FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

TWEEKERKENSTRAAT 2
B-9000 GENT
Tel. : 32 - (0)9 – 264.34.61
Fax. : 32 - (0)9 – 264.35.92

WORKING PAPER

Divining the Level of Corruption
A Bayesian State-Space Approach

Samuel Standaert

April 2013
2013/835

D/2013/7012/06

Divining the level of corruption
a bayesian state-space approach
Working paper

Samuel Standaert[*†]
Department of Economics
Ghent University

April 10, 2013

Abstract

This paper outlines a new methodological framework for combining indicators of corruption. The methodology of the World Governance Indicators is extended to fully make use of the time-structure present in corruption data. The resulting state-space framework is estimated using a Bayesian Gibbs sampler algorithm.

The state-space framework holds many advantages from a practical, an estimation and a theoretical point of view. Most importantly, the indicator significantly increases data availability while at the same time addressing the selection bias issues that plague the CPI and WGI indexes. It produces estimates that are more stable and reliable. Because the estimation framework is transparent and data is entered without any manipulations, the resulting indicator should also be more objective.

Keywords: *Corruption indicators; Bayesian Econometrics; Factor Model; State-Space*

1 Introduction

Researchers looking at the effects or determinants of corruption are faced with the difficulty of having to choose one out of the more than 40 individual indicators available. Each one of these indicators differs in availability both in time and space, exactly what it is trying to measure, and where or with whom it was collected. Because that one indicator that meets all requirements often proves elusive, most studies resort to aggregated indicators of corruption. The

[*]I am indebted to Glenn Rayp, Gerdie Everaert and José Louis Figueroa Oropeza for their invaluable help in writing this paper.

[†]`Samuel.Standaert@UGent.be`

two most used are the Corruption Perceptions Index (CPI) published by Transparency International and the World Governance Indicators' index of corruption (WGI) published by the World Bank.

However, the use of these aggregated indicators is not without criticism, especially when making comparisons over time. Shifts in the indices are not only driven by the level of corruption, but by changes in the methodology and sources as well. Moreover, both indicators only go back to the mid-nineties and their early values suffer from serious selection bias problems (Treisman, 2007).

This paper outlines a new methodology for combining indicators of corruption that significantly improves on the existing ones. The following section reviews the data and methodology used in the CPI and WGI indices and highlights some of their shortcomings. Subsequently the new framework is presented and its results are discussed.

2 Data

An important question when choosing the sources for an indicator of corruption is whether to use incidence or perception-based surveys. The former asks for personal experience with corruption within the last x months, while the latter asks respondents for their opinion on the level of corruption in the country as a whole, or in various branches of the government. Depending on the country, both indicators can come to quite different conclusions (Roca, 2011).

Both the CPI and WGI indexes use only perception-based indicator for a number of reasons. First of all, because most hard data on corruption is essentially flawed. For example, the number of corruption convictions crucially depends on the judicial quality as well as the level of corruption (Lambsdorff, 2005). Kaufmann, Kraay, and Mastruzzi (2010, p. 18) also add that non-perception-based indicators of corruption will capture *de jure* notions of corruption rather than the *de facto* reality.

Nevertheless, some authors argue that experience-based indicators are less prone to bias due to preconceived notions. For example, Arndt and Oman (2006) posit that surveys taking from business people could be skewed against environmental or social protection. However, Kaufmann, Kraay, and Zoido-Lobatón (2004) and Kaufmann, Kraay, and Mastruzzi (2007a) found no evidence of systematic differences between firm-level surveys and other indicators, nor between expert opinions on left-wing versus right-wing governments.

Finally, unlike the perception surveys, experience-based indicators are less likely to be influenced by other people's opinions (Treisman, 2007). While Kaufmann, Kraay, and Mastruzzi (2007b) could find no evidence of a higher correlation among expert opinions than among firm-level surveys, this does not altogether rule cross-correlation out. The firm and household survey data can still influence expert opinions, just as the expert opinions can influence the opinion of business leaders, NGOs, etc.

For comparability's sake, this paper uses a subset of the WGI dataset, including all free publicly available data[1] as well as the International Country Risk Guide's index of corruption (ICRG). As a robustness check, the dataset is then further expanded to include experience-based indicators [note: to be added]. To the extent that both type of indicators are prone to different measurement errors, combining their information could significantly strengthen the combined indicator of corruption.

The corruption perceptions dataset contains 42 variables coming from 21 different sources and spans 211 countries from 1984 to 2010. It includes both survey data as well as expert assessments and can be divided into five groups: cross-country household and firm-level surveys; and expert opinions from NGOs, from commercial risk rating agencies or from governments and multilateral organizations. Appendix A list all 42 indicators and provides summary data. For a more thorough description of the individual sources see Arndt and Oman (2006, p. 52-57).

Cross-country survey of households
- Gallup World Poll
- Latinobarometer

Expert assessment from NGO and think tanks
- Global Integrity
- The Freedom House
- Bertelsmann Transformation Index
- Global Corruption Barometer

Cross-country survey of firms
- Afrobarometer
- Global competitiveness survey
- Vanderbilt University's Americas Barometer
- World Competitiveness Yearbook
- Business Environment and Enterprise Performance Survey

Expert assessment from commercial risk rating agencies
- International Country Risk Guide
- Global Risk Service
- World markets online
- Economist Intelligence Unit
- Political and Economic Risk Consultancy

Expert assessment from governments and multilaterals
- Country Policy and Institutional Assessment
- African Development Bank
- Asian Development Bank
- IFAD Rural Sector Performance Assessment
- Institutional Profiles database

3 Composite indicators of corruption

3.1 Corruption Perceptions Index

Published yearly since 1995 by Transparency International, the Corruptions Perceptions Index (CPI) is probably the best known worldwide indicator. As the name indicates, it combines perception based corruption indicators in order to capture the *'misuse of public power for private benefit'* (Lambsdorff, 2005, p.4). The higher a country's CPI score, the lower its level of corruption. It merges these indicators in the following way:

1. The individual indicators are first standardized using a technique called *matching percentiles*:

    (a) For every source, the countries are ranked from least to most corrupt.

[1] Available at www.govindicators.org.

(b) The highest ranked country is assigned last year's highest score. The second highest ranked gets the second to highest score, and so on.

2. These standardized indicators are then combined using a simple average.

3. Because of the matching percentiles standardization, the standard deviations decreases year after year. For example, if a country is ranked least corrupt in all surveys except in one, it will be this year's highest ranking country, but its score will be below last year's high score. To prevent this from happening while making sure that the values stay within the [0,10] interval, a beta-transformation[2] is used to preserve the standard deviation.

4. Finally, in order to give an indication of the uncertainty of the CPI score, a confidence interval is constructed using bootstrap methods. In every iteration, a new sample is drawn from the sources of each country (with replacement) and the index is recomputed. The more the ranking of the indicators differs, the wider the confidence interval becomes.

The methodology used in creating the CPI index has a number of drawbacks, the most important of which is that it should not be used for comparisons over time (Transparency International, 2012). The reason for this is that it only uses the relative rank data and combines it using a simple average, making the index sensitive to changes in the countries covered and indices used.

Secondly, it does not include countries for which there are less than three sources available in a given year and as a result, the initial years of the index only cover a select group of countries. This selection is not independent of the level of corruption, causing the index to be prone to a selection bias issue (Treisman, 2007). In order to alleviate the availability problem, the data is manipulated in a number of ad-hoc ways. For example, some (but not all) sources are averaged over the last three years, while others are used twice.

Lastly, the confidence interval only reflects the divergence of the underlying indicators. It ignores the differences in reliability of different indicators, assuming that all indicators are without error and independent of one another. Moreover, it does not take the number of sources into account. This could lead to the scenario where a country with only three relatively unreliable sources has a lower confidence interval than a country for which there are dozens of data points available but whose standard deviation is slightly higher.

3.2 World Governance Indicators

The World Governance Indicators' index of corruption measures *'the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private*

[2]The average of all standardized corruption scores, X, is transformed in the following way: $X^\star = 10 * \int_0^1 (X/10)^{\alpha-1}(1-X/10)^{\beta-1}dX$. α and β are chosen such that the resulting mean and standard deviation remain constant over time (Lambsdorff, 2005, p. 8).

interests.' (Kaufmann et al., 2010, p.4). To do this, they combine the perception based corruption indicators using an unobserved components (or factor) model.

3.2.1 Unobserved components model

Denoting the $(1 \times k)$ vector of corruption indicators with $y_{i,t}$, and the unknown 'true' level of corruption with the scalar $\alpha_{i,t}$, they estimate the following model:

$$y_{i,t} = C + Z(\alpha_{i,t} + \epsilon_t^\star) \quad (1)$$
$$\alpha_{i,t} \sim N(0, Q) \quad (2)$$
$$\epsilon_{i,t}^\star \sim N(0, H^\star) \quad (3)$$

for all countries $i = 1, ..., p$ and years $t = 1, ..., n$, with k the number of individual corruption indicators in $y_{i,t}$.

The $(1 \times k)$ vectors C and Z capture both differences in scaling as well as the distribution of values. For example, some indicators might easily assign the highest score, while others reserve that only for a limited numbers of countries.

Finally, $\epsilon_{i,t}^\star$ is an error term with $(k \times k)$ variance matrix, $H^\star$. The measurement errors of different indicators are assumed to be uncorrelated: $E(\epsilon_{i,t}^\star, \epsilon_{j,t}^\star) = 0 \ \forall i \neq j$, or $H^\star$ diagonal. The error term is meant to capture two effects. Firstly, it will account for errors in the data collection process. Secondly, it also corrects for the possibility that the indicators do not measure the overall level of corruption, but a related concept like the level of petty corruption, or the level of corruption in the judiciary.

3.2.2 Estimation

In order to estimate this model, $\alpha_{i,t}$ and $\epsilon_{i,t}^\star$ are assumed to be bivariate normal distributed. This implies that $y_{i,t}$ is a normally distributed variable with mean $Z^\star \alpha_{i,t}$ and variance $Z^\star Z^{\star\prime} + Z^\star H^\star Z^{\star\prime}$.

The data then is split up in a *representative* and *non-representative* group. Simply put, the representative group contains all indicators whose scope either covers the entire population, or represents a random selection of countries. Conversely, the non-representative group contains those indicators that are most likely to suffer from selection bias. For example those focussing on the least developed countries, etc.

In the first step, estimation is done using only the representative group, where the yearly expected value of α is assumed zero. These estimates are then updated with the information from the non-representative group. The advantage of this two-step procedure is that the results from the representative group can be used to assess and correct the bias in the non-representative group without having to make any prior assumptions on the size or direction of the bias.

Finally, the results are rescaled. Firstly, for each year the mean is set to zero and the standard deviation of the corruption values is set to one. Kaufmann et al. (2010) argue that this standardization does not preclude their use in time-series or panel studies because they find no significant evidence of a worldwide trend in corruption. More on this in the next section.

A second rescaling occurs to partly address the selection bias issue. They find that countries added later to the sample on average have lower levels of corruption. However, the mean value of corruption was set to zero for each year. To compensate, the mean value in each year is adjusted, using the values of 2003 as a benchmark.

3.2.3 Results

The inclusion of the error term with indicator-specific variance is a big advantage of the WGI index. It makes it more robust to the inclusion of indicators that are less correlated with the general level of corruption, whether due to measurement errors or because it only measures a related concept (for example, the level of corruption in elected officials). The CPI index on the other hand treats all indicators the same, regardless of their reliability or conceptual suitability.

As was the case with the CPI index, the initial years of the index are available for a select group of countries. The problem is less severe because the index is composed even when only one datasource is available. Nevertheless, from its start in 1996 to 2002 the index is only available every two years. Those values of the individual indicators in the years in between are copy-pasted onto the following year.

3.3 Unit root

Because the level of corruption is in a large part driven by social norms and values, it is expected to show a high degree of persistence. Table 1 confirms this. The augmented Dicky-Fuller test finds that the null-hypothesis of a unit root could not be rejected for at least seven out of ten countries, regardless of the indicator used. This means that for the vast majority of countries, all changes in the level of corruption are permanent.

While this time-dependence is reflected in the values of the CPI and WGI indexes (table 1), they do not make use of it in their estimations. However, by taking the past values into account, the reliability of the estimates can be increased and random 'noise' can be better filtered out from the corruption indicators. Most importantly, it significantly expands the time period for which the level of corruption can be computed. How and why this is the case is explained in the next section.

Table 1: Unit root tests on corruption

indicator	H_0 not rejected	H_0 rejected	% not rejected	indicator	H_0 not rejected	H_0 rejected	% not rejected
WGI	173	44	78.74 %	y_{16}	64	12	84.21%
CPI	148	32	82.22%	y_{17}	63	13	82.89%
y_1	136	7	95.10%	y_{18}	62	13	82.67%
y_2	164	37	81.59%	y_{19}	61	14	81.33%
y_3	148	4	97.37%	y_{20}	64	11	85.33%
y_4	127	15	89.44%	y_{21}	61	13	82.43%
y_5	109	23	82.58%	y_{22}	62	11	84.93%
y_6	44	13	77.19%	y_{23}	30	13	69.77%
y_7	91	11	89.22%	y_{26}	24	4	85.71%
y_8	95	32	74.80%	y_{27}	41	1	97.62%
y_9	64	15	81.01%	y_{28}	18	0	100%
y_{10}	68	10	87.18%	y_{29}	18	0	100%
y_{11}	69	9	88.46%	y_{30}	17	0	100%
y_{12}	63	15	80.77%	y_{31}	14	4	77.78%
y_{13}	61	15	80.26%	y_{32}	12	1	92.31%
y_{14}	73	2	97.33%	y_{34}	28	1	96.55%

Augmented Dicky-Fuller stationarity test on the individual corruption indicators y_i for each country. The first column reports the number of countries for which the H_0 hypothesis of non-stationarity could not be rejected at 5% significance level, second column the number for which it could. Indicators which did not have enough observations to run the test for any country are left out.

4 Methodological framework

4.1 Model

Extending the WGI framework to take the time dependence into account leads to the following system of equations:

$$y_{i,t} = C + Z\alpha_{i,t} + \epsilon_{i,t} \quad (4)$$

$$\alpha_{i,t} = T_i\alpha_{i,t-1} + \nu_{i,t} \quad (5)$$

$$\epsilon_{i,t} \sim N(0, H) \quad (6)$$

$$\nu_{i,t} \sim N(0, Q) \quad (7)$$

Once again, the measurement equation (eq. 4) states that the k indicators of corruption $y_{i,t}$ try to measure the 'true' level of corruption $\alpha_{i,t}$. Corruption is here defined in the same way as in the World Governance Indicators (Kaufmann et al., 2010). The variables i and t respectively represent the different countries and time-periods.

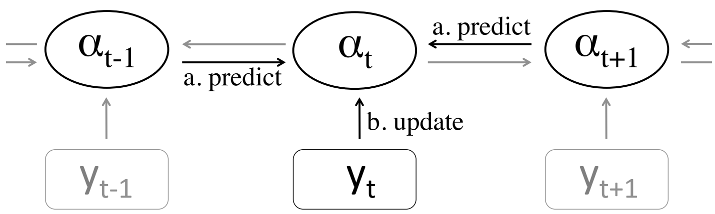
The scaling parameters C and Z can vary over the indicators of corruption, but remain constant over time and country. Similarly, the variance of the error term ϵ can differ over all corruption indicators, but initially the correlation between the error terms of different indicators is ruled out (H diagonal). In a second version of this model, the measurement error is allowed to be correlated between different indicators that come from the same source (H block-diagonal).

The state equation (eq. 5) allows for the 'true' level of corruption to depend on its previous values. This level of dependence can be different for each country, i. If the values for T_i are set to zero this model coincides with that of the WGI index[3] (eq. 2). Lastly, the level of corruption is restricted to be a non-explosive series: $|T| \leq 1$. This rules out a continuously expanding process and ensures that the model converges to a steady solution.

By bringing the time dimension into play, a lot more information is used in the estimation of each corruption value. Figure 1 illustrates this. In the WGI framework, only the current information on corruption, y_t, is used (step b). In the new framework the level of corruption is predicted using both past and future values (step a), after which the information in y_t is used to update that estimate (step b). The importance of step a versus step b will depend on how reliable the corruption indicators (H) are versus how reliable the past values are (Q).

Because α_{t-1} and α_{t+1} also depend on past and future values, all available information will be used to estimate the current level of corruption. Not only does this increase the reliability of each estimate, it also helps smooth out the estimates of corruption. By taking the past and future values of corruption into account, the algorithm is better able to distinguish between changes in the level of corruption and random measurement errors.

Figure 1: Estimation using time dependency



4.1.1 H blockdiagonal

A second extension to the WGI framework centers on the structure of the measurement error. Combining information from different sources only increases the reliability to the extent that those sources are independent. However this level of independence is often called into question (Arndt and Oman, 2006; Treisman, 2007). Nevertheless, both the CPI and WGI indicators assume that the measurement errors of different indicators are completely uncorrelated. The second version of the indicator tries to loosen this restriction.

[3]Substituting $\epsilon^{i,t} \equiv Z\epsilon^{\star}_{i,t}$ in equation 1 will return equation 4.

Allowing all indicators to be cross-correlated is not possible because non-overlapping missing values would reduce the sample size to zero. Instead, only the correlations between error terms of indicators that come from the *same source* are allowed to differ from zero. Firstly, these are the indicators that are a priori most likely to be subject to a shared measurement error. Secondly, because these variables are often available for the same periods and countries, this ensures that the sample size is not reduced too much when drawing values for H[4].

4.2 Estimation

This section aims to provide only a very general overview of the estimation technique. A more detailed explanation can be found in appendix B. For a complete overview of state-space models and how to estimate them, see Kim and Nelson (1999) or Durbin and Koopman (2012).

4.2.1 Gibbs sampling

This model is estimated in a Bayesian framework because of the convenience the Gibbs sampling algorithm provides. In order to solve the model, we need to estimate both the parameters of the state and measurement equation (C, Z, T and H) as well as the level of corruption (α). While it is possible to solve this model using maximum likelihood, the problem quickly becomes very complex as more and more countries are added, especially when the structure of the error term is changed. However, using the Gibbs sampling we can split the estimation up into various subcomponents which can be dealt with one at a time.

Simply put, the Gibbs sampler allows us to draw from a multivariate probability, $p(a,b)$, using only conditional probabilities, $p(a|b)$ and $p(b|a)$. Starting from a random value b_1, draws are taken iteratively from both conditional distributions while conditioning on the last drawn values:

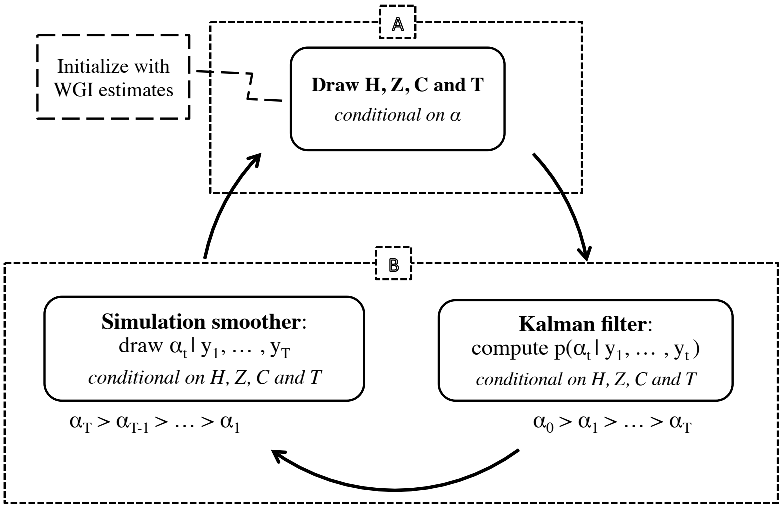
$$a_1 \sim p(a|b_1) \to b_2 \sim p(b|a_1) \to a_2 \sim p(a|b_2) \to ...$$

It can be shown that after a sufficient number of iterations (the burn-in), a_n and b_n represents random draws from the unconditional probability function $p(a,b)$. Using enough random draws, we can then reconstitute the original multivariate probability $p(a,b)$.

In this case, the Gibbs sampler consists of two main components (figure 2). In part A, the parameters of the state and measurement equations are drawn conditional on the values for α. Part B samples from the distribution of the 'true' level of corruption conditional on the parameters on the state and measurement equation.

[4]This gives 26 variables that are allowed to be cross-correlated, devided in four groups. The only exception are y_{30} and y_{33}, respectively the Global Corruption Barometer's (GCB) corruption in customs and in public officials. These indicators are not available in the years the other GCB indicators are.

Figure 2: Estimation flow chart



An additional advantage of the Gibbs sampling algorithm is that it avoids the need to distinguish between *representative* and *non-representative* sources. Nor does it require the assumption that α and ϵ are bivariate normal.

More information on the estimation procedure can be found in the appendix B, which also discusses convergence (B.4). For more information on Bayesian econometrics and Gibbs sampling algorithms, see Lancaster (2004) and Koop, Poirier, and Tobias (2007).

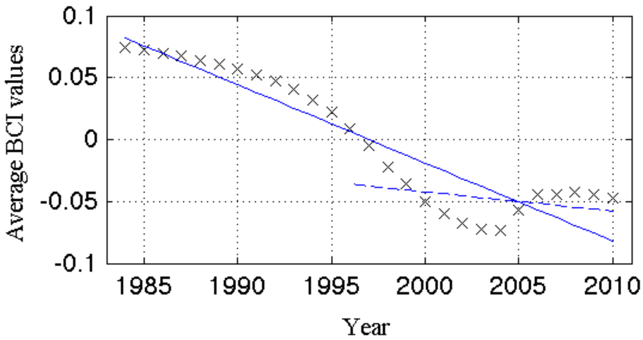
4.2.2 Missing observations

Finally, there is the issue of missing observations. There are different ways of dealing with this in the state-space framework, but they all boil down to the same idea: missing data is replaced by information which is entirely uncertain and consequently holds no value: $y_{missing} = 0$, $Var(\epsilon_{missing}) = \infty$. It allows the model to run uninterruptedly without fundamentally changing the nature of missing data. This, in combination with the time dependency, enables us to significantly increase the number of countries and years covered, without having to impute or otherwise manipulate the data (Kim and Nelson, 1999; Durbin and Koopman, 2012).

4.2.3 Standardization

Following the WGI index the expected value of corruption was standardized such that it has a mean of zero and a standard deviation of one. However,

Figure 3: Evidence of a worldwide trend



Scatterplot of the average BCI corruption value over time. The full line is a linear fit for the entire sample. The dashed line is a linear fit for the sample from 1996 onwards.

Table 2: Evidence of a worldwide trend

	Coefficient$^{(a)}$	Std. error$^{(a)}$	Std. error$^{(a)}$ (corrected)	Observations
Year ≥ 1996	-2.67	0.667***	2.648	3165
Entire sample	-6.282	0.472***	1.984***	5697

Fixed effects regression of the BCI level of corruption on time. Standard errors in round brackets. Standard errors corrected for the uncertainty of the corruption estimate in square brackets. *, **, *** indicate significance at 10, 5 and 1% level. $^{(a)}$All numbers are multiplied by a thousand.

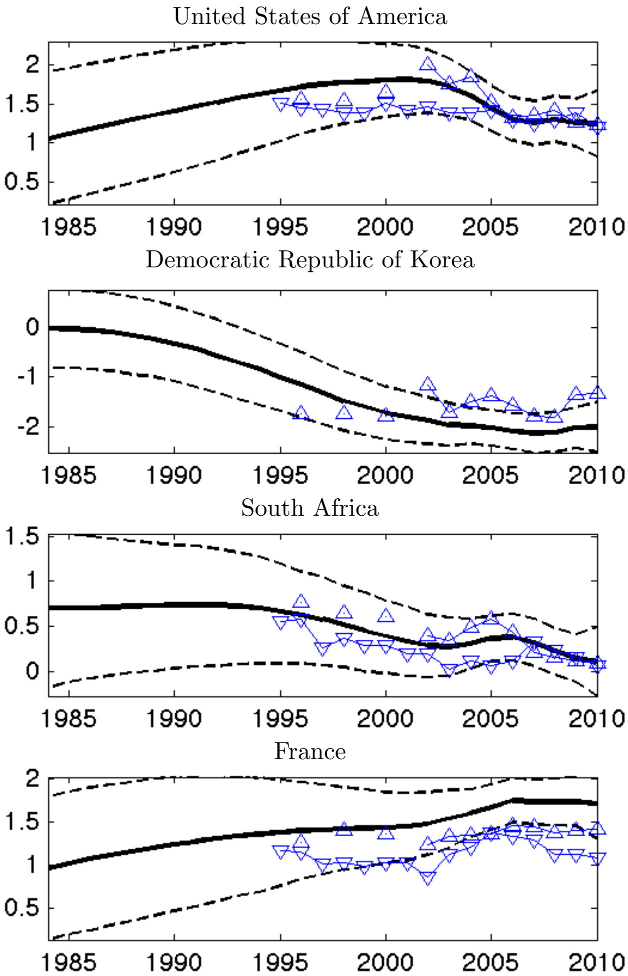
unlike the WGI index, this was done for the entire sample rather than on a yearly basis (appendix B.3). Figure 3 illustrates why yearly normalization is problematic: there is a clear global trend in the corruption values. Normalizing the yearly means would destroy this trend, and invalidate any comparisons over time (table 2). If the uncertainty of the corruption estimates is taken into account, this linear trend is significant for the entire sample. However, this is not the case for the part of the sample covered by the WGI, which is in line with what Kaufmann et al. (2010) found.

5 Results

5.1 H diagonal

Figure 4 plots of the Bayesian Corruption Index (BCI) for four countries alongside the WGI and CPI index. Like the WGI and CPI indexes, a *high* score for

Figure 4: Plot of BCI indicator with 90% confidence interval

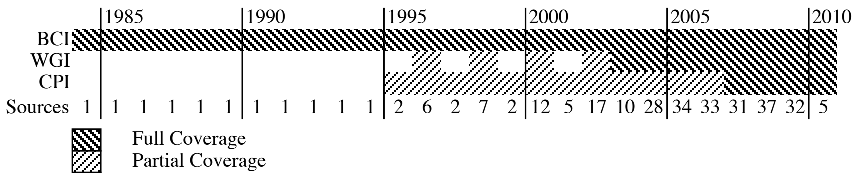


Plot of the BCI estimates, including 90% confidence interval (dotted lines). Values of the standardized CPI (downwards pointed triangles) an WGI indexes (upward pointing triangles) are also included.

the BCI indicator means that there is *little* corruption.

These figures confirm the advantages described in the previous section. First of all, the availability of data is significantly increased as the estimates of corruption go back to 1984 for all countries covered. As the availability of the sources decreases, the confidence interval widens. It should be noted that while

Figure 5: Availability combined corruption indicators and data



the BCI estimates ran until 2010, most indicators are missing from the dataset in that year. With only 5 indicators (relative to 32 in the year before), this explains why the uncertainty of the corruption estimates increases again in the final year (figure 5).

Secondly, the BCI estimates are more stable. Other corruption indicators (both individual and combined) have been criticized as being prone to small jumps in the data that have nothing to do with the level of corruption (Arndt and Oman, 2006; Treisman, 2007). By taking the past and future values of corruption into account these deviations can be filtered out better and the resulting indicator is much smoother.

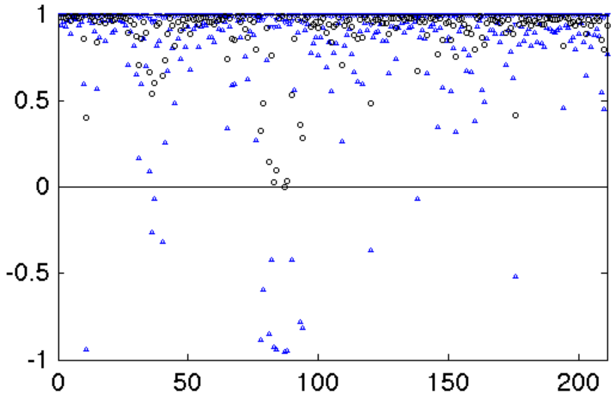
5.1.1 Validity of T

The values for parameter T provide a formal test of the necessity of the time-dependence. As was explained earlier, setting $T = 0$ will reduce the BCI framework to that of the World Governance Indicators. However, figure 6 clearly shows that the hypothesis that this parameter is zero can be rejected for the vast majority of countries: 196 out of 211 at the 1% significance level. In fact, for most countries, T lies close to one. Keep in mind that the values for T are restricted to lie within the [-1,1] interval, to ensure that corruption is a stable, non-explosive time series.

5.1.2 Correlation

Despite these differences, the BCI, WGI and CPI indicators give relatively similar predictions. Table 3 lists the pairwise correlations between the three indexes. Their overall correlation is very high. However, this is almost completely driven by their between-correlations (the correlation between the mean values). The within correlation (between the demeaned values) on the other hand is significantly lower. In other words, while the choice of indicator might not have a

Figure 6: Plot of mean values of T and their 95% confidence interval



Plot of mean values paramter T (circles) and 95% confidence intervals (triangles) for all countries.

large effect on the results in a cross-country study, this changes in time-series or panel studies.

Table 3: Pairwise correlations between BCI, CPI and WGI

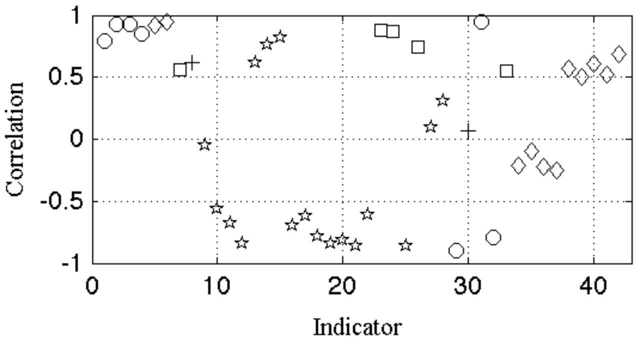
	BCI - WGI	BCI - CPI	CPI - WGI
Between[(1)]	0.9760	0.9641	0.9779
Within[(2)]	0.3147	0.1972	0.3340
Total	0.9710	0.9680	0.9691

[(1)] Between correlation is defined as the correlation between the means of each countries; [(2)] Within correlation is the correlation between the demeaned values of all countries.

Figure 7 shows the correlation of the BCI index with the individual indicators that went into the index. By far, most indicators are very highly positively or highly negatively correlated with the BCI index. This means that a lot of the information contained in these indicators is used. Furthermore, the correlations are fairly similar over the different types of indicators, whether they are surveys or expert opinions. Unlike Arndt and Oman (2006) suggested, all types of sources are represented by the BCI index.

However, there are a number of indicators that are barely correlated with the BCI index. Table 4 filters out the variables whose correlation with the BCI index is less that 0.3 and compares them with the CPI and WGI indicators. For the most part, this group consists of the data from the Business Environment and Enterprise Survey ($y_{35} - y_{38}$). While the correlations with the CPI and WGI indexes are sometimes higher, they are still low, implying that their influence on those indexes is also limited.

Figure 7: Correlations with individual indicators



Correlation of the 43 individual indicators of corruption with the BCI index. ○ denote expert surveys; + household surveys; ◇ firm-level surveys; □ assessments by government or multilateral organizations; and ☆ NGO assessments.

Table 4: Correlations of individual indicators with the combined indexes

	Source	Indicator	BCI	WGI	CPI
y_9	Global Corruption Barometer	Corruption in the media	-0.0462	0.0544	0.0320
y_{28}	Global Integrity	Anti-corruption law	0.0947	0.0848	0.0918
y_{31}	Latinobarometer	Frequency of corruption	0.0646	0.0088	0.0562
y_{35}	BEEPS	Bribe Share	-0.2083	-0.2601	-0.2014
y_{36}	BEEPS	Corruption Frequency	-0.0972	-0.1391	-0.1784
y_{37}	BEEPS	Bribery to get things done	-0.2234	-0.2626	-0.2833
y_{38}	BEEPS	Constraint: corruption	-0.2489	-0.3299	-0.4236

5.1.3 Significant changes in corruption

In order to see whether the level of corruption has changed over time, Kaufmann et al. (2010) suggest as a rule of thumb to look at whether the 90% confidence intervals overlap. If they do not, the change is not big enough to be deemed significant. Using this rule of thumb, they find that from 2000 to 2009 only 8% of their sample experienced a significant change. The same rule is also used for making comparisons of different countries.

The problem is that this rule of thumb ignores the time structure in the corruption data. If corruption did not depend on its previous values, this approximation would return relatively good results. However, most countries have a unit root and as a result the rule makes a lot of type I errors (labeling significant changes as not significant).

Using the data from the Gibbs sampler, we can formally test whether the change in corruption is significant. For example, if in more than 95% of the drawn values of α a country's level of corruption decreases, this change is significant at 5% significance level. This can be extremely useful given the increased importance of changes in governance in for example the allocation international aid (Arndt and Oman, 2006).

Table 5 works out an example. It lists those countries for which there was a significant change in the level of corruption over an (overlapping) 10 year period. Out of a total of 43 changes, the rule of thumb only succeeded in identifying 6, demonstrating the danger of relying on it for policy decisions.

Table 5: Changes in the level of corruption

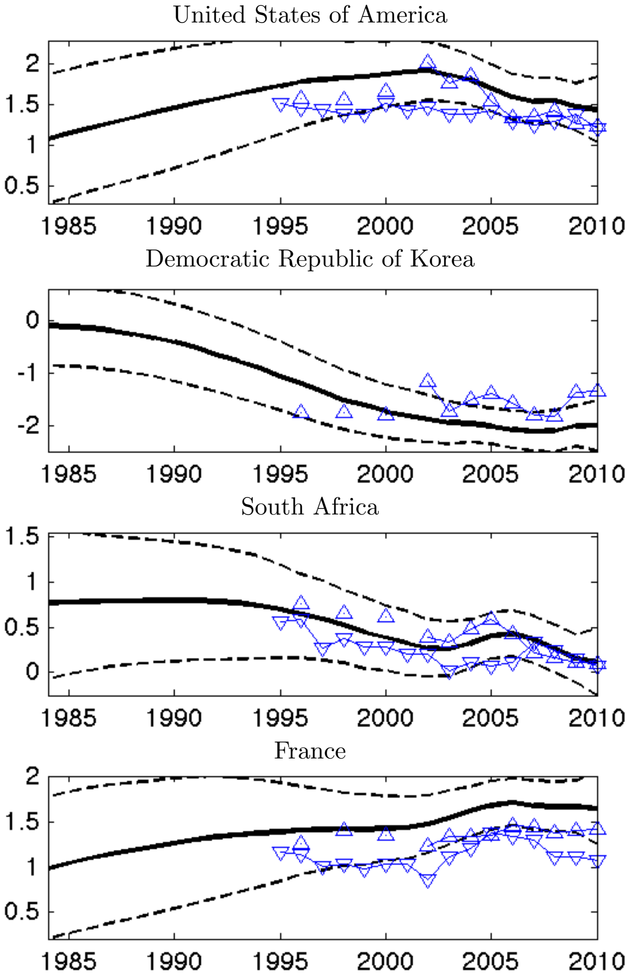
	Deteriorated (BCI decreased)		Improved (BCI increased)	
1984-1995	Iraq		Australia	*New Zealand
	Korea, DR		*Bahamas	Norway
	*Somalia		Canada	Singapore
			*Denmark	*Sweden
			*Finland	*Switzerland
			Iceland	United Kingdom
			Netherlands	
1990-2000	**Korea, DR	*Somalia	**Bahamas	
1995-2005	Brazil	Nicaragua	**Bahamas	Qatar
	*Cote d'Ivoire	Papua New Guinea	Hong Kong	*UAE$^{(a)}$
	*Costa Rica	Somalia		
	*Korea, DR	Zimbabwe		
2000-2009	Greece	*Maladives	Hong Kong	Rwanda
	Madagascar	United States	Iraq	Turkey
			**Qatar	UAE$^{(a)}$

This table list those countries whose level of corruption significantly changed over a 10 year period. Nothing, *, ** indicates significance at 10, 5 and 1%, respectively. $^{(a)}$UAE = United Arab Emirates

5.2 H blockdiagonal

In a second extension to the WGI framework, we allow the measurement error of variables from the same source to be correlated. As before, these results are rescaled such that the expected value has mean zero and standard error of one. These values are referred to as BCI2. As can be seen from figure 8, their plots closely resemble those of the first model. Table 6 confirms that the mean values of the BCI and BCI2 indexes are almost perfectly correlated.

Figure 8: Plot of BCI2 values with 90% confidence interval



Plot of the BCI2 corruption estimates, including 90% confidence interval (dotted lines). Values of the standardized CPI (downwards pointed triangles) an WGI indexes (upward pointing triangles) are also included.

Table 6: Pairwise correlations between BCI2, BCI, CPI and WGI

	BCI2 - WGI	BCI2 - CPI	BCI2 - BCI
Between$^{(1)}$	0.9824	0.9673	0.9965
Within$^{(2)}$	0.3393	0.2107	0.9576
Total	0.9783	0.9710	0.9957

$^{(1)}$ Between correlation is defined as the correlation between the means for each countries; $^{(2)}$ Within correlation is the correlation between the demeaned values for all countries.

However, the biggest difference is not in the mean values, but in the confidence intervals. Table 7 shows the average standard deviation of WGI and both BCI indices. First of all, allowing the standard deviations to be correlated significantly decreases the uncertainty of the BCI corruption estimates. It also decreases the uncertainty relative to the WGI. However, in order to make the right comparison there, BCI indicators first have to be standardized in the same way as the WGI indicator (columns five and six). Secondly, seeing that BCI and BCI2 are available for those countries and years where very little information is available, it is not surprising that their average standard deviation is higher. Restricting the sample to the years when the WGI index is available, it can be seen that the BCI indicators can compute the level of corruption with a greater certainty. This is in spite of the fact that the WGI indicator has access to more, not-publicly-available data.

Table 7: Average standard deviation of WGI, BCI and BCI2

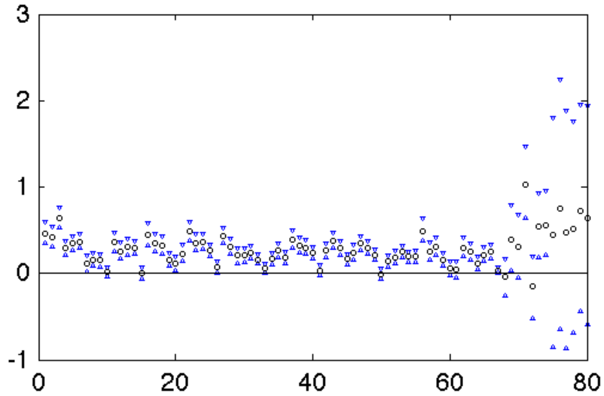
	BCI	BCI2	WGI	BCI yearly$^{(a)}$	BCI2 yearly$^{(a)}$
Total sample	0.3609	0.3387	0.2330	0.4158	0.3886
Year $\geq$ 1996	0.2529	0.2347	0.2330	0.2224	0.2071

Average standard deviation of the BCI, BCI2 and WGI corruption indices. $^{(a)}$These indexes are standardized in the same way as WGI, meaning that *each year* the expected level of corruption has mean zero and standard deviation of one.

5.2.1 Validity H blockdiagonal

Similar to the relationship with the WGI and BCI models, the BCI model can be nested inside that of the BCI2. Setting all non-diagonal elements to zero will render H diagonal and return the BCI model. However, figure 9 shows that the hypothesis that the non-diagonal elements are zero can be rejected in all but 16 out of 80 cases at 1% significance level.

Figure 9: Plot of mean values of the non-diagonal elements of H and their 95% confidence interval



Plot of mean values of the lower triangular elements of H that differ from zero (circles) and 95% confidence intervals (triangles)

5.3 Selection bias

As a number of authors, including Kaufmann et al. (2010), have made clear, the initial values of the WGI and CPI indicators potentially suffer from selection bias issues. Namely, the selection of countries they cover is influenced by those countries' level of corruption. The BCI estimates on the other hand are available for all countries from 1984 onwards, eliminating this bias. Moreover, this makes it possible to formally test whether or not there is a selection bias in the WGI and CPI indicators.

Table 8 lists the results of a logistic regression of the availability of CPI and WGI (1 if available, 0 otherwise) on the values BCI and BCI2 indices. Following Treisman (2007), the real level of GDP from the Penn World tables was also included (Heston, Summers, and Aten, 2012). It confirms the hypothesis of Kaufmann et al. (2010): the higher the level of corruption is the more likely the WGI index is available. The opposite holds for the CPI index, but mostly for its earlier values. Moreover, both indices are also more likely to be available for countries whose GDP is high.

Regardless of the direction of the effect, the fact that the level of corruption influences whether a country is covered or not strongly cautions against using the (early values of the) CPI or WGI indexes in statistical research as they are likely to produce incorrect results.

Table 8: Tests of selection bias in the WGI and CPI indicators

	(1) D_{WGI}	(2) D_{WGI}	(3) D_{CPI}	(4) D_{CPI}
Total sample				
BCI	-0.159***	-	-0.0274	-
	(0.0291)		(0.0302)	
BCI2	-	-0.157***	-	-0.0175
		(0.0292)		(0.0303)
GDP	0.303***	0.305***	1.183***	1.176***
	(0.0515)	(0.0517)	(0.101)	(0.101)
Cnst.	-0.237***	-0.237***	-0.638***	-0.637***
	(0.0304)	(0.0304)	(0.0329)	(0.0329)
Obs	4,893	4,893	4,893	4,893
1966				
BCI	-0.951*	-	0.675***	-
	(0.569)		(0.261)	
GDP	-	-0.935	-	0.658**
		(0.570)		(0.260)
BCI2	1,390**	1,389**	24.93***	24.83***
	(583.6)	(581.4)	(4.516)	(4.515)
Cnst.	-0.596	-0.603	-2.874***	-2.868***
	(0.635)	(0.633)	(0.381)	(0.381)
Obs	188	188	188	188

Logistic regressions of the availability of the WGI and CPI indexes (respectively D_{WGI} and D_{CPI}; 1 if available, 0 otherwise), on the mean values of the BCI and BCI2 indexes, GDP and a constant. Standard errors between brackets. *,**,*** indicates significance at 10, 5 and 1% level

6 Conclusion

The Bayesian Corruption Indicator (BCI) improves on the existing corruption indicators in multiple ways.

From a practical point of view, the BCI indicator can predict the level of corruption with greater certainty while significantly increasing data availability. The possibility of capturing the shared measurement error of certain corruption indicators significantly increases the reliability of the estimates. In addition, by taking changes in methodology and other measurement issues into account, the BCI produces more stable estimates. Most importantly, the BCI index does not suffer from the selection bias issues that plague both the Corruption Perceptions Index (PCI) and the World Governance Indicators (WGI). In the latter two, the level of corruption and GDP significantly influence which countries are covered. Finally, because the estimation of the model returns the entire distribution of corruption, it is also possible to say whether or not the level of corruption significantly increased or decreased in a country.

From an estimation point of view, the underlying assumptions of the BCI model are explicitly stated, making it a very transparent approach. Furthermore, by taking the time-aspect into account, the entire dataset is used in the estimation of each datapoint. This, in combination with the solution to missing data points also eliminates the need for additional assumptions or the need to impute data points. The fact that the individual indicators are entered as is, without any modifications or sub-level aggregations, further increases the objectivity of the BCI index.

Lastly, from a theoretical perspective, the values of the parameters of the state and measurement equation clearly indicate that the additions to the state-space model are valid. Not only does the level of corruption today depend on that of yesterday, in most countries, it clearly shows signs of a unit root. Finally, the variance of the error term also demonstrates that some variables are indeed prone to a shared measurement error. The fact that corruption has a unit root for most countries means that the upmost caution has to be used when it is regressed on other non-stationary series. It also invalidates any regressions on stationary data where the level of corruption was used as they are likely to produce insignificant results.

References

C. Arndt and C. Oman. Uses and abuses of governance indicators. *OECD*, 2006.

C. K. Carter and R. Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

J. Durbin and S. Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press, 2 edition, 2012.

A. Heston, R. Summers, and B. Aten. Penn world table version 7.1. *Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania*, July 2012.

D. Kaufmann, A. Kraay, and P. Zoido-Lobatón. Governance matters iii: Governance indicators for 1996, 1998, 2000 and 2002. *World Bank Economic Review*, 18:253–287, 2004.

D. Kaufmann, A. Kraay, and M. Mastruzzi. The worldwide governance indicators: Answering the critics. *World Bank Policy Research Paper No. 4194*, 2007a.

D. Kaufmann, A. Kraay, and M. Mastruzzi. Governance matter vi: Aggregate and individual governance indicators for 1996-2006. *World Bank Policy Research Paper No. 4280*, 2007b.

D. Kaufmann, A. Kraay, and M. Mastruzzi. The worldwide governance indicators - methodology and analytical issues. *The World Bank*, September 2010.

C.-J. Kim and C. R. Nelson. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. MIT Press, 1999.

G. Koop, D. J. Poirier, and J. L. Tobias. *Bayesian Econometric Methods*. Cambridge University Press, 2007.

J. G. Lambsdorff. The methodology of the 2005 corruption perceptions index. Technical report, University of Passau, 2005.

T. Lancaster. *Introduction to Modern Bayesian Econometrics*. John Wiley & Sons, 2004.

T. Roca. Measuring corruption: perception surveys or victimization surveys. *Groupe d'économie du développement Lare-Efi*, Working paper: DT/167/2011, 2011.

Transparency International. Cpi index - in detail, 2012. URL http://cpi.transparency.org/cpi2011/in_detail/.

D. Treisman. What have we learned about the causes of corruption from ten years of cross-national empirical research? *Annual Review of Political Science*, 10:211–244, 2007.

A Summary of the used corruption indicators

	Source	Indicator	Availability	Mean	St. Dev.
y_1	International Country Risk Guide	Index of corruption	3703	3.0488	1.3507
y_{14}	Country Policy and Institutional Assessment	Transparency, accountability, and corruption in the public sector	381	2.8753	0.6647
y_{27}	Global Integrity	Anti-corruption law	222	91.1604	13.5812
y_{28}		Anti-corruption agency	222	66.6155	19.7827
y_{25}	The Freedom House	Nations in Transit	252	4.3006	1.6691
y_{26}	African Development Bank	Transparency, Accountability and Corruption in public sector	242	3.0775	0.8166
y_{39}	Afrobarometer	Elected leaders	52	0.6868	0.1294
y_{40}		Judiciary	52	0.6659	0.1298
y_{41}		Public officials	52	0.6276	0.1253
y_{42}		Tax/customs officials	52	0.5565	0.1543
y_{33}	Asian Development Bank	Transparency, accountability, and corruption in the public sector	136	3.0551	0.635
y_{34}	BEEPS[(a)]	Bribe Share	110	2.8061	0.788
y_{35}		Corruption Frequency	110	2.5585	1.5191
y_{36}		Bribery to get things done	110	2.7803	0.9804
y_{37}		Constraint: corruption	110	2.7451	0.913
y_{15}	Bertelsmann Transformation Index	Corruption	488	3.4027	2.5294
y_4	Global Risk Service[(b)]	Control of Corruption	1371	0.562	0.2862
y_2	World markets online[(b)]	Control of Corruption	1944	0.5309	0.2548
y_3	Economist Intelligence Unit	Control of Corruption	1481	0.3503	0.3223
y_{13}	Global Corruption Barometer[(c)]	Household bribery	441	0.8739	0.1334
y_{10}		Political parties	448	3.8121	0.5492
y_{11}		Parliament/Legislature	448	3.5335	0.6366
y_9		Media	451	3.0625	0.4291
y_{22}		The military	302	2.8209	0.6639
y_{18}		Education system	308	2.9049	0.6177
y_{12}		Legal system/Judiciary	448	3.4886	0.6996
y_{16}		Medical services	311	3.1405	0.6286
y_{21}		Police	307	3.5463	0.7726
y_{19}		Registry and permit services	308	2.9825	0.6813

Indicator numbering is based on their number of observations, ranking them from most to least; [(a)]Business Environment and Enterprise Performance Surveys; [(b)]Global Insight; [(c)]Transparency International.

	Source	Indicator	Availability	Mean	St. Dev.
y_{17}	BEEPS[a]	Utilities (telephone, electricity, water etc.)	311	2.9045	0.5471
y_{20}		Tax revenue	308	3.186	0.6333
y_{29}		Customs	319	1.763	1.8226
y_{32}		Public Officials	431	1.1732	1.739
y_5	Global competitiveness survey[d]	Control of Corruption	1121	0.55	0.1939
y_8	Gallup World Poll	Control of Corruption	533	0.3386	0.1937
y_7	IFAD Rural Sector Performance Assessments	Accountability, transparency and corruption in rural areas	630	3.6424	0.6193
y_{23}	Institutional Profiles database	Petty corruption	257	2.144	1.058
y_{24}		Large-scale corruption	257	2.1751	0.9676
y_{30}	Latinobarometer	Frequency of corruption	159	0.803	0.14
y_{31}	Political and Economic Risk Consultancy in Asia	Control of Corruption	142	0.3846	0.2483
y_{38}	Vanderbilt University's Americas Barometer	Frequency of Corruption	56	0.2729	0.0791
y_6	World Competitiveness Yearbook[e]	Bribing and corruption	810	4.6814	2.672

Indicator numbering is based on their number of observations, ranking them from most to least; [a]Business Environment and Enterprise Performance Surveys [b]Global Insight [c]Transparency International [d]World Economic forum [e]Institute for Managed Development

B Estimation

B.1 Priors

Because the model is estimated in a Bayesian framework, it is necessary to specify the prior distribution of the parameters. However, since there is no prior information available on the parameters of the measurement equation, Z, C and H, flat probabilities are used for these variables. This means that all values in $\mathbb{R}$ are equally probable ($\mathbb{R}^{\geq 0}$ in the case of the diagonal elements of the variance matrix H). It is important to note that the WGI or CPI indexes cannot be used as sources of prior information, seeing that they are based on the same data used in the estimations.

$$p(Z^j) \propto 1 \tag{8}$$
$$p(C^j) \propto 1 \tag{9}$$
$$p(\log(H)) \propto \mathbb{I}_k \tag{10}$$

with $j = 1, ..., k$, and k the total number of individual indicators.

For the state equation, there is a prior restriction on T_i that its absolute value does not exceed one, for all countries i. This ensures that $\alpha_{i,t}$ is a non-explosive time series, without precluding non-stationary series. Apart from the intuitive reasons for this restriction, it is also necessary from a practical point of view: not imposing it causes the model to no longer converge.

$$p(T) \quad = \quad 0.5 * \mathbb{1}_{|T| \leq 1} \tag{11}$$

Finally, as an identifying assumption, the variance of the state equation, Q, is set to one.

B.2 Gibbs sampler

As was explained, the Gibbs sampler allows us to split the estimation process up in two main blocks, which can then be further divided into a number of easily solvable subroutines:

A. Conditioning on the values for $\alpha = (\alpha_{1,0}, ..., \alpha_{1,n}, ..., \alpha_{p,1}, ..., \alpha_{p,n})'$, the state and measurement equation are reduced to simple linear regressions:

$$\begin{aligned} p(T|\alpha, Q) &\propto .5 * \mathbb{1}_{|T| \leq 1} \\ & \quad N[(\alpha'_{t-1}\alpha_{t-1})^{-1}(\alpha'_{t-1}\alpha_t);\ (\alpha'_{t-1}\alpha_{t-1})^{-1}Q] && (12) \\ p(Z, C|\alpha, y, H) &\propto N[([\alpha, 1]'[\alpha, 1])^{-1}([\alpha, 1]'y); ... \\ & \qquad ([\alpha, 1]'[\alpha, 1])^{-1}H] && (13) \\ p(H|\alpha, y) &\propto iWishart[e'e;\ n] && (14) \end{aligned}$$

with $e = y - T\alpha$ and $iWishart$ the inverse Wishart distribution.

When H is block-diagonal, equations 13 and 14 stay the same. Equation 13 does not have to be estimated using Seemingly Unrelated Regressions since the regressors in all equations are the same : $[\alpha, 1]$. This means that the SUR models is equivalent to equation by equation OLS.

B. Conditional on the parameters of the state and measurement equations, the probability of the 'true' level of corruption can be computed and drawn from using the Carter and Kohn (1994) simulation smoother (Kim and Nelson, 1999).

- *The Kalman filter*
  Starting from a wild guess, $p(\alpha_0) = N(0, \infty)$, the following equations are iteratively solved for $t = 1$ to $t = n$:

$$\begin{aligned} a_{t|t} &= E(\alpha_t|y_1, ..., y_t) \\ &= T * a_{t-1|t-1} + \kappa(y_t - C - ZTa_{t-1|t-1}) && (15) \\ p_{t|t} &= V(\alpha_t|y_1, ..., y_t) \\ &= p_{t|t-1} + \kappa Z p_{t-1|t-1} && (16) \end{aligned}$$

  with $\kappa = p_{t|t-1}Z'(Zp_{t|t-1}Z' + H)^{-1}$; and $p_{t|t-1} = Tp_{t-1|t-1}T' + Q$.

- *Simulation smoother*
  The simulation smoother algorithm is used to draw values for α for each countries one at a time. Starting from the last iteration of the

Kalman filter, draw $\hat{\alpha}_n$ from $N(a_{n|n};\ p_{n|n})$ and iterate backwards from $t = n - 1$ to $t = 1$:

$$\begin{aligned} a_{t|n} &= E(\alpha_t | y_1, ...y_n) \\ &= a_{t|t} + \varsigma(\hat{a}_{t+1|n} - T a_{t|t}) \qquad (17) \\ p_{t|n} &= V(\alpha_t | y_1, ...y_n) \\ &= p_{t|t} + \varsigma(p_{t+1|n} - T p_{t|t} T' - Q)\varsigma' \qquad (18) \end{aligned}$$

with $\varsigma = p_{t|t} T' p_{t+1|t}^{-1}$; and $\hat{a}_{t+1|n}$ a random draw from $N(a_{t+1|n};\ p_{t+1|n})$.

B.3 Standardization

Setting the variance of the state equation, Q, to one gives us mean values for α that lie between -12 and 14. These were normalized such that the expected value for all countries has mean zero and standard deviation one. Each drawn value of α is modified in the following way:

$$BCI^{(j)} = \frac{\alpha_{i,t}^{(j)} - \bar{\alpha}}{\left(\frac{1}{p}\sum_{i=1}^{p}\frac{1}{n}\sum_{t=1}^{n}\left(\bar{\alpha}_{i,t} - \bar{\alpha}\right)^2\right)^{\frac{1}{2}}}$$

with $\alpha_{i,t}^{(j)}$ the value of α for country i at time t in the j^{th} iteration; $\bar{\alpha}_{i,t}$ is the mean of alpha over all iterations, and $\bar{\alpha}$ is the mean over all iterations, years and countries.

The BCI index for country i at time t and its variance is then respectively the mean and variance of $\mathrm{BCI}_{i,t}^{(j)}$ with respect to j.

yearly standardization

The normalization used in the WGI on the other hand is such that the mean values for all countries has a *yearly* mean of zero and standard deviation one.

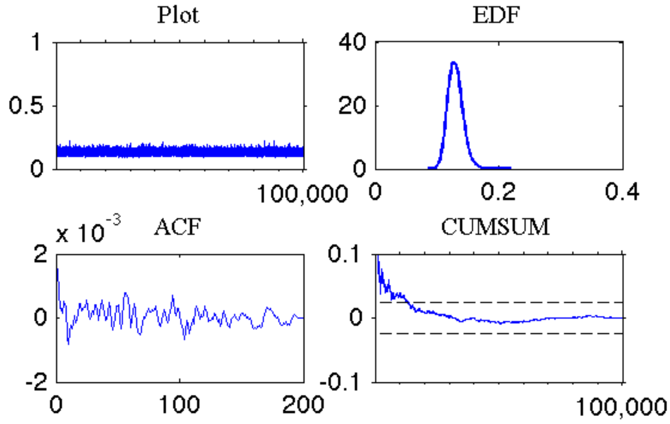
$$BCI^{(j)} = \frac{\alpha_{i,t}^{(j)} - \frac{1}{p}\sum_{i=1}^{p}\bar{\alpha}_{i,t}}{\left(\frac{1}{p}\sum_{i=1}^{p}\left(\bar{\alpha}_{i,t} - \frac{1}{p}\sum_{i=1}^{p}\bar{\alpha}_{i,t}\right)^2\right)^{\frac{1}{2}}}$$

B.4 Convergence

For both models (H diagonal and H block-diagonal), the Gibbs sampler ran 100,000 iterations of which the first 50,000 were discarded as burn-in.

The results of the index only make sense if their probability distributions have converged. Because it is too time-consuming to check the convergence of the close to 6000 values of corruption, the focus was instead on the still close to 300 parameters of the state and measurement equation. Convergence was checked using simple plotted values, autocorrelation functions and a rolling window CUMSUMs. As an example, figure 10 shows these graphs for the sixth

Figure 10: Convergence statistics for $H_{(6,6)}$



Top left: simple plot of all drawn values; top right: the empirical distribution function; bottom left: the autocorrelation function; and bottemright: the rolling window CUMSUM statistic, with 5% significance bounds (window: 1000 draws).

diagonal element of the H matrix, as well as its empirical distribution function. All plots point to a well-behaved, converged distribution, which is what is found for the other parameters as well.