A critical assessment of UNDP's Gender Inequality Index

Abstract: This paper critically reviews the Gender Inequality Index (GII), the new gender-related index proposed by United Nations Development Program (UNDP) in the 2010 Human Development Report. We argue that the particular ways in which the index was constructed limit its usefulness and appropriateness as a global gender inequality index. In particular, we contend that the functional form of the index might be excessively and unnecessarily confusing. Moreover, the inclusion of indicators that compare the relative performance of women vis-à-vis men together with absolute women-specific indicators obscures even more the interpretation of an already complicated index that turns out to penalize the performance of low-income countries. In order to overcome some of the identified limitations we define a new composite index of gender inequality that incorporates the GII variables but uses a much simpler functional form. Our results suggest that the interpretation of GII values should be heavily qualified.

Keywords: Measurement, Gender inequality, Composite Indicator, Human Development.

JEL Classification: D63, J16, O15.

1. Introduction

In 1995, the United Nations Development Program (UNDP) launched a pathbreaking Human Development Report (HDR) focusing on gender. In that report, the Gender-related Development Index (GDI) and the Gender Empowerment Measure (GEM) were presented (see UNDP 1995). These were the first composite indices originally designed to reflect gender disparities in basic capabilities at the world level. The impact of these two measures has been enormous both in academic and non-academic circles and their values have been widely used for the purpose of assessing disparities between women and men all over the world (Dana Schüler 2006). Among other things, these indices were particularly useful to raise awareness on gender related issues in the context of human development.

Despite their relevance, the GDI and GEM have been criticized for their conceptual and methodological limitations, as has been widely acknowledged elsewhere (see Kalpana Bardhan and Stephan Klasen 1999, 2000; A. Geske Dijkstra and Lucia Hanmer 2000; A. Geske Dijkstra 2002; Stephan Klasen 2006; A. Geske Dijkstra 2006; Schüler 2006). As a result of the many limitations of the GDI and GEM, a plethora of gender-related well-being indices has appeared in the literature [[[Endnote#1]]]. In turn, these alternative indices suffer from alternative shortcomings that limit somehow their usefulness and appropriateness as global gender inequality indices (Angela Hawken and Gerardo Munck 2009, Iñaki Permanyer 2010 and Eduardo Bericat 2011 provide extensive and critical reviews on that literature). In an attempt to overcome some of the problems identified by different researchers during the past fifteen years, the 2010 HDR has presented a new measure: the Gender Inequality Index (GII). This index has been designed to capture women's disadvantage in three dimensions – empowerment, economic activity and reproductive health – for 138 countries around the world. The publication of a new global index of gender inequality by UNDP is good news for at least two reasons. On the one hand, the GII brings fresh air by substituting a couple of indices that, despite their importance, have been criticized on many fronts. On the other hand, it further contributes to the debate on gender inequality measurement by incorporating concepts and dimensions that had not been used before in that context at the global level.

While the GII is an interesting novel way of conceptualizing gender inequality, we argue that the particular ways in which the index was constructed limit its usefulness and appropriateness as a global gender inequality index. As we will show, the functional form of the index has been unnecessarily and excessively complicated to satisfy a series of normative properties that are otherwise satisfied by much simpler indices as well. We also suggest that the incorporation of some indicators where the achievements of women and men are compared vis-à-vis each other together with some indicators that are only defined for women obscures even more the conceptualization and interpretation of an already complicated index which, as will be shown in the empirical section, turns out to penalize the performance of low-income countries. Whenever it is feasible, we will propose constructive alternatives to some of the identified limitations. In particular, we define a new composite index of gender inequality that incorporates the GII variables but uses a much simpler functional form that has been recently used in the literature (Iñaki Permanyer 2010; Lourdes Beneria and Iñaki Permanyer 2010; Stephan Klasen and Dana Schüler 2011). Our results indicate that the impression that one gets of gender gaps in development across the world when comparing the values of the new index with those of GII differs substantially. Therefore, careful attention should be paid when working with the Gender Inequality Index presented in the 2010 Human Development Report, whose values should be heavily qualified.

2. Introducing the GII

The Gender Inequality Index has been designed to overcome the most important limitations of the GDI and GEM. For that purpose it has proposed (i) new dimensions to capture gender inequality and (ii) a new functional form to summarize multidimensional information into a real number that can be eventually used to compare countries' performance over time. We will now briefly review these points separately.

2.1. Selection of dimensions and indicators

When choosing the dimensions and indicators to include in the GII, its designers have kept in mind some basic criteria for indicator selection (Amie Gaye, Jeni Klugman, Milorad Kovacevic, Sarah Twigg and Eduardo Zambrano 2010:9). Given the severe quality data limitations at the global level, there are important trade-offs between data relevance/importance and geographical coverage. Fortunately, the GII designers have been able to identify highly relevant indicators for a fairly large amount of countries – 138 – that cover most parts of the world. The final choice of dimensions and indicators is the following:

Dimension 1: Reproductive health. This is an extremely relevant dimension to evaluate individuals' well-being levels that has been completely absent in other well-known UNDP composite indices, like the Human Development Index (HDI), the Human Poverty Index (HPI), the GDI and the GEM [[[Endnote#2]]]. To capture countries' reproductive health situations, two indicators have been used: the maternal mortality ratio (MMR) and the adolescent fertility rate (AFR). The maternal mortality ratio, which is defined as the number of maternal deaths per 100000 live births, captures a leading cause of death and disability among women of reproductive age in developing countries. The adolescent fertility rate is defined as the number of births per 1000 women aged 15-19, so it represents the risk of childbearing among adolescent women. This is an important indicator because women that have children at such young ages see their health and future opportunities in life compromised.

Dimension 2: Empowerment. The notion of women's empowerment gained momentum since the International Population Conference held in Cairo in 1994 and has been gradually incorporated in national and international institutions' everyday parlance. Given the complexity and looseness of its definition, it is a particularly difficult concept to measure (see Sabina Alkire and Solava Ibrahim, 2007). The indicators chosen for the GII are educational attainment (secondary level and above, henceforth SE) and parliamentary representation (henceforth PR). Education is an essential factor that contributes to the creation of knowledge and self-confidence. It is widely acknowledged, both on theoretical and empirical grounds, that literacy-related skills are a necessary condition to escape out of poverty. Education brings empowerment because it strengthens people's capacity to question and act on one's condition and increases accessibility to the information needed to do so. On the other hand, PR is a crude but widely available measure of women's access to the levers of power.

Dimension 3: Economic activity. When measuring economic activity, the GII uses the gender specific labor force participation rates (LFPR). This is an important variable that replaces the problematic gender-specific earned income component that was used both in the GDI and GEM (as has been argued in Bardhan and Klasen 1999, that component was estimated in many countries using rather questionable assumptions). The measurement of the LFPR is much more reliable, even though it is not devoid of certain problems. As mentioned in Gaye et al. (2010:14): '*Labor force participation, as traditionally measured, ignores the important contributions of women in unpaid work and may perpetuate the undervaluing of these critical activities*'. Therefore, much work still needs to be done on the appropriate measurement tools to capture the informal and care economy sectors in which women are typically overrepresented (Nancy Folbre 2006).

2.2. Methodology

Before presenting the methodology used to compute the GII we will introduce some basic notation that will be used throughout the paper. The average achievement levels of women and men in the aforementioned indicators will be coded in a *2×5* matrix referred as *achievement matrix*. A generic element in that matrix, h_{ij}, corresponds to the average achievement level of gender '*i*' in indicator '*j*'. It is assumed that these average achievements are always non-negative: $h_{ij} \geq 0$. The first row in that matrix contains the average achievement levels for women in the five indicators, while men's achievements are coded in the second row. Hence, a generic achievement matrix looks like

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} & h_{14} & h_{15} \\ 1 & 1 & h_{13} & h_{14} & h_{15} \end{pmatrix} = \begin{pmatrix} MMR & AFR & SE_f & PR_f & LFPR_f \\ 1 & 1 & SE_m & PR_m & LFPR_m \end{pmatrix}$$

where the subscripts *f,m* refer to female and male indicators respectively. Given the fact that there are no male counterparts for MMR and AFR, the first two cells in the second row are fixed at 1. This value is interpreted as the reference benchmark that MMR and AFR should attain in case of "perfect gender equality" [[[Endnote#3]]]. The set of achievement matrices will be denoted by *A*. The Gender Inequality Index can thus be formally defined as a function GII:$A \rightarrow$R that for any achievement matrix $H \in A$ assigns a real number GII(H).

The GII has been basically constructed to satisfy two normative assumptions – or axioms: 'symmetry in gaps' and, more importantly, 'association sensitivity'. Symmetry in gaps simply states that the gender gaps favoring women should be treated equally as the gender gaps favoring men. One consequence of adopting this axiom is that the values of the GII do not inform on the average relative position of women vis-à-vis men, an issue that will be further discussed in section 3. On the other hand, association sensitivity is an axiom borrowed from Suman Seth (2009a) adapted to the context of gender inequality measurement. Originally, Seth (2009a) introduced that axiom for the measurement of multidimensional welfare. Loosely speaking, an 'association sensitive social welfare index' can be thought of as a welfare index defined in a *n*-person multidimensional framework that is responsive to changes in association between indicators, that is: penalizing or rewarding those distributions where some individuals perform better than others in *all* dimensions at the same time. When adapted to the measurement of gender inequality, association sensitivity requires an index to be responsive to those changes in association between indicators that end up benefiting one gender over the other in all dimensions at the same time (see section 3 for more details).

With the main purpose of having an association sensitive index, the GII has been defined as follows:

$$GII(H) = 1 - 4\frac{\left[\left(\sqrt[3]{\left(\frac{1}{MMR}\frac{1}{AFR}\right)^{1/2}\left(PR_f SE_f\right)^{1/2} LFPR_f}\right)^{-1} + \left(\sqrt[3]{\left(PR_m SE_m\right)^{1/2} LFPR_m}\right)^{-1}\right]^{-1}}{\sqrt[3]{\left(\sqrt{\frac{1}{MMR}\frac{1}{AFR}}+1\right)\left(\sqrt{PR_f SE_f}+\sqrt{PR_m SE_m}\right)\left(LFPR_f + LFPR_m\right)}} \quad [1]$$

where $H \in A$ is an achievement matrix. For practical purposes, the values of MMR are truncated at 10 (minimum) and 1000 (maximum) before entering into the formula. Analogously, the female parliamentary representation of countries reporting 0 is coded as 0.1 to avoid zero denominators. As shown in Gaye et al (2010), *1-GII* is the ratio between an average of average achievement levels of women and men and an average of average achievements in the different dimensions [[Endnote#4]]. In order to be able to compute averages in the women-specific variables of the reproductive health dimension, their male counterparts are fixed at 1, a value representing a benchmark of "perfect achievement level" that should be attained in case of complete gender equality. The values of GII should be interpreted as the loss in human development due to gender inequality accounting for association – or overlap – between dimensions. Because of the 'symmetry in gaps' axiom, it is not possible to know the relative position of men vis-à-vis women looking at the values of the GII alone. By construction, the values of the index are bounded between 0 and 1. According to the designers of the index:

'*The Gender Inequality Index will be equal to 0 if women and men fare equally well in each dimension. The Gender Inequality Index tends to 1 [...] if the gap between women's and men's achievement is increasing*' (Gaye et al. 2010:34).

While the novel approaches taken in the definition of the GII (introducing both new indicators and new ideas in the aggregation methodology) are to be welcomed for their contribution to the debate on gender inequality measurement, some convincing explanations seem warranted to justify the fearsome appearance of equation [1]. This critical issue will be analyzed in detail in the following section.

3. Critical Review

In this section we will critically review some of the main shortcomings of the GII. We start analyzing the chosen dimensions and indicators and then proceed with the aggregation methodology and its adverse consequences.

Chosen dimensions and indicators.

The fact that the earned income component is not present in the GII (as opposed to what happens with the GDI and GEM) is certainly good news. As argued in Bardhan and Klasen (1999:992-993), that component has to be estimated in a large set of countries using very questionable assumptions, thus producing highly unreliable results. Moreover, that component does not account for inequalities in intra-household distribution of resources, an important component in the analysis of gender inequality. In contrast, the labor force participation rate is a much more reliable estimate of economic participation that, unfortunately, fails to capture the informal and care economy sectors in which women are typically overrepresented. Given the growing body of literature on the care economy, the introduction of variables reflecting inequalities related to this sector may be a project whose time has come (see Folbre 2006; Lourdes Beneria 2008).

Another issue that is worth mentioning is the fact that the GII combines both well-being and empowerment indicators into a single measure. This is in contrast with the separation established by UNDP in the mid-90s with the creation of a gender-related index that only included well-being indicators on the one hand – the GDI – and another measure that included empowerment indicators on the other – the GEM. We contend that, for the sake of conceptual clarity, it might have been advisable to construct composite indices with either well-being or empowerment variables separately. In this respect, the parliamentary representation component (an empowerment indicator) might not fit conceptually with the other well-being indicators included in the GII. Moreover, there is a further technical issue already suggested in Klasen and Schüler (2011:21) that renders the PR component problematic. In order to reach gender parity, the share of women in the parliament should equal the share of women in the corresponding population. Therefore, in a country where the share of women was larger than the share of men, a 50-50 parliamentary partition among men and women would under-represent the latter. In order to correct this problem it would be necessary to introduce somehow the population shares of women and men into the GII formula. This issue will be addressed in section 4.

An innovative aspect of the GII is the inclusion of reproductive health variables like MMR and AFR, which is surely motivated by the desire of incorporating a dimension that is essential for women's well-being levels in the assessment of gender inequality. Unfortunately, the fact that the GII incorporates indicators that are computed for women and men (SE, PR, LFPR) together with indicators that are women-specific (MMR, AFR) does lead to serious conceptual and methodological problems that will be analyzed in detail. Loosely speaking, if all indicators were women-specific we might be talking about something like a "women status measure" – that is: an index that could be used to assess women's *absolute* achievement levels –, and if all indicators were available and comparable for women and men we might have the possibility of constructing something like a "gender inequality measure" – that is: an index that could be used to assess the *relative* position of women vis-à-vis men. Including the two kinds of indicators simultaneously, the GII becomes an odd mixture that is halfway between both concepts, thus obscuring even more the interpretation of an already complicated index. Moreover, there are many other methodological problems derived from this mixture of indicators to which we now turn.

Because of this indicators mixture, there are different properties one would naturally expect to see in a gender inequality index that are simply not satisfied. To start with, it is far from clear why, on the one hand, increases in MMR and AFR do systematically represent a worsening of gender inequality levels while, on the other hand, decreases in women's education or labor force participation do not necessarily represent a worse state of affairs *as long as* men's education and labor force participation decrease by the same amount. In other words: why should one allow a deterioration of women's education and economic participation to be compensated by an equivalent deterioration in men's corresponding dimensions while at the same time not allowing for any kind of compensation when a deterioration in women's reproductive health conditions takes place?. Secondly, when women and men fare equally well in each dimension, one would expect that the index reaches a normatively desirable extreme of its range distribution (typically a maximum or a minimum depending on whether the index measures gender equality or inequality). As a matter of fact, according to Gaye et al. (2010, p. 34), '*the GII will be equal to 0 if women and men fare equally well in each dimension*'. However, given the fact that the GII incorporates variables that are meaningful for women only, this is simply not true. Consider any achievement matrix $H \in A$ where the achievement levels for women and men in PR, SE and LFPR are exactly the same. It is straightforward to check that the only way to have GII(H)=0 for such an achievement matrix is to impose (MMR)(AFR)=1, an overly unrealistic assumption that, as a matter of fact, is not observed in any of the countries with available data nor is very likely to be observed in any country at all. In general, whenever (MMR)(AFR)≠1, one has that GII(H)>0, a disturbing result for the Gender Inequality Index. To illustrate, consider a hypothetical country with $PR_f = PR_m$, $SE_f = SE_m$, $LFPR_f = LFPR_m$ and with the lowest MMR and AFR observed in the sample of countries for which data is available (this is MMR=10 (truncated) and AFR=3.8 observed in The Netherlands). In that case, such hypothetical country would have a GII value well above 0 (GII≈0.15, that is: a 15% of its maximal (potential) value of 1). Even more disturbingly, if both MMR and AFR were allowed to continuously approach the value of zero (that is: to approach a state of affairs with absence of maternal mortality and adolescent fertility), the values of the GII would approach the normatively undesirable value of 1.

An intended but debatable feature of the GII is the following: for any achievement matrix $H \in A$ where the achievement levels for women and men in PR, SE and LFPR are exactly the same, the values of GII(H) are strictly increasing in MMR and AFR. For illustrative purposes, consider the following (hypothetical) achievement matrices:

$$H_1 = \begin{pmatrix} 300 & 50 & x & y & z \\ 1 & 1 & x & y & z \end{pmatrix}; H_2 = \begin{pmatrix} 10 & 5 & x & y & z \\ 1 & 1 & x & y & z \end{pmatrix}$$

where *x,y,z* represent any meaningful achievement levels. It can be easily checked that GII(H_1)=0.57>GII(H_2)=0.17 for any *x,y,z*. While the proponents of the index might rightly argue that it makes sense to "penalize" those countries with bad reproductive health conditions for women, it is fair to say that countries' performance in those areas is influenced by a myriad of factors other than gender-related issues *tout court*. Risks associated with childbearing vary tremendously globally and locally within countries, reflecting differences in access to and use of health services, social and cultural practices affecting access to healthcare, socio-economic levels and public health policies. Therefore, while it is true that gender norms and practices exert an important influence on MMR and AFR values, these are by no means the only influencing factors. For instance, both MMR and AFR are strongly and negatively associated to countries GDP per capita (see Figures 1,2). Other things being equal, richer countries have better health facilities and communications infrastructure that could contribute to reduce MMR. As can be seen in Figure 1, there are marginally increasing returns (in terms of MMR reduction) to increases in the GDP per capita. Analogously, richer countries tend to have education and production systems that discourage pregnancies at very young ages, thus lowering the corresponding AFRs (Figure 2). As a consequence, the GII is implicitly penalizing poorer countries for certain structural relationships that are not always explicitly related to gender norms nor discrimination against women. To avoid this problem it would be necessary to factor out the "pure" effects that gender-related norms and practices have on the MMR and AFRs, an overly complex exercise that would probably be vulnerable to both methodological and conceptual criticisms.

[[[Figure 1]]]

[[[Figure 2]]]

In this context, it is insightful to explore the relationship between the GII and the GDP per capita: it is shown in Figure 3. It can be observed that a strong, negative and roughly linear relationship exists between them. The correlation coefficient between these two variables is remarkably high (in absolute terms), r= –0.87, therefore suggesting that the information conveyed by the GII is not substantially different from the GDP per capita [[[Endnote#5]]]. This is reminiscent of the close relationship between the GDI and the GDP per capita that has been highlighted and criticized in other papers (see Dijkstra and Hanmer 2000:48). The fact that MMR and AFR are so strongly related to the GDP per capita and that these indices have no male counterpart does contribute decisively to the strong negative relationship between GII and GDP per capita. If MMR and AFR were dropped altogether from the GII (therefore only using the variables with female and male achievement levels: SE, PR and LFPR), the correlation coefficient between such "capped GII" and GDP per capita would drop to –0.34, therefore indicating a much weaker relationship between gender inequality and per capita income levels. This issue will be discussed in further detail in section 4.

[[[Figure 3]]]

An extremely important consequence of just incorporating women-specific health variables is that the health status of men is completely disregarded from the evaluation of gender inequality levels performed by the GII. Even if GII designers claim that the value of 1 must be interpreted as the "perfect achievement level" in MMR and AFR that should be attained in case of perfect gender equality, for practical purposes when interpreting the meaning of the index it is *as if* male's average health status was artificially fixed at its highest possible level. However, it can be shown that in those countries with high maternal mortality the average health conditions of men tend not to be very good either. Figure 4 shows that higher values of MMR are strongly associated with lower male life expectancies. If certain male health status variables were incorporated in the GII, it would turn out that the health gender gap in those countries would not be as large as the values of MMR alone would suggest. As a matter of fact, if some male-specific variables are fixed at the "perfect achievement level" in a gender inequality measure that is supposed to capture the relative performance of women vis-à-vis men, the values of the aggregate index will naturally tend to be dominated precisely by those variables (which will tend to have larger gaps than the others). Hence, it should come as no surprise the finding that, '[*according to the GII values*] *reproductive health is the largest contributor to gender inequality around the world*' announced in the 2010 and 2011 Human Development Reports (chapters 5 and 3 respectively). This argument is not to deny the importance of reproductive health variables in the assessment of gender inequality levels, but rather to emphasize that the lack of male health status variables inevitably biases the results penalizing those countries with high maternal mortality levels (which, in turn, have lower GDP per capita levels). The decision of not incorporating the male and female counterparts for each dimension has therefore many pernicious consequences for the GII.

[[[Figure 4]]]

Aggregation methodology.

As the different UNDP reports and notes on the Gender Inequality Index make clear from the very beginning, the GII has been built on the same framework as the new Human Development Index and, particularly, the Inequality-adjusted Human Development Index (IHDI): both the IHDI and GII take into account the losses in human development due to different kinds of existing inequalities. The fact that the new family of UNDP indices presented in the 2010 Human Development Report have been crafted with the same underlying methodology is helpful to project a unified and coherent vision on what human development is and how it should be measured (analogous efforts to have methodologically coherent measures were also made in the 'first wave' UNDP indices: the HDI, GDI, GEM and HPI). However, while that methodology might be useful to generate social welfare measures, it might not be very appropriate for the measurement of gender inequality. To start with, the GII is supposed to measure '*losses in human development due to gender inequality*' and the IHDI '*adjusts the Human Development Index (HDI) for inequality in distribution of each dimension across the population*'. In the hypothetical case there were no inequalities in the underlying distribution, the IHDI would be equal to the new HDI. For this reason, the new HDI is supposed to measure the 'maximal' or 'potential' human development (for details see Technical note #2 in 2010 Human Development Report) and the IHDI can be thought as a welfare function adjusted downwards for existing inequalities. Analogously, when there are no disparities between women and men, the GDI is actually equal to the old HDI. However, in the case of the GII, it is not clear what the underlying welfare function is. When women and men fare equally well in each dimension, its designers – wrongly – claim it should take a value of zero, so there should be no welfare loss; but welfare loss from what 'maximal' or 'potential' measure? Clearly, it is not the HDI – as the individual variables that compose these measures are completely different – so the conceptual foundations of the GII do not seem to be completely specified.

A distinctive feature of the GII is that it has been designed to satisfy the axioms of 'symmetry in gaps' and 'association sensitivity'. It is precisely the adoption of strong versions of these axioms that generates the peculiar functional form of the index shown in equation [1]. Now, what is the rationale for introducing an association sensitive index in the context of gender inequality measurement?

1. To start with, an association sensitive gender inequality index is responsive to those distributional changes that end up benefiting one gender over the other in all indicators at the same time. In other words, if one defines the achievement

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} & h_{14} & h_{15} \\ 1 & 1 & h_{13} & h_{14} & h_{15} \end{pmatrix}$$

$$H' = \begin{pmatrix} \max\{h_{11},1\} & \max\{h_{12},1\} & \max\{h_{13},h_{23}\} & \max\{h_{14},h_{24}\} & \max\{h_{15},h_{25}\} \\ \min\{h_{11},1\} & \min\{h_{12},1\} & \min\{h_{13},h_{23}\} & \min\{h_{14},h_{24}\} & \min\{h_{15},h_{25}\} \end{pmatrix}$$

$$H'' = \begin{pmatrix} \min\{h_{11},1\} & \min\{h_{12},1\} & \min\{h_{13},h_{23}\} & \min\{h_{14},h_{24}\} & \min\{h_{15},h_{25}\} \\ \max\{h_{11},1\} & \max\{h_{12},1\} & \max\{h_{13},h_{23}\} & \max\{h_{14},h_{24}\} & \max\{h_{15},h_{25}\} \end{pmatrix}$$

and if H≠H′,H≠H″ [[[Endnote#6]]] then one has that GII(H)≠GII(H′) and GII(H)≠GII(H″) whenever GII is an association sensitive index. Technically speaking, H′ and H″ are obtained from H after an 'association increasing transfer' of the dimensions where we compare the achievement levels of women and men (see Seth 2009a,b). This is an interesting property that allows identifying those situations where one gender might be systematically discriminated against the other.

With regard to this property, it is important to point out that the direction of change (i.e: improvement or deterioration) of an association sensitive index under an association increasing transfer depends on whether all couples of attributes we are taking into account are substitutes or complements. In the *n*-person multidimensional welfare context, when all couples of attributes are assumed to be substitutes (resp. complements), association increasing transfers should decrease (resp. increase) overall welfare (see Seth 2009a,b). The choice made in the GII of penalizing association increasing transfers implicitly assumes that *all* couples of attributes are substitutes with an elasticity of substitution equal to one. To the best of our knowledge, however, this assumption is not based on empirical findings but has rather been taken for practical reasons. On the other hand, it is fair to say that under similar circumstances – and in face of the daunting task of measuring the degree of complementarity or substitutability between attributes in a multidimensional setting – other well-known papers have taken a similar approach because the substitutability assumption seems to be ethically justified in many cases [[[Endnote#7]]]. Importantly for our purposes in this paper, we would like to emphasize that GII is by no means the only index satisfying the axiom of association sensitivity; as will be shown in section 4 there are much simpler indices satisfying this reasonable property too.

2. Another motivation to introduce association sensitive indices is that, in the context of multidimensional *n*-person welfare measurement, such indices are capable – under certain restrictive assumptions (see below) – of targeting the specific individual *and* specific dimension that should receive priority assistance to maximize the increase of social welfare in case an extra indivisible dollar was available to the corresponding policy-maker (see Seth 2009a,b for details). As argued by Seth, this useful decision-making rule would not be possible without the property of association sensitivity. Adapted to the context of gender inequality measurement, an association sensitive index is capable – under the same restrictive assumptions as before – of identifying the gender and dimension that should be the target of social welfare programs within each country.

With respect to this property, it is important to emphasize that an association sensitive index belonging to the family of indices presented in Seth (2009a,b) can be used for targeting purposes only when *all* couples of attributes have exactly the same degree of complementarity or substitutability between them. In case this restrictive underlying assumption failed to be true, the decision-rule advocated by Seth (2009a,b) might target the wrong person and/or dimension that should receive priority assistance. As before, we are not aware of any attempt to assess empirically the relationship between the different dimensions, so the validity and usefulness of the GII for those targeting purposes is by no means warranted.

On this targeting issue, however, we would like to point out that the problem of identifying the most deserving recipient of extra assistance is way easier in the context of gender inequality where we only compare two groups: women and men. To make things even simpler, in most cases there is no doubt whatsoever on whether it is women or men that should be targeted for priority assistance: in 76% of the countries with available GII data men score higher average achievement levels than women in *every* indicator. In the other 24% of countries with available data, women slightly perform better than men in the education component only, that is: in one out of the five indices that compose the GII. Therefore, in the vast majority of cases, there is hardly any doubt on whether priority assistance should focus on women or men, so the value-added of this targeting property is seriously called into question in the context of gender inequality measurement.

Another important axiom in which the GII is implicitly based is 'symmetry in gaps'. As a consequence, the GII does not take into account the direction of the gender gaps (i.e: whether they favor women or men) so it is not possible to determine the relative position of men vis-à-vis women from the values of GII alone. Rather, the GII focuses on welfare losses due to gender inequalities. Therefore the GII does not measure gender inequality *per se*, but incorporates implicitly the gender gaps to assess losses in aggregate welfare: it could then be argued that the name of the index is misleading [[[Endnote#8]]]. We further argue that the insensitivity to the direction of the gender gaps would make more sense in other contexts where the underlying dimensions over which we compare the performance of women and men were not normatively desirable. For instance, if we compare the presence/absence of women and men in certain non-ordered categorical groups – consider the sex segregation indices according to geographical locations or different economic sectors (see Maria Charles and David Gursky 1995, 1998) – the direction of those gender gaps does not necessarily imply a worse state of affairs for the underrepresented group. On the other hand, in contexts where the underlying dimensions are normatively desirable – as is the case with the GII –, we contend that the direction of the gender gaps are essential pieces of information that should be taken into account in our evaluative exercises. The achievement distributions of women and men that are observed in a given year, say 2010, are not the result of a sudden "random allocation process" but, to a great extent, they are rather the result of discriminative norms and practices against women that have been pervasive both in time and space. A symmetric treatment of the gender gaps that only cares about the absolute differences between women and men but not on who benefits from them would seem to be more in line with the former point of view rather than the latter.

Finally, given the fact that the GII is interested in measuring the welfare loss that can be accounted for by gender inequalities (regardless of whether these inequalities favor women or men), one would naturally expect that a permutation of the achievement levels between women and men in all dimensions – that is: permute PR_f with PR_m, SE_f with SE_m and $LFPR_f$ with $LFPR_m$ – should leave the values of the index unaffected. As can be readily checked out with any achievement matrix, this is simply not true because of the lack of male indicator counterparts for MMR and AFR.

Implications

Because of the efforts to use a methodology that is consistent with the one that has been used in the construction of other new UNDP indices – those presented in the 2010 HDR – and because of the overreliance on the association sensitivity axiom, the GII is an unnecessarily confusing index. The true meaning of the values of the index is not crystal clear (what does it *really* mean 'loss in human development due to gender inequality accounting for association – or overlap – between dimensions'?), particularly for those practitioners not acquainted with relatively sophisticated concepts from economic theory jargon, like welfare function, association increasing or decreasing transfers, complementarity, substitutability and the like. Should the GII values be interpreted as a percentage loss with respect to some 'maximal' or 'potential' human development level – which has not been specified anywhere? Echoing a past comment of Amartya Sen on Theil's inequality index (James Foster and Amartya Sen 1997), we can conclude that the GII 'is not a measure that is exactly overflowing with intuitive sense'.

Another adverse consequence of using association sensitive indices is that it is not possible to decompose the values of the aggregate index into the respective contributions of its subcomponents. This point has already been emphasized and discussed in Seth (2009b:394). In that paper, the author ponders whether this limitation is compensated by the purported ability of association sensitive indices of targeting the individuals and dimensions that deserve more urgently an extra assistance. In the light of the present discussion – where the usefulness of this "targeting ability" is put into question in the context of gender inequality measurement (see above) – the property of subcomponent decomposability seems to be more attractive. Furthermore, we contend that the possibility of decomposing the aggregate value of a composite index into the corresponding contributions of its subcomponents is an important property for at least three reasons: i) Its intrinsic interest, ii) Its usefulness for a precise and detailed exploration of the internal structure of the index, and iii) When satisfied, it is a good argument against those who argue that the values of composite indices are like black boxes in which apples and oranges are lumped together to generate a meaningless measure (see OECD 2008, Martin Ravaillon 2010). In section 4 we propose alternative gender inequality indices that are subcomponent decomposable.

To sum up: the different arguments presented so far seriously put into question the validity of the methodology used to construct the GII. In the next section we present some constructive alternatives that try to overcome some of the identified shortcomings.

4. Some constructive proposals

In the previous section we basically identified two kinds of problems: the first one related to the chosen aggregation function and the second one to the mixture of absolute/women-specific and relative/"woman vs men" indicators into a single formula. In order to remedy the first problem we propose to use much simpler and intuitive indices. In this respect, the functional form of the indices used in Permanyer (2010), Beneria and Permanyer (2010) and Klasen and Schüler (2011) can be particularly useful. Let us denote by x_i, y_i the average female and male achievement levels in indicator 'i' and let $I_M=\{i \mid x_i < y_i\}$ be the list of indicators for which the corresponding gender gap strictly favor men. With this notation, the new indices can be constructed using the following functional forms:

$$GRS = \prod_{i=1}^{n} \left(\frac{x_i}{y_i} \right)^{w_i} \qquad [2]$$

$$WD = \prod_{i \in I_M} \left(\frac{x_i}{y_i} \right)^{w_i} \qquad [3]$$

where n is the number of indicators we are taking into account and w_i is the weight attached to indicator 'i' (representing its relative importance vis-à-vis other indicators). The acronym GRS stands for Gender Relative Status index and WD for Women Disadvantage index [[[Endnote#9]]]. The interpretation of the values of these indices is simple and clear: GRS is just an average of all gender gaps; whenever GRS<1 men are on average better-off than women and when GRS>1 women are on average better-off than men. The main problem with the GRS is that it combines into a single formula gender gaps running in opposite directions. This can muddy the waters because of the possibility of compensation between dimensions that can lead to a distorted picture of the existing levels of gender inequality [[[Endnote#10]]]. This problem is avoided using WD, an index that only averages the gender gaps favoring men [[[Endnote#11]]]. The values of WD are an average ratio of women's vs men's achievement levels in those dimensions where men outperform women, so they can be interpreted as a measure of the extent to which women are disadvantaged with respect to men. Given the fact that in most countries men outperform women in all (or almost all) well-being dimensions, the differences between GRS and WD are negligible (see below).

The new GRS and WD indices are much simpler than the GII – compare equations [1], [2] and [3] – and they are association sensitive too: it is straightforward to check that both indices are responsive to association increasing transfers. Moreover, as will be shown below, WD has the further advantage of being decomposable by subcomponents [[[Endnote#12]]] – thus allowing to know the percent contribution of each individual subcomponent to the aggregate value of the index.

Recall that the GRS and WD do not leave room for women-specific indicators, thus resulting in conceptually clearer measures that avoid many of the problems identified in the previous section. The problem, then, is what to do with the women-specific components of the GII. A first drastic alternative might be simply to drop altogether those components from the index and construct a new version that uses the GII methodology but only including the variables SE, PR and LFPR. As mentioned before, such an index will be called 'capped GII' and denoted as $\overline{GII}$. Unfortunately, this might be too crude as the important health dimension disappears completely from our assessment. Ideally, it would be desirable to have meaningful reproductive health indicators for men too, as this is a crucial dimension of human well-being that has been typically neglected in global assessments of gender inequality. However, the ways in which reproductive health issues affect the lives of women and men are completely different, so it is far from clear how such an indicator should be constructed for men in a way that it was meaningfully comparable with a women's reproductive health indicator [[[Endnote#13]]]. Because of the lack of appropriate data, we suggest to replace the interesting but problematic reproductive health variables by the 'classic' gender-specific life expectancies. While imperfect, these variables are routinely collected with reasonable quality in almost all countries of the world and have been used for fifteen years in the GDI and other well-known gender inequality indices presented in the literature (e.g: Dijkstra and Hanmer's (2000) Relative Status of Women; Beneria and Permanyer's (2010) Gender Relative Status and Women Disadvantage indices; Klasen and Schüler's (2011) Gender Gap Measure). The normalized gender-specific life expectancy indices [[[Endnote#14]]] are denoted by LEI_f and LEI_m.

In section 3 we mentioned that the 'parliamentary representation' component – PR – was problematic both on conceptual and technical grounds. As before, a simple drastic alternative would be to drop that component from the index, so as to have new measures with well-being variables only. However, in order to ensure a more meaningful comparison with the original GII formulation, we have decided to keep the PR component in the new GRS and WD measures. To reflect the idea suggested in Klasen and Schüler (2011:21) that gender parity is achieved when gendered parliamentary shares equal gendered population shares, we need to introduce the variables POP_f, POP_m, the female and male shares in the population. With these variables, the new GRS will be defined as

$$GRS_w = \left(\frac{SE_f}{SE_m}\right)^{w_1} \left(\frac{PR_f / POP_f}{PR_m / POP_m}\right)^{w_2} \left(\frac{LFPR_f}{LFPR_m}\right)^{w_3} \left(\frac{LEI_f}{LEI_m}\right)^{w_4} \qquad [4]$$

where the powers $w=(w_1,w_2,w_3,w_4)$ introduced in the formula reflect the weights that are attached to the different dimensions. WD is defined with the same variables and weights but with the functional form shown in equation [3]. Because of the variables included in equation [4], the new GRS index can be seen as a mixture of the GGM and GGM3 measures introduced in Klasen and Schüler (2011).

The choice of weights w_i for a composite index like GRS or WD is a particularly controversial and sensitive issue with important normative implications that has no perfect solution. Traditionally, UNDP indices assign the same weight to each dimension, implicitly assuming that all dimensions are equally relevant. In this context, if one assigns the weight 1/3 to each of the three dimensions and takes into account the fact that the 'Empowerment' dimension has two equally weighted subindicators (SE and PR), the corresponding weights for equation [4] would be $w_1=1/6, w_2=1/6, w_3=1/3, w_4=1/3$.

Given the important ethical implications that the choice of a particular weighting scheme entails, it is important to perform some sensitivity test in order to ascertain the robustness of the ranking one is working with. While the equal weighting approach might be attractive at first sight on grounds of simplicity, it might not be very appropriate in the present context where some variables exhibit much larger variability than others. To illustrate: the observed variability (in terms of standard deviations) of the gap in parliamentary representation is seventeen times as large as the variability of the gap in life expectancy. If equal weights are assigned to all dimensions, the values of the composite index are largely driven by the values of the dimensions with largest variability [[[Endnote#15]]]. In order to reduce the extent of this problem, we will choose weights whose magnitudes are inversely proportional to the standard deviation of the corresponding variable. This simple procedure – which has already been used in the construction of the World Economic Forum's Global Gender Gap index (see Hausmann et al. 2007) – yields the following list of weights that will be used in this paper: $w_1^* = 0.12, w_2^* = 0.04, w_3^* = 0.13, w_4^* = 0.71.$ The rationale for choosing these weights is that it is worse to have a gap of a certain size in a variable with small variability (i.e.: life expectancy) than having the same gap in a variable with large variability (i.e.: parliamentary representation).

Despite the fact that the weights w_i and w_i^* are so disparate, the corresponding rankings derived from them are quite similar [[[Endnote#16]]]. Given the fact that the country rankings obtained from both weighting schemes is not very different, in the next section we will present our empirical results for GRS and WD based on the values of w_i^* alone. Among the two, this set of weights seems more attractive on normative grounds.

4.1. Empirical results

Because of the different theoretical foundations upon which the GII and the new indices are based, the corresponding values will differ too. To compare the behavior of GII, $\overline{GII}$, GRS and WD we present the corresponding two-way scatterplots in Figure 5. The similarity between GRS and WD values is remarkable. Since the values of GRS and WD only differ in case women outperform men in some well-being dimensions, this illustrates the underprivileged situation of women vis-à-vis men in most dimensions incorporated into the indices. For the sake of completeness, Figure 5 also compares the distribution of the previous indices with that of the capped GII. It is interesting to observe the strong negative relationship between the $\overline{GII}$ on the one hand and GRS and WD on the other. These two groups of indices are conveying relatively similar ordinal information, but the dispersion is larger for those countries with higher gender inequality levels. Given the fact that $\overline{GII}$ is defined using SE, PR, LFPR only while GRS and WD further incorporate the life expectancy component, this strong relationship suggests at least two things: (i) the contribution of the life expectancy component to the aggregate GRS and WD values might not be very large – an issue that will be further investigated below –, and (ii) when the absolute/women-specific components are dropped from the GII, measuring gender inequality per se or measuring welfare losses due to gender inequality is basically the same exercise [[[Endnote#17]]].

Figure 5 also shows the negative relationship between GII and WD: this is expected because, loosely speaking, GII increases with larger gender gaps while WD decreases. The dispersion observed in their joint distribution is considerable, thus illustrating the different ways in which both indices operate. While Scandinavian and Middle East countries tend to occupy the normatively desirable and undesirable positions of both GII and WD rankings respectively, there are huge variations in between. In this respect, it is particularly illuminating to explore in more detail the extent to which countries change their ranking position when shifting from the GII to the WD values. For that purpose, Figure 6 plots the logged GDP per capita vs the change in ranking positions that is observed when shifting from GII to WD values. A clear negative relationship is observed, therefore suggesting that high-income countries tend to occupy lower (i.e: "worse") positions with WD and that low-income countries tend to occupy lower positions with GII. To illustrate: low-income countries like Rwanda, Malawi, Kyrgyzstan or Liberia are among the greatest beneficiaries when shifting from the GII to the WD ranking. On the contrary, high-income countries like Bahrain, Japan, Slovenia, Libyan Arab Jamahiriya, Greece, Malta and Kuwait are greatly affected by that shift – loosing more than 40 positions. These results empirically support the claim that GII penalizes low-income countries for poor performances in reproductive health indicators that are not entirely explained by gender-discriminative practices.

[[[Figure 5]]]

[[[Figure 6]]]

For illustrative purposes, we explore the values of WD and the percent contribution of its subcomponents to the aggregate values of the index using data from the 2010 HDR: they are shown in Figure 7. The values of WD range from 0.98 to 0.56, the five top performers and their WD values are Sweden (0.98), Finland (0.977), Norway (0.96), Denmark (0.95) and The Netherlands (0.949), and the five worst performers are Mali (0.67), Niger (0.66), Saudi Arabia (0.62), Yemen (0.58) and Afghanistan (0.56). Interestingly, Rwanda occupies the sixth position in the WD ranking, a somewhat surprising result when compared to other conceptually related rankings [[[Endnote#18]]]. This is basically due to the fact that: (i) Rwanda is one of the three countries in the world where women's labor force participation rate is officially higher than men's, and (ii) Rwanda is the only country in the world where the share of parliamentary seats occupied by women is officially larger than that of men. These exceptional figures are be partly explained by the Rwandan Genocide that took place in 1994 (when the killing ended there were twice as many women as men in Rwanda, and while the gap has since narrowed, more than a third of households are still headed by women; moreover, women make up 55% of the workforce and own about 40% of businesses). In contrast, for almost all other world countries with available data the gender gap in economic participation and parliamentary representation is quite high. While GII advocates might reasonably argue that Rwanda should not merit such a privileged position because of women's poor health status, it is worth pointing out that men's health status in that country is extremely poor as well (Rwanda occupies the 10th worst position in the world in terms of male life expectancy), so the corresponding health gender gap as measured with gender-specific life expectancies is not that large after all. The examples of the Scandinavian countries and Rwanda clearly illustrate that gender equality can be achieved by an equally good or an equally bad performance of women and men in the different dimensions one is taking into account – albeit the case of Rwanda is certainly exceptional. Since the values of gender equality indices alone are not informative on the situation of women and men in absolute terms, it is widely acknowledged that such information should be complemented with many other contextual and qualitative measures as well (Dijkstra and Hanmer 2000, Beneria and Permanyer 2010).

As illustrated in Figure 7, the percent contributions of the economic participation and parliamentary representation gaps are relatively large, particularly for those countries with higher gender equality levels (i.e.: larger WD values). At lower gender equality levels, the health and education gaps tend to have a greater contribution to WD values. Aggregating over all world countries with available data, we find that the average contribution of the different components are: 47% for PR, 28% for LFPR, 13% for SE and 12% for LE. In other words, the gender inequality levels measured by WD are largely driven by the parliamentary representation component – a variable that has been criticized elsewhere for excluding political participation at the community and local levels. At the other extreme, the education and health components play a less influential role in the aggregate values of the index. This is basically due to the fact that the gaps in PR and LFPR not only have larger variability (an issue that has been controlled for with the choice of the w_i^*) but also have larger average size.

[[[Figure 7]]]

In the previous section we explored the relationship between GDP per capita and GII, and found it to be fairly strong (*r*=–0.87). In other words, the values of the GII do not offer much differentiated information with respect to the one conveyed by the GDP per capita, partly because the two women-specific subindicators are also strongly related to the income component (see Figures 1 and 2). Figure 8 shows the scatterplot between the log of GDP per capita values against those of WD.

Interestingly, the dispersion in this case is larger – particularly at lower levels of per capita income –, there are many outliers and the corresponding correlation coefficient is lower in absolute terms: *r*=0.62. That is: WD is more successful in the task of capturing new information that is not encapsulated in the GDP. This is illustrated by comparing the performances of Saudi Arabia and Rwanda: the former country scores one of the lowest values on WD but it has a very large GDP per capita while the latter scores a high value on WD but has an extremely low GDP per capita.

[[[Figure 8]]]

5. Summary and concluding remarks

This paper provides a critical assessment of the new Gender Inequality Index presented in the 2010 Human Development Report. The new index is an interesting contribution to the debate on gender inequality measurement. On the one hand it incorporates important reproductive health variables that were not used in previous UNDP composite indices, and on the other hand it proposes a new methodology to aggregate multidimensional information into a single dimensional index. The inclusion of reproductive health variables seems promising at first sight as it brings to the fore an essential component of women's well-being that, among other things, has been included as one of UNDP's Millennium Development Goals (MDGs). However, their inclusion in the index creates more problems than it solves. Overall we have identified two major flaws in the construction of the index for which some remedy might be sought. To start with, the functional form of the index is neither particularly intuitive nor user-friendly. While the choice of that functional form is in line with the methodology used in the construction of the new HDR measures (i.e: the new HDI and IHDI), the conceptual foundations of the index are ill-specified: ¿what is the 'maximum' or 'potential' welfare level that should be achieved in case women and men fared equally well in all dimensions? ¿What is the underlying welfare function? This has not been specified anywhere. Moreover, the GII has been unnecessarily complicated in order to satisfy certain normative properties that are otherwise satisfied by much simpler indices as well. In addition, the GII incorporates both women-specific indicators (i.e.: absolute) and "woman vs men" indicators (i.e.: relative) into a single formula. This creates a host of important problems – both conceptual and methodological. Conceptually, the mixture of absolute and relative indicators obscures even more the interpretation of an already complicated index. This choice produces an index that, among other things: (i) penalizes low-income countries for poor performances in reproductive health indicators that are not entirely explained by the gender-related norms or discriminative practices against women that the GII purports to measure, (ii) does not reach the expected or normatively desirable value whenever women and men fare equally in all indicators, (iii) allows deteriorations in women's education and economic participation to be compensated by equivalent deteriorations in men's corresponding dimensions but somewhat arbitrarily does not allow for any such compensation when a deterioration in women's reproductive health conditions occurs, and (iv) completely disregards men's average health statuses – which are also essential pieces of information to be incorporated in a comprehensive assessment of gender inequality levels. Therefore, the values of the new GII index should be heavily qualified.

Whenever possible we have offered some alternatives to the identified shortcomings. We suggest substituting the complicated GII methodology by a much simpler index – WD – that has been recently proposed in Beneria and Permanyer (2010) and Klasen and Schüler (2011). WD is just an average of the gender gaps in which men outperform women (for practical purposes this is tantamount to virtually all gender gaps). Moreover, this index only uses relative indicators that compare women and men achievement levels, thus avoiding the many problems assailing the GII. WD incorporates gender-specific life expectancies at birth – an indicator that has been widely used in previous global gender inequality assessments – as an imperfect substitute of the women-specific components of the GII (Maternal Mortality Ratio and Adolescent Fertility Rate). It turns out that WD is moderately and positively correlated with the GDP per capita – as opposed to what happens with the GII, which is biased in favor of high-income countries –, thus illustrating its success in the task of capturing new information that is not encapsulated in the GDP. In contrast with the GII, WD has the further advantage of being decomposable by subcomponents, thus facilitating the understanding of the internal structure of the index and allowing to easily compute the percent contribution of each individual subcomponent to the aggregate value of the index. Using data from the 2010 HDR, we find that the values of WD are largely driven by the gender gap in parliamentary representation – an indicator that has been criticized elsewhere for excluding political participation at the community and local levels.

According to our results from the empirical section, it turns out that the ordinal information provided by WD and a "capped version" of the GII where the women-specific indicators have been removed from the index is very similar, that is: both indices rank world countries pretty much in the same way. This is a remarkable result suggesting that when the absolute/women-specific components are dropped from the GII, measuring gender inequality *per se* or measuring welfare losses due to gender inequality is essentially the same exercise. In a way, this could be used as an argument to reinforce the idea that, after all, it might not be necessary to introduce the sophisticated and somewhat confusing GII methodology since a much simpler and intuitively appealing index like WD would lead to analogous results.

One of the major weaknesses of the WD index presented in this paper is the lack of information concerning reproductive health conditions for both women and men. Given the fact that reproductive health issues affect the lives of women and men in completely different ways, it is a conceptual and methodological challenge to construct the corresponding internationally comparable gender-specific variables that could be plugged in gender inequality indices as the ones discussed in this paper. Despite this difficulty, we contend that the quest for such components is an important topic for the forthcoming research agenda that is needed to design more meaningful global gender inequality indices.

REFERENCES

Alkire, Sabina and Solava Ibrahim. 2007. “Agency and Empowerment: A Proposal for Internationally Comparable Indicators.” *Oxford Development Studies* 35(4):379–403.

Bardhan, Kalpana and Stephan Klasen. 1999. “UNDP’s Gender-Related Indices: A Critical Review.” *World Development* 27(6):985–1010.

––––––. 2000. “On UNDPs revision to the Gender-related Development Index.” *Journal of Human Development* 1(2):191–195.

Beneria, Lourdes. 2008. “The Crisis of Care, International Migration, and Public Policy.” *Feminist Economics* 14(3):1--21.

Beneria, Lourdes and Iñaki Permanyer. 2010. “The Measurement of Socio-economic Gender Inequality Revisited.” *Development and Change* 41:375–399.

Bericat, Eduardo. 2011. “The European gender equality index: Conceptual and analytical issues.” *Social Indicators Research*, doi: 10.1007/s11205-011-9872-z.

Branisa, Boris, Stephan Klasen and Maria Ziegler. 2009. “Background Paper: The Construction of the Social Institutions and Gender Index (SIGI).” University of Goettingen, Germany, March 6, 2009. Available at: http://www.oecd.org/dataoecd/49/19/42295804.pdf

Charles, Maria and David Gursky. 1995. “Models for describing the underlying structure of sex segregation.” American Journal of Sociology 100:931–971.

––––––. 1998. “The past, present and future of sex segregation methodology.” *Demography* 35(4):497–504.

D’Agostino, Marcello and Valentino Dardanoni. 2009. “The measurement of rank mobility.” *Journal of Economic Theory* 144(4):1783–1803.

Dijkstra, A. Geske and Lucia C. Hanmer. 2000. “Measuring Socio-economic Gender Inequality: Towards an Alternative to the UNDP Gender-Related Development Index.” *Feminist Economics* 6(2):41–75.

Dijkstra, A. Geske. 2002. “Revisiting UNDP’s GDI and GEM: Towards an Alternative.” *Social Indicators Research* 57:301–38.

Dijkstra, A. Geske. 2006. “Towards a Fresh Start in Measuring Gender Inequality: A Contribution to the Debate.” *Journal of Human Development* 7(2):275–83.

Duclos, Jean-Yves, David Sahn and Stephen Younger. 2006. “Robust Multidimensional Poverty Comparisons.” *Economic Journal* 116(514):943–968.

Folbre, Nancy. 2006. “Measuring Care: Gender, Empowerment, and the Care Economy.” *Journal of Human Development* 7(2):183–99.

Foster, James and Amartya Sen. 1997. On Economic Inequality, Extended Edition. Oxford: Clarendon Press.

Forsythe, Nancy, Roberto P. Korzeniewick and Valerie Durrant. 2000. "Gender Inequalities and Economic Growth: A Longitudinal Evaluation." *Economic Development and Cultural Change* 48(3):573–617.

Gaye, Amie, Jeni Klugman, Milorad Kovacevic, Sarah Twigg and Eduardo Zambrano. 2010. "Measuring key disparities in human development: the Gender Inequality Index." Human Development Research Paper 2010/46.

Hausmann, Ricardo, Laura D. Tyson and Saadia Zahidi. 2007. *The Global Gender Gap Report*. Switzerland: World Economic Forum.

Hawken, Angela and Gerardo Munck. 2009. "Cross-National Indices with Gender-Differentiated Data: What Do They Measure? How Valid Are They?" Technical background paper for the forthcoming UNDP Asia Pacific Human Development Report on gender. UNDP.

Jütting, Johannes P., Christian Morrisson, Jeff Dayton-Johnson and Denis Drechsler. 2008. "Measuring Gender (In)Equality: The OECD Gender, Institutions and Development Data Base." *Journal of Human Development* 9(1):65–86.

Klasen, Stephan. 2006. "UNDP's Gender-related Measures: Some Conceptual Problems and Possible Solutions." *Journal of Human Development* 7(2):243–74.

Klasen, Stephan and Dana Schüler. 2011. "Reforming the Gender-Related Index and the Gender Empowerment Measure: Implementing Some Specific Proposals." *Feminist Economics* 17(1):1–30.

McGillivray, Mark. 1991. "The Human Development Index: yet another redundant composite development Indicator?" *World Development* 19:1461–1468.

Morrisson, Christian and Johannes Jütting. 2005. "Women's discrimination in developing countries: a new data set for better policies." *World Development* 33(7):1065–1081.

OECD. 2008. *Handbook on constructing composite indicators: Methodology and user guide*. OECD, Paris.

Permanyer, Iñaki. 2008. "On the Measurement of Gender Equality and Gender-related Development Levels." *Journal of Human Development* 9(1):87–108.

––––––. 2010. "The Measurement of multidimensional gender inequality: continuing the debate." *Social Indicators Research* 95(2):181–198.

Plantenga, Janneke, Hugo Figueiredo, Chantal Remery and Mark Smith. 2003. "Towards an EU Gender Equality Index." a feasibility study commissioned by and presented to the European Commission, Utrecht School of Economics, Utrecht.

Ravaillon, Martin. 2010. "Mashup Indices of Development." World Bank Policy Research Working Paper No. 5432.

Schüler, Dana. 2006. "The Uses and Misuses of the Gender-related Development Index and Gender Empowerment Measure: A Review of the Literature." *Journal of Human Development* 7(2):161–81.

Seth, Suman. 2009a. "A class of association sensitive multidimensional welfare indices." OPHI Working Paper No. 27.

––––––. 2009b. "Inequality, Interactions and Human Development." *Journal of Human Development and Capabilities* 10:375–396.

Social Watch. 2004. Social Watch Annual Report 2004 (Montevideo: Instituto del Tercer Mundo).

United Nations Development Program (UNDP). 1995, 2010. Human Development Report. New York: Oxford University Press.

United Nations Economic Commission for Africa (UNECA). 2004. The African Gender and Development Index (Addis Ababa, Ethiopia: Economic Commission for Africa).

Unterhalter, Elaine. 2006. Measuring Gender Inequality in Education in South Asia (Kathmandu: The United Nations Children's Fund [UNICEF], Regional Office for South Asia; and United Nations Girls' Education Initiative [UNGEI]).

White, Howard. 1997. "Patterns of Gender Discrimination: An Examination of the UNDP's Gender Development Index." Mimeo, The Hague: Institute of Social Studies.

Endnotes

Endnote#1: See, for instance, the Gender Equality Index (Howard White 1997), the Gender Inequality Index (Nancy Forsythe, Roberto Korzeniewick and Valerie Durrant 2000), the Relative Status of Women index (Dijkstra and Hanmer 2000), the Standardized Index of Gender Equality (Dijkstra 2002), the European Union Gender Equality Index (Janneke Plantenga, Hugo Figueiredo, Chantal Remery and Mark Smith 2003), the African Gender and Development Index (UNECA 2004), the Gender Equity Index (Social Watch 2004), the Social Institutions and Gender Index (Christian Morrisson and Johannes Jütting 2005, Johannes Jütting, Christian Morrisson, Jeff Dayton-Johnson and Denis Drechsler 2008, Boris Branisa, Stephan Klasen and Maria Ziegler 2009), the Gender Equality in Education Index (Elaine Unterhalter 2006), the Global Gender Gap Index (Ricardo Hausmann, Laura Tyson and Saadia Zahidi 2007), the Multidimensional Gender Equality Index (Iñaki Permanyer 2008), the Gender Relative Status and Women Disadvantage indices (Lourdes Beneria and Iñaki Permanyer 2010), the European Gender Equality Index (Eduardo Bericat 2011) and the Gender Gap Measure (Stephan Klasen and Dana Schüler 2011).

Endnote#2: Population Action International (PAI) has recently proposed the construction of the Reproductive Risk Index (RRI), which is the weighted average of nine reproductive health indicators. However, that index has not been designed to capture gender inequality explicitly (even if certain gender roles and norms clearly influence its values), so it will not be discussed in this paper.

Endnote#3: Even if the absence of maternal mortality and adolescent fertility means that MMR=AFR=0, the choice of that number as a reference benchmark entails several technical problems (like the obtention of formulas with zero denominators), so it has not been incorporated in the definition of the GII. It must be pointed out that in the 2011 Human Development Report, the reference benchmark value for MMR was changed from 1 to 10. However, this does not change any of the issues raised in this paper.

Endnote#4: The numerator in equation [1] is the hyperbolic mean of the average achievement levels of women and men, and the denominator is the geometric mean of the average achievement levels in the different dimensions.

Endnote#5: In a similar context, McGillivray (1991) criticized the – by then – recently released HDI for its high correlation with the GDP per capita. The paper basically concluded that the information provided by the HDI was essentially redundant when compared to the income variable.

Endnote#6: This condition simply states that in the original achievement matrix H, the achievement vectors of women and men do not vector-dominate each other, that is: the achievement levels of one gender are not systematically higher than the achievement levels of the other for all indicators.

Endnote#7: Jean-Yves Duclos, David Sahn and Stephen Younger (2006) assume that all couples of attributes are substitutes in the context of multidimensional poverty orderings. However, that assumption needs to be imposed because of methodological constraints. In that paper, the authors point out to some cases in which the attributes might well be complementary: "*For instance, for a poverty analysis in the dimensions of education and nutritional status of children, there are production complementarities because better-nourished children learn better. If this complementarity is strong enough, it may overcome the usual ethical judgment that favors the multiply-deprived, so that overall poverty would decline by more if we were to transfer education from the poorly nourished to the better nourished, despite the fact that it increases the correlation of the two measures of wellbeing*" (Duclos, Sahn and Younger 2006, p. 950).

Endnote#8: These arguments are echoing the ideas presented in Dijkstra (2006). In that paper, the author criticizes the use of HDI-GDI or 1-(GDI/HDI) as measures of gender inequality, stating that they do "*not measure gender inequality per se, but instead the reduction in welfare due to gender inequalities*" (Dijkstra 2006:280). Exactly the same happens here, the only difference being that the welfare function used in the HDI and GDI allowed for perfect substitutability between dimensions while the underlying welfare function for the GII is association sensitive and does not allow for perfect substitutability.

Endnote#9: In Klasen and Schüler (2011), the functional forms GRS and WD are labeled as "Gender Gap Measure" (GGM) and "capped GGM" respectively.

Endnote#10: GRS would not be able to distinguish between a country where women and men fare equally well in each dimension and another one with some large gender gaps favoring women and some equally large gender gaps favoring men. This problem has also been mentioned in Klasen (2006) and further explored in Beneria and Permanyer (2010).

Endnote#11: As already argued in Beneria and Permanyer (2010:382) and Klasen and Schüler (2011:11) there are good reasons to focus on these gaps only.

Endnote#12: Given the fact that $WD = \sum_{i \in I_M} w_i \ln(x_i / y_i)$, the contribution of component '*i*' to the aggregate value of WD can be measured as $w_i \ln(x_i / y_i) / \ln(WD)$. Recall that such decomposition would not make much sense in the case of GRS because the opposing directions of the gender gaps and the corresponding contributions cancel out each other.

Endnote#13: Gender-specific HIV youth prevalence rates, for instance, is a candidate with a fairly good geographical coverage (data is available for about 140 countries) but more research is needed to ascertain its appropriateness in a global gender inequality index. On the other hand, adult, infant and under-five gender-specific mortality rates are typically higher for males, but, as is well known, this cannot be attributed to gender-discriminative practices against them.

Endnote#14: According to the definitions used in the GDI, these normalized indices are defined as: LEI_f=(LE_f −27.5)/60; LEI_m =(LE_m −22.5)/60.

Endnote#15: The greater influence of one component over the others is reminiscent of the results found in Bardhan and Klasen (1999). In that paper, the authors found that gender inequality penalty in the GDI were overwhelmingly accounted for by penalties for earned-income gaps. In order to correct this problem, the authors suggested alternative remedies.

Endnote#16: In order to measure the similarity/dissimilarity between rankings we have used a rankings distance function borrowed from Marcello D'Agostino and Valentino Dardanoni (2009). The function is defined as $d(R,R') = \left(\sum_{i=1}^{n} (R_i - R_i')^2\right) \big/ \left((n^3 - n)/3\right)$, where R_i, R_i' are the ranking positions of country '*i*' in rankings *R* and *R'* respectively and *n* is the number of countries being ranked (see D'Agostino and Dardanoni 2009 for details). That function takes a minimal value of 0 whenever the two rankings we compare are the same and a maximal value of 1 whenever we compare a given ranking with its opposite ranking. It turns out that the distance between rankings obtained from GRS_w and GRS_{w^*} equals 0.06, and the distance between rankings obtained from WD_w and WD_{w^*} equals 0.05. Therefore, the choice between *w* and *w** does not entail major differences in the corresponding country rankings.

Endnote#17: This is confirmed by the fact that when the original $\overline{GII}$ weights are used in the construction of GRS and WD (i.e.: w_1=1/6,w_2=1/6,w_3=1/3,w_4=1/3), the country rankings that are obtained from the values of $\overline{GII}$, GRS and WD are almost the same.

Endnote#18: For instance, Rwanda ranks in the 83rd position according to the GII values (this is basically due to the appallingly high MMR prevalent in that country). According the 2009 GDI values, Rwanda ranks in the 139th position (in turn, this is due to its extremely low GDP per capita).