz function, the graph of the Robin Hood index with respect to the Gini coefficient increases from the origin, has initial slope approximately equal to 0.75, and terminates at the point $(1,1)$. As Lorenz functions approach a discontinuous curve as G approaches unity and because most income densities in applications have $G < 0.95$, Robin Hood index approximation is examined in the present investigation for Gini coefficients between 0.0 and 0.95.

Plots of the Robin Hood index with respect to the Gini coefficient for these four Lorenz curves are shown in Figures 4 and 5 along with the approximate formula (6). The approximate formula (6) is generally within 5% of the Robin Hood index curves for all these Lorenz functions. In addition, the maximum differences between approximate formula (6) and these one-parameter Lorenz-function Robin Hood indexes is: 6.7% for the Pareto function, 4.4% for the Log-normal function, 4.1% for the Weibull function, and 6.2% for the Lamé function.

5 Comparison with results from a stochastic income-wealth model

Another check on the approximate Robin Hood index formula (6) is made in this section by comparing the approximate formula with Robin Hood indexes obtained for two additional income distributions. The distributions are derived from a stochastic income-wealth model which is similar to many models of in-

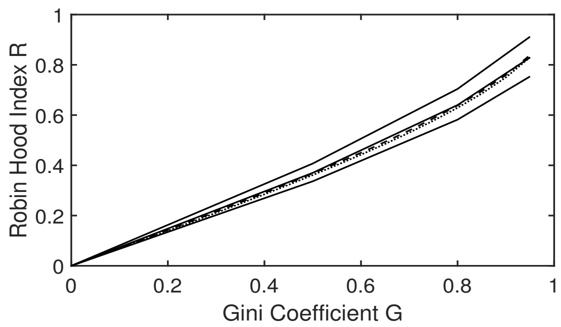

Figure 5: Lamé (dotted) and Weibull (dashed) Robin Hood indexes compared with Eq. 6 (solid) and curves 10% above and below Eq. (6) (solid).

come and wealth such as those developed and investigated in [22, 23, 24, 25]. In this model, the economy is assumed to consist of a large population of individuals having differing individual salaries and investment proficiencies. The stochastic model considered has the form

$$\begin{cases} I(t,X) = s(X) + r(X)K(t,X) + f_D\big(\bar{I}(t) - I(t,X)\big), \\ \dfrac{\partial W(t,X)}{\partial t} = I(t,X) - \gamma_2 W(t,X) - \gamma_3 I(t,X) - c(t,X), \end{cases} \tag{7}$$

where $\bar{I}(t) = \int_0^1 I(t,x)\,dx$, $K(t,X) = \gamma_1 W(t,X)$, and where $X \sim \mathbb{U}(0,1)$ is a uniformly distributed random variable on $[0,1]$. In model (7), the stochastic process $I(t,X)$ is the rate of income, $W(t,X)$ is the wealth, and $K(t,X)$ is the capital investment at time t. The value of X ranges from 0 to 1 and identifies individuals in the population with a certain salary level and investment proficiency. The capital $K(t,X)$ is assumed to be proportional to the wealth $W(t,X)$ for all individuals and all time. The salary rate for each individual is assumed to be $s(X)$. Income obtained by investment depends on how the individual's assets are invested and is equal to $r(X)K(t,X)$ where $r(X)$ is rate of capital growth. For example, $I(t,x_i)$ is the income rate for the ith individual in the population, $s(x_i)$ is the salary rate, and $r(x_i)$ is the individual's rate of capital investment where $0 \le x_i \le 1$. The parameter $f_D \ge 0$ is a linear income redistribution parameter with no redistribution if $f_D = 0$ and increasing redistribution as $f_D > 0$ increases. The average income of the population is denoted $\bar{I}(t)$ at time t. Taxes and costs on wealth and income are calculated using flat-rate constants $0 \le \gamma_2, \gamma_3 < 1$. The parameter $c(t,X)$ is a rate of consumption for individuals in the economy. It is assumed, as in [26], that individuals consume a constant fraction of their wealth, i.e., $c(t,X) = \gamma_4 W(t,X)$. A sample of N individuals in the economy are described by N realizations of model (7) where the individuals' wealth and rate of income, $W_i(t)$ and $I_i(t)$ for $i = 1,2,\dots,N$, satisfy the differential equation system:

$$\begin{cases} I_i(t) = s_i + r_i K_i(t) + f_D\big(\bar{I}(t) - I_i(t)\big), \\ dW_i(t)/dt = I_i(t) - \gamma_2 W_i(t) - \gamma_3 I_i(t) - c_i(t), \end{cases}$$

where $\bar{I}(t) = \sum_{i=1}^N I_i(t)/N$, $K_i(t) = \gamma_1 W_i(t)$, and where x_i for $i = 1,2,\dots,N$ are uniformly distributed random numbers on $[0,1]$ and, for example, $I_i(t) = I(t,x_i)$.

To determine Lorenz functions and study how Robin Hood indexes vary with Gini coefficient, the equilibrium income distribution is considered where wealth $W(t,X)$, rate of capital $K(t,X)$, and rate of income $I(t,X)$ are not changing with time, specifically, $W(t,X) = W(X)$, $K(t,X) = K(X)$, and $I(t,X) = I(X)$ in (7). At equilibrium, model (7) has the form:

$$\begin{aligned} I(X) &= s(X) + \gamma_1\gamma_5 r(X) I(X) + f_D(\bar{I} - I(X)), \\ W(X) &= \gamma_5 I(X), \quad K(X) = \gamma_1 W(X) = \gamma_1\gamma_5 I(X), \end{aligned} \tag{8}$$

where $\bar{I} = \int_0^1 I(x)\,dx$ and $\gamma_5 = (1-\gamma_3)/(\gamma_2+\gamma_4)$. As $K(X)$ and $W(X)$ are proportional to income $I(X)$, only $I(X)$ needs to be explicitly examined. Average wealth satisfies $\bar{W} = r_1\bar{I}$ and so, the total wealth of the individuals is proportional to the total income for model (8).

For model (8), $I(X)$ satisfies:

$$I(X) = \frac{s(X) + f_D\bar{I}}{1 + f_D - \gamma_1\gamma_5 r(X)}. \tag{9}$$

Two model cases are considered in the present investigation. First, it is assumed that rate of investment, $r(X)$, is the same for all individuals in which case $r(X) = r$ and Eq. (9) reduces to

$$I(X) = a_1 s(X) + a_2 \text{ for constants } a_1, a_2. \tag{10}$$

In the second case, it is assumed that salary rate, $s(X)$, is the same for all individuals in which case $s(X) = s$ and Eq. (9) reduces to

$$I(X) = \frac{b_1}{1 - b_2 r(X)} \text{ for constants } b_1, b_2. \tag{11}$$

For the first model case, it is assumed that the salary is distributed exponentially. Specifically, $p(s) = \beta\exp(-\beta(s-s_1)$ for $s_1 \le s < \infty$ is the probability density of salary s. It follows that for x uniformly distributed on $[0,1]$, then $s(x)$ has the form $s(x) = s_1 - \log(1-x)/\beta$. Solving (10) for $I(x)$, $I(x) = I_1 - a_1\log(1-x)/\beta$ where $I_1 = a_1 s_1 + a_2$, indicating that income is also exponentially distributed. This case's type of income probability density therefore agrees with the type of income density observed by several investigators (e.g., [17, 27, 28]). As $I(x)$ is increasing in x for $0 \le x < 1$, x is the fraction of the population with rate of income less than or equal to $I(x)$. Thus, the fraction of total income for the lowest income fraction x of the population is equal to the Lorenz function $L(x,G) = \int_0^x I(u)\,du / \int_0^1 I(u)\,du$ and is given by:

$$L(x,G) = x + c(1-x)\log(1-x), \text{ for } 0 \le x \le 1 \quad (12)$$

for a constant c. The value of c is restricted to $0 \le c \le 1$ to ensure non-negativity of the Lorenz function. As $1-G = 2\int_0^1 L(x,G)\,dx$, then $c = 2G$. Thus, the Gini coefficient G is restricted to values between 0 and 0.5 for this case. The maximum of $x - L(x,G)$ on $[0,1]$ occurs at $x^* = 1-e^{-1}$ and the Robin Hood coefficient, $R(G)$, for this case is equal to $R(G) = 2G/e^1 \approx 0.736G$ for $0 \le G \le 0.5$. Hence, this model case with exponentially distributed salaries agrees closely with the approximate formula (6) for Gini coefficients less than 0.5.

For the second model case, it is assumed that the rate of investment probability density is a linear function. That is, $p(r) = br$ for $0 \le r \le r_1 = \sqrt{2/b}$ for a positive constant $b > 0$. It follows that $r(x) = r_1 x^{1/2}$ for x uniformly distributed on $[0,1]$. Solving for $I(x)$,

$$I(x) = \frac{a}{1-cx^{1/2}}$$

where a, c are positive constants. As $I(x)$ is increasing in x for $0 \le x < 1$, x is the fraction of the population with rate of income less than or equal to $I(x)$. Thus, the fraction of total income for the lowest income fraction x of the population is equal to the Lorenz function $L(x,G) = \int_0^x I(u)\,du / \int_0^1 I(u)\,du$ and is given by:

$$L(x,G) = \frac{cx^{1/2} + \log(1-cx^{1/2})}{c + \log(1-c)}, \text{ for } 0 \le x \le 1. \quad (13)$$

The Gini coefficient, $G = 1 - 2\int_0^1 L(x,G)\,dx$, for the Lorenz curve in (13) is equal to:

$$G = \frac{2c + c^2 - \frac{1}{3}c^3 + (2-c^2)\log(1-c)}{c^3 + c^2\log(1-c)}, \quad (14)$$

where $0 < c < 1$. When the value of the Gini coefficient G is specified, then the value of c can be calculated using equation (14) which in turn determines the Lorenz curve in (13). The maximum of $x - L(x,G)$ occurs at $x^* = \Big(1/c + c/\big(2c + 2\log(1-c)\big)\Big)^2$ and the Robin Hood index is given by $x^* - L(x^*,G)$. The graph of the Robin Hood index for this case is compared in Figure 6 to the approximate formula (6).

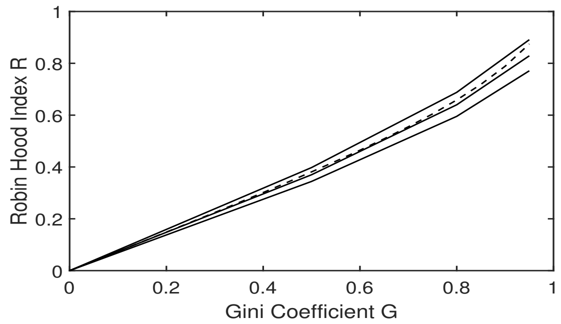

Figure 6: Model case 2 Robin Hood indexes (dashed) compared with approximate formula (6) (solid) and curves 10% above and below formula (6) (solid).

Good agreement is seen between the two curves with a maximum difference of 5.5%.

In addition, the income probability density for this case can be derived, for example, from $I(x) = a/(1-cx^{1/2})$. Let $z = f(x) = a/(1-cx^{1/2})$ and, as $f(x)$ is increasing for $0 \le x \le 1$, then $x = f^{-1}(z) = (1-a/z)^2/c^2$. Let $g(z) = f^{-1}(z)$. Finally, let $p(I)$ be the probability density of income rate I. Then, for $\hat{I} \in [I_{min}, I_{max}]$ and ΔI small,

$$\begin{aligned} p(\hat{I})\Delta I &= \mathbb{P}\big(\hat{I} < I < \hat{I} + \Delta I\big) \\ &= \mathbb{P}\big(g(\hat{I}) < X < g(\hat{I} + \Delta I)\big) \\ &= g'(\hat{I})\Delta I = \frac{2}{c^2}\left(\frac{a}{\hat{I}^2} - \frac{a^2}{\hat{I}^3}\right)\Delta I. \end{aligned}$$

Thus, the income probability density $p(I)$ for the second case is equal to:

$$p(I) = \frac{2}{c^2}\left(\frac{a}{I^2} - \frac{a^2}{I^3}\right) \text{ for } I_{min} \le I \le I_{max}, \quad (15)$$

where $I_{min} = a$, and $I_{max} = a/(1-c)$.

6 Comparison with income data for four different nations

To test the approximate formula (6), the formula is compared with Robin Hood indexes calculated from the income distribution data available from each of four countries, Finland, United States, Brazil, and South Africa. These four countries exhibit a wide range of Gini coefficients. In 2016, Finland, United States, and Brazil had Gini coefficients of 0.271, 0.411, and 0.533, respectively. Over the period from 1991 to 2014, South Africa reached a maximum Gini coefficient of 0.648 in 2005. Income distribution data for these four countries are listed in Table 1.

Table 1: Cumulative income fractions of four countries for eight cumulative population fractions with Gini coefficients 0.648 for South Africa, 0.533 for Brazil, 0.411 for United States, and 0.271 for Finland. (The income distribution data and the Gini coefficient values are from [29].)

Cumulative Population Fraction	South Africa 2005	Brazil 2016	United States 2016	Finland 2016
0.00	0.000	0.000	0.000	0.000
0.10	0.010	0.011	0.018	0.039
0.20	0.026	0.033	0.052	0.094
0.40	0.073	0.107	0.155	0.234
0.60	0.148	0.228	0.308	0.409
0.80	0.290	0.420	0.533	0.634
0.90	0.458	0.579	0.696	0.776
1.00	1.000	1.000	1.000	1.000

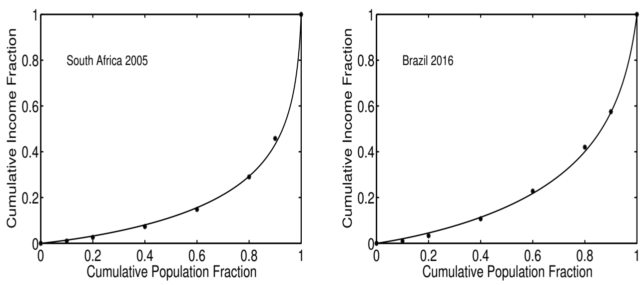

Figure 7: Income data of South Africa and Brazil with Gini coefficients 0.648, and 0.533, respectively. Lorenz curves (13) shown have c values of 0.99598 (South Africa) and 0.98276 (Brazil) corresponding to the respective Gini coefficients.

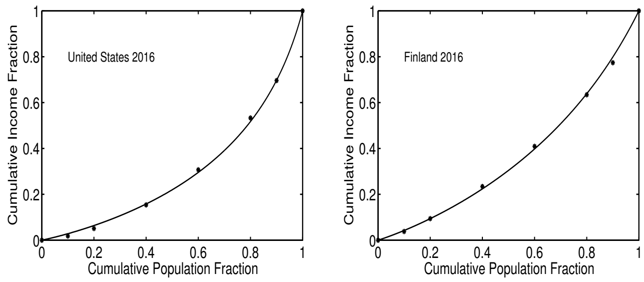

Figure 8: Income data of United States and Finland with Gini coefficients 0.411 and 0.271, respectively. Lorenz curves (13) shown have c values of 0.94719 (United States), and 0.84894 (Finland) corresponding to the respective Gini coefficients.

First, the Lorenz function (13), derived in the previous section, is compared in Figures 7 and 8 with the income distribution data for these four nations. The values of c needed for the Lorenz function (13) are calculated using equation (14) for the four values of the Gini coefficients. For Finland, United States, Brazil, and South Africa, the Gini coefficients 0.271, 0.411, 0.533, and 0.648 result in values of c equal to 0.84894, 0.94719, 0.98276, and 0.99598, respectively. The Lorenz function (13) consistently provides a good fit to the income data for the four countries. For all four countries, the maximum root mean square error between the data points in Table 1 and the Lorenz curves of Eq. (13) is 0.0103. In comparison, the maximum root mean square errors between the data and the standard one-parameter Lorenz curves of Pareto, Log-normal, Weibull, and Lamé [14, 15, 19, 20, 21] are 0.0380, 0.0155, 0.0343, and 0.0177, respectively, for all four countries.

The Robin Hood index values for the four countries are: 0.5103, 0.3839, 0.2929, and 0.1910, calculated using polynomial interpolants of the Lorenz curve income data. The Robin Hood indexes using formula (6) for the four Gini coefficients are: 0.5032, 0.3997, 0.3041, and 0.2005 which are all within 5% of the Robin Hood indexes calculated from the income data. A graph of approximate formula (6) is illustrated in Figure 9 along with the Robin Hood index values for the four countries.

7 Comparison with several reported results

Generally, Gini coefficients and Robin Hood indexes are strongly correlated in reported investigations in the literature [11, 30, 31, 32]. For example, the analysis of [30] indicates that income inequality measures (such as the Gini coefficient and the Robin Hood in-

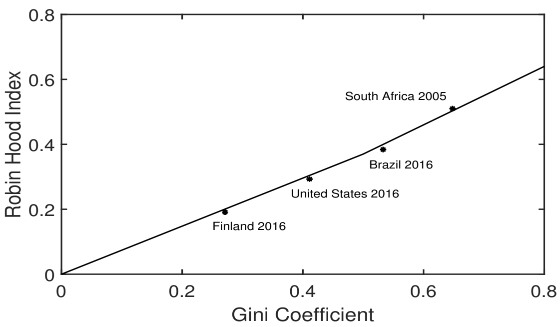

Figure 9: Robin Hood indexes for South Africa, Brazil, United States, and Finland graphed with the approximate Robin Hood index formula (6).

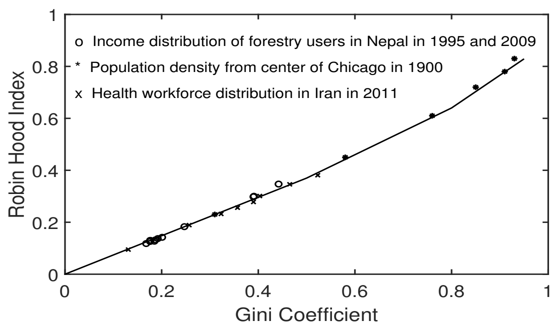

Figure 10: Robin Hood indexes for three different investigations plotted with the approximate Robin Hood index formula (6).

dex) behave very similarly and are highly correlated [11]. Indeed, it has been observed for certain Lorenz functions, such as the Lamé function, that the Robin Hood index is approximately linear with respect to the Gini coefficient (see, e.g., [18]).

In this section, the results of three previous investigations are compared with the approximate Robin Hood index formula (6). The first study [33] examines the population density with distance from the center of Chicago in 1900. The second study [34] compares distributions of maternal and child health-related workforces in provinces of Iran during 2010-2012. The third study [35] examines income distributions of users of three community forests in Nepal with pre-community forestry usage occurring in 1995 and post-community forestry usage in 2009. The Gini coefficients and Robin Hood indexes reported for these studies are illustrated in Figure 10 along with the approximate formula (6). The approximate formula agrees well with the reported Robin Hood index values for these three investigations. Formula (6) is within 5% of the Robin Hood index values for twenty-four of the twenty-six points and within 6% and 7% for the two remaining points.

8 Summary

The Gini coefficient is a widely-used measure of income inequality but is defined in a mathematical sense. The Robin Hood index, which is not as widely used as the Gini coefficient, is equal to the proportion of the population's total income that, if redistributed, would give perfect income equality. An approximate formula for the Robin Hood index in terms of the Gini coefficient is useful in obtaining a better understanding of income distributions.

An approximate formula for the Robin Hood index, R, in terms of the Gini coefficient, G, is determined for values of G between 0 and 0.95. The approximation is developed from 100,000 Lorenz curves that are randomly generated based on 100 twenty-parameter families of income distributions. The formula is within 5% (within 10%) of the Robin Hood indexes for 81% (for 97%) of the randomly-generated income distributions examined.

The approximate formula is tested against Robin Hood indexes of commonly-used one-parameter Lorenz curves, income data of several countries, and reported results of Robin Hood indexes. The approximate formula is also tested against results of a stochastic income-wealth model which is introduced in the present investigation. The continuous piecewise-linear approximation is generally within 5% of Robin Hood indexes of standard one-parameter Lorenz curves, income distribution data, and the stochastic income-wealth model.

Acknowledgement

The author acknowledges the reviewers' helpful comments on improving the paper.

References:

[1] C. Gini. *Variabilitá e Mutuabilitá: Contributo allo Studio delle Distribuzioni e delle Relazioni Statistichee*. Tipografia di Paolo Cuppini, Bologna, 1912.

[2] M. Schneider. The discovery of the Gini coefficient: was the Lorenz curve the catalyst? *History of Political Economy*, 53:115–141, 2021.

[3] Gaetano Pietra. On the relationships between variability indices (Note I). *Metron*, 72(1):5–16, 2014. English translation of 1915 article "Delle relazioni tra gli indici di variabilitá (Nota I)".

[4] F. Porro and M. Zenga. Decompositions by sources and by subpopulations of the Pietra in-

dex: two applications to professional football teams in Italy. *AStA Advances in Statistical Analysis*, pages 1–28, 2021.

[5] E. M. Hoover. The measurement of industrial localization. *Rev. Econ. Stat.*, 18:162–171, 1936.

[6] R. McVinish and R. J. G. Lester. Measuring aggregation in parasite populations. *Journal of the Royal Society Interface*, 17:20190886, 2020.

[7] R. Schutz. On the measurement of income inequality. *Am. Econ. Rev.*, 41:107–122, 1951.

[8] J. Park, Y. Kim, and A.-J. Ju. Measuring income inequality based on unequally distributed income. *Journal of Economic Interaction and Coordination*, 16:309–322, 2021.

[9] A. B. Atkinson and J. Micklewright. *The Distribution of Income in Eastern Europe*. European University Institute, Florence, 1992. EUI Working Paper ECO; 1992/87.

[10] S. Kozuharov and V. Petkovski. The Impact of Social Transfers on Inequality Measured by Gini Index: The Example of Macedonia. *UTMS Journal of Economics*, 9:49–61, 2018.

[11] F. G. De Maio. Income inequality measures. *J Epidemiol Community Health*, 61:849–852, 2007.

[12] J. Fellman. *Mathematical Analysis of Distribution and Redistribution of Income*. Science Publishing Group, New York, 2015.

[13] J. Fellman. Income inequality measures. *Theoretical Economics Letters*, 8:557–574, 2018.

[14] J. M. Sarabia, V. Jordá, and C. Trueba. The Lamé class of Lorenz curves. *Communications in Statistics, Theory and Methods*, 46(11):5311–5326, 2017.

[15] F. Cowell. *Measuring Inequality. LSE Handbooks on Economics Series*. Prentice Hall, London, 1995.

[16] A. S. Ackleh, E. J. Allen, R. B. Kearfott, and P. Seshaiyer. *Classical and Modern Numerical Analysis: Theory, Methods, and Practice*. Chapman & Hall / CRC Press, Bocal Raton, Florida, 2010.

[17] A. Drăgulescu and V. M. Yakovenko. Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. *Physica A*, 299:213–221, 2001.

[18] J. M. Henle, N. J. Horton, and S. J. Jaku. Modelling inequality with a single parameter. In D. Chotikapanich, editor, *Modeling Income Distributions and Lorenz Curves. Economic Studies in Equality, Social Exclusion and Well-Being, Vol. 5*, pages 255–269. Springer, New York, 2008.

[19] S. P. Jenkins. Pareto models, top incomes and recent trends in UK income inequality. *Economica*, 84:261–289, 2017.

[20] M. Krause. Parametric Lorenz curves and the modality of the income density function. *Review of Income and Wealth*, 60:905–929, 2014.

[21] M. Lubrano. The econometrics of inequality and poverty. Lecture 4: Lorenz curves, the Gini coefficient and parametric distributions. *Manuscript available online at http://www. vcharite. univ-mrs. fr/PP/lubrano/poverty. htm*, 2013.

[22] A. Alesina and G.-M. Angeletos. Fairness and redistribution. *The American Economic Review*, 95:960–980, 2005.

[23] A. H. Meltzer and S. F. Richard. A rational theory of the size of government. *Journal of Political Economy*, 89:914–927, 1981.

[24] T. Romer. Individual welfare, majority voting, and the properties of a linear income tax. *Journal of Public Economics*, 4:163–185, 1975.

[25] S. A. Shaikh. Equitable distribution of income with growth in an Islamic economy. *International Journal of Islamic Economics and Finance Studies*, 3:6–14, 2017.

[26] C. I. Jones. Pareto and Piketty: The macroeconomics of top income and wealth inequality. *Journal of Economic Perspectives*, 29:29–46, 2015.

[27] A. Drăgulescu and V. M. Yakovenko. Evidence for the exponential distribution of income in the USA. *The European Physical Journal B*, 20:585–589, 2001.

[28] Y. Tao, X. Wu, T. Zhou, W. Yan, Y. Huang, H. Yu, B. Mondal, and V. Yakovenko. Exponential structure of income inequality: evidence from 67 countries. *Journal of Economic Interaction and Coordination*, 14:345–376, 2019.

[29] Knoema. Gini coefficient and Lorenz curve around the world. https://knoema.com/crmndag/gini-coefficient-and-lorenz-curve-around-the-world, 2016. Accessed: 2021-02-06.

[30] I. Kawachi and B. P. Kennedy. The relationship of income inequality to mortality: Does the choice of indicator matter? *Social Science & Medicine*, 45:1121–1127, 1997.

[31] B. P. Kennedy, I. Kawachi, and D. Prothrow-Stith. Income distribution and mortality: cross-sectional ecological study of the Robin Hood Index in the United States. *British Medical Journal*, 312:1004–1007, 1996. Also the published erratum: (1996) *British Medical Journal* 312, 1194.

[32] L. Shi, J. Macinko, B. Starfield, J. Wulu, J. Regan, and R. Politzer. The Relationship Between Primary Care, Income Inequality, and Mortality in US States, 1980–1995. *J Am Board Fam Pract*, 16:412–422, 2003.

[33] J. E. Cohen. Measuring the concentration of urban population in the negative exponential model using the Lorenz curve, Gini coefficient, Hoover dissimilarity index, and relative entropy. *Demographic Research*, 44:1165–1184, 2021.

[34] R. Honarmand, M. Mozhdehifard, and Z. Kavosi. Geographic distribution indices of general practitioners, midwives, pediatricians, and gynecologists in the public sector of Iran. *Electronic Physician*, 9:4584–4589, 2017.

[35] B. Khanal. Is community forestry decreasing the inequality among its users? Study on impact of community forestry on income distribution among different user groups in Nepal. *International Journal of Social Forestry*, 4:139–152, 2011.