An objective corruption risk index using public procurement data

Abstract

In order to address the lack of reliable indicators of corruption, this article develops a composite indicator of high-level institutionalised corruption through a novel 'Big Data' approach. Using publicly available electronic public procurement records in Hungary, we identify "red flags" in the public procurement process and link them to restricted competition and recurrent contract award to the same company. We use this method to create a corruption indicator at contract level that can be aggregated to the level of individual organizations, sectors, regions and countries. Because electronic public procurement data is available in virtually all developed countries from about the mid-2000s, this method can generate a corruption index based on objective data that is consistent over time and across countries. We demonstrate the validity of the corruption risk index by showing that firms with higher corruption risk score had relatively higher profitability, higher ratio of contract value to initial estimated price, greater likelihood of politicians managing or owning them, and greater likelihood of registration in tax havens, than firms with lower scores on the index. In the conclusion we discuss the uses of this data for academic research, investigative journalists, civil society groups, and small and medium business.

1. INTRODUCTION

Various corruption indices have received considerable academic, policy, and media attention, at least partially due to the central role the underlying phenomena play in the quality of democratic governance, the provision of public goods, economic growth, and equality. Some international organisations regularly monitor corruption in their member countries (European Commission, 2011) while many donor agencies tie funding to performance on governance indicators including corruption (Andersson and Heywood, 2009; Radelet, 2002, 2003).

In the absence of robust 'objective' measures, there are three major sources of corruption indicators to date: 1) surveys of corruption perceptions and attitudes (which are most widely used); 2) reviews of institutional and legal frameworks; and 3) detailed analyses and audits of individual cases. Each of these has deficiencies leaving us without any reasonably reliable and valid indicator of corruption suitable for comparing countries over time or exploring within country diversity. If these indices are not valid and reliable, and are instead contaminated by past economic performance (so that "successful" countries are rated as less corrupt) then their use to inform aid disbursement is not only problematic but risks producing the perverse results of denying funding to countries that need it most (Kurtz & Schrank, 2007b).

In order to fill some of the gap between the demand for reliable and valid corruption indices and the state of the data currently available, the goal of this paper is to develop a novel measure of institutionalised grand corruption which: 1) rests on a thorough understanding of the corrupt rent extraction process; 2).solely derives from objective data describing behaviour; 3) allows for consistent temporal comparisons

within and across countries, and 4) can be replicated for many countries using pre-existing data. Such high standards of indicator building are met using a novel 'Big Data' approach.

We develop a measure of grand corruption in public procurement.[1] In public procurement institutionalised grand corruption refers to the allocation and performance of public procurement contracts by bending prior explicit rules and principles of good public procurement in order to benefit a closed network while denying access to all others (Mungiu-Pippidi, 2006; North, Wallis, & Weingast, 2009; Rothstein & Teorell, 2008).

The paper is structured as follows: first, the literature on corruption measurement is reviewed; second, the proposed novel measurement approach is presented; third, Hungarian data and variables are summarized; fourth, the composite corruption risk index (CRI) is constructed; fifth, we discuss validity. Finally, we conclude with a discussion of possible applications.

2. LITERATURE ON MEASURING GRAND CORRUPTION

Available indicators of corruption are either flawed or too narrow for testing theories of grand corruption and developing effective solutions to it. By and large, corruption indicators derive from: (1) Surveys of attitudes, perceptions and experiences of corruption among different stakeholders (e.g. general population, firms, experts); (2)

---

[1] Other types of grand corruption can be found in privatization, altering regulation, large scale smuggling operations, etc. Public procurement is arguably one of the most important types as for example it amounts to roughly one third of public spending in OECD countries.

Reviews of institutional features controlling corruption in countries or individual organisations; and (3) Audits and investigations of individual cases.

Among perception and attitude surveys, the two most widely used are the World Bank's Control of Corruption (Kaufmann, Mastruzzi, and Kraay, 2010) and Transparency International's Corruption Perceptions Index (Transparency International, 2012a). Both of these have received extensive criticism applicable to any similar survey (Andersson & Heywood, 2009; Lambsdorff, 2006). Critics point out that perceptions may or may not be related to actual experience (Rose and Peiffer, 2012). They can be driven by general sentiment reflecting, for example, prior economic growth (Kurtz & Schrank, 2007a, 2007b) or media coverage of high profile corruption cases (Golden and Picci, 2005). Arguably, perceptions of grand corruption are even more unreliable than perceptions of everyday corruption since experts and citizens have almost no direct experience of this type of corruption. Furthermore, these indicators are produced from non-representative surveys, therefore representativeness bias is likely to occur (i.e. capturing the views of a particular group rather than the whole population), in addition to reflexivity bias (i.e. respondents influenced by prior and future measurements) exaggerated by small sample sizes (Golden and Picci, 2005). In addition, these indicators vary surprisingly little over time even when large changes in underlying governance structures occure suggesting that they are too insensitive to change (Arndt & Oman, 2006; Mungiu-Pippidi, 2011). Surveys of experiences with low-level bribery, such as the Quality of Government Institute's regional survey (Charron, Dijkstra, and Lapuente, 2010), address some of the weaknesses of perception surveys, but fall short of a sufficient data source also. One major problem is non-response or false response to sensitive questions such as giving or receiving bribes. Most importantly, only a tiny fraction of

the population has direct experience with grand corruption limiting the use of this method.

Reviews of institutions controlling corruption (e.g. OECD, 2009; Transparency International, 2012b), while crucial in understanding the determinants of corruption, are, by design, not measuring corruption directly. In the absence of a precisely measured outcome variable, they have to rely on untested theories on which institutional features work.

Scientific analyses and audits of individual cases are highly reliable in establishing both petty and grand corruption, however, their narrow scope make them of only limited use for comparative purposes. In addition, data from courts and law enforcement agencies can be used to create indices of corruption only to a limited degree because courts and law enforcement agencies have little capacity to investigate large number of cases and there is a high risk of capture in corrupt countries.

2.1 Objective measures of corruption

Some authors, recognising the deficiencies of the above indicators, developed objective measures which rely on directly observable indicators of behaviour that likely indicate corrupt behaviour (for an overview see Annex A). These studies investigate corruption in various contexts such as elections and high level politics or welfare services and redistributive programs. For example, Olken (2007) uses independent engineers to review road projects and calculates the amount and value of missing inputs to determine corruption. These indicators are very narrow, and would be very expensive to generate over time and across space which represents their key weakness compared to surveys of corruption perceptions. In an ideal

scenario scalable objective indicators and surveys of corruption experiences are combined to provide a fuller picture of the hidden and diverse phenomena.

More closely associated with our approach are those studies which focus on corruption in public procurement and bidding markets. For example, Golden and Picci (2005) propose a new measure of corruption based on the difference between the quantity of infrastructure and public spending on it among 20 regions in Italy.[2] Our index is inspired by other authors that use red-flags in public procurement records as proxy measures for corruption. These include the use of exceptional procedure types (Auriol, Flochel, and Straub, 2011) or clear scoring rules (Hyytinen, Lundberg, and Toivanen, 2008) or political connections of winning companies (Goldman, Rocholl, and So, 2013).

While these papers inspired our approach and point in the right direction, they cannot be scaled up to allow for temporal comparisons across countries and organisations. The reason is that they rely on a single indicator which may or may not be the primary vehicle for corrupt rent extraction depending on the regulatory framework in place (Olken and Pande, 2012). For example, corruption linked to exceptional procedure types may be easily removed by simply deleting the procedure from the procurement law. However it is unlikely that this alone would change the underlying corrupt phenomena much, as this technique can be replaced by another method of restricting competition (Auriol et al., 2011). Instead of relying on a single indicator, these and further elementary indicators have to be combined for meaningful temporal and international comparisons.

---

[2] It would be extremely difficult to use this method in a large cross-national framework, because it requires detailed knowledge of the cost of constructing public works across regions within a country (Golden and Picci 2005: 43), It is also restricted to construction, so could not be used to measure corruption in other sectors. Moreover, it is not appropriate for time-series data, as it "is not a measure of the flow of corrupt transactions." (Ibid: 43).

3. THE MEASUREMENT APPROACH

3.1 Corrupt rent extraction in public procurement

In order to indirectly measure institutionalised grand corruption, its underlying logic must be contrasted with a competitive market logic. Institutionalised grand corruption's primary aim is earning corruption rents. Corruption rents in public procurement can be earned when the winning contractor is a pre-selected company which earns extra profit due to higher than market price for the delivered quantity and/or quality.

The winning company has to be pre-selected in order to control rent extraction in an institutionalised manner. This rules out occasional dishonesty where the company is lured into corruption during the public procurement process. Extra profit has to be realised in order to create the pot of money from which rents can be paid.

In order to measure extra profit; price, delivered quantity, and quality of deliveries has to be known with high precision. However, none of these three can adequately be measured. Price and quantity are publicly available, but they are comparable only for homogenous products such as electricity without laborious case-by-case analysis and even then it is difficult to arrive at accurate estimates. Quality cannot be reliably observed in official records without using expensive expert knowledge. Hence, we can only analyse the process of awarding contracts to pre-selected companies.

Competition has to be eliminated or tilted in order to award the contract to the pre-selected company. Bypassing competition can be done in three primary forms, each corresponding to a phase of the public procurement process: 1) limiting the set of

bidders (submission phase); 2) unfairly assessing bidders (assessment phase); and 3) ex-post modifying conditions of performance[3] (delivery phase).

On the one hand, these three elementary corruption strategies can be combined in any way to reach the final desired outcome. For example, some bidders may be excluded with a tightly tailored eligibility criteria while the remaining unwanted bidders can simply be unfairly scored on subjective scoring items. On the other hand, once the desired outcome has been achieved at a given stage, there is no need for further corrupt actions that would increase the risk of detection with no additional benefit. For example, if the only company submitting a valid bid is the pre-selected company there is no need to modify contract content later to increase price.

3.2 Measurement model

Utilizing a public procurement database it is possible to measure a host of elementary indicators in relation to each of the above three stages of public procurement from which a composite indicator can be built. In order to most adequately model the company selection process, measurement is carried out on the level of individual contract awards. Aggregation to organisation level can also be carried out to link procurement data to company or bureaucratic characteristics.

Likely outcomes of corrupt procurement procedures are defined for each of the above three main phases (see section 5.1). Indicators of likely corruption techniques to achieve these outcomes in each phase are also defined, which constitute the inputs for corrupt contract award and completion (see Authors, 2013a).

---

[3] While modifying contract conditions does not belong to the set of company selection techniques, it can be part of an arsenal supporting the selection of the 'right' company. For example, the pre-selected company wins in a competitive process by promising low price and high quality knowing that later contract modifications will allow it to earn the agreed corruption rent.

The corrupt contract award process is modelled using multiple regression linking likely corruption inputs (e.g. eligibility criteria tailored to one company) to likely corruption outcomes (e.g. only one company submitting a bid) while controlling for alternative explanations (e.g. number of competitors on the market). Our models linking corrupt inputs to outcomes in public procurement directly follow from the definition of institutionalised grand corruption. Hence, they explain recurrent contract award to a pre-selected company with those corruption techniques that typically serve as means for corruptly eliminating competitors. Given that the theoretical model postulates that both inputs and outputs of the corruption process as organised by the corrupt network, no causal claim is made based on the models, even though temporal sequence of inputs and outputs could lend some support to a causal interpretation.

The explanatory model linking corruption inputs to outcomes delivers a set of coefficients that represent the strength of association (not causal impact!) between each underlying likely corruption input and likely corruption outcome. Reliability of elementary corruption indicators is defined using their regression coefficients, as those corruption inputs which are more powerful in predicting probable corruption outcomes are more likely to signal corruption rather than noise. Falsely indicating corruption is minimised by dropping those indicators which didn't prove to be powerful and significant predictors in the model and assigning lower component weights to those whose effect is only moderate.

Likely corrupt outcomes, having no regression coefficients, receive a weight of 1 reflecting their benchmark status in modelling the corruption process. Based on our knowledge of corruption practices, likely corruption outcomes measure most directly the underlying corrupt transactions hence their benchmark status. Every powerful-

enough corruption input receives a weight between 0 and 1, reflecting the size of its regression coefficient and our theoretical understanding of how corruption inputs lead to outcomes. This means that all weights are scaled compared to likely corruption outcomes. For more on weighting see section 6.1.

For comparison across time and countries, both the list of components and component weights should be kept constant unless there are differences in the regulatory framework or types of corrupt practices warranting any deviation. This is because some corruption inputs may be unused in some countries and periods while widely used in others. Giving different weights balances the validity of the composite indicator within its context with keeping measurement consistent across time and countries; a balancing act which is easiest if regulations and typical forms of corrupt practices are very similar. As corruption techniques can substitute for each other, the different component weights reflect institutional features impacting on the form but not the substance of institutionalised grand corruption (For details of a comparative appplication of this index see Authors, 2013c). This implies conducting laborious calculations for fitting this approach to each context such as country, following the approach set out for Hungary.

Using the weights obtained from the measurement model, elementary indicators are simply summed to produce the corruption risk composite indicator of individual transactions (CRI). Summation reflects the view that different combinations of elementary corruption techniques can equally render a contract corrupt; while multiple signs of corruption indicate higher corruption risks. As sophisticated actors can achieve corrupt control of a tender even with recourse to a single corruption technique implies that CRI is only a lower bound estimate of 'true' corruption, a significant limitation. CRI is defined as follows:

$$CRI^i = \Sigma_j \ w_j * CI_j^{\ i} \tag{1}$$

$$\Sigma_j \ w_j = 1 \tag{2}$$

$$0 \leq CRI^i \leq 1 \tag{3}$$

$$0 \leq CI_j^i \leq 1 \tag{4}$$

where CRI^i stands for the corruption risk index of contract i, $CI_j^{\ i}$ represents the jth elementary corruption indicator observed in the tender of contract i, and w_j represents the weight of elementary corruption indicator j. Elementary corruption indicators can be either corruption inputs or outputs.

Higher level units' such as organisations' CRI can be obtained by calculating the arithmetic average of their transactions' CRI in a given period (it is also possible to use contract values for weighting). An added value of aggregating CRI to a higher unit of observation such as an issuer of tenders is that it further increases our confidence in CRI. An organisation consistently displaying high CRI is likely to be a corrupt organisation rather than simply a victim of random fluctuations in the data.

4. DATA

The database derives from Hungarian public procurement announcements of 2009-2012 (this database is referred to as PP henceforth). The data represent a complete database of all public procurement procedures conducted under Hungarian Public Procurement Law. PP contains variables appearing in 1) calls for tenders, 2) contract award notices, 3) contract modification notices, 4) contract completion announcements, and 5) administrative corrections notices. As not all of these kinds

of announcements appear for each procedure, for example depending on procedure type, we only have the variables deriving from contract award notices consistently across every procedure. Comparable datasets exist or can be constructed from public records in most developed countries including every EU member state, Russia, and the US at least since 2008 (for examples see Annex B).

The place of publication of these documents is the Public Procurement Bulletin which is accessible online[4]. As there is no readily available database, we used a crawler algorithm[5] to capture the text of every announcement. Then, applying a complex automatic and manual text mining strategy, we created a structured database that contains variables with clear meaning and well-defined categories. As the original texts available online contain a range of errors, inconsistencies, and omissions, we applied several correction measures to arrive at a database of sufficient quality for scientific research. For a full description of database development, see Authors (2012a) in Hungarian and in somewhat less detail Authors (2012b) in English.

A potential limitation of our database is that it only contains information on public procurement procedures under the Hungarian Public Procurement Law as there is no central depository of other contracts. The law defines the minimum estimated contract value for its application depending on the type of announcing body and the kind of products or services to be procured (for example, from 1 January 2012, classical issuers have to follow the national regulations if they procure services for more than 8 million HUF or 27 thousand EUR). By implication, PP is a biased sample of total Hungarian public procurement of the period, containing only the

---

[4] See: http://www.kozbeszerzes.hu/nid/KE (in Hungarian)
[5] For a gentle introduction see: http://en.wikipedia.org/wiki/Web_crawler

larger and more heavily regulated cases. This bias makes PP well suited for studying more costly and more high stakes corruption where coverage is close to complete. In spite of any such limitations, the scale of data available is daunting, over 53,000 contracts awarded worth over 11 billion EUR, amounting to over 3% of total GDP between 2009-2012.

5. BUILDING BLOCKS: THE CORRUPTION PROCESS' OUTCOMES AND INPUTS

5.1 Indicators of corruption outcomes

The selection of "red-flags" to be included in our composite corruption risk index was informed by a comprehensive review of the literature and fieldwork in Hungary with people familiar with corruption in public procurement. This included 14 interviews with officials, contractors, and procurement advisors conducted in 2012-2013. This fieldwork yielded a large list of potential "red-flags". Our indicator includes 13 of these "red-flags" that are predictive of likely corrupt outcomes in regression analysis. In order to minimize the odds that our "red-flags" are indicating incompetence, inexperience, or unusual procurement markets, we restrict our analysis to competitive markets and experienced tenderers.

The key outcome of institutionalised corruption in public procurement is recurrently awarding contracts to a pre-selected company. We use three outcome indicators to match this corrupt outcome as closely as possible (see Table 1). First, the lack of competition during the bid submission phase is indicated by a single bid submitted as it allows the public officials to award the contract at a higher than market price. Second, during the assessment phase excluding all the 'unwanted' bidders leads to

a very similar non-competitive outcome. Hence, we use the exclusion of all but one received bids as an indication of a potentially corrupt outcome. Finally, in order to capture the recurrent award of contracts to the same company, we use the winner's share within the issuer's contracts as an indication of a corrupt tendering outcome. If corruption is institutionalised it is likely to lead to a high market concentration further underpinning the view that competition is tilted to the advantage of connected bidders.

Table 1. Summary of outcome indicators

phase	indicator name	Definition
submission	single bidder	1=1 bid received, 0=more than 1 bid received
assessment	exclusion of bids	1=1 bid NOT excluded, 0=more than1 bid NOT excluded
overall	winner's share of issuer's contracts	Previous 12-month total contract value of winner / 12-month total awarded contract value (by issuer)

5.1.1 Single bidder

Issuers of tenders are free to choose the bidder of their preference; however, they are mandated to maximise value for money, most importantly through soliciting competing bids. Corruption arises when competition is blocked in order to earn a corruption rent. The most obvious signal that there was absolutely no competition for a public contract is when a tender received only one bid (Amaral, Saussier, & Yvrande-Billon, 2009). Interview evidence from Hungary suggests that tenders with only 2-3 bids are also highly likely to be prone to corruption, as one public procurement adviser working in the industry for over a decade put it: "it is easy, just bring two friends with whom we can agree on the exact content of their bids". Focusing only on single bidder contracts is, therefore, a conservative approach in line with the goal of delivering a lower bound estimate of large-scale corruption.

There are two potential criticisms to this indicator: 1) The single bidder indicator also signals corruption in cases when there was truly only one bidder capable of performing the task, but no corruption took place. While this is a serious weakness of

the indicator, it is considered to be only of marginal magnitude as the overwhelming majority of products procured by governments are ordinary and widely produced such as office stationery, cars, national roads, or IT support services (less than 5 percent of contracts were awarded on markets with 3 or fewer companies). 2) Some authors contend that a single bidder has no incentive to give a bribe (Soreide, 2002). However, in an environment of systemic corruption, a single bidder tender is the ideal outcome created by colluding bidders and issuers, especially if the same single bidder wins contracts repeatedly (see section 5.1.3).

5.1.2 Exclusion of all but one bidder

It is possible that a corrupt issuer didn't manage to deter all but one bidder from submitting a bid, in which case it can still award the contract to the 'well-connected' bidder if it manages 1) to exclude the bids of all unwanted bidders on administrative or formal grounds (Heggstad and Froystad, 2011); or 2) to unfairly assess the bids to favour a particular bidder. As there is no data in public records for the latter, the assessment phase's corruption outcome indicator captures only the former. Having a single valid bid tender is heavily associated with corruption for the same reasons as for single submitted bid. This similarity between the two measures, is also supported by regression results.

5.1.3 Winner's share of issuer's contracts

While there is no separate indicator for the delivery phase, we develop a likely corruption outcome measure for the public procurement corruption process as a whole. The ultimate goal of large-scale institutionalised corruption is to repeatedly award contracts to the same company or companies controlled by the corrupt group (Heggstad and Froystad, 2011). By implication, winner's share of issuer's previous contracts indicates the likelihood of such corruption. As the primary location of

collusion and capture is the individual public organisation disbursing public funds, this variable is defined as the ratio of contract value the winner won from a given issuer to the total value of contracts awarded by the given issuer throughout the 12-month period prior to the tender being analysed.

Using winner's share within issuer's contracts (or winner's contract share as we will call it to remain succinct) as a corruption indicator is likely to suffer from disturbances in periods when a new dominant group takes control of public organisations with its new clients, for example when a new government comes into office. Changes of dominant, captor groups are expected to be rare events, hence, this downward bias may only be moderate (and controlling for year of contract award in the below regressions captures much of this potential bias). Moreover, this indicator also underestimates corruption when the corrupt network uses multiple companies for extracting rents such as multiple companies owned by one group or independent companies orchestrating their bids.

5.2 Indicators of corruption inputs

These (likely) corrupt outcomes described above are achieved through a variety of corruption "inputs" or techniques to eliminate competitors. The international literature and interviews in Hungary revealed 30 such techniques which could be reliably calculated using publicly available public procurement records (see Authors 2013a). However, we only retained 14 of them for the composite index building as these turned out to be significant and substantial predictors of all three outcome variables in regression analyses, suggesting fit with a corrupt rent extraction logic.

Table 2 summarizes the 14 corruption input factors included in our composite corruption risk index (descriptive statistics in Annex C). Some of the variables are

continuous, and we had to determine thresholds over/under which we considered them a "red-flag." We describe this process of determining thresholds below.

Table 2. Summary of corruption inputs (higher score indicates greater likelihood of corruption)

phase	indicator name	indicator definition
submission	Single bidder contract*	0=more than one bid received 1=ONE bid received
	Call for tender not published in official journal	0=call for tender published in official journal 1=NO call for tenders published in official journal
	Procedure type	0 =open procedure 1=invitation procedure 2=negotiation procedure 3=other procedures (e.g. competitive dialogue) 4=missing/erroneous procedure type
	Relative length of eligibility criteria	number of characters of the eligibility criteria MINUS average number of characters of the given market's eligibility criteria
	Length of submission period	number of days between publication of call for tenders and submission deadline
	Relative price of tender documentation	price of tender documentation DIVIDED BY contract value
	Call for tenders modification	0=call for tenders NOT modified 1=call for tenders modified
assessment	Exclusion of all but one bid	0=at least two bids NOT excluded 1=all but one bid excluded
	Weight of non-price evaluation criteria	proportion of NON-price related evaluation criteria within all criteria
	Annulled procedure re-launched subsequently**	0=contract awarded in a NON-annulled procedure 1=contract awarded in procedure annulled, but re-launched
	Length of decision period	number of working days between submission deadline and announcing contract award
delivery	Contract modification	0=contract NOT modified during delivery 1=contract modified during delivery
	Contract lengthening	relative contract extension (days of extension/days of contract length)
	Contract value increase	relative contract price increase (change in contract value/original, contracted contract value)

* The single bidder indicator is simultaneously an outcome of the submission phase and an input to the assessment phase.

** Combining annulations by the issuer and the courts

Table 3 describes the logic linking corruption inputs to corruption outputs (for more details see [Authors 2013a]). Single received bid and single valid bid outcomes are discussed jointly because the theoretical considerations are very similar and the regressions reveal largely the same findings.

The expectations are formulated in a general linear form. For example, the shorter the submission period is the more likely that only one bid was received. However, many of the continuous variables are indeed not a continuous measure of corruption

risks, rather there are critical thresholds beyond which corruption risks greatly increase. For example, a submission period of 5 days compared to 15 days is likely to convey higher corruption risks while a submission period of 35 days compared to 45 days may carry little to no information regarding corruption. By implication, behind any of our linear hypotheses lies the expectation of finding the thresholds which best capture spikes in the probability of a corruption outcome hence corruption risks.

In every case, the input variables are defined in a way that their higher values are expected to signal higher corruption risks. However, some of the corruption inputs are typically used as 'corrective action' later on in the procurement process to fix the failed attempts at bending competition earlier. These factors are expected to have negative association with corruption outcomes of earlier stages. For example, if only the well-connected company submitted a bid there is no need for subsequently modifying the contract as the corrupt bidder could set the price and quality allowing for corrupt rent extraction. However, if there was real competition at the submission phase the well-connected bidder is likely to be forced to submit a competitive bid with little scope for earning extra profit; hence the need for subsequent contract modification.

6. REGRESSION SPECIFICATIONS

This section discusses the regressions modelling institutionalised grand corruption in public procurement and derives component weights for composite indicator building..

The regressions' primary purpose is to determine whether a red-flag should be included into the composite index, as well as its weight in the index. As stated above, only those indicators that predict a single bid or exclusion of all but one bid and/or

the winner's share of issuer's contract in the previous 12 months, are included into the index. We also use the size of the regression coefficients to determine the weight of the technique in the index, since we want those techniques that are the greatest predictors of corruption outcomes to have a higher weight. Adopting this method will facilitate cross-national comparisons, as different techniques will be utilized differentially in various institutional settings. The regressions also provide the primary source of internal validity of the composite indicator. For outcomes single received bid and single valid bid, we used binary logistic regression; while for the winner's contract share outcome, we used linear regression.

Table 3. Summary of the expected direction of and grounds for the relationships between corruption inputs and outputs

Phase	INPUT/OUTPUT	single received / valid bid		winner's share within issuer's contracts	
		direction	Reason	direction	reason
Submis-sion	Single bidder contract	**not relevant**	not relevant	**+**	Single received bid contracts make it easier for issuers to repeatedly award contracts to the same well-connected company.
	Call for tender not published in official journal	**+**	Not publishing the call for tenders in the official journal makes it less likely that eligible bidders notice the bidding opportunity and bid.	**+**	Not publishing the call for tenders in the official journal weakens competition allowing the issuer to more easily award contracts repeatedly to a well-connected company.
	Procedure type	**+**	Non-open procedures, which are less transparent and require less open competition, create more opportunities to limit the range of bids received and to exclude bids.	**+**	Non-open procedures, which are less transparent and require less open competition, create more opportunities for issuers to repeatedly award contracts to the same well-connected company.
	Relative length of eligibility criteria	**+**	Lengthy, hence complex, eligibility criteria allows issuers to tailor the tender to a single company and to exclude unwanted bids.	**+**	Lengthy, hence complex, eligibility criteria allows issuers to benefit a well-connected company, for example by keeping less competitive bidders in competition.
	Short submission period	**+**	A short submission period leaves less time hence make it harder for non-connected companies to bid and to submit a bid.	**+**	A short submission period leaves less time hence make it harder for non-connected companies to bid successfully whereas a well-connected firm can use its inside knowledge to win repeatedly.
	Relative price of documentation	**+**	Relatively expensive tender documentation makes bidding more expensive and hence deters bidders from bidding except for the well-connected company which is close to certain of its success.	**+**	Relatively pricey tender documentation weakens competition allowing the issuer to more easily award contracts repeatedly to a well-connected company.
	Call for tenders modification	**+**	Modifying call for tenders allows for excluding unwanted bidders by changing eligibility criteria once the interested bidders are known.	**+**	Strategic modification of the call for tenders favours the well-connected company further increasing its market share.
Assess-ment	Exclusion of all but one bid	**not relevant**	not relevant	**+**	Single valid bid contracts make it easier for issuers to repeatedly award contracts to the same well-connected company.
	Weight of non-price evaluation criteria	**+**	Non-price related evaluation criteria tend to be more subjective, allowing issuers to favour the well-connected company. Apparently unfair assessment criteria deters bidders.	**+**	Non-price related evaluation criteria tend to be more subjective, allowing issuers to favour the well-connected company, hence repeatedly awarding contracts to the same company.
	Annulled procedure re-launched subsequently*	**-**	If unwanted bidders couldn't be deterred from bidding and their bids couldn't be excluded, annulling and subsequently re-launching the tender allows issuer to correct its failed attempt to eliminate competition.	**+**	If unwanted bidders couldn't be deterred from bidding and their bids couldn't be excluded, annulling and subsequently re-launching the tender allows issuer to more successfully award the contract to a well-connected company.
	Length of decision period	**+**	Overly lengthy decision period signals extensive legal challenges to the tender, suggesting that the issuer attempted to limit competition.	**+**	Lengthy decision periods signal extensive legal challenge to the tender, suggesting that the issuer wants to award the contract to a well-connected company.
Delivery	Contract modification	**-**	If competition couldn't be eliminated, the well-connected firm can still win with a competitive offer, but subsequent contract modification(s) still allow it to collect extra profit.	**+**	Contract modification(s) suggests that the issuer corruptly favour a well-connected company, potentially repeatedly.
	Contract lengthening	**-**	If competition couldn't be eliminated, the well-connected firm can still win with a competitive offer, but subsequent contract lengthening still allows it to collect extra profit.	**+**	A contract lengthening suggests that the issuer corruptly favour a well-connected company, potentially repeatedly.
	Contract value increase	**-**	If competition couldn't be eliminated, the well-connected firm can still win with a competitive offer, but subsequent contract value increase still allows it to collect extra profit.	**+**	A contract value increase suggests that the issuer corruptly favour a well-connected company, potentially repeatedly.

The regressions are also used to account for non-linearity in the indicators of corruption inputs, that is identifying indicator values beyond which corruption risks considerably increase or jump. Hence, one value range was identified for each continuous corruption input which could be deemed in line with a market or economic logic; furthermore, at least one other value range was defined which was considered as an outlier compared to the prevailing market norm while also being strongly associated with the likely corruption outcomes This approach led to a definition of red flags which is determined by the empirical patterns rather than evoking arbitrary thresholds (Kenny and Musatova, 2010; Pricewaterhouse Coopers, 2013).

In any regression, a significant and large coefficient is interpreted as indicating that the given corruption input is typically used for reaching the corruption output even after taking into account alternative explanations, such as contract size or length, and all other corruption inputs. This still means that it can be used for other, non-corrupt purposes in atypical cases; conversely, all the non-significant and weak explanatory factors may still be used for corrupt purposes, albeit only exceptionally.

Component weights of the composite indicator are derived from regression coefficients; whereby, the larger coefficient means a higher component weight. This reflects the view that the more often a corruption input is used in combination with corruption outcomes the more confident we can be that institutionalised grand corruption lies behind its use.

6.1 Modelling corrupt rent extraction: component weights

Regression models were built based on the above theoretical expectations by entering each corruption input and controls step-by-step, entering first those inputs

which characterise the earliest tender phase such as publication of call for tenders and entering finally those which come into play the last such as contract completion. All those inputs were dropped from the models which were insignificant and/or too small to matter. Here, only final regression results are reported for the sake of brevity. The regressions are fitted only on markets with at least 3 different winners in 2009-2012 to reduce the false positives from our measure of corrupt outcomes (single bidder contract). As the validity of all three outcome variables crucially hinges on the availability of suitable competitors, robustness checks are presented in Annex E excluding markets with fewer different winners throughout 2009-2012. The conclusions are substantially the same on the restricted samples too.[6] We also restricted the analysis to experienced bidders (defined as Awarding at least 3 contracts over 12 months), to reduce the possibility that the red-flags were due to inexperience.

Thresholds in each continuous variable were identified in an iterative process: first, a model was fitted using the linear continuous predictor; second, discrete jumps in residual values were identified using residual distribution graphs. For example, average residual values of the regression using all the control variables plus the linear continuous measure of the relative price of documentation for predicting single received bid are depicted in Figure 1, left panel. It clearly indicates that there are three distinctive groups of relative document prices. For the lowest region, ranging between approximately the 24th and 40th percentiles, the model overestimates the probability of a single received bid, while it is the opposite case for the region between the 70th and 100th percentiles. These suggest at least three distinct

---

[6] Further robustness checks were done excluding issuers which awarded few contracts. Regression results confirm the robustness of models on samples with issuers awarding at least 5, 10, and 50 contracts. Regression outputs can be obtained from the authors.

categories. The right panel of Figure 1 shows the same residual distribution after the categorical measure of relative document price replaced its continuous version in the model with categories following the cut-points identified earlier. No clear pattern remains in the residual distribution, suggesting most non-linearity has been accounted for by the categorical measure of relative document price. A similar procedure was followed in the case of every continuous variable; if necessary completing multiple iterations of searching for thresholds.

In order to preserve the full population of observations, we always included a missing category in every categorical variable. In addition, this also helped measuring corruption inputs as concealing relevant tender information from bidders or the wider public often serves as a corruption technique.

Figure 1. Mean regression residuals by two-percentiles of relative price of documentation, left panel: linear prediction; right panel: prediction after taking into account non-linearity
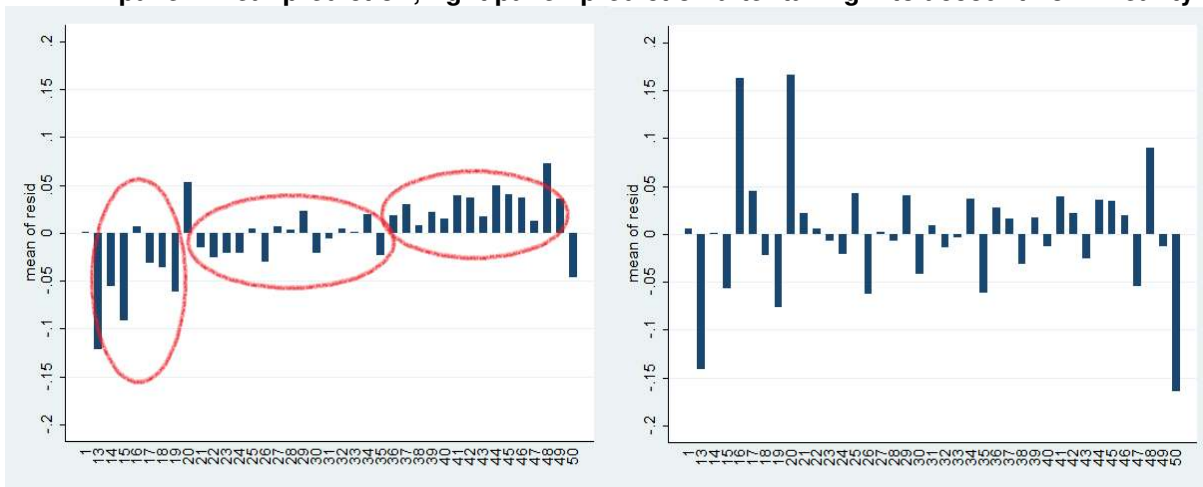
Source: PP

When deciding on whether a variable is significant in the model, we used significance values from Monte Carlo random permutation simulations (Good, 2006), even though standard Fisher significance tests would have led to the same conclusions in most cases. This is because standard Fisherian significance tests are appropriate for statistical inference from a random sample to a population. However,

24

our public procurement database contains the full population of interest. While some observations have been removed purposefully from the public domain hence from the database (a corruption risk on its own which is certainly far from random) this cannot be reflected by Fisher significance tests. Permutation tests are widely used in the natural as well as the social sciences, for example in social network analysis where data typically relates to full populations and observations are not independent of each other (Borgatti, Everett, and Johnson, 2013). The Monte Carlo random permutation simulation randomly reassigns the outcome variable to observations multiple times and calculates the regression coefficients each time. By doing so, it obtains a distribution of each regression coefficient when the outcome is truly random. The probability of the actual test statistic falling outside this random distribution, therefore, represents the probability of observing the relationship when the outcome is truly random. A low significance level indicates that it is highly unlikely that the observed regression coefficient could be the result of a random process – a very intuitive interpretation.

Three different regressions are reported in Table 4, two binary logistic regressions on single received bid and single valid bid, following the same structure:

$$\Pr(\text{single bidder}_i = 1) = \frac{1}{1 + e^{-Z_i}} \tag{5}$$

$$Z_i = \beta_0 + \beta_{1j} S_{ij} + \beta_{2k} A_{ik} + \beta_{3l} D_{il} + \beta_{4m} C_{im} + \varepsilon_i \tag{6}$$

where single bidder$_i$ equals 1 if the ith contract awarded had only one bidder and 0 if it has more; Z_i represents the logit of a contract being a single bidder contract; β_0 is the constant of the regression; S_{ij} is the matrix of j corruption inputs of the submission phase for the ith contract such as length of submission period; A_{ik} stands

for the matrix of k corruption inputs of the assessment phase for the ith contract such

weight of non-price evaluation criteria; D_{il} stands for the matrix of l corruption inputs

of the delivery phase for the ith contract such contract lengthening; C_{im} stands for the

matrix of m control variables for the ith contract such as the number of competitors

on the market; ε_i is the error term; and β_{1j}, β_{2k}, β_{3l}, and β_{4m} represent the vectors of

coefficients for explanatory and control variables.

In addition to the two logistic regression models in Table 4, a linear regression on

winner's share within issuer's contracts is reported following the structure:

$$Y_i = \beta_0 + \beta_{1j}S_{ij} + \beta_{2k}A_{ik} + \beta_{3l}D_{il} + \beta_{4m}C_{im} + \varepsilon_i \tag{7}$$

where Y_i represents winner's share within issuer's contracts; β_0 is the constant of the

regression; S_{ij} is the matrix of j corruption inputs of the submission phase for the ith

contract such as length of submission period; A_{ik} stands for the matrix of k corruption

inputs of the assessment phase for the ith contract such weight of non-price

evaluation criteria; D_{il} stands for the matrix of l corruption inputs of the delivery phase

for the ith contract such contract lengthening; C_{im} stands for the matrix of m control

variables for the ith contract such as the number of competitors on the market; ε_i is

the error term; and β_{1j}, β_{2k}, β_{3l}, and β_{4m} represent the vectors of coefficients for

explanatory and control variables.

The main differences among regressions are the outcome variables. Each

regression includes the full list of controls and predictors having non-missing values

in the given sample. Control variables account for the most important alternative

explanations to our corrupt outcomes such as low administrative capacity and

product market idiosyncrasies, in particular (for descriptive statistics of all control

variables see Annex F): 1) type of product procured using 40 different CPV[7] divisions which control for differences in technology and market standards; 2) number of unique winners throughout 2009-2012 on the product market using a matrix of 820 CPV categories at level 3 and 4 geographical regions using NUTS[8] definitions which makes sure that our findings on single bidders and winner's share within issuer's contracts are not driven by the low number of competitors available on the market;[9] 3) year of contracting which by and large proxies the changes in the legal framework and government in power; 4) log real contract value[10] (2009 constant prices) controlling for the differences emanating from contract size and complexity; 5) contract length in years controlling for the differences emanating from contract size and complexity; 6) whether the contract is a framework contract which have specific regulations and procedural rules[11]; and 7) issuer type (e.g. central, regional), sector (e.g. education, health care), and status (private vs public) controlling for the regulatory as well as the institutional specificities of different issuers.

To make the cleanest test of our index we only run the regressions in competitive markets with experienced issuers. We restrict the regressions to markets with at least 3 unique winners throughout 2009-2012. These markets are defined by a matrix of 820 CPV categories at level 3 and 4 geographical regions using NUTS definitions. Level 3 CPV categories identify markets such as medical equipment or beverage and spice crops; NUTS

---

[7] CPV=Common Procurement Vocabulary. For more info see: http://simap.europa.eu/codes-and-nomenclatures/codes-cpv/codes-cpv_en.htm

[8] NUTS=Nomenclature of territorial units for statistics. For more info see: http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/introduction

[9] A potential criticism against excluding such markets is that they may be the most corrupt markets biasing results. However, taking the full period of 2009-2012 means that there was a change of national government and many changes in local and regional administrations. Hence, the probability that the same market remained captured by the same 1 or 2 firms is very low implying that the observed low number of unique winners reflects market specificity rather than corruption. In addition, models run on the full sample revealed qualitatively identical cut-points and coefficents.

[10] Log contract values are used instead of actual contract values because the contract value distribution is highly skewed with a few large contracts distorting results.

[11] For details see:? http://ec.europa.eu/internal_market/publicprocurement/docs/explan-notes/classic-dir-framework_en.pdf

regions denote West, Central, and Eastern Hungary plus the country as a whole for national markets. So the matrix of these two dimensions creates markets such as medical equipment in Western Hungary or beverage and spice crops in Eastern Hungary. To ensure the tenderers are experienced we restrict the analysis to issuers awarding at least 3 contracts in the 12 months period prior to the contract award in question.

Regression Results

By and large, our hypotheses are supported by regressions, warranting the construction of a composite indicator reflecting systematically corrupt public procurement (Table 4). The explanatory power of models is moderate, R^2 ranging between 0.1 and 0.2, which partially reflects the diversity of corruption techniques in this turbulent period of Hungary, while also indicating the need for including further corruption techniques, especially those which require further public data disclosure (e.g. evaluation scores per bidder).

First, the single received or valid bid is a powerful predictor of winner's share within issuer's contracts. Those contracts with a single bid tend to be awarded to winners with 1.8 percent higher share within issuer's previous contracts on average compared to contracts with more than one bids. This significant effect confirms that restricting the number of bids to one can support corrupt rent extraction on a recurrent basis. The magnitude of the impact is modest which is not surprising as restricting competition at the submission phase is only one of many ways to circumvent competition in public procurement.

Second, not publishing the call for tenders in the official journal increases the probability of single received and valid bids and the winner's contract share in every regression in line with expectations. For example, in model 1 and 2, it increases the

average probability of a single received bid contract award by 12.1-14.0 per cent, which is one of the strongest impact across models.

Third, every non-open procedure type carries a higher corruption risk than open procedures in terms of single received and valid bids and winner's contract share, supporting and further refining our theoretical expectations. Other, exceptional procedures carry the highest corruption risks adding 2.9 percent to winner's share within issuer's contracts compared to open procedures. Invitation and negotiation procedures are powerful and significant predictors in the regressions explaining single bidder contracts, but they have weak or counterintuitive impacts in the winner's contract share regressions that suggests that their main effect is likely to come through number of bidders. Invitation procedures appear to have about twice as strong effect on the probability of a single bidder contract award (6.0-7.1 percent) as negotiation procedures (3.0-5.8 percent).

Fourth, relative length of eligibility criteria behaves as expected with lengthier, thus more complex, criteria associated with higher probability of a single bidder contract and higher winner contract share. The effect of criteria length around the market average length seems weak, but positive indicating that there may be markets where complex criteria is frequently used to deter bidders. Criteria lengths considerably higher than market average are especially strongly associated with higher probability of single bidder contracts and higher winner contract share. For example, criteria length above market average by 520-2639 characters[12] increases probability of a single received bid by 6.3-10.4 percent and the winner's share within issuer's contracts by 1.3 percent compared to the shortest criteria-length group. Interestingly,

---

[12] Standard deviation of character lengths from the population mean is 3435 for the whole 2009-2012 period. So, eligibility criteria 2639 characters above its market average is about three quarters standard deviation difference.

the call for tenders which are published, but don't contain eligibility criteria at the section where it is prescribed by law, are associated with especially high corruption risks: 1.8-9.0 percent higher probability of single received bid contract compared to the shortest character length group. This signals that making eligibility criteria less visible deters bidders.

Fifth, the shorter the submission period the higher the probability of single received and valid bids and winner contract share in line with expectations. This relationship appears in distinct jumps around legally prescribed thresholds and the abuse of weekends[13]. The exceptionally short submission period abusing weekends is one of the most powerful predictors in all of the models. It increases the winner's share within issuer's contracts by 7.6 percent and the probability of single valid bid by 19.8-21.6 percent. Similar to criteria length, not displaying visibly and clearly the submission deadline is associated with very high corruption risks, for example 16 percent higher probability of single received bid. As the effect is negligible on winner contract share, this corruption technique's impact arises primarily in the submission phase.

Sixth, more expensive tender documents increase both the probability of single bidder contracts and winner contract share in line with expectations. Compared to free documentation, document prices between 0.04-0.1 percent of the contract value increase the probability of single received bid by 2.9 percent and increase winner's share within issuer's contracts by 3.5 percent. Even more expensive tender documents have a stronger impact in the single bidder regressions, but insignificant and small effect in the winner contract share regression. This indicates that their

---

[13] Abuse of weekends is possible as legally required time periods are defined in calendar days so the effective time companies would have for bid preparation can further be decreased by including weekends and national holidays in the submission period.

main effect is exercised in the submission phase. The effect of the cheapest tender documentation is ambiguous across regressions. Missing tender documentation price is insignificant in most regressions. Therefore, these categories receive a zero weight in the composite indicator.

Seventh, call for tenders modifications behave according to expectations only for the period of the previous government (before 01/05/2010)[14], that is it increases the probability of single bidder contracts and the winner's market share. While it takes on a considerable significant negative coefficient under the current government' period. These differences signal the changing role call for tenders modifications may play in corrupt rent extraction in response to changing regulatory (e.g. new Public Procurement Law entering into force soon after the new government entered into force) and political climate such judicial review of modifications (interviews indicate that the regulations and practice of judicial review of procurement tenders changed considerably after the new government entered office). Call for tenders modifications receive a positive weight in the composite indicator only for the pre-May 2012 period reflecting a conservative approach.

---

[14] Restricted sample results are not reported here. Regression outputs can be obtained from the authors.

Table 4. Regression results on contract level, 2009-2012, average marginal effects reported for models 1-2 and unstandardized coefficients for model 3, nr. of winners >=3

models	1	2	3
Independent vars / dependent vars	single received bid	single valid bid	winner's 12 month market share
single received/valid bid			0.018***
no call for tenders published in official journal	0.14***	0.121***	0.039***
procedure type			
ref. cat.=open procedure			
1=invitation procedure	0.071***	0.06***	-0.032*
2=negotiation procedure	0.03***	0.058***	0.009*
3=other procedures	0.274***	0.258***	0.029***
4=missing/erroneous procedure type	0.028***	0.017	-0.008
relative length of eligibility criteria			
ref.cat.=length<-2922.125			
1= -2922.125<length<=520.704	0.046***	0.019	0.001
2= 520.704<length<=2639.729	0.104***	0.063***	0.013
3= 2639.729<length	0.124***	0.071***	0.014
4= missing length	0.09***	0.018***	0.048***
short submission period[15]			
ref.cat.=normal submission period			
1=accelerated submission period	0.022***	0.007	0.014***
2=exceptional submission period	0.09***	0.084***	0.047***
3=except. submission per. abusing weekend	0.216***	0.198***	0.076***
4=missing submission period	0.16***	0.028	-0.009
price of tender documentation relative to contract value			
ref.cat.= relative price=0%			
1= 0% <relative price<=0.04%	-0.01	-0.042***	0.062***
2= 0.040% <relative price<=0.01%	0.029**	-0.005	0.035***
3= 0.01% <relative price<=0.211%	0.031***	0.008	0.009
4= 0.211% <relative price	0.049***	0.012	0.000
5=missing relative price	-0.001	-0.017	-0.008*
call for tenders modified	-0.036***	-0.043***	0.017***
weight of non-price evaluation criteria			
ref.cat.= only price			
2= 0<non-price criteria weight<=0.4	-0.019***	-0.034***	-0.002
3= 0.4<non-price criteria weight<=0.556	0.069***	0.05***	0.028***
4= 0.556<non-price criteria weight<1	0.076***	0.075***	0.038***
5=only non-price criteria	0.001	-0.012	0.007***
procedure annulled and re-launched	-0.112***	-0.031*	
length of decision period in days			
ref.cat.= 44<decision period<=182			
1= decision period<=32	0.078***	0.117***	0.013**
2= 32<decision period<=44	0.032***	0.047***	0.016***
4= 182<decision period	0.147***	0.161***	0.046***
5= missing decision period	-0.02	-0.016	0.022*
contract modified during delivery	-0.004	-0.024***	0.015***
contract extension(length/value)			
ref.cat.=c.length diff.<=0 AND c.value diff.<=0.001			
2=0<c.length d.<=0.162 OR 0.001<c.value d.<=0.24	-0.061***	-0.026	-0.01
3= 0.162<c. length diff. OR 0.24<c.value diff.	-0.017	0.000	-0.006
4= missing (with contr. completion ann.)	-0.022**	-0.017*	-0.002
5= missing (NO contr. completion ann.)	-0.011*	0.005	0.003
constant included in each regression; control variables: product market (cpv divisions); number of winners on the market (market defined by cpv level 4 and nuts 1) year of contract award; log real contract value; contract length; framework contract; issuer type, sector, and status (public or private)			
N	52390	42607	20653
pseudo-R2 (logistic reg) *or* R2 (linear reg.)	0.104	0.099	0.243

*Source: PP; Note: * p<0.05; ** p<0.01; *** p<0.001; P(permute) significance levels are obtained by Monte Carlo random permutation simulations (200 permutations) using stata 12.0*

---

[15] Exact number of days used for these categories differ per procedure type and year reflecting legally permissible deadlines. As an example, for open procedures before 2011, normal period is longer than 21 days, accelerated period is between 21 and 14 days, exception period is between 14 and 5 days (excluding the weekend), and exceptional period abusing the weekend is less than 5 days including the weekend.

Eight, the effect of the weight of non-price evaluation criteria turned out to be somewhat different from expectations. Instead of a clearly positive relationship, we found an inverted U-shape relationship. This can be interpreted using our interview evidence: stipulating only or predominantly price-related evaluation criteria warrants fair competition, hence, it is associated with lower corruption risks. While majority subjective criteria suggests rigged competition deterring bidders and increasing winner contract share. Only non-price evaluation criteria combined with fixed price is most likely complying with certain industry standards such as IT procurement without signalling heightened corruption risks (Authors, 2013a). Hence, only the two categories with positive coefficient receive non-zero weights in the composite indicator.

Ninth, annulling and re-launching procedures has the expected sign for both single received and single valid bid outcomes, but its effect cannot be determined on winner contract share as annulled tenders would bias contract share figures. Annulling a contract award is associated with 3.1-11.2 percent lower probability of single bidder contract award, that is contract awards are annulled and re-launched more often when there were multiple bidders. This is completely contradictory to the prescriptions of the EU Public Procurement Directive or the Hungarian Public Procurement Law, but in line with a corrupt rent extraction logic.

Tenth, the effects of decision period length on probability of single bid and winner contract share are both somewhat different from our expectations. It seems that the relationship follows a U-shaped pattern with average decision period lengths (between 40th and 90th percentile) having the lowest corruption risk. Compared to this reference category, exceptionally long decision periods and exceptionally short decision periods are both associated with high corruption risks. Decision periods

longer than 182 working days result in 14.7-16.1 percent higher probability of single bid contract and 4.6 percent higher winner's share within issuer's contracts. Decision periods shorter than 32 working days are associated with 7.8-11.7 percent higher probability of single bid contract and 1.3 percent higher winner contract share. Decision periods between 32 and 44 working days have a somewhat weaker effect than exceptionally short decision periods. These results suggest that there are two mechanisms at play. First, exceptionally short decision periods may indicate rushed through decisions and the corresponding high corruption risks. Second, exceptionally long decision periods may signal multiple legal challenges and troubled decision making hence high corruption risks. While the missing category is significant in some models, its effect is far from clear, thus, it cannot be included in the composite indicator.

Eleventh, contract modification has the expected relationships with probability of single bid and winner contract share albeit effect sizes are small in general and insignificant for model 1-2. Modifying contract at least once after contract award is associated with 2.4 percent lower probability of single valid bid and 1.5 percent higher winner's share within issuer's contracts. This indicates that a competitive contract award procedure may necessitate contract modification to assure rent extraction.

Twelfth, increasing contract length and increasing the contract value after contract award had to be considered together due to low number of relevant observations. These two techniques can be combined in as much as they represent two parallel methods for increasing the profitability of a contract, that is making delivery cheaper by extending the completion deadline or making price higher by increasing contract

value. Contract extension (length/value) displays the expected relationships, but effects are insignificant for the winner contract share regression.

Compared to contracts which were performed within the timeframe of delivery and original contract price (less than 0.1 percent value increase), contracts with 0-16.2 percent longer delivery period or 0.1-24 percent higher contract value were associated with 2.6-6.1 percent lower probability of single received bid. For contracts which were extended even more the effects are insignificant which may signal that excessive project overruns are more often due to non-corrupt reasons such as low state capacity. For contracts whose contract completion announcement didn't contain the prescribed final contract length or final contract value information, the probability of single bid was 1.1 percent lower which is a moderately strong impact. This suggests that competitive tendering makes it more necessary to hide the final total time of performance which potentially deviated from the original contractual terms. Hence, contract extensions of moderate magnitude and missing information are included in the composite indicator.

Based on these regression results and prior theory, the variables and their categories could be selected which will make up the composite Corruption Risk Index (CRI) (for descriptive statistics of each component see Annex D). Component weights have to reflect theory, quantitative as well as qualitative results, and the authors' assessment of the precision of these. First, all three corruption outcomes are part of CRI because their strong theoretical backing and close conceptual association with the definition of institutionalised grand corruption. We assign outcome variables a weight of 1 reflecting their benchmark status. Second, qualitative evidence clearly underlines that any of the corruption inputs (i.e. corruption techniques) is sufficient on its own to render a procurement procedure

corrupt. Therefore, each significant and substantial corruption input receives the weight of 1. In order to reflect coefficient sizes of categories in each corruption input as well as acknowledging the imprecision of point estimates, we ranked categories of each variable with the most impactful category receiving weight 1 and the others proportionately lower weights. For example, if there are four significant categories of a variable, then they would get weights 1, 0.75, 0.5, and 0.25. Finally, we normed each component weight so that the resulting composite indicator falls between 0 and 1 (Table 5). This was achieved in two steps: component weights were divided by the total number of components (N=13), then the resulting score was divided by its observed maximum (CRI[raw]=0.805). This rescaling assures that the minimum (maximum) of the score corresponds to the lowest (highest) corruption risks observed. The upper end of the scale may be too conservative as the combined presence of 3-4 corruption inputs and/or outputs (CRI=0.27-0.36) is already almost certainly very corrupt according to our interviewees[16].

Once components are identified using the above regressions, alternative techniques could also be considered for deriving composite score(s), for example using predicted outcomes (risk prediction) (for a highly problematic, but similar approach see Pricewaterhouse Coopers, 2013), developing corruption risk profiles using clustering techniques, or employing item response theory to extrapolate from observed characteristics to latent corruption risks. These avenues remain for further research as our aim was to develop a CRI score that is intuitive and straightforward.

Using a simple weighted sum of inputs and outputs of the corrupt rent extraction process to derive CRI is methodologically preferred to the above alternatives for at

---

[16] Calculating CRI for court decisions which established corruption in public procurement could serve as a more robust upper bound for the CRI scale.

least two reasons. First, additivity reflects our increasing certainty in corruption to take place rather than any need for combining corruption techniques for reaching corrupt goals (recall, even a single corruption input is enough on its own to render a tender fully corrupt), while any technique can be substituted for by another one. In particular the simultaneous presence of inputs and outputs represent reliable measurement of corruption risk. Second, the specified models and indicators only imperfectly capture corruption, there are omitted corruption techniques while the included techniques are only measured with considerable error. Hence, including inputs as well as outputs of the corruption process increases the measurement reliability of the composite score as different omissions and measurement errors are present on the two sides of the equation. Think for example about capturing single bidding as an outcome without any associated corruption technique due to omitted measure for the latter. In this case, including single bidding in CRI is superior to reporting no corruption risk at all.

Table 5. Component weights of CRI reflecting variable and category impact on corruption outcomes, normed to have an overall sum of 1

Variable	component weight
single received/valid bid	0.096
no call for tenders published in official journal	0.096
procedure type	
ref. cat.=open procedure	0.000
1=invitation procedure	0.048
2=negotiation procedure	0.072
3=other procedures	0.096
4=missing/erroneous procedure type	0.024
relative length of eligibility criteria	
ref.cat.=length<-2922.125	0.000
1= -2922.125<length<=520.7038	0.024
2= 520.7038<length<=2639.729	0.048
3= 2639.729<length	0.072
4= missing length	0.096
short submission period	
ref.cat.=normal submission period	0.000
1=accelerated submission period	0.048
2=exceptional submission period	0.072
3=except. submission per. abusing weekend	0.096
4=missing submission period	0.024
relative price of tender documentation	
ref.cat.= relative price=0	0.000
1= 0<relative price<=0.0004014	0.000
2= 0.0004014<relative price<=0.0009966	0.096
3= 0.0009966<relative price<=0.0021097	0.064
4= 0.0021097<relative price	0.032
5=missing relative price	0.000
call for tenders modification(only before 01/05/2010)	0.096
weight of non-price evaluation criteria	
ref.cat.= only price	0.000
2= 0<non-price criteria weight<=0.4	0.000
3= 0.4<non-price criteria weight<=0.556	0.048
4= 0.556<non-price criteria weight<1	0.096
5=only non-price criteria	0.000
procedure annulled and re-launched subsequently	0.096
length of decision period	
ref.cat.= 44<decision period<=182	0.000
1= decision period<=32	0.064
2= 32<decision period<=44	0.032
4= 182<decision period	0.096
5= missing decision period	0.000
contract modified during delivery	0.096
contract extension(length/value)	
ref.cat.= c.length diff.<=0 AND c.value diff.<=0.001	0.000
2= 0<c. length d.<=0.162 OR 0.001<c.value d.<=0.24	0.096
3= 0.162<c. length diff. OR 0.24<c.value diff.	0.000
4= missing (with contr. completion ann.)	0.048
5= missing (NO contr. completion ann.)	0.000
winner's market share	0.096

Source: PP

Note: If the call for tenders or contract fulfilment announcements are missing, the index is reweighted to only reflect the available variables (i.e. proportionately increasing the weight of observed variables).

6.2 Validating the corruption risk index

Validating CRI will take several years of work in multiple countries using many different indicators. Here only a few simple validation procedures can be reported. First, we look at the cross-sectional distribution of CRI to see if it follows any standard distribution underpinning its use in quantitative analyses such a linear regression. Second, the co-variation between CRI and a range of external hard measures of corruption risks is explored.

First, applying the weights specified in Table 5, each contract receives a corruption risk index (CRI) falling into a 0–1 band. CRI of each awarded contract follows an approximately normal distribution, albeit it has a long tail to the right. These contracts with CRI higher than approximately 0.45-0.50 represent particularly high corruption risks and hence deserve attention in later research. The distribution also implies that CRI can be treated as a continuous variable in later regression analyses (same applies to organisation-level analysis).

Second, we utilize four different external indicators for validating CRI, each of which is expected to co-vary with CRI in a particular way if CRI is a valid and reliable measure of corruption. All the tests are confirmatory.

Profitability

First, highly profitable companies are expected to display higher CRI because the primary aim of institutionalised grand corruption is to generate extra profit considerably above market average. However, we believe this relationship is likely to be only of moderate magnitude and probabilistic as high corruption companies are often hiding their profits and turnover through offshore companies, chains of

subcontractors, and tax fraud. These have been confirmed by interviews in Hungary and international research (Cole and Tran, 2011a, 2011b).

Simple comparisons of companies falling in the terciles of annual profit margin reveal a relationship in line with expectations (Figure 2). Percentile comparisons are preferable to simple correlations as corruption and profitability may have a non-linear relationship (linear correlation coefficient is only 0.05). Companies of highest profit margin (higher than 5.15%) display higher CRI than any other company group, but the difference is especially large when compared to the group of lowest profit margin companies (annual profit margin<1.44%): 0.024 points higher CRI score or 10% percent more risky (0.024/0.24).

Figure 2. Mean CRI by profit margin terciles, 2009-2012, N=3357
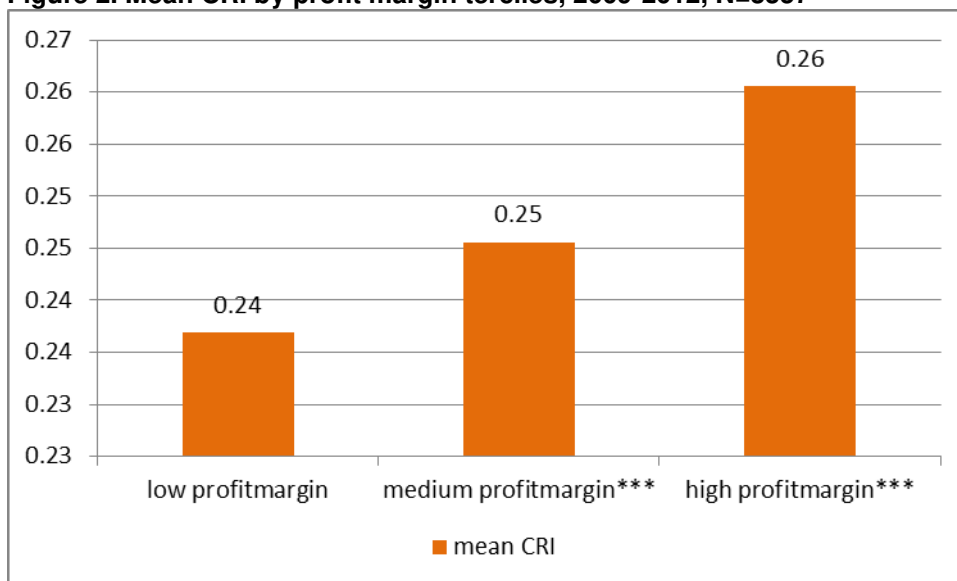
Source: PP
Note: p<0.05; ** p<0.01; *** p<0.001 designate the significance of the difference from the "low CRI" group. Significance levels computed using Monte-Carlo random permutations (300 repetitions) with stata 12.0*

Ratio of price of final contract to price advertised in tender

Second, further supporting the link between rent extraction and CRI, the ratio of the actual contract price to contract value to the advertised price in the tender is explored. As detailed information on unit prices is missing, the only way to determine

how expensive public procurement was is to compare the originally estimated contract value with the final contract value. Higher values of the ratio of contract value to estimated contract value indicate more expensive tenders (Ishii, 2009; Morozov & Podkolzina, 2013; Padhi & Mohapatra, 2011). By implication, this price ratio is expected to be positively associated with corruption risks.

We find the expected relationship. The price ratio is positively correlated with CRI (corr=0.14) which indicates that prices are higher in high corruption risk tenders. This relationship is also pronounced when looking at the terciles of the price ratio which avoids the pitfalls for assuming a linear relationship (Figure 3). Contracts with price ratios above 1 display somewhat lower average CRI than contracts with price ratios between 0.9 and 1 which underlines the nonlinear relationship between prices and corruption risks.

Figure 3. Mean CRI by price ratio terciles, 2009-2012, N=28 390
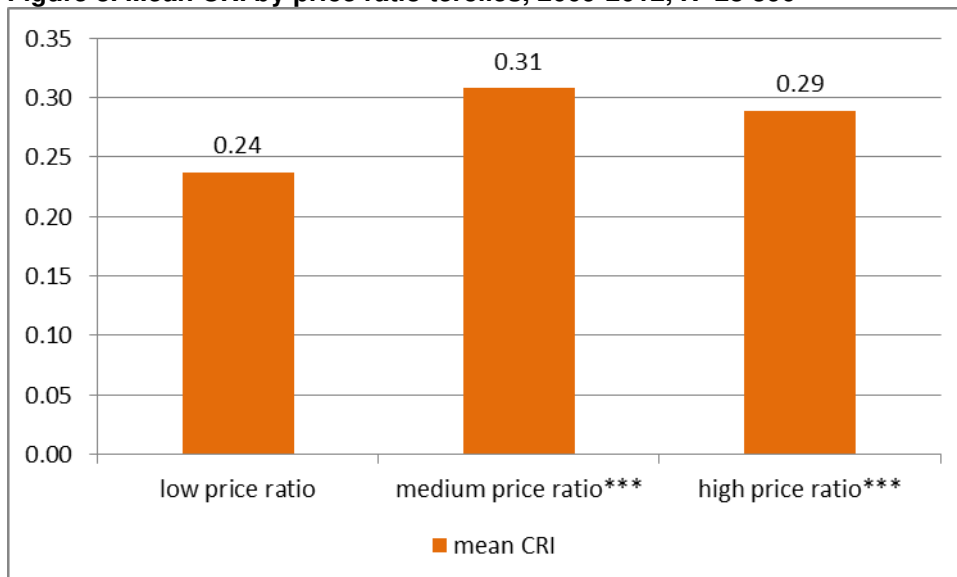
Source: PP
Note: p<0.05; ** p<0.01; *** p<0.001; Significance levels computed using Monte-Carlo random permutations (300 repetitions) with stata 12.0*

Political ties

Third, we expect that companies with political connections to display higher corruption risks as the primary vehicle for maintaining institutionalised grand

corruption is to have strong ties between powerful political and business actors (Coviello and Gagliarducci, 2010; Grodeland, 2005). We mapped the owners and manager of each company winning in 2009-2012 (15 percent of companies were either unidentifiable or we lacked the relevant data) and matched them with key political officeholders of public organisations existing in the period (for full list of institutions and offices see Table 6). The matching was done between more than 35000 owners/managers of winning firms and more than 10000 political officeholders based on full name. Matching solely on name is obviously prone to random error that is nevertheless set aside for the present analysis by assuming that name frequency is not correlated with CRI. Those companies that have or had at least one owner or manager holding a political office at any point in time were designated as politically connected firms.

Table 6. List of institutions and positions of the political office holder database, 2010-2011

Institution	Position
Ministries	minister, secretary of state, vice-secretary of state, ministerial councillor,
Constitutional court	members and leaders
County courts	president, vice- president
Supreme court	President, vice-president, spokesperson
Prosecutors' Office	Chief prosecutor, vice-chief prosecutor, spokesperson
Municipalities	Major, vice-major, notary
County governments (new "kormányhivatal" too)	president, vice-president, notary
Regional police	Chief
National police headquarters	Chief, vice-chief, spokesperson
Minority governments	president, vice-president, head of office head of secretary
National medical service	Chief doctor, chief pharmacist
National Healthcare Fund	Director, vice-director
Army headquarters	Marshal, Vice-marshal
Treasury	President, vice-president, head of finances
Tax Administration	President, vice-president, spokesperson
Office of the president	President of the state, heads of every bureau of the office
State Audit Office	President, vice-president, chief director, director of finances
Regional Development Councils	presidents, member of governing committee
Office of the parliament	Head of office, heads of offices
Ombudsmen offices	Ombudsmen, heads of offices
National headquarters of Prisons	National chief, national vice-chief,
Competition Authority	President, vice-president, head of secretary
Central statistical office	president, vice-president
Other regulatory agencies and background institutes	top-management (2-3 positions)

In line with our expectations, politically connected firms have a higher CRI by 0.01 on average than companies without political connections. While this difference is relatively small,[17] increasing the precision of identifying political connections could shed more light on the validity of CRI. The magnitude of group differences may also signal that direct political connections serve as a means to corruption only in some cases while in others political ties operate through seemingly unrelated brokers and a range of intermediaries.

Registration in tax havens

Finally, we expect that high corruption companies are strongly associated with money laundering and secretive company registration arrangements (Gounev and Bezlov, 2010). These are most straightforwardly measured by the Financial Secrecy Index (FSI) of the countries where winning companies or their owners are registered. The FSI measures to what degree a country's company registry, tax, and financial regulatory framework facilitates secrecy and provides for tax heavens exclusively based on objective data on regulations and their implementation (Tax Justice Network, 2013). In line with prior expectations, companies registered in tax heavens (FSI>58.5) are of considerably and significantly higher corruption risks as measured by CRI compared to companies registered in relatively transparent countries (Figure 4). The difference is quite substantial (ΔCRI=0.02) even though relatively few companies are registered outside Hungary (only slightly more than 2 percent winning firms).

---

[17] This difference remains unchanged when taking into account the number of managers with different names in the database suggesting that the relationship is not an artefact of matching politicians and businessman based on name only.

Figure 4. Average CRI scores of companies of foreign origin according to the Financial Secrecy Index of the domicile country, 2009-2012, N=414
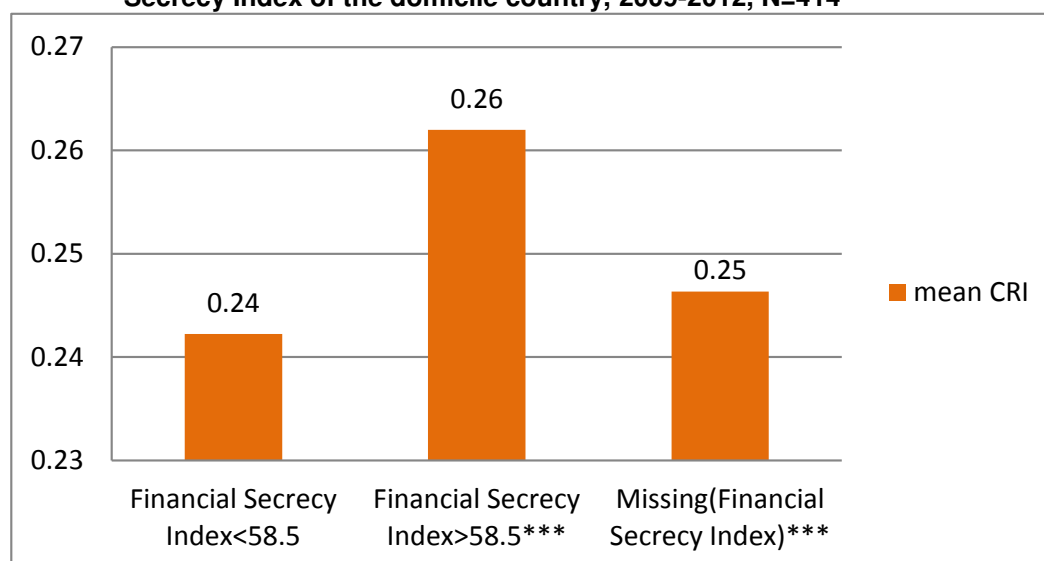
8. CONCLUSIONS

The analysis demonstrated that it is feasible and fruitful to measure corruption at the micro-level based exclusively on 'objective' behavioural data. Initial evidence confirms the validity of CRI. The great advantage of our approach is that a large amount of data is available for research across every developed country starting from about 2008, opening up a completely new horizon for comparative corruption research. Such comparative research will also benefit from a measure of corruption that avoids the pitfalls of subjective indicators as well as prior 'objective' indicators like institutional arrangements. Even if one is happy with existing measures of corruption, the CRI adds an entirely new measure that can be used to validate other analyses.

Following this article's methodology institutionalised grand corruption can be measured on the level of individual contracts and tenders in about 50 countries around the globe since 2008-2009. This opens up a whole new universe of scientific and policy applications on many levels. The CRI is calculated at the individual contract level, and can be scaled up to the tendering organization, the sector, local jurisdiction, region or the country. This flexibility opens up a large number of hypotheses that can be tested, but we mention only a few. First, data on tendering organizations can be combined with organizational level data to test various theories of corruption. For example, are higher salaries for bureaucrats more conducive to clean government? Perhaps there is a non-linear relationship between bureaucratic size/pay and corruption? Are bigger bureaucracies more prone to corruption than smaller ones? Since the CRI is valid over time, scholars can also evaluate how various reforms in different sectors affect corruption. Does deregulation or privatization increase or decrease corruption? Since the CRI can be calculated at the local level, scholars can estimate the impact of political contestation on corruption. Since it can also be aggregated by funding source, scholars can investigate whether external funds (e.g. EU funds) are more prone to corruption than internally generated funds.

Beyond the academy, the CRI can be used to evaluate single regulatory or organisational changes such as tightening reporting requirements or introducing integrity management, providing data to evaluate what works to reduce corruption. The CRI could also guide regulators in where to spend their limited resources for conducting audits of contracts and companies. In addition to these and many other academic and policy applications, with a little work the CRI can be made available to citizens, civil society groups and journalists to hold politicians and political parties

accountable for corrupt behaviour. It could also be used by businesses in deciding where to spend their resources in preparing bids for tenders.

In sum, the CRI can revolutionize the study of corruption, and provide an invaluable resource for limiting corruption in the real world. We hope that it inspires other attempts to utilise the vast pool of data now available in digital format ("big data") to generate objective indicators of various components of state activity and capacity.

BIBLIOGRAPHY

Akey, P. (2013). Valuing Changes in Political Networks: Evidence from Campaign Contributions to Close Congressional Elections. *Available at SSRN 2336131*.

Amaral, M., Saussier, S., & Yvrande-Billon, A. (2009). Auction procedures and competition in public services: The case of urban public transport in France and London. *Utilities Policy*, *17*, 166–175.

Andersson, S., & Heywood, P. M. (2009). The Politics of Perception: Use and Abuse of Transparency International's Approach to Measuring Corruption. *Political Studies*, *57*, 746–767.

Arndt, C., & Oman, C. (2006). *Uses and Abuses of Governance Indicators*. Paris: OECD.

Auriol, E., Flochel, T., & Straub, S. (2011). *Public Procurement and Rent-Seeking: The Case of Paraguay* (No. 11-224). TSE Working Papers: 11-224, Toulouse: Toulouse School of Economics (TSE).

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing Social Networks*. London: SAGE Publications.

Charron, N., Dijkstra, L., & Lapuente, V. (2010). *Mapping Quality of Government in the European Union: A Study of National and Sub-National Variation* (No. 2010:22). Gothenburg.

Cole, S., & Tran, A. (2011a). *Bribes for Sales: Evidence from the Second Set of Books*.

Cole, S., & Tran, A. (2011b). Evidence from the Firm: A New Approach to Understanding Corruption. In S. Rose-Ackerman & T. Søreide (Eds.), *International Handbook on the Economics of Corruption, volume two*. Cheltenham, UK: Edward Elgar Publishing.

Coviello, D., & Gagliarducci, S. (2010). *Building Political Collusion: Evidence from Procurement Auctions* (No. 4939). IZA DP No. 4939, Bonn: Institute for the Study of Labor (IZA).

European Commission. (2011). *Establishing an EU Anti-corruption reporting mechanism for periodic assessment ("EU Anti-corruption Report")*. Brussels.

Fazekas, M., Chvalkovská, J., Skuhrovec, J., Tóth, I. J., & King, L. P. (2014). Are EU funds a corruption risk? The impact of EU funds on grand corruption in Central and Eastern Europe. In A. Mungiu-Pippidi (Ed.), *The Anticorruption Frontline. The ANTICORRP Project, vol. 2.* (pp. 68–89). Berlin: Barbara Budrich Publishers.

Fazekas, M., & Tóth, I. J. (2012a). *Hibák, javítások és előzetes eredmények - magyarországi közbeszerzések 2010-2011*. Corruption Research Centre, Budapest.

Fazekas, M., & Tóth, I. J. (2012b). *Public Procurement, Corruption and State Capacity in Hungary – objective measures and new insights*. Corruption Research Centre, Budapest.

Fazekas, M., Tóth, I. J., & King, L. P. (2013). *Corruption manual for beginners: Inventory of elementary "corruption techniques" in public procurement using the case of Hungary* (No. CRC-WP/2013:01). CRC-WP/2013:01,Corruption Research Centre, Budapest.

Golden, M. A., & Picci, L. (2005). Proposal for a New Measure of Corruption, illustrated with Italian data. *Economics & Politics*, *17*(1), 37–75. doi:10.1111/j.1468-0343.2005.00146.x

Goldman, E., Rocholl, J., & So, J. (2013). Politically Connected Boards of Directors and The Allocation of Procurement Contracts. *Review of Finance*, *17*(5), 1617–1648. doi:10.1093/rof/rfs039

Good, P. I. (2006). *Resampling Methods A Practical Guide to Data Analysis* (3rd editio.). Berlin: Birkhauser.

Gounev, P., & Bezlov, T. (2010). *Examining the links between organised crime and corruption*. Sofia: Center for the Study of Democracy.

Grodeland, A. B. (2005). Bulgaria, Czech Republic, Romania and Slovenia: The Use Of Contacts and

Informal Networks in Public Procurement. In OECD (Ed.), *Fighting Corruption and Promoting Integrity in Public Procurement* (pp. 59–76). Paris: OECD.

Heggstad, K. K., & Froystad, M. (2011). *The basics of integrity in procurement*. Bergen, Norway: U4 Anti-Corruption Resource Centre Chr. Michelsen Institute (CMI).

Hyytinen, A., Lundberg, S., & Toivanen, O. (2008). *Politics and Procurement: Evidence from Cleaning Contracts* (No. 233). HECER Discussion paper No. 233.

Ishii, R. (2009). Favor exchange in collusion: Empirical study of repeated procurement auctions in Japan. *International Journal of Industrial Organization*, *27*(2), 137–144.

Kaufmann, D., Mastruzzi, M., & Kraay, A. (2010). *The Worldwide Governance Indicators. Methodology and Analytical Issues* (No. 5430). Washington, DC.

Kenny, C., & Musatova, M. (2010). *"Red Flags of Corruption" in World Bank Projects: An Analysis of Infrastructure Contracts* (No. Policy Research Working Paper 5243) (p. Policy Research Working Paper 5243, World Bank, Washington DC). Washington, DC.

Kurtz, M. J., & Schrank, A. (2007a). Growth and Governance: A Defense. *Journal of Politics*, *69*(2), 563–569.

Kurtz, M. J., & Schrank, A. (2007b). Growth and Governance: Models, Measures, and Mechanisms. *The Journal of Politics*, *69*(2), 538–554.

Lambsdorff, J. G. (2006). Measuring Corruption – The Validity and Precision of Subjective Indicators (CPI). In C. Sampford, A. Shacklock, C. Connors, & F. Galtun (Eds.), *Measuring Corruption* (pp. 81–100). Aldershot, UK: Ashgate.

Morozov, I., & Podkolzina, E. (2013). *Collusion detection in procurement auctions* (No. WP BRP 25/EC/2013). Moscow.

Mungiu-Pippidi, A. (2006). Corruption: Diagnosis and Treatment. *Journal of Democracy*, *17*(3), 86–99.

Mungiu-Pippidi, A. (Ed.). (2011). *Contextual Choices in Fighting Corruption: Lessons Learned*. Oslo: Norwegian Agency for Development Cooperation.

North, D. C., Wallis, J. J., & Weingast, B. R. (2009). *Violence and Social Orders. A Conceptual Framework for Interpreting Recorded Human History*. Cambridge, UK: Cambridge University Press.

OECD. (2009). *Towards a Sound Integrity Framework: Instruments, Processes, Structures and Conditions for Implementation*. Paris: OECD.

Olken, B. A. (2007). Monitoring Corruption: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy*, *115*(2), 200–249. doi:10.1086/517935

Olken, B. A., & Pande, R. (2012). *Corruption in Developing Countries*. Boston, Massachusetts.

Padhi, S. S., & Mohapatra, P. K. (2011). Detection of collusion in government procurement auctions. *Journal of Purchasing and Supply Management*, *17*(4), 207–221.

Pricewaterhouse Coopers. (2013). *Identifying and reducing corruption in public procurement in the EU*. Brussels: PricewaterhouseCoopers and Ecorys.

Radelet, S. (2002). *Qualifying for the Millennium Challenge Account*.

Radelet, S. (2003). Bush and Foreign Aid. *Foreign Affairs*, *82*(5), 104–117.

Rose, R., & Peiffer, C. (2012). *Paying Bribes to Get Public Services: A Global Guide to Concepts and Survey Measures* (No. SPP 494). Glasgow, UK: Centre for the Study of Public Policy.

Rothstein, B., & Teorell, J. (2008). What Is Quality of Government? A Theory of Impartial Government Institutions. *Governance*, *21*(2), 165–190.

Soreide, T. (2002). *Corruption in public procurement. Causes, consequences and cures*. Bergen, Norway.

Tax Justice Network. (2013). *Financial Secrecy Index 2013. Methodology*. Chesham, UK.

Transparency International. (2012a). *Corruption Perceptions Index 2012*. Berlin: Transparency International.

Transparency International. (2012b). *NIS Assessment Toolkit*. Berlin: Transparency International.

ANNEXES

ANNEX A - SUMMARY OF SELECTED STUDIES USING OBJECTIVE INDICATORS OF CORRUPTION

Table A1. Summary of selected studies using objective indicators of corruption

paper	indicator used	Country	year	sector	potential for international comparison	part of CRI*
Auriol et al. (2011)	Exceptional procedure type	Paraguay	2004-2007	general procurement	HIGH If procedure definitions can be aligned, international comparisons can be made widely	Yes
Bandiera, Prat, and Valletti (2009)	Price differentials for standard goods purchased locally or through a national procurement agency	Italy	2000-2005	various standardized goods (e.g. paper)	LOW Price data is not readily available in most countries, many countries don't have national procurement agencies, national procurement agencies are likely to be captured in many countries.	No
Coviello and Gagliarducci (2010)	Number of bidders Same firm awarded contracts recurrently Level of competition	Italy	2000-2005	general procurement	HIGH Number of bidders, recurrent contract award, and competitiveness of bids are available in many countries.	Yes
Di Tella and Schargrodsky (2003)	Difference in prices of standardized products such as ethyl alcohol	Brazil	1996-1997	health care	MEDIUM Detailed product-level price and quantity information is not readily available across many countries, but can be collected.	No
Ferraz and Finan (2008)	Corruption uncovered by federal audits of local government finances	Brazil	2003	federal-local transfers	LOW high quality audits, not influenced by powerful corrupt groups are unlikely to be available in many countries.	No
Golden and Picci (2005)	Ratio of physical stock of infrastructure to cumulative spending on infrastructure	Italy	1997	infrastructure	MEDIUM It is hard to compute comparable value of the stock of physical capital across countries different in the quality of infrastructure and geography.	No
Goldman et al. (2013)	Political office holders' position on company boards	USA	1990-2004	general procurement	HIGH Company contract volumes can be estimated in many countries and publicly listed companies political connections can be traced relatively easily.	No
Hyytinen et al. (2008)	Number and type of invited firms Use of restricted procedure	Sweden	1990-1998	cleaning services	HIGH Both number of bidders and procedure types are readily	Yes

					available in many countries.	
Olken (2006)	Difference between the quantity of in-kind benefits (rice) received according to official records and reported survey evidence	Indonesia	1998-1999	welfare spending	MEDIUM It is possible to design user surveys across a wide range of countries to track actual receipts, although it may be expensive.	No
Olken (2007)	Differences between the officially reported and independently audited prices and quantities of road construction projects	Indonesia	2003-2004	infrastructure (roads)	LOW Auditing large numbers of projects by independent engineers is costly and unlikely to allow for cross-country comparisons.	No
Klasnja (2015)	Single bidder procedures Non-open procedure types	Romania	2008-2012	general procurement	HIGH If procedure definitions and bidding conditions can be aligned, international comparisons can be made widely	Yes
Reinikka and Svensson (2004)	Difference between block grants received by schools according to official records and user survey	Uganda	1991-1995	education	MEDIUM It is possible to design user surveys across a wide range of countries to track actual receipts, although it may be expensive.	No

*CRI=Corruption Risk Index, developed in this paper

ANNEX B - AVAILABILITY OF PUBLIC PROCUREMENT DATA

Table B1. Overview of contract-level public procurement data availability in selected countries and regions, 2000-2012

Country	Data-source	Key online source	Minimum threshold (2012, classical issuer, services, EUR)[18]	Period	Availability
Czech Republic	Ministerstvo pro místní rozvoj ČR	http://www.isvzus.cz/usisvz/	39,000	2006-2012	structured data readily available and partially cleaned
EU	Tenders Electronic Daily	http://ted.europa.eu/	130,000	2005-2012	structured data partially available and cleaned
Germany	Bund.de- Verwaltung Online	http://www.bund.de/DE/Ausschreibungen/ausschreibungen_node.html	130,000[19]	2010-2012[20]	raw data available, not cleaned
Hungary	Közbeszerzési Értesítő	http://www.kozbeszerzes.hu/	27,300	2005-2012	structured data available and partially cleaned
Romania	eLicitatie	http://www.e-licitatie.ro/	30,000	2007-2012	raw data available, not cleaned
Russia	Goszakupki	www.zakupki.gov.ru	2,500	2006-2012[21]	structured data partially available and cleaned
Slovakia	Úrad pre verejné obstarávanie	http://tender.sme.sk/en/	30,000	2005-2012	structured data readily available and partially cleaned
UK	UK Contracts Finder	http://www.contractsfinder.businesslink.gov.uk/	11,600	2000-2012	raw data available, not cleaned
US	Federal Procurement Data System - Next Generation	https://www.fpds.gov/fpdsng_cms/	1,850	2004-2012	structured data readily available and partially cleaned

---

[18] National currencies are converted into EUR using official exchange rates of 5/2/2013 of the European Central Bank.
[19] It was increased from 30,000 EUR during the economic crisis.
[20] Earlier data have to be requested from the relevant bodies.
[21] 2006-2010 only for some regions.

ANNEX C – DESCRIPTIVE STATISTICS OF CORRUPTION INPUTS

Table C1. Descriptive statistics of corruption inputs, 2009-2012, markets with at least 3 unique winners

	mean	min	max	sd	N
Single bidder contract	0.301	0.00	1.00	0.46	51012
Exclusion of all but one bid	0.367	0.00	1.00	0.48	41277
Call for tender not published in official journal	0.388	0.00	1.00	0.49	51823
Length of submission period in days	27.080	2.00	751.00	12.62	29885
Relative price of tender documentation	0.003	0.00	0.20	0.01	16743
Call for tenders modification	0.109	0.00	1.00	0.31	31726
Annulled procedure re-launched subsequently	0.061	0.00	1.00	0.24	55217
Weight of non-price evaluation criteria	0.216	0.00	1.00	0.33	51823
Length of decision period in days	126.960	0.00	1404.00	168.31	28605
Contract modification	0.189	0.00	1.00	0.39	51823
Contract lengthening	0.014	-0.97	30.29	0.26	16238
Contract value increase	0.079	-0.80	5.00	0.53	6547

Source: PP

Table C2. Distribution of procedure type, 2009-2012, markets with at least 3 unique winners

	N	%
open	31,007	59.83
invitation	906	1.75
negotiation	9,510	18.35
other	5,760	11.11
missing/error	4,640	8.95
Total	51,823	100

Source: PP

ANNEX D – DESCRIPTIVE STATISTICS OF CRI COMPONENTS

Table D1. Descriptive statistics of CRI components, 2009-2012, markets with at least 3 unique winners

CRI component name	mean	median	sd	N
single bidder	0.353	0.000	0.478	45,652
no call for tenders published	0.336	0.000	0.472	45,865
procedure type	0.279	0.000	0.385	45,865
length of eligibility criteria	0.311	0.250	0.348	45,865
submission period	0.074	0.000	0.173	45,865
document relative price	0.126	0.000	0.277	45,865
call for tenders modification	0.026	0.000	0.158	45,865
weight of non-price criteria	0.110	0.000	0.279	45,844
Cancellation & relaunch	0.073	0.000	0.260	45,865
decision period	0.185	0.000	0.313	45,865
contract modification	0.197	0.000	0.397	45,865
price incr./contr. lengthening	0.095	0.000	0.217	45,865
winner contract share	0.127	0.071	0.169	43,749

Table D2. Pairwise linear correlation coefficients among CRI components, 2009-2012, markets with at least 3 unique winners

	single bidder	no call for tenders published	procedure type	length of eligibility criteria	submission period	document relative price	call for tenders modification	weight of non-price criteria	Cancellation & relaunch	decision period	contract modification	price incr./contr. lengthening	winner contract share
single bidder	1												
no call for tenders published	0.0117*	1											
procedure type	0.1176*	0.4217*	1										
length of eligibility criteria	0.0110*	-0.6350*	-0.2336*	1									
submission period	0.0259*	-0.3048*	-0.0055	0.0845*	1								
document relative price	-0.0002	-0.3227*	-0.1279*	0.1432*	0.0526*	1							
call for tenders modification	-0.0026	-0.1152*	-0.0103*	-0.0228*	0.1306*	0.0316*	1						
weight of non-price criteria	0.0093*	-0.1856*	-0.1024*	0.5362*	0.0468*	0.0194*	-0.0124*	1					
Cancellation & relaunch	0.0961*	-0.1246*	-0.0003	0.1082*	0.0809*	0.0619*	0.0757*	0.0153*	1				
decision period	0.0841*	-0.4211*	-0.2086*	0.2610*	0.0016	0.1671*	-0.0421*	0.0686*	0.0587*	1			
contract modification	-0.0007	-0.1722*	-0.0198*	0.1477*	0.0635*	0.0695*	0.1205*	0.0503*	0.1558*	0.0683*	1		
price incr./contr. lengthening	0.0141*	-0.1202*	-0.006	0.0744*	0.0709*	-0.0477*	0.0459*	0.0353*	0.1142*	0.1114*	0.1745*	1	
winner contract share	0.0386*	0.0022	0.0767*	0.0577*	0.0546*	0.1250*	0.0098*	0.0154*	0.0824*	-0.0159*	0.1135*	-0.0346*	1

*Note: *=significant at 5% level*

ANNEX E - ROBUSTNESS CHECKS

The most convincing alternative explanation to this paper's interpretation of regressions as models of corrupt contract award states that products and services bought by public agencies are highly specific. Therefore, both single bidder and high share of the winner within the issuer's contracts are driven by the lack of adequate suppliers rather than corruption. In order to control for this important confounding factor each regression contains the number of winners on the market throughout 2009-2012 as an explanatory factor. In addition, this annex reports regressions on restricted samples which include contracts for products and services procured on markets with more than 2, 9, and 37 winners in 2009-2012. The cut-points 2 and 37 were defined using the same technique of identifying thresholds in continuous variables as spelled out in section 6.1. The cut-point of 9 was added arbitrarily in order to display an intermediary value.

To define the number of adequate competitors on a market, an appropriate definition of market has to be found. We defined markets along two dimensions: 1) the nature of product or service procured, and 2) the geographical location of contract performance. CPV codes differentiate over 3000 products and services as detailed as eggs (03142500-3) or potatoes (03212100-1). While we aim at being conservative in market definition, such level of detail is surely excessive. Exploiting the hierarchical nature of CPV classification, level-4 categories were selected as suitable for market definition, because the distribution of winners throughout 2009-2012 suggested that there are a large number of markets with a fairly small winners. Contracts were awarded in 820 level-4 CPV categories such as crops, products of market gardening and horticulture (0311) or construction materials (4411). Even

though Hungary is a relatively small country interviewees suggested that there may be geographical frontiers of markets. Hence, we used 3 NUTS-1 regions plus the whole country to define markets along a geographical dimension (national reach typically requires an extensive set of local offices warranting an effective market barrier). Taken together, these resulted in 820*4=3280 distinct markets.

To define how many suitable competitors a market has, we simply calculated the number of unique winners of each market throughout 2009-2012. This is a conservative estimate as bidders who never won, for example because they were too expensive, but submitted valid bids were not taken into account. As some companies may have gone bankrupt or been bought by others, this estimation strategy may also be somewhat upward biased; therefore in some regressions we excluded markets with very many competitors.

The below tables demonstrate the robustness of our models to excluding markets with specific products and services (Tables E1 and E2). Each of the findings in these alternative specifications remain unchanged compared to the main regressions, while indicators of goodness of fit improve somewhat.

Table E1. Regression results on contract level, 2009-2012, average marginal effects reported for models 1-2 and unstandardized coefficients for model 3, nr. of winners >=38

models	1	2	3
Independent vars / dependent vars	single received bid	single valid bid	winner's 12 month market share
single received/valid bid			0.027***
P(Fisher)			0.000
P(permute)			0.000
no call for tenders published in official journal	0.131***	0.128***	0.057***
P(Fisher)	0.000	0.002	0.000
P(permute)	0.000	0.000	0.000
procedure type			
ref. cat.=open procedure			
1=invitation procedure	0.06***	0.058***	-0.021
P(Fisher)	0.206	0.339	0.471
P(permute)	0.000	0.000	0.195
2=negotiation procedure	0.03***	0.063***	0.013
P(Fisher)	0.074	0.002	0.235
P(permute)	0.000	0.000	0.055
3=other procedures	0.3***	0.281***	0.031***
P(Fisher)	0.000	0.000	0.003
P(permute)	0.000	0.000	0.000
4=missing/erroneous procedure type	0.039***	0.026***	-0.008
P(Fisher)	0.017	0.165	0.379
P(permute)	0.000	0.000	0.275
relative length of eligibility criteria			
ref.cat.=length<-2922.125			
1= -2922.125<length<=520.7038	0.033***	0.009	0.014
P(Fisher)	0.227	0.784	0.233
P(permute)	0.000	0.420	0.175
2= 520.7038<length<=2639.729	0.106***	0.07***	0.022
P(Fisher)	0.001	0.052	0.114
P(permute)	0.000	0.000	0.070
3= 2639.729<length	0.116***	0.068***	0.025
P(Fisher)	0.001	0.087	0.106
P(permute)	0.000	0.000	0.025
4= missing length	0.057***	-0.008***	0.041*
P(Fisher)	0.132	0.841	0.052
P(permute)	0.000	0.000	0.015
short submission period			
ref.cat.=normal submission period			
1=accelerated submission period	0.025***	0.009	0.015***
P(Fisher)	0.028	0.530	0.045
P(permute)	0.000	0.260	0.010
2=exceptional submission period	0.089***	0.065***	0.012
P(Fisher)	0.006	0.090	0.514
P(permute)	0.000	0.000	0.500
3=except. submission per. abusing weekend	0.193***	0.153***	0.039
P(Fisher)	0.004	0.013	0.423
P(permute)	0.000	0.000	0.520
4=missing submission period	0.163***	0.047*	-0.014
P(Fisher)	0.000	0.308	0.641
P(permute)	0.000	0.015	0.495
relative price of tender documentation			
ref.cat.= relative price=0			
1= 0<relative price<=0.0004014	-0.013	-0.047***	0.056***
P(Fisher)	0.531	0.053	0.010
P(permute)	0.295	0.000	0.000
2= 0.0004014<relative price<=0.0009966	0.016	-0.019	0.038***
P(Fisher)	0.455	0.418	0.015
P(permute)	0.195	0.175	0.000
3= 0.0009966<relative price<=0.0021097	0.031***	-0.005	0.012
P(Fisher)	0.135	0.839	0.388
P(permute)	0.005	0.720	0.245
4= 0.0021097<relative price	0.055***	0.015	0.003
P(Fisher)	0.009	0.482	0.803
P(permute)	0.000	0.160	0.765

models	2	4	5
5=missing relative price	0.005	-0.02	-0.012*
P(Fisher)	0.828	0.416	0.304
P(permute)	0.620	0.065	0.180
call for tenders modified	-0.02*	-0.016	0.005
P(Fisher)	0.288	0.538	0.610
P(permute)	0.030	0.105	0.515
weight of non-price evaluation criteria			
ref.cat.= only price			
2= 0<non-price criteria weight<=0.4	0.005	-0.017***	-0.003
P(Fisher)	0.718	0.316	0.722
P(permute)	0.405	0.000	0.585
3= 0.4<non-price criteria weight<=0.556	0.091***	0.069***	0.047***
P(Fisher)	0.000	0.000	0.000
P(permute)	0.000	0.000	0.000
4= 0.556<non-price criteria weight<1	0.102***	0.086***	0.045***
P(Fisher)	0.000	0.000	0.000
P(permute)	0.000	0.000	0.000
5=only non-price criteria	-0.002	-0.009	0.001
P(Fisher)	0.900	0.615	0.893
P(permute)	0.840	0.360	0.865
procedure annulled and re-launched	-0.098***	-0.027*	
P(Fisher)	0.001	0.422	
P(permute)	0.000	0.035	
length of decision period			
ref.cat.= 44<decision period<=182			
1= decision period<=32	0.067***	0.119***	0.014*
P(Fisher)	0.000	0.000	0.110
P(permute)	0.000	0.000	0.020
2= 32<decision period<=44	0.023***	0.042***	0.021***
P(Fisher)	0.067	0.003	0.019
P(permute)	0.000	0.000	0.000
4= 182<decision period	0.147***	0.187***	0.05***
P(Fisher)	0.000	0.000	0.005
P(permute)	0.000	0.000	0.000
5= missing decision period	-0.024*	-0.022	0.032**
P(Fisher)	0.249	0.418	0.112
P(permute)	0.010	0.060	0.005
contract modified during delivery	-0.003	-0.029***	0.023***
P(Fisher)	0.765	0.028	0.001
P(permute)	0.545	0.000	0.000
contract extension(length/value)			
ref.cat.=c.length diff.<=0 AND c.value diff.<=0.001			
2=0<c. length d.<=0.162 OR 0.001<c.value d.<=0.24	-0.063***	-0.026	-0.011
P(Fisher)	0.000	0.269	0.445
P(permute)	0.000	0.110	0.475
3= 0.162<c. length diff. OR 0.24<c.value diff.	-0.015	0.011	-0.008
P(Fisher)	0.468	0.605	0.523
P(permute)	0.335	0.520	0.575
4= missing (with contr. completion ann.)	-0.008	-0.007*	-0.001
P(Fisher)	0.634	0.707	0.883
P(permute)	0.340	0.395	0.825
5= missing (NO contr. completion ann.)	-0.013*	0.007	0.005
P(Fisher)	0.252	0.594	0.582
P(permute)	0.030	0.255	0.380

constant included in each regression; control variables: product market (cpv divisions); number of winners on the market (market defined by cpv level 4 and nuts 1) year of contract award; log real contract value; contract length; framework contract; issuer type, sector, and status (public or private)

	2	4	5
N	36977	30365	13019
R2/pseudo-R2	0.1101	0.1024	0.2558

Source: PP

Note:* p<0.05; ** p<0.01; *** p<0.001; clustered standard errors clustered by issuer for P(Fisher), Monte Carlo random permutation simulations for P(permute) (200 permutations) using stata 12.0

Table E2. Regression results on contract level, 2009-2012, average marginal effects reported for models 1-2 and unstandardized coefficients for model 3, nr. of winners >=110

models	1	2	3
Independent vars / dependent vars	single received bid	single valid bid	winner's 12 month market share
single received/valid bid			0.034***
P(Fisher)			0.000
P(permute)			0.000
no call for tenders published in official journal	0.136***	0.114***	0.032
P(Fisher)	0.001	0.010	0.150
P(permute)	0.000	0.000	0.085
procedure type			
ref. cat.=open procedure			
1=invitation procedure	0.054***	0.05**	-0.054*
P(Fisher)	0.304	0.451	0.196
P(permute)	0.000	0.010	0.030
2=negotiation procedure	0.023**	0.056***	0.032***
P(Fisher)	0.208	0.009	0.051
P(permute)	0.005	0.000	0.000
3=other procedures	0.309***	0.287***	0.037***
P(Fisher)	0.000	0.000	0.007
P(permute)	0.000	0.000	0.000
4=missing/erroneous procedure type	0.037***	0.02	-0.004
P(Fisher)	0.062	0.376	0.741
P(permute)	0.000	0.080	0.660
relative length of eligibility criteria			
ref.cat.=length<-2922.125			
1= -2922.125<length<=520.7038	0.029*	-0.004	0.008
P(Fisher)	0.345	0.896	0.565
P(permute)	0.015	0.785	0.605
2= 520.7038<length<=2639.729	0.093***	0.056***	0.02
P(Fisher)	0.006	0.121	0.247
P(permute)	0.000	0.000	0.195
3= 2639.729<length	0.107***	0.052**	0.027*
P(Fisher)	0.003	0.178	0.140
P(permute)	0.000	0.005	0.035
4= missing length	0.039***	-0.009***	0.018
P(Fisher)	0.325	0.829	0.527
P(permute)	0.000	0.000	0.380
short submission period			
ref.cat.=normal submission period			
1=accelerated submission period	0.025***	0.006	0.014
P(Fisher)	0.062	0.715	0.177
P(permute)	0.000	0.605	0.060
2=exceptional submission period	0.086***	0.062**	0.015
P(Fisher)	0.006	0.120	0.660
P(permute)	0.000	0.005	0.585
3=except. submission per. abusing weekend	0.204***	0.169**	-0.027
P(Fisher)	0.008	0.016	0.501
P(permute)	0.000	0.005	0.765
4=missing submission period	0.165***	0.053*	0.004
P(Fisher)	0.001	0.273	0.907
P(permute)	0.000	0.010	0.885
relative price of tender documentation			
ref.cat.= relative price=0			
1= 0<relative price<=0.0004014	-0.007	-0.063***	0.036
P(Fisher)	0.765	0.029	0.168
P(permute)	0.615	0.000	0.070
2= 0.0004014<relative price<=0.0009966	0.014	-0.04*	0.022
P(Fisher)	0.555	0.146	0.269
P(permute)	0.255	0.015	0.140
3= 0.0009966<relative price<=0.0021097	0.032*	-0.029	-0.004
P(Fisher)	0.193	0.258	0.834
P(permute)	0.020	0.070	0.735
4= 0.0021097<relative price	0.069***	0.009	-0.005
P(Fisher)	0.005	0.707	0.768
P(permute)	0.000	0.540	0.700
5=missing relative price	0.01	-0.039***	-0.038***
P(Fisher)	0.687	0.146	0.033
P(permute)	0.305	0.000	0.000
call for tenders modified	-0.028***	-0.02	0
P(Fisher)	0.118	0.456	0.989
P(permute)	0.000	0.095	0.990

models	2	4	5
weight of non-price evaluation criteria			
ref.cat.= only price			
2= 0<non-price criteria weight<=0.4	-0.005	-0.031***	-0.008
P(Fisher)	0.729	0.087	0.456
P(permute)	0.425	0.000	0.270
3= 0.4<non-price criteria weight<=0.556	0.077***	0.048***	0.049***
P(Fisher)	0.000	0.017	0.007
P(permute)	0.000	0.000	0.000
4= 0.556<non-price criteria weight<1	0.124***	0.102***	0.077***
P(Fisher)	0.000	0.000	0.000
P(permute)	0.000	0.000	0.000
5=only non-price criteria	0.014	0.005	-0.004
P(Fisher)	0.355	0.795	0.751
P(permute)	0.115	0.675	0.720
procedure annulled and re-launched	-0.076***	-0.025	
P(Fisher)	0.007	0.445	
P(permute)	0.000	0.100	
length of decision period			
ref.cat.= 44<decision period<=182			
1= decision period<=32	0.033***	0.089***	0.005**
P(Fisher)	0.015	0.000	0.688
P(permute)	0.000	0.000	0.610
2= 32<decision period<=44	0.019*	0.03**	0.01
P(Fisher)	0.212	0.051	0.441
P(permute)	0.035	0.005	0.305
4= 182<decision period	0.143***	0.159***	0.055***
P(Fisher)	0.000	0.001	0.013
P(permute)	0.000	0.000	0.000
5= missing decision period	-0.035***	-0.038***	0.016
P(Fisher)	0.088	0.177	0.461
P(permute)	0.000	0.000	0.440
contract modified during delivery	0.001	-0.023**	0.022***
P(Fisher)	0.922	0.102	0.015
P(permute)	0.835	0.005	0.000
contract extension(length/value)			
ref.cat.=c.length diff.<=0 AND c.value diff.<=0.001			
2=0<c. length d.<=0.162 OR 0.001<c.value d.<=0.24	-0.048**	-0.01	-0.022
P(Fisher)	0.012	0.719	0.252
P(permute)	0.005	0.580	0.225
3= 0.162<c. length diff. OR 0.24<c.value diff.	-0.035*	-0.005	-0.023
P(Fisher)	0.119	0.858	0.192
P(permute)	0.025	0.790	0.185
4= missing (with contr. completion ann.)	0.002	0.015	0
P(Fisher)	0.900	0.457	0.995
P(permute)	0.830	0.195	0.985
5= missing (NO contr. completion ann.)	-0.009	0.011	-0.01
P(Fisher)	0.454	0.416	0.372
P(permute)	0.195	0.190	0.220

constant included in each regression; control variables: product market (cpv divisions); number of winners on the market (market defined by cpv level 4 and nuts 1) year of contract award; log real contract value; contract length; framework contract; issuer type, sector, and status (public or private)

	2	4	5
N	25813	21584	7806
R2/pseudo-R2	0.1272	0.1148	0.2448

Source: PP

Note:* p<0.05; ** p<0.01; *** p<0.001; clustered standard errors clustered by issuer for P(Fisher), Monte Carlo random permutation simulations for P(permute) (200 permutations) using stata 12.0

ANNEX F – DESCRIPTIVE STATISTICS OF CONTROL VARIABLES

Table F1. Descriptive statistics of control variables, 2009-2012, markets with at least 3 unique winners

Variable name	mean	min	max	sd	N
number of unique winners	144.827	3.00	674.00	170.57	47768
log real contract value	16.034	9.04	25.08	1.85	49867
contract length in years	1.178	0.00	35.00	1.68	34684
PPPs/framework contracts	0.114	0.00	1.00	0.32	51823

Source: PP

Table F2. Distribution of issuer type, 2009-2012, markets with at least 3 unique winners

type of issuer	N	%
central organisation	10,057	19.41
public utility	2,252	4.35
regional/local organisation	25,945	50.06
supported body	2,114	4.08
body established by public law	5,282	10.19
other	4,498	8.68
missing	1,675	3.23
total	51,823	100

Source: PP

Table F3. Distribution of issuer main sector, 2009-2012, markets with at least 3 unique winners

main sector of issuer	N	%
general public services	17,922	34.58
education	4,828	9.32
health	1,506	2.91
culture, sports	1,997	3.85
environment, agriculture	1,379	2.66
energy	1,211	2.34
finance, post	1,111	2.14
welfare	1,222	2.36
transport	784	1.51
defence, policing	955	1.84
water, sewage, waste	437	0.84
other	3,553	6.86
missing	14,918	28.79
total	51,823	100

Source: PP

Table F4. Distribution of issuer status, 2009-2012, markets with at least 3 unique winners

status of issuer	N	%
private	15,050	29.04
public	36,398	70.24
missing	375	0.72
total	51,823	100

Source: PP

Table F5. Distribution of contract award year, 2009-2012, markets with at least 3 unique winners

year of contract award	N	%
2009	10,587	20.43
2010	17,487	33.74
2011	13,708	26.45
2012	10,041	19.38
total	51,823	100

Source: PP

Table F6. Distribution of main market of contract, 2009-2012, markets with at least 3 unique winners

main market of contract	N	%
administration, defence and social security services	56	0.11
agricultural machinery	355	0.69
agricultural, farming, fishing, forestry and related products	118	0.23
agricultural, forestry, horticultural, aquacultural and apicultural services	434	0.84
architectural, construction, engineering and inspection services	4,275	8.32
business services: law, marketing, consulting, recruitment, printing and security	4,739	9.22
chemical products	226	0.44
clothing, footwear, luggage articles and accessories	161	0.31
construction structures and materials; auxiliary products to construction	510	0.99
construction work	17,862	34.76
education and training services	1,489	2.9
electrical machinery, apparatus, equipment and consumables; lighting	223	0.43
financial and insurance services	625	1.22
food, beverages, tobacco and related products	4,123	8.02
furniture, furnishings, domestic appliances and cleaning products	1,120	2.18
health and social work services	516	1
hotel, restaurant and retail trade services	717	1.4
industrial machinery	539	1.05
installation services (except software)	54	0.11
it services: consulting, software development, internet and support	980	1.91
laboratory, optical and precision equipment (excl. glasses)	815	1.59
leather and textile fabrics, plastic and rubber materials	31	0.06
machinery for mining, quarrying, construction equipment	74	0.14
medical equipment, pharmaceuticals and personal care products	3,121	6.07
mining, basic metals and related products	13	0.03
musical instruments, sport goods, games, toys, handicraft, art mat. and accessories	357	0.69
office and computing machinery, equipment and supplies ex. furniture and software	1,906	3.71
other community, social and personal services	173	0.34
petroleum products, fuel, electricity and other sources of energy	823	1.6
postal and telecommunications services	114	0.22
printed matter and related products	688	1.34
public utilities	14	0.03
radio, television, communication, telecommunication and related equipment	249	0.48
real estate services	367	0.71
recreational, cultural and sporting services	246	0.48
repair and maintenance services	715	1.39
research and development services and related consultancy services	110	0.21
security, fire-fighting, police and defence equipment	88	0.17
services related to the oil and gas industry	7	0.01
sewage-, refuse-, cleaning-, and environmental services	1,046	2.04
software package and information systems	341	0.66
supporting and auxiliary transport services; travel agencies services	34	0.07
transport equipment and auxiliary products to transportation	775	1.51
transport services (excl. waste transport)	155	0.3
total	51384	100

Source: PP

REFERENCES FOR ANNEXES

Auriol, E., Flochel, T., & Straub, S. (2011). *Public Procurement and Rent-Seeking: The Case of Paraguay* (No. 11-224). TSE Working Papers: 11-224, Toulouse: Toulouse School of Economics (TSE).

Bandiera, O., Prat, A., & Valletti, T. (2009). Active and Passive Waste in Government Spending: Evidence from a Policy Experiment. *American Economic Review*, *99*(4), 1278–1308.

Coviello, D., & Gagliarducci, S. (2010). *Building Political Collusion: Evidence from Procurement Auctions* (No. 4939). IZA DP No. 4939, Bonn: Institute for the Study of Labor (IZA).

Di Tella, R., & Schargrodsky, E. (2003). The Role of Wages and Auditing during a Crackdown on Corruption in the City of Buenos Aires. *Journal of Law and Economics*, *46*(1), 269–292.

Ferraz, C., & Finan, F. (2008). Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes. *The Quarterly Journal of Economics*, *123*(2), 703–745. doi:10.1162/qjec.2008.123.2.703

Golden, M. A., & Picci, L. (2005). Proposal for a New Measure of Corruption, illustrated with Italian data. *Economics & Politics*, *17*(1), 37–75. doi:10.1111/j.1468-0343.2005.00146.x

Goldman, E., Rocholl, J., & So, J. (2013). Politically Connected Boards of Directors and The Allocation of Procurement Contracts. *Review of Finance*, *17*(5), 1617–1648. doi:10.1093/rof/rfs039

Hyytinen, A., Lundberg, S., & Toivanen, O. (2008). *Politics and Procurement: Evidence from Cleaning Contracts* (No. 233). HECER Discussion paper No. 233.

Klasnja, M. (2015). Corruption and the Incumbency Disadvantage: Theory and Evidence. *Journal of Politics*, *forthcomin*.

Olken, B. A. (2006). Corruption and the costs of redistribution: Micro evidence from Indonesia. *Journal of Public Economics*, *90*(4–5), 853–870.

Olken, B. A. (2007). Monitoring Corruption: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy*, *115*(2), 200–249.

Reinikka, R., & Svensson, J. (2004). Local Capture: Evidence From a Central Government Transfer Program in Uganda. *The Quarterly Journal of Economics*, *119*(2), 678–704.