the review of income and wealth

Review of Income and Wealth
Series 63, Number 4, December 2017
DOI: 10.1111/roiw.12286

MEASURING INEQUALITY BY ASSET INDICES: A GENERAL APPROACH WITH APPLICATION TO SOUTH AFRICA

by Martin Wittenberg and Murray Leibbrandt*

University of Cape Town

Asset indices are widely used, particularly in the analysis of Demographic and Health Surveys, where they have been routinely constructed as "wealth indices." Such indices have been externally validated in a number of contexts. Nevertheless, we show that they often fail an internal validity test, that is, ranking individuals with "rural" assets below individuals with no assets at all. We consider from first principles what sort of indexes might make sense, given the predominantly dummy variable nature of asset schedules. We show that there is, in fact, a way to construct an asset index which does not violate some basic principles and which also has the virtue that it can be used to construct "asset inequality" measures. However, there is a need to pay careful attention to the components of the index. We show this with South African data.

JEL Codes: D63, I32

Keywords: asset indicators, inequality measurement

1. Introduction

Asset indices have become widely used since Filmer and Pritchett (2001) described a simple way to calculate them. Their use really took off once the Demographic and Health Surveys incorporated the calculation of a "wealth index" with the release of each dataset (Rutstein and Johnson, 2004). A Google Scholar search (April 18, 2014) came up with 13,900 "hits" on "DHS wealth index," 2,434 citations of the article by Filmer and Pritchett (2001), 591 citations of the paper by Rutstein and Johnson (2004) documenting the creation of the DHS index. The main use of the indices in this vast literature is in creating wealth rankings, separating the "rich" from the "poor" as ingredients for more substantive analyses.

Several articles, including the original piece by Filmer and Pritchett (2001), have tried to validate these indices against *external* criteria, for example, incomes or expenditures. A recent review (Filmer and Scott, 2012) concludes that "the use of an asset index can clearly provide useful guidance to the order of magnitude of

Note: We have benefited from useful comments from David Lam and seminar participants at the University of Michigan, as well as from audience members at the UNU–WIDER conference on Inequality—Measurement, trends, impacts, and policies, Helsinki, September 2014. We would also like to thank Conchita D'Ambrosio and two anonymous referees for feedback which improved the paper markedly. Of course, we remain responsible for all remaining errors.

*Correspondence to: Martin Wittenberg, 3.48.2, School of Economics Building, Middle Campus, University of Cape Town, Lover's Walk, Rondebosch 7701, South Africa (Martin.Wittenberg@uct.ac.za).

rich–poor differentials" (p. 389), although the asset indices measure a different concept than per capita consumption. Indeed, the paper devotes attention to the question of under which circumstances the two measures will provide the most similar rankings, arguing that this will occur when per capita expenditures are well explained by observed household and community characteristics and when "public goods" are more important in household expenditures than "private ones" such as food. In other work, we have ourselves argued that asset indices do a good job of proxying for income differences (Wittenberg, 2009, 2011).

None of this literature has examined whether the asset indices calculated in the traditional way make sense *internally*, that is, according to a number of simple criteria such as that individuals that have more (of anything) should be ranked higher than individuals that have less. In particular, little attention has been paid to the problems created by the predominantly dummy variable nature of asset schedules. We show that this is not just a theoretical issue but that, in a number of cases, DHS wealth indices exhibit anomalous rankings.

One additional issue that has been lamented in some contexts is that the way in which these indices are typically calculated precludes the use of traditional inequality measures. One might think that if it makes sense to talk about inequality in incomes or wealth, it would certainly make sense to think about inequality in asset holdings (McKenzie, 2005; Bhorat and van der Westhuizen, 2013). Nevertheless, the manipulation of traditional indices is not a viable strategy (Wittenberg, 2013): a different approach is needed. As we show below, it is when we consider the particular problems of calculating inequality measures with dummy variables that many problems with the creation of asset indices crystallize. However, we show that these problems are not insuperable. Indeed, an approach due to Banerjee (2010) for dealing with multidimensional inequality can be used to create such asset indices, as we will show below.

We show that this approach is easy to implement and we apply it to South African data. This provides a new perspective on the evolution of South African inequality which is somewhat at odds with the literature measuring inequality with money-metric approaches. We think it is likely that the asset approach reveals genuine improvements over time, although the reduction in inequality is unlikely to be as dramatic as the Gini coefficients calculated on the asset indices suggest. We think that more detailed asset inventories would moderate some of the conclusions. Indeed, one of our key points is that asset indices need to be approached with some caution—churning out "wealth indices" in semi-automated ways, without considering in detail what the individual scores suggest, is likely to be problematic.

The plan of the paper is as follows. In Section 2, we provide a very brief overview of the theoretical literature dealing with asset indices. We follow on by enunciating several principles for the creation of such indices in Section 3. We refer to these as "principles" since our approach is not fully axiomatic. Our approach is more heuristic—investigating what happens when we apply different approaches to simple data and considering whether the answers make sense. We do this in Sections 4–6, where we consider first the case of a single binary variable and then we progressively consider more complicated cases. In each case, we consider both the index itself and what it might mean to estimate inequality with it. Having set

out what we consider to be a defensible approach, we turn to applying it to DHS data in Section 7. Finally, we consider what assets may tell us about the evolution of inequality in South Africa from 1993 to 2008.

The chief contributions of our paper to the literature are both negative and positive. On the negative side, we show that there are anomalies embedded deep in the predominant approaches for creating asset indices, which users should be aware of before blithely adopting them. On the positive side, this paper: (1) describes how to construct an asset index that is internally coherent; (2) shows that inequality measures on this index are well defined and have reasonable interpretations; (3) provides some perspective on the "art" of index construction; and (4) provides a fresh perspective on South African inequality.

2. Literature Review

McKenzie (2005, p. 232) suggests that the idea of using the first principal component of a set of asset variables as an index for "wealth" has been around in the social science literature for a long time. Its use, however, has become common only after the publication of Filmer and Pritchett (2001) and the subsequent adoption of the method in the release of the DHS "wealth indices" (Rutstein and Johnson, 2004). The basic idea of principal components is to find the linear combination of the asset variables that maximizes the variance of this combination. More formally, if we have k random variables $a_1, \ldots, a_k$, each standardized to be of mean zero and variance one, the objective is to rewrite these as

$$\begin{aligned} a_1 &= v_{11}A_1 + v_{12}A_2 + \ldots + v_{1k}A_k, \\ a_2 &= v_{21}A_1 + v_{22}A_2 + \ldots + v_{2k}A_k, \\ &\vdots \\ a_k &= v_{k1}A_1 + v_{k2}A_2 + \ldots + v_{2k}A_k, \end{aligned} \tag{1}$$

where the A_i are unobserved components, created so as to be orthogonal to each other. Writing this in vector notation as

$$\mathbf{a} = \mathbf{V}\mathbf{A},$$

it follows that the covariance matrix (here equal to the correlation matrix $\mathbf{R}$) is given by

$$\begin{aligned} E(\mathbf{a}\mathbf{a}') &= E(\mathbf{V}\mathbf{A}\mathbf{A}'\mathbf{V}'), \\ \mathbf{R} &= \mathbf{V}\boldsymbol{\Phi}\mathbf{V}', \end{aligned}$$

where $\boldsymbol{\Phi} = E(\mathbf{A}\mathbf{A}')$. Note that $\boldsymbol{\Phi}$ is diagonal since the unobserved components are assumed to be orthogonal to each other. We need to impose some normalization in order to get a determinate solution. Let $\boldsymbol{\Phi}$ be the matrix of eigenvalues and $\mathbf{V}$ the orthonormal matrix of eigenvectors, and assume that $\mathbf{V}$ is ordered so that the

eigenvector associated with the largest eigenvalue is listed first. We can then solve for **A**, to obtain

$$\mathbf{A}=\mathbf{V}'\mathbf{a}.$$

In particular,

$$A_1=v_{11}a_1+v_{21}a_2+\ldots+v_{k1}a_k. \tag{2}$$

We will refer to this as the PCA index. By assumption, $\text{var}(A_1)=\lambda_1$, the first eigenvalue, and we can show that no other linear combination of the a_i variables will achieve a greater variance (Wittenberg, 2009, pp. 5–6).

If the asset variables a_i do not have unit variance and zero mean, they are first standardized, so that the equation for the first principal component will be given by

$$\begin{aligned} A_1 &= v_{11}\left(\frac{a_1-\bar{a}_1}{s_1}\right)+v_{21}\left(\frac{a_2-\bar{a}_2}{s_2}\right)+\ldots+v_{k1}\left(\frac{a_k-\bar{a}_k}{s_k}\right) \\ &= \frac{v_{11}}{s_1}a_1+\frac{v_{21}}{s_2}a_2+\ldots+\frac{v_{k1}}{s_k}a_k-c, \end{aligned} \tag{3}$$

where the coefficients v_{i1} are the elements of the eigenvector $\mathbf{v}_1$ associated with the largest eigenvalue λ_1 of the correlation matrix **R** of the a_i variables. The constant c is the weighted sum of the means, which ensures that A_1 has a zero mean.

The use of the first principal component was defended by Filmer and Pritchett (2001) on a "latent variable" interpretation of equations (1): A_1 is whatever explains most of what is common to $a_1, \alpha_2, \ldots, a_k$ and it makes most sense to think of this as "wealth." Other authors have taken this formulation more seriously and have suggested that other procedures, such as factor analysis, be used to retrieve the common latent variable (Sahn and Stifel, 2003).[1] Although the procedure produces a different index than the PCA one, in practice the indices calculated by both approaches are highly correlated, particularly since authors using this approach seem to restrict themselves to extracting only one factor and eschew the "orthogonal rotations" that produce arbitrarily many solutions.

Reviews of the procedure have focused on several issues. First, if the assets are measured mainly through categorical variables, then the index defined through equation (2) is intrinsically discrete. The more assets and the more integer-valued variables (e.g. number of rooms) that are included in the index, the smoother the resulting index will be and the better will be its potential to differentiate finer gradations of poverty (McKenzie, 2005). Second, if categorical variables with multiple categories are included (e.g. water access), then the resulting group of dummy variables will be internally correlated with each other in ways that will influence the construction of the index. The more categories, the more dummy variables and the more this group influences the overall index. As a result, some authors have used multiple correspondence analysis instead (Booysen *et al.*,

[1]For a more detailed discussion of the factor analysis approach, see Wittenberg (2009).

2008). Unfortunately, it cannot accommodate continuous variables. In practice, the PCA index is also highly correlated with the MCA index. An additional point is that some of the categories will inevitably feature as "bads" and so should definitely receive a negative weight (Sahn and Stifel, 2003). This is, however, different to the cases that we consider later, where "goods" get assigned negative scores.

A third issue which has received some attention is whether or not the index should include infrastructure variables (such as access to water and sanitation). Houweling *et al*. (2003) tested the PCA index rankings for sensitivity to the assets included. They were concerned about the fact that the infrastructure assets might have independent effects on the outcome of interest, in particular child mortality. They show that the rankings change somewhat as some of the "assets" are stripped out. Thus there are important judgments to be made in deciding which assets to include or exclude in an asset index.

Several authors have tried to validate asset indices against external benchmarks. We have already referred to the review article by Filmer and Scott (2012). They found that different techniques for constructing asset indices tended to get results that were highly correlated with each other, but in some cases differing from the rankings implied by per capita consumption. This is not thought to be a problem in principle, since it is possible that assets may be a more reliable indicator of long-run economic well-being. They may also be measured with less error (Filmer and Pritchett, 2001; Sahn and Stifel, 2003).

One noteworthy finding in Filmer and Scott (2012) is that urban–rural differences tend to be more marked when using asset indices than when using per capita expenditure. Consumption/expenditure is felt to be a better indication of longer-run money-metric well-being than income, and thus the high aggregate correlation between asset indices and consumption is not that surprising. But this makes the sharp urban–rural divergence between these two measures noteworthy. It could be due to the fact that wealth is more concentrated than consumption, but perhaps it is also due to the fact that many of the household durable goods that make up asset schedules (e.g. televisions and refrigerators) require electricity, which tends to be more accessible in urban areas. Indeed, we have argued that both principal components and factor analysis will tend to extract an index which is a hybrid of "wealth" and "urbanness" (Wittenberg, 2009). We will show below that the asset index values rural assets (in particular, livestock) negatively, thus making rural asset holders look poorer than they should. We will suggest that the urban–rural differences are actually exaggerated by the indexes.

3. Principles for the Creation of Asset Indices

Intuitively, all the justifications for the creation of an asset index rely on the idea that higher asset holdings should convert into a higher index number and, conversely, a higher index number should imply greater wealth. This is a simple, yet obvious, internal consistency requirement. We shall refer to this as the *monotonicity* principle. In order to outline this more rigorously, we first define what we mean by an asset and an asset index.

We define assets as goods that provide (potentially) a stream of benefits. An asset variable A_j will be a random variable such that a_j is either the quantity or the value or the presence/absence of the asset. This excludes "bads." We also therefore do not allow a_j to be negative.

Definition 1. *Let* $(a_1, a_2, \ldots, a_k) \in \Re^k$ *be a vector of asset holdings. The function* $A : \Re^k \to \Re$ *defined for all possible asset holdings is called an asset index.*

Typically, we will restrict attention to linear asset indices, that is, indices that can be written in the form $A(a_1, a_2, \ldots, a_k) = v_1 a_1 + v_2 a_2 + \ldots + v_k a_k$.

Principle 2. *Let* $A(a_1, a_2, \ldots, a_k)$ *be an asset index. The asset index is monotonic if and only if*

$$(a_1, a_2, \ldots, a_k) \geq \left(a_1^*, a_2^*, \ldots, a_k^*\right) \Rightarrow A(a_1, a_2, \ldots, a_k) \geq A\left(a_1^*, a_2^*, \ldots, a_k^*\right).$$

Note that this is a fairly weak condition. It does, not, for instance, rule out "inferior" assets. For instance, if we had an asset schedule that listed different types of stoves—for example, electric, paraffin, coal, or gas—the corresponding "ownership" vectors might be recorded as $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$ and $(0, 0, 0, 1)$, respectively. Since none of these vectors is numerically bigger than the other, there is no restriction on how the asset index should rank them either. However, if these are not recorded as mutually exclusive categories, then an individual that owned *both* an electric stove and a gas stove should receive a higher asset index than one that owns only an electric stove.

The second principle that we require is that the index must be ratio-scale, that is, it must have an *absolute zero*. This is indispensable if we want to calculate inequality measures on the index, since it is not valid to calculate "shares" (required to construct, for instance, the Lorenz curve) if the variable is not ratio-scale. It implies in particular that the index must be able to recognize individuals or households that own nothing.

Principle 3. *Let* $A(a_1, a_2, \ldots, a_k)$ *be an asset index. The asset index has an absolute zero if and only if*

$$A(0, 0, \ldots, 0) = 0.$$

Obviously, this principle is violated by all of the current asset indices, except those that simply sum up the number of assets. Nevertheless, it is sensible to maintain that if the notion of "asset holdings" is to have any meaning, it is only in relation to individuals that do not have any. Even for purely ranking exercises, it is conceptually necessary that it makes sense to define the "have-nots" and that they should rank at the bottom.

Assuming that the previous two principles hold, it then makes sense to consider inequality measures on the space of asset index measures.

Principle 4. *We will say that the asset inequality measure I is robust if it can be applied to asset vectors of dummy variables as well as to continuous ones.*

Robustness is not a conceptual requirement, but it is desirable nonetheless, given that the asset information is typically dummy variable based. Theoretically, there is no reason why one should not construct different types of measures for different types of data. It is, however, much simpler if the approach can accommodate these differences. One big advantage of robust measures is that we know what the measures mean when the underlying data are of the continuous type that are treated in standard social welfare accounts. When these measures are applied to dummy variables, however, the interpretation becomes more complex. Robustness in this case means that the "standard" and "non-standard" treatments are part of the same continuum, so that if the measurement of the variable were to improve over time, we would only need to tweak our approach rather than switch completely. It is easier to see what this means by turning directly to the simplest case of all.

4. One Binary Variable

Consider first the case where we have precisely one binary variable, for example, we know whether or not the respondent owns a television set. Note that in this case the only possible "asset index" is the variable itself. Note also that we cannot analyse these data "from first principles" according to the typical axioms of inequality measurement, since these types of data will not support the "principle of transfers"—it is impossible to take away an asset from person j and give it to person i without them changing places in the distribution. Furthermore, such a "trade" (by the principle of anonymity) would leave the distribution precisely unchanged—and ratio-scale independence does not hold either, since rescaling of the variable does not provide a valid asset distribution.

4.1. *Standard Inequality Measures*

Many of the standard inequality measures (e.g. Atkinson indices) will not provide valid answers in the presence of zeroes. Nevertheless, some do, with the Gini coefficient the most common example. It is instructive to consider what the Gini of such a variable would measure. Assume that there are n_0 observations with zeroes and n_1 ones. Let the proportion of ones be p, that is, $p=\frac{n_1}{N}$, where $N=n_0+n_1$. The Gini coefficient[2] is simply $1-p$.

This is not an unattractive choice as a measure of inequality: if everyone has the asset, then the Gini is zero; as $p \to 0$—that is, the asset becomes concentrated in a smaller and smaller group—the index approaches one. It is obvious that given the paucity of information in the binary variable any "measure" of inequality must be, in some sense, a function of p.

There are some alternatives. For instance, the coefficient of variation applied to the binary variable would yield $\sqrt{\frac{1-p}{p}}$. This again yields a measure of zero when $p = 1$, but in this case the index of inequality approaches $+\infty$ as $p \to 0$.

[2]Wagstaff (2005) provides a discussion of "concentration indices" for the case where the dependent variable is binary. This value for the Gini coefficient is a special case of his more general result.

Obviously, both measures break down at $p = 0$. Indeed, in a world in which nobody has the "asset," it seems hard to define what inequality in the possession of that asset would mean. It is also worth noting that both measures give meaningful results only if the variable records the possession of a "good." If the variable measures a deprivation, it should be recoded first.

4.2. *The Cowell–Flachaire Measures*

An alternative to the cardinally based measures is the approach for ordinal variables proposed by Cowell and Flachaire (2012). These require us to measure the status of everyone in the distribution, where this is simply the count of everyone of equal rank or lower ("downward" measure) or, alternatively, everyone of equal rank or higher ("upward" measure). Both are expressed as proportions of the population. The vector of status measures $\mathbf{s}=(s_1, s_2, \ldots, s_N)$ is then used to calculate an inequality measure, relative to a "reference" status, which Cowell and Flachaire suggest should be set to 1. The inequality measures then become:

$$I_\alpha = \begin{cases} \dfrac{1}{\alpha(\alpha-1)}\left[\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} s_i^\alpha - 1\right] & \text{if } \alpha \neq 0, 1, \\ -\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \log s_i & \text{if } \alpha = 0. \end{cases}$$

A virtue of this set of measures is that it is invariant to the way in which the ordinal variable is "cardinalized," since the cardinalization will not affect the ranking of individuals in the distribution.

In the case of our binary variable, we obtain the following status values:

Value	s_i ("Downward")	s_i' ("Upward")
1	1	p
0	$1-p$	1

Consequently, the "downward" measure would be

$$I_\alpha = \begin{cases} \dfrac{(1-p)}{\alpha(\alpha-1)}[(1-p)^\alpha - 1] & \text{if } \alpha \neq 0, 1, \\ -(1-p)\log(1-p) & \text{if } \alpha = 0, \end{cases}$$

while the "upward" measure would be

$$I_\alpha = \begin{cases} \dfrac{p}{\alpha(\alpha-1)}[p^\alpha - 1] & \text{if } \alpha \neq 0, 1, \\ -p\log p & \text{if } \alpha = 0. \end{cases}$$

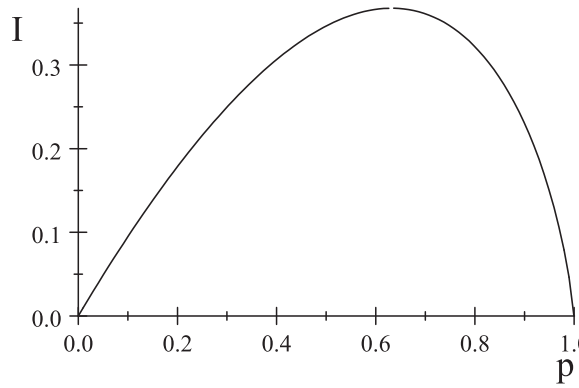

Figure 1. The Cowell–Flachaire I_0 index as a function of p

Low values of α emphasize inequality at the bottom of the distribution—the deprivation of those without the assets is felt more—while for α values close to one what happens at the top is more accentuated. Note that when $p = 0$ or $p = 1$, inequality is zero. Indeed, by considering the second derivative of I_a, it is clear that this measure of inequality has an inverse "U"-shaped curve, as shown in Figure 1 for the case $\alpha = 0$.

The variance of the distribution (which is also sometimes used as a measure of inequality) also exhibits this sort of pattern, with a low index of inequality near $p = 0$ and $p = 1$.

4.3. *The Meaning of Asset Inequality*

The difference in the behavior of the two groups of inequality "measures," namely a monotonic decrease in inequality as p goes from near zero to one versus inverse "U" shaped, raises fundamental questions about how we interpret the contrast between the "haves" and the "have-nots." In the Gini and coefficient of variation interpretation, that gulf is the central feature of the distribution—so if 99 percent of the population are lacking the asset but 1 percent have it, that is the most salient fact about the distribution. In the Cowell–Flachaire view, if most of the population shares the deprivation, then most outcomes are very similar to each other, that is, there is not a lot of inequality.

Which of these perspectives is right? Consider a "satisfaction with life" variable that has been measured on a Likert scale ranging from 1 (very dissatisfied) to 5 (very satisfied). Let 99 percent of the population record a "3" (i.e. neutral) but let 1 percent rate above that. This variable could be dichotomized as a 0/1 binary variable, with the "satisfied" responses scored as 1 while those below are recorded as zero. This distribution probably should not rate as very unequal, so in this context the Cowell–Flachaire measure seems more reasonable than the equivalent Gini. Note, however, that the Cowell–Flachaire measure is invariant to linear translation—that is, we would get the same measure whether 99 percent responded "3" and 1 percent "4" or whether 99 percent answered "4" and 1

percent "5," or even 99 percent "1" and 1 percent "5." Indeed, the reason why these all give the same distributional measure is that the conversion of the underlying phenomenon into a cardinal measure is arbitrary.

The central question is therefore to what extent the binary variable is an *arbitrary* coding of the underlying distribution. The crucial difference is not so much what the "1" codes for (since that could stand for almost any value), but whether the "0" can be thought of as absolute. Indeed, as we noted in the previous section, the Gini coefficient is sensible only if the variable is ratio-scale, that is, if the zero is absolute. The reason why the Gini scores inequality so highly when p is low is that the gulf between having *nothing* and having *something* is enormous. This is true, however, only if the "0" is really nothing and "1" signals the real possession of an asset (e.g. a car). Some of the variables typically used in the construction of asset indices need to be thought about very carefully in this context. For instance, a dummy variable for "tiled roof" obviously really measures the presence or absence of a "tiled roof." Nevertheless, the absence of a tiled roof does not imply the absence of all roofs; whether or not the gap between owning a thatched roof and a tiled roof is as vast as the gap between having nothing and having something is debatable.

Nonetheless, many of the assets do measure material gaps—ownership of a car or of a television are examples. Some infrastructure variables arguably also satisfy this criterion. The presence or absence of water in the house may be such a salient difference that the "0" really denotes a key absence. For variables such as these, the Gini measure seems closer to our intuition of how we would think about "asset inequality."

We take two points away from this discussion. First, one needs to think quite carefully about what variables one wants to include in one's measure of "asset inequality." If the variables in question are, at best, ordinal quality-of-life measures (e.g. "tiled roof"), then the appropriate "inequality measure" needs to be an ordinal one, like the Cowell–Flachaire approach. Second, if the binary variable really captures the presence or absence of a real asset, then the behavior of the Gini coefficient accords more closely with our intuition of "asset inequality." Nevertheless, we accept that this is a judgment issue and that different analysts might come to different conclusions.

5. Two Binary Variables

We now turn to consider the case in which we have two binary variables. We could obviously analyse both variables separately, but we might want to combine the information to arrive at some overall measure of "asset inequality." There are several potential ways of doing this. First, we could combine the two variables into one scale (an "asset index") and then apply some inequality measure to that scale. Depending on whether we think of the scale as giving us cardinal or ordinal values, we could use either a standard inequality measure or the Cowell–Flachaire ordinal measures. Second, we could utilize some of the approaches in the "multidimensional inequality" literature.

5.1. *Some Preliminaries*

First, however, we will rehearse some of the issues that make the two variable case more complicated. To make the discussion more precise, let us presume that the empirical information on the two binary variables X_1 and X_2 is contained in the following matrix:

$$\mathbf{X}=[\mathbf{x}_1 \quad \mathbf{x}_2]=\begin{bmatrix} \mathbf{0}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} \\ \mathbf{0}_{n_3} & \mathbf{1}_{n_3} \\ \mathbf{1}_{n_4} & \mathbf{1}_{n_4} \end{bmatrix}, \tag{4}$$

where $\mathbf{0}_{n_j}$ is the n_j null vector $[0,0,\ldots,0]'$ and $\mathbf{1}_{n_k}$ is the n_k vector of ones $[1,1,\ldots,1]'$. Let $N=\sum_j n_j$, and let $p_1=\frac{n_2+n_4}{N}$, $p_2=\frac{n_3+n_4}{N}$, and $p_{12}=\frac{n_4}{N}$; that is, p_1 is the proportion of the population that owns asset 1, p_2 is the proportion that owns asset 2, and p_{12} is the proportion owning both. Without loss of generality, let us assume that $p_1 \geq p_2$, that is, the second asset is rarer in the population than the first.

Besides the general case, we will also consider the polar cases:

- special case 1—
  $n_2=n_3=0$, that is, $\mathbf{x}_1=\mathbf{x}_2=[\mathbf{0}_{n_1} \quad \mathbf{1}_{n_4}]'$ with $p_1=p_2=p_{12}$; and
- special case 2—
  $n_1=n_4=0$, in which case $\mathbf{x}_1=\mathbf{1}-\mathbf{x}_2$ with $p_1=1-p_2$, and $p_{12}=0$.

What distinguishes the cases is that the correlation between the two variables is positive in the former, while it is negative in the latter. The literature on multidimensional inequality measurement speaks about a "correlation increasing majorization" (e.g. Tsui, 1999, p. 150). Intuitively, the second case, in which everyone has an asset, should be less unequal than the first, in which some people have nothing and some have everything. In general, we would like a measure of inequality that is true to that intuition.

We now turn to the first method, that of combining the two variables into one scale.

5.2. *Creating an Asset Index*

As noted above, one of the most common ways of creating an asset index is by means of principal components. Applying the PCA formula mechanically, we can derive the values of the asset index in terms of p_1, p_2, and p_{12} (see Appendix A.1 in the online supporting information; in particular, the table). Several insights follow from an examination of those formulae. Trivially, since the mean of the variables (by construction) is zero and they include positive and negative values, we cannot use traditional inequality measures on these values. Second, however, the range of the index is a function of the ranges of the standardized variables $\tilde{x}_1$ and $\tilde{x}_2$. Those are of the form $\sqrt{\frac{1-p_1}{p_1}}+\sqrt{\frac{p_1}{1-p_1}}$. These are unbounded near zero and

one and follow a "U" shape, with a minimum at $p_1 = \frac{1}{2}$. As an "inequality statistic," the range (and hence dispersion) of the asset index therefore works *inversely* to the Cowell–Flachaire statistic for the univariate case. It is unlikely to communicate useful information about real inequality in the distribution of assets. This is the contrary to the intuition, expressed for instance by McKenzie (2005), that the dispersion of the index could be a measure of inequality.

A third point emerges from the fact that the "weight" assigned to asset 1 is the sign of the correlation coefficient between the two standardized assets, which is negative if $p_{12} < p_1 p_2$. Indeed, whenever $p_{12} < p_1 p_2$, the asset scores give the following ranking: $(1,0) \prec (0,0)$ and $(1,1) \prec (0,1)$; that is, a person who has the more common asset is always ranked *below* a person who does not have the asset (see the second column of the table in Appendix A.1, in the online supporting information). How can this possibly make sense? The problem arises from the fact that the principal components analysis correctly isolates the negative correlation between the two assets. But the PCA procedure is intended to isolate what is common to both; this quandary is resolved by interpreting x_1 as a "bad" instead of a genuine asset. Given the philosophy of the PCA approach this is understandable, but it is problematic in this context nonetheless.

Indeed, it is not difficult to construct examples where the first asset becomes such an intense "bad" that a person having *no* assets gets a higher score than an individual with *both* assets. We show one example in Table 1. Indeed, whenever $1 - p_1 < p_2 < p_1$, the PCA rankings will produce such a perverse outcome. Table 1 also shows that the principal components method is not unique in this regard: the most popular alternatives, namely factor analysis (with one factor) and multiple correspondence analysis, produce precisely the same perverse ranking.

Is this case relevant for empirical analyses? There are, in fact, many practical examples where "assets" acquire negative weights in principal components procedures. In the South African case (as shown below), ownership of cattle is frequently negatively correlated with the ownership of other asset types, mainly because cattle are a typically "rural" asset while the other assets require a connection to the electricity grid.

Given these issues, it would be wise to restrict the construction of these type of asset indices to situations where the assets are positively correlated—although that would be a non-trivial limitation. Nevertheless, even in these cases, the question of how to deal with the negative values created by the estimation process remains. One possible response would be to add, to all index values, a positive

TABLE 1

An Example Where Conventional Asset Indices Produce Perverse Rankings

Assets	Proportion	PCA	FA	MCA
(1, 0)	0.5	−1.0403	−0.53318	−0.89322
(1, 1)	0.2	0.40312	0.20661	0.34613
(0, 0)	0.1	0.50277	0.25769	0.43170
(0, 1)	0.2	1.9461	0.99748	1.6711

Example: $p_1 = 0.7$, $p_2 = 0.4$, $p_{12} = 0.2, p_{12} < p_1 p_2$

constant that is big enough to ensure that only non-negative values remain. Linear translations of this sort have been used in some cases (Sahn and Stifel, 2000; Bhorat and van der Westhuizen, 2013). The problem is that while these shifts maintain the rankings, the Gini coefficients are not invariant to such transformations. Indeed, Lorenz dominance on subgroups can be reversed (see Wittenberg, 2013).

5.3. *Creating an Ordinal Scale*

With two binary variables, there are four possible outcomes, that is, four possible values for any combined scale. If we can rank the value of the two assets—for example, if having the x_1 asset is preferred to having the x_2 asset—then we know that $(1,0) \succ (0,1)$. But then we can rank all outcomes in the obvious way, that is, $(1,1) \succ (1,0) \succ (0,1) \succ (0,0)$. Even if we cannot cardinalize these bundles, we could use the Cowell–Flachaire approach to create inequality measures. Indeed, one of their examples is precisely of this form.

One problem with this approach is that it does not take into account the correlation between the two assets. Consider, for instance, two assets that can be ranked (e.g. ownership of a car versus ownership of a television) and assume that precisely half of the population have cars and the other half have televisions. Assume now that we "redistribute" the televisions to those who have cars, that is, we now have half of the population that have cars and televisions and half that have nothing. The Cowell–Flachaire measures will report the *same* inequality measures before and after, despite the fact that the distribution has become a lot more unequal.

Additional problems arise if there are more than two binary variables, because we would then need to know how owning two lower-ranking assets (e.g. a television and a cell phone) rank relative to owning only the most desirable asset (e.g. a car). There are 2^k possible bundles with k binary variables and we would need to be able to rank all of them. In practice, with more than two variables this approach is likely to be intractable. Consequently, we will not pursue these Cowell–Flachaire measures further.

6. The Multidimensional Case

6.1. *The Multidimensional Gini*

The literature dealing with multidimensional indices tends to approach the issue axiomatically. However, currently this theoretical work does not extrapolate well to contexts in which multidimensionality is defined by vectors of binary variables. Given that such situations are very much the context of this paper, we use the multidimensional Gini proposed by Banerjee (2010). As the approach is based on the Gini coefficient, it can cope better with zeroes than approaches based on generalized entropy measures (e.g. Tsui, 1999).

The procedure harks back to the first approach, that is, creating a linear combination of the variables on which the Gini coefficient is then estimated. Again, the weights on the components are given by the elements of an eigenvector

of a cross-product matrix, but in this case the variables are not demeaned, so that the moments, as it were, are calculated around zero rather than the mean. As a result, the weights are compelled to be positive. Banerjee proves that when applied to standard continuous non-negative variables, this approach provides an inequality index that satisfies all of the key axioms, but also shows increasing inequality if a "correlation increasing transfer" occurs.

More concretely, Banerjee suggests that variables should first be divided by their mean. In our case, transforming the original data matrix **X** (given in equation (4)), we obtain

$$\mathbf{A}=\begin{bmatrix} \mathbf{0}_{n_1} & \mathbf{0}_{n_1} \\ \frac{\mathbf{1}}{\mathbf{p}_1 n_2} & \mathbf{0}_{n_2} \\ \mathbf{0}_{n_3} & \frac{\mathbf{1}}{\mathbf{p}_2 n_3} \\ \frac{\mathbf{1}}{\mathbf{p}_1 n_4} & \frac{\mathbf{1}}{\mathbf{p}_2 n_4} \end{bmatrix},$$

where $\frac{\mathbf{1}}{\mathbf{p}_j n_k}$ should be interpreted in the obvious way as the n_k column vector containing $\frac{1}{p_j}$ in every element. It follows that

$$\frac{1}{N}\mathbf{A}'\mathbf{A}=\begin{bmatrix} \frac{1}{p_1} & \frac{p_{12}}{p_1 p_2} \\ \frac{p_{12}}{p_1 p_2} & \frac{1}{p_2} \end{bmatrix}.$$

A (non-normalized) eigenvector associated with the maximal eigenvalue is

$$\begin{bmatrix} \frac{p_2-p_1+\sqrt{(p_2-p_1)^2+4p_{12}^2}}{2p_{12}} & 1 \end{bmatrix}',$$

provided that $p_{12} \neq 0$. If $p_{12}=0$, then the maximal eigenvalue is $\frac{1}{p_2}$ with associated eigenvector $[0 \quad 1]$. The "index values" (up to a multiplicative constant) for this case are given in Table 2.

If $p_{12} \neq 0$ (and still assuming that $p_2 \leq p_1$), we can show that the index will order the asset bundles as $(0,0) \prec (1,0) \prec (0,1) \prec (1,1)$. However, when $p_{12}=0$—that is, the vectors of asset holdings are completely orthogonal—then the first asset gets a weight of zero, that is, $y(0,0)=y(1,0)=0$. This does not create perverse rankings, but it does mean that asset one is completely ignored. This illustrates why Banerjee in his proofs requires that there be at least one individual that owns positive quantities of all assets. In this context, this would require $p_{12} > 0$. This is undoubtedly a limitation, although one not nearly as severe as the requirement that the assets be positively correlated.

Table 2 also gives the Gini coefficient for this case. In fact, we show in the online Appendix (Section A.2) that the formula

$$G = 1 - p_2 - (p_1 + p_2 - 2p_{12})(p_1 - p_{12})u_1 - p_{12}(p_1 - p_2)u_1, \tag{5}$$

where

$$u_1 = \frac{y_1}{p_1 y_1 + p_2 y_2}, \tag{6}$$

is valid for any asset index which scores the assets as $y(0,0)=0$, $y(1,0)=y_1$, $y(0,1)=y_2$, $y(1,1)=y_1+y_2$. The formula is interesting, because it shows that $1-p_2$ is an upper bound for the Gini, as the expressions in brackets in the third and fourth terms both have to be non-negative. So the proportion of the less common asset is the key determinant for inequality overall. In the extreme case where $p_1=p_2=p_{12}$—that is, where the society splits into two groups, one which owns nothing and one which owns both assets—the upper bound is reached. Indeed, it is also reached in the case we have ruled out, where $p_{12}=0$, because then asset 1 is scored as having value zero, that is, $u_1=0$. It turns out that the behavior of the asset index and the associated Gini coefficient depend critically on p_{12}. The rarer p_{12} is, the more the procedure down-values the first asset. This is accentuated by the size of the gap between p_1 and p_2.

One interesting special case is if $p_1=p_2$. Then the Gini approaches $1-2p_2$ as $p_{12} \to 0$, that is, it treats the two assets equally and inequality gets measured according to who has any assets versus who has none. This is an attractive property, although the probability of finding such a balanced relationship in any "real world" application is zero. Nevertheless, the limiting value of $1-2p_2$ serves as a lower bound to the Gini coefficients that can be achieved.

6.2. *Some Provisional Lessons*

The key lesson is that the process of deriving weights for the asset index needs to be handled with care in any analysis of asset well-being and, in particular, in the analysis of asset inequality. The conventional PCA, FA, or MCA procedures can yield negative weights. Simply dropping these variables from the

TABLE 2

Index Values for the Uncentered Principal Components Procedure of Banerjee

Bundle	Probability	Value (y)	Case 1: $p_1=p_2$
$(0,0)$	$1-p_1-p_2+p_{12}$	0	0
$(1,0)$	p_1-p_{12}	$\frac{p_2-p_1+\sqrt{(p_2-p_1)^2+4p_{12}^2}}{2p_{12}p_1}$	$\frac{1}{p_2}$
$(0,1)$	p_2-p_{12}	$\frac{1}{p_2}$	$\frac{1}{p_2}$
$(1,1)$	p_{12}	$\frac{p_2-p_1+\sqrt{(p_2-p_1)^2+4p_{12}^2}}{2p_{12}p_1}+\frac{1}{p_2}$	$\frac{2}{p_2}$
Gini	$1-p_2-(p_1+p_2-2p_{12})(p_1-p_{12})u_1-p_{12}(p_1-p_2)u_1$		$1-2p_2+2p_{12}-\frac{p_{12}^2}{p_2}$
	where $u_1=\frac{p_2-p_1+\sqrt{(p_2-p_1)^2+4p_{12}^2}}{p_1\left(p_2-p_1+\sqrt{(p_2-p_1)^2+4p_{12}^2}+2p_{12}\right)}$		

analysis (if they are genuine assets) is likely to skew the results in other ways. The uncentered PCA (UC PCA) of Banerjee can handle these cases, provided that ownership of these assets is not completely orthogonal to that of the other assets. Nevertheless, in situations where the overlap of asset holdings is relatively small, these unconventional assets may be down-valued. It seems important to inspect both the asset scores and the resulting rankings before doing any substantive analysis. These are not mere theoretical niggles. We have focused at length on the measurement of asset inequality, and have shown that these same limitations of PCA, FA, and MCA procedures lie behind their inability to provide useful applied measures of asset inequality. In contrast, UC PCA can provide such measures.

In the next two sections, we further explore these lessons using two South African case studies. In the next section, we use South African DHS data to interrogate the asset scores that result from all of the latent variable approaches including UC PCA. We find examples of negative asset values for each of PCA, FA, and MCA. The UC PCA approach does not produce these negative values and we are able to compare the vectors of values across these techniques and go on to show the potential costs of simply dropping such negative values in analysis using PCA, FA, or MCA.

This analysis affirms that while the UC PCA approach has some limitations, it has much to recommend it. In the context of this paper, its greatest strength is that it results in an asset index that contains only non-negative values. It therefore satisfies the standard axioms of inequality analysis and can be used for inequality analysis. We show this briefly in Section 7, and in Section 8 we implement a fuller example of this by exploring changes in asset inequality in South Africa in post-apartheid South Africa.

7. Application to the DHS Wealth Indices

A 1998 Demographic and Health Survey (MEASURE DHS, 1998) allows us to apply the above discussion in the South African context in a way that explores the similarities and differences between the UC PCA approach on the one hand, and the PCA, FA, or MCA approaches on the other. The general approach to the creation of the DHS wealth indices is outlined in the paper by Rutstein and Johnson (2004). As many assets as possible are used, including country-specific ones. The actual coefficients underlying the index can be accessed on the DHS website.[3] For the South African case, the coefficients on several key variables are shown in the first column of Table 3. Note that these are the coefficients on the untransformed variables, that is, v_i/s_i (see equation (3)). The most important point for our purposes is the fact that the coefficients on the two livestock variables (possession of a donkey or horse, and possession of sheep or cattle) are both negative. It follows that individuals that have no assets will rank above individuals that have only donkey and/or cattle. Indeed, if we search for the poorest individuals (according to the wealth index), they invariably own livestock.

[3]At http://www.dhsprogram.com/topics/wealth-index/Wealth-Index-Construction.cfm

TABLE 3

Constructing Asset Indices Using the 1998 Demographic and Health Survey for South Africa

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables	DHS WI	UC PCA	UC PCA2	PCA	PCA2	MCA	FA
Water in house	0.185	0.209	0.565	0.708	0.707	0.329	0.289
Electricity	0.181	0.0814	0.220	0.663	0.657	0.300	0.265
Radio	0.097	0.0515	0.140	0.467	0.477	0.206	0.113
Television	0.165	0.101	0.273	0.678	0.680	0.312	0.301
Refrigerator	0.184	0.136	0.369	0.735	0.738	0.343	0.413
Bicycle	0.097	0.600	1.401	0.490	0.501	0.233	0.137
Motorcycle	0.166	52.57		0.788	0.821	0.412	0.193
Car	0.172	0.490	1.202	0.766	0.777	0.368	0.320
Rooms	NA*	0.0176	0.0482	0.0977	0.105	CAT	0.0221
Telephone	0.195	0.378	0.989	0.813	0.818	0.387	0.397
PC	0.207	4.984	14.42	0.967	0.982	0.481	0.296
Washing machine	0.203	0.654	1.696	0.870	0.877	0.421	0.452
Donkey/horse	−0.089	2.836	4.523	−0.293		−0.118	−0.0849
Sheep/cattle	−0.115	0.291	0.509	−0.375		−0.156	−0.0909
Infrastructure vars	Y						
Constant	Y	N	Y	Y	Y	Y	Y
Observations	12,136	12,136	12,136	12,136	12,136	12,136	12,136

Notes: *The DHS wealth index uses occupants per room rather than number of rooms.

DHS WI, DHS wealth index; UC PCA, uncentered principal components analysis; PCA, principal components analysis; MCA, multiple correspondence analysis; FA, factor analysis.

In order to investigate this further, we categorize individuals in terms of their possession (or otherwise) of "real" assets. We exclude building materials from the list and include only water piped inside the house and access to electricity. The list, with the corresponding summary statistics, is shown in Table 4. The minimal possible asset holding corresponds to one room with nothing else. Households in the

TABLE 4

Means of the Asset Variables Used in the South African DHS

	Mean	Robust Standard Error
Water in house	0.391	0.015
Electricity	0.652	0.017
Radio	0.803	0.006
Television	0.578	0.012
Refrigerator	0.507	0.014
Bicycle	0.170	0.006
Motorcycle	0.019	0.002
Car	0.252	0.011
Rooms	2.213	0.021
Telephone	0.282	0.012
PC	0.064	0.005
Washing machine	0.214	0.012
Donkey/horse	0.024	0.003
Sheep/cattle	0.100	0.007

Notes: Estimates are weighted to the population using the sample weights. Standard errors adjusted for clustering. All variables are binary except for "Rooms."

DHS with such minimal assets could have a large range of "wealth index" numbers, depending on the building material of which their accommodation was made. Interestingly, however, 13 percent of individuals who had a *higher* asset holding (typically owning livestock as well as having more rooms), nevertheless had a *lower* wealth index than the mean score among those with no moveable possessions. Indeed, the richest person among those with no water in the house, no electricity, one room, and no durables was better off (according to the wealth index) than 47 percent of individuals that had at least something on top of one room.

In order to explore the relationship between livestock ownership and other forms of assets further, we constructed a series of asset indices using our more restrictive list of assets. Besides the uncentered principal components index (labelled UC PCA in Table 3), we also constructed indices using PCA, MCA, and FA. The first point to note is that the negative weighting on livestock ownership is a feature of each of the latter three approaches. The coefficients shown in Table 3 are those on the untransformed variables, that is, before any standardization.

The second point to note is that the UC PCA also has its bizarre feature: in this case, it is the extremely large implied coefficient on ownership of a motorcycle. The reason for this is that the coefficient is v_i/μ_i, where v_i is the score from the principal components calculation and μ_i is the mean of the variable. We divide by μ_i due to the standardization suggested by Banerjee. As Table 4 shows, motorcycles are owned by very few South Africans, and consequently the score becomes inflated in ways which are unlikely to reflect their real asset status. Consequently, we decided to drop this variable and recalculate the index (the results are shown in column 3). Ownership of a personal computer now gets the highest score, although its magnitude is not as outlandish as that for the motorcycle.

Similarly, we also recalculated the PCA index without the livestock variables, to provide the fairest comparison between the two techniques. This, however, did not have much of an impact on the remaining coefficients, as can be seen by comparing columns 4 and 5 in Table 3. It will, of course, remove the anomalies noted earlier. Individuals owning livestock will now appear indistinguishable from individuals owning nothing. What is the impact of this for the identification of deprivation?

One simple check is to divide the population up into quintiles according to the two indices and see how well they compare. Table 5 performs that analysis.

TABLE 5

Comparing the Quantiles of the Uncentered versus the Usual PCA

	Quantiles of PCA 2					
Quantiles of UC PCA2	1	2	3	4	5	Total
1	2,368	482	0	0	0	2,850
2	530	1,145	748	0	0	2,423
3	34	429	1,277	586	0	2,326
4	0	66	275	1,463	399	2,203
5	175	104	55	84	1,912	2,330
Total	3,107	2,226	2,355	2,133	2,311	12,132

TABLE 6
Correlations Between the Different Asset Indices

	DHS WI	PCA	PCA 2	MCA	FA	UC PCA	UC PCA 2
DHS WI	1						
PCA	0.9435	1					
PCA 2	0.9337	0.9974	1				
MCA	0.94	0.999	0.9973	1			
FA	0.9449	0.9968	0.9952	0.9959	1		
UC PCA	0.3059	0.3862	0.3995	0.3956	0.362	1	
UC PCA 2	0.6247	0.7391	0.7559	0.747	0.7234	0.4539	1

We see that there are some key differences. The starkest contrast is provided by the 175 households which are rated in the bottom quintile according to the PCA index but are rated at the top of the UC PCA. Looking at the means of the asset variables, it emerges that all of them owned horses/donkeys, 76 percent of them also owned sheep or cattle, and 75 percent of them also owned a radio. Ownership of horses and/or donkeys is a significant asset according to the uncentered PCA. Perhaps the coefficient is on the large side, but it is unlikely that households that own both types of livestock should truly be ranked among the poorest of the poor (the bottom 20 percent). Of course, the original PCA index would have ranked many of these households below the "poorest of the poor" (given the negative value on those assets).

In Table 6, we present the correlation matrix between the different asset indices. Although we have used fewer assets in our version of the principal components scores, they are still highly correlated with the wealth index released with the DHS. The PCA, FA, and MCA approaches end up highly correlated. The two uncentered PCA indices show much lower correlations. The first of these has very low correlations with all the indices, since motorcycle owners receive such high scores that the entire distribution is highly skewed (95 percent of all scores are below 8, whereas motorcycle owners score above 50). The second shows correlations of 0.75 with the PCA index that does not weight livestock negatively—but correspondingly lower correlations with the others that maintained that negative weighting.

The obvious implication of all of this is that the standard asset indices will tend to find higher urban–rural contrasts in poverty than the uncentered PCA. This is shown clearly in Table 7. In each case, we have classified the bottom 40 percent of individuals as "poor" according to the DHS wealth index, the PCA 2 index and the second uncentered PCA index. It is clear that there is a strong urban–rural poverty gradient. Nevertheless, the DHS wealth index accentuates this contrast, while the uncentered PCA index finds more urban poverty and less rural poverty. This should not be surprising given the negative valuation of rural assets in the DHS wealth index and the strong positive valuations of urban infrastructure.

Interestingly, calculating the Gini coefficient on the asset scores of the UC PCA we find (in Table 8) strong asset inequality in South Africa in 1998, not dissimilar to the magnitude of income inequality (Leibbrandt *et al.*, 2010). Furthermore, as this table also suggests, there were strong inequalities within rural areas, a finding that many South Africans will find plausible.

TABLE 7
Proportion Poor in Different Types of Localities, According to Different Asset Indices

	DHS		PCA 2		UC PCA 2	
Localities	Mean	Linearized Standard Error	Mean	Linearized Standard Error	Mean	Linearized Standard Error
Capital, large city	0.098	0.013	0.146	0.014	0.198	0.015
Small city	0.178	0.024	0.220	0.021	0.275	0.022
Town	0.204	0.031	0.291	0.032	0.372	0.033
Countryside	0.720	0.020	0.648	0.019	0.597	0.016

Note: "Poor" defined as the bottom 40 percent in terms of the index.

8. Asset Inequality in South Africa, 1993–2008

We now turn to consider the evolution of asset inequality in South Africa using two nationally representative surveys conducted under the auspices of SALDRU at the University of Cape Town. The first of these is the Project for Statistics on Living Standards and Development (PSLSD), conducted in 1993, and the second is the first wave of the National Income Dynamics Study (NIDS). These studies have already been used to investigate changes in money-metric income inequality over the period (Leibbrandt *et al.*, 2010). It has been found that over this period, money-metric inequality started at very high levels and remained at those high levels.

Both of these surveys are nationally representative general living standards instruments that gathered detailed information on incomes, expenditures, and assets, as well as education, health, and other dimensions of well-being. The literature on money-metric inequality has been useful in giving detailed attention to the comparability of the incomes and expenditure in these two surveys over time (Leibbrandt *et al.*, 2010).

The two datasets provide good coverage of household assets. However, they differ in asset registries. In total, 31 assets categories exist, of which the NIDS contains 29 and the PSLSD contains 19. The NIDS does not include an electrical kettle or the presence of a geyser in its asset register. Some of the assets not included in the PSLSD are due to technological progress. Assets such as computers and cell phones were not as prominent in 1993 as they are now and thus were not included. Furthermore, the NIDS includes greater detail with regard to

TABLE 8
Asset Inequality Measured by the Gini Coefficient Using the UC PCA2 Index

Group	Estimate	Standard Error	Lower Bound	Upper Bound
1: capital, large city	0.566	0.009	0.549	0.583
2: small city	0.538	0.014	0.511	0.566
3: town	0.569	0.023	0.524	0.614
4: countryside	0.609	0.014	0.582	0.636
Population	0.623	0.007	0.610	0.636

Note: Statistics calculated using the DASP package.

TABLE 9
Asset Holdings in 1993 and 2008

Over	Mean	Linearized Standard Error
electricity		
1993	0.459	0.024
2008	0.779	0.020
pipedwater		
1993	0.506	0.027
2008	0.697	0.025
radio		
1993	0.811	0.008
2008	0.694	0.012
TV		
1993	0.477	0.018
2008	0.703	0.017
fridge		
1993	0.399	0.020
2008	0.609	0.020
motor		
1993	0.247	0.016
2008	0.220	0.018
livestock		
1993	0.110	0.011
2008	0.100	0.011
landline		
1993	0.242	0.018
2008	0.143	0.015
cellphone		
2008	0.807	0.011
phoneany		
1993	0.242	0.018
2008	0.827	0.010

Source: 1993 PSLSD, 2008 NIDS wave 1.

transportation assets (such as motorcycles, boats, and donkey carts) as well as agricultural assets (such as tractors, ploughs, and grinding mills), which are not included in the PSLSD. However, the PSLSD has the advantage of not only including ownership of assets but also the quantity of each asset owned.

In order to look at asset inequality over time, we need to calculate a pooled index for the two periods first, so that we are using the same scores for the assets in each period. This limits us to assets that were asked for in both periods. The descriptive statistics presented by Bhorat and van der Westhuizen (2013) suggest that there has been considerable progress over the period. Table 9 presents the statistics as calculated on our data.

One immediately evident issue is that the prevalence of landlines has gone down as the availability of cell phones has become ubiquitous. If this measurement issue is not addressed, it will result in a spurious decrease in assets over time. Indeed, given the relative rarity of landlines in the later period, these would become erroneously marked as valuable assets instead of as assets whose utility is actually in decline. Consequently, we collapse landline and cell phone ownership into an omnibus "any phone" variable. The coefficients on the assets implied by our uncentered PCA asset index are given in Table 10.

TABLE 10
Coefficients on the Asset Variables Used in the Pooled UC PCA Index

electricity	0.515
pipedwater	0.536
radio	0.353
TV	0.655
phoneany	0.789
fridge	0.800
motor	2.265
livestock	3.170

When we use this asset index to construct Lorenz curves, we obtain the result shown in Figure 2. The Lorenz curves show clear evidence that asset inequality fell considerably, and this is confirmed by the Gini coefficients, which fell markedly, from 0.47 in 1993 to 0.29 in 2008. As a reflection of the fact that these Lorenz curves and Gini coefficients were estimated from the pooled UCPC measure, the pooled or Population Lorenz curve is plotted in the figure too.

The fact that asset inequality should have declined is not surprising given that the statistics shown in Table 9 show strong increases in access to assets between 1993 and 2008. This is not universally true—motor cars, for instance, remain relatively rare. Nevertheless, the penetration of television, cell phones, refrigerators, and electricity suggest that asset holdings have certainly increased. By contrast, the money-metric measures suggest very little change. Part of the problem, of course, is that if the whole distribution shifts upward by an equiproportionate amount of money, the measured money-metric inequality will remain static. Note, however, that dummy variables cannot be rescaled in this way. The way in which we measure asset inequality will make asset ownership more common at the bottom, leave it unchanged at the top, and thus reduce inequality.

It is also true, of course, that all measurements are contingent on the schedules that are employed. One note of caution in this regard is appropriate. The asset inequality measure for 1998 that we calculated for the DHS is significantly higher than either the 1993 or 2008 measures that we have just considered. The main reason for this is that the asset schedule for 1998 included assets such as "personal computer," which allowed a better contrast to be drawn between high earners and the rest.[4] While we are sure that access to assets has spread and that in this sense asset inequality has decreased, the magnitude of the initial level of inequality and the size of the decrease are probably not as dramatic as suggested by the Gini coefficients that we have reported.

9. Conclusion

In this paper, we have argued that asset indices can be interesting and powerful tools for analysing social trends. However, doing so in an unreflective and automatic way is unlikely to provide useful insights. We have drawn particular

[4]Of course, the case of the motorcycle should remind us that some of these contrasts can be overdrawn.

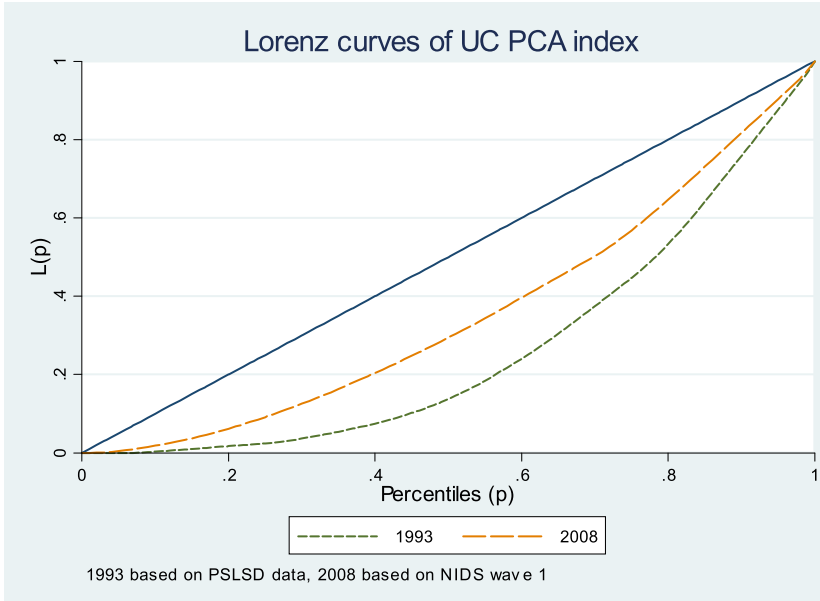

Figure 2. Lorenz curves for the UC PCA index [Colour figure can be viewed at wileyonlinelibrary.com]

attention to the fact that the standard approaches often value access to a good such as livestock negatively, implying that the household would be better off without access to that good. Proceeding in this way, for instance, has obscured real asset holdings in rural areas in the South African case. We go on to show how this has led to an exaggerated sense of rural deprivation and a lack of appreciation for deprivation in urban areas. This is not just about South Africa. DHS weights are available for a range of other African countries and these show that it is common for there to be negative weights on rural assets such as landholding and cattle.[5] There is a large literature showing that it is these very assets that are stores of value or wealth in many rural African contexts, which seems to provide strong support to the salience and importance of this point.

A related focus of this paper has been to link these problematic properties of widely used asset indices to the limitations of these indices in measuring asset inequality. So, the standard application of these indices may also have obscured real inequality within rural areas. But this has been hard to ascertain up to this point, as these indices do not allow for the measurement of asset inequality.

Our analysis has gone on to suggest that it is possible to create asset indices in ways that allow the calculation of Gini coefficients. To that end, we have used the method suggested by Banerjee for the calculation of "multidimensional Gini coefficients" using continuous data. Our application suggests that the technique can work well, provided that care is taken in ensuring that some rare assets do

[5]Examples include Burundi (2012), DRC (2013–14), Equatorial Guinea (2011), Eritrea (2002), Ghana (2014), Mali (2006), Mauretania (2000–1), Morocco (2003–4), Namibia (2013), Niger (2012), Rwanda (2014–15), Sierra Leone (2008), Swaziland (2006–7), and Zambia (2013–14).

not distort the index. In general, then, whether researchers are using our proposed approach to deriving an asset index and measuring inequality, we have shown clearly in this paper that such indices should not be used without scrutinizing the implied coefficients.

We have used nationally representative survey data from 1993 and 2008 to derive an uncentered principal components analysis asset index for South Africa spanning the initial 15 years of post-apartheid South Africa. We have used this index to analyse how asset inequality has changed in South Africa between these two years. We have plotted comparable Lorenz curves and derived comparable Gini coefficients. The Lorenz curves show unambiguously that asset inequality has declined sharply over time. The Gini coefficients give a sense of the extent of this decline. This picture of falling asset inequality contrasts sharply with the money-metric analysis of inequality over the same period. The latter narrative is one of very high inequality in 1993 that does not fall over the post-apartheid years. Substantively, our empirical work suggests that the money-metric approach to inequality measurement in South Africa may have obscured the real progress in large portions of the population and in important dimensions of inequality.

Still, this stark difference does prompt some reflections on the limitations of the scope of our analysis in this paper. We have focused on the derivation of asset indices and asset inequality from a binary view of assets: whether households do or do not have access to them. Such are the data that we have in many developing countries and, as a consequence, such asset indices are very common in the international literature; thus justifying the focus of the paper.

Nonetheless, this binary view of the world misses the complexity arising from a more continuous approach. We cannot differentiate between households that have many instances of an asset (such as TVs) and those that only have one. Nor does it take account of the differing quality or values of the assets or the real returns that they bestow on the household. In short, our "asset indices" fall considerably short of true wealth indices.

Finally, it is worthwhile pointing out that "asset indices" are used toward many different ends: trying to identify the poor and deprived; measuring the gap between the rich and the poor; and ranking households in terms of their quality of life. Measures such as the Cowell–Flachaire one are aimed at only one of these objectives and perform well in that context. The fact that Cowell–Flachaire does not perform as well in the situation that we analyse is not to deny that utility. Indeed, it is too much to expect that one type of index could address all of the issues listed and do so equally well. The same is true of the UC PCA index. It is not a tool that will work in all contexts. Our point that asset indices should not be used in an unreflective way also extends here: it is vital to think about what goes into the index as well as how it is assembled.

References

Banerjee, A. K., "A Multidimensional Gini Index," *Mathematical Social Sciences*, 60, 87–93, 2010.

Bhorat, H. and C. van der Westhuizen, "Non-monetary Dimensions of Well-Being in South Africa, 1993–2004: A Post-Apartheid Dividend?" *Development Southern Africa*, 30, 295–314, 2013.

Booysen, F., S. van der Berg, M. von Maltitz, and G. du Rand, "Using an Asset Index to Assess Trends in Poverty in Seven Sub-Saharan African Countries," *World Development*, 36, 1113–30, 2008.

Cowell, F. A. and E. Flachaire, "Inequality with Ordinal Data," mimeo, London School of Economics, 2012; available at http://darp.lse.ac.uk/pdf/IneqOrdinal.pdf

Filmer, D. and L. H. Pritchett, "Estimating Wealth Effects without Expenditure Data—or Tears: An Application to Educational Enrollment in States of India," *Demography*, 38, 115–32, 2001.

Filmer, D. and K. Scott, "Assessing Asset Indices," *Demography*, 49, 359–92, 2012.

Houweling, T. A., A. E. Kunst, and J. P. Mackenbach, "Measuring Health Inequality among Children in Developing Countries: Does the Choice of the Indicator of Economic Status Matter?" *International Journal for Equity in Health*, 2, 2003.

Leibbrandt, M., I. Woolard, A. Finn, and J. Argent, "Trends in South African Income Distribution and Poverty since the Fall of Apartheid," Social, Employment and Migration Working Papers 101, OECD, 2010; available at https://doi.org/10.1787/5kmms0t7p1ms-en

McKenzie, D. J., "Measuring Inequality with Asset Indicators," *Journal of Population Economics*, 18, 229–60, 2005.

MEASURE DHS, "Demographic and Health Survey, South Africa 1998" (dataset), 1998; available at http://www.measureDHS.com

Rutstein, S. O. and K. Johnson, "The DHS Wealth Index," DHS Comparative Reports 6, ORC Macro, Calverton, MD, 2004.

Sahn, D. E. and D. Stifel, "Poverty Comparisons Over Time and Across Countries in Africa," *World Development*, 28, 2123–55, 2000.

———, "Exploring Alternative Measures of Welfare in the Absence of Expenditure Data," *Review of Income and Wealth*, 49, 463–89, 2003.

Southern Africa Labour and Development Research Unit, "Project for Statistics on Living Standards and Development, 1993" (dataset), Southern Africa Labour and Development Research Unit, distributed by DataFirst, Cape Town, South Africa, 1993.

Southern Africa Labour and Development Research Unit, "National Income Dynamics Study 2008, Wave 1" (dataset), version 5.3, Southern Africa Labour and Development Research Unit, distributed by DataFirst, Cape Town, South Africa, 2015.

Tsui, K., "Multidimensional Inequality and Multidimensional Generalized Entropy Measures: An Axiomatic Derivation," *Social Choice and Welfare*, 16, 145–57, 1999.

Wagstaff, A., "The Bounds of the Concentration Index when the Variable of Interest is Binary, with an Application to Immunization Inequality," *Health Economics*, 14, 429–32, 2005.

Wittenberg, M., "Weighing the Value of Asset Proxies: The Case of the Body Mass Index in South Africa," Working Paper 39, SALDRU, University of Cape Town, Cape Town, 2009; available at http://www.opensaldru.uct.ac.za/handle/11090/16

———, "Estimating Expenditure Impacts without Expenditure Data using Asset Proxies," *Economics Letters*, 110, 122–5, 2011.

———, "Non-monetary Dimensions of Well-Being: A Comment," *Development Southern Africa*, 30, 826–9, 2013.

Supporting Information

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

Appendix A Derivations
A.1 The PCA Index in the Case of Two Binary Variables
A.2 The Gini Coefficient for the Bivariate Case
Figure A.1: Calculating the Gini Coefficient in the Bivariate Case, where $y(0;0)=0$