

The Tyranny of International Index Rankings*

Bjørn Høyland

Karl Moene

Fredrik Willumsen[†]

April, 2009

Abstract

International index rankings are popular, but dangerous. They stimulate rank-seeking behavior by emphasizing country differences where similarity is dominant. We concentrate the discussion around the popular Human Development Index, Freedom House, and Doing Business. Most of the rankings in these three indexes appear to be misleading, not because of wrong indicators, but because the estimation of the scores ignores inherent uncertainty. When we re-estimate them with a method that captures this uncertainty, it is clear that the practice of comparing adjacent countries is a rather courageous activity.

“Looking-glass upon the wall, Who is fairest of us all?”

1 Introduction

One can hardly open a newspaper without finding a reference to an international index. International country rankings provide an instant idea of the relative success of a country vis-à-vis other countries in the world. Their appeal lies in their simplicity. Their users need no more statistical knowledge than readers of the sports pages in the newspapers. Just as boxers, baseball players, and national football teams are ranked according to their performance, countries are ranked according to their ability to provide a high standard of living, democratic rights, and an appealing business environment. Just as pundits use sport rankings to place their bets for the weekend, journalists use country rankings in their search for an easy way to finish their Saturday commentary and policy makers use the country indexes to guide their decisions over own policy and evaluate other countries.

*We thank Marie Ingvaldsen, Carl Henrik Knutsen, Jo Thori Lind, and seminar participants at the PRIO brownbag seminar for useful comments. While carrying out this research, we have been associated with the ESOP centre at the Department of Economics, University of Oslo. ESOP is supported by The Research Council of Norway.

[†]Høyland: Department of Political Science, University of Oslo. Moene and Willumsen: Department of Economics, University of Oslo. E-mail addresses: bjorn.hoyland@stv.uio.no, k.o.moene@econ.uio.no, f.h.willumsen@econ.uio.no

It seems like we are blessed with a tool that everybody can understand and that is appropriate for a wide range of purposes. If only . . .

Our criticism of international country rankings concerns their accuracy. We single out three indexes for scrutiny: the United Nations' *Human Development Index*, the Freedom House's *Freedom in the World*, and the World Bank's *Doing Business Index*. The three are chosen not because they are constructed in a particularly bad manner, but because they cover important aspects of social, economic, and political life, and because they are popular and widely noticed. The Human Development Index ranks countries according to health, knowledge, and material resources; Doing Business ranks countries on the basis of how 'business-friendly' the regulatory environment is; and Freedom House creates an annual index of the political rights and civil liberties enjoyed in different countries.

The rankings are precisely reported, widely publicized, and closely followed also when ranking is meaningless because of the high level of uncertainty. We claim that each of the rankings draws attention to one specific, but rather arbitrary, ranking within large groups of countries. Much of the focus can be on differences where similarity is dominant. Paradoxically, the popularity of the indexes may simply be a reflection of this weakness as people seem obsessed by differences when everything is more or less on an even level. The most exciting index rankings may therefore be the most uncertain, and the attention that such rankings receive in the international community is often inversely related to their accuracy. This misplaced emphasis is a consequence of the tyranny of international index rankings.

While we do not believe that all involved fully believe that the rankings portrait an accurate picture, a particular favorable or unfavorable position on one of these rankings are most likely to be widely publicized and therefore impossible to ignore. Governments stand to lose by not observing the rankings and commenting upon them. To attack an index is never appropriate for them. When the ranking is unfavorable, an attack would just make things worse; when the ranking is favorable the praise is too tempting.

In Norway, for instance, prime ministers from different political parties regularly insist that the United Nations has chosen Norway as the best country to live in—based on its position in the Human Development Index (at least prior to 2007). To give an indication of the hidden uncertainty in the human development rankings, let us consider the four countries Iceland, Norway, Canada, and Australia, denoted INCA. The four countries were the top 4 countries on the Human Development Index in 2007. The observed performance indicators, we suggest, are uncertain signals of the underlying quality, here "Human Development". It is easy to calculate, given the results in section 5, that of all countries in the world, the four are more likely to belong to the top 10 human development performers than not. The index, however, rank them as top 4 in the order INCA in 2007, conveying the picture that Iceland is best, and thus clearly better than Norway, which is clearly better than Canada, and so on.

How interesting is this ranking compared to the information, neither conveyed nor discussed, that the four countries have about the same chances of being among the top 10 and that none of them have a higher chance than $3/4$ of being in this group? In this case the probability that INCA is the “true” top rank order must be extremely low. Just the probability that all four are among the top 10 is less than $1/3$, and if all sequences of the four countries are almost equally likely, the probability that INCA is the true top ranking must be around one percent or so.

Similarly for the other indexes: media, policy makers and researchers often end up discussing the deep causes of a slight alteration in the internal rankings. What is even more problematic, policy makers may design policies more to improve their rankings than to improve their real performance. Governments may be tempted to engage in what we denote “rank-seeking” behavior to improve the relative standing on the indexes more than the situation on the underlying phenomena.

For example, the director-general of the Malaysian Industrial Development Authority was quoted saying that “Malaysia aims to move from the 24th to a top 10 position in the World Bank’s ‘Doing Business’ ranking list. We continue to ask ourselves what it will take to reach the top 10, and are we willing to do what it takes to get there.” (Asia in Focus, Jan. 8 2007). Similarly, Macedonia placed a one-page advert in the Economist’s annual forecasting report “The World in 2008”, where the key message was that Macedonia had improved their position on the Doing Business index.

These policymakers may be misguided. Human development, political freedom, and good business environment are all performances that cannot be measured directly, and any measure is bound to contain noise. How can we infer the quality of performance in a manner that takes the inherent uncertainty into account?

The obvious idea is to start out from a set of observable country outcomes that one thinks are partial signals of the relevant performance of the countries vis-à-vis each other. Next, one can aggregate them into a number or an index for each country. Finally, one rank countries by this index. This is the approach that the three indexes follow. They do not, however, consider the important question as to how much uncertainty is carried over from *the noisy signals* of performance to the aggregate index score and the rankings based on that index score.¹

There are several sources of noise in the rankings. Some of them we ignore in order to put the indexes in the most favorable light, and to err on the safe side with respect to the measured uncertainty. We therefore assume that the indicators chosen are the correct ones,² that they are measured accurately, and that there is only one single underlying

¹For other weaknesses of relying on rankings see the overview in Søreide (2006).

²The indexes may or may not incorporate the most relevant information, but we will not enter into the quarrel over which variables should be included. There has been a great deal of discussions about this for all the indexes used in this paper; for the Human Development Index see for example Sagar and Najam (1998), for Freedom House see among others Munck and Verkuilen (2002), and for Doing Business

dimension. The remaining uncertainty, that indicators are noisy signals of performance, is the one that we focus on. Our assessment is therefore a lower bound on the true uncertainty. We do, however, check whether the data support the view that the performance measure is one-dimensional. For the Human Development Index and Freedom House this view is supported, while for Doing Business the picture is more blurry.

But how can we observe the uncertainty in the signals? Imagine that we ranked every country on each of the available indicators for a particular index. If it were the case that we obtained the same ranking in all of these sub-rankings, then the indicators would simply be deterministic representations of the same underlying performance variable, and a ranking based on these would therefore be exact. If the rankings differed a lot across the sub-rankings, however, it would make more sense to think of the indicators as stochastic representations of the same underlying performance variable, and the level of variability across the rankings based on the sub-indicators would be a potential measure of the degree of the uncertainty contained in the signals of the underlying performance variable. This is the approach we pursue in this paper; the uncertainty we estimate is based on the observations that the countries differ in their rankings on the different sub-indicators, that translate into uncertainty over what their overall ranking really is.

More specifically, the goal of this paper is to estimate the uncertainty contained in the index scores and rankings using a Bayesian latent variable approach. Each of the indicators are considered as signals of the underlying performance level. Our emphasis is on how we can extract the underlying level of performance and the inherent uncertainty from these signals. The general problem of measuring unobserved performance and how the indexes have solved this problem is described in section 2. We use a latent variable approach to assess the uncertainty in the index scores. The approach is presented in section 3. We then give a brief description of the data in section 4. The main results of the paper are the estimated uncertainty contained in the index scores and the rankings for the different indexes. We also evaluate to what extent the different indexes really express different things: Do countries with a business friendly environment tend to score high on political freedom and human development? Are the empirical relationships between human development, political freedom and business-friendliness tight or lax according to the indexes? Is there a tendency in the indexes that good things come together? These results are presented in section 5. Finally, section 6 concludes.

2 The problem: measuring unobserved performance

To avoid unnecessary abstractions we illustrate how to interpret the indexes by focusing on one of them, the Human Development Index (HDI). It should be noted, however, that HDI

see for example Lee, McCann, and Torm (2008) and Hampel-Malagrosa and Frickenstein (2008). There are also problems with comparing these indexes over time, see for example Morse (2003).

is perhaps the least suited of the indexes for the basic principle in our approach, that the indicators are uncertain signals of the same underlying performance measure. Despite this, we choose to use HDI for illustration as it is the simplest and most transparent index of the three, and, for good and bad, therefore gives a stark representation of our approach.

The Human Development Index emphasizes three different “aspects” of human development: a long and healthy life, knowledge, and a decent standard of living. How should we interpret the index?

2.1 Two approaches

In contrast to our approach, we could insist that the HDI actually *is* human development—or that human development *is* the HDI. The index makers have approached the experts, one might imagine, who have told them that HDI is it. But the most prominent scholars on the topic, Anand and Sen (1994, p. 1), who also are leading proponents of human development, motivate the index quite differently when they claim that “[h]uman beings are the real end of all activities, and development must be centered on enhancing their achievements, freedom, and capabilities. It is the lives that they lead that is of intrinsic importance, not the commodities or incomes they happen to possess.” Accordingly, human development must be about how people can lead good lives and develop to their full potential. It is therefore related to self-worth, dignity and creativity—none of which can be completely defined by the rise or fall of national incomes, life expectancy, and schooling. Clearly, the assertion ‘HDI equals human development’ cannot be a reasonable interpretation. The observed factors that enter the index, however, may of course be correlated with human development in one way or another.

One way that the observed factors can be correlated with human development is via their links to the immediate *causes* of good performance. An ambitious way to try to capture the underlying performance would be to model the economic, social, and political mechanisms that determine income, longevity, and schooling, and to base the comparisons across countries on a ranking of how human-development-friendly the underlying mechanisms are. In doing this one would of course have to incorporate mutual interdependencies of the three dimensions: Health and knowledge are inputs to the production of the material standard of living; a high material standard of living can help finance health and education; there is a political economy and a governance structure that affect all three dimensions, and so on.

Another way that the observed factors can be correlated with human development is via their links to the immediate *consequences* of good or bad human development. One way to proceed would then be to consider the available indicators of the the three dimensions: a long and healthy life can be measured by life expectancy at birth; knowledge can be measured by literacy rates and school enrollment; a decent standard of living

can be measured by per capita income of the country. This is the approach chosen by the index-makers of the Human Development Index. Then the index goes further by collapsing the three dimensions into one aggregate number for each country, with an obvious loss of information. Yet this aggregation can be justified exactly by the existence of mutual interdependencies between the different dimensions, but instead of trying to model the interdependencies one try to tap the information that each indicator contains about the unmeasurable human development performance, here by taking the average of the performance on the different indicators. Taking this view on what the indexes measure hence amounts to viewing each indicator as an uncertain signal of the underlying variable (“human development”) that they all are partial observations of. We build on this interpretation of what the indexes measure, but instead of just taking the average of the performance on the different indicators, we try to tap the information that the different indicators contain on the underlying variable in an optimal way.

3 Our approach: measuring latent variables

As already stated observable indicators can be thought of as signals of the underlying performance. In this section we show how we can use the different indicators to trace out the underlying performance. We use the Human Development Index as an example when discussing the model and the estimation; what we show hold true for the other indexes as well.

3.1 Model and identification

Let the true, unobserved performance of country i be denoted by θ_i , and let the observable outcomes or indicators be denoted by x_{ij} , where j denotes one of several outcomes. It is reasonable to think of x_{ij} as a *signal* of the unobserved θ_i , a signal which hopefully can be used to make inferences on the unobserved performance of country i (relative to country k).

Consider the following specification of how the true, unobserved performance is related to the observable indicators:

$$x_{ij} = \alpha_j + \beta_j \theta_i + \epsilon_{ij}, \quad \text{var}(\epsilon_{ij}) = \sigma_{\epsilon,j}^2, \quad (1)$$

where α_j and β_j are constants and ϵ_{ij} is an error term. Clearly, in this specification the parameter β_j is important; it measures index j 's ability to discriminate between countries with respect to the latent variable θ_i . As we have specified it, the indicators observed for each country are related via the common underlying level of human development in the country, θ_i . The goal is to use this relationship to try to tap the different observed indicators for the information they contain on θ_i .

By looking at equation (1), we realize that everything that is on the right-hand side of the equation is unknown; although this looks like an equation readily set out for estimation via Ordinary Least Squares it is not, as the “explanatory variable” θ_i is not observed. So how can we estimate α, β, σ^2 , and θ using only information contained in x ?³ Intuitively, we are going to exploit the correlation between the different indicators across countries to estimate α_j, β_j and σ_j^2 , and use the within-country variation to estimate the different θ_i 's. For the model to be identified, however, we first need to assume a structure of the underlying distribution of performance (the θ_i 's) across countries. We assume that the performance is distributed with mean 0 and variance 1. It should be noted that this is not nearly as restrictive as it may look like at first sight; as only relative performance matter and this performance does not have any natural scale, we are free to normalize it as we wish.

Since we exploit the correlation between the different indicators to estimate the weights that should be put on them in estimating the index, the β 's, it is important to note that the error terms in equation (1) must not be correlated. Correlation within the error terms on two or more indicators will make us put more weight on these indicators relative to the others, as the correlation will make us think that these indicators tap heavily from the underlying variable θ_i , while in fact they only tap information from the correlated error term. This implies that it is important to have indicators that are unrelated to each other other than through the underlying quality variable, θ_i . By looking at the data we have at hand, it is not obvious that this conditional independence assumption, that, conditional of θ_i the observed indicators are unrelated, holds. One example is the Human Development Index. The Human Development Index have three different classes of measures, schooling, life expectancy, and GDP per capita, but have two schooling indicators (enrollment and literacy rates) while only one indicator in the two other groups. Hence, by just using the four raw variables directly, we will not take into account that two of the variables may be more closely related than the others for reasons that are not contained in θ_i . Fortunately, it is straightforward to take such grouping effects into account, by augmenting equation (1) with a group-specific error term $\delta_{i,k(j)}$:

$$x_{i,j} = \alpha_j + \beta_j \theta_i + \delta_{i,k(j)} + \epsilon_{ij}, \quad \text{var}(\epsilon_{ij}) = \sigma_{\epsilon,j}^2, \quad \text{var}(\delta_{i,k(j)}) = \sigma_{\delta,k(j)}^2, \quad (2)$$

where δ is a group k -specific shock that creates correlation between the indicators j in group k for reasons other than θ_i . As we estimate variance of the group-specific shock, $\sigma_{\delta,k(j)}^2$, we of course still allow for the possibility that, even though a priori it seemed reasonable that the indicators were clustered, the variance of the group-specific shock is estimated to be 0.

³ α, β, σ^2 denote the vectors of respectively $\alpha_j, \beta_j, \sigma_j^2$ for all j , θ denote the vector of θ_i 's, x_i denote the vector of the observable indicators for country i , and finally x denote the matrix of the observed indicators over all countries.

A key issue is of course whether we can identify all the parameters in (2) with the data we have. We will use the Human Development Index to show that we in fact have identification. As will be clear below, if we have identification for the Human Development Index, we will also have identification for the other models. The reason for this is that the extra moments we get by adding more indicators more than make up for the the extra parameters that need to be estimated when we add more indicators.

Again, the Human Development Index contains four measures of performance: school enrollment, literacy rate, GDP per capita, and life expectancy. Assume now for simplicity that school enrollment and literacy rate are contained in the variables x_{i1} and x_{i2} , GDP per capita in x_{i3} , and life expectancy in x_{i4} . Exploiting that $\text{var}(\theta) = 1$ by assumption, the covariance matrix is

$$\text{cov}(x_i) = \begin{pmatrix} \beta_1^2 + \sigma_{\epsilon,1}^2 + \sigma_\delta^2 & & & & \\ \beta_2\beta_1 + \sigma_\delta^2 & \beta_2^2 + \sigma_{\epsilon,2}^2 + \sigma_\delta^2 & & & \\ \beta_3\beta_1 & \beta_3\beta_2 & \beta_3^2 + \sigma_{\epsilon,3}^2 & & \\ \beta_4\beta_1 & \beta_4\beta_2 & \beta_4\beta_3 & \beta_4^2 + \sigma_{\epsilon,4}^2 & \end{pmatrix}$$

where x_i is the vector consisting of the four variables in the Human Development Index and σ_δ^2 is the covariance among the error terms for the two schooling measures. We immediately see that by using the empirical covariance matrix, we are able to identify β, σ_ϵ^2 , and σ_δ^2 .⁴ From the mean of the variables we then identify α . Finally, using the identified parameters and x_i , we can identify θ_i .⁵

The statistical approach just described can be contrasted to the one applied by the index makers. Again we use the Human Development Index as an illustration. By using similar notation, one should ask how the index makers infer the unobserved performance from the observed indicators, i.e. how do they choose α_j and β_j ? The Human Development Index is constructed using the average of three development outcomes for each country: health, as measured by life expectancy; knowledge and skills, as measured by a weighted average of functional literacy and combined elementary and secondary net school-enrollment rates; and resources, as measured by the level of real per capita income. The formula for calculating the index is

$$\theta_i = \frac{1}{3} \sum_{j=1}^3 \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})}$$

⁴We have ten unique elements in the covariance matrix, and nine parameters ($\beta_1, \beta_2, \beta_3, \beta_4, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2, \sigma_\delta^2$). In this case we hence have one over-identifying restriction.

⁵It should be noted that we only have *local* identification, meaning that if both β and θ are multiplied by -1 the results will be the same. This is entirely unproblematic; as previously stated the scale of the performance is arbitrary, we can therefore normalize it such that the best performers have positive values and the poorest performers have negative values, remembering that we have normalized θ such that $\text{var}(\theta) = 1$ and $E(\theta) = 0$.

To cast this formula in terms of the model in equation (1), the HDI parameters β_j and α_j for $j = 1, 2, 3$ are set as follows

$$\beta_j = \max x_j - \min x_j, \quad \alpha_j = \min x_j, \quad \sum_{j=1}^3 \frac{\epsilon_{ij}}{\max x_j - \min x_j} = 0,$$

where the max and the min in principle are the highest and lowest observed outcomes across all countries.⁶

First note that the three sub-indexes, and hence also the aggregate index, are bounded between 0 and 1. By normalizing the observed variables x_{ij} such that maximum value of each of the sub-indexes is 1 and the minimum value 0, it readily follows that the approach behind the Human Development Index is basically to set α_j equal to 0 and $\beta_j = 3$, i.e. what the Human Development Index implicitly assumes is that all the sub-indicators are equally good at discriminating between countries. Again, our approach is fundamentally different, as we try to estimate the discrimination parameters from the available data.

3.2 Continuous, ordinal, count, and censored outcomes

In the presentation of the model in the previous section, we assumed that the indicators were continuous variables. By looking at the indicators for the different indexes, the full list of indicators for all the indexes is given in Appendix B, we immediately realize that this is not a valid assumption. Several of the indicators are as if they are censored, an example is the literacy rate in the Human Development Index, some of them are of an ordinal nature, for example the indicator for “Rule of law” in Freedom House, while others are counts, such as the number of documents needed for importing in Doing Business. It is important to take these properties of the indicators into account when we estimate the model, as a misspecification such as fitting a model suitable for continuous variables on an ordinal indicator may give rise to greater estimated uncertainty in the index scores and the rankings.

So how do we take this into account in the estimation? For the ordinal variables, we follow the approach in Treier and Jackman (2008) and Clinton and Lewis (2008), and estimate an ordered choice model. We also here take into account within-group correlation for the different variables as described in section 3.1. For the variables that are as if they were censored, we modify the likelihood function to take this censoring into account, following Gelman and Hill (2006, p. 404-5). Finally, the count data is modeled using a Poisson regression type model with overdispersion.

⁶The Human Development Index does not always use the observed max and min from the data, but rather rely on so-called “goalposts”, a theoretical max/min for the different indicators. They also use the natural logarithm rather than the level of GDP per capita. For simplicity, we still refer to this variable as GDP per capita. For more information, see Technical Note 1 in HDR (2007).

3.3 Estimation

We use a Bayesian approach to estimate the model. The goal of the estimation is to characterize the joint posterior probability distribution over all the parameters in the model given the observed indicators, i.e. to estimate $p(\Theta|x)$ where $\Theta = (\alpha, \beta, \sigma, \theta)$. This posterior probability distribution is, according to Bayes theorem, proportional to the product of the likelihood of the data given the parameters and the prior distribution over the parameters: $p(\Theta|x) \propto p(x|\Theta)p(\Theta)$. We use non-informative priors on all the parameters; the priors we use are given in Table 2 in Appendix A.2. By using non-informative priors, we put no a priori restrictions on the values that the parameters can or are likely to take, and the identification of the parameters therefore rely solely on the data through the likelihood function $p(x|\Theta)$. Clearly, the likelihood function $p(x|\Theta)$ comes from the model we specify for how the observed indicators are linked to the underlying performance and the assumed distribution of the error terms. The model we estimate is, as previously stated, an augmented version of the model from Section 3.1; all details about the model are given in Appendix A. Again, the reason that we augment the model to take censoring, ordinal outcomes, and count data into account, is that the precision with which we can estimate the underlying quality scores will depend crucially on whether the model not misspecified, that is whether the likelihood function we use is the correct one.

We use a Markov Chain Monte Carlo (MCMC) algorithm implemented in WinBUGS⁷ to calculate the posterior from the likelihood function and the priors. The MCMC algorithm, after it has converged, provides us with a random sample from the joint posterior distribution. This sample is what we use for statistical inferences, for example for construction credibility intervals as in Figures 1, 3, and 5 and for testing hypotheses as in Figures 2, 4, and 6. For brevity, all computational details are relegated to Appendix A.

4 Data and definitions

We briefly describe the data here; in Appendix B we provide more details on the data and how we classify the different variables.

The Human Development Index

The Human Development Index contains four indicators: life expectancy at birth, school enrollment, literacy rates, and GDP per capita. School enrollment and literacy rates are on a 0-100 scale, while life expectancy is in years, and GDP per capita is in 2005 PPP US \$.

⁷WinBUGS is available at <http://www.mrc-bsu.cam.ac.uk/bugs/>.

Freedom House

Freedom House collects information about political rights and civil liberties. The “Political Rights” pillar has three ordinal scaled indicators: the electoral process, political pluralism and participation, and the functioning of government. The “Civil Liberties” pillar has four ordinal scaled indicators: freedom of expression and belief, associational and organizational rights, rule of law, and personal autonomy and individual rights. All indicators are based on questionnaires.⁸

Doing Business

The scores on the Doing Business indicators are based on a survey of international companies operating in the major economic center of a country. The indicators include information on a variety of aspects, from the degree of investor protection to the amount protection that workers enjoy.

More precisely, Doing Business collect indicators on 10 different stages of a business’s life: starting a business, dealing with licenses, employing workers, registering property, getting credit, protecting investors, paying taxes, trading across borders, enforcing contracts, and closing a business. Each of these sub-indexes has one or more indicators. The sub-index for starting a business, for example, contains a variable counting the number of procedures required for starting a business, the number of days it takes to start a business, cost (as % of income per capita), and minimum capital requirement (as % of income per capita). The other sub-indexes contain similar variables; all the variables in the different groups are given in Appendix B.

5 Results: uncertain rankings

To see how well each of the indexes distinguishes among the countries, we estimate the underlying performance and its corresponding uncertainty, i.e. how much uncertainty is carried over from the noisy signals to index scores and rankings.

5.1 Rankings

The rankings are based on the index scores. As our critique of the indexes is on their focus on rankings, we first present the uncertainty in the rank positions for all countries, before illustrating the uncertainty in the rankings by looking at the probabilities that particular countries are among the top or bottom ten countries on the indexes.⁹ We start with the

⁸Unfortunately, Freedom House does not release the underlying data material used to generate the indicators. This is unfortunate for us, as this info could have helped reduce the estimated uncertainty in our rankings.

⁹The actual scores are presented in Appendix C.

Human Development Index, followed by Freedom House and Doing Business.

The Human Development Index

As previously stated, the Human Development Index consists of 4 indicators: Life expectancy at birth, school enrollment, literacy rates, and (log of) GDP per capita. School enrollment and literacy rates have a maximum value at 100, and several countries obtain the maximum score on these indicators. We take this censoring into account in the estimation procedure, see section 3.2 and Appendix A. As stated in section 3.1, it is problematic for our method of inference if the indicators we use are correlated for reasons other than the underlying performance measure (here “Human Development”). For the Human Development Index, this is a problem for the two schooling-related indicators. The Human Development Index solves this issue by assigning fixed weights to each of the indicators: Life expectancy at birth and (log of) GDP per capita are given a weight of $1/3$ each, while school enrollment and literacy rates are given respectively a weight of $2/9$ and $1/9$ each. As we estimate these weights, we have to take this correlation among the error terms into account in a different way: the method we use is to allow for a group-specific error term and estimate the variance of the group-specific error component. More details on the identification of this parameter is given in section 3.1.

The ranking based on the re-estimation of the index score is given in Figure 1. We see that the Human Development Index distinguishes well among the countries at the lower end of the ranking; the Democratic Republic of Congo and the Central African Republic are clearly doing worse than most other countries in terms of Human Development. As a device to distinguish between the 100 least developed countries in the world, the Human Development Index is quite successful.

At the other end of the scale, we find Australia, Canada, Norway, and Iceland. It should be noted that there are tiny differences, if any, between the countries at the top of the ranking. Nevertheless, in pack-patting celebratory political speeches, leaders of advanced economies brag about their rank among nations on the Human Development Index. The latter years have seen the throne being passed among Canada, Iceland, and Norway.

For less developed countries, however, it seems that the indicators do a better job in distinguishing the countries. This is also seen in Figure 2. In the Figure we display the countries for which we cannot reject the hypothesis that they are among the respectively top and bottom 10 countries in the world in terms of Human Development at a 5 per cent level.¹⁰ While 31 countries have at least a 5 per cent chance of being among the top 10 countries in Human Development, 22 countries have at least a 5 per cent chance of being

¹⁰Since we are using non-informative priors, the probabilities can also be seen as the p-values of the classical test H_0 : the country is among the top/bottom ten versus the alternative hypothesis that it is not.

Uncertainty in HDI ranking

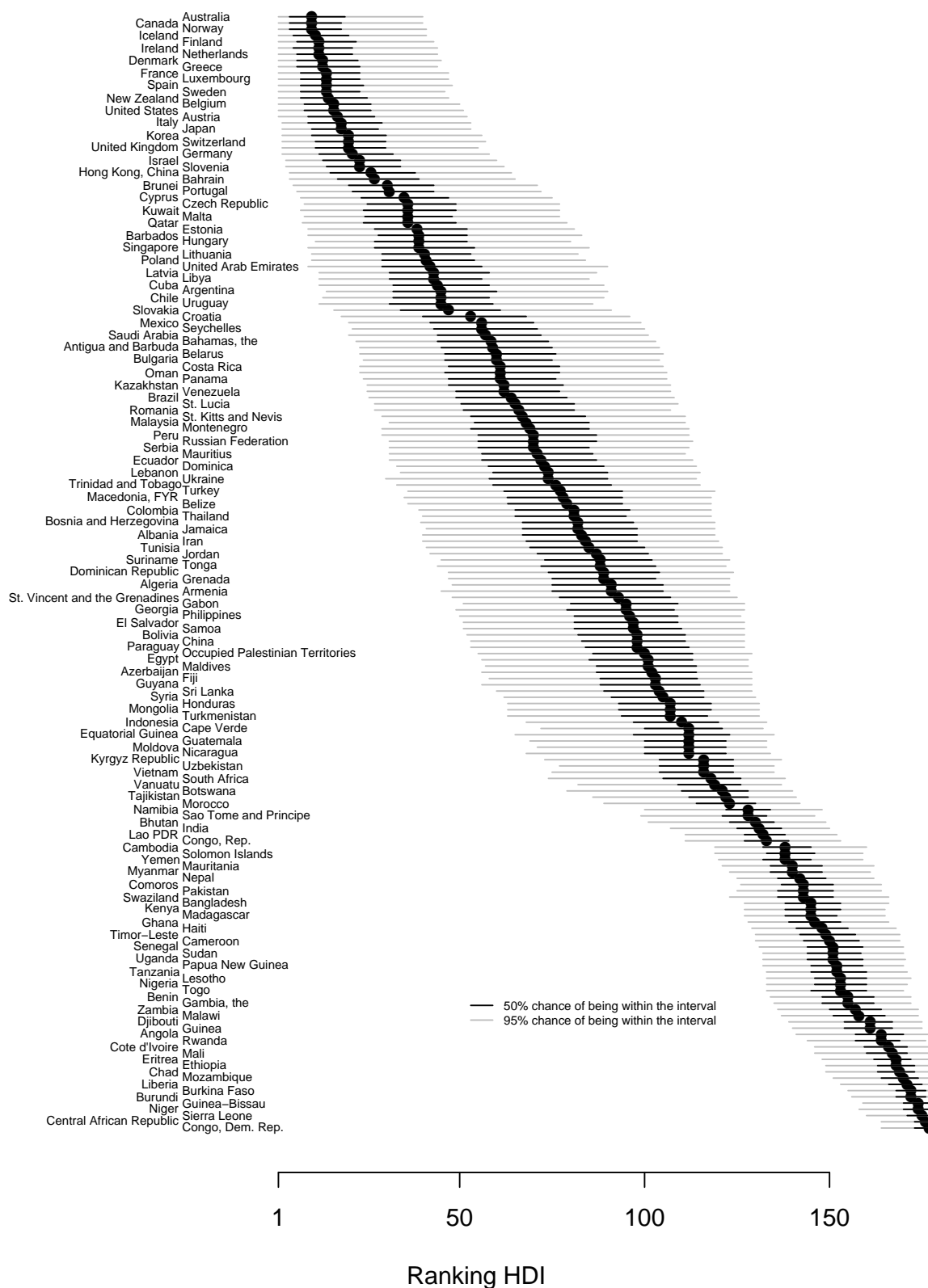
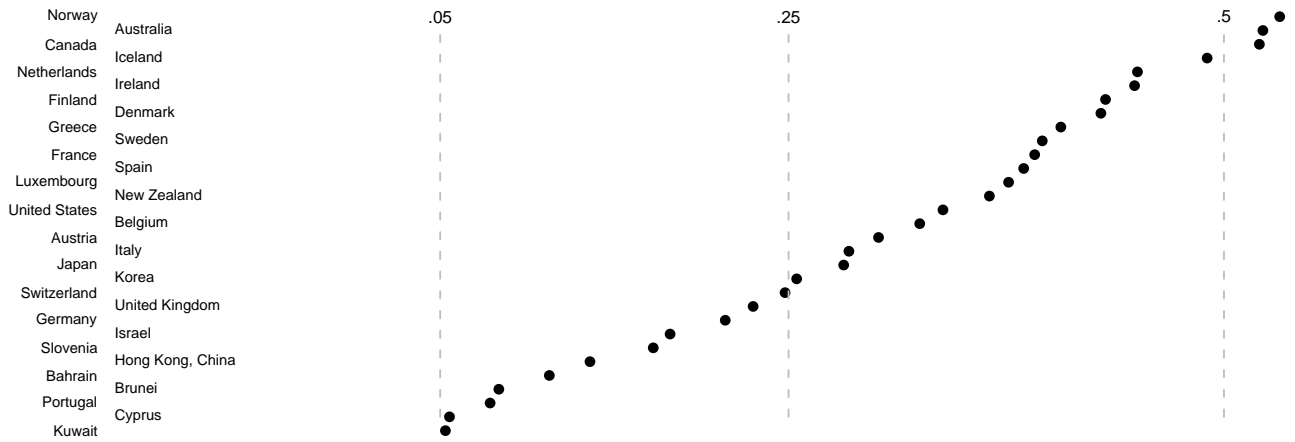


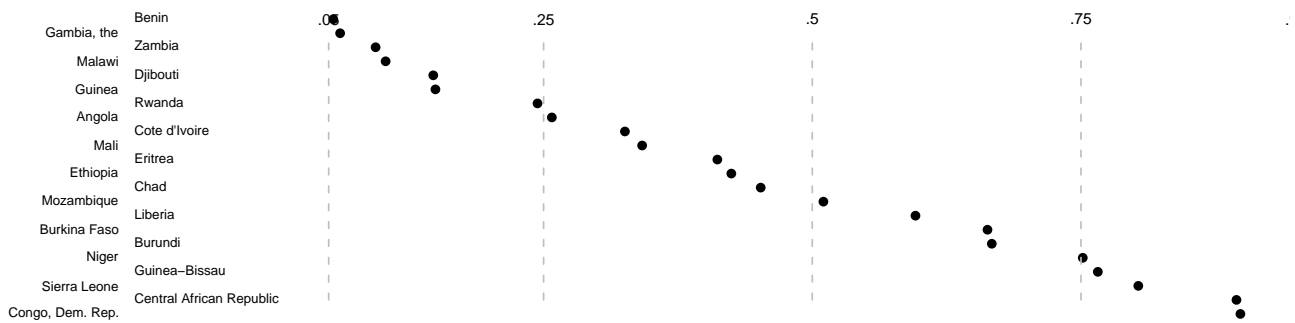
Figure 1: Estimated rankings for the Human Development Index. Posterior inferences. The black circle indicates the median ranking, the black line gives the rank position interval for which the country is more likely to fall inside than outside, and the gray line indicate the ranking interval for which the country is more than 95 per cent certain to be located within.

HDI – Probability of being among top 10



No. of countries: 31

HDI – Probability of being among bottom 10



No. of countries: 22

Figure 2: Rank uncertainty for the Human Development Index: countries that have more than a 5 per cent chance of being among the respectively top and bottom 10.

among the bottom 10 countries on the Human Development Index. No country has more than a 75 per cent chance of being among the top 10, while 5 countries have more than a 75 per cent chance of being among the bottom 10. The Central African Republic and the Democratic Republic of Congo are the worst countries out, with an almost 95 per cent chance of being among the 10 least developed countries in the world.

Freedom House

Next, we consider Freedom House, an index commonly used to measure levels of democracy. Freedom House report values on 7 ordinal variables. As previously stated, we estimate an ordinal response model for the Freedom House data, taking group-specific correlation into account.

The rankings based on the estimated index scores for Freedom House are given in Figure 3. Freedom House distinguishes better along a wider range of the underlying dimension than the Human Development Index does. It is only at the very top of the scale that the index clearly fails to distinguish between countries. There is a group of 7 countries that, for all practical purposes are identical in the underlying dimension. The reason for this is that these 7 countries all score the maximum value on all the indicators that go into the index. For this reason it is very hard to know how good these countries really are, and the estimated uncertainty cannot be ignored.

Figure 4 shows that 17 countries have at least a 5 per cent chance of being among the 10 most democratic countries in the world, while 16 countries have at least a 5 per cent chance of being among the 10 least democratic countries in the world. Freedom House is hence capable of narrowing down the top and bottom countries fairly accurately. The least free countries in the world are Turkmenistan, North Korea, Uzbekistan, and Somalia. The probability that North Korea is among the 10 least free countries in the world is estimated to be very close to 1.

Doing Business

Finally, we consider Doing Business, the index on the quality of the regulatory environments around the world. The results are presented in Figure 5. Our estimates suggest that there is considerable uncertainty around the ranking, in particular for countries in the middle 80 per cent of the scale. In contrast to the key message of the precise ranking published in the Doing Business report, it is clear that the index does not do a very good job in distinguishing between most of the regulatory environments in the world. While the rankings, after taking uncertainty into account, clearly distinguishes the best economies from the worst, it does not distinguish particularly well between the economies that are somewhat in between. There is a large group of more than 100 countries, among which it is almost impossible to identify any differences.

As Figure 6 shows, 20 countries have at least a 5 per cent chance of providing one of the top ten business environments in the world. There are nine countries with more than a 50 per cent chance of belonging to this group, of which four countries—Singapore, New Zealand, United States, and Hong Kong—are almost surely among the top ten. Figure 6 also reveals that the situation at the bottom mirrors the one at the top: 20 countries have at least a 5 per cent chance of having among the 10 least inviting business environments in the world. Nine countries have more than a 50 per cent chance of belonging to this group, of which four countries—Congo, Zimbabwe, Iraq, and Chad—are almost surely among the 10 economies with the worst business environments in the world.

Uncertainty in Freedom House ranking

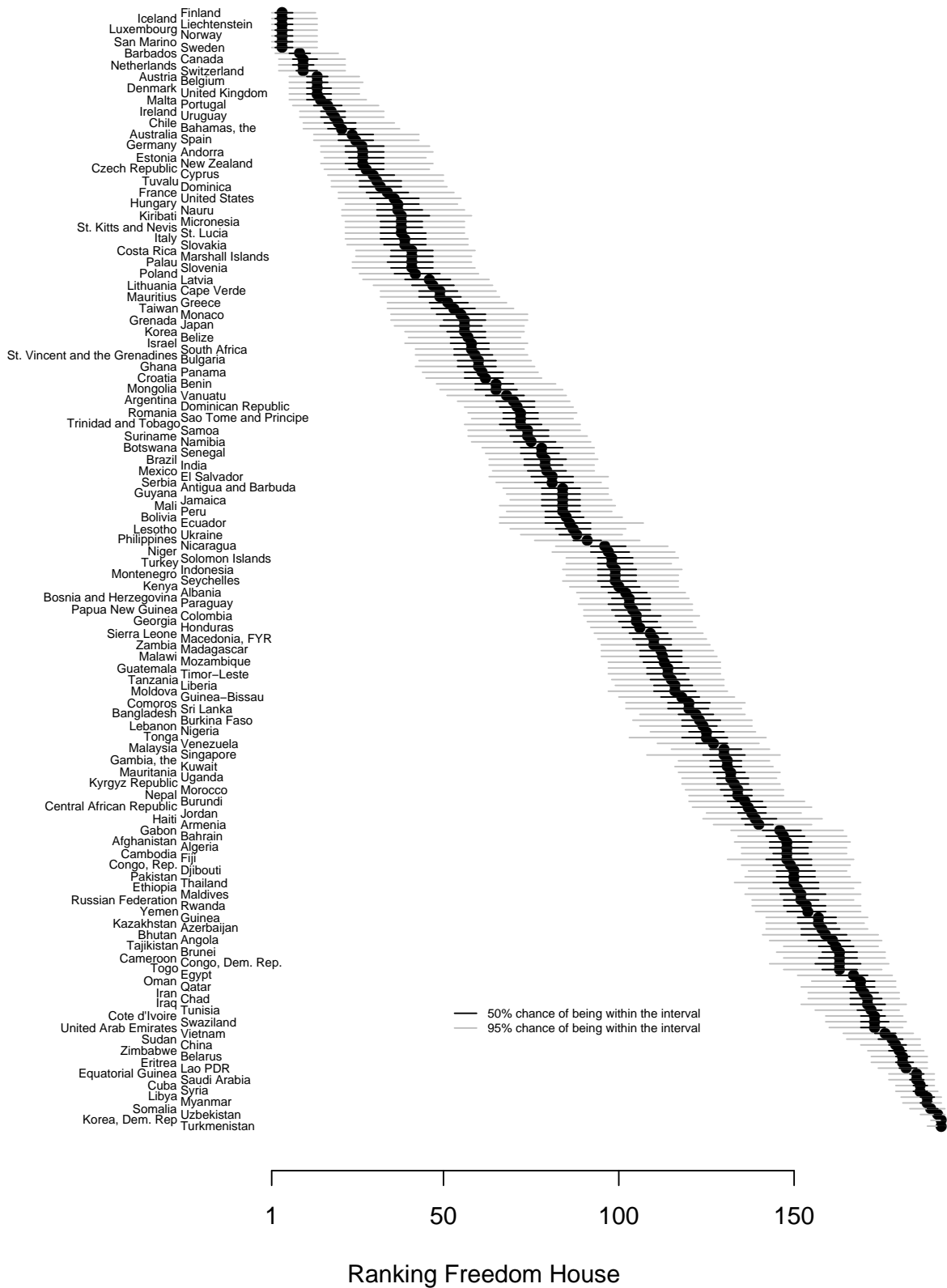


Figure 3: Estimated rankings for Freedom House.

Posterior inferences. The black circle indicates the median ranking, the black line gives the rank position interval for which the country is more likely to fall inside than outside, and the gray line indicate the ranking interval for which the country is more than 95 per cent certain to be located within.

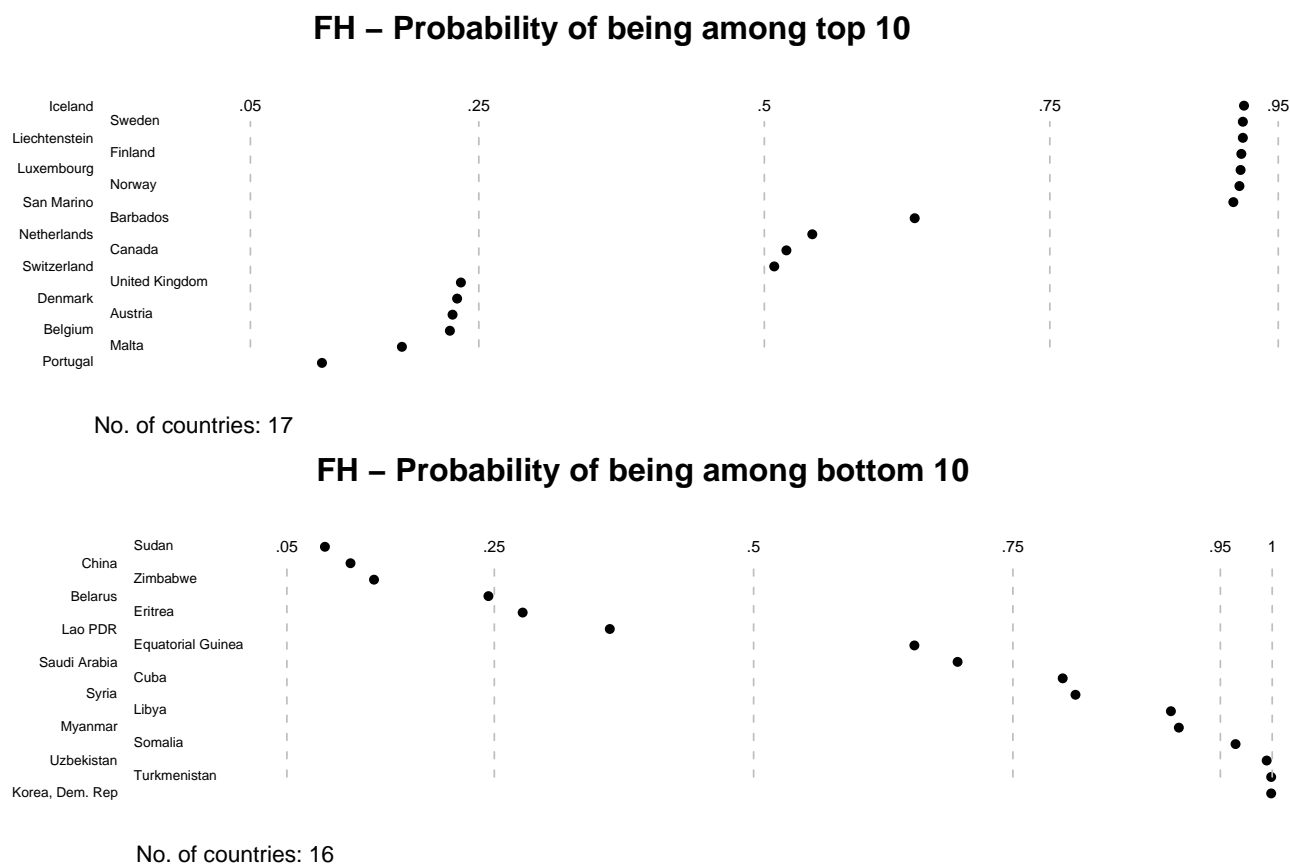


Figure 4: Rank uncertainty for Freedom House: countries that have more than a 5 per cent chance of being among the respectively top and bottom 10.

5.2 Do the indexes capture the same thing?

Figure 7 shows the correlation between the rankings on the different indexes. The position of the countries on each of the indexes is evaluated using the median of the estimated rankings for each of the countries. The plots on the diagonal of the matrix show the density of the median rankings for respectively Doing Business, Freedom House, and the Human Development Index. The off-diagonal elements shows scatterplots between the different indexes. The dots are median rankings for each country, the solid line is a non-parametric (lowess) regression, and the dashed line is a linear regression line.

First, we note that although the correlations between all the different indexes are positive, there is substantial spread in the degree of correlation. For all pairs, the correlation is much stronger for higher rankings (closer to 1) than for lower rankings. The extreme case is the correlation between the Human Development Index and Freedom House. These are highly correlated at the upper end of the scale (closer to 1, i.e. in the bottom left corner of the figure). But as demonstrated by the flatness and even turning of the local regression line, there is no or negative correlation between these two measures at the lower half of the scale. The correlations between Doing Business and the other two indexes are positive

Uncertainty in Doing Business ranking

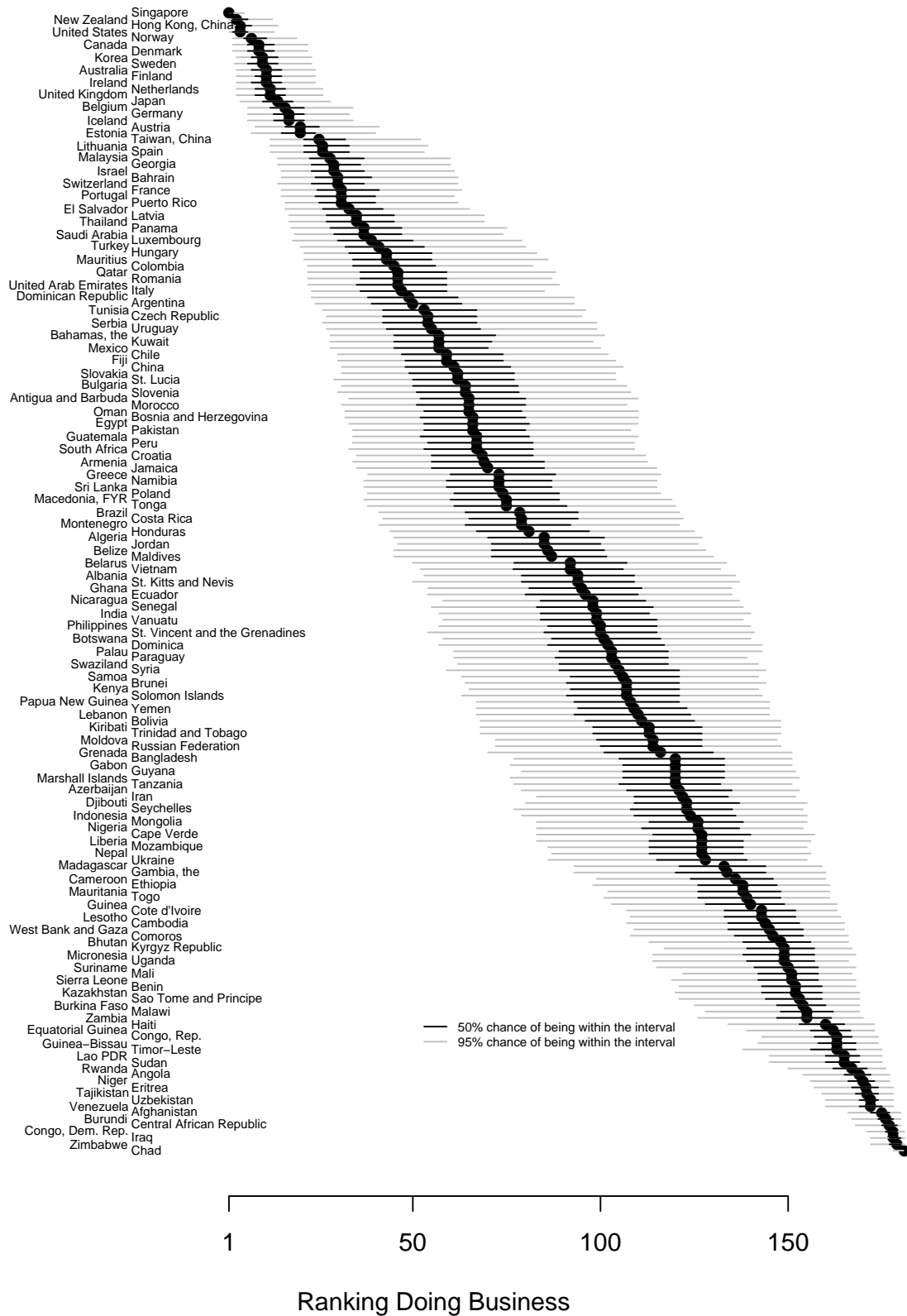


Figure 5: Estimated rankings for Doing Business.

Posterior inferences. The black circle indicates the median ranking, the black line gives the rank position interval for which the country is more likely to fall inside than outside, and the gray line indicate the ranking interval for which the country is more than 95 per cent certain to be located within.

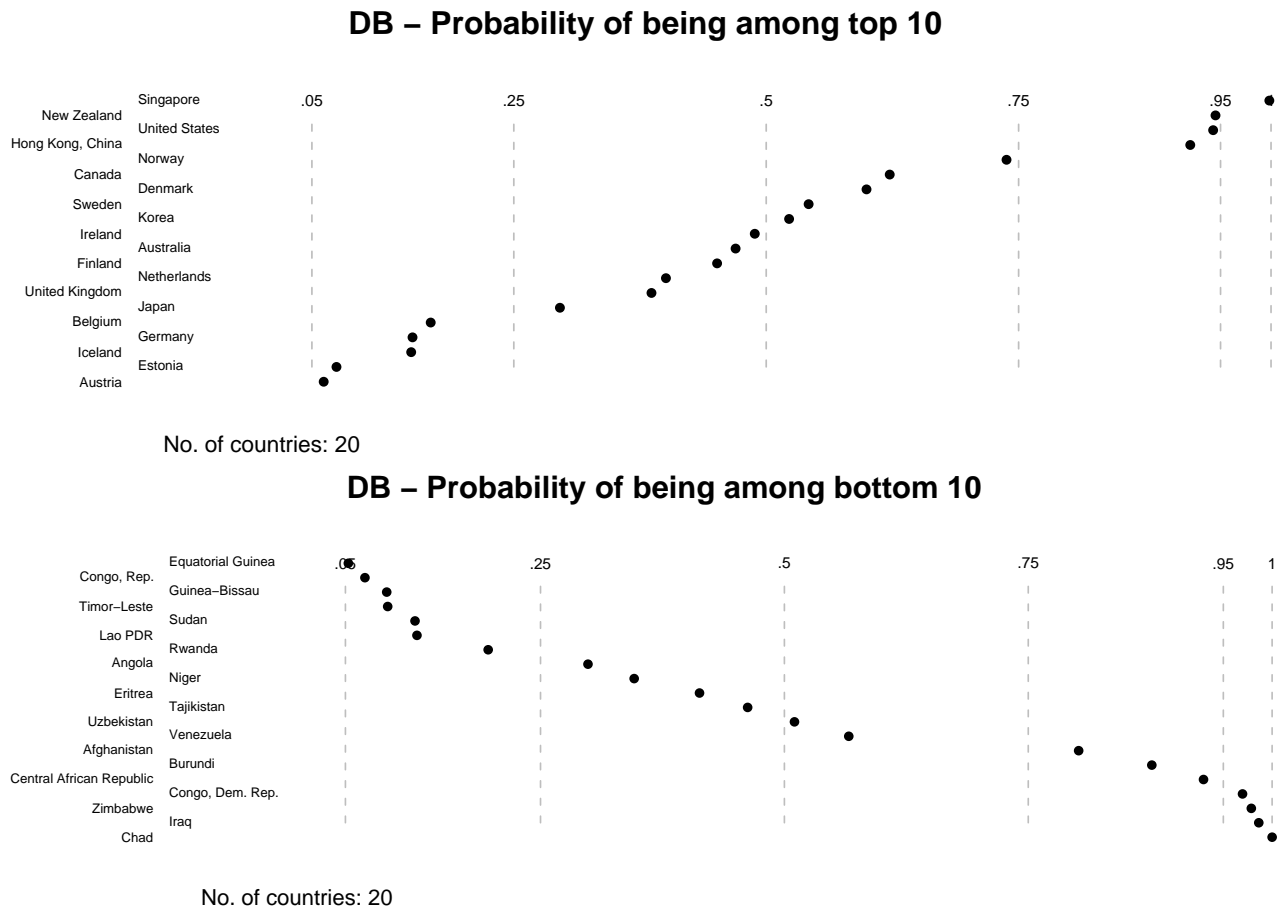


Figure 6: Rank uncertainty for Doing Business: countries that have more than a 5 per cent chance of being among the respectively top and bottom 10.

in both instances, although the positive correlation is much sharper in Doing Business versus the Human Development Index than between Doing Business and Freedom House.

5.3 One or several dimensions?

Finally, as the three indexes are far from perfect in distinguishing between countries we check whether the data support the fitting of a one-dimensional model. We use standard principal components analysis to check whether the data really support the use of one single index to capture the performance of the countries, or whether performance is in fact many-dimensional.

Table 1 gives a summary of a principal components analysis on the three indexes. Row 1 in the three sub-tables contains the square root of the eigenvalue of the correlation matrix of the indicators for each index (denoted “Standard deviation” in the table), a measure of the amount of variation explained by the component in question. A standard rule of thumb, first advocated by Kaiser (1960), is to select the number of components such that we only retain components that have eigenvalues above 1. This is the same as

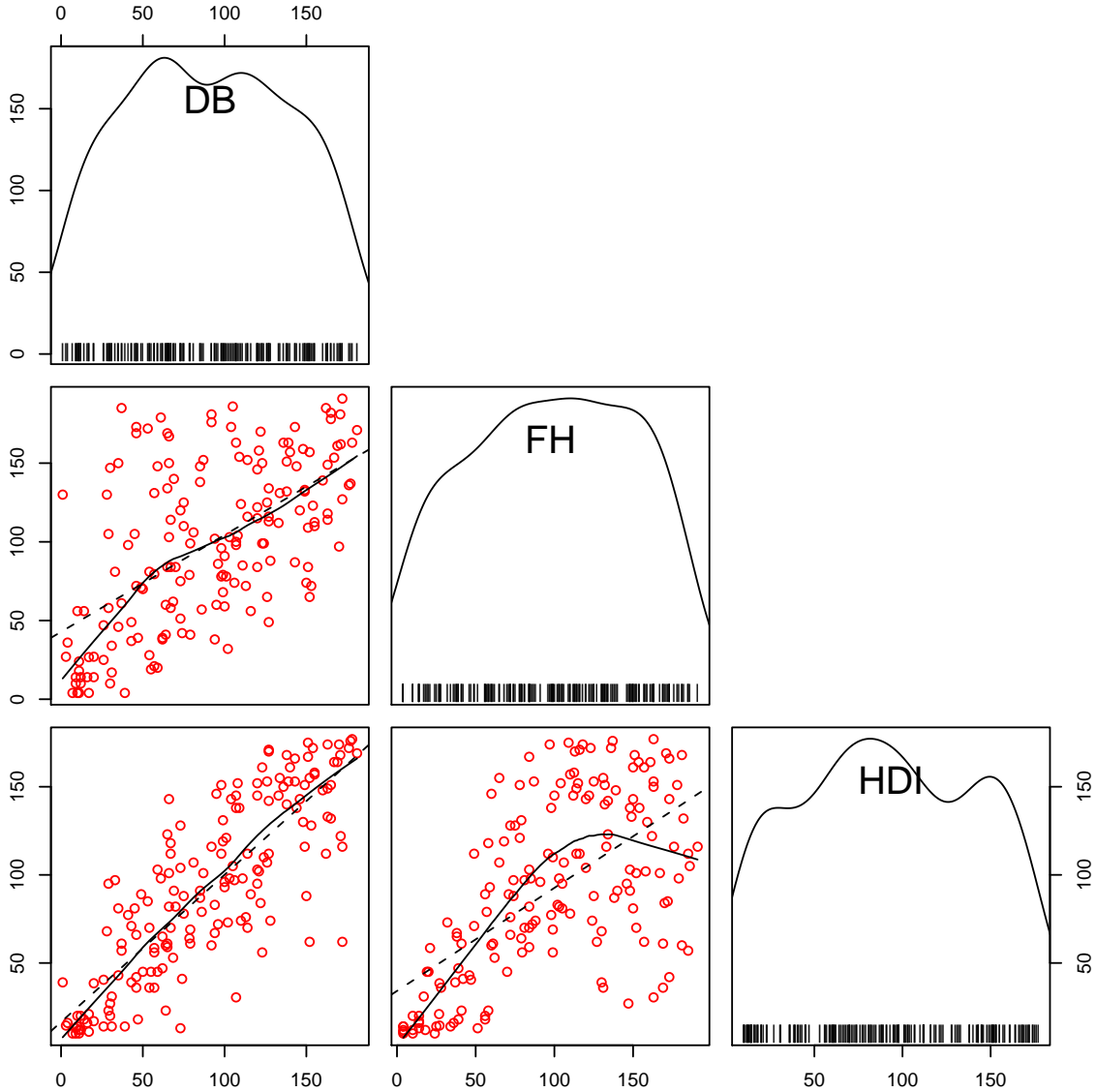


Figure 7: Scattermatrix for the rankings on the different indexes.

All countries are evaluated at the median ranking on each of the indexes. Best possible rank is 1. The plots on the diagonal of the matrix show the density of the median rankings for respectively Doing Business (top/left), Freedom House (center), and the Human Development Index (bottom/right). The off-diagonal plots show the correlation in rankings between the indexes. The dashed line is a regression line, while the solid line is a non-parametric regression line (lowess).

saying that we only want to have principal components that are able to explain at least all the variation in one indicator, since the variance of the indicators is normalized to 1 before doing the principal component analysis.

We see that both the Human Development Index and Freedom House have only one principal component. Doing Business, however, has as many as 9 components according to this simple test. This is of course no big surprise, with the original data selection from the Doing Business report there are 29 indicators going into the calculation of the index. It hence looks like the indicators are rather heterogeneous in what they capture.¹¹

¹¹We have also checked for the number of components by using the scree-plot method of Cattell (1966);

	PC1	PC2	PC3	PC4						
Standard deviation	1.8127	0.5683	0.4640	0.4194						
Proportion of Variance	0.8215	0.0807	0.0538	0.0440						
Cumulative Proportion	0.8215	0.9022	0.9560	1.0000						
	PC1	PC2	PC3	PC4	PC5	PC6	PC7			
Standard deviation	2.5283	0.5293	0.3278	0.2876	0.2308	0.2154	0.1939			
Proportion of Variance	0.9132	0.0400	0.0154	0.0118	0.0076	0.0066	0.0054			
Cumulative Proportion	0.9132	0.9532	0.9686	0.9804	0.9880	0.9946	1.0000			
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Standard deviation	2.64	1.69	1.53	1.37	1.20	1.14	1.12	1.05	1.01	0.95
Proportion of Variance	0.24	0.10	0.08	0.06	0.05	0.05	0.04	0.04	0.04	0.03
Cumulative Proportion	0.24	0.34	0.42	0.48	0.53	0.58	0.62	0.66	0.69	0.72

Table 1: Principal components of the Human Development Index (top), Freedom House (middle), and Doing Business (bottom).

We conclude from this that for the Human Development Index and the Freedom House, a single underlying dimension is supported. For Doing Business, however, the picture is more blurry. The potential nine components that we identify for Doing Business do not correspond to the sub-categories of the index. This provides yet another warning against putting too much emphasis on the ranking of countries based on this index.

6 Conclusion

We have demonstrated that the Human Development Index, Freedom House, and the Doing Business index all provide country rankings that are wildly uncertain for the great majority of the countries ranked. The rankings therefore end up emphasizing imaginary differences between countries as if they were distinct and real.

Our assessment is a lower bound on the true uncertainty. We have assumed, contrary to what many critics believe, that the indexes incorporate just the right indicators. We have also assumed that the indicators are measured without error. In spite of this abstraction from obvious uncertainty, we claim that the international index rankings are highly imprecise, yet presented as highly precise. The imprecision is caused by the high level of noise introduced when the index scores are calculated.

Yet, country rankings are used as guides for economic, social, and legal reforms by governments, NGOs and international organizations. Their popularity is not without cost. The Doing Business team, for instance, is proud of the real policy consequences: “Publishing comparative data on the ease of doing business inspires governments to reform. Since its start in October 2003 the *Doing Business* project has inspired or informed 113 reforms around the world. In 2006 Georgia targeted the top 25 list and used *Doing* the results are the same.

Business indicators as benchmarks of its progress. It now ranks 18 on the ease of doing business, and the government has set an even more ambitious goal. Saudi Arabia and Mauritius have targeted the top 10. Both have made tremendous progress: Saudi Arabia now ranks 23, and Mauritius 27” (World Bank, 2007, p. 7).

All this may be nothing more than *rank-seeking* behavior, whenever the link between the observed indicators and the underlying performance is highly uncertain. The Doing Business index is in this respect particularly problematic. First of all, it invites policy reforms, the success of which are assessed solely by improved ranking. Secondly, the large number of indicators contained in the index makes it easy to undertake cosmetic interventions to improve a country’s ranking without altering the underlying performance.

In sum, whenever the scores of international index rankings are taken literally, the indexes may be poor guides for policies as each link between indicators and scores are noisy and uncertain, but presented as certain. The belief in accuracy in the presence of inaccuracy may lead to a shift in focus among reformers from what really counts to what the makers of these rankings count. This is the tyranny of international index rankings.

References

- ANAND, S., AND A. K. SEN (1994): “Human Development Index: Methodology and Measurement,” Occasional Papers 12, Human Development Report Office, New York.
- CATTELL, R. B. (1966): “The Scree Test For The Number Of Factors,” *Multivariate Behavioral Research*, 1(2), 245–276.
- CLINTON, J. D., AND D. E. LEWIS (2008): “Expert Opinion, Agency Characteristics, and Agency Preferences,” *Political Analysis*, 16(1), 3–20.
- GELMAN, A., AND J. HILL (2006): *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- HAMPEL-MALAGROSA, A., AND J. FRICKENSTEIN (2008): “Taking the woman’s perspective: gender risks of regulatory reforms,” *Enterprise Development & Microfinance*, 19(3), 204–19.
- HDR (2007): “Human Development Report 2007/2008—Fighting climate change: Human solidarity in a divided world,” United Nations Development Programme, New York.
- KAISER, H. F. (1960): “The Application of Electronic Computers to Factor Analysis,” *Educational and Psychological Measurement*, 20(1), 141–51.

- LEE, S., D. MCCANN, AND N. TORM (2008): “The World Bank’s “Employing Workers” index: Findings and critiques—A review of recent evidence,” *International Labour Review*, 147(4), 416–32.
- MORSE, S. (2003): “For better or for worse, till the human development index do us part?,” *Ecological Economics*, 45, 281–96.
- MUNCK, G. L., AND J. VERKUILEN (2002): “Conceptualizing and Measuring Democracy,” *Comparative Political Studies*, 35(1), 5–34.
- SAGAR, A. D., AND A. NAJAM (1998): “The human development index: a critical review,” *Ecological Economics*, 25(3), 249–264.
- SØREIDE, T. (2006): “Is it wrong to rank? A critical assessment of corruption indices,” CMI Working Papers WP 2006: 1, CMI (Chr. Michelsen Institute), Bergen, Norway.
- TREIER, S., AND S. JACKMAN (2008): “Democracy as a Latent Variable,” *American Journal of Political Science*, 52(1), 201–217.
- WORLD BANK (2007): “Doing Business 2008: Comparing Regulation in 178 Economies,” The World Bank, Washington, D.C.

A Model and estimation

A.1 Model—Continuous, ordinal, count, and censored outcomes

All the observable indicators we have in our data are one of the following kinds: i) continuous with a clearly defined scale (GDP per capita, life expectancy, etc.), ii) continuous but as if censored (literacy rates, enrollment rates, etc.), iii) discrete count data (the number of documents needed to fill out to export or import a good etc.), iv) discrete data of an ordinal nature (scores on scales that does not have a clear-cut meaning). The point of our estimation is to estimate the parameters of the model given in equation (1).

For the ordinal variables, we follow the approach in Treier and Jackman (2008) and Clinton and Lewis (2008), and estimate an ordered choice model. The idea is to still think of θ_i as the (continuous) underlying quality of the countries, but that the observable indicator now is observed in steps only. The steps are assumed to be ordered, in the sense that x_{ij} is non-decreasing in θ_i , holding the error term constant and assuming that $\beta_j > 0$.

More precisely,

$$x_{ij} = \begin{cases} 1 & \text{if } \alpha_j + \beta_j \theta_i + \varepsilon_{ij} \in (-\infty, \tau_{j1}^*] \\ 2 & \text{if } \alpha_j + \beta_j \theta_i + \varepsilon_{ij} \in (\tau_{j1}^*, \tau_{j2}^*] \\ \vdots & \\ K_j & \text{if } \alpha_j + \beta_j \theta_i + \varepsilon_{ij} \in (\tau_{j,K_j-1}^*, \infty) \end{cases}$$

where K_j denotes the maximum number of values observed for the ordinal variable. Assuming the error terms are distributed according to the cdf F , it is immediate that

$$\begin{aligned} \Pr [x_{ij} = 1] &= F(\tau_{j1} - \beta_j \theta_i) \\ \Pr [x_{ij} = 2] &= F(\tau_{j2} - \beta_j \theta_i) - F(\tau_{j1} - \beta_j \theta_i) \\ &\vdots \\ \Pr [x_{ij} = K_j] &= 1 - F(\tau_{j,K_j-1} - \beta_j \theta_i), \end{aligned}$$

where $\tau_{jk} = \tau_{jk}^* - \alpha_j$. We assume that ε_{ij} is logistically distributed, which means that the model for the ordered outcomes is a standard ordered logit, but with an unobserved right-hand side. We also here control for correlation among the indicators within the same group, see section 3.1.

We also take censoring into account, by modifying the likelihood function in the appropriate way. If x_{ij} is top-censored at 100, say, the likelihood function for an observation is (assuming that the error terms are normally distributed)

$$p(x_{ij} | \alpha, \beta, \sigma^2) = \begin{cases} \phi((x_{ij} - \alpha_j - \beta_j \theta_i) / \sigma) & \text{if } x_{ij} < 100 \\ \Phi((\alpha_j + \beta_j \theta_i - 100) / \sigma) & \text{if } x_{ij} = 100, \end{cases}$$

where ϕ is the density function and Φ is the cumulative density function of the normal distribution, see for example Gelman and Hill (2006, p. 404). We take this censoring into account using the $I()$ -construct in WinBUGS, following Gelman and Hill (2006, p. 404-5).

For the count data, we estimate a overdispersed Poisson regression-type model. Assuming that the left-hand side variable x_{ij} only take positive integer values, the regression model we use has the form

$$\begin{aligned} x_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\ \log(\lambda_{ij}) &= \alpha_j + \beta_j \theta_i + \epsilon_{ij}. \end{aligned}$$

A.2 Estimation

The estimation is performed using WinBUGS.¹² All computer codes used in the paper are available at request. Our approach is a latent variable approach, where we estimate the underlying dimension given the observed indicators; see the previous section and section 3. The model we fit to different indexes is the model given in equation (1), with care taken to what kinds of variables are observed on the left-hand side of the equation.

Priors

We use diffuse priors on all the structural parameters; the priors are given in Table 2. By this we mean that we put no a priori restrictions on the values that the parameters can or are likely to take, the identification of the parameters rely only on the data via the likelihood function.

Computational Details

We run 3 chains each consisting of 300,000 iterations for all of the indexes. The first 100,000 iterations of each of the chains are discarded as burn-in, in order to ensure that our results are not driven by the starting values. The starting values for θ_i were obtained by running a standard factor analysis and calculating the Bartlett scores. Starting values for the β_j 's were set to 1, which ensured that all the chains were on the same scale—remember that we only have local identification, so if both f and β are multiplied by -1 the results will be exactly the same. For the same reason we set the starting values for the δ_{ik} -shocks to 0 for the HDI. The remaining starting values were generated at random in WinBUGS.

Each 200th subsequent iteration is recorded. This is in order to ensure low autocorrelation between each of the sampled values. The traceplots show no sign of non-convergence, and formal tests of convergence support the assessment that the models have converged: \hat{R} is below 1.05 for all parameters. There is no indication of serious autocorrelation problems.

B Classification of variables

The Human Development Index

Classification of the variables in the Human Development Index can be found in Table 3.

Freedom House

Classification of the variables in Freedom House can be found in Table 4.

¹²Available at <http://www.mrc-bsu.cam.ac.uk/bugs/>.

Index	Variable	Prior
The Human Development Index	θ	$N(0, 1)$
	(α_j, β_j)	$N(\mathbf{0}, 10^7 \times I_2)$
	$1/\sigma_j^2$	$\Gamma(.05, .05)$
	δ_{ik}	$N(0, \sigma_{\delta,k}^2)$
	$\sigma_{\delta,k}^2$	Uniform on $[0, 100]$
	Freedom House	θ
β_j		$N(0, 1/.15)$
τ_{j1}		$N(0, 1/.15)$
τ_{jk}		Exponential w/ rate 2
δ_{ik}		$N(0, \sigma_{\delta,k}^2)$
$\sigma_{\delta,k}^2$		Uniform on $[0, 100]$
Doing Business		θ
	(α_j, β_j)	$N(\mathbf{0}, 10^7 \times I_2)$
	$1/\sigma_j^2$	$\Gamma(.05, .05)$
	δ_{ik}	$N(0, \sigma_{\delta,k}^2)$
	$\sigma_{\delta,k}^2$	Uniform on $[0, 3]^a$
	τ_{j1}	$N(0, 1/.15)$
	τ_{jk}	Exponential w/ rate 2
	ϵ_{ij}	$N(0, \sigma_{\epsilon_{ij}}^2)$
	$\sigma_{\epsilon_{ij}}^2$	Uniform $[0, 5]$

^a We have also fitted the model using $[0, 100]$. The results were the same, but it converged slower. None of the estimated parameters are close to the boundaries of the distribution.

Table 2: Priors for the different parameters

Group	Variable	Type
Schooling	School enrollment	Continuous, censored at 100
	Literacy	Continuous, censored at 100
GDP	Log GDP per capita	Continuous
Life expectancy	Life expectancy	Continuous

Table 3: Classification of variables, the Human Development Index

Group	Variable	Type
Political Rights	A: Electoral Process	Ordered
	B: Political Pluralism and Participation	Ordered
	C: Functioning of Government	Ordered
Civil Liberties	D: Freedom of Expression and Belief	Ordered
	E: Associational and Organizational Rights	Ordered
	F: Rule of Law	Ordered
	G: Personal Autonomy and Individual Rights	Ordered

Table 4: Classification of variables, Freedom House

Doing Business

Classification of the variables in Doing Business can be found in Table 5.

Group	Variable	Type
Starting a Business	Procedures (number)	Count
	Time (days)	Count
	Cost (% of income per capita)	Continuous
	Min. capital (% of income per capita)	Continuous, censored at 0
Dealing with Licenses	Procedures (number)	Count
	Time (days)	Count
	Cost (% of income per capita)	Continuous
Employing Workers	Difficulty of Hiring Index	Continuous, censored at 0 and 100
	Rigidity of Hours Index	Continuous, censored at 0 and 100
	Difficulty of Firing Index	Continuous, censored at 0 and 100
	Firing costs (weeks of wages)	Continuous, censored at 0
Registering Property	Procedures (number)	Count
	Time (days)	Count
	Cost (% of property value)	Continuous
Getting Credit	Credit Information Index	Ordered
	Private bureau coverage (% of adults)	Continuous, censored at 0 and 100
	Public bureau coverage (% of adults)	Continuous, censored at 0 and 100
	Legal Rights Index	Ordered
Protecting Investors	Disclosure Index	Ordered
	Director Liability Index	Ordered
	Shareholder Suits Index	Ordered
Paying Taxes	Payments (number)	Count

Group	Variable	Type
	Time (hours)	Count
	Profit tax rate	Continuous, censored at 0
	Labor tax rate	Continuous, censored at 0
	Other taxes	Continuous, censored at 0
Trading Across Borders	Documents for export (number)	Count
	Time for export (days)	Count
	Cost to export (US\$ per container)	Continuous
	Documents for import (number)	Count
	Time for import (days)	Count
	Cost to import (US\$ per container)	Continuous
Enforcing Contracts	Procedures (number)	Count
	Time (days)	Count
	Cost (% of debt)	Continuous
Closing a Business	Time (years)	Continuous
	Cost (% of estate)	Continuous
	Recovery rate (cents on the dollar)	Continuous

Table 5: Classification of variables, Doing Business

C Estimated latent scores

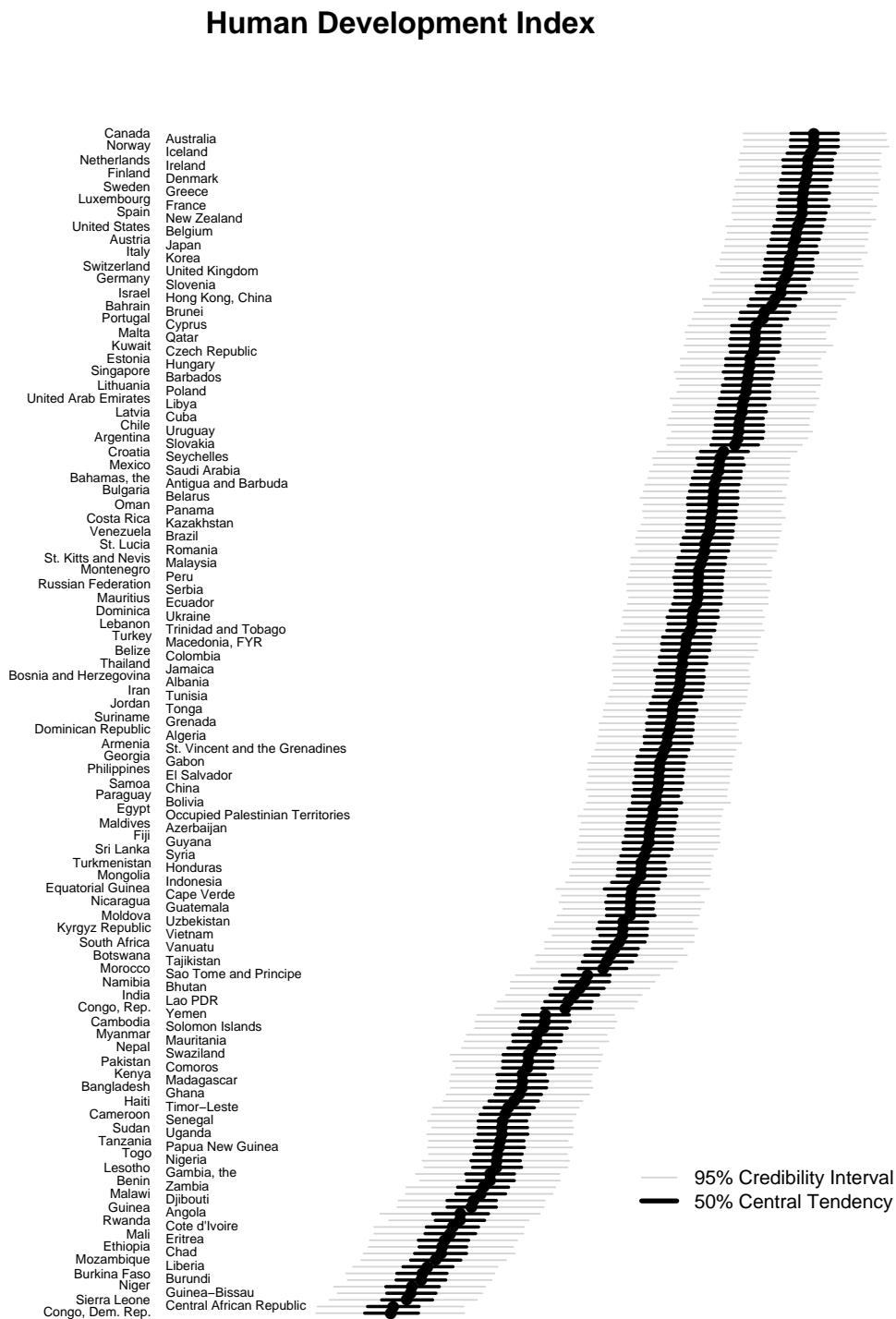


Figure 8: Estimated positions for the Human Development Index. The black circle indicates the median value, the black line the central tendency, and the gray line the 95 per cent highest posterior density region for θ , the position of the countries on the Human Development Index.

Freedom House

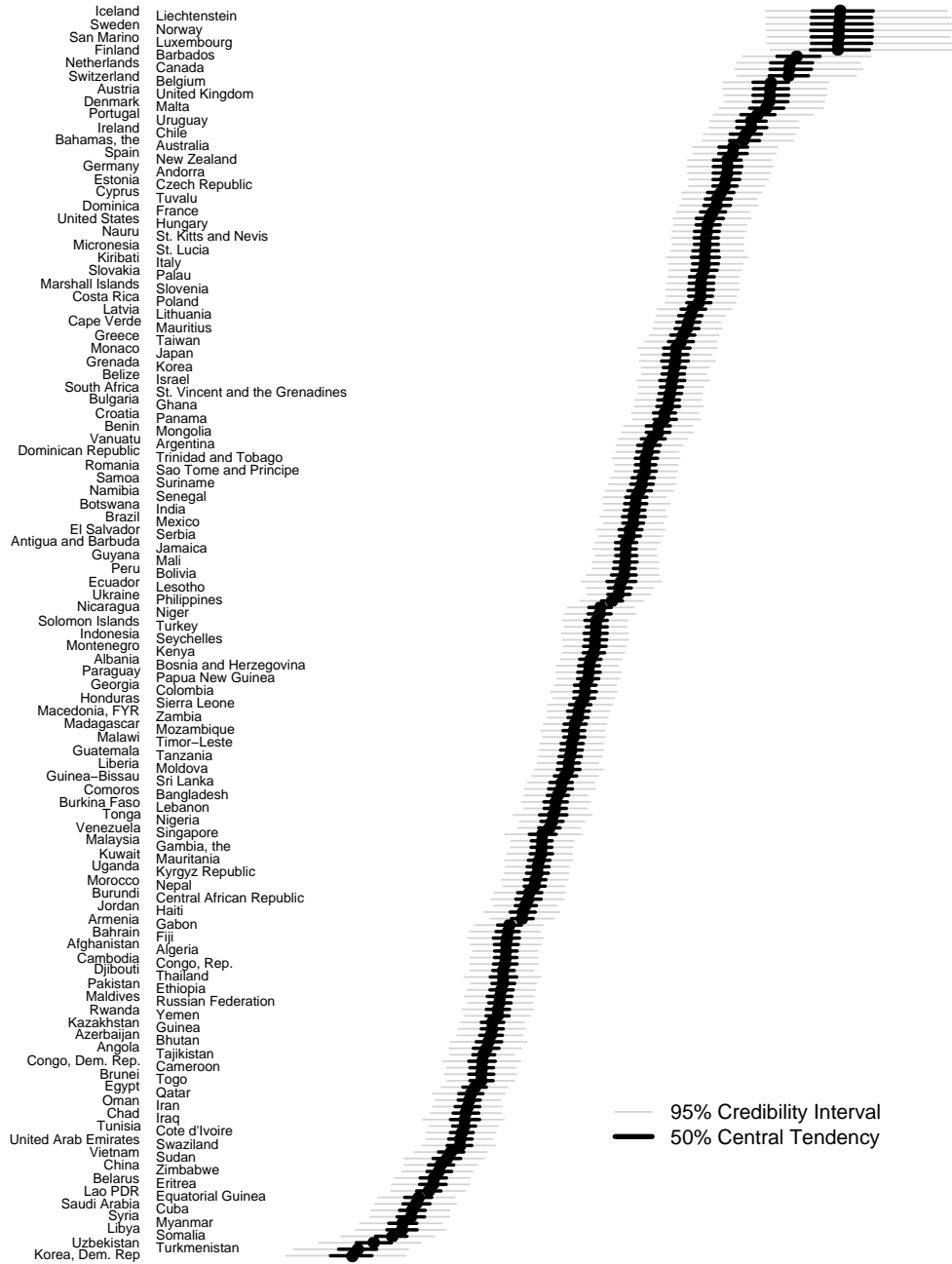


Figure 9: Estimated positions for Freedom House.

The black circle indicates the median value, the black line the central tendency, and the gray line the 95 per cent highest posterior density region for θ , the position of the countries on Freedom House.

Doing Business

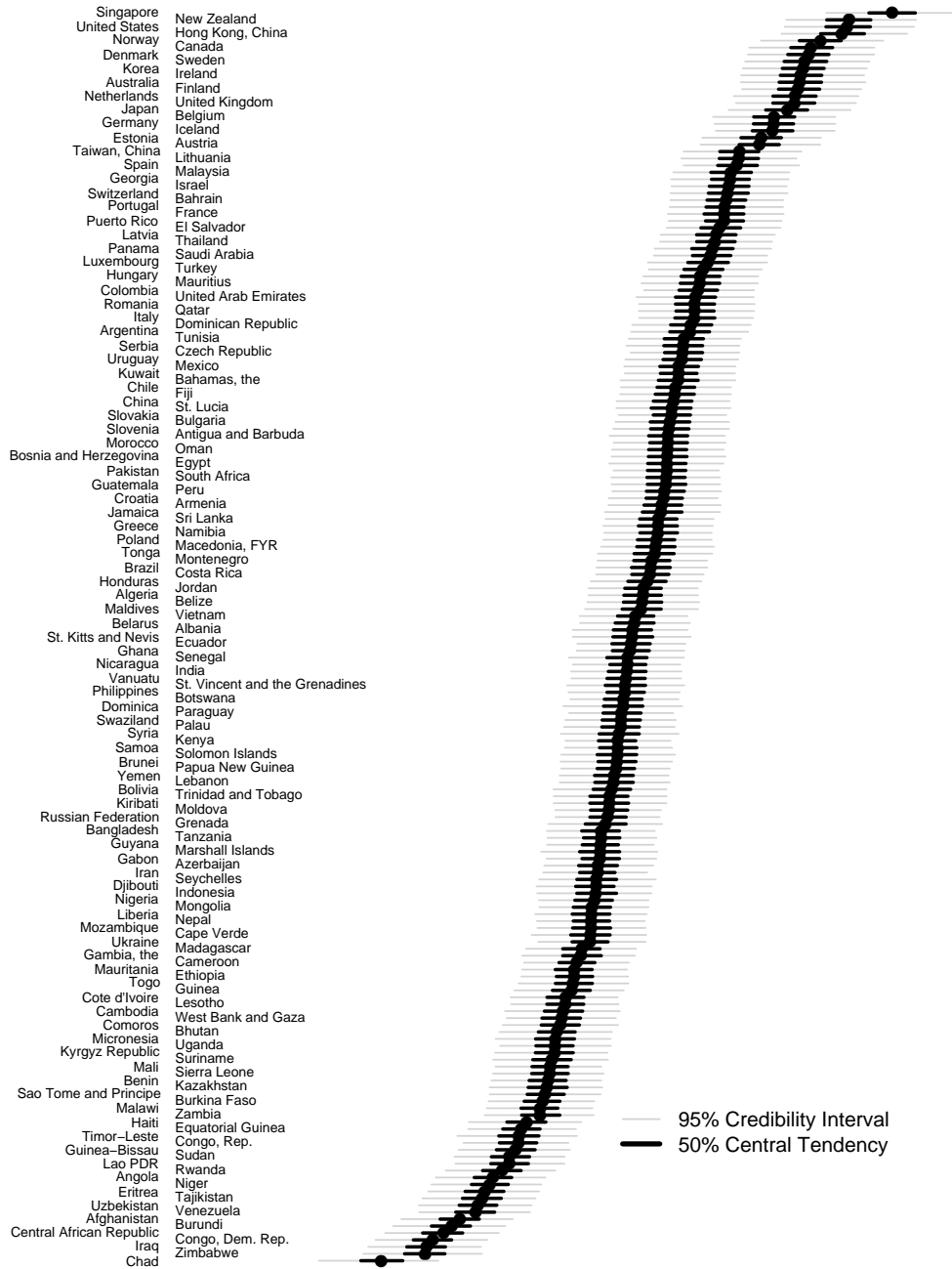


Figure 10: Estimated positions for Doing Business. The black circle indicates the median value, the black line the central tendency, and the gray line the 95 per cent highest posterior density region for θ , the position of the countries on Doing Business.