Check for
updates

Corruption around the world: an analysis by partial least squares—structural equation modeling

Roberto Dell'Anno[1]

Abstract
This paper contributes to the empirical research on corruption in three ways. From a methodological viewpoint, it applies partial least squares–structural equation modeling to estimate an index of perceived corruption around the world—hereinafter structural corruption perception index (S-CPI). This approach allows one to estimate corruption as a multidimensional latent variable by complex cause-effect relationships between observed and/or unobserved variables. From a positive viewpoint, it estimates comparable S-CPI scores in 165 countries from 1995 to 2016, using a model specification based on existing theory of and empirics on the causes and consequences of corruption. In terms of policy implications, helpful hints on which are the most effective channels for fighting corruption are provided.

1 Introduction

Corruption is usually defined as "the abuse of public office for private gain" (World Bank 1997, p. 8). Extensive scholarly research has identified the several effects of corruption on socioeconomic systems. In particular, since the late 1990s, the empirical economics literature has exponentially[1] expanded owing to the raising quality and availability of data on (perceived) corruption. This literature highlights three main criticisms. The first one refers to the reliability of the indexes on (perceived) corruption utilized to describe the

---

[1] Searching for the word "corruption" in the titles of the documents indexed by Scopus database (Subject Areas: Social Sciences; Economics, Econometrics and Finance), returns 366 responses between 1989 and 1998, 1480 in the 1999–2008 period and 5005 in the decade running from 2009 to 2018.

✉ Roberto Dell'Anno
  rdellanno@unisa.it

[1] Department of Economics and Statistics – CELPE, University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, SA, Italy

magnitudes of corrupt activities. A recent critical viewpoint raises significant doubts about whether the perceptions-based indicators are reliable proxies for actual corruption (e.g., Seligson 2006; Razafindrakoto and Roubaud 2010; Donchev and Ujhelyi 2014; Treisman 2015; Ning 2016). For instance, Treisman (2015) considers the differences in countries' perceived corruption scores as for the most part correlated with national cultural stereotypes or with wider media coverage of, e.g., corruption scandals, rather than the actual extent of corrupt activities.

A second criticism refers to the common practice of treating "corruption as unidimensional and as synonymous to bribery" (Philp 2015, p. 19). According to Andersson (2017), other forms of corruption (e.g., favoritism, improper interference, conflicts of interest) usually more common in developed countries are partially neglected by usual corruption-perception indexes that focus essentially on bribery. A growing body of literature has pointed out the existence of different "forms" of corruption. For instance, Dincer and Johnston (2019) distinguish between legal and illegal corruption, relying on the nature of public official's gains in exchange for providing specific benefits to private individuals or groups. Specifically, illegal corruption occurs when public office is abused for private gains in the form of cash or gifts to a government official. On the other hand, legal corruption occurs when the abuse of power is for political gains in the form of campaign contributions to or endorsements by a government official (e.g., lobbying activity).[2] A different taxonomy of corruption has also been proposed, e.g., "High level" or "Grand" corruption versus "Low level" or "Petty" corruption. "Grand" corruption refers to misconduct at the top by leading politicians and that category comprises both illegal and legal corruption. "Petty" corruption refers to underhand payments to expedite administrative procedures: bribes to avoid fines or to "speed up" waiting lists for public services, and so on); it usually involves administrators and bureaucrats. Accordingly, taking into account that the degree to which bribery can serve as a proxy for overall corruption varies depending on the nature of a political system and the extent of economic development, Andersson (2017) concludes that, in established democracies with highly developed economies and low corruption, the accuracy of conventional perceived-corruption indexes[3] may be particularly poor.[4]

The third criticism refers to the evidence that older and more recent studies on corruption often contradict each other. Consequently, doubts arise about the reliability of the estimated corruption indexes because of statistical inconsistencies. The predominant explanation for the conflicting findings points out that as, at least partially, the discrepancies are consequences of the more sophisticated econometric approaches, larger datasets, or both, available for recent analyses (Dimant and Tosato 2018).

---

[2] It is beyond the scope of the present research to examine the relationship between lobbying and illegal corruption. For a review of that literature, see Lambsdorff (2002), Damania et al. (2004), Campos and Giovannoni (2007), Gokcekus and Sonan (2017) and Goldberg (2018).

[3] The deduction is based on analysis of Corruption Perceptions Index of Transparency International and the Control of Corruption variable of World Bank in the Sweden context. Specifically, "In such settings, bribery is more likely only the tip of the corruption iceberg, and undue influence and conflicts of interest are more frequent occurrences" (Andersson 2017, p. 70).

[4] Kaufmann and Vicente (2011) point out that, mainly for developed countries, inadequate empirical attention has been paid to legal types of corruption. Recently, Gokcekus and Sonan (2017) and Dincer and Johnston (2019) contribute to filling that gap by estimating the sizes of and the relationship between legal and illegal corruption in a cross-state panel for the United States. Unfortunately, at this time, no cross-country panel data are available for extending their analysis to a global level.

From a methodological viewpoint, the present article aims to contribute to the debate by focusing on the last two criticisms above. In particular, in order to deal with the second criticism, I apply a statistical approach that considers corruption to be a multidimensional phenomenon. As such, I aim to estimate an overall perception of corruption index (i.e., taking account of both "legal" and "illegal" as well as "grand" and "petty" corruption).[5]

As the third criticism concerns, I apply an estimation method—i.e., Partial Least Squares estimation approach to Structural Equation Modeling (PLS–SEM)—which has two main advantages over the previous empirical analyses. First, it is able to translate into testable relationships the economic hypotheses regarding the causes and consequences of corruption by means of a unified statistical approach—the so-called "structural model" of the SEM. The second advantage of a SEM consists in treating (perceived) corruption as an unobservable variable (i.e., a latent construct) that interacts in complex ways with several other unobserved socioeconomic factors (e.g., institutional variables) and observable variables. In that sense, I aim to improve reliability of the estimates of perceived corruption by reducing measurement errors.

To the best of my knowledge, the study at hand is the first attempt to estimate an index of perceived corruption by PLS–SEM[6]—hereinafter a *structural corruption perception index* (*S-CPI*).

Essentially, PLS–SEM is a system of interdependent equations estimated using both factor analysis and multiple regression techniques until the model converges adequately by an iterative method.

From a positive viewpoint, the contribution of the present study consists in providing an updated, wide-ranging and comparable meta-index of perceived corruption for 165 countries using annual data over the 1995–2016 period.

In terms of policy implications, I will identify the main factors affecting corruption—by decomposing the total effect of the causes on corruption in direct and indirect effects—and which of them are the most effective channels for fighting corruption by conducting an importance-performance map analysis.

The paper is organized as follows. The next section summarizes the causes and consequences of corruption, providing the theoretical background for model specification. Section 3 explains the empirical approach and provides a formal representation of the PLS–SEM. Section 4 reports empirical results and discusses the findings and policy

---

[5] An active debate is underway about whether legal (e.g., lobbying, political contribution) and illegal corruption are substitutes or complements; the results are still inconclusive (Shepsle 2017; Goldberg 2018). The main argument that they are substitutes relies on the idea that lobbying enables the lobbyist to change the rules, thus making corruption redundant (Harstad and Svensson 2011). The rationale that they are complements relies on the idea that legal and illegal corruption may be considered to be two sides of the same coin: on the one hand, legal corruption may be seen as a long-term investment aimed at influencing politicians to change the rules of the game; on the other hand, illegal corruption may be considered to be a short-term investment, directed to influencing public officials to find ways around the existing rules (Gokcekus and Sonan 2017).

[6] Similar to the research herein, Dreher et al. (2007) estimate an index of perceived corruption with structural equation modeling—however, several differences arise in terms of: (1) estimation method—they estimate the model by a covariance-based approach, while I apply a PLS approach; (2) model specification—they estimate corruption with a multiple indicators and multiple causes (MIMIC) model, while I apply a broader structural model specification; (3) exhaustiveness of measurement and structural models - they define one latent variable (i.e., corruption) with five observable causes and four observable indicators, while I define 11 latent variables and, for each of those constructs, I specify a distinct measurement model, implying 47 manifest variables; (4) extensiveness of corruption indexes—Dreher et al. (2007) estimate an index of corruption that covers 100 countries over the 1976–1997 period; the index herein covers 165 countries over the 1995–2016 period.

implications. Section 5 concludes. Two online appendixes describe the dataset and report annual *S-CPI* scores for all 165 countries.

2 Theoretical background causes, consequences and indicators of corruption

Duncan (1975, p. 149), describing the SEM, stated that "the meaning of the latent variable depends completely on how correctly, precisely and comprehensively the causal and indicator variables correspond to the intended semantic content of the latent variable". Thus, the reliability of the estimates of the key latent variable (i.e., perceived corruption) depends completely on what causes and consequences are selected in specifying the model. Accordingly, following the literature on corruption[7] and data availability, I specify a model with nine latent variables[8] and 42 observed indicators. In SEM terminology, the system of statistical relationships explaining how latent variables (causes, consequences and indicators of perceived corruption) are related with each other is defined as the structural or "inner" model. The systems of equations—so-called "blocks"—in which each latent variable is connected to a subset of manifest variables constitute the measurement or "outer" model—in the SEM. Table 1 summarizes the main theoretical hypotheses supporting the specification of structural model.

According to the unobservable and/or multidimensional nature of the potential causes and consequences of corruption, I define the latent variables as "reflective" (i.e., the observed indicators of a construct are considered to be caused by that construct) or as "formative" (i.e., the manifest variables are considered to be the causes of the latent variable). For the sake of brevity, in the next section, I report details on the measurement model of the key latent construct only (i.e., corruption). As for the measurement models of other latent variables, the definitions and sources of observations on all manifest variables are provided in the Appendix A1.

2.1 Indicators of corruption: the measurement model for the S-CPI index

The latent variable "Corruption" (*S-CPI*) is measured by a reflective model based on some of the most widely known cross-country indexes that account for the magnitude of perceived corruption as reflected in the opinions of panels of national experts and business people. Specifically, the five indicators are: (a) the *Corruption Perceptions Index* published by the Transparency International (2017) (*CPI Rev*); the original index—perceptions of the extent of corruption as seen by business people, risk analysts and the general public—is rescaled so that the scores are higher scores when the level of perceived of corruption

---

[7] In chronological order, I refer the reader to Ades and Di Tella (1997), Tanzi (1998), Rose-Ackerman (1999), Jain (2001), Aidt (2003), Svensson (2005), Lambsdorff (2006, 2007), Serra (2006), Treisman (2007, 2015), Goel and Nelson (2010), Enste and Heldman (2017) and Dimant and Tosato (2018).

[8] More specifically, the model includes 20 latent constructs but 11 of these constructs have a single indicator with a loading coefficient fixed equal to 1 (see Table 2). Accordingly, those 11 (formative) constructs are equal to their corresponding single manifest indicators. These specifications of measurement models make it possible to estimate the path coefficients of observable variables (i.e., oil rents, decentralization, colonial and religion dummies—see Table 3) of structural models that, by definition, only include latent constructs.

Table 1 Theoretical hypotheses on the causes and consequences of corruption

Hp	Theoretical hypothesis	References	According to this research[a]
1	High levels of media freedom (*MediaFreed*) reduce corruption	Enste and Heldman (2017), Brunetti and Weder (2003), Kalenborn and Lessmann (2013)	Yes
2	Higher levels of education (*Educ*) reduce corruption	Melgar et al. (2010)	Yes
3	Higher quality of democracy (*Democracy*) reduces corruption	Sandholtz and Koetzle (2000), Treisman (2000), Kunicova and Rose-Ackerman (2005), Pellegata (2012)	Yes
4	Higher quality of regulation (*Regulation*) reduces corruption	Tanzi (1998)	Yes
5	Abundance of natural resources (*NaturalRes*; *OilRent*) increases corruption	Sachs and Warner (1997), La Porta et al. (1999), Ades and Di Tella (1999), Bhattacharyya and Hodler (2010)	No; Yes (Oil)
6.a	Higher size of public sector (*SizePubSec*) increases corruption	Goel and Nelson (1998), Lambsdorff (2006), Arvate et al. (2010)	No
6.b	Higher size of public sector (*SizePubSec*) reduces corruption	Goel and Budak (2006), Goel and Nelson (2010)	Yes
7	Higher degree of decentralization (*Decentraliz*) is associated with lower levels of corruption	Shleifer and Vishny (1993), Fisman and Gatti (2002), Arikan (2004), Lessmann and Markwardt (2010), Dell'Anno and Teobaldelli (2015)	No
8	The more fractionalized (*Fractionaliz*) is the society, the higher is corruption	Mauro (1995), Pellegrini and Gerlagh (2008), Treisman (2000), Dincer (2008)	Yes
9	British colonial heritage (*British, US, AustralC*) is associated to lower corruption than countries with a different colonial heritage.	Treisman (2000), Swamy et al. (2001), Serra (2006), Goel and Nelson (2010)	No/yes
10	Higher the belonging of hierarchical religions (i.e. *Protest* [−] versus *Cath* [+], *Musl* [+]) the higher is corruption	La Porta et al. (1997), Treisman (2000), Serra (2006), Paldam (2001), North et al. (2013)	Yes
11	Higher levels of corruption reduce socio-economic development (*SocioEconDev*)	Tanzi and Davoodi (2001), Aidt (2009)	Yes

In parenthesis the labels of latent variables

[a]The last column answer to the question: is this hypothesis validated by this empirical analysis?

increases. (b) The *Bayesian Corruption Indicator* estimated by Standaert (2015) (*Bayesian Corr*). It is a composite index of the perceived overall level of corruption combining information from 20 different surveys and more than 80 different survey questions. (3) The *Political Corruption index* (*Political Corr*) is equal to the average of *public sector corruption index* as estimated by Coppedge et al. (2017) and Pemstein et al. (2017) in the "Varieties of Democracy (V-Dem)" Project. (4) The *Control of Corruption index* (*Control Corr. Rev*) is extracted from the Worldwide Governance Indicators database of the World Bank and measures perceptions of corruption. (5) *Freedom from corruption* (*Freedom Corr. Rev*) is the (rescaled) index of *Freedom from corruption* published by the Heritage Foundation (2017). (6) The (rescaled) "ICRG Indicator of Quality of Government" (*ICRG. Rev*) included in the International Country Risk Guide indicators and produced by the PRS Group (2018). It is calculated as the complement to the mean of the ICRG variables "Corruption", "Law and Order" and "Bureaucracy Quality".

3 The statistical approach: partial least squares: structural equation modeling

SEM is a multivariate statistical approach that subsumes a whole range of standard multivariate analytical methods, including regression and factor analysis. It enables the researcher simultaneously to estimate complex causal relationships among latent (unobservable) and manifest (observable) variables. SEM is extensively applied in different fields, such as business, marketing, management, psychology, social and, more recently, in macroeconomics research (e.g., Dell'Anno 2007; Dreher et al. 2007; Ruge 2010; Dell'Anno and Dollery 2014; Buehn et al. 2018).

Two approaches to estimating a SEM are possible: a covariance-based approach (CB–SEM) and partial least squares (PLS–SEM).[9] The differences between CB–SEM and PLS–SEM estimation methods of SEM parameters mainly relate to different data characteristics and the researcher's objectives (Richter et al. 2016). According to Faizan et al. (2018), PLS–SEM is especially promising when both the assumption of a multinormal distribution is violated and the theory relied on to explain the phenomenon requires modelling complex interactions with many latent constructs. For Esposito Vinzi et al. (2010b), PLS–SEM has the advantage, compared to the CB-SEM, that no strong assumptions with respect to the distributions, sample size and measurement scale, are required. However, those advantages must be considered in light of some disadvantages. For example, the absence of any distributional assumptions implies that scholars cannot rely on the classic parametric inferential framework (Chin 1998; Tenenhaus and Esposito Vinzi 2005). PLS–SEM in fact applies the jackknife and bootstrap resampling methods to derive empirical confidence intervals and for testing hypotheses on statistical coefficients. For this reason, "the emphasis [of PLS] is more on the accuracy of predictions than on the accuracy of estimation" (Esposito Vinzi et al. 2010b, p. 52). Similarly, Shmueli et al. (2016) state that PLS–SEM, by focusing on the explanation of variances rather than covariances, makes it a prediction-oriented approach to SEM. Another drawback is that the absence of a global

---

[9] Extensive reviews of the PLS approach to SEM are given in Chin (1998), Tenenhaus and Esposito Vinzi (2005), Esposito Vinzi et al. (2010a), Hair et al. (2016, 2017, 2019) and Faizan et al. (2018). The benefits and limitations of partial least squares path modeling (PLS) is still an open issue. On opposite side of the debate is Rönkkö et al. (2016).

optimization criterion in PLS–SEM implies the absence of measures of overall model fit. The lack of such measures limits PLS–SEM's usefulness for theory testing and for comparing alternative model structures (Hair et al. 2012). Hair et al. (2019) provide some guidelines to identify the best approach to estimate a SEM model. Following Hair et al.'s (2019) hints,[10] I consider the PLS approach as preferable to the CB method for estimating the proposed SEM.

3.1 The PLS–SEM model for estimating the structural corruption perception index

In this section, I provide a formal representation of the PLS–SEM based on existing theory and empirical evidence on the causes and consequences of corruption. Moreover, the structural (or inner) model of PLS–SEM allows one to model both the direct effects of the "causes" of corruption, but also the interactions among them. Accordingly, the inner model of PLS–SEM specification may be described by the system of Eq. (1):

$$
\begin{aligned}
S\text{-}CPI_{it} = {}& \beta_{21} MediaFreed_{it} + \beta_{31} Educ_{it} + \beta_{41} Democracy_{it} + \beta_{51} Regulation_{it} \\
& + \beta_{61} NaturalRes_{it} + \beta_{71} OilRent_{it} + \beta_{81} SizePubSec_{it} + \beta_{91} Decentraliz_{it} \\
& + \beta_{10,1} Fractionaliz_{it} + \beta_{11,1} FrenchC_{it} + \beta_{12,1} PortugC_{it} + \beta_{13,1} SpanC_{it} \\
& + \beta_{14,1} ItalianC_{it} + \beta_{15,1} Bel\&DutC_{it} + \beta_{16,1} British, US, Austl_{it} \\
& + \beta_{17,1} Catholic_{it} + \beta_{18,1} Protestant_{it} + \beta_{19,1} Muslim_{it} + \varsigma_{1,it} \\
MediaFreed_{it} = {}& \beta_{32} Educ_{it} + \varsigma_{2,it} \\
Democracy_{it} = {}& \beta_{24} MediaFreed_{it} + \beta_{34} Educ_{it} + \varsigma_{4,it} \\
Regulation_{it} = {}& \beta_{35} Democracy_{it} + \varsigma_{5,it} \\
NaturalRes_{it} = {}& \beta_{76} OilRent_{it} + \varsigma_{6,it} \\
Decentraliz_{it} = {}& \beta_{10,9} Fractionaliz_{it} + \varsigma_{9,it} \\
Catholic_{it} = {}& \beta_{11,17} FrenchC_{it} + \beta_{12,17} PortugC_{it} + \beta_{13,17} SpanC_{it} + \beta_{14,17} ItalianC_{it} + \varsigma_{17,it} \\
Protestant_{it} = {}& \beta_{15,18} Bel\&DutC_{it} + \beta_{16,18} British, US, Austl_{it} + \varsigma_{18,it} \\
SocEconDev_{it} = {}& \beta_{1,20} S\text{-}CPI_{it} + \beta_{3,20} Educ_{it} + \beta_{6,20} NaturalRes_{it} + \varsigma_{20,it},
\end{aligned}
$$

(1)

where the subscript $i = 1,...,165$ indicates the country and $t = 1995,...,2016$ denotes the year.

In the system (1), the first equation accounts for the direct effects of the causes on corruption; in the second through eighth equations, I model the interactions among the causes of corruption. The path-coefficients estimated in those seven equations allow me to estimate indirect (mediated) effects of the causes on *S-CPI*. Finally, the last equation accounts for the consequences of corruption on the socioeconomic system. It is included in the model to improve the reliability of the estimates in accordance with Duncan's (1975) remark (i.e., the meaning of the latent variable depends completely on how precisely I select causes and indicators in the SEM specification). In that sense, including within the

---

[10] Specifically (1) I aim to predict an index of perceived corruption; (2) the network of relationships between corruption and its potential economic, cultural, and institutional determinants is complex; (3) the specified model includes more formatively measured constructs; (4) the availability of several alternative indicators for measuring variables that intrinsically are unobservable and/or multidimensional and (5) the violation of the multivariate normality assumption.

empirical model the effect of corruption on socioeconomic development allows me to better describe the target latent construct (i.e., perceived corruption).

The dataset used for the empirical analysis is extracted from "The Quality of Government Standard Dataset" collected by Teorell et al. (2018). All variables are scaled to have zero means and unit variances.

Owing to the prediction-oriented focus of the proposed SEM, I deal with missing values in the dataset by applying different missing data treatments. First, I apply the pairwise deletion method.[11] That option is chosen because it retains as much information as possible. The second treatment is based on interpolating the missing values. I use three different datasets in that empirical analysis as a function of the replacement used: (1) a dataset, labelled "MV", wherein pairwise deletion is applied with no replacement; (2) a dataset, labelled "I", for which I first replace missing values by linear interpolation—i.e., calculated using the last valid value before the missing value and the first valid value afterwards—later I apply pairwise deletion; (3) a dataset, labelled "IFB", in which I apply, in the following order, linear Interpolation (I), "forward" interpolation (i.e., I use the last observed value to replace subsequent missing values of the same country) and "backward" interpolation (i.e., I impute the newest observation to replace earlier missing observation of the same country) and, lastly, pairwise deletion.

4 Empirical results

I estimate several PLS–SEM specifications[12] using three missing data treatments (MV, I and IFB). Taking into account that the results are robust to alternative missing data treatments and model specifications, for the sake of brevity, I report estimates based only on the IFB dataset and two models: the broadest model specification (Model 1) and a restricted model (Model 2) in which the determinants of corruption that cannot be affected by policy-makers (i.e., *Colonial Heritage* and *Religion belonging*) and the "consequence" of corruption (i.e., *Socioeconomic Develop)* are excluded in order to focus on normative interpretations. Accordingly, Model 1 is predictive (i.e., to explain the S-CPI index), while Model 2 is applied to derive policy implications by conducting an importance-performance map analysis (IPMA) (Ringle and Sarstedt 2016).

Once the SEM models have been specified and the PLS-algorithm generates the estimates,[13] Hair et al. (2019) suggest first to evaluate the reliabilities and validities of the latent variables in the outer models and, only if the outer models are reliable, evaluating the reliability of inner model.

Accordingly, to assess the reflective outer models, I test: (1) the reliability of reflective indicator—outer loadings should be larger than 0.708; (2) internal consistency

---

[11] That option deletes those observations for which values are missing in each pair of manifest variables.

[12] The estimates are calculated by the "SmartPLS 3.0" software developed by Ringle et al. (2015).

[13] A controversial issue of bias (and potential remedies) arises when use PLS–SEM to estimate reflective models. According to Sarstedt et al. (2016), the PLS algorithm is preferable to CB and PLSc, when it is not known whether the data's nature is common factor- or composite-based. Other studies (e.g., Dijkstra and Henseler 2015), state that PLSc is preferable to PLS. In the following, I choose to report PLS estimates instead of consistent PLS estimates (PLSc) and a bootstrapping routine applied to correct estimated coefficients on the reflective constructs (Dijkstra and Henseler 2015). That choice is supported by evidence that the differences between findings based on PLSc and PLS estimates are negligible and that the latent scores used to calculate the S-CPI are not affected by the decision.

Table 2 Outer loadings and weights (*p* values in parenthesis)—assessment statistics

Reflective construct	Manifest var.	Model 1			Model 2		
		Loadings	ρ_A	AVE	Loadings	ρ_A	AVE
Corruption	Bayesian corr.	0.914 (0.000)	0.98	0.91	0.907 (0.000)	0.97	0.91
	Freedom corr. rev.	0.904 (0.000)			0.902 (0.000)		
	ICRG rev.	0.949 (0.000)			0.948 (0.000)		
	CPI rev.	0.942 (0.000)			0.933 (0.000)		
	Political corr.	0.886 (0.000)			0.879 (0.000)		
	Control corr. rev.	0.949 (0.000)			–		
Education	Educ. 25–34 fem.	0.929 (0.000)	0.93	0.86	0.927 (0.000)	0.93	0.88
	Second. educ. fem.	0.942 (0.000)			0.939 (0.000)		
	Tertiary educ.	0.907 (0.000)			0.911 (0.000)		
Media freedom	Press freedom	0.886 (0.000)	0.79	0.71	0.886 (0.000)	0.79	0.71
	Media corr.	0.888 (0.000)			0.888 (0.000)		
	Internet access	0.708 (0.000)			0.709 (0.000)		
Quality of democracy	Institutional. dem.	0.876 (0.000)	0.84	0.76	0.876 (0.000)	0.84	0.76
	Participatory dem.	0.931 (0.000)			0.931 (0.000)		
	Political stability	0.748 (0.000)			0.748 (0.000)		
Quality of regulation	Econom. freedom	0.917 (0.000)	0.85	0.81	0.917 (0.000)	0.85	0.81
	Business freedom	0.770 (0.000)			0.771 (0.000)		
	Intern. trade freed.	0.908 (0.000)			0.908 (0.000)		
Size of public sect.	Gov. tax. ent.	0.853 (0.000)	0.80	0.71	0.852 (0.000)	0.80	0.71
	Gov. freedom. rev	0.805 (0.000)			0.805 (0.000)		
	Gov. consumption	0.795 (0.000)			0.796 (0.000)		
SocioEconomic dev.	Human capital I.	0.909 (0.000)	0.95	0.81	–	–	–
	Human develop. I.	0.951 (0.000)			–		
	Access electricity	0.848 (0.000)			–		
	Real GDP percapita	0.722 (0.000)			–		
	Life expectancy	0.894 (0.000)			–		
Formative construct	**Manifest var.**	**Weights**		**VIF**	**Weights**		**VIF**
Fractionalization	Ethnic fract.	0.958 (0.000)		1.9	0.966 (0.000)		2.0
	Language fract.	0.081 (0.031)		2.1	0.069 (0.059)		2.1
	Religion fract.	−0.324 (0.000)		1.1	−0.319 (0.000)		1.1
Natural resources	Gas prod. value	−0.150 (0.000)		1.5	−0.228 (0.000)		1.5
	Oil prod. value	1.082 (0.000)		1.5	1.118 (0.000)		1.5
Oil rents	Oil rents	1(fixed)		–	1(fixed)		–
Decentralization	Autonom. reg.	1(fixed)		–	1(fixed)		–
British, US, Austral. c.	British, US, Austral. c.	1(fixed)		–	–		–
Belgian, Dutch c.	Belgian, Dutch c.	1(fixed)		–	–		–
French colony	French colony	1(fixed)		–	–		–
Never colony	Never colonized	1(fixed)		–	–		–
Portuguese colony	Portuguese c.	1(fixed)		–	–		–
Spanish colony	Spanish c.	1(fixed)		–	–		–
Catholic	% of catholics	1(fixed)		–	–		–
Muslim	% muslims	1(fixed)		–	–		–

Table 2 (continued)

Formative construct	Manifest var.	Weights	VIF	Weights	VIF
Protestant	% protestants	1(fixed)	–	–	–

reliability—ρ_A falls between the thresholds 0.70 and 0.95; (3) convergent validity—the *average variance extracted* (AVE) of each construct is 0.50 or larger; (4) discriminant validity assessment—representing the extent to which the construct is empirically distinct from other constructs—Henseler et al. (2015) suggest that a heterotrait-monotrait (HTMT) value below 0.90 provides evidence for discriminant validity between a given pair of reflective constructs. To assess the formative outer models, I analyze: (1) the indicator weights' statistical significances—p values should be less than 0.05 and (2) indicator collinearity—*variance inflation factors* (VIFs) of 5 or above indicate potential collinearity problems.

Table 2 reports the outer loadings and weights and assessment statistics for the reflective and formative models.

Table 2 shows that every outer loading is statistically significant and with a value larger than 0.71; ρ_As are higher than 0.75. Corruption reveals some problems of indicator redundancy—because ρ_A is larger than 0.95—I consequently have excluded "Control Corr. Rev.*"* from model 2 to reduce indicator redundancy; convergent (AVE) and discriminant validity (HTMT) are satisfactory.[14] The formative latent constructs return satisfactory assessment statistics for both models.

Once the reliability and validity of the outer models have been positively evaluated, the second step in assessing a PLS–SEM consists of evaluating the inner (or structural) model. Table 3 reports the standardized path coefficients and standardized total effects for each latent construct on corruption.

Table 3 shows that the estimated path coefficients are qualitatively robust to the two specifications, with the only exception being the direct effect of "*Education*" on "*Corruption*", which is statistically significant only in Model 1. I find that the path coefficients (i.e., direct effects) carry the expected signs with some exceptions: lower *Decentralization*, higher *Fractionalization*, abundance of *Natural Resources* and *British colonial heritage* are not associated with more corruption.

In particular, on the one hand, countries with higher *Quality of Regulation, Quality of Democracy, Media Freedom, Natural Resources, Education* (only for model 1)*, Fractionalization,* higher population percentages of Protestants and countries with Belgian, Dutch or French colonial heritages are perceived as being less corrupt. On the other hand, higher levels of *Decentralization, Oil Rent,* higher population percentages of Catholics and countries with Italian colonial heritages are associated with higher levels of corruption. Lastly, looking at the consequences of corruption, my findings validate the common finding that more corrupt countries show lesser S*ocioeconomic development.*

Following Hair et al. (2019), in addition to (1) the statistical significances of standardized path coefficients in assessing the inner model, I check (2) collinearity among latent constructs—VIFs of more than five are indicative of probable collinearity issues; (3) the

---

[14] To save space, I omit reporting the matrices. In brief, the analysis reveals that the hypothesis of discriminant validity holds for estimated model 1 (2) because only 2 (1) HTMT values of 153 (28) estimated ratios exceed the threshold.

Table 3 Path coefficients (direct effects) and total effects on corruption—inner model

Cause → corruption		Model 1 (direct effect)	Model 2 (direct effect)	Model 1 (total effect)	Model 2 (total effect)
Media freedom	β_{21}	−0.188 (0.000)	−0.263 (0.000)	−0.528 (0.000)	−0.540 (0.000)
Education	β_{31}	−0.050 (0.000)	−[a]	−0.446 (0.000)	−0.394 (0.000)
Quality of democracy	β_{41}	−0.144 (0.000)	−0.063 (0.000)	−0.434 (0.000)	−0.353 (0.000)
Quality of regulation	β_{51}	−0.435 (0.000)	−0.435 (0.000)	−0.435 (0.000)	−0.435 (0.000)
Natural resources	β_{61}	−0.061 (0.000)	−0.072 (0.000)	−0.061 (0.000)	−0.072 (0.000)
Oil rent	β_{71}	0.072 (0.000)	0.084 (0.000)	0.037 (0.000)	0.043 (0.000)
Size of public sector	β_{81}	−0.240 (0.000)	−0.272 (0.000)	−0.240 (0.000)	−0.272 (0.000)
Decentralization	β_{91}	0.067 (0.000)	0.066 (0.000)	0.067 (0.000)	0.066 (0.000)
Fractionalization	$\beta_{10,1}$	−0.027 (0.005)	0.054 (0.000)	0.015 (0.125)	0.042 (0.000)
French colony	$\beta_{11,1}$	−0.033 (0.001)	–	−0.038 (0.000)	–
Portuguese colony	$\beta_{12,1}$	−0.004 (0.532)	–	0.007 (0.257)	–
Spanish colony	$\beta_{13,1}$	0.010 (0.378)	–	0.049 (0.000)	–
Italian colony	$\beta_{14,1}$	0.037 (0.000)	–	0.031 (0.000)	–
Belgian and Dutch colony	$\beta_{15,1}$	−0.034 (0.000)	–	−0.039 (0.000)	–
British, US, Australian colony	$\beta_{16,1}$	−0.013 (0.195)	–	−0.020 (0.052)	–
% catholic	$\beta_{17,1}$	0.072 (0.000)	–	0.072 (0.000)	–
% protestant	$\beta_{18,1}$	−0.108 (0.000)	–	−0.108 (0.000)	–
% muslim	$\beta_{19,1}$	−0.021 (0.084)	–	−0.021 (0.082)	–

Cause → cause (direct effect)		Model 1	Model 2
Education → media freedom	β_{32}	0.647 (0.000)	0.649 (0.000)
Education → quality of democracy	β_{34}	0.125 (0.000)	0.123 (0.000)
Education → socioeconomic development	$\beta_{3,20}$	0.738 (0.000)	–
Media freedom → quality of democracy	β_{24}	0.783 (0.000)	0.784 (0.000)
Quality of democracy → quality of regulation	β_{35}	0.667 (0.000)	0.667 (0.000)
Oil rent → natural resources	β_{76}	0.569 (0.000)	0.571 (0.000)
Natural resources → socioeconomic development	$\beta_{6,20}$	0.096 (0.000)	–

Table 3 (continued)

Cause → cause (direct effect)		Model 1	Model 2
Fractionalization → decentralization	$\beta_{10,9}$	− 0.176 (0.000)	− 0.176 (0.000)
French colony → %catholic	$\beta_{11,17}$	− 0.067 (0.000)	–
Portuguese colony → %catholic	$\beta_{12,17}$	0.160 (0.000)	–
Spanish colony → %catholic	$\beta_{13,17}$	0.549 (0.000)	–
Italian colony → %catholic	$\beta_{14,17}$	− 0.076 (0.000)	–
Belgian, Dutch colony → %protestant	$\beta_{15,17}$	0.050 (0.000)	–
British, US, Australian colony → %protestant	$\beta_{16,17}$	0.057 (0.000)	–
Corruption → consequences		Model 1	Model 2
Corruption → socioeconomic development	$\beta_{1,20}$	− 0.229 (0.000)	–

p values in parentheses

[a]Education is omitted in model 2 because *p* value > 0.05. Total effect = direct effect + indirect effect, where the latter indicates the (standardized) impact of a latent construct via other latent variables on the target construct (i.e., corruption)

coefficient of multiple determination (R^2)[15]; (4) *cross-validated redundancy*, also known as the Stone-Geisser Q^2—which assesses the inner model's predictive relevance[16]; and (5) the model's predictive power (PP)—by checking if the PLS–SEM analysis yields higher prediction errors in terms of Root Mean Square Error (RMSE) than the linear regression model (LM).[17] Table 4 reports assessment inner statistics and criteria for model selection among a finite set of models—i.e., the Bayesian Information Criteria (BIC) and Akaike's Information Criterion (AIC).[18]

As far as the criteria for inner model assessment are concerned, Table 4 shows probable collinearity issues only for "*Corruption*" in the model 1. The explained variance of the key variable in the analysis at hand (i.e., corruption) has a large R^2 (about 0.80). Looking at the Stone-Geisser Q^2, the most predictive relevance is associated with "*Corruption*", while "*Quality of Democracy*", "*Quality of Regulation*", "*Media Freedom*", "*Catholics*" and "*Socioeconomic Development*" all have "high" or "medium" predictive relevance. All of the latent variables, with the exception of "*Protestants*", reveal high predictive power (PP) in estimating the observed indicators. In conclusion, taking also into account the BIC and the AIC metrics, model 1 is considered to be the best specification for predicting latent scores, i.e., *S-CPI*.[19]

Following the current literature, I standardize the estimated latent scores of "perceived corruption" ($\hat{x}_{it}$) in order to derive an index ranging between 0 and 100. The standardization is based on the following formula:

$$S\text{-}CPI_{it} = 100 \frac{\hat{x}_{it} - \underset{\forall i, \forall t}{Min}\left(\hat{x}_{it}\right)}{\underset{\forall i, \forall t}{Max}\left(\hat{x}_{it}\right) - \underset{\forall i, \forall t}{Min}\left(\hat{x}_{it}\right)} \tag{2}$$

where for Model IFB,[20] the following values are obtained: $\underset{\forall i, \forall t}{Min}\left(\hat{x}_{it}\right) = -2.581$ and $\underset{\forall i, \forall t}{Max}\left(\hat{x}_{it}\right) = 1.693$. Focusing on the "extreme cases", I find that the four nations with the smallest indexes of perceived corruption are Denmark, Finland, New Zealand and Sweden. On the other side of the ranking, the most corrupt countries are Somalia, the Democratic Republic of Congo, Iraq and North Korea.[21] In terms of the time trends of *S-CPI*, Fig. 1 shows some representative countries: Italy and South Korea (i.e., countries representative of developed economies with relatively high levels of perceived corruption); Germany,

---

[15] R^2s larger than 0.75, 0.50 and 0.25 indicate large, medium and moderate amounts of explained variations in endogenous construct.

[16] A Q^2 exceeding 0.5 reveals the large predictive relevance of given latent variables, while when Q^2 is negative, no evidence of predictive relevance is found (Cohen 1988).

[17] The rule of thumb is that if PLS yields a larger RMSE than LM for all, the majority of, the minority of, or the same number, or none of the observed indicators, then PLS has no, low, medium, or high predictive power, respectively.

[18] The model with the smallest BICs and AICs is preferred.

[19] However, because of the robustness of results across similar specifications, S-CPI scores are not significantly affected by that choice.

[20] Although the different treatments of missing values don't markedly change the rankings of countries between the S–CPI indexes estimated by the original dataset (MV) and the datasets with replacement (I and IFB) and their correlations are quite high (99.7% and 98.9%), the S–CPI scores are biased because missing values are more prevalent during the first decade of time range (1995–2005) and during the last 2 years of the sample (2015–2016).

[21] The annual estimates (1995–2016 for 165 countries) of S-CPI are reported in Appendix A.2.

Table 4 Assessment of structural model

	Model 1				Model 2			
	VIF	R^2	Q^2	PP	VIF	R^2	Q^2	PP
Corruption	2(28)	**0.81**	**0.66**	**6(6)**	**0(8)**	**0.78**	**0.63**	**6(6)**
Decentralization	**0(1)**	0.03	0.03	**1(1)**	**0(1)**	0.03	0.03	**1(1)**
Media freedom	**0(1)**	0.42	**0.27**	**3(3)**	**0(1)**	0.42	**0.27**	**3(3)**
Natural resources	**0(1)**	0.27	0.15	**2(2)**	**0(1)**	0.27	0.15	**2(2)**
Quality of democracy	**0(2)**	**0.77**	**0.54**	**3(3)**	**0(2)**	**0.77**	**0.54**	**3(3)**
Quality of regulation	**0(1)**	0.46	**0.35**	**3(3)**	**0(1)**	0.46	**0.35**	**3(3)**
Catholics	**0(4)**	0.41	**0.41**	**1(1)**	–	–	–	–
Protestants	**0(2)**	0.01	0.07	0(1)	–	–	–	–
Socio-economic development	**0(3)**	**0.88**	**0.64**	**4(5)**	–	–	–	–
Model comparisons								
BIC	**− 5775**				− 5408			
AIC	**− 5893**				− 5464			

VIF columns report the number of VIF values larger than 5 for each latent construct in parenthesis the number of calculated VIF; Q^2 is the "construct cross-validated redundancy" score calculated by Smart PLS 3; PP (predictive power) reports the number of manifest indicators estimated by PLS–SEM with lower RMSE than LM, in parenthesis the total number of manifest variables included in the measurement model

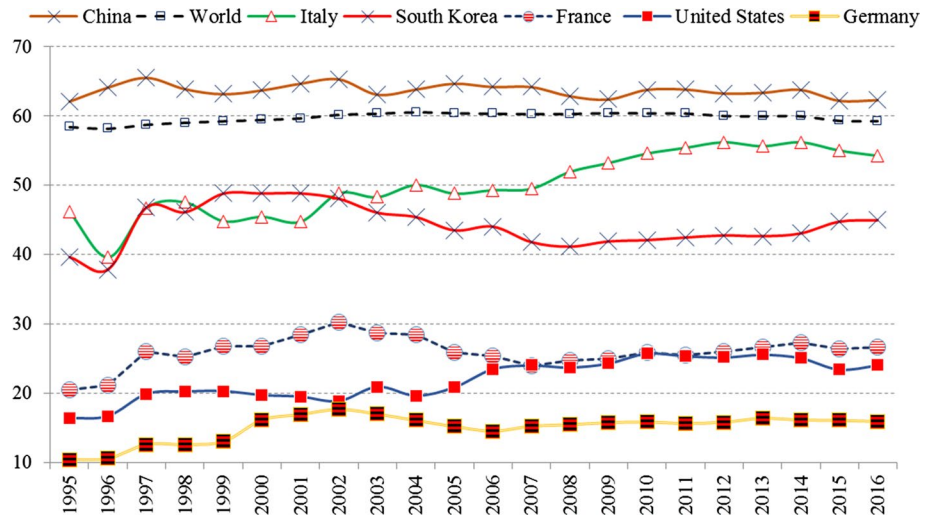Bold numbers indicate "medium" (or better) assessment statistics

Fig. 1 Some annual estimates of Standardized S-CPI

United States and France (i.e., developed countries with relatively low levels of perceived corruption) and China (i.e., a developing economy with a relatively high perception of corruption).

Table 5 Comparison of perceived corruption indexes

Standirdized ind. (0–100)	RMSE	MAE	S-CPI	CPI rev	Control corr. rev	Bayesian corr.	Freedom corr. rev	ICRG rev	Political corr.
S-CPI	–	–	**1.000**						
CPI rev	2.04	4.18	0.979	**1.000**					
Control corr. rev	4.24	3.24	0.984	0.974	**1.000**				
Bayesian corr.	7.60	5.79	0.947	0.938	0.928	**1.000**			
Freedom corr rev	8.22	5.76	0.937	0.955	0.921	0.871	**1.000**		
ICRG rev	12.31	9.95	0.926	0.892	0.909	0.8528	0.835	**1.000**	
Political corr	14.39	11.52	0.915	0.875	0.910	0.820	0.814	0.817	**1.000**

To conclude the descriptive analysis, Table 5 compares the standardized S-CPI and the most widely known existing indexes of perceived corruption.[22]

The root mean square error, mean absolute error and the correlation matrix reveal that the corruption perceptions index published by Transparency International and the control of corruption index published by Worldwide Governance Indicators are more similar to the S-CPI. However, taking into account that the S-CPI covers more than 30% (22%) of country-level scores over the 1995–2016 period than the corruption perceptions index or the control of corruption index and, moreover, that its scores are validated by statistical and economic theories, I conclude that the S-CPI can be considered to be a superior data source for empirical analysis.

4.1 Policy implications

In terms of policy implications, normative inferences as to which are the most effective channels for fighting corruption can be drawn by conducting an importance-performance map analysis (IPMA) (Ringle and Sarstedt 2016) and a partial least squares multi-group analysis (PLS–MGA) on model 2.

The IPMA extends the standard SEM results based on the total effects of the latent constructs on *Corruption* by taking the performance of each determinant into account. That approach makes it possible to identify the causes that have a relatively high "Importance" for *Perceived Corruption* (i.e., those latent variables that have larger total effects on the target construct), but also a relatively low "Performance" (i.e., low average latent variable scores). Graphically, the importance-performance map reports the (unstandardized) total effects on the x-axis to measure the "Importance" and, on the y-axis, the average rescaled latent variable scores to measure the "Performance".[23] For the interpretation of the results, Ringle and Sarstedt (2016) point out as the constructs in the lower right area of the IPMA are characterized by high importance for the target construct, but reveal low performance, they should be considered to be particularly relevant for policy action (i.e., there are placed the potential first-best policies for deterring corrupt practices). Figure 2 shows the IPMA map and the four quadrants that identify the priority order for policy actions.

According to the IPMA, the main policy implications can be summarized as follows:

(1) Reducing corruption is hard because no "first-best" policies that, affecting some causes of corruption—with relatively low performance (i.e., below the average of 30.1—horizontal line) and particularly high importance (i.e., total effect above the average of 0.9—vertical line)—reduce a country's perceived corruption markedly;

(2) Looking at the second priority for policy actions aiming to reduce corruption, IPMA suggests improving *Quality of Democracy*, reducing (ethnic, linguistic and religious) *Fractionalization* and fostering *Media Freedom;*

(3) More *Education* and *Decentralization,* on the one hand, have relatively low "importance" in curbing *Corruption*, but, on the other hand, both reveal relatively low per-

---

[22] The original indexes are standardized in order to range over the same scale of the S–CPI (i.e., 0–100).

[23] In order to fulfil the requirements for conducting the IPMA, I have taken the total effects in absolute values, such that higher values represent positive effects for the meaning of the key latent construct. In Fig. 2 the original signs of the total effect on perceived corruption are reported in parentheses.
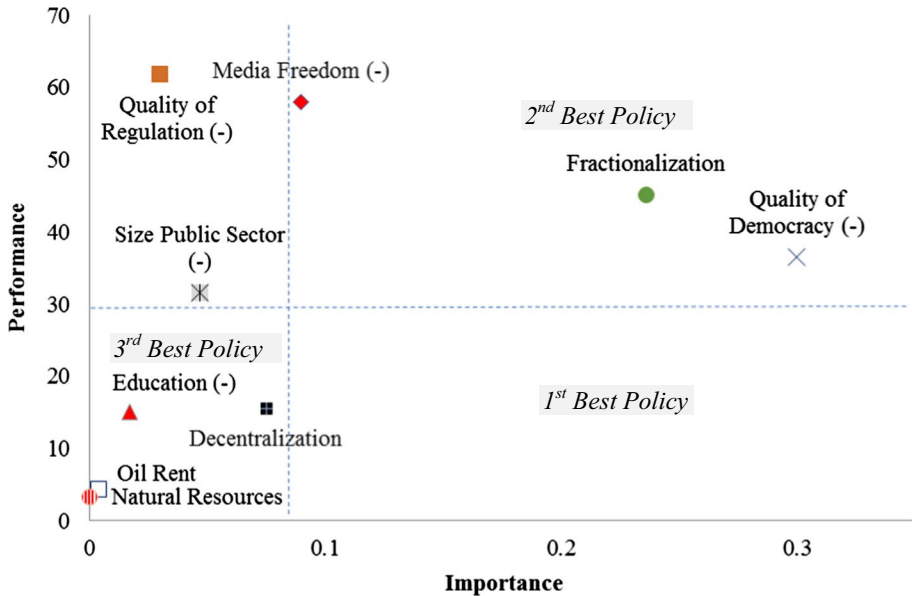
Fig. 2 Importance-performance map of perceived corruption

formance. *Education* and *Decentralization* therefore are potentially relevant for policy actions, but have less significant marginal effects;

(4) *Quality of Regulation* and *Size of Public Sector* have lower importance and larger performance on *Corruption* than the average; hence, a decision maker should prioritize the above-mentioned alternative policies to reduce *Corruption*;

(5) *Oil Rents* and *Natural Resources* have negligible (unstandardized) total effects on *Corruption*.

In order to explore the overall validity of the just listed policy implications, I conduct a multi-group analysis (MGA) by clustering the global sample in subgroups based on geographical areas and estimating inner and outer coefficients for each subgroup separately. Table 6 reports the standardized total effects of each potential determinant of corruption.

The findings support the hypothesis that the order of priority for policy actions (see the "Rank" values in Table 5) change according to the geographical areas considered. Focusing attention on the main results, *Media Freedom* has the largest effect in reducing *Corruption* all around the world, with the exception of North Africa, the Middle East, Latina America and the Caribbean, where the priority is improving the *Quality of Democracy*. More *Education* contributes to reducing *Corruption*, with the exception of Western Europe and North America for which the two phenomena are statistically uncorrelated. Similarly, the *Size of public Sector* is negatively correlated with *Corruption,* but this correlation does not materialize in South, East and South-East Asia. As far as the sign of the effect of *Natural Resources* on *Corruption* concerns, estimates for that group-specific analysis corroborates the indeterminateness of the empirical literature. Indeed, the direction of effect depends on the specific geographical group studied. Abundance Natural resource abundance has a negative effect on corruption in North Africa and the Middle East; it carries a positive sign in Europe, post-Soviet Union, Sub-Saharan Africa and North America; while it does not

Table 6 Standardized total effects of corruption by geographical area—model 2

	World [rank]	EE, PSU [rank]	LA, C [rank]	NA, ME [rank]	SSA [rank]	WE, NA [rank]	ASIA [rank]
Media freedom	-0.54*** [1]	-0.684*** [1]	-0.502*** [2]	-0.221*** [5]	-0.444*** [1]	-0.544*** [1]	-0.457*** [1]
Q. of regulation	-0.435*** [2]	-0.314*** [4]	-0.483*** [3]	-0.487*** [2]	-0.368*** [3]	-0.463*** [2]	-0.454*** [2]
Education	-0.394*** [3]	-0.382*** [3]	-0.177*** [6]	-0.162*** [8]	-0.322*** [4]	-0.086 [8]	-0.329*** [3]
Q. of democracy	-0.353*** [4]	-0.430*** [2]	-0.558*** [1]	-0.547*** [1]	-0.412*** [2]	-0.337*** [3]	-0.282*** [5]
Size public sector	-0.272*** [5]	-0.268*** [5]	-0.354*** [4]	-0.382*** [3]	-0.285*** [5]	-0.238*** [4]	0.042 [8]
Natural resources	-0.072*** [6]	0.065*** [7]	-0.049 [8]	-0.169*** [6]	0.072*** [8]	0.118** [6]	-0.012 [8]
Decentralization	0.066*** [7]	0.039* [8]	0.143*** [7]	0.334*** [4]	-0.035* [9]	0.071*** [7]	0.119*** [7]
Oil rent	0.043*** [8]	0.002 [8]	0.008 [8]	0.160*** [9]	0.219*** [6]	-0.03 [8]	0.127*** [6]
Fractionaliz.	0.042*** [9]	-0.098*** [6]	0.203*** [5]	-0.166*** [7]	0.212*** [7]	-0.14*** [5]	-0.318*** [4]
# countries	**165**	**26**	**26**	**20**	**45**	**23**	**25**

Based on PLS-MGA, the cells where the estimates of total effect based on the geographical group are statistically different (*p* value < 0.05) from the total effect estimated on the full dataset (World) are highlighted in italics

EE,PSU Eastern Europe and post-Soviet Union, *LA,C* Latina America and the Caribbean, *NA,ME* North Africa and the Middle East, *SSA* Sub-Saharan Africa, *WE,NA* Western Europe and North America, *ASIA* South Asia, East Asia, South-East Asia, The Pacific

*p value < 0.10

**p value < 0.05

***p < 0.01

have a statistically significant effect in Latin America, the Caribbean, South, East, South-East Asia or the Pacific. *Oil Rent* has the largest positive effect on *Corruption* in North Africa, the Middle East and Sub-Saharan Africa, while that determinant of corruption does not have a statistically significant effect in Europe, post-Soviet Union, Latin America, Caribbean or North America.

In the second step of MGA—which it is not reported here for the sake of brevity—the statistical tests of differences in group-specific coefficients reveal that these differences often are statistically significant across geographical areas. The normative implication is that the efficacies of policy actions significantly differ from country to country. Given that the effectiveness of policies to reduce a specific type of corruption significantly varies from actions against other types of corruption, a policy maker should select a strategy based on empirical analysis and best practices of countries with similar institutional and economic development because the magnitudes of different types of corruption (e.g., grand, petty, legal, illegal) depend on socioeconomic and institutional development as well.

5 Conclusions

This research examines the causes and consequences of corruption by adopting partial least squares—structural equation modeling (PLS–SEM). Approaching corruption as a latent construct, I estimate an index of perceived corruption in 165 countries from 1995 to 2016.

From a methodological perspective, the analysis of empirical relationships between constructs that are not directly observed (e.g., corruption), intrinsically multidimensional (e.g., institutional quality, economic development), or both, predicting an overall index of perceived corruption makes the PLS–SEM approach worthwhile for the relevant strand of literature. The methodology presented herein allows researchers to estimate the determinants of corruption in a unified framework that relies on the existing theory and empirical evidence on corruption. It is made possible by the opportunity that SEM supplies to specify simultaneously both the determinants that affect corruption directly, indirectly, or both as well as the effects of corruption on a country's socioeconomic performance—the "structural" or "inner" model of the SEM. On the other hand, SEM allows one to exploit currently available indexes of perceived corruption as complementary observable measurements of corruption—the "measurement" or "outer" model. However, the proposed statistical approach shares two of the problems most relevant in the empirical literature. First is the problem of the divergence between "perceived" and "actual" corruption". Second, the PLS–SEM provides unsatisfactory solutions to the problem of endogeneity. Specifically, it is likely that some variables, identified in the model as "causes" of corruption, also are influenced by the perceived magnitude of corruption, which depends, e.g., on institutional quality. Therefore, I suggest caution in assessing the relationships between institutional explanatory variables and perceived corruption as one-way causal links instead of bi-directional interactions that generate feedback loops.

On the positive side, the estimated S-CPI has two main advantages over existing indexes of perceived corruption. First, it provides estimates of perceived corruption by exploiting not only the existing measures, but also combines elements of the extant theoretical and empirical literature on the causes and consequences of corruption within a unified framework. Second, it reduces measurement errors in two ways: (a) by using several indicators for each "unobservable" variable (e.g., corruption; quality of institutions; socioeconomic development). Accordingly, the proposed index can be thought of as a "meta-index". (b)

By following the conventional statistical remedy to enlarge the sample size in order to reduce measurement errors. Specifically, I consider about 160,000 observations (coming from 47 manifest variables concerning 165 countries over a period of 22 years). Those two correlated strategies make my findings robust to alternative model specifications and to strategies for replacing missing values.

On the normative side, I derive some policy implications from PLS–SEM findings by analyzing the estimated direct and indirect (i.e., mediated by other potential causes) effects and by conducting the importance-performance map analysis proposed by Ringle and Sarstedt (2016).

In general terms, I find that focusing only on direct effects may be misleading. For instance, *Quality of Regulation* has the largest direct effect in reducing *Corruption* (−0.44), followed by *Media Freedom* (−0.54) and *Quality of Democracy* (−0.06), but, once the indirect effects are taken into account, the ranking of the most important causes of corruption change as follows: *Media Freedom* (−0.54); *Quality of Regulation* (−0.44); *Education* (−0.39); *Quality of Democracy* (−0.35). Furthermore, by conducting an IPMA to identify the most effective policy actions for fighting corruption, I find that a decision maker should primarily be concerned with (in descending order): *Quality of Democracy*, *Fractionalization*, *Media Freedom*, *Decentralization*, *Education*, *Size of Public Sector and Quality of Regulation*. For other determinants (e.g., *Natural resources*, *Oil Rents*), that are often considered as important causes of corruption in the existing empirical literature, my results do not validate those conclusions.

The last step of the empirical analysis consists in implementing a multi-group analysis by clustering the global sample of countries in subgroups based on geographical areas and estimating the total effect on each subgroup separately. According to that analysis, the estimated effects of the causes on corruption vary significantly across geographical areas; consequently, policy actions also should differ from country to country. The rationale is that different types of corruption (i.e., "grand" and "petty", "legal" and "illegal") exist and their relative importance depends on economic development, the quality of institutions, cultural background, and so on. Accordingly, the best policies for fighting corruption consist in taking action first on the most important causes (i.e., with the largest estimated total effects) and with the most room for improvement. Each of the dimensions of policy action should be estimated on sub-groups of homogenous (in terms of institutional quality and economic development) countries.

Compliance with ethical standards

References

Ades, A., & Di Tella, R. (1997). The new economics of corruption: A survey and some new results. *Political Studies, 45,* 496–515.

Ades, A., & Di Tella, R. (1999). Rents, competition, and corruption. *American Economic Review, 89,* 982–994.

Aidt, T. S. (2003). Economic analysis of corruption: A survey. *Economic Journal, 113,* F632–F652.

Aidt, T. S. (2009). Corruption, institutions, and economic development. *Oxford Review of Economic Policy, 25*(2), 271–291.

Andersson, S. (2017). Beyond unidimensional measurement of corruption. *Public Integrity, 19*(1), 58–76.

Arikan, G. (2004). Fiscal decentralization: A remedy for corruption? *International Tax and Public Finance, 11*(2), 175–195.

Arvate, P. R., Curi, A. Z., Rocha, F., & Miessi Sanches, F. A. (2010). Corruption and the size of government: Causality tests for OECD and Latin American countries. *Applied Economics Letters, 17*(10), 1013–1017.

Bhattacharyya, S., & Hodler, R. (2010). Natural resources, democracy and corruption. *European Economic Review, 54*(4), 608–621.

Brunetti, A., & Weder, B. (2003). A free press is bad for corruption. *Journal of Public Economics, 87*(7–8), 1801–1824.

Buehn, A., Dell'Anno, R., & Schneider, F. (2018). Exploring the dark side of tax policy: An analysis of the interactions between fiscal illusion and the shadow economy. *Empirical Economics, 54*(4), 1609–1630.

Campos, N. F., & Giovannoni, F. (2007). Lobbying, corruption, and political influence. *Public Choice, 131*(1–2), 1–21.

Chin, W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern Methods for Business Research* (pp. 295–336). Mahwah, NJ: Lawrence Erlbaum Associates, Publisher.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum.

Coppedge, M., Gerring, J., Lindberg, S. I., Skaaning, S., Teorell, J., Altman, D., Bernhard, M., Fish, M. S., Glynn, A., Hicken, A., Knutsen, C. H., Krusell, J., Lührmann, A., Marquardt, K. L., McMann, K., Mechkova, V., Olin, M., Paxton, P., Pemstein, D., Pernes, J., Petrarca, C. S., von Römer, J., Saxer, L., Seim, B., Sigman, R., Staton, J., Stepanova, N., & Wilson S. (2017). V-Dem [Country-Year/Country-Date] Dataset v7.1. Varieties of Democracy (V-Dem) Project.

Damania, R., Fredriksson, P. G., & Mani, M. (2004). The persistence of corruption and regulatory compliance failures: Theory and evidence. *Public Choice, 121,* 363–390.

Dell'Anno, R. (2007). Shadow economy in Portugal: An analysis with the MIMIC approach. *Journal of Applied Economics, 10*(2), 253–277.

Dell'Anno, R., & Dollery, B. (2014). Comparative fiscal illusion: A fiscal illusion index for the European Union. *Empirical Economics, 46,* 937–960.

Dell'Anno, R., & Teobaldelli, D. (2015). Keeping both corruption and the shadow economy in check: The role of decentralization. *International Tax and Public Finance, 22*(1), 1–40.

Dijkstra, T. K., & Henseler, J. (2015). Consistent partial least squares path modeling. *MIS Quarterly, 39*(2), 297–316.

Dimant, E., & Tosato, G. (2018). Causes and effects of corruption: What has past decade's empirical research taught us? *A survey. Journal of Economic Surveys, 32*(2), 335–356.

Dincer, O. C. (2008). Ethnic and religious diversity and corruption. *Economics Letters, 99*(1), 98–102.

Dincer, O., & Johnston, M. (2019). The search for the most CORRUPT State in the US. In L. Eileen (Ed.), *Political corruption* (pp. 39–48). New York, US: Greenhaven Publishing.

Donchev, D., & Ujhelyi, G. (2014). What do corruption indices measure? *Economics and Politics, 26*(2), 309–331.

Dreher, A., Kotsogiannis, C., & McCorriston, S. (2007). Corruption around the world: Evidence from a structural model. *Journal of Comparative Economics, 35*(3), 443–466.

Duncan, O. D. (1975). *Introduction to structural equation models*. New York: Academic Press.

Enste, D., & Heldman, C. (2017). Causes and consequences of corruption: An overview of empirical results. IW-Reports 2/2017, Institut der deutschen Wirtschaft Köln (IW)/Cologne Institute for Economic Research.

Esposito Vinzi, V., Chin, W. W., Henseler, J., & Wang, H. (2010a). *Handbook of partial least squares: Concepts, methods and applications. Springer handbooks of computational statistics*. Berlin: Springer.

Esposito Vinzi, V., Trinchera, L., & Amato, S. (2010b). PLS path modeling: From foundations to recent developments and open issues for model assessment and improvement. In V. Esposito Vinzi, et al.

(Eds.), *Handbook of partial least squares: Concepts, methods and applications springer handbooks of computational statistics* (Vol. 2, pp. 47–82). Berlin: Springer.

Faizan, A., Rasoolimanesh, S. M., Sarstedt, M., Ringle, C. M., & Ryu, K. (2018). An assessment of the use of partial least squares structural equation modeling (PLS–SEM) in hospitality research. *International Journal of Contemporary Hospitality Management, 30*(1), 514–538.

Fisman, R., & Gatti, R. (2002). Decentralization and corruption: Evidence across countries. *Journal of Public Economics, 83,* 325–345.

Goel, R. K., & Budak, J. (2006). Corruption in transition economies: Effects of government size, country size and economic reforms. *Journal of Economics and Finance, 30*(2), 240–250.

Goel, R. K., & Nelson, M. A. (1998). Corruption and government size: A disaggregated analysis. *Public Choice, 97*(1–2), 107–120.

Goel, R. K., & Nelson, M. A. (2010). Causes of corruption: History, geography and government. *Journal of Policy Modeling, 32*(4), 433–447.

Gokcekus, O., & Sonan, S. (2017). Political contributions andcorruption in the United States. *Journal of Economic Policy Reform, 20*(4), 360–372.

Goldberg, F. (2018). Corruption and lobbying: Conceptual differentiation and gray areas. *Crime, Law and Social Change, 70*(2), 197–215.

Hair, J. F., Hollingsworth, C. L., Randolph, A. B., & Chong, A. Y. L. (2017). An updated and expanded assessment of PLS–SEM in information systems research. *Industrial Management & Data Systems, 117,* 442–458.

Hair, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling (PLS–SEM)* (2nd ed.). Thousand Oaks, CA: Sage.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS–SEM. *European Business Review, 31*(1), 2–24.

Hair, J. F., Sarstedt, M., Ringle, C. M., & Mena, J. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the Academy of Marketing Science, 40*(3), 414–433.

Harstad, B., & Svensson, J. (2011). Bribes, lobbying, and development. *American Political Science Review, 105*(1), 46–63.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135.

Heritage Foundation. (2017). *Index of economic freedom. The Heritage Foundation, Washington*. Data retrieved February 5, 2018. http://www.qog.pol.gu.se. https://doi.org/10.18157/QoGStdJan18.

Jain, A. K. (2001). Corruption: A review. *Journal of Economic Surveys, 15,* 71–121.

Kalenborn, C., & Lessmann, C. (2013). The impact of democracy and press freedom on corruption: Conditionality matters. *Journal of Policy Modeling, 35*(6), 857–886.

Kaufmann, D., & Vicente, P. C. (2011). Legal corruption. *Economics and Politics, 23*(2), 195–219.

Kunicova, J., & Rose-Ackerman, S. (2005). Electoral rules and constitutional structures as constraints on corruption. *British Journal of Political Science, 35*(4), 573–606.

La Porta, R. L., de Silanes, F. L., Shleifer, A., & Vishny, R. (1999). The quality of government. *Journal of Law Economics and Organization, 15*(1), 222–279.

La Porta, R. L., Lopez-De-Silanes, F., Shleifer, A., & Vishny, R. W. (1997). Trust in large organisations. *The American Economic Review, Papers and Proceedings, 87*(2), 333–338.

Lambsdorff, J. G. (2002). Corruption and rent-seeking. *Public Choice, 113,* 97–125.

Lambsdorff, J. G. (2006). Causes and consequences of corruption: What do we know from a cross-section of countries? In S. Rose-Ackerman (Ed.), *International handbook on the economics of corruption* (pp. 3–51). Cheltenham, UK: Edward Elgar.

Lambsdorff, J. G. (2007). *The institutional economics of corruption and reform: Theory, evidence, and policy*. Cambridge: Cambridge University Press.

Lessmann, C., & Markwardt, G. (2010). One size fits all? Decentralization, corruption, and the monitoring of bureaucrats. *World Development, 38*(4), 631–646.

Mauro, P. (1995). Corruption and growth. *Quarterly Journal of Economics, 110*(3), 681–712.

Melgar, N., Rossi, M., & Smith, T. W. (2010). The Perception of corruption. *International Journal of Public Opinion Research, 22*(1), 120–131.

Ning, H. (2016). Rethinking the causes of corruption: Perceived corruption, measurement bias, and cultural illusion. *Chinese Political Science Review, 1*(2), 268–302.

North, C. M., Orman, W. H., & Gwin, C. R. (2013). Religion, corruption, and the rule of law. *Journal of Money, Credit and Banking, 45*(5), 757–779.

Paldam, M. (2001). Corruption and religion. *Adding to the Economic Model. Kyklos, 54*(2/3), 383–414.

Pellegata, A. (2012). Constraining political corruption: An empirical analysis of the impact of democracy. *Democratization, 20*(7), 1195–1218.

Pellegrini, L., & Gerlagh, R. (2008). Causes of corruption: A survey of cross-country analyses and extended results. *Economics of Governance, 9*(2), 245–263.

Pemstein, D., Marquardt, K. L., Tzelgov, E., Wang, Y., Krusell, J., & Miri, F. (2017). *The V-Dem measurement model: Latent variable analysis for cross-national and cross-temporal expert-coded data* (2nd ed.). University of Gothenburg, Varieties of Democracy Institute: Working Paper No. 21.

Philp, M. (2015). The definition of political corruption. In P. M. Heywood (Ed.), *Routledge handbook of political corruption* (pp. 17–29). Abingdon, UK: Routledge.

PRS Group. (2018). *International country risk guide*. Political Risk Services.

Razafindrakoto, M., & Roubaud, F. (2010). Are international databases on corruption reliable? A comparison of expert opinion surveys and household surveys in Sub-Saharan Africa. *World Development, 38*(8), 1057–1069.

Richter, N. F., Cepeda Carrión, G., Roldán, J. L., & Ringle, C. M. (2016). European management research using partial least squares structural equation modeling (PLS–SEM): Editorial. *European Management Journal, 34,* 589–597.

Ringle, C. M., & Sarstedt, M. (2016). Gain more insight from your PLS–SEM results: The importance-performance map analysis. *Industrial Management & Data Systems, 116*(9), 1865–1886.

Ringle, C. M., Wende, S., & Becker, J.-M. (2015). *SmartPLS 3*. Boenningstedt: SmartPLS GmbH.

Rönkkö, M., McIntosh, C. N., Antonakis, J., & Edwards, J. R. (2016). Partial least squares path modeling: Time for some serious second thoughts. *Journal of Operations Management, 47,* 9–27.

Rose-Ackerman, S. (1999). *Corruption and government: Causes*. Consequences and Reform: Cambridge University Press.

Ruge, M. (2010). Determinants and size of the shadow economy—a structural equation model. *International Economic Journal, 24*(4), 511–523.

Sachs, J. D., & Warner, A. M. (1997). Sources of slow growth in African economies. *Journal of African Economies, 6*(3), 335–376.

Sandholtz, W., & Koetzle, W. (2000). Accounting for corruption: Economic structure, democracy, and trade. *International Studies Quarterly, 44*(1), 31–50.

Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., & Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: Where the bias lies! *Journal of Business Research, 69*(10), 3998–4010.

Seligson, M. A. (2006). The measurement and impact of corruption victimization: Survey evidence from Latin America. *World Development, 34*(2), 381–404.

Serra, D. (2006). Empirical determinants of corruption: A sensitivity analysis. *Public Choice, 126,* 225–256.

Shepsle, K. A. (2017). *Rule breaking and political imagination*. Chicago, United States: The University of Chicago Press.

Shleifer, A., & Vishny, R. (1993). Corruption. *The Quarterly Journal of Economics, 108*(3), 599–617.

Shmueli, G., Ray, S., Velasquez Estrada, J. M., & Chatla, S. B. (2016). The elephant in the room: Evaluating the predictive performance of PLS models. *Journal of Business Research, 69*(10), 4552–4564.

Standaert, S. (2015). Divining the level of corruption: A bayesian state-space approach. *Journal of Comparative Economics, 43*(3), 782–803.

Svensson, J. (2005). Eight questions about Corruption. *Journal of Economic Perspectives, 19*(3), 19–42.

Swamy, A., Knack, S., Lee, Y., & Azfar, O. (2001). Gender and corruption. *Journal of Development Economics, 64,* 25–55.

Tanzi, V. (1998). Corruption around the world: Causes, consequences, scope and cures. *IMF Staff Papers, 45,* 559–594.

Tanzi, V., & Davoodi, H. R. (2001). *Corruption, growth, and public finances*. IMF Working Paper 182.

Tenenhaus, M., & Esposito Vinzi, V. (2005). PLS regression, PLS path modeling and generalized procrustean analysis: A combined approach for PLS regression, PLS path modeling and generalized multi-block analysis. *Journal of Chemometrics, 19,* 145–153.

Teorell, J., Dahlberg, S., Holmberg, S., Rothstein, B., Pachon, N. A., & Svensson, R. (2018). The quality of government standard dataset, version Jan18. University of Gothenburg: The Quality of Government Institute. Retrieved February 5, 2018. http://www.qog.pol.gu.se (https://doi.org/10.18157/qogstdjan18).

Transparency International. (2017). *Corruption perceptions index. Transparency International*. Data retrieved February 5, 2018. http://www.qog.pol.gu.se. https://doi.org/10.18157/QoGStdJan18.

Treisman, D. (2000). The causes of corruption: A cross-national study. *Journal of Public Economics, 76,* 399–457.

Treisman, D. (2007). What have we learned about the causes of corruption from 10 years of crossnational empirical research? *Annual Review of Political Science, 10,* 211–244.

Treisman, D. (2015). What does cross national empirical research reveal about the cause of corruption. In P. M. Heywood (Ed.), *Routledge handbook of political corruption* (pp. 95–109). New York: Routledge.

World Bank. (1997). *Helping countries combat corruption: The role of the World Bank*. Washington, DC: World Bank Group.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.