Working Paper Series

Shaikh and Ragab's 'Incomes of the Vast Majority': Some additions and extensions

Peter J. Lambert
S. Subramanian

ECINEQ WP 2015 - 354

ECINEQ 2015 - 354
January 2015

www.ecineq.org

Shaikh and Ragab's 'Incomes of the Vast Majority': Some additions and extensions*

Peter J. Lambert†
University of Oregon, USA

S. Subramanian
Madras Institute of Development Studies, India

Abstract

The 'Vast majority of incomes ratio (VMIR)' R_0 is the ratio of the average income μ_0 of a poorest majority p_0 of the population to the overall average income μ. Another measure of equality is $E_0 \equiv (1 - G)$ where G is the Gini coefficient of inequality of the distribution. Shaikh and Ragab (2007, 2008), employing a wide variety of data sets, have shown (among other things), that when p_0 is 70 per cent, R_0 serves as a uniformly stable and excellent approximation of E_0. We extend the Shaikh-Ragab findings in two directions. We explore some analytics of the relationship between R_0 and E_0 for a lognormal distribution; and we derive a couple of other VMIRs and related measures of welfare which are distinct from but inspired by the Shaikh-Ragab formulations.

Keywords: Vast Majority of Incomes Ratio, lognormal distribution, Gini coefficient, Lorenz curve, Sen's welfare index.

*Alterative *AMS* (2000) Subject Classification: Primary 91B82, 62P20. Secondary 91C05.

†**Contact details:** (P.J. Lambert) Department of Economics, University of Oregon Eugene, OR 97403-1285, U.S.A. Email: plambert@uoregon.edu. (S. Subramanian) Madras Institute of Development Studies, 79, 2nd Main Road, Gandhinagar, Adyar, Chennai – 600 016, India. Email: ssubramanianecon@gmail.com.

1 Introduction

It is well-recognized that per capita national income measures are averages which conceal wide disparities in populations. There is agreement, too, that international economic comparisons should not ignore inequality. But there has been considerable debate on how exactly to bring inequality into the picture (Gruen and Klasen, 2008, p. 213). Shaikh and Ragab (2007, 2008) explore a combination of information on national income and its distribution, the so-called *VMI ratio*, or per capita income of the vast majority as a fraction of per capita national income, across countries and across time. They find an empirically robust international relation, which they call the '1.1 Rule', as well as a new interpretation of the Gini coefficient (*G*). In this note, we explore Shaikh and Ragab's findings further.

2 Shaikh and Ragab's '1.1 Rule' and '1.0 Rule'

Let *VMIR(x)* be the ratio of the per capita income of the bottom *x* per cent of the population to the overall average income. We might take $x = 70$, 80 or 90, for example. One can think of *VMIR(x)* as a measure of the degree of income equality. Another such measure is (1-*G*), where *G* is the Gini coefficient. Both of these of course vary substantially across countries and across time, in accordance with varying social and historical determinants of inequality. Shaikh and Ragab's finding is quite striking: the ratio of *VMIR(80)* to (1-*G*) is extraordinarily stable across countries, and in a country through time, with an average value of about 1.1 and variations which seldom go beyond ± 5 percent of this. Similarly, the ratio of *VMIR(70)* to (1-*G*) is approximately 1.0. They characterize these empirically robust relations as the '1.1 Rule' and '1.0 Rule' respectively.

3 Interpretations of the Gini coefficient

According to Shaikh and Ragab's findings, (1-*G*) represents the relative disposable per capita income of the first seventy percent of a nation's population; equivalently, *G* represents the percentage difference between national income per capita and the per capita income of the first 70 percent of the income distribution. Sen's well-known (1976) proposal is to rank countries' standards of living by the inequality-discounted net national income per capita, $\mu(1-G)$, where μ is per capita income overall.

This measure is tied to traditional social welfare theory.[1] Shaikh and Ragab's 1.0 Rule, though advocating the same ranking, is free from social welfare connotations, assessing national progress purely in terms of the per capita income of the bottom 70 percent of the population. In terms of the 1.1 Rule, 1.1 times Sen's index is equal to the per capita income of the bottom 80 percent of the population. For varying *x*, *VMIR(x)* is a representation of the economic situation of the poorest *x* percent of the population. For values of $x = 70$, 80 or 90, for example, we are talking of the economic situation of the vast majority. That the Gini coefficient should somehow be implicated in the economic situation of the vast majority is a somewhat inevitable outcome on

[1] As Shaikh and Ragab point out, the social welfare function approach 'requires strong theoretical assumptions about individual behavior and psychology, about appropriate measures of individual well-being such as utility, about … aggregation … and about the effects of income, inequality, education, etc.' They cite Fleurbaey and Mongin (2005) on all of this.

reflection, and in the following sense. The Gini coefficient of a pure exponential pdf is a constant, but the overall Gini depends also on the fraction of total income which accrues to the very rich. Hence the overall Gini can be viewed as an index of the relative income of the very rich, and its variations across countries and through time mark the changes in the relative fortunes of this particular segment of society.

More generally, Shaikh and Ragab report finding that, across a large sample of countries in the WIDER-UNU-World Bank database, and at least for lower quantile groups comprising between 50 and 90 percent of the whole population, the per capita income of the group is proportional to inequality-discounted average income per capita overall, through a constant of proportionality which is dependent solely on the population proportion x. Let that constant of proportionality be $a(x)$, so that $VMIR(x)\big/(1-G) \approx a(x),\ \ 50 < x < 90.$[2]

4 Further analysis and implications

The lognormal income distribution is unimodal, and skewed to the right, and is popular in labor economics. The Lorenz curve and Gini coefficient for this distribution are in terms of the parameter σ, which is the standard deviation of log incomes. The estimated variance σ^2 of log incomes is commonly used as an inequality measure by labor economists (but see Foster and Ok, 1999, on this).

We investigated the behavior of the ratio $VMIR(p_0)\big/(1-G) = a(p_0)$, where p_0 is a generic "percentage of the vast majority" value, for the lognormal distribution using a wide range of values of σ. See Figure 1. Over that wide range of σ values, $VMIR(p_0)\big/(1-G)$ is not constant, nor even monotonic in σ, but in the restricted range 0.49 < σ < 0.64, which is realistic[3] and is marked on the horizontal axis in Figure 1, the model accounts quite well for the 1.1 and 1.0 Rules, in that $VMIR(80\%)/(1-G)$ is close to a flat value of 1.1, and $VMIR(70\%)/(1-G)$ is close to 1.0.

The relationship between $VMIR(p_0)$ and $(1-G)$ is positive over the range 0.49 < σ < 0.64 for all p_0, but of course differs for different values of p_0: see Figure 2, parts a and b, which also shows the influence of changing inequality (σ) on the rate of change of $VMIR(p_0)$ with

[2] Shaikh and Ragab report coefficients *a(x)* of 1, 1.1, 1.27 for x = 70, 80, 90 respectively, for all countries in their sample and for all time periods in each country. They note that these findings are "so robust that they constitute general empirical rules". We should regard them as rules of thumb, not laws. The sixty percent cumulative population proportion also gives rise to a tolerably good empirical rule of thumb. Shaikh and Ragab show that an "econophysics" approach to income distribution, as discussed by Dragulescu and Yakovenko (2001), can be used to predict both the level as well as the international and intertemporal constancy of the "1.1 Rule". According to this approach, the income distribution is fitted with two distinct pdfs, the exponential applicable to the first 97-99 percent of the population and the Pareto for the top 3 percent (Dragulescu and Yakovenko, 2002, pp. 1-2). The interpretation is that income from wages and salaries yields additive diffusion, while income from investments and capital gains yields multiplicative diffusion (Silva and Yakovenko, 2004, p. 6).

[3] This range of σ-values is used in Lambert (2011) and corresponds to the range 0.65 < 1 - G < 0.73 for one minus the Gini coefficient. Shaikh and Ragab (2008) report international values 0.30 < 1 - G < 0.70.

respect to inequality $(1-G)$; this rate of change is decreasing in σ for $p_0 < 60$, increasing in σ for $p_0 > 70$, and roughly invariant to changes in σ for $60 < \sigma < 70$.

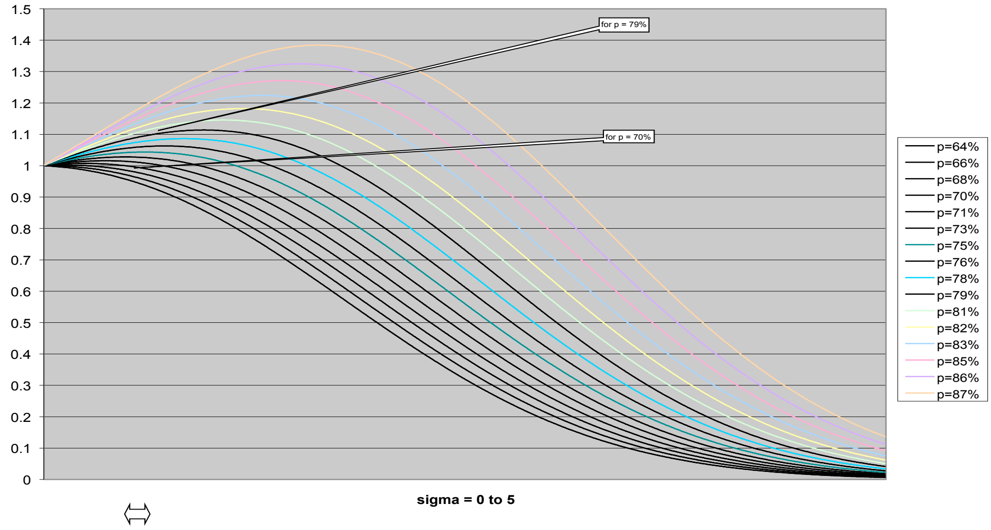

Figure 1: $VMIR/(1-G)$ plotted against σ for a full range of % of the vast majority

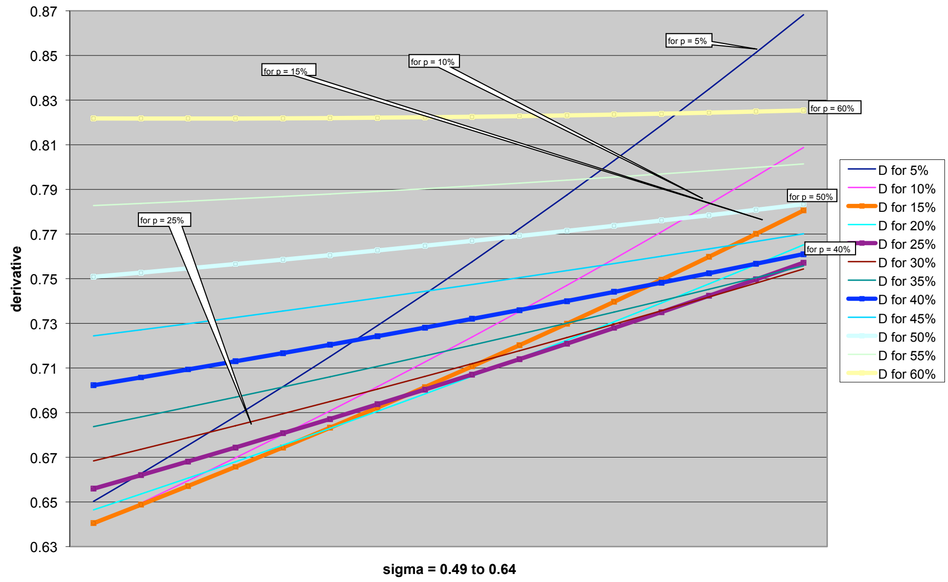

Figure 2(a): $dVMIR/d(1-G)$ plotted against σ (for low % of the vast majority)

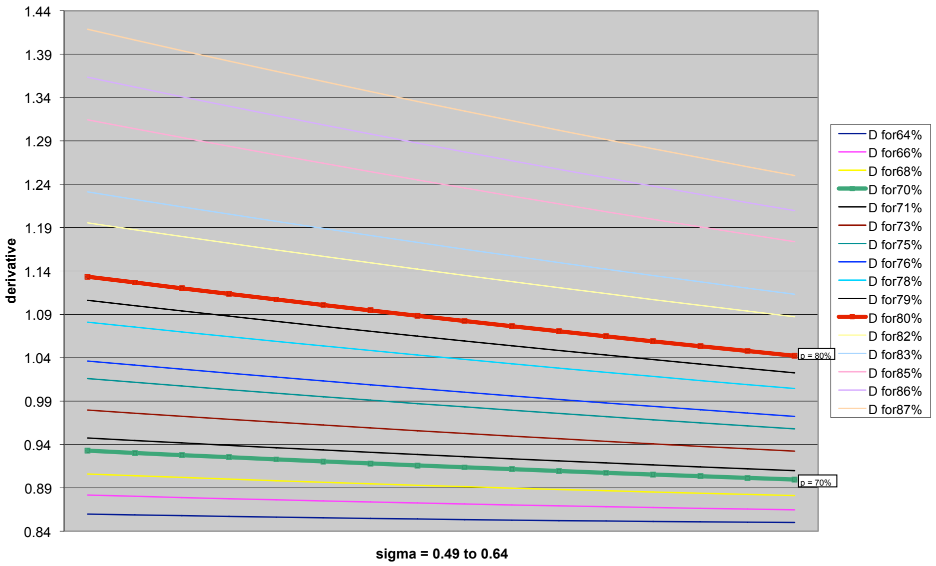

Figure 2(b): $dVMIR/d(1-G)$ plotted against σ (for high % of the vast majority)

Mathematical analysis reveals more. Let $F(x|\theta,\sigma^2)$ be the distribution function for the lognormal income distribution, say $x \approx LN(\theta,\sigma^2)$, and let Φ be the distribution function for the standard normal distribution, so that $\ell n(x) \approx N(\theta,\sigma^2)$ and

(1) $$F(x|\theta,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\ell n(x)} e^{-\left[\frac{(t-\theta)^2}{2\sigma^2}\right]} dt = \Phi\left(\frac{\ell n(x)-\theta}{\sigma}\right)$$

As in Lambert (1989, page 45), the Lorenz curve *L(p)* is defined by

(2) $$p = F(y|\theta,\sigma^2) \Rightarrow L(p) = F(y|\theta+\sigma^2,\sigma^2)$$

and the Gini coefficient G is defined by

(3) $$1-G = 2\left[1-F(e^{\theta+\frac{\sigma^2}{\sqrt{2}}}|\theta,\sigma^2)\right]$$

Let $p_0 = F(y_0|\theta,\sigma^2)$ be the proportion of the population in the vast majority, so that

(4) $$VMIR(p_0) = \frac{L(p_0)}{p_0} = \frac{F(y_0|\theta+\sigma^2,\sigma^2)}{F(y_0|\theta,\sigma^2)} = \frac{\Phi(t-\sigma)}{\Phi(t)} \quad \& \quad 1-G = 2\left[1-\Phi\left(\frac{\sigma}{\sqrt{2}}\right)\right]$$

where $t = \frac{\ell n(y_0)-\theta}{\sigma} = \Phi^{-1}(p_0)$. Therefore, when σ changes, the responses of $VMIR(p_0)$ and $(1-G)$ are as follows:

(5) $$\frac{dVMIR(p_0)}{d\sigma} = \frac{-\phi(t-\sigma)}{\Phi(t)} \quad \& \quad \frac{d(1-G)}{d\sigma} = -\sqrt{2}\phi\left(\frac{\sigma}{\sqrt{2}}\right)$$

where $\Phi'(u) = \phi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \quad \forall u$ is the standard normal density function; both of these responses are negative. Writing R_0 for $VMIR(p_0)$ and E_0 for $(1-G)$, from (5) we have: $\frac{dR_0}{dE_0} = \frac{\phi(t-\sigma)}{\sqrt{2}p_0\phi\left(\frac{\sigma}{\sqrt{2}}\right)} > 0$ (given that $\Phi(t) = p_0$); R_0 and E_0 move in the same direction.

The Lorenz curve $L(p)$ plots the cumulative income share of the poorest *p*th fraction of the population for every $p \in [0,1]$. The generalized Lorenz curve is $GL(p) = \mu L(p)$ (Shorrocks, 1983). Shorrocks calls the curve $GL(p)/p$ the "COMIC" and sees it as a "useful analytical device". $GL(p)/p$ is the mean income of the lowest *p* per cent of income recipients; in our terms, it is $\mu VMIR(p)$.[4] Shorrocks says that it represents a "cumulated mean income curve, which seems appropriate to abbreviate to COMIC. COMICS are non-decreasing functions of *p*, rising from the lowest income y_1 (when $p = 1/n$) to the mean μ (when $p = 1$). COMICs drawn for two distributions enable an immediate comparison to be made from both the viewpoint of a Rawlsian (for whom only the left-hand end of the graph would be relevant) and that of someone who is indifferent to the distribution of any aggregate income (for whom only the right-hand end-points would be relevant)". The COMIC $\mu VMIR(p)$ strikes a balance between these two extreme welfare stances. Indeed, by paying attention to the entire generalized Lorenz curve, as opposed to only its extremities, one attends to considerations of *both* equity and efficiency, as attested by the fact that the area under the generalized Lorenz curve is just Sen's (1976) welfare index $\mu(1-G)$.

5 An interesting parallel

We present now another set of results, similar in flavor but with significant differences, which seem to be a near-relation of the *VMIR*. We begin by identifying an income level x^* and a proportion $p^* = F(x^*)$ such that $L(p^*) = 1 - p^*$, as shown in Figure 3. Namely, p^* is the abscissa value at the unique point, labeled D in the Lorenz diagram, where the Lorenz curve $q = L(p)$ intersects the line $q = 1 - p$. (Hence, in particular, $p^* > ½$). Also shown in Figure 3 is a tangent line to the Lorenz curve at $(p^*, L(p^*))$, whose slope we know to be y^*/μ where μ is mean income; and another (dashed) Lorenz curve through $(p^*, L(p^*))$, one which has two linear segments and lies inside $L(p)$. Call it $L^*(p)$. It is in fact the Lorenz curve which would obtain if all incomes below y^* were replaced by their mean, and if all incomes above y^* were similarly replaced by their mean, i.e. it is the between-groups Lorenz curve in the case of a partition of the population into two groups, those whose incomes are below y^* and those whose

[4] Similarly, $L(p)/p = VMIR(p)$.

incomes are above y^*. If y^* were the poverty line, these groups would comprise the poor and the non-poor respectively (and, since $p^* > \frac{1}{2}$, the poor would be in the majority). Let G be the Gini coefficient for the Lorenz curve $L(p)$, and let G^* be the Gini coefficient for the Lorenz curve $L^*(p)$. Notice that G^* is a lower bound on G. In terms of the labeling of Figure 3, the following measurements follow readily from geometry and/or trigonometry:[5]

(6) $AD = \sqrt{2}(1-p^*)$, $AC = 1/\sqrt{2}$, $CD = (2p^*-1)/\sqrt{2}$, $G^* = 2.\Delta ODB = 2p^*-1 < p^*$

Further, we may decompose G^* across the two groups, to yield

(7) $G = G^* + p^*(1-p^*)G_1^* + (1-p^*)p^*G_2^*$

where G_1 and G_2 are the within-group Gini coefficients (this result is due to Bhattacharya and Mahalanobis,1967: note that there is no overlap between these two groups). The geometry of (7) is apparent when we realize that, in Figure 3, G_1 is twice the area between *OD* and the first part of *L*(*p*), and G_2 is twice the area between *BD* and the second part of *L*(*p*).

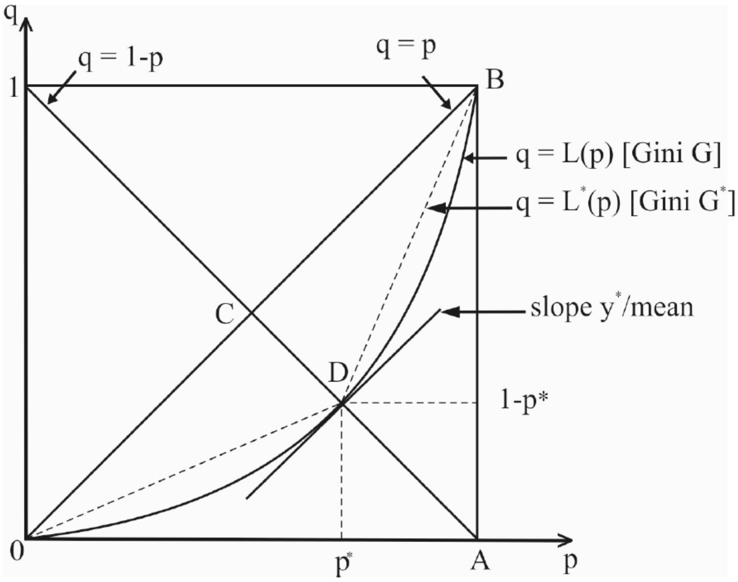

Figure 3: Diagrammatic Aid to an Understanding of the Discussion in Section 5

Let μ^* be the average income in the poorer group, so that, in view of what has gone before,

(8) $\mu^* = \mu L(p^*)/p^* = \mu(1-p^*)/p^*$.

Now from (6), $\mu^* = \mu\left\{1 - \frac{G^*}{p^*}\right\} = \mu\{1 - I^*\}$, say, where $0 < I^* < 1$; the average income of those having incomes not exceeding x^* can be interpreted as a 'Sen-type' welfare index. If x^* is the poverty line, we are talking about average income among the poor.

5 This and some of the subsequent analysis draws upon Subramanian's (2010) 'Tricks with the Lorenz curve'. See also Osmani's (1982) 'The algebra of the Lorenz curve', appendix 2.

We contend that the ratio measure $R^* \equiv {}^{\mu^*}\!/_{\mu} = 1 - I^* \equiv E^*$ (say), and the income measure $\mu^* = \mu(1 - I^*)$, are analogous to the Shaikh and Ragab measures $R_0 (= E_0)$ and $\mu_0 (= \mu(1 - G))$ respectively, with clearly similar welfare interpretations.

6 A second parallel

Now consider Figure 4, in which another Lorenz curve, labeled $\tilde{L}(p)$, is formed, by connecting O and B with E, which has coordinates $(\tilde{p}, 1 - \tilde{p})$ = $\left(\frac{1+G}{2}, \frac{1-G}{2}\right)$. Some elementary geometry and trigonometry will establish that the value of the Gini coefficient for $\tilde{L}(p)$ is also *G*. Note that $\tilde{p}$ always represents a majority of the population: by construction, the point E will coincide with the point *C*, where the two diagonals of the unit square intersect, when $G = 0$ (in which case, $\tilde{p} = ½$); and E will coincide with A when $G = 1$ (in which case, $\tilde{p} = 1$). Hence in general, $\tilde{p} \in [½, 1]$[6].

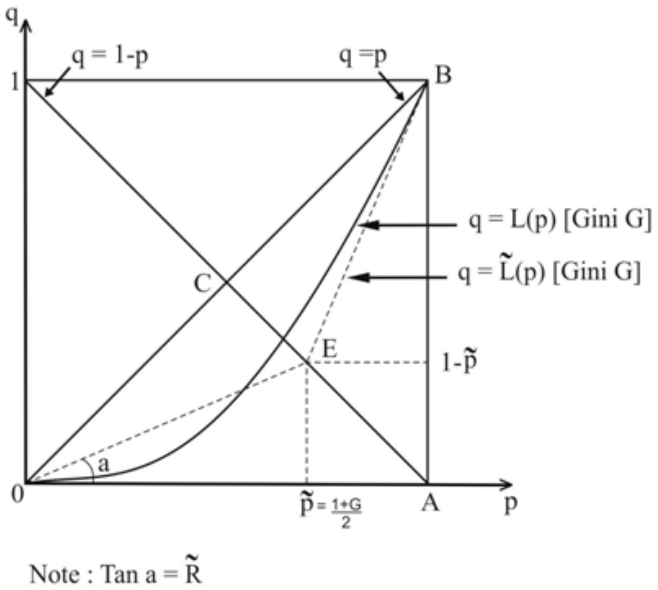

Figure 4: Diagrammatic Aid to an Understanding of the Discussion in Section 6

Let $\tilde{\mu}$ stand for the mean income of the poorest $\tilde{p}$ fraction of the population, and μ (as before) for the mean income of the entire population. Let $\tilde{R}$ be a short-hand for a new Majority of Incomes Ratio, defined as $\tilde{R} \equiv \tilde{\mu}/\mu$. Define the quantity $\tilde{E} \equiv (1 - G)/(1 + G)$. $\tilde{E}$ is a well-

[6] It might be as well to clarify that Figure 4 differs significantly from Figure 3. In Figure 4, *E* is *not* a point on the Lorenz curve $L(p)$ (whereas in Figure 3, the analogous point connecting with *O* and *B* by dotted straight lines, D, is). Both D and E are on the alternative diagonal.

defined measure of equality, which is declining in the inequality measure G (note that $\frac{d\tilde{E}}{dG} = -2/(1+G)^2 < 0$). It is easy to see that $\tilde{\mu}/\mu = (1-\tilde{p})/\tilde{p} = (1-G)/(1+G)$, whence

(8) $\tilde{R} = \tilde{E}$.

This indicates that the Majority of Incomes Ratio $\tilde{R}$ can be expressed as a measure of equality $\tilde{E}$ which is related to the Gini coefficient of inequality, in a manner reminiscent of the connections, reviewed earlier, between R_0 and E_0, and R^* and E^*, respectively. Next, let $\tilde{I}$ be an inequality measure given by $\tilde{I} = 1 - \tilde{E} \equiv 2G/(1+G)$. It is easy to see that

(9) $\tilde{\mu} = (1-\tilde{I})\mu$.

Hence $\tilde{\mu}$, like the other two average Incomes of the Majority μ_0 and μ^* which we have reviewed earlier, can also be interpreted as a Sen-type welfare function.

Briefly, while the Shaikh-Ragab results point to an empirically stable relationship between a Majority of Incomes Ratio and a measure of equality, as also between an average of a Majority of Incomes and a Sen-type welfare measure, we have analytically derived similar relationships between analogous quantities - relationships which hold not just empirically but deterministically, i.e. in a non-stochastic manner.

Finally, define the binary relation of strict Lorenz dominance $\succ_L$ as follows: for all cumulative distribution functions F and H, we shall say $F \succ_L H$ if and only if the Lorenz curve for F lies everywhere above the Lorenz curve for H. Then, it can be verified that R^* preserves order according to $\succ_L$, in the sense that, for all distributions F and H, if $F \succ_L H$, then R^* for the F-distribution is greater than R^* for the H-distribution. Similarly, it can be verified that $\tilde{R}$ preserves order according to the Gini coefficient G, in the sense that, for all distributions F and H, if the Gini for the F-distribution is smaller than the Gini for the H-distribution, then $\tilde{R}$ for the F-distribution is greater than $\tilde{R}$ for the H-distribution.

7 Concluding remark

Shaikh and Ragab have brought our attention to measures of distribution and welfare, and the relationship between them, which are insightful and productive; we, in this note, have sought to extend the Shaikh-Ragab findings in ways which we hope are helpful and interesting.

References

BHATTACHARYA, N. and B. MAHALANOBIS (1967). Regional disparities in household consumption in India. *Journal of the American Statistical Association,* vol. 62 (317), pp. 143–161.

DRAGULESCU, A. and V. M. YAKOVENKO (2001). Evidence for the exponential distribution of income in the USA. *The European Physical Journal B*, vol. 20, 585-589.

DRAGULESCU, A. and V. M. YAKOVENKO (2002). Statistical mechanics of money, income and wealth: a short survey. *arXiv:cond-mat/0211175*, vol. 1, pp. 1-4.

FLEURBAEY, M. and P. MONGIN (2005). The news of the death of welfare economics is greatly exaggerated. *Social Choice and Welfare*, 25(2), pp. 381-418.

FOSTER, J.E. and E. OK (1999). Lorenz dominance and the variance of logarithms. *Econometrica*, vol. 67, pp. 901-907.

GRUEN, C. and S. KLASEN (2008). Growth, inequality and welfare: comparisons across space and time. *Oxford Economics Papers*, 60(2), pp. 212-236.

LAMBERT, P.J. (1989). The Distribution and Redistribution of Income: A Mathematical Analysis. Oxford, Basil Blackwell.

LAMBERT, P.J. (2011). Positional equity and equal sacrifice: design principles for an EU- wide income tax? Pages 241-259 in Fleurbaey, M., M. Salles and J.A. Weymark (eds.) *Social Ethics And Normative Economics: Essays in Honour of Serge- Christophe Kolm.* Springer-Verlag series *Studies in Choice and Welfare.* Berlin and Heidelberg, 2011.

OSMANI, S.R. (1982). *Economic inequality and group welfare: a theory of comparison with application to Bangladesh*. Oxford: Clarendon Press.

SEN, A. (1976). Real national income. *Review of Economic Studies*, 43(1), pp. 19-39.

SHAIKH, A. and A. RAGAB (2007). An international comparison of the incomes of the vast majority. *SCEPA Working Paper 2007-3*, Schwartz Center for Economic Policy Analysis, The New School, New York.

SHAIKH, A. and A. RAGAB (2008). The vast majority income (VMI): a new measure of global inequality. *Policy Research Brief No. 7*, International Poverty Centre, Brazil

SHORROCKS, A.F. 1983: Ranking income distributions. *Economica*, vol. 50, pp. 1-17.

SILVA, C.A. and V.M. YAKOVENKO (2004). Temporal evolution of the ‘thermal’ and ‘superthermal’ income classes in the USA during 1983-2001. *Europhysics Letters*, vol. 3(31), pp. 1-7.

SUBRAMANIAN, S. (2010). Tricks with the Lorenz curve. *Economics Bulletin*, vol. 30(2), pp. 1594-1602.

YAKOVENKO, V.M. (2007). Statistical mechanics approach to econophysics. *Encyclopedia of Complexity and System Science* (ed. R. A. Meyers), Springer 2009.