

Measuring world governance: revisiting the institutions hypothesis

Mehmet Pinar

Edge Hill University *

December 16, 2013

Abstract

We consider the weighting scheme that results in a best-case scenario in the construction of the World Governance Indicators (WGIs), a proxy of institutional quality. To do that we use an approach that relies on consistent tests for stochastic dominance efficiency of a given index with respect to all possible indices constructed from a set of individual components. The test statistics and the estimators are computed using mixed integer programming methods. The results show that the equally-weighted (fixed weights) composite WGI index is not the best-case scenario and that governance indicators at different years should be weighted differently. Furthermore, we revisit the institutions hypothesis in the empirical growth literature, where institutional quality is the main determinant of long term development. We find that not only institutions matter for economic development, but also geography and macroeconomic policies do affect economic development directly.

JEL Classifications: C12; C13; C14; C15; O1; O57

Key Words: World Governance Indicators; Institutions; Geography; Openness; Nonparametric Stochastic Dominance; Mixed Integer Programming

*Business School, Edge Hill University, St Helens Road, Ormskirk, Lancashire, L39 4QP, United Kingdom. Tel.: +44 1695657629; fax: +44 1695584675. E-mail:pinarmeh@edgehill.ac.uk

1 Introduction

One of the crucial questions in economics is what determines income differences among countries. Why are there such large and persistent income differences among countries? Growth differences between countries are accounted for by differences in human capital, physical capital and technology (e.g. Mankiw et al. 1992). However, why is a poor country unable to replicate a rich country to achieve higher human, physical capital and technology levels? Over the last two decades there has been a lot of research devoted to better understanding the long term determinants of economic development. In that respect, institutions, international trade and geography are found to be the factors that affect income differences between countries in the long term.

One camp argues that institutional quality is the most important determinant of income differences among countries. North (1990) pointed out the importance of institutions as a main factor of economic development. Hall and Jones (1999) pointed out the importance of institutions and linked institutions to colonial time and the origin of the colonizer. Acemoglu et al. (2001) found that many countries' institutional quality was shaped at colonization time and persisted over time and the difference in institutional quality is the main reason behind cross-country income per capita differences. Engerman and Sokoloff (1997), Easterly and Levine (2003), Rodrik et al. (2004), Bosker and Garretsen (2009) are among the many papers that contributed to the debate of the effect of institutions on economic development in the long term.

Another camp argues that geographical characteristics of the country are the most direct determinants of development. Sachs (2003) pointed out the importance of the disease environment that acts as a barrier to economic development for the poorest countries. Bloom and Sachs (1998) and Gallup et al. (1998) found that the disease environment has an effect on economic growth and suggest that a disease environment is geographically specific after controlling for institutions. Sachs (2003) shows malaria transmission directly affects the level of per capita income after controlling for the quality of institutions. Since malaria transmission is strongly affected by ecological conditions, Sachs (2003) concluded that geography affects economic development directly.

International trade is also considered to be one of the main determinants of economic development. Frankel and Romer (1999) argue that trade openness has a causal link to economic development. They use geographical factors to obtain trade openness by summing the predicted bilateral trade shares and found that it has a direct effect on economic development. Diamond (1997), Sachs and Warner (1995), Dollar and Kraay (2003) are some other papers

which emphasized the impact of international trade on development.

There has been a lot of research done on what is called “deep” determinants of economic development. However, there have been many criticisms on how institutional quality is measured. As Glaeser et al. (2004) pointed out, how institutional quality is measured is the root question. Among the most commonly used institutional quality measures are the World Governance Indicators (WGI hereafter). Our present study will try to contribute to this debate.

Kaufmann, Kraay and Zoido-Lobaton (1999), KKZ hereafter, first establish a new database of governance indicators. They constructed six aggregate indicators corresponding to six basic governance concepts: voice and accountability (VA hereafter), political stability (PS hereafter), government effectiveness (GE hereafter), regulatory quality (RQ hereafter), rule of law (RL hereafter) and control of corruption (CC hereafter). Over the years these aggregate indicators were updated, see KKZ (2002), Kaufmann, Kraay and Mastruzzi, KKM hereafter, (2004, 2005, 2006, and 2007a). KKM (2008) lastly updated WGI, six aggregate governance indicators for the years 1996, 1998, 2000 and 2002-2007 annually, covering 202 countries. These aggregate indicators are based on hundreds of specific and disaggregated individual variables measuring various dimensions of governance, taken from 35 data sources provided by 32 different organizations.

In the empirical literature, WGIs are used as composite (equally-weighted) indices in two ways: first use is the WGI composite index for a given year (i.e. averaging six governance indicators at a given year),¹ its second use is the indicator composite index (i.e. taking average of a specific indicator: VA, GE, PS, RQ, RL or CC over the whole period or a specific indicator at a given year).²

Using both the WGI and indicator composite indices with fixed equal weights in the empirical literature suffers from many shortcomings as it ignores the dependence among governance indicators, it introduces double counting of highly correlated indicators and it is also plagued by measurement error problems. In this paper we will follow an approach for the construction of best-case scenario of aggregate indices based on stochastic dominance efficiency (SDE hereafter) analysis that avoids the problems mentioned above.³

Constructing a best-case scenario governance index based on SDE analysis has advantages, since it provides an efficient index resulting from the least variable combination of components

¹See, for example, Neumayer (2002); Easterly and Levine (2003); Beck and Laeven (2006); Easterly (2007); Durlauf et al. (2008) for the similar use of the world governance indicators.

²See, for example, Dollar and Kraay (2003); Rodrik et al. (2004); Bulte et al. (2005); Harms and Lutz (2006); Aidt et al. (2008); Brunnschweiler (2008); Bosker and Garretsen (2009) for the similar use of the world governance indicators.

³Space restrictions preclude us to present the best-case scenario WGIs for each year but we present the best-case scenario for each governance indicator. The results for the best-case scenarios for each year are available upon request from the author.

that maximize the level of governance over time for each country or group of countries. Secondly, economic theory is agnostic in terms of offering us strong guidance about functional form of preferences and distributions of the different components of governance indicators, and relatively large data sets are available, so that nonparametric analysis can let the data “speak for themselves”. The best-case scenario of the governance index refers to the fact that it gives the greatest value of governance level over time for a given probability. In other words, we will construct an index with those weights that will make it stochastically dominate all other competitor indices.

It is worth mentioning that the best-case scenario weighting scheme that we obtain in this paper is derived from the nature of the stochastic dominance (SD hereafter) optimization problem and refers to the “measured” level of governance and not the “true” governance. In our analysis, we take the choice of governance indicators as given and we look at what constitutes the most optimistic scenario given the choice of these governance indicators. This criterion gives emphasis on reducing the overall volatility and as such favoring more equal governance levels. An index that is the least volatile produces the most meaningful comparisons across countries under consideration and any movements over time will be less subjected to measurement error (i.e. change in scores to upward or downward shift will be minimized). In the next section we discuss some of the issues and shortcomings with the use of the equally-weighted governance indices and discuss how the best-case scenario institutional quality indices handle those shortcomings.

An important reason why SD has not been applied before (based on its theoretical attractiveness) in the construction of indices is that until recently, SD could only be tested pair-wise. This restriction was limiting the scope of SD tests, because indices are constructed from a set of components and they effectively face infinitely many choice alternatives. Barrett and Donald (2003) developed pair-wise SD comparisons that relied on Kolmogorov-Smirnov type tests developed within a consistent testing environment. This offers a generalization to Anderson (1996), Beach and Davidson (1983) and Davidson and Duclos (2000) who have looked at second order SD using tests that rely on pair-wise comparisons made at a fixed number of arbitrary chosen points. This is not a desirable feature since it introduces the possibility of test inconsistency. Davidson and Duclos (2000) have discussed the importance of first, second and third order stochastic dominance concepts (SD1, SD2, and SD3 respectively) between income distributions for social welfare and poverty rankings of distributions. As the above mentioned tests are pair-wise in nature, Bawa et al. (1985) extend on the concept of convex stochastic dominance which is first introduced by Fishburn (1974), and provided tests that satisfy necessary and sufficient conditions if more than two choice alternatives are compared. However, the

Bawa et al. (1985) approach only considers a finite number of choice alternatives, not a full diversification.

On the other hand, recently Post (2003) and Kuosmanen (2004) introduced the notion of SDE, which allow for full diversification. Both the Post (2003) and Kuosmanen (2004) approaches are based on derivations of ranked observations under an independent identical distribution (iid) assumption of the asset returns. In other words, both Post (2003) and Kuosmanen (2004) offer a sampling scheme assuming iid observations and therefore do not allow for possible data dependence such as GARCH effects often encountered in high frequency returns (see e.g. Linton and Whang 2012). Scaillet and Topaloglou (2010), hereafter ST, not only allow for full diversification similar to the Post (2003) and Kuosmanen (2004) methodologies but also extend the Barrett and Donald (2003) and Horvath et al. (2006) approaches by allowing for time-dependent data. In other words, the ST methodology allows for full diversification for serially correlated time-dependent data by extending the existing approaches mentioned above. As the WGI is serially correlated over time and there are infinitely many alternative weighting schemes to choose from, we employ the ST methodology to obtain the best-case scenario governance indices. We use the ST approach to test whether the equally-weighted governance indicators are the best-case scenarios or there exist alternative weighing schemes that dominate any other alternative weighting scheme by using the same set of attributes (i.e. VA, PS, GE, RQ, RL and CC over time) to construct the best-case scenario governance indices.⁴ We derive the best-case scenarios of governance indicators that maximize the measured governance levels in a first-order stochastic dominance sense. The best-case scenario governance indicators therefore maximize the measured achieved governance levels and also achieve the minimum variability among its competitors.⁵

To sum up, in this paper, firstly we obtain the best-case governance indices for a study specific sample of countries that have been used as institutional quality proxy in the literature and therefore, tackling the issue of excess variability in the construction of equally-weighted gov-

⁴Similar applications employed by Pinar et al. (2013) to construct the best-case scenario of Human Development Index and Agliardi et al. (2012) to construct a riskiest sovereign risk index for the emerging countries. The current study differs from the above mentioned ones in two ways. Firstly, both Pinar et al. (2013) and Agliardi et al. (2012) examine the absolute levels of human development and sovereign risk respectively. However, in the current paper we examine the relative levels of governance indicators. In that case we derive the least volatile governance indices and therefore these indices will be less subject to measurement error. Secondly, the current study examines the effect of the use of the proposed indices in the relevant literature as institutional quality proxies and provides further insights on how measurement could be an important issue.

⁵As each governance indicator is bounded between -2.5 and +2.5, higher measured governance levels for more countries suggest a distribution that is negatively skewed and therefore having lower variability across countries and over time.

ernance indices by using stochastic dominance efficiency. Secondly, we revisit three prominent studies, namely Easterly and Levine (2003), EL hereafter, Rodrik et al. (2004), and Bosker and Garretsen (2009), BG hereafter, with the proposed best-case governance indices as institution proxies. Rather than using an average of six governance indicators as an institutional quality (e.g. Easterly and Levine 2003) or certain indicator(s) for a given year (e.g. Rodrik et al. 2004 and Bosker and Garretsen 2009), we use the best-case scenario institutional quality indices as proxies for institutions and revisit the institutions hypothesis. The reason we have chosen these studies is to capture the different specifications that tackled the institutions hypothesis. EL (2003) pointed out the importance of endowments' impact on economic development through institutions, Rodrik et al. (2004) examined a larger sample and find that "institutions rule" over geography and international trade for the economic development of a country and BG (2009) point out the importance of not only a country's own institutions but also the neighboring countries' institutions as the main determinant of economic development.

The findings of this paper are as the following. We find that for different set of countries, the equally-weighted institutional quality index does not capture the best-case scenario. One can construct many other hybrid indices that dominate the equally-weighted governance index. With the most optimistic view of the relative levels of governance, across countries and over time, we derive an institutional quality index representing the highest possible governance. Furthermore, this index as the least volatile both across countries and over time would be less subject to measurement error. Secondly, those best-case scenario governance indices for each set of country samples are used as institutional quality proxies to revisit the findings of the recent literature. We find that not only the country's own and neighboring countries' institutions matter for the long term economic development but also the geography and policies matter directly. We find that the valid instrumental variable for institutions will vary with the use of different institutional quality proxies and one would need to choose a valid instrument before proceeding with the application at hand.

The remainder of the paper is as follows. In section 2 we present the use of world governance indicators and the debate regarding the use of governance indicators and the purposes that it serves. In section 3, we examine the main framework of analysis, we define the notions of stochastic dominance efficiency and we discuss the general hypothesis for stochastic dominance of any order. In section 4, we present some simulations to show the importance of the empirical distribution of components over time for the efficiency of the index. In section 5 we present an empirical application. We first look at the data and offer descriptive statistics, and we use the ST methodology to ascertain whether governance composite indices for each indicator over time are the best-case scenarios with the use of equal weights, or whether we can obtain an

alternative weighting with the most optimistic weights for each year for the different constituent components of the index for each study specific samples. In section 6, after obtaining study specific best-case institutional quality indices, we revisit the institutions hypothesis in the three prominent studies' baseline estimations. Finally, section 7 concludes.

2 Literature

There has been a vibrant debate on the use, purpose and construction of WGI. The WGI is used as a composite index of various attributes with fixed weights that measure countries' institutional quality. A serious shortcoming is that the construction of all the above hybrid measures, as in the case of the separate analysis of single attributes, ignores the association among the various attributes. Another shortcoming for the use of composite indices is the issue of double counting when the components are highly correlated. "When using an equal weights, composite index with averaging, it may happen that – by combining variables with a high degree of correlation – an element of double counting may be introduced to the index: if two collinear indicators are included in the composite index with a weight of w_1 and w_2 , the unique dimension that the two indicators measure would have weight $(w_1 + w_2)$ in the composite index. The response has often been to test indicators for statistical correlation – using the Pearson correlation coefficient (Manly, 1994) – and choose only indicators which exhibit a low degree of correlation, e.g. principal component analysis, or to adjust weights correspondingly, e.g. giving less weight to correlated indicators" (Nardo et al. 2005, p. 21). Knack (2006) points out that the aggregate WGIs are constructed from different sources by weighting sources differently where the sources that tend to be more highly correlated with the other sources are given greater weight. However, Knack (2006) points that the highly correlated sources getting greater weight is not desirable if the measurement error is correlated among sources (i.e. measurement errors are not independent). Moreover, since each world governance indicator is highly correlated with the other ones and each indicator is highly correlated over the years⁶, it is suggested that the composite index ranks countries in a manner not dissimilar from the way any governance indicator ranks them. These correlations would seem to lie at the heart of the indicated redundancy, both in terms of rankings and values, of the composite indices.

Another shortcoming for WGIs with fixed equal weights is measurement error. Both earlier and currently updated aggregate level of governance indicators have reported margin of errors

⁶Each aggregate governance indicator is highly correlated with the other aggregate governance indicators at a given year. Moreover, each governance indicator outcomes at a given year is highly correlated with the following year outcomes of that indicator. The correlation coefficients, both the simple and the Spearman-rank one are very high (above 0.9) and very significant.

due to permanent measurement errors among all indicators of governance. However, other than the reported margins of error, there are three other sources of measurement error. First of all, aggregate governance indicators rely on the perception of governance by the public sector, the private sector, NGO experts, firms and citizens. Glaeser et al. (2004) point out those governance indicators are ex-post outcomes, highly correlated with the level of economic development, rather than political constraints per se. Therefore, WGIs are plagued by recent economic performances and/or policy rather than being influenced by actual governance levels. Kurtz and Schrank (2007) find that GDP per capita maintains a substantively and statistically important relationship to government effectiveness and once investment, education, inflation and the last two year rates of economic growth effects are controlled for; government effectiveness no longer fosters subsequent rates of economic growth. Therefore, both Glaeser et al. (2004) and Kurtz and Schrank (2007) point out that the economic development of a country had an effect on the governance indicators, not vice versa.

The second source of measurement error is related to the construction of WGIs over time. Arndt and Oman (2006), and Knack (2006) indicate that comparisons cannot be made since WGIs are scaled to have zero mean and one standard deviation at a given year. As more countries are covered over time and all aggregate governance indicators are standardized to have a zero mean and unit standard deviation for a certain year, there is a need for adjustments to earlier periods to maintain the zero mean and unit standard deviation. As a result, changes in scores over time are related to updating scores rather than actual improvements or deteriorations in governance. Moreover, Knack (2006) suggests that updating governance scores to maintain the zero mean and unit variance is the source of the change of the relative rankings of the counties.

The third source of the measurement error is that the individual disaggregated data are different for each country over time and updating scores will make the different data sources become more correlated with each other over time and as such they will result in weights that will produce inefficient fixed weight indices of the aggregate level of governance, as the latter will become more variable. Knack (2006) state that indicators for 2002 and 2004 are based on the same set of sources for only 4 of the 27 countries in Eastern European and Central Asia (ECA) counties, therefore the comparison between two countries or a country over time may be based on a totally independent source of data and the changes in scores over time may be caused by the correction of the scores over time. Therefore, Knack (2006) investigates the effect of updating scores on Eastern Europe and Central Asia countries and finds that updating scores will make the different data sources become more correlated with each other over time.

In this paper, we question the use of composite governance indices in the empirical growth

literature (e.g. averaging governance levels over time for each respective indicator). As governance levels for all indicators are highly correlated over time, assigning equal weights to governance levels of each year might produce a double counting problem. Moreover, it is important to deal with the presence of measurement error as it creates high variability for the fixed weight indices. If the governance levels for a given indicator are highly correlated over time, more weight should be given to a governance level of a particular year that has a different empirical distribution function (i.e. governance levels of that year release different information than the ones of other years) and that is also the least volatile one over time (i.e. governance levels of that year are less subject to score updating and measurement error). KKM (2007b) claim that particular weights given to different sources in order to construct the aggregate indicators do not make a significant difference in overall estimates of governance. However, this claim is not true for the use of composite indices of aggregate indicators. We will show empirically that the equally-weighted composite indices are not the best-case scenarios and there are many alternative composite indices that dominate the equally-weighted ones.

3 SD Efficiency Testing

We consider a strictly stationary process $\{\mathbf{Y}_t; t \in \mathbb{Z}\}$ taking values in \mathbb{R}^n . The observations consist of a realization of $\{\mathbf{Y}_t; t = 1, \dots, T\}$. These data correspond to observed values of the given governance indicator's levels over time (i.e. nine different given governance indicator levels over time, e.g. VA in 1996, 1998, 2000, 2002, 2003, 2004, 2005, 2006, and 2007). We denote by $F(\mathbf{y})$, the continuous cdf of $\mathbf{Y} = (Y_1, \dots, Y_n)'$ at point $\mathbf{y} = (y_1, \dots, y_n)'$. Let us consider a hybrid composite index with a weighting vector $\boldsymbol{\lambda} \in \mathbb{L}$ where $\mathbb{L} := \{\boldsymbol{\lambda} \in \mathbb{R}_+^n : \mathbf{e}'\boldsymbol{\lambda} = 1\}$ with \mathbf{e} being a vector of ones. This means that all the different year outcomes of the given governance indicator have positive weights and that these weights sum up to one. Let us denote by $G(z, \boldsymbol{\lambda}; F)$ the cdf of the hybrid index value $\boldsymbol{\lambda}'\mathbf{Y}$ at point z given by $G(z, \boldsymbol{\lambda}; F) := \int_{\mathbb{R}^n} \mathbb{I}\{\boldsymbol{\lambda}'\mathbf{u} \leq z\} dF(\mathbf{u})$.

3.1 Tests for SD Efficiency of different indices

SDE is a direct extension of SD to the case where full diversification is allowed. In that setting we derive statistics to test for SDE of the equally-weighted governance composite index (with the vector of equal weights denoted by $\boldsymbol{\tau}$) with respect to all possible combinations of weighting schemes ($\boldsymbol{\lambda}$) constructed from the set of year outcomes of given governance indicator.⁷

⁷We have defined above $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$ to be different weighting vectors that are associated with different indices. In the discussion that follows we use $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$ interchangeably with the index that they

In this paper we test whether the use of equally-weighted governance indicator, τ , is the best-case scenario, in the sense that it gives the maximum value and lower variability of measured governance levels across countries and over time or whether we can construct another composite index λ (alternative weighting scheme) from the set of year outcome levels that dominates it.

The distribution of the hybrid index λ dominates the distribution of the index τ stochastically at first-order (SD1) if, for any argument z , $G(z, \tau; F) \geq G(z, \lambda; F)$. If z denotes a governance level, then the inequality in the definition means that the proportion of countries in distribution λ with value of governance smaller than z is not larger than the proportion of such countries in τ . In other words, there is at least as high a proportion of governance in λ as in τ . If the composite index λ dominates the index τ at first order, then there are always more countries having relative levels of governance below a given governance level, z , in τ than in λ , so that λ achieves higher relative levels of measured governance for more observations than τ .

The objective function that we use is the following:

$$Max_{z, \lambda} [G(z, \tau; F) - G(z, \lambda; F)]$$

The above maximization results in the best-case scenario (most optimistic) governance index λ constructed from the set of year outcomes of that governance indicator in the sense that it reaches the highest level of measured governance for a given probability, implying that the number of observations having a relative governance level above a given argument z is maximized.

As pointed out in Pinar et al. (2013), it is worth mentioning that SD is considerably more general than mean-variance analysis (MVA) which only looks at the first two moments of the two distributions under comparison. On the other hand, Post (2003) suggested that the MVA is consistent with expected utility theory if and only if investor preferences and/or empirical distributions obey certain conditions. On the other hand the SDE methodology allows for full diversification considering all moments and therefore offers more robust findings. Moreover, SDE has additional advantages over MVA as the latter allows for possible trade-off between two moments and might classify inferior portfolios as efficient and inefficient portfolios as optimal (see e.g. Post and Versijb 2007). Therefore, the MVA analysis only looks into a dominant relation with a higher mean and lower variance, whereas the SDE methods consider all possible moments. Only in the case where one compares two normal distributions does SD reduce to mean-variance analysis. This is also true for PCA which is based on the consideration of the second moment alone after standardizing for a common mean. However, the assumption of

represent.

normality for each component is difficult to support empirically. In contrast, SD analysis takes into account the whole distribution, not only the mean and the variance. Hence, one could expect significant differences between the SD efficient index and the mean-variance efficient index when more realistic assumptions are made concerning the distributions of the different components. SD is attractive because it is effectively nonparametric as no explicit specification of a utility function or probability distribution functional form is required (Post 2003; Post and Versijb 2007; Wong et al. 2008; Scaillet and Topaloglou 2010). In addition, the entire probability density function is taken into account rather than a finite number of moments so it can be considered less restrictive and more robust.

When studying governance measures, certain criteria need to be satisfied. The SD1 criterion corresponds to all types of utility functions as long as they are non-decreasing in governance levels. SD1 only relies on the fact that people are rational in the sense that they prefer more rather than less governance (also known as the monotonicity axiom). Therefore, a sensible aggregate measure should increase in any indicator which is a social ‘good’, and decrease in any indicator which represents a social ‘bad’. Accordingly, for aggregate governance indices containing only social ‘good’ indicators (i.e. different set of governance measures), one hybrid outcome as expressed by the index λ should be ranked higher than that of another hybrid outcome expressed by the index τ if at least one country is better off in λ than in τ , and no one is worse off. So, SD1 of τ by λ means that λ corresponds to a higher measured relative governance than τ .

When there is no governance index λ that dominates the given index τ at first-order, we move to the SD2 criterion. The objective function that we use is the following:

$$Max_{z,\lambda} \int_{-\infty}^z G(u, \tau; F) du - \int_{-\infty}^z G(u, \lambda; F) du$$

This maximization results in the best-case governance index λ constructed from the set of year outcomes of a given governance indicator in the sense that it also gives the greatest value of governance for a given probability.

Overall, the ST methodology ensures the regularity conditions of nonsatiation (first-order SDE) and risk aversion (second-order SDE). The best-case scenario must be first-order stochastic dominance efficient for all governance indicators that use a nonsatiable decision maker (see Fishburn 1974; Bawa et al. 1985; Post 2003; Kuosmanen 2004; Scaillet and Topaloglou 2010 among many others) which is appropriate for both risk averse and risk lover individuals as utility function may have concave and convex segments respectively. In other words, first-order SDE is more general as it corresponds to all types of utility functions as long as

they are non-decreasing. We first test for first-order SDE of the equally-weighted governance index to ascertain whether it is a best-case scenario governance index or whether there exists alternative weighting schemes that dominate it. Therefore, first-order dominance will suggest whether the equally-weighted index constitutes a rational choice or not. If it is dominated by alternative weight schemes, then it will never be chosen independently of whether the decision maker is risk averse, risk neutral or risk lover. If there is no alternative weighting scheme, $\boldsymbol{\lambda}$, that dominates the equally-weighted one, $\boldsymbol{\tau}$, we move to the second-order SDE that allows for a nonsatiable and additionally risk-averse decision maker which permits a selection of weights for the best-case scenario.

We can further define for $z \in \mathbb{R}$:

$$\begin{aligned}\mathcal{J}_1(z, \boldsymbol{\lambda}; F) &:= G(z, \boldsymbol{\lambda}; F), \\ \mathcal{J}_2(z, \boldsymbol{\lambda}; F) &:= \int_{-\infty}^z G(u, \boldsymbol{\lambda}; F) du = \int_{-\infty}^z \mathcal{J}_1(u, \boldsymbol{\lambda}; F) du, \\ \mathcal{J}_3(z, \boldsymbol{\lambda}; F) &:= \int_{-\infty}^z \int_{-\infty}^u G(v, \boldsymbol{\lambda}; F) dv du = \int_{-\infty}^z \mathcal{J}_2(u, \boldsymbol{\lambda}; F) du,\end{aligned}$$

and so on.

From Davidson and Duclos (2000) Equation (2), we know that

$$\mathcal{J}_j(z, \boldsymbol{\lambda}; F) = \int_{-\infty}^z \frac{1}{(j-1)!} (z-u)^{j-1} dG(u, \boldsymbol{\lambda}, F),$$

which can be rewritten as

$$\mathcal{J}_j(z, \boldsymbol{\lambda}; F) = \int_{\mathbb{R}^n} \frac{1}{(j-1)!} (z - \boldsymbol{\lambda}'\mathbf{u})^{j-1} \mathbb{I}\{\boldsymbol{\lambda}'\mathbf{u} \leq z\} dF(\mathbf{u}).$$

The general hypotheses for testing SDE of order j of $\boldsymbol{\tau}$, hereafter SDE_j , can be written compactly as:

$$\begin{aligned}H_0^j &: \mathcal{J}_j(z, \boldsymbol{\tau}; F) \leq \mathcal{J}_j(z, \boldsymbol{\lambda}; F) \text{ for all } z \in \mathbb{R} \text{ and for all } \boldsymbol{\lambda} \in \mathbb{L}, \\ H_1^j &: \mathcal{J}_j(z, \boldsymbol{\tau}; F) > \mathcal{J}_j(z, \boldsymbol{\lambda}; F) \text{ for some } z \in \mathbb{R} \text{ or for some } \boldsymbol{\lambda} \in \mathbb{L}.\end{aligned}$$

Under the null hypothesis H_0^j there is no governance index $\boldsymbol{\lambda}$ constructed from the set of year outcomes of a given governance indicator that dominates the given equally-weighted index $\boldsymbol{\tau}$ (i.e. equally-weighted year outcomes of a given governance indicator) at order j . In this case, the function $\mathcal{J}_j(z, \boldsymbol{\tau}; F)$ is always lower than the function $\mathcal{J}_j(z, \boldsymbol{\lambda}; F)$ for all possible

hybrid indices $\boldsymbol{\lambda}$ for any argument z . Under the alternative hypothesis H_1^j , we can construct a governance index $\boldsymbol{\lambda}$ that for some arguments z , the function $\mathcal{J}_j(z, \boldsymbol{\tau}; F)$ is greater than the function $\mathcal{J}_j(z, \boldsymbol{\lambda}; F)$. In other words, the main focus of the hypotheses is to test the efficiency of the equally-weighted index (i.e. whether the equally-weighted index is the best-case scenario at all levels of governance). If the null is rejected, then the evaluated index (i.e. equally-weighted index) is not the best-case scenario.

We consider the weighted Kolmogorov-Smirnov type test statistic

$$\hat{S} := \sqrt{T} \sup_{z, \boldsymbol{\lambda}} \left[\mathcal{J}_j(z, \boldsymbol{\tau}; \hat{F}) - \mathcal{J}_j(z, \boldsymbol{\lambda}; \hat{F}) \right],$$

and a test based on the decision rule:

$$\text{reject } H_0^j \quad \text{if } \hat{S}_j > c_j,$$

where c_j is some critical value (see section 2 of Scaillet and Topaloglou 2010 for the derivation of the test). In order to make the result operational, we need to find an appropriate critical value c_j . Since the distribution of the test statistic depends on the underlying distribution, we rely on block bootstrap methods to simulate p-values. The use of block bootstrapping is designed for time series applications and we used block bootstrap in the present case to guard the findings against the time dependence of the panel structure; however, different set of bootstrapping approaches could also be employed in future studies to account for panel dependence.⁸

We derive statistics to test for SDE of the equally-weighted governance index with respect to all possible combinations of weighting schemes constructed from the set of year outcomes of a given governance indicator. In the next section, before moving to the empirical analysis of WGIs, we present some simulation experiments to evaluate the importance of different distributional component characteristics in the derivation of best-case scenario weights.

4 Simulations

We present some simulations to demonstrate the importance of empirical distribution of components over time. We present simulation results for two different experiments. In each case we have six different components (e.g. those components could be considered as governance levels of a given indicator over a six year period) and all of them are normally distributed as in

⁸Test statistics are obtained for different governance levels considering all possible weight combinations which require mixed integer and linear programming for the first- and second-order SDE test statistics. See section 4 of the ST for the mathematical formulation details.

the case of the construction of each governance indicator at a given period. Then those components are used to construct the equally-weighted composite index by using 500 observations. The simulation results are reported in Table 1.

In the first experiment, we simulate six components which are normally distributed with each component having different mean, 0.25, 0.2, 0.15, 0.1, 0.05 and 0 respectively and same standard deviation, 1, and we proceed to construct an equally-weighted composite index from these simulated components. The results of the first panel of Table 1, show that there are 494 composite indices for a given “z” point that dominate the equally-weighted composite index and it is clear that the highest mean component has the greatest impact. In that case, the highest mean component’s weight (the first component) from these 494 dominant indices is on average 0.422. On the other hand the other components with respective means 0.2, 0.15, 0.1, 0.05 and 0 have weights 0.285, 0.122, 0.068, 0.062 and 0.041 respectively. The above simulation results suggest that using the equally weighted composite index is not the best-case scenario index and it is dominated by many other potential component indices with different weights. One can see that when the means of each component are different, with the same standard deviations, the component with highest mean has the greatest impact in the construction of the best-case index. In the second and third cases in the first panel, where we decrease the mean differences among the components (e.g. in the second case, the gap between means is 0.03 and in the third case, this gap is 0.01), we can see that the component with the highest mean has the greatest impact in the construction of the best-case index, however, this impact decreases as the gap between the means of the different components decreases. Similarly, when the mean gap between components decreases, the weight assigned to each component is getting closer to each other. Finally, for the fourth case, we have six components which are normally distributed and all have same mean of “0” and same standard deviation of “1”. There is no composite index for a given “z” point that dominates the equally-weighted composite index where we can conclude that the average weighting is the best-case scenario. Therefore, when the means are close to each other and each component has the same variability as any other, the equally-weighted index becomes the best-case scenario since each component will have the same empirical distribution and will contribute equally to the composite index. Since the index is bounded between -2.5 and +2.5, the higher mean component has the least standard deviation in the upper tails. Therefore, not only the component with higher mean is getting more weight but also does the component which is least variant.

In the second experiment, we allow the six components to have the same mean values (i.e. 0), but some of them have different standard deviations. In the first case of the second experiment, we allow some of the components to have different standard deviations (i.e. each

component's standard deviation is set to be 1, 1, 1, 0.8, 0.5 and 0.2 respectively). The results of the first case of this experiment show that there are 41 composite indices for a given "z" point that dominate the equally-weighted composite index and it is clear that the least variable component has the greatest impact. In that case, the least variable component's weight (the last component) from these 41 dominant indices is on average 0.384. In the third case, we allow some of the components to have different means and different standard deviations. First three variables have one standard deviation and last three variables have 0.8 standard deviation. First and fourth variable, second and fifth variable, third and sixth variable have mean values 0.2, 0 and -0.2 respectively. In that case, the component with higher mean and least standard deviation has the highest weight in the construction of the best-case scenario. The highest mean and least variable component's weight (the fourth component) is on average 0.394. Moreover, the component with lowest mean and highest standard deviation has the lowest weight in the construction of the best-case scenario. The lowest mean and most variable component's weight (the third component) is on average 0.053. There is a trade off between mean and variance. For example, the second and fifth components have the same mean but fifth component has lower standard deviation than the second component. The gap between the first and second component's weights in the construction of the best-case scenario index is 0.154, but the gap between the weights of the first and fifth component is 0.12.

Overall, we find that under normality, when all components have different means but the same standard deviation, the component with the highest mean has the greatest impact (weight) in the construction of the best-case index. When the gap between the means of the different components decreases, then the weights of these components become similar to each other. Finally, if all components have the same distribution (i.e. zero mean and a variance of one), then averaging the components is the best-case scenario since all distribution of components is equally informative and there is no other weighting scheme that dominates the given index. On the other hand, when each component has the same mean and different standard deviations, then the least variable component has the greatest impact in the construction of the best-case scenario index. When both mean and standard deviations vary, then the component with the highest mean and the lowest variability relative to the other components has the greatest impact in the construction of the most optimistic index. There is a trade off between mean and standard deviation and in the case where one component has the highest mean and highest standard deviation and the other component has the second highest mean but the least variability, then the gap between the components' weights will be less in the construction of the best-case scenario index.

The SD approach maximizes the distributional distance between the given (equally-weighted)

index and any possible alternative. In comparing distributions, we know that the mean plays the key role, followed by variability. The mean indicates the better governance level offered, while the standard deviation shows the volatility of the governance level. The best-case scenario index obtained from the SDE approach is the one that offers the greatest governance level over time and at the same time exhibits discernibly the most stable governance level.

5 Empirical Application

5.1 Data and Descriptive Statistics

We use the World Bank’s World Governance Indicators (see KKM, 2008 where the units of the aggregate governance indicators will be those of a standard normal random variable, i.e. with zero mean, unit standard deviation, and ranging approximately from -2.5 to 2.5 with higher scores corresponding to better governance). The mean of the governance estimates for each indicator at a given year is set to zero, and the standard deviation is one. In particular, for each indicator at a given year, KKM (2008) subtract the sample mean (across countries) from each country, and divide by the sample standard deviation (across countries).

Each governance indicator composite index represents the simple arithmetic average of the nine year entry of that indicator:

$$\text{Indicator composite index}_I = \frac{I_{1996} + I_{1998} + I_{2000} + I_{2002} + I_{2003} + I_{2004} + I_{2005} + I_{2006} + I_{2007}}{9}$$

while $I = VA, PS, GE, RQ, RL, CC$

Table 2 presents the descriptive statistics for each individual governance indicator over time and the equally-weighted composite index obtained by averaging the governance levels over time. One can see that each indicator composite index has lower standard deviation than the standard deviation of the indicator at a given year as expected. However, in order to have a balanced data set, countries that do not have all components for any given year need to be excluded from the analysis. Therefore, even though the original indicators are constructed to preserve the zero mean and unit standard deviation for the full sample size, in practice when certain countries are selected for a specific study and some observations are excluded to form a balanced data set, the distribution of indicators no longer has the property of zero mean and unit standard deviation. In the next section, we will examine the efficiency results for equally-weighted governance indices for each study specific sample size.

5.2 Results for the efficiency of the equally-weighted governance indices

In this section, we provide the results of the SD efficiency tests. We find that the equally-weighted index is not the best-case scenario for all different sample sizes as there are alternative weighting schemes at different governance levels that dominate the evaluated index. Therefore, it is clear that the equally-weighted index is not the best-case scenario index for different sample sizes. Below, we will offer the results for each sample case and give the best-case scenario for each sample, and also provide comparisons between the equally-weighted index (i.e. evaluated index) and the proposed best-case scenario index for all sample cases.

In the EL (2003) study, the sample size is 72 countries. Table 3A summarizes the results when equally-weighted governance indices (i.e. evaluated composite VA, PS, GE, RQ, RL and CC indices which are averaged over time) are tested. The first column of Table 3A offers the governance indicator that is being evaluated, whereas the second panel gives the number of countries that have a balanced data set for the sample. After conducting tests, we find that the over time equally-weighted index for each governance indicator is not the best-case scenario as there are many other indices λ consisting of the nine years that dominate the evaluated composite index. For each governance level, z , there is an alternative weighting scheme that results in the best-case scenario for that given governance level.

To derive a weighting scheme that is the best-case scenario, we average the dominating weighting schemes that are found for different z levels. In order to test whether these proposed best-case scenarios are best cases, we implemented the proposed test for these first-stage best-case scenarios.⁹ The first-stage best-case scenarios are also dominated by alternative schemes but fewer times (i.e. first-stage best-case scenario is a better case for more governance levels when compared to the equally-weighted one). Therefore, we continue implementing iterations of testing the best-case scenarios. The third column of Table 3A offers the number of iterations took to reach the best-case scenarios for each governance level. On average, it takes 5-6 iterations to reach the best-case scenario for a given governance indicator. After implementing iterations of the test, we find the best-case scenario for each governance level that is not dominated by any alternative weighting scheme. For example, the best-case VA index is obtained when governance levels in 1996 and 2004 are weighted by 96.5% and 3.5% respectively. Similarly, the best-case scenario weighting schemes for the remaining governance indicators are given in each row of Table 3A for the EL (2003) sample.

To make the results with the best-case scenario more apparent, we offer comparisons between

⁹We thank an anonymous referee for suggesting us to test for the first-stage best-case scenarios.

the evaluated index (i.e. equally-weighted governance index) and the best-case scenarios for each governance indicator in Table 3B. We provide comparisons between the evaluated and the best-case scenarios for the first three quartiles of the empirical distribution of the evaluated index. The second column offers the governance levels of the equally-weighted index at these first three quartiles. Finally, the last two columns offer the number of observations that are above these governance levels at different quartiles. For example, there are 56, 37, and 19 countries above -0.872, -0.220, and 0.329 governance levels with the equally-weighted VA index. However, there are always more countries above these governance levels with the best-case scenario weighting scheme (57, 38 and 21 countries respectively). Similarly, Table 3B offers the comparison between the evaluated and best-case scenario index for the remaining governance indicators.

Moving to the Rodrik et al. (2004) study, there are three different sample sizes that are used, namely the Acemoglu sample, the extended Acemoglu sample and a larger sample with each sample having 64, 79 and 137 countries respectively. For all these sample sizes, we find that the equally-weighted indicator index is not the best-case governance index for each governance indicator. Tables 4A, 5A, and 6A offer the best-case weighting schemes for each governance indicator for each sample with 64, 79, and 137 countries respectively. Similar to that of Table 3A, we offer the number of countries for each application and the number of iterations of testing employed to obtain the best-case scenarios. Tables 4B, 5B, and 6B offer the comparisons between evaluated and the best-case scenarios for the first three quartiles of the empirical distribution of the evaluated index. We find that there are always more countries that are above a given governance level with the best-case scenario when compared with the evaluated index.

Lastly, the BG (2009) paper uses two different samples. One sample is used for choosing the instrumental variable and consists of 79 countries. The other sample is used for baseline estimation and consists of 147 countries. We also find that the equally-weighted indicator index is not the best-case governance index for each governance indicator. Tables 7A and 8A summarize the best-case scenario weighting schemes for each governance indicator for the 79 and 147-country samples. Tables 7B and 8B offer the comparisons between the evaluated index and the best-case scenario index when the number of countries (observations) above a given governance level is given for the first three quartiles of the empirical distribution of the evaluated index. We find that the best-case scenario indices for each governance level result in a higher number countries achieving higher governance levels.

The above results suggest that using the equally-weighted index is not the best-case scenario because it is stochastically dominated by many other potential composite indices with different weights for each year of a given governance indicator. We find that different years

contribute differently to the construction of the most optimistic governance index. In the next section, we revisit the findings of the three prominent studies by using the best-case scenario indices as a proxy for the institutional quality.¹⁰

It is important to point out that the weights assigned to each year for each governance indicator and the degree of efficiency may change over time as pointed out by Post (2003). Firstly, with the additional releases of WGIs, the weights assigned to each year for each respective governance index might change. The best-case governance indices for any given set of countries, i.e. best-case governance indices for the pre-determined samples, can vary over time. Secondly, if different sample sizes are used, each year might yield different weights for the best-case scenario governance indices. Therefore, the weights assigned to each year for the best-case governance indices need to be re-estimated periodically.¹¹

6 Revisiting the Institutions Hypothesis

6.1 Case 1: Easterly and Levine (2003)

EL (2003) find that geographic endowments explain economic development through their impact on institutions but do not have direct impact on economic development. On the other side, EL found no impact of policies on economic development once institutions are controlled for. The main question that we pose here is the following: Are the findings of the EL study still valid once we use our proposed best-case governance indices as proxies for institutional quality and/or are there some differences? Do policies matter once we use our proxies for institutional quality? Do endowments affect economic development directly or through channels other than their impact on institutions?

We find that when the best-case governance indices are used as institutional quality proxy, institutions are still significant after controlling for legal origin, religion and ethnolinguistic

¹⁰The next sections summarize the findings of the three studies when the best-case scenario governance indices are used as institutional proxies. However, detailed findings of the three studies are available upon request from the author.

¹¹To assess the robustness of the findings in the next sections, we also test whether the equally-weighted governance indices are the best-case scenario indices or not for the full sample size which consists around 200 observations. We find that the equally-weighted indices are not the best-case scenarios for all governance indices when the full sample size is used. Even though weights assigned to each year in the case of full sample size differ from the ones that are found with different sample sizes, the findings that are presented in the next sections do not change significantly when the best-case scenario governance indices for the full sample size are used as institutional proxies. Especially, the different results with the revisited papers still hold when the best-case scenario indices for the full sample size are used. Full sample size best-case scenario governance indices and detailed findings of three studies with those indices are available upon request from the author.

diversity and use of different endowments as instruments. However, over-identification tests offer different results compared with those in the EL Table 4. We do find that the data sometimes reject the hypothesis that endowments only explain the economic development through their ability to explain institutional development. Therefore, once some of the best-case governance indices are used as institution index, endowments are not valid instruments and endowments do explain economic development through other channels rather than institutions alone.

In their Table 5, EL further control for the main macroeconomic policies (i.e. inflation, real exchange rate overvaluation and trade openness) and find that only institutions are the main significant variable. Once we use best-case scenario institutional quality indices, we find that the institutions are still significant. Therefore, our findings with the best-case scenario indices still favor the institutions approach. However, we also find that policies do matter. Inflation, trade openness, real exchange rate overvaluation enter significantly in the regressions. Inflation enters with the “wrong” sign, but trade openness and real exchange rate overvaluation enter significantly and with the “correct” sign. Once some of the best-case institutional quality indices are used, trade openness has a positive impact and overvaluation has a negative impact on development. On the other side, we also find that once the most optimistic institutional quality indices are used, in some cases, the over-identification test is rejected, so settler mortality and latitude do explain economic development through other channels that are different from the institutions.

In Table 6 of EL (2003), they find that institutions are still significant and none of the policy variables enter the regressions significantly at 10% level when policies are treated as endogenous variables. However, we find that for some specifications, once our proposed institutional quality indices are used, institutions are not significant any more, something that contradicts the institutions hypothesis. On the other side, we find that trade openness affects economic development significantly. When institutions and trade openness are used in the regressions, we found that once the best-case PS indices are used as proxy for institutional quality, trade openness is significant at 10% level. Moreover, when institutions, trade openness, religion and French legal origin are used in the regressions, we find that trade openness significantly affects development when the best-case PS and VA institutional quality indices are used.

Finally, when the institution index and trade openness are used in the regressions, we find that once the best-case GE, RL and CC indices are used as a proxy for the institutions, the over-identification test is rejected at the 5% level. Moreover, when the institutions and the real exchange rate overvaluation are used in the regressions, we find that the over-identification test is also rejected at the 5% level when the best-case GE, RL and CC indices are used. Over-identification test rejection means that endowments do explain economic development through

other channels that are different than institutions alone and therefore they cannot act as valid instrumental variables for institutions.

6.1.1 Case 2: Rodrik et al. (2004)

Rodrik et al. (2004) investigate the contributions of the deep determinants of economic development, namely geography, institutions and international trade to income level differences around the world. Rodrik et al. (2004) find that institutions “trumps” since institutions have a direct impact on economic development with different sample sizes and specifications. Once institutions are controlled for, geography has indirect effect on economic development through its impact on institutions. On the other hand, once institutions are controlled for, international trade is always insignificant and with the “wrong” sign.

Rodrik et al. (2004) differs from the EL (2003) study in a couple of ways. Firstly, Rodrik et al. (2004) have larger sample sizes (79 and 137 countries) whereas EL (2003) only consider 72 countries. Secondly, Rodrik et al. (2004) consider international trade in their main estimation, but EL (2003) did not test its impact on economic development (see footnote 3 in Rodrik et al. 2004 for further details).

In the Rodrik et al. (2004) study, geography is treated as an exogenous variable and measured by absolute latitude. Institution is treated as an endogenous variable and is measured with the rule of law in 2000, taken from world governance indicators. Depending on the sample size, settler mortality (sample with 64 and 79 countries) or percentage of population speaking English and percentage of population speaking European language (sample of 137 countries) are used as instrumental variable(s) for the institutions. Finally, international trade is treated as an endogenous variable and is measured as the ratio of trade to GDP. Predicted trade shares from Frankel and Romer (1999) are used as an instrumental variable for international trade.

Since the Rodrik et al. (2004) paper uses scores of the rule of law 2000 from world governance indicators as an institution proxy, the main questions that we pose are: Are the findings of the Rodrik et al. (2004) still valid once we use our proposed best-case scenario indices as proxies for institutional quality and/or are there some differences? Do institutions “trump” after all? Does geography or international trade have a direct impact on economic development?

In their baseline findings, Table 3 of Rodrik et al. (2004), they find that the institutions “trump” with different specifications and sample sizes. On the other hand, when the best-case scenario of governance indices are used, for the samples of 64-countries and 79-countries, except for the best-case PS index, all most optimistic institutional quality indices confirm the findings of Rodrik et al. (2004), where institutions “trump” over integration and geography. Once the best-case PS index is used as a proxy for institutions, the institutions are not significant for the

64- and 79-country case which contradicts the institutions hypothesis. Moreover, we find that Rodrik et al. (2004) findings for the large sample are not robust when some of the best-case governance indices are used as institutional proxy. We find that geography has a direct impact on economic development once the most optimistic VA and RQ indices are used as a proxy which is consistent with the geography approach. Geography has not only an indirect effect through institutions but also has a direct impact on economic development. Finally, unlike the Rodrik et al. (2004) findings, when the best-case VA, PS and RQ institutional quality indices are used, the over-identification test is not rejected at the 10% level for the large sample, which means that percentage of population speaking English and percentage speaking European language are valid instruments for the institutions. Since both instruments are valid for some best-case institutional quality indices, then one can exploit the findings in the large sample case. In the Rodrik et al. (2004) findings, the large sample findings are ignored since in their case the instruments were not passing the over-identification test. However, when the best-case VA and RQ indices are used, not only are instruments valid for the large sample but also geography has a direct impact on economic development, something that favors the geography approach.

Following Rodrik et al. (2004) Table 4, Rodrik et al. examined the inter-relations between institutions and international trade for the samples of 79-countries and 137-countries. Rodrik et al. (2004) find that for the 79-country sample, geography affects institutions significantly and positively. Countries that are in higher latitudes trade less and countries that have better institutions trade more. On the other side, for the large sample case, the only significant IV estimate is that geography affects institutions significantly and positively. When the majority of the best-case institutional quality indices are used, our findings confirm Rodrik et al. (2004) findings. However, we have two major differences from their findings. Firstly, once the best-case PS index is used, then international trade does affect institutions significantly and positively for the 79-country sample. Secondly, once the best-case VA index is used, institutions are not significant any more in affecting international trade for the 79-country sample.

6.1.2 Case 3: Bosker and Garretsen (2009)

BG (2009) mainly concentrated on absolute geography, measured as the absolute latitude of a country; institutions of a country, measured with the rule of law in 2000 from world governance indicators; institutions of neighboring countries, measured as the average of institution proxy for neighboring countries. BG (2009) find that it is not only the country's own institutions that matter for economic development but also the neighboring countries' institutions are important for development of that country.

In the BG (2009) study, the country's own institution proxy and the neighboring countries'

institutions are treated as an endogenous variable; therefore, at least two instrumental variables are needed for instrumental variable estimation.¹² There are three possible instrumental variables suggested by the literature, settler mortality rate (available only for 79 countries), percentage of population speaking English and percentage of population speaking a European language (sample of 147 countries). In order to exploit the larger sample case, BG (2009) uses percentage of population speaking European language as a valid instrument for institutions (see section 3.3 and Appendix B of the BG study for detailed procedure of the choice of the instrumental variable). Once the percentage of population speaking a European language is used as an instrument for institutions, BG (2009) find that not only the country's own institutions that matter for economic development but also the neighboring countries' institutions as well. Absolute geography on the other hand has an indirect effect on economic development through its effect on institutions.

Since the BG (2009) paper uses scores of the rule of law in 2000 from the world governance indicators as an institutional proxy, the questions we pose are the following: Are the findings of BG (2009) still valid once we use our proposed best-case scenario indices as proxies for institutional quality and/or are there some differences? Is the percentage of population speaking European language still a valid instrument or is it valid to use other instruments? Do institutions of a country or neighboring institutions matter for economic development after all? Does geography have a direct impact on economic development?

First, we examine the choice of the instrumental variable. When the majority of the best-case governance indices are used as proxies for institutions, our findings confirm BG (2009) findings that the percentage of population speaking a European language is a valid instrument. However, we find that different instruments are valid once we use some other best-case governance indices as proxies for institutional quality. Firstly, when we use the best-case VA, PS and RQ indices as institutional quality proxies, we find that using both the percentage of population speaking English and the percentage of speaking a European language are valid instruments since over-identification tests are not rejected. On the other hand, using all possible three instruments together is a valid option when the best-case PS and RQ governance indices are used.

With the 147-country sample, when the best-case VA, PS and RQ indices are used, both language instruments are valid. For the remaining best-case institutional quality indices, using the percentage of population speaking European language is a valid instrument. Overall, the validity of instruments changes depending on which best-case governance index is used.

¹²BG (2009) cited Rodrik et al. (2004) paper that the available language instruments for large sample do not pass over-identification test, whereas in the previous section, we find that the language instruments are valid for some of the best-case institutional quality indices. However, we continue with BG specification.

Using the percentage of population speaking European language is not always valid and the percentage of population speaking English is also a valid instrument when some of the best-case institutional quality indices are used.

Finally, we replicate the baseline findings of the BG paper (see Table 3 of BG paper for their baseline findings). Our findings confirm BG (2009) findings that a country's own institutions as well as neighboring countries' institutions matter for the development of a country and absolute geography, climate or disease environment, have an indirect effect on economic development through their effect on institutions when the majority of the best-case scenario indices are used as institution proxies. However, when the best-case RQ and VA institutional quality indices are used and the country's own institutions are accounted for, the absolute geography has a direct effect on economic development. On the other hand, when a country's own institutions and neighboring countries' institutions are accounted for and the best-case VA institutional quality index is used, geography has a direct effect on economic development.

7 Conclusion

In this paper we conduct consistent tests for stochastic dominance efficiency at any order of a given equally-weighted index for each governance indicator when it is compared to all possible governance indices constructed from over time levels of governance. We find that the equally-weighted indicator composite indices (i.e. average of each specific governance indicator over time) are not the best-case scenarios for governance. We find that each year contributes differently in the construction of the most optimistic governance indices for each governance measure for each specific sample of countries used in empirical literature (Easterly and Levine 2003; Rodrik et al. 2004; and Bosker and Garretsen 2009).

Moreover, we use the best-case scenario indices as proxies for institutional quality to check for robustness of the results obtained in three prominent applications where the best-case indices are used to assess the link between economic development and institutional quality. With some of the best-case scenario institutional quality proxies, we confirm the findings of the EL (2003) study. However, for some other best-case institutional quality indices, the EL (2003) findings are not robust. We find that for some specifications, the institutions are not significant when the most optimistic governance indices are used, something that contradicts the institutions hypothesis. We also find that policies do matter for certain specifications once some of the best-case institutional quality indices are used. Moreover, endowments are not valid instruments for institutions for all of the best-case institutional quality indices, since we find that endowments do affect economic development through channels that are different than institutional quality.

For the Rodrik et al. (2004) paper, for the samples of 64-countries and 79-countries, we confirm the findings of Rodrik et al. (2004), where institutions “trump” over integration and geography with the exception of the best-case PS index. When the best-case PS index is used, institutions are not significant for the 64 and 79-country samples, a contradiction of the institutions hypothesis. Overall, when some of the best-case institutional quality indices are used, the Rodrik et al. (2004) findings for the large sample case are not robust. We find that once the best-case VA and RQ indices are used, geography has a direct impact on economic development which is consistent with the geography approach. Geography has not only an indirect effect through institutions but also a direct impact on economic development. Moreover, we find that the percentage of population speaking English and the percentage of population speaking European language are valid instruments for institutions for the large sample when some of the best-case institutional quality indices are used.

Finally, when the best-case indices are used as institutional quality proxies, our findings confirm the BG (2009) findings. A country’s own institutions as well as neighboring countries’ institutions matter for the development of that country and absolute geography has an indirect effect on economic development through their effect on institutions. However, once the best-case VA index is used, we find that absolute geography has a direct effect on economic development. Moreover, we find that validity of the instruments used depends on the choice of the best-case scenario index used.

Overall, with many of the best-case scenarios of the institutional quality indices, we confirm the findings in the literature that the institutional quality of a country and that of its neighboring countries are the primary reason for economic development and geography has an indirect effect on economic development through its impact on institutions. However, there are certain cases where we find that not only the country’s own and neighboring institutions matter for economic development but also geography and policies do matter directly for economic development. Finally, valid instruments for institutions vary with the use of different best-case institutional quality indices. One has to be careful with using instruments for institutions since the results vary with different institutional quality indices. Moreover, instruments used in literature are not valid for some cases and instruments which are considered as not being valid are actually found to be valid. One can see that the findings in the literature for the institutional hypothesis are fragile, and therefore more careful research is needed to be taken before policy recommendations are made.

As it was suggested by a referee, there is a methodological caveat with our approach in that we test the standard benchmark of equal weights as adopted and applied in the literature. Yet, the dominating indices when averaged to form a new “optimal index” may generally be

dominated by some other combinations. To improve on that, one can use the efficiency tests by Post (2003), Kuosmanen (2004), Kopa and Chovanec (2008), and Kopa and Post (2011) to identify a dominating alternative. We leave that interesting approach for future research.

Finally, it is important to note that WGIs are updated every year and some earlier governance scores are adjusted over time, and as such the weights assigned to each year for each respective governance index might change. In other words, the best-case governance indices for any given set of countries can vary over time. Moreover, each study uses different set of countries for their analysis and the best-case scenario governance indices for those samples are potentially different. Therefore, the best-case scenario weights need to be re-estimated periodically and/or a robustness analysis with different and full sample size best-case scenarios should be considered in any future analysis.

Acknowledgements This research is part of author's PhD thesis at the Department of Economics of the University of Guelph. The author wishes to thank his PhD supervisor Thanasis Stengos, and committee members, Michael Hoy, Ege Yazgan, Alex Maynard, Ilias Tsiakas, and James Amegashie for their helpful suggestions and comments on improving the paper. The author is also grateful to anonymous referees for their constructive comments and suggestions.

References

- Aidt T, Dutta J, Sena V (2008) Governance regimes, corruption and growth: theory and evidence. *J Comp Econ* 36:195-220
- Acemoglu D, Johnson S, Robinson JA (2001) The colonial origins of comparative development: an empirical investigation. *Am Econ Rev* 91:1369-1401
- Agliardi E, Agliardi R, Pinar M, Stengos T, Topaloglou N (2012) A new country risk index for emerging markets: a stochastic dominance approach. *J Emp Financ* 19:741-761
- Anderson G (1996) Nonparametric tests of stochastic dominance in income distributions. *Econometrica* 64:1183-1193
- Arndt C, Oman C (2006) Uses and abuses of governance indicators. Organization for Economic Cooperation and Development (OECD) Development Center Study.
- Barrett GF, Donald SG (2003) Consistent test for stochastic dominance. *Econometrica* 71:71-104
- Bawa VS, Bodurtha Jr. JN, Rao MR, Suri HL (1985) On determination of stochastic dominance optimal sets, *J Financ* 40:417-431
- Beach C, Davidson R (1983) Distribution-free statistical inference with Lorenz curves and income shares. *Rev Econ Stud* 50:723-735
- Beck T, Laeven L (2006) Institution building and growth in transition economies. *J Econ Growth* 11:157-186
- Bloom DE, Sachs JD (1998) Geography, demography, and economic growth in Africa. *Brookings Pap Econ Act* 2:207-295
- Bosker M, Garretsen H (2009) Economic development and geography of institutions. *J Econ Geogr* 9:295-328
- Brunnschweiler CN (2008) Cursing the blessings? Natural resource abundance, institutions, and economic growth. *World Dev* 36:399-419
- Bulte EH, Damania R, Deacon RT (2005) Resource intensity, institutions and development. *World Dev* 33:1029-1044
- Davidson R, Duclos J-Y (2000) Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica* 68:1435-1464
- Diamond J (1997) *Guns, germs, and steel: the fates of human societies*. W.W. Norton, New York
- Dollar D, Kraay A (2003) Institutions, trade and growth. *J Monet Econ* 50:133-162
- Durlauf SN, Kourtellos A, Tan CM (2008) Are any growth theories robust? *Econ J* 118:329-346
- Engerman SL, Sokoloff, KL (1997) Factor endowments, institutions, and differential paths of

growth among new world economies: a view from economic historians of the United States. In Haber S (ed) *How Latin America fell behind*. Stanford University Press, Stanford, pp 260-304

Easterly W, Levine R (2003) Tropics, germs, and crops: how endowments influence economic development. *J Monet Econ* 50:3-39

Easterly W (2007) Inequality does cause underdevelopment: Insights from a new instrument. *J Dev Econ* 84:755-776

Fishburn PC (1974) Convex stochastic dominance with continuous distribution functions. *J Econ Theory* 7:143-158

Frankel JA, Romer D (1999) Does trade cause growth? *Am Econ Rev* 89:379-399

Gallup JL, Mellinger AD, Sachs JD (1998) Geography and economic development. NBER working paper no. 6849

Glaeser EL, La Porta R, Lopez-de-Silanes F, Shleifer A (2004) Do institutions cause growth? *J Econ Growth* 9:271-303

Hall RE, Jones C (1999) Why do some countries produce so much more output per worker than others? *Q J Econ* 114:83-116

Harms P, Lutz M (2006) Aid, governance and private foreign investment: Some puzzling findings for the 1990s. *Econ J* 116:773-790

Horvath L, Kokoszka P, Zitikis R (2006) Testing for stochastic dominance using the weighted McFadden-type statistic. *J Econometrics* 133:191-205

Kaufmann D, Kraay A, Zoido-Lobaton P (1999) Aggregating governance indicators. World bank policy research working paper 2195

Kaufmann D, Kraay A, Zoido-Lobaton P (2002) Governance matters II - Updated indicators for 2000/01. World bank policy research working paper 2772

Kaufmann D, Kraay A, Mastruzzi M (2004) Governance matters III: Governance indicators for 1996, 1998, 2000, and 2000. *World Bank Econ Rev* 18:253-287

Kaufmann D, Kraay A, Mastruzzi M (2005) Governance matters IV: Governance indicators for 1996-2004. World bank policy research working paper 3630

Kaufmann D, Kraay A, Mastruzzi M (2006) Governance matters V: Aggregate and individual governance indicators for 1996-2005. World bank policy research working paper 4012

Kaufmann D, Kraay A, Mastruzzi M (2007a) Governance matters VI: Aggregate and individual governance indicators for 1996-2006. World bank policy research working paper 4280

Kaufmann D, Kraay A, Mastruzzi M (2007b) The worldwide governance indicators project: Answering the critics. World bank policy research working paper 4149

Kaufmann D, Kraay A, Mastruzzi M (2008) Governance matters VII: Aggregate and individual governance indicators for 1996-2007. World bank policy research working paper 4654

- Knack S (2006) Measuring corruption in Eastern Europe and Central Asia: a critique of the cross-country indicators. World bank policy research working paper 3968
- Kopa M, Chovanec P (2008) A second-order stochastic dominance portfolio efficiency measure. *Kybernetika* 44:243-258
- Kopa M, Post T (2011) A general test for portfolio efficiency. SSRN Working Paper, available at SSRN: <http://ssrn.com/abstract=1824174>
- Kuosmanen T (2004) Efficient diversification according to stochastic dominance criteria. *Manage Sci* 50:1390-1406
- Kurtz MJ, Schrank A (2007) Growth and governance: models, measures and mechanisms. *J Polit* 69:538-554
- Linton O, Whang, Y-J (2012) Testing for the stochastic dominance efficiency of a given portfolio. CEMMAP Working Paper No. CWP27/12
- Mankiw NG, Romer D, Weil DN (1992) A contribution to the empirics of economic growth. *Q J Econ* 107:407-437
- Manly BFJ (1994) *Multivariate statistical methods: a primer*. Chapman & Hall, London
- Nardo M, Saisana M, Saltelli A, Tarantola S, Hoffman A, Giovanni E (2005) *Handbook on constructing composite indicators: methodology and user guide*. OECD Statistics working paper 2005/3
- Neumayer E (2002) Is good governance rewarded? A cross-national analysis of debt forgiveness. *World Dev* 30:913-930
- North DC (1990) *Institutions, institutional change and economic performance*. Cambridge University Press, New York
- Pinar M, Stengos T, Topaloglou N (2013) Measuring human development: a stochastic dominance approach. *J Econ Growth* 18:69-108
- Post T (2003) Empirical tests for stochastic dominance efficiency. *J Financ* 58:1905-1932
- Post T, Versijp P (2007) Multivariate tests for stochastic dominance efficiency of a give portfolio. *J Financ Quant Anal* 42:489-516
- Rodrik D, Subramanian A, Trebbi F (2004) Institutions rule: the primacy of institutions over geography and integration in economic development. *J Econ Growth* 9:131-165
- Sachs JD (2003) Institutions matter, but not for everything. *Financ Dev* 40:38-41
- Sachs JD, Warner AM (1997) Fundamental sources of long-run growth. *American Economic Review* 87:184-188
- Scaillet O, Topaloglou N (2010) Testing for stochastic dominance efficiency. *J Bus Econ Stat* 28:169-180
- Wong W-K, Phoon K, Lean H (2008). Stochastic dominance analysis of Asian hedge funds.

Table 1: Simulations

| Case 1a: $X_i \sim N(\mu_i, 1)$ where $\mu_i = 0.25, 0.2, 0.15, 0.1, 0.05$ and 0 for $i = 1, 2, 3, 4, 5$ and 6 respectively | | | | | | | |
|---|------------------------------|-------------------------------|-------|-------|-------|-------|-------|
| | | X1 | X2 | X3 | X4 | X5 | X6 |
| Distance | Number of dominating indices | Average of dominating indices | | | | | |
| 0.12505 | 494 | 0.422 | 0.285 | 0.122 | 0.068 | 0.062 | 0.041 |

| Case 1b: $X_i \sim N(\mu_i, 1)$ where $\mu_i = 0.15, 0.12, 0.09, 0.06, 0.03$ and 0 for $i = 1, 2, 3, 4, 5$ and 6 respectively | | | | | | | |
|---|------------------------------|-------------------------------|-------|-------|-------|-------|-------|
| | | X1 | X2 | X3 | X4 | X5 | X6 |
| Distance | Number of dominating indices | Average of dominating indices | | | | | |
| 0.07505 | 434 | 0.366 | 0.275 | 0.132 | 0.091 | 0.079 | 0.058 |

| Case 1c: $X_i \sim N(\mu_i, 1)$ where $\mu_i = 0.05, 0.04, 0.03, 0.02, 0.01$ and 0 for $i = 1, 2, 3, 4, 5$ and 6 respectively | | | | | | | |
|---|------------------------------|-------------------------------|-------|-------|-------|-------|-------|
| | | X1 | X2 | X3 | X4 | X5 | X6 |
| Distance | Number of dominating indices | Average of dominating indices | | | | | |
| 0.02411 | 347 | 0.269 | 0.244 | 0.162 | 0.134 | 0.102 | 0.089 |

| Case 1d: $X_i \sim N(0, 1)$ for $i = 1, 2, 3, 4, 5$ and 6 respectively | | | | | | | |
|--|------------------------------|-------------------------------|-----|-----|-----|-----|-----|
| | | X1 | X2 | X3 | X4 | X5 | X6 |
| Distance | Number of dominating indices | Average of dominating indices | | | | | |
| 0.0000 | 0 | n/a | n/a | n/a | n/a | n/a | n/a |

| Case 2: $\mu_i = 0, \sigma_i = 1, 1, 1, 0.8, 0.6, 0.2$ for $i = 1, 2, 3, 4, 5$ and 6 respectively | | | | | | | |
|---|------------------------------|-------------------------------|-------|-------|-------|-------|-------|
| | | X1 | X2 | X3 | X4 | X5 | X6 |
| Distance | Number of dominating indices | Average of dominating indices | | | | | |
| 0.00002 | 41 | 0.030 | 0.024 | 0.026 | 0.199 | 0.337 | 0.384 |

| Case 3: $\mu_i = 0.2, 0, -0.2, 0.2, 0, -0.2$ and $\sigma_i = 1, 1, 1, 0.8, 0.8$ and 0.8 for $i = 1, 2, 3, 4, 5$ and 6 respectively | | | | | | | |
|--|------------------------------|-------------------------------|-------|-------|-------|-------|-------|
| | | X1 | X2 | X3 | X4 | X5 | X6 |
| Distance | Number of dominating indices | Average of dominating indices | | | | | |
| 0.01998 | 420 | 0.249 | 0.095 | 0.053 | 0.394 | 0.129 | 0.080 |

Table 2: Data and Descriptive Statistics

| Voice and Accountability | | | | | | | | | | |
|---------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----------------|
| | 1996 | 1998 | 2000 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | Composite Index |
| Sample | 193 | 193 | 193 | 193 | 193 | 193 | 193 | 193 | 193 | 193 |
| Mean | -0.05 | -0.05 | -0.05 | -0.05 | -0.05 | -0.04 | -0.05 | -0.05 | -0.05 | -0.05 |
| Median | -0.04 | -0.09 | -0.10 | -0.09 | -0.03 | -0.08 | -0.07 | -0.02 | -0.06 | -0.08 |
| Skewness | -0.16 | -0.04 | -0.06 | -0.10 | -0.09 | 0.00 | -0.05 | -0.19 | -0.20 | -0.06 |
| Std. Dev. | 1.02 | 1.01 | 1.02 | 1.01 | 1.02 | 1.02 | 1.01 | 1.01 | 1.01 | 1.00 |

| Political Stability | | | | | | | | | | |
|----------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----------------|
| | 1996 | 1998 | 2000 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | Composite Index |
| Sample | 179 | 179 | 179 | 179 | 179 | 179 | 179 | 179 | 179 | 179 |
| Mean | -0.11 | -0.11 | -0.11 | -0.11 | -0.12 | -0.14 | -0.14 | -0.13 | -0.12 | -0.12 |
| Median | 0.04 | -0.04 | -0.03 | -0.02 | -0.06 | -0.08 | -0.06 | -0.03 | 0.04 | -0.08 |
| Skewness | -0.67 | -0.57 | -0.44 | -0.31 | -0.30 | -0.40 | -0.39 | -0.49 | -0.63 | -0.36 |
| Std. Dev. | 1.01 | 1.01 | 1.01 | 1.00 | 1.00 | 0.97 | 0.97 | 0.98 | 0.98 | 0.95 |

| Government Effectiveness | | | | | | | | | | |
|---------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----------------|
| | 1996 | 1998 | 2000 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | Composite Index |
| Sample | 181 | 181 | 181 | 181 | 181 | 181 | 181 | 181 | 181 | 181 |
| Mean | -0.04 | -0.04 | -0.03 | -0.04 | -0.03 | -0.04 | -0.05 | -0.04 | -0.05 | -0.04 |
| Median | -0.25 | -0.23 | -0.19 | -0.29 | -0.26 | -0.25 | -0.28 | -0.23 | -0.29 | -0.26 |
| Skewness | 0.80 | 0.44 | 0.51 | 0.60 | 0.55 | 0.58 | 0.46 | 0.52 | 0.49 | 0.60 |
| Std. Dev. | 0.99 | 1.00 | 1.00 | 1.01 | 1.01 | 1.02 | 1.01 | 1.01 | 1.01 | 0.98 |

| iv) Regularity Quality | | | | | | | | | | |
|-------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----------------|
| | 1996 | 1998 | 2000 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | Composite Index |
| Sample | 182 | 182 | 182 | 182 | 182 | 182 | 182 | 182 | 182 | 182 |
| Mean | -0.05 | -0.05 | -0.04 | -0.05 | -0.05 | -0.06 | -0.06 | -0.05 | -0.04 | -0.05 |
| Median | 0.08 | -0.10 | -0.10 | -0.17 | -0.20 | -0.26 | -0.24 | -0.18 | -0.21 | -0.20 |
| Skewness | -0.58 | -0.21 | -0.08 | 0.12 | 0.10 | 0.15 | 0.07 | 0.06 | 0.06 | 0.03 |
| Std. Dev. | 0.99 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.96 |

| Rule of Law | | | | | | | | | | |
|--------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----------------|
| | 1996 | 1998 | 2000 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | Composite Index |
| Sample | 170 | 170 | 170 | 170 | 170 | 170 | 170 | 170 | 170 | 170 |
| Mean | -0.08 | -0.08 | -0.10 | -0.11 | -0.11 | -0.11 | -0.13 | -0.13 | -0.12 | -0.11 |
| Median | -0.24 | -0.25 | -0.32 | -0.36 | -0.40 | -0.29 | -0.38 | -0.48 | -0.44 | -0.41 |
| Skewness | 0.36 | 0.33 | 0.38 | 0.46 | 0.44 | 0.42 | 0.46 | 0.49 | 0.49 | 0.47 |
| Std. Dev. | 1.01 | 1.01 | 1.01 | 1.01 | 1.03 | 1.03 | 1.02 | 1.02 | 1.02 | 1.00 |

| vi) Control for Corruption | | | | | | | | | | |
|-----------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----------------|
| | 1996 | 1998 | 2000 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | Composite Index |
| Sample | 153 | 153 | 153 | 153 | 153 | 153 | 153 | 153 | 153 | 153 |
| Mean | -0.02 | -0.02 | -0.01 | -0.01 | 0.00 | -0.04 | -0.05 | -0.04 | -0.04 | -0.03 |
| Median | -0.27 | -0.34 | -0.31 | -0.33 | -0.34 | -0.35 | -0.36 | -0.33 | -0.39 | -0.35 |
| Skewness | 0.71 | 0.66 | 0.67 | 0.78 | 0.72 | 0.73 | 0.76 | 0.81 | 0.91 | 0.81 |
| Std. Dev. | 1.06 | 1.05 | 1.06 | 1.07 | 1.06 | 1.07 | 1.05 | 1.05 | 1.05 | 1.04 |

| Indicator | Number of observations | Number of iterations | 1996 | 1998 | 2000 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
|-----------|------------------------|----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| VA | 72 | 6 | 0.965 | 0.000 | 0.000 | 0.000 | 0.000 | 0.035 | 0.000 | 0.000 | 0.000 |
| PS | 72 | 5 | 0.000 | 0.032 | 0.966 | 0.000 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 |
| GE | 71 | 5 | 0.001 | 0.061 | 0.011 | 0.000 | 0.861 | 0.066 | 0.000 | 0.000 | 0.000 |
| RQ | 72 | 5 | 0.989 | 0.003 | 0.000 | 0.000 | 0.001 | 0.005 | 0.001 | 0.001 | 0.000 |
| RL | 72 | 6 | 0.992 | 0.000 | 0.000 | 0.000 | 0.002 | 0.004 | 0.002 | 0.000 | 0.000 |
| CC | 63 | 7 | 0.078 | 0.000 | 0.130 | 0.008 | 0.775 | 0.009 | 0.000 | 0.000 | 0.000 |

| Quartile | Governance level | Number of observations with the equally-weighted index that are above a given governance level | Number of observations with the best-case scenario index that are above a given governance level |
|---------------------------------|------------------|--|--|
| Voice and accountability | | | |
| Q1 | -0.872 | 56 | 57 |
| Q2 | -0.220 | 37 | 38 |
| Q3 | 0.329 | 19 | 21 |
| Political stability | | | |
| Q1 | -1.240 | 56 | 59 |
| Q2 | -0.320 | 37 | 39 |
| Q3 | 0.180 | 19 | 23 |
| Government effectiveness | | | |
| Q1 | -0.834 | 56 | 57 |
| Q2 | -0.395 | 36 | 38 |
| Q3 | 0.065 | 18 | 19 |
| Regulatory quality | | | |
| Q1 | -0.670 | 56 | 58 |
| Q2 | -0.271 | 37 | 48 |
| Q3 | 0.270 | 19 | 28 |
| Rule of law | | | |
| Q1 | -0.902 | 56 | 58 |
| Q2 | -0.514 | 37 | 41 |
| Q3 | 0.098 | 19 | 21 |
| Control of corruption | | | |
| Q1 | -0.824 | 47 | 48 |
| Q2 | -0.414 | 31 | 33 |
| Q3 | 0.114 | 15 | 17 |

| Indicator | Number of observations | Number of iterations | 1996 | 1998 | 2000 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
|-----------|------------------------|----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| VA | 64 | 4 | 0.016 | 0.000 | 0.000 | 0.000 | 0.006 | 0.972 | 0.000 | 0.000 | 0.006 |
| PS | 64 | 4 | 0.003 | 0.014 | 0.920 | 0.009 | 0.000 | 0.052 | 0.002 | 0.000 | 0.000 |
| GE | 64 | 5 | 0.000 | 0.020 | 0.006 | 0.001 | 0.896 | 0.075 | 0.002 | 0.000 | 0.000 |
| RQ | 64 | 3 | 0.957 | 0.026 | 0.000 | 0.000 | 0.004 | 0.010 | 0.000 | 0.000 | 0.003 |
| RL | 64 | 5 | 0.970 | 0.011 | 0.000 | 0.000 | 0.000 | 0.015 | 0.004 | 0.000 | 0.000 |
| CC | 64 | 6 | 0.020 | 0.000 | 0.008 | 0.000 | 0.954 | 0.005 | 0.013 | 0.000 | 0.000 |

| Quartile | Governance level | Number of observations with the equally-weighted index that are above a given governance level | Number of observations with the best-case scenario index that are above a given governance level |
|---------------------------------|------------------|--|--|
| Voice and accountability | | | |
| Q1 | -0.873 | 48 | 52 |
| Q2 | -0.347 | 32 | 38 |
| Q3 | 0.322 | 16 | 18 |
| Political stability | | | |
| Q1 | -1.104 | 48 | 51 |
| Q2 | -0.345 | 32 | 34 |
| Q3 | 0.167 | 16 | 19 |
| Government effectiveness | | | |
| Q1 | -0.816 | 48 | 50 |
| Q2 | -0.414 | 32 | 35 |
| Q3 | 0.066 | 16 | 17 |
| Regulatory quality | | | |
| Q1 | -0.625 | 48 | 52 |
| Q2 | -0.259 | 32 | 46 |
| Q3 | 0.271 | 16 | 29 |
| Rule of law | | | |
| Q1 | -0.939 | 48 | 51 |
| Q2 | -0.539 | 32 | 35 |
| Q3 | 0.022 | 16 | 20 |
| Control of corruption | | | |
| Q1 | -0.835 | 48 | 49 |
| Q2 | -0.468 | 32 | 33 |
| Q3 | -0.051 | 16 | 18 |

Table 5A: Rodrik et al. (2004) Extended Acemoglu Sample best-case scenario governance indices for each indicator

| Indicator | Number of observations | Number of iterations | 1996 | 1998 | 2000 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
|-----------|------------------------|----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| VA | 79 | 5 | 0.943 | 0.000 | 0.000 | 0.003 | 0.000 | 0.048 | 0.003 | 0.003 | 0.000 |
| PS | 79 | 5 | 0.000 | 0.000 | 0.003 | 0.000 | 0.000 | 0.995 | 0.000 | 0.002 | 0.000 |
| GE | 78 | 5 | 0.000 | 0.006 | 0.000 | 0.017 | 0.843 | 0.134 | 0.000 | 0.000 | 0.000 |
| RQ | 79 | 6 | 0.998 | 0.000 | 0.000 | 0.000 | 0.001 | 0.001 | 0.000 | 0.000 | 0.000 |
| RL | 79 | 6 | 0.992 | 0.000 | 0.000 | 0.000 | 0.004 | 0.000 | 0.002 | 0.000 | 0.002 |
| CC | 70 | 7 | 0.001 | 0.000 | 0.000 | 0.004 | 0.984 | 0.005 | 0.004 | 0.000 | 0.002 |

Table 5B: Comparisons between evaluated and best-case scenario index for Extended Acemoglu sample

| Quartile | Governance level | Number of observations with the equally-weighted index that are above a given governance level | Number of observations with the best-case scenario index that are above a given governance level |
|---------------------------------|------------------|--|--|
| Voice and accountability | | | |
| Q1 | -1.017 | 60 | 63 |
| Q2 | -0.354 | 40 | 45 |
| Q3 | 0.300 | 20 | 22 |
| Political stability | | | |
| Q1 | -1.146 | 60 | 65 |
| Q2 | -0.345 | 40 | 41 |
| Q3 | 0.181 | 20 | 21 |
| Government effectiveness | | | |
| Q1 | -0.878 | 60 | 63 |
| Q2 | -0.427 | 40 | 42 |
| Q3 | -0.069 | 20 | 23 |
| Regulatory quality | | | |
| Q1 | -0.688 | 60 | 61 |
| Q2 | -0.312 | 40 | 52 |
| Q3 | 0.171 | 20 | 34 |
| Rule of law | | | |
| Q1 | -0.989 | 60 | 63 |
| Q2 | -0.540 | 40 | 41 |
| Q3 | 0.001 | 20 | 22 |
| Control of corruption | | | |
| Q1 | -0.876 | 52 | 53 |
| Q2 | -0.468 | 34 | 35 |
| Q3 | -0.044 | 17 | 20 |

| Indicator | Number of observations | Number of iterations | 1996 | 1998 | 2000 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
|-----------|------------------------|----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| VA | 137 | 5 | 0.004 | 0.000 | 0.000 | 0.000 | 0.000 | 0.995 | 0.001 | 0.000 | 0.000 |
| PS | 137 | 7 | 0.000 | 0.001 | 0.991 | 0.005 | 0.002 | 0.001 | 0.000 | 0.000 | 0.000 |
| GE | 136 | 5 | 0.172 | 0.163 | 0.361 | 0.303 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 |
| RQ | 137 | 4 | 0.282 | 0.567 | 0.037 | 0.025 | 0.033 | 0.037 | 0.005 | 0.000 | 0.014 |
| RL | 137 | 6 | 0.980 | 0.020 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| CC | 125 | 7 | 0.178 | 0.022 | 0.581 | 0.217 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 |

| Quartile | Governance level | Number of observations with the equally-weighted index that are above a given governance level | Number of observations with the best-case scenario index that are above a given governance level |
|---------------------------------|------------------|--|--|
| Voice and accountability | | | |
| Q1 | -0.866 | 105 | 108 |
| Q2 | -0.097 | 69 | 70 |
| Q3 | 0.740 | 36 | 37 |
| Political stability | | | |
| Q1 | -0.924 | 105 | 111 |
| Q2 | -0.090 | 69 | 75 |
| Q3 | 0.511 | 36 | 41 |
| Government effectiveness | | | |
| Q1 | -0.720 | 105 | 110 |
| Q2 | -0.235 | 69 | 72 |
| Q3 | 0.651 | 36 | 37 |
| Regulatory quality | | | |
| Q1 | -0.576 | 105 | 112 |
| Q2 | -0.156 | 69 | 80 |
| Q3 | 0.695 | 36 | 39 |
| Rule of law | | | |
| Q1 | -0.832 | 105 | 106 |
| Q2 | -0.329 | 69 | 80 |
| Q3 | 0.730 | 36 | 37 |
| Control of corruption | | | |
| Q1 | -0.756 | 94 | 98 |
| Q2 | -0.313 | 63 | 68 |
| Q3 | 0.718 | 31 | 33 |

Table 7A: Bosker and Garretsen (2009) IV Selection Sample best-case scenario governance indices for each indicator

| Indicator | Number of observations | Number of iterations | 1996 | 1998 | 2000 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
|-----------|------------------------|----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| VA | 79 | 6 | 0.940 | 0.000 | 0.000 | 0.000 | 0.000 | 0.058 | 0.002 | 0.000 | 0.000 |
| PS | 79 | 6 | 0.004 | 0.000 | 0.002 | 0.000 | 0.000 | 0.989 | 0.000 | 0.005 | 0.000 |
| GE | 78 | 4 | 0.072 | 0.024 | 0.065 | 0.000 | 0.640 | 0.199 | 0.000 | 0.000 | 0.000 |
| RQ | 79 | 6 | 0.996 | 0.000 | 0.000 | 0.000 | 0.001 | 0.001 | 0.001 | 0.001 | 0.000 |
| RL | 79 | 5 | 0.980 | 0.000 | 0.000 | 0.000 | 0.000 | 0.014 | 0.006 | 0.000 | 0.000 |
| CC | 69 | 6 | 0.001 | 0.000 | 0.000 | 0.003 | 0.992 | 0.003 | 0.002 | 0.000 | 0.000 |

Table 7B: Comparisons between evaluated and best-case scenario index for IV Selection Sample

| Quartile | Governance level | Number of observations with the equally-weighted index that are above a given governance level | Number of observations with the best-case scenario index that are above a given governance level |
|---------------------------------|------------------|--|--|
| Voice and accountability | | | |
| Q1 | -0.959 | 60 | 63 |
| Q2 | -0.354 | 40 | 46 |
| Q3 | 0.300 | 20 | 22 |
| Political stability | | | |
| Q1 | -1.146 | 60 | 67 |
| Q2 | -0.355 | 40 | 42 |
| Q3 | 0.181 | 20 | 21 |
| Government effectiveness | | | |
| Q1 | -0.909 | 60 | 65 |
| Q2 | -0.458 | 40 | 41 |
| Q3 | -0.069 | 20 | 23 |
| Regulatory quality | | | |
| Q1 | -0.714 | 60 | 61 |
| Q2 | -0.312 | 40 | 53 |
| Q3 | 0.171 | 20 | 33 |
| Rule of law | | | |
| Q1 | -1.012 | 60 | 63 |
| Q2 | -0.550 | 40 | 42 |
| Q3 | 0.001 | 20 | 23 |
| Control of corruption | | | |
| Q1 | -0.876 | 52 | 53 |
| Q2 | -0.468 | 34 | 35 |
| Q3 | -0.044 | 17 | 20 |

Table 8A: Bosker and Garretsen (2009) Large Sample best-case scenario governance indices for each indicator

| Indicator | Number of observations | Number of iterations | 1996 | 1998 | 2000 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
|-----------|------------------------|----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| VA | 146 | 6 | 0.988 | 0.000 | 0.000 | 0.000 | 0.000 | 0.012 | 0.000 | 0.000 | 0.000 |
| PS | 144 | 8 | 0.000 | 0.847 | 0.153 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| GE | 146 | 4 | 0.465 | 0.175 | 0.181 | 0.053 | 0.037 | 0.069 | 0.010 | 0.003 | 0.007 |
| RQ | 147 | 4 | 0.182 | 0.336 | 0.401 | 0.000 | 0.000 | 0.034 | 0.034 | 0.007 | 0.006 |
| RL | 137 | 6 | 0.055 | 0.931 | 0.000 | 0.003 | 0.003 | 0.000 | 0.000 | 0.005 | 0.003 |
| CC | 122 | 7 | 0.998 | 0.000 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 |

Table 8B: Comparisons between evaluated and best-case scenario index for Large Sample

| Quartile | Governance level | Number of observations with the equally-weighted index that are above a given governance level | Number of observations with the best-case scenario index that are above a given governance level |
|---------------------------------|------------------|--|--|
| Voice and accountability | | | |
| Q1 | -0.804 | 109 | 110 |
| Q2 | 0.001 | 73 | 76 |
| Q3 | 0.810 | 38 | 40 |
| Political stability | | | |
| Q1 | -0.924 | 109 | 116 |
| Q2 | -0.015 | 73 | 74 |
| Q3 | 0.700 | 38 | 41 |
| Government effectiveness | | | |
| Q1 | -0.672 | 109 | 112 |
| Q2 | -0.196 | 73 | 74 |
| Q3 | 0.560 | 38 | 40 |
| Regulatory quality | | | |
| Q1 | -0.576 | 109 | 115 |
| Q2 | -0.113 | 73 | 83 |
| Q3 | 0.666 | 38 | 40 |
| Rule of law | | | |
| Q1 | -0.812 | 105 | 109 |
| Q2 | -0.210 | 69 | 73 |
| Q3 | 0.783 | 34 | 37 |
| Control of corruption | | | |
| Q1 | -0.756 | 91 | 92 |
| Q2 | -0.268 | 61 | 64 |
| Q3 | 0.825 | 31 | 32 |