Beyond the human development index: a stochastic spanning methodology

Mehmet Pinar[a], Thanasis Stengos[b], Nikolas Topaloglou[c]

Abstract

The well-known Human Development Index (HDI) goes beyond single measure of well-being comparisons by obtaining a composite index by combining achievements in education, income and health dimensions. Yet, many still argue that dimensions covered by the HDI do not reflect overall well-being of societies and new indicators should be included to the HDI. This paper offers a new methodology, stochastic spanning, to test the inclusion of new dimensions to the HDI. This methodology takes the current measurement of HDI as a benchmark and tests whether spanning occurs or not by examining the distributions of composite scores obtained from a particular expansion of the feasible choice set or not. We use this methodology to test the inclusion of twelve indicators of governance to the benchmark components of the HDI and find that inclusion of some governance indicators to the list of components of the HDI leads to welfare improvements.

JEL classifications: C14; I31; O15

Keywords: well-being; human development; composite indices; institutions; nonparametric stochastic dominance

---

[a] Business School, Edge Hill University, St Helens Road, Ormskirk, Lancashire L39 4QP, United Kingdom; Tel: 0044 1695 657629; Fax: 0044 1695 584675; e-mail: mehmet.pinar@edgehill.ac.uk

[b] Department of Economics, University of Guelph, N1G 2W1, Guelph, Ontario, Canada; e-mail: tstengos@uoguelph.ca

[c] Department of International European & Economic Studies, Athens University of Economics and Business, 76, Patision Street, GR10434, Athens, Greece; e-mail: nikolas@aueb.gr

1 Introduction

Well-being is inherently a multidimensional concept and it is commonly measured in that way (see, e.g., Fleurbaey 2009; Fleurbaey and Blanchet, 2013) as policy makers propose different frameworks to measure sustainable development that rely on multidimensional indices (Hall et al., 2010). For example, the European Commission's 'Going beyond GDP' initiative aims to assess development indicators that go beyond GDP to measure the progress in the social, environmental, and economic aspects of development in a multidimensional manner. Stiglitz et al. (2009) proposed eight dimensions beyond GDP to measure multidimensional well-being, and these include standard of living, health, education, political voice and governance, among the other dimensions. Furthermore, there is an extensive list of composite indices that is being developed to measure the overall multidimensional welfare progress of countries (see Singh et al., 2012 for a recent overview of large set of composite indices). Some of these composite indices have been obtained by simple aggregation of different welfare dimensions. For example, the environmental sustainability index (ESI) is a measure of the overall progress towards environmental sustainability (Esty et al. 2005) measured by the arithmetic average of environmental indicators, while the Human Development Index (HDI) of the United Nations Development Programme (UNDP) is a geometric mean of the education, life expectancy and GNI per capita (Malik, 2013). Furthermore, the FEEM sustainability index (FEEM SI) aggregates 19 indicators under three main pillars – environmental, social and economic – through weights selected by experts (see Pinar et al. 2014) and OECD's Better Life Index (BLI) is aggregated through preferences of individuals on different well-being indicators (see OECD, 2011; Boarini and D'Ercole, 2013; Durand, 2015).

The above mentioned composite indices serve to convey a quick message to stakeholders and policy makers about the strengths and weaknesses of current policies that promote development (see e.g., Ness et al., 2007). Yet, they rely on implicit value judgements (i.e., selection of the indicators, normalization of indicators, and weight allocation to indicators) and as such they have drawn criticisms in the literature (see Decancq and Lugo, 2013 for a comprehensive discussion on the topic). For instance, the construction of HDI received many criticisms with respect to the selection of indicators (e.g., Ranis et al. 2006 suggested inclusion of more indicators to the already existing set of indicators; Bravo, 2014 analyzed the inclusion of environmental dimension to the HDI; additional dimensions beyond the three components of the HDI are suggested by Stiglitz et al., 2009, among many others), normalization of indicators prior to aggregation (see e.g., Noorbakhsh, 1998 and Pollesch and Dale, 2016 for further discussion on how normalization plays a role on implicit weights and affects the composite scores), and weight allocation across the well-being dimensions (see e.g., the effects of choice of weights on both the ranking and composite achievement levels are examined by Cherchye et al., 2008; Permanyer, 2011; Foster et al., 2013; Pinar et al. 2013; Tofallis 2013; Athanassoglou, 2015, among many others)[1].

In this paper, we will adopt a data driven methodology that allows for the possible inclusion of additional well being components to HDI based on stochastic dominance (SD) spanning analysis. More specifically, we will examine whether the inclusion of some other indicator to the already existing composite HDI leads to an improvement in well being by comparing the empirical achievement distribution of the benchmark HDI with that of the new augmented index. The comparison of the empirical achievement distributions with and without the included indicator will be done using SD spanning, which extends the SD efficiency methodology to test whether inclusion of an indicator to the existing composite index (i.e., HDI) leads to welfare improvements. SD is a powerful framework of analysis that has been used in a wide variety of applications in economics, finance and statistics (see e.g., Levy (2015) for an overview and references). Due to its non-parametric attractiveness, SD is particularly appealing for comparisons of variables with asymmetric profiles (e.g., income, life expectancy, human capital

[1] Since the choice of weights can be considered as paternalistic (see e.g., Decancq et al., 2015), alternative multidimensional well-being comparisons based on equivalent incomes were offered by Fleurbaey and Blanchet (2013) and Decancq and Schokkaert (2016).

distributions among different countries). Davidson and Duclos (2000), Barrett and Donald (2003) and Linton et al. (2005), among others, develop statistical tests for such pairwise comparisons. Recent developments in the SD methodology has given rise to multivariate well-being comparison of countries (see, e.g., Duclos et al., 2006), yet none of these papers allowed for differential weights to be assigned to each dimension in the multivariate context. In the finance literature, a more general, multivariate problem is that of testing whether a given portfolio is stochastically efficient relative to all mixtures of a discrete set of alternatives (Post, 2003; Kuosmanen, 2004; Roman et al., 2006). Post and Versijp (2007), Scaillet and Topaloglou (2010), Linton et al. (2014) and Post and Poti (2017) address this problem with various proposed SD efficiency (SDE) tests. The latter are used to examine the existence of alternative ways of combining assets that dominate the market. In recent papers, Pinar et al. (2013, 2017) used the SDE methodology to obtain the best-case scenario weighting scheme for the HDI with the sub-indices used prior and after the 2010 edition. In their paper, Pinar et al. (2013) found that if one were to weight the education index relatively more than the pre-determined equal weights, that would tend to result in a more optimistic way of measuring welfare,where the education index would be assigned a higher implicit weight. This approach results in a new HDI with weights that are given to the component(s) where most countries have been most successful in fulfilling. However, since the majority of countries would have already achieved good levels of literacy and enrolment ratios (i.e., the indicators used to measure the education index prior to the 2010 edition of Human Development Report), it would appear that these indicators would not serve any longer much of a purpose for relative welfare comparisons. In that case the optimistic scenario would not be very useful from a policy perspective as it showcases a weighted average of components where most countries would be already successful[2]. The opposite case would be to derive the pessimistic scenario which would highlight the components with the

[2]In a recent paper by Pinar et al. (2017), the new measurement of the education index (measured by the mean of years of schooling for adults aged 25 years and above, and expected years of schooling for children of school entering age) receives relatively lower implicit weights.

least success. In such a case, HDI rankings using the pessimistic weights would give incentives to countries that lag behind to make improvements on the respective components where they are least successful in achieving.

The concept of SD spanning has been recently introduced by Arvanitis et al. (2018) as a model-free alternative to mean-variance spanning (Huberman and Kandel, 1987). In the context of HDI, SD spanning occurs if there is no benefits from a particular expansion of the given feasible choice set (i.e., income, life expectancy, education) or to put it differently, the exclusion of some other potential component would not worsen-off the distributional achievements of countries. The null hypothesis in the SD spanning testing framework is that the introduction of a new component of human development is not welfare enhancing. Consequently, we will test the null hypothesis that the components of the benchmark HDI provide the best welfare outcome available when compared with alternative sets which also include one of each of twelve different indicators that measure different aspects of governance (or institutional quality). From a behavioral perspective, spanning implies that the problem presentation can be simplified by excluding the redundant choice alternatives. If the inclusion of an indicator leads to a welfare improving augmentation of the choice set of indicators, this indicator's inclusion would imply rejection of the null hypothesis as the original set does not form a spanning set. On the other hand, not rejecting the null hypothesis would suggest that the inclusion of the additional indicator would not lead to any welfare gain. SD spanning can be considered as an assessment tool to test the inclusion of indicators and allow policy makers to implement changes in the indicator list that make up the given index (in our case the HDI). In other words rejection of the benchmark spanning set would allow policy makers to augment the set of welfare components beyond the standard ones and assess the relative progress of a given country, if well being were to be measured as a weighted average of the components in the new enlarged set. In this study, we focus on the second-order stochastic dominance (SSD), because the assumptions of non-satiation and risk aversion are generally well accepted in the context of expected utility theory and Yaari's (1987) dual theory of risk, and are robust to common transformations of well-being outcomes and probabilities.

The remaining part of the paper is organized as follows. Section 2 provides a discussion why inclusion of governance as part of the composite development index is conceptually a good way forward. Section 3 presents the stochastic spanning framework. Section 4 presents the data, and the results obtained with the empirical application of the SD spanning methodology. Finally section 5 concludes. The appendix also presents the statistical theory underlying the stochastic spanning tests and the computational strategy for the test statistic.

2 Governance and well-being

In this section, we provide a discussion on why the inclusion of governance measures into the list of potential HDI components is worthwhile. There has been an extensive discussion on the importance of governance that links it to Sen's capability approach (Sen, 1985; 1987; 1999). It has been argued that the quality of the legal system and the presence of political rights could foster freedom of thought and political participation that would improve the capabilities of individuals (see e.g., Robeyns, 2005 for detailed discussion and survey on social and formal institutions' relationship with the capabilities approach). Stiglitz et al. (2009) also highlight the importance of political voice and governance (concepts that are closely linked with institutional quality) in shaping freedom of choice and speech and that better rule of law and legislative guarantees would enhance the quality of life of all citizens (see section 4.4 of the report for further discussion on how political voice and governance reinforce better quality of life for citizens). The importance of governance has also been emphasized by the United Nations Assembly (2013) and it has been pointed out that the "implementation of a post-2015 development agenda will depend, critically, on effective governance capacities" (p. 33, UN System Task Team on the

Post-2015 UN Development Agenda, 2012a)[3].

Beyond the conceptual importance of governance, there exists an extensive empirical literature that identifies its importance for social and economic outcomes. For instance, it has been long argued that the quality of institutions is one of the main factors that explains the long-term income differences across countries (see e.g., Acemoglu et al., 2001; Easterly and Levine, 2003; Rodrik et al., 2004; Bosker and Garretsen, 2009; Acemoglu et al., 2014; Pinar, 2015 among many others). In a seminal paper, Acemoglu et al. (2001) argued that the disease environment faced by the European settlers promoted different colonization strategies that led to a different sets of institutions, and these institutional differences are found to be the main reason for the income per capita differences between countries. Rodrik et al. (2004) extended Acemoglu et al. (2001) sample and also found that the quality of the institutions constitutes the primary reason for the cross-country income per capita differences. Bosker and Garretsen (2009) found that better institutions lead to higher long-term income per capita even after controlling for the effect of the neighboring institutions. In a recent application, Barro (2015) found that countries with better rule of law have higher growth rates suggesting that the econometric problems posed by country fixed effects may not be serious in samples within a long time frame, something that contradicts the findings of Acemoglu et al. (2008, 2009) and is consistent with the modernization theory of Lipset (1959)[4].

It has been found that better institutions not only lead to long-term economic development but also aid is more effective in countries with good policies. The Monterrey consensus suggested that countries with sound institutions make a more effective use of foreign aid (World Bank, 2003). In a seminal paper, Collier and Dollar (2002) found that the effectiveness of foreign aid depends on the quality of policies and institutions, whereas Kosack (2003) found that foreign

[3]See also UN System Task Team on the Post-2015 UN Development Agenda (2012b) for the detailed importance of governance for the eradication for poverty and inequality, among many other factors.

[4]Institutional quality also led to an increased foreign capital flows and foreign direct investment (see e.g., Busse and Hefeker, 2007; Alforo et al., 2008; Papaioannou, 2009; Azemar and Desbordes, 2013 among many others), which in turn reinforces economic growth (see, e.g., Borensztein et al., 1998; Haskel et al., 2007).

aid is effective in improving the quality of life in countries with more democratic institutions, suggesting that aid allocation should be combined with the democratization effort of a given country (see also Burnside and Dollar, 2004; Dollar and Levin, 2006; Chong et al., 2009; Tebaldi and Mohan, 2010; Roodman, 2012). Overall, it has been found that aid empowers the poor in a good institutional setting and improves the functionings and capabilities of individuals[5].

The main argument is that countries with better governance have better economic and social conditions, which then increase the capability and well-being of the individuals living in these countries. Therefore, inclusion of the governance dimension to the already existing component list of HDI would be an important step. However, given that there are many proxies of governance (or institutional quality), the question then arises of which one of these proxies would be suitable for inclusion? In this paper, we examine a list of different possible governance proxies that could be included in the list of components of HDI.

3 Stochastic Spanning

SD is traditionally applied for comparing a pair of two distributions of given characteristics, and SD efficiency is a direct extension of SD to the case where full diversification is allowed. This is a multivariate problem of testing whether a given combination of characteristics (an index) is stochastically efficient relative to all mixtures of a discrete set of alternative characteristics (alternative indices). Pinar et al. (2013) use this methodology to test for SD efficiency of the official HDI (the given index) with respect to all possible combination of weighting schemes (the set of indices) constructed from the set of components. Stochastic spanning is a generalization of SD efficiency because it involves the comparison of two sets of alternatives,

[5]One should note that there has been an on-going discussion on how aid should be allocated that suggests that donors should also take into account other factors beyond good policies (see e.g., recent discussions on this issue by Bourguignon and Platteau, 2013; Guillaumont et al., 2017). However, our main concern for this paper is that countries with good governance use aid more effectively and governance promotes a better environment for individuals living in these countries.

while SD efficiency is a special case where one of the two choice sets is a singleton. We adopt stochastic spanning to test whether the inclusion of governance proxies to the original HDI components may lead to welfare improvements. If we were to add one such governance proxy as an additional component and we fail to reject the null of spanning, then this additional component will be redundant, whereas if we reject spanning, then there are optimal combinations of the augmented index that includes all four components that dominate any combination of income, health, and education. In that case, the given governance indicator leads to welfare improvements and should be included in measuring welfare. To summarize, the idea behind stochastic spanning is that enlarging the set of potential outcomes does not lead to welfare gains. If there are gains then spanning is rejected. Below we will provide a formal presentation of stochastic spanning as it applies to the construction of an augmented HDI. In the appendix we present the testing and computational framework that underlies our approach. The welfare universe consists of M components with outcomes $X := (x_1, \ldots, x_M)$ with support bounded by $X^M := [\underline{x}, \overline{x}]^M ,: -\infty < \underline{x} < \overline{x} < +\infty$. X can be chosen arbitrarily if it is a superset of the maximal support of the base components. The data correspond to observed values of the M different constituent components of well-being (or welfare). The components are treated as random variables with a discrete, state-dependent, joint probability distribution characterized by R mutually exclusive and exhaustive scenarios with probabilities $p_r > 0, r = 1, ..., R$. The feasible combinations of components are represented by a bounded polyhedral set, M-simplex $\Lambda := \left\{\lambda \in \mathbb{R}_+^M : 1_M^{\mathrm{T}}\lambda = 1\right\}$.

Let $F : R^M \to [0, 1]$ denote the continuous joint cumulative distribution functions (cdf's) of X and $F(y, \lambda) := \int 1(X^{\mathrm{T}}\lambda \le y)dF(X)$ the marginal cdf for combination $\lambda \in \Lambda$ where 1 is the indicator function. In order to define SD spanning and efficiency, we use the following integrated cdf:

$$\mathbf{F}^{(2)}(\mathbf{x}, \lambda) := \int_{-\infty}^{x} \mathbf{F}(\mathbf{y}, \lambda)\mathbf{dy} = \int_{-\infty}^{x} (\mathbf{x} - \mathbf{y})\mathbf{dF}(\mathbf{y}, \lambda). \tag{3.1}$$

This measure corresponds to Bawa's (1975) first-order lower-partial moment, or expected shortfall, for a given threshold $x \in X$.

This study focuses on the effects of changing the set of benchmark components of the HDI (i.e., income, health, and education). For this purpose, we introduce a non-empty polyhedral subset $K \subset \Lambda$. A polyhedral structure is analytically convenient and arises naturally if we remove some of the base components or tighten the linear constraints which define L.

Stochastic spanning: The set of components Λ is second-order stochastically spanned by subset $K \subset \Lambda$ if all combinations $\lambda \in \Lambda$ are weakly second-order stochastically dominated by some combinations of components $\kappa \in K$:

$$(\kappa \succeq_F \lambda \ \kappa \in \mathrm{K}) : \forall \lambda \in \Lambda \Longleftrightarrow$$

$$((G(x, \kappa, \lambda; F) \leq 0 : \forall x \in \mathcal{X}) : \kappa \in \mathrm{K}) : \forall \lambda \in \Lambda \tag{3.2}$$

$$G(x, \lambda, \tau; F) := F^{(2)}(x, \lambda) - F^{(2)}(x, \tau). \tag{3.3}$$

We will use $R(\Lambda) := \{\mathrm{K} \subseteq \Lambda : (\kappa \succeq_F \lambda \ \kappa \in \mathrm{K}) : \forall \lambda \in \Lambda\}$ to denote all relevant subsets that span Λ. Spanning occurs if and only if $K \in R(\Lambda)$. $R(\Lambda)$ is non-empty because it includes at least Λ; a span $K \in R(\Lambda)$ may itself be spanned by another span $K' \in R(K) \subseteq R(\Lambda)$.

Stochastic spanning occurs if the enlargement $(\Lambda - K)$ does not change the efficient set (i.e., the most optimistic combination of the sub-components of the HDI, see Pinar et al., 2013; 2017 for the details), that is:

$$K \in R(\Lambda) \Longleftarrow E(\Lambda) \subseteq K. \tag{3.4}$$

The reverse relation generally does not hold, because the weak dominance relation does not possess the antisymmetric property. In other words, $E(\Lambda)$ always spans Λ, but it may be reducible by excluding equivalent elements. Consequently, $E(\Lambda) \subseteq K$ is a sufficient but not necessary condition for $K \in R(\Lambda)$. In addition, the sufficient condition $E(\Lambda) \subseteq K$ is not practical, because $E(\Lambda)$ is generally non-convex and disconnected, which makes it difficult to identify all its elements and test the sufficient condition directly. On the contrary, a small polyhedral span $K \in R(\Lambda)$ could be used as a practical approximation to the intractable efficient set $E(\Lambda)$.

We use the following scalar-valued functional of the population cdf as a measure for deviations from stochastic spanning:

$$\eta(F) := \sup_{\lambda\in\Lambda} \inf_{\kappa\in\mathrm{K}} \sup_{x\in\mathcal{X}} G(x,\kappa,\lambda;F). \tag{3.5}$$

The outer maximization searches for a feasible combination $\lambda \in \Lambda$ that is not weakly dominated by a combination $\kappa \in K$. If $\eta(F) = 0$, then no such combination of components exists and K spans Λ; if $\eta(F) > 0$, then stochastic spanning does not occur.

Alternatively, the stochastic spanning measure can be reformulated in terms of expected utility:

$$\eta(F) = \sup_{\lambda\in\Lambda;u\in\mathcal{U}_2} \inf_{\kappa\in\mathrm{K}} \mathbb{E}_F\left[u\left(X^{\mathrm{T}}\lambda\right) - u\left(X^{\mathrm{T}}\kappa\right)\right]; \tag{3.6}$$

$$\mathcal{U}_2 := \left\{u \in \mathcal{C}^0 : u(y) = \int_{\underline{x}}^{\overline{x}} w(x)r(y;x)dx \ w \in \mathcal{W}\right\}; \tag{3.7}$$

$$r(y;x) := (y-x)1(y \leq x),\ (x,y) \in \mathcal{X}^2. \tag{3.8}$$

In this formulation, U_2 is a set of normalized, increasing and concave utility functions that are constructed as convex mixtures of elementary Russell and Seo (1989) ramp functions $r(y;x), : x \in X$. Stochastic spanning ($\eta(F) = 0$) occurs if no risk averter ($u \in U_2$) benefits from the enlargement $(\Lambda - K)$. The lower bound represents the potential benefit of the enlargement to a risk-neutral decision maker with utility function $u(y) = (y - \overline{x})$.

4 Empirical Application

4.1 Data

We will use the United Nations Development Program's HDI and its components - health, education, and income indices for 2010, 2011, 2012, 2013, 2014, and 2015. The HDI is obtained as the geometric average of the three sub-indices, where each index is obtained through a normalization procedure by setting minimum and maximum (goalposts) in order to set the values between 0 and 1:

$$\text{Dimension index} = \frac{\text{Actual value - Minimum value}}{\text{Maximum value - Minimum value}}$$

The health sub-index is measured by life expectancy (LE) at birth and the normalized sub-index outcomes are obtained by using minimum and maximum goalposts of 20 and 85 years respectively. Hence, the health index (HI) outcomes of a given country is obtained by using the following normalization procedure $HI = \frac{LE-20}{85-20}$ where LE is the life expectancy at birth for a given country.

The education sub-index is measured by the mean years of schooling (MYS) for adults aged 25 years and the expected years of schooling (EYS) for children of school entering age. The index values for MYS and EYS (MYSI and EYSI respectively) are obtained by using a minimum value of zero and maximum values of 15 and 18 years such as, $MYSI = \frac{MYS-0}{15-0}$ and $EYSI = \frac{EYS-0}{18-0}$, respectively. Then, two indices are combined into an education index (EI) using the arithmetic mean, i.e., $EI = \frac{MYSI+EYSI}{2}$.

The standard of living dimension is measured by gross national income per capita. The minimum and maximum goalposts for gross national income (GNI) per capita are \$100 and \$75,000, respectively. The income index (II) is then calculated using the normalization procedure $II = \frac{\text{ln (GNI per capita) - ln (100)}}{\text{ln (75000)-ln (100)}}$[6].

[6] We use the arithmetic average of the HDI for our analysis since our SD spanning methodology tests

To test for inclusion to the set of HDI dimensions, we use twelve proxies for governance (institutional quality) that are extensively used in the empirical literature. In particular, we use the corruption perceptions index (CPI) from the Transparency International; democracy index from the Polity IV database (which is a combined index of institutionalized democracy and autocracy indices); property rights from the Heritage Foundation (HF hereafter) database (Miller and Kim, 2016) ; the overall economic freedom index, judicial independence, protection of property rights, legal system and property rights, extra bribes/favoritism, and regulation components from the Economic Freedom of the World (EFW) database of the Fraser Institute (FI hereafter) (Gwartney et al., 2016) and the property rights, civil liberties, and their combined overall score from the Freedom House (FH) database. Table 1 lists the details of these proxies, and each proxy's range. Indicators with a high score imply better governance in a given country (i.e., better property rights, higher judicial independence, better democracy, less of corruption, and so on). We standardize each of the institutional proxies to range between 0 and 1 (similar to that of HDI components) with a higher score representing a better institutional quality outcome[7].

Table 2 presents the descriptive statistics for all the normalized sub-components for 2015. Among the sub-components of the HDI, the health index has the highest mean (median) followed by the income and education indices. However, the average achievements in institutional quality components show variation across proxies. For instance, we have three sub-components that measure the enforcement and protection of property rights (i.e., property rights components from the HF, FI, and FH). The average level of protection of property rights with the FH proxy is relatively higher compared to the other proxies from the HF and FI. Also, the

whether inclusion of additional institutional quality proxy leads to spanning or not by using a linear optimization problem. However, there is a little difference in HDI scores of countries when composite HDI scores are calculated with the use of arithmetic or geometric mean (e.g., median difference in composite HDI is only 0.004 in 2015) and different way of calculating final composite index does not affect the qualitative findings of this paper.

[7]For instance, we divide the corruption perceptions index by 100. On the other hand, we divide institutionalized autocracy with 10 and subtract this from 1 where a higher score presents a less autocratic regime.

former proxy displays negative skewness whereas other two proxies display positive skewness suggesting that distributions of these proxies are fairly asymmetric. Another interesting feature of the descriptive statistics is that the original components of the HDI always displayed negative skewness, whereas some of the institutional quality proxies display negative and some others display positive skewness. These features of the data sets justify our model-free approach that uses information beyond the second moment of the distributions in question.

Table 3 presents the correlation coefficients between the sub-components of HDI (i.e., health, education, and income indices) with the other institutional quality proxies in 2015. It has been argued that if the dimensions are highly and positively correlated, any index constructed by using these dimensions would be redundant (see e.g., McGillivray, 2005; Cahill, 2005; Bérenger and Verdier-Chouchane, 2007; Foster et al., 2013 among many others that examine the redundancy of the HDI by using correlation analysis). Since the correlation coefficients between sub-components of the HDI and other institutional quality proxies are relatively low compared to the correlation coefficients between sub-components of the HDI, and also given the variation exhibited by the distributions of the institutional quality proxies (i.e., differences in mean, standard deviation and skewness measures across governance indicators), we can expect that the inclusion of an additional institutional quality proxy to the original set of sub-components of HDI may provide additional welfare enhancing information beyond that provided by the sub-components of the HDI.

4.2 Results

In this subsection, we present our findings on testing the null hypothesis of SD spanning, namely that the inclusion of additional institutional quality proxies to the existing set of HDI components does not result in a welfare improvement. We proceed by first finding the SD efficient weights for the components of the benchmark HDI, health, education, and income (i.e., the implicit weights that lead to the most optimistic welfare measurement, see Pinar et al., 2013). For all the years, we find that the health dimension receives a relatively higher implicit weight suggesting it constitutes a cost-effective way of achieving a higher level of "human development". The education index receives the least weight among the three dimensions, since the resources needed to make improvements in that dimension are relatively higher (see Pinar et al., 2017 for further discussion on this). Once we obtain the SD efficiency weights, we test whether the inclusion of the governance proxies, one by one, leads to spanning. Tables 4, 5, 6, 7, and 8 present the results when we include CPI from the Transparency International, democracy index from Polity IV, property rights from the HF, six components from the FI, and three components from the FH to the sub-components of the HDI for the years between 2010 and 2015, respectively. Our results differ depending on the type of institutional quality proxy used. In particular, we find that for eight of the twelve institutional quality proxies (i.e., CPI, property rights index from HF, judicial independence, property rights, legal system and property rights and bribes components from the FI, civil liberties and property rights and civil liberties components from the FH), we fail to reject spanning for all years. In other words, inclusion of these governance indicators does not lead to welfare improvements irrespective of the weight allocation across sub-components. However, with the polity IV democracy index, economic freedom and regulation indices from the FI, we reject the null hypothesis of spanning for all years. This means that inclusion of these indicators to the component list of the HDI leads to improved welfare outcomes. Finally, with property rights from the FH, we reject spanning in 2013, but fail to reject it in other years.

What does it mean if we reject spanning with the inclusion of a given governance indicator? In that case there exists a combination of the four components (i.e., the three components of HDI and the additional governance component), where the distribution of the augmented HDI scores second-order stochastically dominates the distribution of the HDI without the governance proxy. Hence, the inclusion of the governance indicator results in welfare gains. Table 9 presents the combinations of components for the cases where spanning with the governance indicators are rejected between 2010 and 2015. The results suggest that if one were to include democracy index from Polity IV, and the regulation and economic freedom indices from FI into the component list of HDI, their inclusion would lead to a better welfare outcome by combining some of the sub-components of the HDI with these governance indicators in all years. For instance, if one were to combine the democracy index from Polity IV, with the health and income components in 2015 with weights of 0.12, 0.80, and 0.08, this would second-order dominate any combination of the three original components of the HDI. Similarly, a combination of the health index and economic freedom index from the FI (or the regulation index from the FI) in 2015 with weights of 0.49 and 0.51 (0.54 and 0.46) would lead to a distribution of achievements that second-order dominates the distribution of any combination of the three original components of the HDI. To summarize, when spanning is rejected for a given governance indicator, this would suggest that the inclusion of that governance indicator leads to distributional gains in welfare outcomes. To show these improvements in the welfare distributions, in Figures 1, 2, and 3 we graph the empirical distribution of the equally-weighted HDI scores in 2015 without the governance indicator (i.e., empirical cumulative distribution of the benchmark HDI scores), presented as ECDF_HDI, and the empirical distribution of the composite index scores with the proposed weights in Table 9 augmented with the democracy index from Polity IV, economic freedom and regulation indices from the FI in 2015 (which are presented as ECDF_HDIGOV), respectively. Figure 1 shows that the inclusion of the democracy index to the sub-components of the HDI leads to major improvements in the HDI scores where there is a clear first-order dominance over the benchmark HDI. The empirical distributions of the composite indices obtained with economic freedom and regulation indices lay below the distribution of benchmark HDI at the initial scores but cut the empirical distribution of the benchmark HDI from below at the higher values of HDI. In that case, there is no first-order dominance of the index with economic freedom and regulation indices, but there is a second-order dominance over the empirical distribution of the benchmark HDI. Finally, in 2015, with the remaining nine governance indicators, there is no alternative way of combining the governance indicator with the sub-components of the HDI that would lead to a welfare improvement, implying that these indicators could be disregarded as they do not offer welfare gains.

It is worth noting that as most of the indices that lead to major improvements in welfare outcomes (i.e., no-spanning cases) are obtained with the combination of health and governance indicators, then most countries experience high outcomes in health, democracy, economic freedom and regulation indices. In the next section, we provide a further ranking analysis to examine how the inclusion of the above three indicators leads to these welfare improvements in HDI scores.

4.3 Ranking analysis

In this subsection, we provide a further analysis of countries in 2015 with the inclusion of different governance indicators that lead to additional welfare gains. In particular, when we do the comparisons, we use the composite index outcomes obtained with the equally-weighted HDI and the index outcomes obtained with the weights reported in Table 9 that lead to highest welfare outcomes. It is worth mentioning that the rankings obtained in both cases are highly correlated since the governance indicators are positively correlated with the sub-components of the HDI (see Table 3)[8]. However, the composite welfare outcomes obtained are very different in terms of achievement levels (see e.g., Figures 1, 2, and 3). Furthermore, there exist major rank reversals when these governance indicators are included in the set of components. Table 10 presents the rankings of the top 20 countries using the composite index outcomes obtained with the governance indicators with the weights given in Table 9 and the equally-weighted HDI in 2015. For the rankings obtained with the democracy index, Japan, Italy, Spain, Sweden and Luxembourg moved to the top 10 positions, while they were ranked 14th, 23rd, 24th, 12th and

[8]Spearmann rank correlation coefficients between the benchmark HDI with the composite index obtained with democracy index, economic freedom and regulation indices are 0.90, 0.89, and 0.90, respectively.

15th respectively with the benchmark HDI. Also, with the composite indices obtained with the inclusion of economic freedom and regulation indices, Hong Kong moved to the top ranking with these indices compared to its 11th position with the benchmark HDI.

We observe major rank reversals between the augmented composite indices and the benchmark HDI even though the rankings obtained with the governance indicators are positively and highly correlated with the benchmark HDI rankings. Panels A and B of Table 11 presents the 20 countries that moved to a higher and lower ranking position in 2015 relative to their position based on the benchmark HDI. The major rank reversals that we observe when governance components are included suggest that there are major differences in governance achievements across countries. For instance, Nicaragua would have moved 48, 50, and 43 positions upward with the composite indices obtained with the democracy, economic freedom and regulation indices compared to the benchmark equally-weighted HDI, respectively. On the other hand, Saudi Arabia would have moved 66, 53 and 45 positions downward for the same comparison, respectively. Countries that experience major upward (downward) movements in their rankings with the governance indicators are mainly developing countries that have relatively higher (lower) health and governance outcomes.

Since the composite indices augmented with democracy, economic freedom and regulation indices second-order stochastically dominate the benchmark HDI, that implies that the sum of the shortfalls below a given welfare level is relatively lower with the augmented composite index when compared to the benchmark HDI. To further demonstrate the magnitude of these improvements we also obtain and compare the total shortfalls for the given augmented composite indices and that of the benchmark HDI. Table 12 reports the total shortfalls of these composite indices below a given welfare level. The first column reports the threshold welfare levels, whereas, the remaining three columns present the total sum of shortfalls of composite achievement levels below a given welfare threshold for each of the three augmented composite indices with the total sum of shortfalls of the benchmark HDI in parenthesis for each case. For instance, when we look at the sum of shortfalls with the composite index obtained with the use of democracy index when compared to that for the benchmark HDI in parentheses, these are 0.109, 1.084, 4.473, 10.937, 21.669, and 35.916 (0.957, 4.423, 9.939, 18.641, 31.025, and 46.283) below 0.5, 0.6, 0.7, 0.8, 0.9 and 1, respectively. A similar pattern is observed with the composite index obtained with economic freedom and regulation indices where there are always lower sums of shortfalls below a given level with these indices compared to the total shortfalls with the benchmark HDI achievements.

In this section we examined the inclusion of governance indicators to the list of components in the HDI. At first glance, since the correlation of indices obtained with the governance indicators with the equally-weighted HDI is high and positive, one might have suspected that the inclusion of these governance indicators may have been redundant based on the redundancy literature (see e.g., McGillivray, 2005; Foster et al., 2013). However, when we look at the composite achievement levels in detail, we find that weighting some of the governance indicators and some components of the HDI (mainly health index) differently, results in higher welfare outcomes compared to the benchmark HDI, as the new augmented indices second-order stochastically dominate the benchmark HDI (and also any combination of the three components of HDI). Furthermore, when we compare the rankings obtained with the composite indices that include governance indicators with the equally-weighted HDI, we also observe major rank reversals for most developing countries. This suggests that the inclusion of some of the governance indicators not only leads to marked increase in welfare outcomes but also provides additional variation in the country ranking analysis.

5 Concluding remarks

In this paper we applied SD spanning testing to examine the inclusion of additional indicators to the component list of the HDI. In particular, we tested for the inclusion of twelve governance indicators to the component list of the HDI, since governance is deemed to be a socially and economically important factor for development. The SD spanning tests allowed us to examine whether the inclusion of any governance indicator may provide additional welfare gains to the benchmark HDI index. We compared the empirical distribution of any combination of the three dimensions of the HDI with the empirical cumulative distribution of any combination of sub-components of the HDI and governance indicator and we found that when democracy index from the Polity IV, economic freedom and regulation indices from the FI are combined with some of the sub-components of the HDI (mainly the health index), these indices second-order stochastically dominate the cumulative distribution of any combination of the three components of the benchmark HDI (which also includes the equally-weighted benchmark HDI). This suggests that the inclusion of these governance indicators to measure the well-being across countries leads to marked welfare gains across the countries in our sample.

We further computed country rankings when these three governance indicators are included in the list of the standard benchmark HDI components. We found that even though the rankings obtained with the composite indices that include governance indicators were positively and highly correlated with the rankings obtained with the equally-weighted HDI, most countries exhibited major rank reversals using the two types of composite indices. Furthermore, we also showed that the composite achievements with the governance indicators were relatively higher than those with the equally-weighted HDI.

In the current paper we only examined the inclusion of governance indicators to the list of components of the HDI. Other sets of important factors may be also tested for inclusion to the HDI, such as the set of environmental factors. We leave the SD spanning testing of such additional factors for future research.

6 References

Acemoglu, D., Johnson, S., Robinson, J.A., (2001). The colonial origins of comparative development: an empirical investigation. American Economic Review, 91(5), 1369-1401.

Acemoglu, D., Simon, J., Robinson, J.A., Yared, P., (2008). Income and Democracy. American Economic Review, 98, 808-842.

Acemoglu, D., Simon, J., Robinson, J.A., Yared, P., (2009). Reevaluating the Modernization Hypothesis. Journal of Monetary Economics, 56(8), 1043-1058.

Acemoglu, D., Gallego, F.A., Robinson, J.A., (2014). Institutions, Human Capital, and Development. Annual Review of Economics, 6(1), 875-912.

Alfaro, L., Kalemli-Ozcan, S., Volosovych, V., (2008). Why doesn't capital flow from rich to poor countries? An empirical investigation. Review of Economics and Statistics, 90, 347-368.

Athanassoglou, S., (2015). Multidimensional welfare rankings under weight imprecision: a social choice perspective. Social Choice and Welfare, 44(4), 719-744.

Arvanitis, S., Hallam, M., Post, T., Topaloglou, N. (2018). Stochastic spanning. Journal of Business and Economic Statistics, forthcoming. https://doi.org/10.1080/07350015.2017.1391099

Azemar, C., Desbordes, R., (2013). Has the Lucas Paradox been fully explained? Economics Letters, 121, 183-187.

Barrett, G.F., Donald, S.G., (2003). Consistent tests for stochastic dominance. Econometrica, 71, 71-104.

Barro, R.J., (2015). Convergence and Modernisation. Economic Journal, 125(585), 911-942.

Berenger, V., Verdier-Chouchane, A. (2007). Multidimensional measures of well-being: standard of living and quality of life across countries. World Development, 35(7), 1259-1276.

Boarini, R., D'Ercole, M.M., (2013). Going beyond GDP: An OECD Perspective*. Fiscal Studies, 34, 289-314.

Borensztein, E., De Gregorio, J., Lee, J-W. (1998). How does foreign direct investment affect economic growth? Journal of International Economics, 45(1), 115-135.

Bosker, M., Garretsen, H., (2009). Economic development and geography of institutions. Journal of Economic Geography, 9, 295-328.

Bourguignon, F., Platteau, J.P., (2013). Aid effectiveness revisited: The trade-off between needs and governance. Working paper. Paris School of Economics (PSE) and Centre for Research in Economic Development (CRED), University of Namur.

Bravo, G., (2014). The Human Sustainable Development Index: New calculations and a first critical analysis. Ecological Indicators, 37, 145-150.

Burnside, C., Dollar, D. (2004). Aid, policies, and growth: Revisiting the evidence. Washington, DC: World Bank, Policy research working paper 3251.

Busse, M., Hefeker, C. (2007). Political risk, institutions and foreign direct investment. European Journal of Political Economy, 23(2), 397-415.

Cahill M.B. (2005), Is the Human Development Index Redundant? Eastern Economic Journal, 31(1), 1-5.

Cherchye, L., Ooghe, E., & van Puyenbroeck, T. (2008). Robust human development rankings. Journal of Economic Inequality, 6(4), 287-321.

Chong, A., Gradstein, M., Calderon, C. (2009). Can foreign aid reduce income inequality and poverty? Public Choice, 140(1), 59-84.

Collier, P., Dollar, D. (2002). Aid allocation and poverty reduction. European Economic Review, 46(8), 1475-1500.

Davidson, R., Duclos, J.-Y., (2000). Statistical inference for stochastic dominance and for the measurement of poverty and inequality. Econometrica, 68, 1435-1464.

Decancq, K., Fleurbaey, M., Schokkaert, E., (2015). Inequality, income, and well-being. In A.B. Atkinson & F. Bourguignon (Eds.), Handbook of income distribution (pp. 67-140).

Amsterdam: Elsevier.

Decancq, K., Lugo, M.A., (2013). Weights in Multidimensional Indices of Wellbeing: An Overview. Econometric Reviews, 32(1), 7-34.

Decancq, K., Schokkaert, E., (2016). Beyond GDP: Using equivalent incomes to measure well-being in Europe. Social Indicators Research, 126, 21-55.

Dollar, D., Levin, V. (2006). The Increasing Selectivity of Foreign Aid, 1984–2003. World Development, 34(12), 2034-2046.

Duclos, J.-Y., Sahn, D.E., Younger, S.D., (2006). Robust multidimensional poverty comparisons. Economic Journal, 116(514), 943-968.

Durand, M. (2015). The OECD Better Life Initiative: How's Life? and the Measurement of Well-being. Review of Income and Wealth, 61(1), 4-17.

Easterly, W., Levine, R., (2003). Tropics, germs, and crops: how endowments influence economic development. Journal of Monetary Economics, 50, 3-39.

Esty, D.C., Levy, M., Srebotnjak, T., de Sherbinin, A., (2005). 2005 Environmental sustainability index: benchmarking national environmental stewardship. Yale Center for Environmental Law & Policy, New Haven.

Fleurbaey, M., (2009). Beyond the GDP: The quest for a measure of social welfare. Journal of Economic Literature, 47(4), 1029-1075.

Fleurbaey, M., Blanchet, D., (2013). Beyond GDP: Measuring Welfare and Assessing Sustainability. Oxford University Press, New York.

Foster, J.E., McGillivray, M., Seth, S., (2013). Composite indices: Rank robustness, statistical association, and redundancy. Econometric Reviews, 32(1), 35-56.

Guillaumont, P., Guillaumont Jeanneney, S., Wagner, L., (2017). How to Take into Account Vulnerability in Aid Allocation Criteria and Lack of Human Capital as Well: Improving the Performance Based Allocation. World Development, 90, 27-40.

Gwartney, J., Lawson, R., Hall, J., 2016. Economic Freedom of the World: 2016 Annual Report. Fraser Institute.

Hall, J., Giovannini, E., Morrone, A., Ranuzzi, G., (2010). A framework to measure the progress of societies, OECD Statistics Working Papers 2010/5, OECD Publishing.

Haskel, J.E., Pereira, S.C., Slaughter, M.J. (2007). Does inward foreign direct investment boost the productivity of domestic firms? Review of Economics and Statistics, 89(3), 482-496.

Huberman, G., Kandel, S., (1987), Mean-Variance Spanning. Journal of Finance, 42, 873-888.

Kosack, S., (2003). Effective Aid: How Democracy Allows Development Aid to Improve the Quality of Life. World Development, 31(1), 1-22.

Kuosmanen, T., (2004). Efficient diversification according to stochastic dominance criteria. Management Science, 50, 1390-1406.

Levy, H. (2015). Stochastic dominance: Investment decision making under uncertainty. Springer.

Linton, O., Maasoumi, E., Whang, Y.-J., (2005). Consistent testing for stochastic dominance under general sampling schemes. Review of Economic Studies, 72(3), 735-765.

Linton, O., Post, T., Whang, Y.-J., (2014). Testing for the Stochastic Dominance Efficiency of a Given Portfolio. Econometrics Journal, 17, 59-74.

Lipset, S.M. (1959). Some social requisites of democracy: economic development and political legitimacy. American Political Science Review, 53(1), 69-105.

Malik, K., (2013). Human development report 2013: The rise of the global south - human progress in a diverse world. United Nations, New York.

McGillivray, M. (2005). Measuring non-economic well-being achievement. The Review of Income and Wealth, 51(2), 337-364.

Miller, T., Kim, A.B., (2016). 2016 Index of economic freedom, Institute for Economic Freedom and Opportunity, The Heritage Foundation.

Ness, B., Urbel-Piirsalu, E., Anderberg, S., Olsson, L., (2007). Categorising tools for sustainability assessment. Ecological Economics, 60, 498-508.

Noorbakhsh, F., (1998). A modified human development index. World Development, 26, 517-528.

OECD, (2011). How's Life? Measuring Well-being. OECD Publishing, Paris. Available at: http://dx.doi.org/10.1787/9789264121164-en.

Papaioannou, E., (2009). What drives international financial flows? Politics, institutions and other determinants. Journal of Development Economics, 88, 269-281.

Permanyer, I., (2011). Assessing the robustness of composite indices rankings. Review of Income and Wealth, 57(2), 306-326.

Pinar, M. (2015). Measuring world governance: revisiting the institutions hypothesis. Empirical Economics, 48(2), 747-778.

Pinar, M., Cruciani, C., Giove, S., Sostero, M., (2014). Constructing the FEEM sustainability index: A Choquet integral application. Ecological Indicators, 39, 189-202.

Pinar, M., Stengos, T., Topaloglou, N., (2013). Measuring human development: A stochastic dominance approach. Journal of Economic Growth, 18(1), 69-108.

Pinar, M., Stengos, T., Topaloglou, N., (2017). Testing for the implicit weights of the dimensions of the Human Development Index using stochastic dominance. Economics Letters, 161, 38-42.

Pollesch, N.L., Dale, V.H., (2016). Normalization in sustainability assessment: Methods and Implications. Ecological Indicators, 130, 195-208.

Post, T., (2003). Empirical Tests for Stochastic Dominance Efficiency. Journal of Finance, 58, 1905-1932.

Post, T., Poti, V. (2017). Portfolio analysis using stochastic dominance, relative entropy and empirical likelihood. Management Science, 63, 153-165.

Post, T., Versijp, P. (2007). Multivariate tests for stochastic dominance efficiency of a given portfolio. Journal of Financial and Quantitative Analysis, 42, 489-516.

Ranis, G., Frances, S., Samman, E., (2006). Human Development: Beyond the Human Development Index. Journal of Human Development and Capabilities, 7, 323-358.

Robeyns, I., (2005). The Capability Approach: a theoretical survey, Journal of Human

Development, 6(1), 93-117.

Rodrik, D., Subramanian, A., Trebbi, F., (2004) Institutions rule: the primacy of institutions over geography and integration in economic development. Journal of Economic Growth, 9, 131-165.

Roman, D., Darby-Dowman, K. and Mitra, G., (2006). Portfolio construction based on stochastic dominance and target return distributions. Mathematical Programming, 108, 541-569.

Roodman, D., (2012). An Index of Donor Performance. Working Paper 67, Center for Global Development, Washington, DC.

Scaillet, O., Topaloglou, N., (2010). Testing for stochastic dominance efficiency. Journal of Business and Economic Statistics, 28(1), 169-180.

Sen, A. (1985) Commodities and Capabilities, North Holland, Amsterdam.

Sen, A. (1987) The standard of living. In G. Hawthorn (ed.), The Standard of Living, Cambridge University Press, Cambridge.

Sen, A. (1999) Development as Freedom, Knopf, New York.

Singh, R.K., Murty, H.R., Gupta, S.K., Dikshit, A.K., (2012). An overview of sustainability assessment methodologies. Ecological Indicators, 15, 281-299.

Stiglitz J.E., Sen, A., Fitoussi J.-P., (2009). Report by the commission on the measurement of economic performance and social progress. Available at: http://www.stiglitz-sen-fitoussi.fr/en/index.htm.

Tebaldi, E., Mohan, R., (2010). Institutions and Poverty. The Journal of Development Studies, 46(6), 1047-1066.

Tofallis, C., (2013). An automatic-democratic approach to weight setting for the new human development index. Journal of Population Economics, 26(4), 1325-1345.

UN System Task Team on the Post-2015 UN Development Agenda (2012a). Realizing the Future We Want for All Report to the Secretary-General. Available via: http://www.un.org/en/developmen

UN System Task Team on the Post-2015 UN Development Agenda (2012b). Development and development. Available via: http://www.un.org/millenniumgoals/pdf/Think%20Pieces/7_governance

World Bank (2003). Supporting sound policies with adequate and appropriate financing: implementing the Monterrey consensus at the country level. World Bank, Washington, DC.

Yaari, M.E., (1987). The dual theory of choice under risk. Econometrica 55, 95-115.

Appendix A: Statistical Theory

We use the test statistic developed in Arvatinis et al. (2018). Let $q(\eta_\infty, 1-\alpha)$ denote the $(1-\alpha)$ quantile of the distribution of η_∞ for any significance level $\alpha \in]0,1[$. The basic decision rule to reject $\mathbf{H}_0$ against $\mathbf{H}_1$ if and only if $\eta_T > q(\eta_\infty, 1-\alpha)$ is infeasible due to the dependence of $q(\eta_\infty, 1-\alpha)$ on the latent cdf of F. However, feasible decision rules can be obtained by using a subsampling procedure to estimate $q(\eta_\infty, 1-\alpha)$ from the data.

To implement the subsampling procedure, we begin by generating $(N - b_N + 1)$ maximally overlapping subsamples of $b_N \in \mathbb{N}_1$ consecutive observations, $s_{b_N;N,n} := (X_s)_{s=n}^{n+b_N-1}$, $n = 1, \cdots, N - b_N + 1$, and compute test scores $\eta_{b_N;N,n} = \sqrt{b_N}\eta(F_{b_N;N,n})$ for each subsample, where $F_{b_N;N,n}$ denotes the empirical joint cdf constructed from $s_{b_N;N,n}$, $n = 1, \cdots, N - b_N + 1$. The distribution of subsample test scores can be described by the following cdf and quantile function:

$$S_{N,b_N}(y) := \frac{1}{N - b_N + 1} \sum_{n=1}^{N-b_N+1} 1(\eta_{b_N;N,n} \leq y); \tag{6.1}$$

$$q_{N,b_N}(1-\alpha) := \inf_y \left\{ y : S_{N,b_N}(y) \geq 1 - \alpha \right\}. \tag{6.2}$$

Our decision rule is to reject the null $\mathbf{H}_0 : \eta(F) = 0$ against the alternative $\mathbf{H}_1 : \eta(F) > 0$ at a significance level of $\alpha \in]0,1[$ if and only if $\eta_N > q_{N,b_N}(1-\alpha)$, or, equivalently, $1 - S_{N,b_N}(\eta_N) < \alpha$.

This subsampling routine is asymptotically exact and consistent under reasonable assumptions on the subsample length and significance level.

Although the test has asymptotically correct size, simulation exercises show that the quantile estimates $q_{N,b_N}(1-\alpha)$ may be biased and sensitive to the subsample size b_N in finite samples of realistic dimensions (M). To correct for small-sample bias and reduce the sensitivity to the choice of b_N, we propose a regression-based bias-correction method that is motivated by our observations from simulation exercises. For a given significance level α, we compute the quantiles $q_{N,b_N}(1-\alpha)$ for a reasonable range of the subsample size b_N. Next, we estimate the intercept and slope of the following regression line using OLS regression analysis:

$$q_{N,b_N}(1-\alpha) = \gamma_{0;N,1-\alpha} + \gamma_{1;N,1-\alpha}(b_N)^{-1} + \nu_{N;1-\alpha,b_N}. \tag{6.3}$$

Finally, we estimate the bias-corrected $(1-\alpha)$-quantile as the OLS predicted value for $b_N = N$:

$$q_N^{BC}(1-\alpha) := \hat{\gamma}_{0;N,1-\alpha} + \hat{\gamma}_{1;N,1-\alpha}(N)^{-1}. \tag{6.4}$$

Since $q_{N,b_N}(1-\alpha)$ converges in probability to $q(\eta_\infty, 1-\alpha)$ and $(b_N)^{-1}$ converges to zero as $N \to 0$, $\hat{\gamma}_{0;N,1-\alpha}$ converges in probability to $q(\eta_\infty, 1-\alpha)$ and the asymptotic properties are not affected. However, computational experiments show that the bias-corrected method is more efficient and more powerful in small samples.

Appendix B: Computational Strategy

According to Arvanitis et al. (2018), the test statistic can be written:

$$\eta_N = \sqrt{N} \sup_{u \in \mathcal{U}_2} \left(\sup_{\lambda \in \Lambda} \mathbb{E}_{F_N} \left[u \left(X^{\mathrm{T}} \lambda \right) \right] - \sup_{\kappa \in \mathrm{K}} \mathbb{E}_{F_N} \left[u \left(X^{\mathrm{T}} \kappa \right) \right] \right). \tag{6.5}$$

The term in parentheses is the difference between the solutions to two standard convex optimization problems of maximizing a quasi-concave objective function over a polyhedral feasible set. The analytic complexity of computing η_N stems from the search over all admissible utility functions ($\mathcal{U}_2$). However, the utility functions are univariate, normalized, and have a bounded domain ($\mathcal{X}$). As a result, we can approximate $\mathcal{U}_2$ with arbitrary accuracy using a finite set of increasing and concave piecewise-linear functions in the following way.

We partition $\mathcal{X}$ into N_1 equally spaced values as $\underline{x} = z_1 < \cdots < z_{N_1} = \overline{x}$, where $z_n := \underline{x} + \frac{n-1}{N_1-1}(\overline{x} - \underline{x})$, $n = 1, \cdots, N_1$; $N_1 \geq 2$. Instead of an equal spacing, the partition could also be based on percentiles of the welfare distribution. Similarly, we partition the interval $[0, 1]$, as $0 < \frac{1}{N_2-1} < \cdots < \frac{N_2-2}{N_2-1} < 1$, $N_2 \geq 2$. Using this partition, let

$$\underline{\eta_N} := \sqrt{N} \sup_{u \in \underline{\mathcal{U}_2}} \left(\sup_{\lambda \in \Lambda} \mathbb{E}_{F_N} \left[u \left(X^{\mathrm{T}} \lambda \right) \right] - \sup_{\kappa \in \mathrm{K}} \mathbb{E}_{F_N} \left[u \left(X^{\mathrm{T}} \kappa \right) \right] \right); \tag{6.6}$$

$$\underline{\mathcal{U}_2} := \left\{ u \in \mathcal{C}^0 : u(y) = \sum_{n=1}^{N_1} w_n r(y; z_n) : \boldsymbol{w} \in \mathrm{W} \right\}; \tag{6.7}$$

$$\mathrm{W} := \left\{ \boldsymbol{w} \in \left\{ 0, \frac{1}{N_2 - 1}, \cdots, \frac{N_2 - 2}{N_2 - 1}, 1 \right\}^{N_1} : \sum_{n=1}^{N_1} w_n = 1 \right\}. \tag{6.8}$$

Every element $u \in \underline{\mathcal{U}_2}$ consists of at most N_2 linear line segments with knots at N_1 possible outcome levels. Clearly, $\underline{\mathcal{U}_2} \subset \mathcal{U}_2$ and $\underline{\eta_N}$ approximates η_N from below as we refine the partition ($N_1, N_2 \to \infty$). The appealing feature of $\underline{\eta_N}$ is that we can enumerate all $N_3 := \frac{1}{(N_1-1)!} \prod_{i=1}^{N_1-1} (N_2 + i - 1)$ elements of $\underline{\mathcal{U}_2}$ for a given partition, and, for every $u \in \underline{\mathcal{U}_2}$, solve the two embedded maximization problems in (7.2) using linear programming (LP):

$$c_{0,n} := \sum_{m=n}^{N_1} \left(c_{1,m+1} - c_{1,m}\right) z_m; \tag{6.9}$$

$$c_{1,n} := \sum_{m=n}^{N_1} w_m; \tag{6.10}$$

$$\mathcal{N} := \{n = 1, \cdots, N_1 : w_n > 0\} \bigcup \{N_1\}. \tag{6.11}$$

For any given $u \in \underline{\mathcal{U}_2}$, $\sup_{\lambda\in\Lambda} \mathbb{E}_{F_N} \left[u\left(X^{\mathrm{T}}\lambda\right)\right]$ is the optimal value of the objective function of the following LP problem in canonical form:

$$\max N^{-1} \sum_{n=1}^{N} y_n \tag{6.12}$$

$$\text{s.t.} : y_n - c_{1,n} X_n^{\mathrm{T}} \lambda \leq c_{0,n}, : n = 1, \cdots, N; n \in \mathcal{N};$$

$$\sum_{i=1}^{M} \lambda_i = 1;$$

$$\lambda_i \geq 0, : i = 1, \cdots, M;$$

$$y_n : \text{free}, : n = 1, \cdots, N.$$

The LP problem always has a feasible and finite solution and has $\mathcal{O}(N+M)$ variables (where N and M represent number of countries and dimensions) and constraints, making it small for typical data dimensions. Our application in Section 4 is based on the number of countries and dimensions (e.g., $M = 4$, $N = 188$), and uses small LP problems, which is perfectly manageable with modern-day computer hardware and solver software.

The total run time of all computations for our application amounts to several working days on a standard desktop PC with a 2.93 GHz quad-core Intel i7 processor, 16GB of RAM and using MATLAB with the external Gurobi Optimizer solver.

Table 1. List of governance indicators			
Governance measure	Source	Range	Reference
Corruption perceptions index	Transparency International	0-100	https://www.transparency.org/cpi
Democracy index	Polity IV	-10 - +10	http://www.systemicpeace.org/inscrdata.html
Property rights	The Heritage Foundation	0-10	Miller and Kim (2016) http://www.heritage.org/index/
Economic Freedom	Fraser Institute	0-10	Gwartney et al. (2016) https://www.fraserinstitute.org/economic-freedom/dataset
Judicial independence	Fraser Institute	0-10	Gwartney et al. (2016) https://www.fraserinstitute.org/economic-freedom/dataset
Protection of property rights	Fraser Institute	0-10	Gwartney et al. (2016) https://www.fraserinstitute.org/economic-freedom/dataset
Legal System & Property Rights	Fraser Institute	0-10	Gwartney et al. (2016) https://www.fraserinstitute.org/economic-freedom/dataset
Extra payments/bribes/favouritism	Fraser Institute	0-10	Gwartney et al. (2016) https://www.fraserinstitute.org/economic-freedom/dataset
Regulation	Fraser Institute	0-10	Gwartney et al. (2016) https://www.fraserinstitute.org/economic-freedom/dataset
Property rights	Freedom house	0-40	https://freedomhouse.org/report/freedom-world/freedom-world-2017
Civil liberties	Freedom house	0-60	https://freedomhouse.org/report/freedom-world/freedom-world-2017
Overall score (Property rights + Civil liberties)	Freedom house	0-100	https://freedomhouse.org/report/freedom-world/freedom-world-2017

Table 2. Summary statistics					
	Mean	Median	Standard deviation	Skewness	Obs.
Health index	0.790	0.822	0.128	-0.668	188
Education index	0.639	0.659	0.174	-0.344	188
Income index	0.687	0.702	0.180	-0.289	188
CPI	0.430	0.370	0.201	0.790	162
Polity IV – Democracy index	0.718	0.850	0.307	-1.008	156
HF - Property rights	0.422	0.350	0.250	0.654	177
FI - Economic Freedom	0.679	0.686	0.091	-0.675	158
FI – Judicial independence	0.501	0.475	0.211	0.312	151
FI - Property rights	0.560	0.536	0.166	0.281	151
FI - Legal system & Property rights	0.525	0.508	0.156	0.415	158
FI - Extra payments/bribes/favouritism	0.445	0.411	0.177	0.782	150
FI - Regulation	0.699	0.703	0.108	-0.689	158
FH - Property rights	0.601	0.675	0.320	-0.392	188
FH - Civil liberties	0.604	0.617	0.274	-0.223	188
FH - Property rights & Civil liberties	0.603	0.635	0.290	-0.292	188

Table 3. Correlation coefficients between governance proxies and sub-components of the HDI				
	Health	Education	Income	Obs.
Health	1.000***	0.806***	0.795***	188
Education	0.806***	1.000***	0.839***	188
Income	0.795***	0.839***	1.000***	188
CPI	0.674***	0.705***	0.717***	162
Polity IV - Democracy	0.300***	0.333***	0.179**	156
HF - Property rights	0.624***	0.636***	0.662***	177
FI - Economic Freedom	0.580***	0.593***	0.559***	158
FI - Judicial	0.486***	0.511***	0.570***	151
FI - Property rights	0.528***	0.539***	0.599***	151
FI - Legal system & Property rights	0.654***	0.713***	0.691***	158
FI - Extra payments/bribes/favouritism	0.565***	0.537***	0.604***	150
FI - Regulation	0.423***	0.506***	0.506***	158
FH - Property rights	0.487***	0.526***	0.395***	188
FH - Civil liberties	0.536***	0.593***	0.469***	188
FH - Property rights & Civil liberties	0.520***	0.569***	0.441***	188

Table 4. SD spanning test for the inclusion of the CPI to the HDI				
Year	Test Statistic	Critical value (95%)	Number of countries	Result
2015	0.000	0.000	162	Spanning
2014	0.000	0.000	170	Spanning
2013	0.000	0.000	172	Spanning
2012	0.000	0.000	171	Spanning
2011	0.000	0.000	178	Spanning
2010	0.000	0.000	174	Spanning

Table 5. SD spanning tests for the inclusion of democracy index of the Polity IV to the HDI				
Year	Test Statistic	Critical value (95%)	Number of countries	Result
2015	0.007	0.004	156	Reject Spanning
2014	0.046	0.020	157	Reject Spanning
2013	0.067	0.025	156	Reject Spanning
2012	0.005	0.004	156	Reject Spanning
2011	0.011	0.007	158	Reject Spanning
2010	0.033	0.029	157	Reject Spanning

Table 6. SD spanning tests for the inclusion of property rights of the HF to the HDI				
Year	Test Statistic	Critical value (95%)	Number of countries	Result
2015	0.000	0.000	177	Spanning
2014	0.000	0.000	177	Spanning
2013	0.000	0.000	176	Spanning
2012	0.000	0.000	176	Spanning
2011	0.000	0.000	176	Spanning
2010	0.000	0.000	176	Spanning

Table 7. SD spanning tests for the inclusion of sub-components of the FI to the HDI				
Panel A. SD spanning tests for inclusion of the economic freedom component				
Year	Test Statistic	Critical value (95%)	Number of countries	Result
2015	0.032	0.031	158	Reject Spanning
2014	0.057	0.055	158	Reject Spanning
2013	0.062	0.057	156	Reject Spanning
2012	0.074	0.069	152	Reject Spanning
2011	0.532	0.073	152	Reject Spanning
2010	0.074	0.073	152	Reject Spanning
Panel B. SD spanning tests for inclusion of the judicial independence component				
2015	0.000	0.000	151	Spanning
2014	0.000	0.000	150	Spanning
2013	0.000	0.000	148	Spanning
2012	0.000	0.000	143	Spanning
2011	0.000	0.000	142	Spanning
2010	0.000	0.000	140	Spanning
Panel C. SD spanning tests for inclusion of the property rights component				
2015	0.000	0.000	151	Spanning
2014	0.000	0.000	149	Spanning
2013	0.000	0.000	150	Spanning
2012	0.000	0.000	143	Spanning
2011	0.000	0.000	142	Spanning
2010	0.000	0.000	140	Spanning
Panel D. SD spanning tests for inclusion of the legal system and property rights component				
2015	0.000	0.000	158	Spanning
2014	0.000	0.000	158	Spanning
2013	0.000	0.000	156	Spanning
2012	0.000	0.000	152	Spanning
2011	0.000	0.000	152	Spanning
2010	0.000	0.000	152	Spanning
Panel E. SD spanning tests for inclusion of the extra payments/bribes/favouritism component				
2015	0.000	0.000	150	Spanning
2014	0.000	0.000	145	Spanning
2013	0.000	0.000	148	Spanning
2012	0.000	0.000	143	Spanning
2011	0.000	0.000	142	Spanning
2010	0.000	0.000	140	Spanning
Panel E. SD spanning tests for inclusion of the regulation component				
2015	0.049	0.039	158	Reject Spanning
2014	0.089	0.087	158	Reject Spanning
2013	0.088	0.085	156	Reject Spanning
2012	0.107	0.100	152	Reject Spanning
2011	0.123	0.109	152	Reject Spanning
2010	0.144	0.118	152	Reject Spanning

Table 8. SD spanning tests for the inclusion of sub-components of the FH to the HDI				
Panel A. SD spanning tests for the inclusion of the property rights component				
Year	Test Statistic	Critical value (95%)	Number of countries	Result
2015	0.000	0.008	188	Spanning
2014	0.000	0.011	188	Spanning
2013	0.005	0.004	188	Reject Spanning
2012	0.000	0.012	188	Spanning
2011	0.000	0.010	187	Spanning
2010	0.001	0.011	187	Spanning
Panel B. SD spanning tests for the inclusion of the civil liberties component				
2015	0.000	0.008	188	Spanning
2014	0.000	0.009	188	Spanning
2013	0.000	0.005	188	Spanning
2012	0.000	0.013	188	Spanning
2011	0.001	0.013	187	Spanning
2010	0.001	0.011	187	Spanning
Panel C. SD spanning tests for the inclusion of the property rights & civil liberties component				
2015	0.000	0.009	188	Spanning
2014	0.000	0.011	188	Spanning
2013	0.000	0.005	188	Spanning
2012	0.000	0.013	188	Spanning
2011	0.001	0.012	187	Spanning
2010	0.001	0.013	187	Spanning

Table 9. Combinations of components where the spanning is rejected between 2010 and 2015					
Year	Included governance proxy	Health	Education	Income	Governance
2015	Democracy index from Polity IV	0.80	0.00	0.08	0.12
2015	Economic Freedom of the FI	0.49	0.00	0.00	0.51
2015	Regulation index of the FI	0.54	0.00	0.00	0.46
2014	Democracy index from Polity IV	0.76	0.00	0.12	0.12
2014	Economic Freedom of the FI	0.46	0.00	0.00	0.54
2014	Regulation index of the FI	0.44	0.00	0.00	0.56
2013	Democracy index from Polity IV	0.80	0.00	0.08	0.12
2013	Economic Freedom of the FI	0.48	0.00	0.00	0.52
2013	Regulation index of the FI	0.44	0.00	0.00	0.56
2013	Property rights from the FH	0.90	0.00	0.00	0.10
2012	Democracy index from Polity IV	0.84	0.00	0.02	0.14
2012	Economic Freedom of the FI	0.39	0.00	0.00	0.61
2012	Regulation index of the FI	0.38	0.00	0.00	0.62
2011	Democracy index from Polity IV	0.89	0.00	0.00	0.11
2011	Economic Freedom of the FI	0.31	0.00	0.00	0.69
2011	Regulation index of the FI	0.41	0.00	0.00	0.59
2010	Democracy index from Polity IV	0.89	0.00	0.00	0.11
2010	Economic Freedom of the FI	0.30	0.00	0.00	0.70
2010	Regulation index of the FI	0.35	0.00	0.00	0.65

Table 10. Country rankings with composite indices obtained with governance indicators and benchmark HDI in 2015					
Country	Rankings with democracy index (benchmark HDI)	Country	Rankings with economic freedom (benchmark HDI)	Country	Rankings with regulation index (benchmark HDI)
Japan	1 (14)	Hong Kong	1 (11)	Hong Kong	1 (11)
Switzerland	2 (2)	Singapore	2 (4)	New Zealand	2 (13)
Italy	3 (23)	Switzerland	3 (2)	Singapore	3 (4)
Australia	4 (3)	New Zealand	4 (13)	Switzerland	4 (2)
Spain	5 (24)	Australia	5 (3)	Canada	5 (10)
Sweden	6 (12)	Ireland	6 (8)	Australia	6 (3)
Luxembourg	7 (15)	Canada	7 (10)	Ireland	7 (8)
Canada	8 (9)	United Kingdom	8 (15)	Iceland	8 (9)
Norway	9 (1)	Chile	9 (35)	Sweden	9 (14)
Netherlands	10 (6)	Japan	10 (16)	Japan	10 (16)
New Zealand	11 (11)	Sweden	11 (14)	United States	11 (12)
France	12 (18)	Netherlands	12 (6)	Denmark	12 (6)
Austria	13 (21)	Spain	13 (26)	Netherlands	13 (7)
Chile	14 (31)	Norway	14 (1)	United Kingdom	14 (15)
Germany	15 (5)	Luxembourg	15 (17)	Belgium	15 (21)
Ireland	16 (8)	Finland	16 (22)	Luxembourg	16 (17)
Korea, Republic	17 (16)	Austria	17 (23)	Malta	17 (32)
Finland	18 (20)	Israel	18 (19)	Norway	18 (1)
United Kingdom	19 (13)	Korea, Republic	19 (18)	Germany	19 (5)
Portugal	20 (35)	Germany	20 (5)	Qatar	20 (31)

Table 11. Major rank reversals with the composite indices obtained with governance indicators

Panel A. Countries that experienced largest upward movements in their rankings with the composite indices with governance indicators

Country	Change in ranking with democracy index	Country	Change in ranking with economic freedom index	Country	Change in ranking with regulation index
Nicaragua	48	Nicaragua	50	Rwanda	56
Cape Verde	41	Guatemala	48	Nicaragua	43
Honduras	37	Honduras	47	Jamaica	42
Jamaica	33	Rwanda	43	Macedonia	40
Guatemala	31	Jamaica	35	Fiji	39
Niger	31	Cambodia	35	Bhutan	39
Peru	28	Peru	32	Honduras	36
El Salvador	28	Georgia	31	Niger	34
Lebanon	27	Albania	31	Jordan	32
Senegal	27	Armenia	30	Timor-Leste	29
Costa Rica	26	El Salvador	30	Malaysia	28
Albania	26	Bhutan	30	Cambodia	27
Tunisia	26	Gambia	30	Uganda	26
Mexico	25	Costa Rica	29	Bangladesh	25
Macedonia	25	Chile	26	Belize	24
Paraguay	25	Liberia	26	Vietnam	24
Solomon Islands	24	Jordan	25	Georgia	23
Dominican Rep.	22	Cape Verde	25	Cape Verde	23
Malawi	22	Mauritius	24	Burundi	23
Nepal	21	Dominican Rep.	24	Peru	22

Panel B. Countries that experienced largest downward movements in their rankings with the composite indices with governance indicators

Country	Change in ranking with democracy index	Country	Change in ranking with economic freedom index	Country	Change in ranking with regulation index
Saudi Arabia	66	Venezuela	82	Venezuela	83
Kazakhstan	60	Argentina	70	Brazil	59
Belarus	59	Saudi Arabia	53	Argentina	55
Qatar	45	Russia	49	Russian	55
Azerbaijan	42	Iran	45	Iran	48
UAE	41	Algeria	45	Saudi Arabia	45
Bahrain	39	Ukraine	44	Trinidad & Tobago	38
Kuwait	39	Libya	39	Bolivia	38
Russia	39	Trinidad & Tobago	38	Algeria	35
Equatorial Guinea	38	Gabon	38	Libya	32
Turkmenistan	37	Brazil	33	Egypt	29
Swaziland	35	Congo, Republic	32	Congo, Republic	29
Uzbekistan	34	Kuwait	31	Ukraine	28
Oman	31	Azerbaijan	30	Greece	26
Angola	31	South Africa	30	Gabon	26
Iran	30	Greece	29	Angola	25
South Africa	29	Kazakhstan	28	Barbados	24
Singapore	25	Angola	28	South Africa	24
Gabon	25	Brunei	25	Ecuador	22
Cameroon	22	Barbados	24	Zimbabwe	22

Table 12. Sum of shortfalls below a given threshold welfare level with the composite indices that have governance indicator and benchmark HDI			
Level of composite index	Sum of shortfalls below a given threshold composite index with democracy index (benchmark HDI)	Sum of shortfalls below a given threshold with composite index with economic freedom index (benchmark HDI)	Sum of shortfalls below a given threshold with composite index with regulation index (benchmark HDI)
0.5	0.109 (0.957)	0.027 (1.030)	0.035 (1.030)
0.6	1.084 (4.423)	0.808 (4.278)	0.639 (4.278)
0.7	4.473 (9.939)	4.281 (9.459)	3.628 (9.459)
0.8	10.937 (18.641)	12.403 (17.787)	11.082 (17.787)
0.9	21.669 (31.025)	26.138 (30.040)	23.863 (30.040)
1.0	35.916 (46.283)	41.865 (45.456)	39.483 (45.456)
Number of total observations	156	158	158
Notes: Each column offers the total sum of shortfalls of composite achievement levels below a given welfare threshold with the composite indices with governance indicator where the same information is provided within parenthesis for the benchmark HDI.			

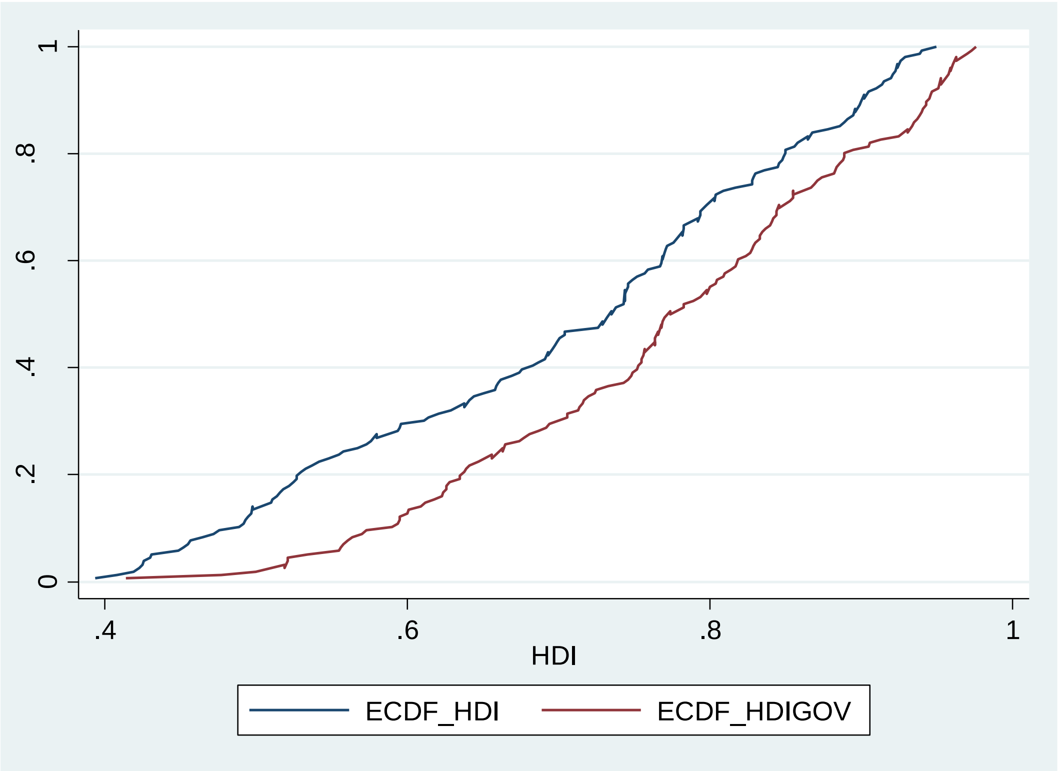

Figure 1. Empirical cumulative distribution of the benchmark HDI and HDI scores with democracy index from the Polity IV (ECDF_HDI and ECDF_HDIGOV, respectively).

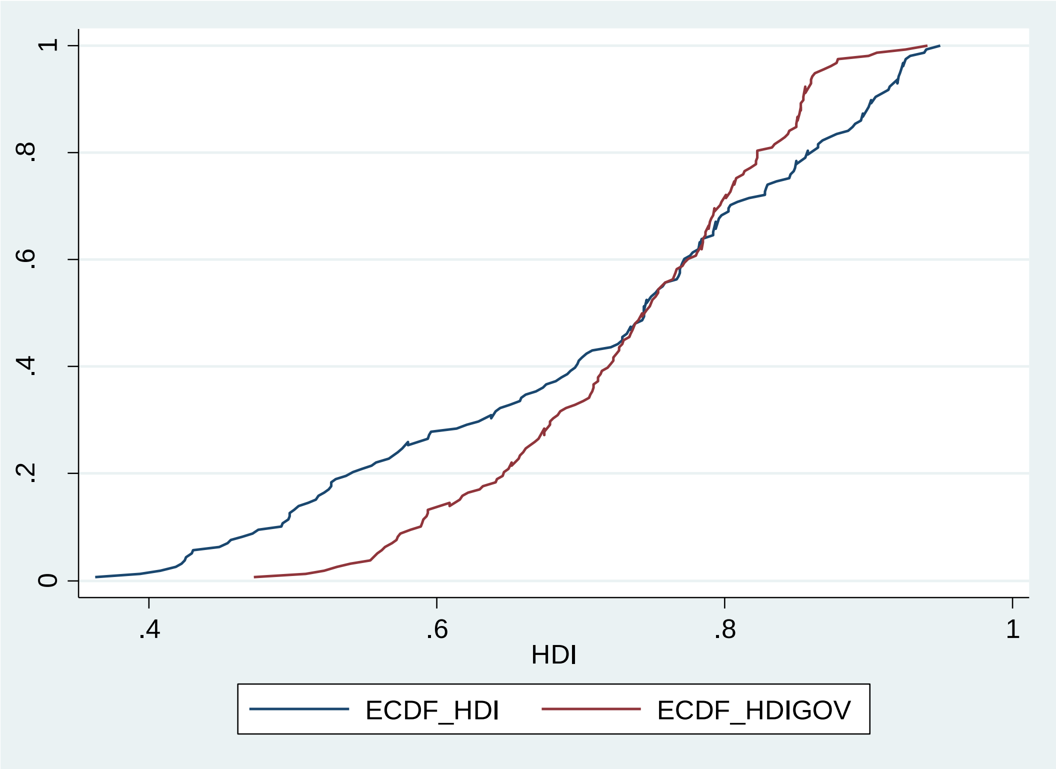

Figure 2. Empirical cumulative distribution of the benchmark HDI and HDI scores with economic freedom index from FI (ECDF_HDI and ECDF_HDIGOV, respectively).

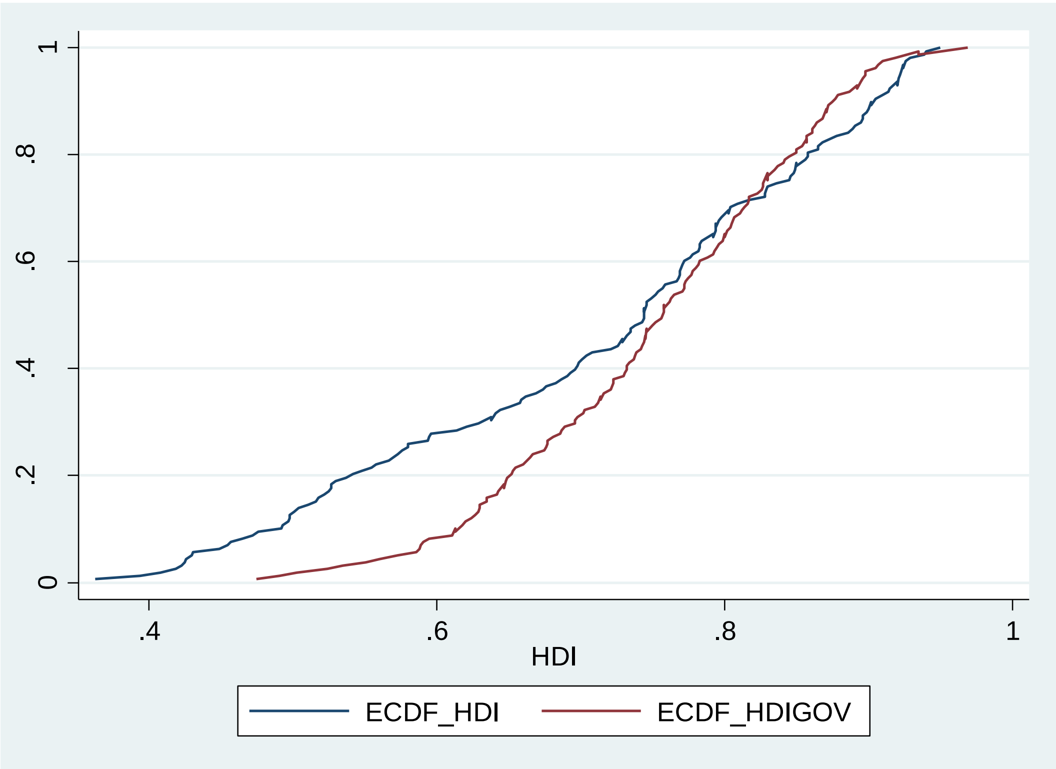

Figure 3. Empirical cumulative distribution of the benchmark HDI and HDI scores with regulation index from the FI (ECDF_HDI and ECDF_HDIGOV, respectively).