



Government
Transparency
Institute

Mihály Fazekas¹, Elizabeth Dávid-Barrett², Aly Abdou³, Olivier Basdevant⁴

The Corruption Cost Tracker

Quantifying the costs of corrupt contracting and the savings to be made from reform

Government Transparency Institute Reports: GTI-R/2020:02
Policy Brief

Version 1.0

September 2020, Budapest, Hungary

¹ Central European University and Government Transparency Institute

² University of Sussex

³ Government Transparency Institute

⁴ International Monetary Fund

We gratefully acknowledge the financial support of DFID UK Anti-Corruption Evidence Programme and the European Commission's Horizon 2020 research funding for the initial data collection and research work. We are also grateful for the IMF's Anti-Corruption Challenge for funding additional data collection and analytical work, underlying this policy brief. The views expressed in this policy brief are those of the authors and do not necessarily represent the views of the organizations the authors work at.

Challenge: Public Procurement Vulnerable to Corruption, Leading to Massive Losses for Governments

Public procurement constitutes about one-third of government spending or 13 trillion USD per year. It is highly vulnerable to corruption with estimates of losses amounting to 10-20%.¹ Corruption in public procurement can lead to²:

- Overpriced public procurement contracts contributing to larger budget deficits.
- Unfinished, sub-standard delivery or lower than contracted quantity leading to the need for further public expenditure or lost revenue.
- Adverse impact on growth through lower efficiency of public investment, lower quality of public services and higher volatility in markets with substantial public presence.

With Covid-19-related spending largely channeled through procurement systems, tackling corruption has become even more important. However, identifying where corruption takes place and prioritizing impactful anti-corruption policies is notoriously difficult.

Solution: The Corruption Cost Tracker to Deter Corruption and Inform Anti-Corruption Strategies

Our global intelligence tool, the Corruption Cost Tracker (CCT)³, builds an evidence base about where corruption risks lie in public procurement, their costs, and the benefits of reform in terms of savings. The Corruption Cost Tracker is an interactive online tool, with dashboards for Corruption Risk Analysis, Spending Analysis, Efficiency Gains and Policy Scenarios. Each of the dashboards allows users to undertake their own analysis in these areas by comparing sectors, regions, years, and buyer types.

The CCT will assist national authorities, the staff of the IMF and the World Bank, and the relevant civil society organisations in identifying priority areas for governance and anti-corruption reforms. In particular, the CCT can help inform these reforms by providing a granular identification of where public procurement in a given country is the most vulnerable to corruption, and what are the potential gains to be expected from reforms.

Data and methodology

The CCT aims to use data on 50 million contracts from 47 countries⁴ where the project consortium has already advanced with data collection and data processing and - if funds allow - to expand further. The pilot version is developed for five countries – Uganda, Paraguay, Indonesia, Georgia, and Romania –

¹ See for example: https://www.rand.org/pubs/research_reports/RR1483.html

² <https://www.sciencedirect.com/science/article/abs/pii/S0965856417311199>

³ <https://public.tableau.com/profile/gti1940#!/vizhome/Corruptioninpublicprocurement/Overviewofcountries?publish=yes>

⁴ For full country list and details on the datasets see:

https://public.tableau.com/profile/mihaly.fazekas#!/vizhome/GTIDataScope/GTI_DataScope_national



allowing users to test the methodology and explore the relevance of the proposed indicators in a diverse set of contexts (for details on data see Annex A). We exclusively use publicly available, official government data. Since such data sources are often poorly structured, we also quality check the datasets and standardize them so that they can reliably be used for large-scale data analysis. The pilot dataset for 5 countries includes over 1.5 million contracts capturing 15-55% of total procurable expenses depending on the country.

The CCT dashboard builds on corruption risk indicators and corruption cost estimates calculated using well-established methods.⁵ The corruption risk indicators that we developed proxy corruption by identifying high-risk situations where open and fair competition has been curtailed in order to benefit a favoured firm. For example, when only one firm submits a bid on an otherwise competitive market and the bid advertisement period was only 1 working day, the chances are higher that tendering decisions were driven by corruption. We kept risk indicators as comparable as possible across countries to assure international comparability, while some deviations remain due to specificities of the local context (e.g. too few foreign suppliers in Paraguay, Indonesia, and Uganda for the tax haven indicator to be statistically meaningful) (Table 1).

TABLE 1: CRI COMPONENTS BY COUNTRY

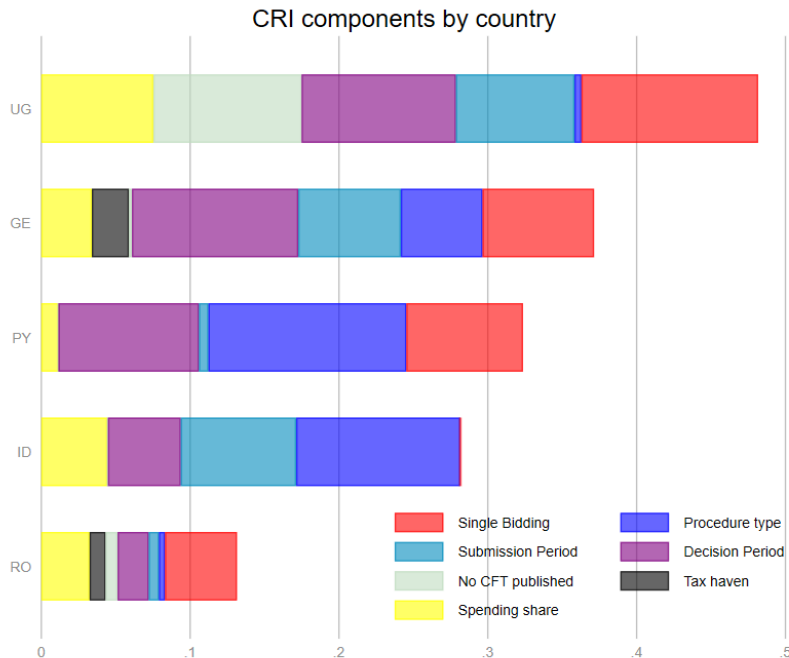
Indicator group	Red flag	GE	PY	ID	RO	UG
Tendering risk	Non-open procedure type	x	x	x	x	x
Tendering risk	Lack of call for tender publication	x	x	x	x	x
Tendering risk	Short bid submission period	x	x	x	x	x
Tendering risk	Length of decision period	x	x	x	x	x
Tendering risk	Single bidder contract	x	x	x	x	x
Supplier risk	Supplier registered in tax haven	x			x	
Supplier risk	Spending concentration (by organization, by year)	x	x	x	x	x

We carried out a series of econometric tests identifying the best parameters for our indicators (e.g. how many days would count as a very short advertisement period in different contexts) and also validating them. All these indicators are also confirmed by proven cases and economic theories of crime. In order to use a robust risk indicator, we aggregate 7 red flags into a composite score by simply averaging them (where 0 is lowest corruption risk and 1 highest); we call this the Corruption Risk Index (Figure 1). For a detailed description, including precise indicator definitions and validity regression see Annex B.

⁵ See for example: <https://doi.org/10.1017/S0007123417000461>



FIGURE 1. CORRUPTION RISK INDEX COMPONENTS FOR EACH COUNTRY



Our methodology also links corruption risks to spending based on econometric modelling which estimates the price sensitivity of awarded contracts to corruption risks (for details see Annex C). We predict the size of discounts offered by the winning firm compared to the auction reference price (that is typically the maximum budgetary allocation for a given purchase) based on corruption risks while controlling for year, contract value, main market, buyer location, and buyer type on the contract level. Finally, these models allow us to bridge our large-scale micro-level dataset with macro aggregates such as budget deficit and to offer different macro spending estimates based on different risk levels in each country and sector.

How to use the Corruption Cost Tracker

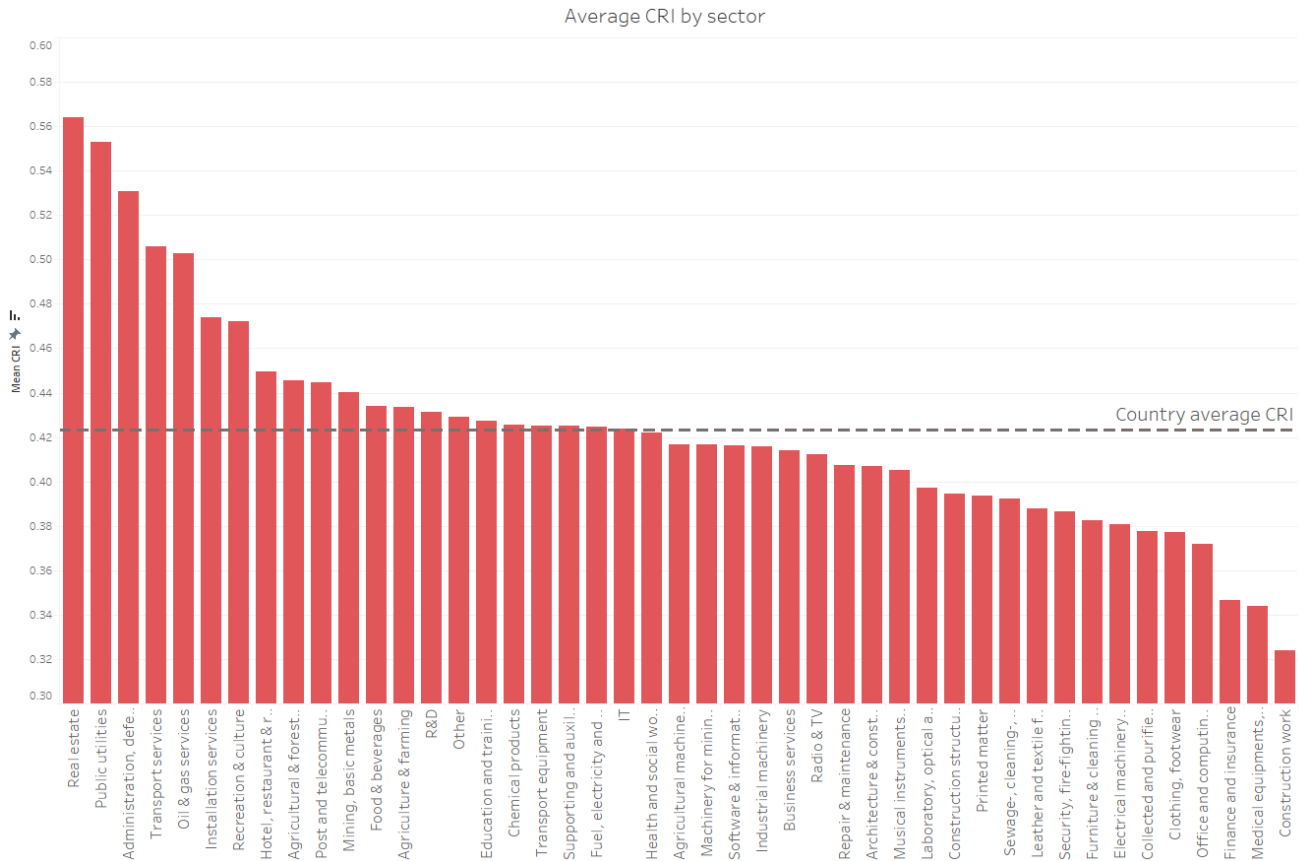
1. Identify corruption risks in public procurement on the meso-level

The CCT⁶ can be used to compare the levels of corruption risk within countries by sector, region, year, and public organisation type (e.g. municipal administrations). Figure 2 shows the mean Corruption Risk Index (0 is the lowest risk and 1 the highest) in Georgia by main sectors. We can see that the highest risk sectors are real estate, public utilities, and public administration and defence. On the other end of the spectrum, medical equipment and pharmaceuticals is one of the lowest risk sectors.

⁶ <https://public.tableau.com/profile/gti1940#!/vizhome/Corruptioninpublicprocurement/Overviewofcountries?publish=yes>



FIGURE 2. AVERAGE CORRUPTION RISK INDEX BY SECTOR, GEORGIA, 2010-2019

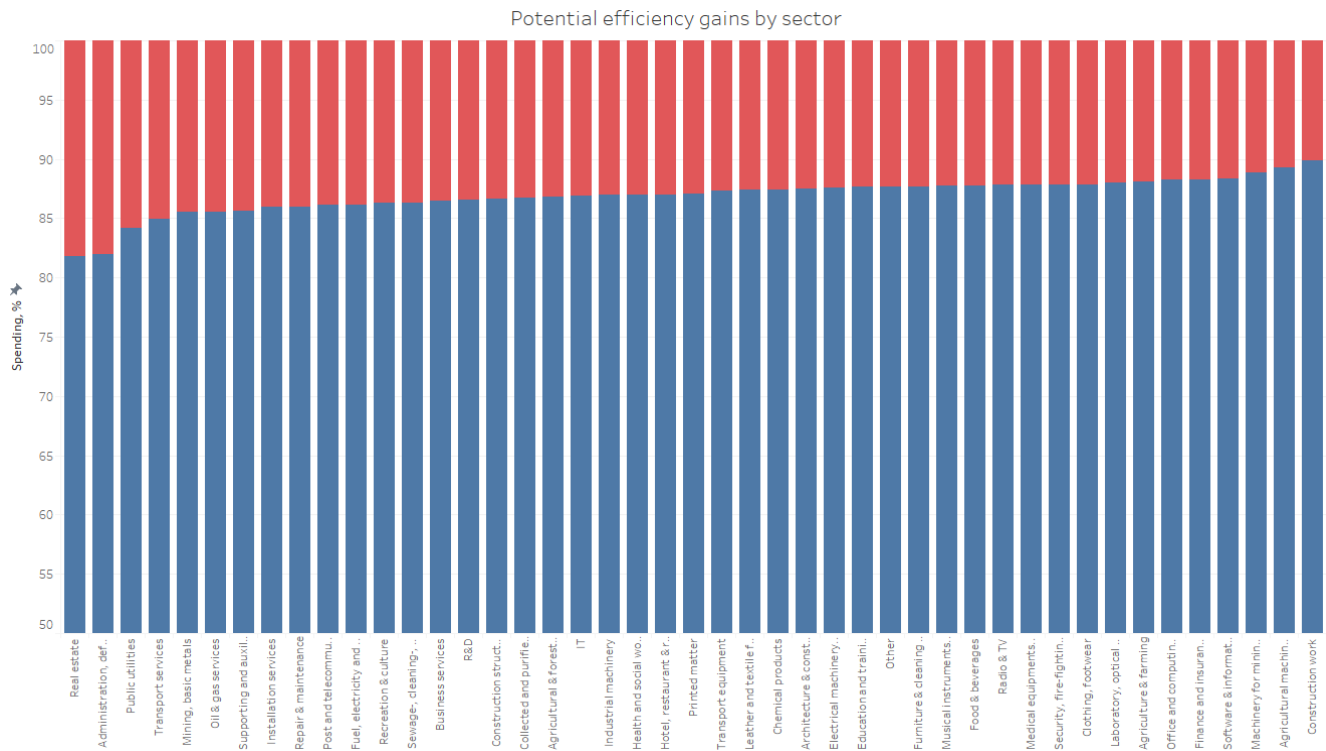


2. Tracking the financial impacts of corruption risks on the meso-level

To help policy dialogue in deciding which areas to focus on, the CCT enables comparisons of the potential efficiency gains to be made from reducing corruption by sector, region, year, and public organisation type. These estimates are based on our price modelling results which offer predictions of public procurement spending according to corruption risk levels. Figure 3 allows us to rank sectors in a country by the percentage of potential efficiency gains to be made. According to this model, some of the largest savings can be achieved in, for example, public utilities or transport services.



FIGURE 3. AVERAGE EFFICIENCY GAIN BY ELIMINATING CORRUPTION, BY SECTOR, GEORGIA, 2010-2019



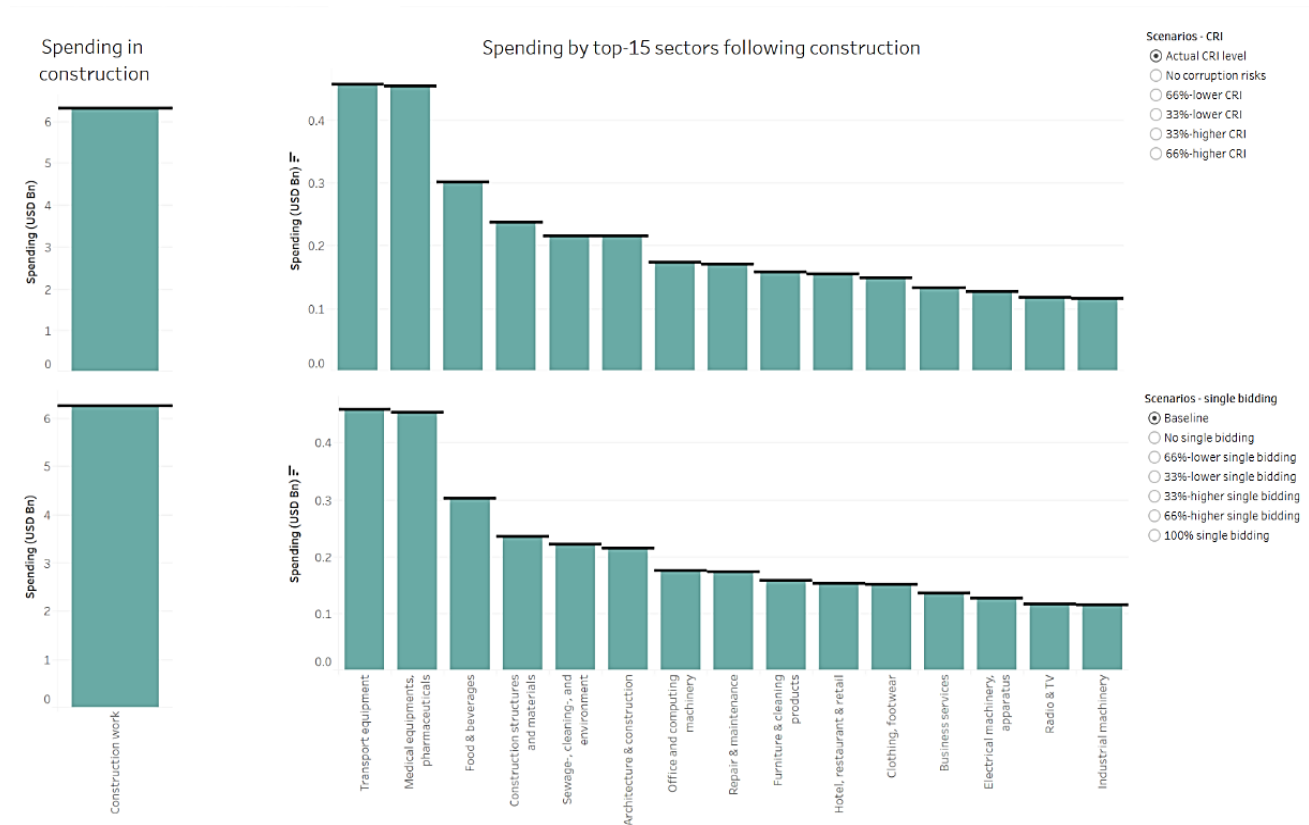
3. Evaluating Specific Procurement Reforms on Risks and Savings

The spending implications of different corruption risk levels can be further probed through different policy scenarios reflecting varying levels and types of corruption risks and the associated public procurement spending. Based on the results of overpricing models, we can make predictions about how sectoral procurement spending would vary if the level of Corruption Risk Index changed or the prevalence of a single risk factor moved.

Figure 4 shows how the tool can be used to evaluate different reforms. Here, we show estimates for the sectoral spending impact of reducing the composite Corruption Risk Index by one-third, two-thirds, etc. or reducing single bidding by one-third, etc. For example, users could compare the financial benefits of reducing the incidence of single bidding which is the strongest predictor of prices in Georgia. Further, actionable risk factors and their price impacts can be added to the dashboard such as online advertisement of tenders, choice of procedure type or awarding contracts to tax haven registered companies. Such scenarios could inform policy decisions about how best to reduce budget deficits without compromising public service quality.



FIGURE 4. SAVINGS SCENARIOS BY SECTOR ACCORDING TO DIFFERENT CORRUPTION RISK INDEX AND SINGLE BIDDING LEVELS, GEORGIA, 2010-2019





Technical Annexes

Annex A. Data Description

Section A.1 Technical description

We developed an automated web crawler to scrape data from each of the sources presented in Table A.1. In brief, our methodology is composed of the following steps: we crawl a collection of HTML, XML, and CSV outputs from the sources. We then structure each publication from its original format to a uniform structured data template⁷. Next, the data is formatted through the conversion of structured text to standard data types (numbers, dates, enumeration values) including cleaning nonsensical values or ballast information. We then link all the information which describes the same tender, where a tender ideally begins with one Call for Tenders (or more) followed by one Contract Award (or more) and completed by a series of payments or Contract Completion Announcement. We also take into account if any modifications or cancellations occur to the tender at any point during the process. After successfully linking related publications, we then reconcile all linked data records to create a single best image of a public tender covering its whole tendering cycle (importantly, this is the step where we reconcile conflicting information or fill in empty fields if available in a related notice).

The data is then subject to an extensive validation process through manually cross-checking the records in the database with the source publications. After successfully validating the datasets, we subject the data to a second round of cleaning, and standardize buyer and supplier names. Whenever needed, we used HERE REST APIs⁸ to generate missing locations for buyers for example in Georgia. For Indonesia and Uganda, we also implemented a multi step token-based string matching algorithm for observations with missing tender product codes. We used a combination of tender title, lot title, and/or product description to match them with CPV-2008 product codes⁹. For full technical documentation and codes see: <https://github.com/digiwhist/backend>.

⁷ <https://docs.google.com/spreadsheets/d/13pGIpt47sMBnZ68E-N-hMLiErpDB1CQwZzd2MXllq5U/edit#gid=623190471>

⁸ <https://developer.here.com/develop/rest-apis>

⁹ <https://simap.ted.europa.eu/cpv>



TABLE A.2: DATA DESCRIPTION

		Georgia	Romania	Indonesia	Paraguay	Uganda
Number of observations		202,299	620,261	682,070	142,878	47,641
Year		2011-2019	2007-2020	2012-2018	2010-2020	2016-2020
Nr. of buyer		2,833	9,710	4,146	434	190
Nr. of suppliers		18,203	47,533	93,292	13,277	10,810
Log (Contract Value)	Mean	9.544743	10.0351	20.25113	17.83497	15.22688
	Standard Deviation	1.702274	2.650291	1.210582	1.883986	2.012175
	Min	5.774551	3.020947	14.25336	8.569976	0
	Max	16.11579	21.96522	28.43628	28.00678	24.72635
	Missing Rate	0.94%	12.10 %	1.14 %	6.16 %	0.82 %
Relative Price	Mean	0.8758014	0.7181427	0.931469	8.678997	8527.829
	Standard Deviation	0.1510052	0.4570516	3.117477	175.063	1668938
	Min	0.0063097	3.01e-06	0.0001671	0	6.80e-13
	Max	1.587389	5.601896	995.8755	14905.97	3.58e+08
	Missing Rate	3.69 %	43.73 %	1.17 %	70.87%	3.21 %
Buyer types	PUBLIC_BODY	39.19 %	9.64 %			1.22 %
	REGIONAL_AUTHORITY	27.77 %	2.69 %	34.13 %		0.12 %
	NATIONAL_AUTHORITY	15.03 %	11.00 %	9.83 %		57.77 %
	OTHER	18.01 %	16.14 %	6.79 %	7.01 %	0.86 %
	EUROPEAN_AGENCY	0.00 %	0.00 %			
	NATIONAL_AGENCY		0.29 %			14.51 %
	REGIONAL_AGENCY		0.40 %			
	UTILITIES		1.59 %			
	ARMED_FORCES			0.04 %		0.99 %
	INDEPENDENT_AGENCY			6.13 %		3.88 %
	LOCAL_BODY			41.83 %		10.17 %
	FEDERAL GOVERNMENT				21.25 %	
	FEDERAL_BODY				18.23 %	
	STATE_GOVERNMENT				5.13 %	10.47 %
	STATE FACILITY					
	MUNICIPAL_GOVERNMENT				18.30 %	
	UNIVERSITY				10.05 %	
	HOSPITAL				0.81 %	
	JUSTICE				6.48 %	
	NATIONAL_FUNDS & BANK				2.80 %	
	Missing rate (%)	0.00 %	58.25%	1.25 %	9.95 %	0 %
Markets	Levels	45	45	43	44	24
	Missing rate	0.05%	2.95 %	14.10 %	5.37 %	64.77 %
Contract Type	SUPPLIES		66.93%	18.50 %		
	SERVICES		18.43 %	24.55 %		
	WORKS		13.20 %	56.94 %		
	Missing rate (%)		1.44 %	0.00 %		
Corruption Risk Index (CRI)	Mean	0.3852011	0.1970716	0.296058	0.2715395	0.4723657
	Standard Deviation	0.1706613	0.2150351	0.12833	0.1577541	0.2018105
	10 th percentile	0.1668591	0	0.1	0.0904423	0.2
	90 th percentile	0.6	0.5224221	0.5	0.4305008	0.7058015
Single Bidding	% single bidding = 0	48.99 %	68.33 %	99.92 %	41.16 %	31.41 %



	% single bidding = 1	51.01 %	31.50 %	0.08 %	19.67 %	68.57 %
	% single bidding = Missing	0.00 %	0.17 %	0.00 %	39.17 %	0.02 %
Procedure Type (red flag = corr_proc)	% corr_proc = 0	23.62 %	83.70 %	10.33 %	19.15 %	96.23 %
	% corr_proc = 1	75.41 %	16.30 %	47.55 %	19.81 %	1.77 %
	% corr_proc = 2	0.96 %	-	41.88 %	60.79 %	1.99 %
	% corr_proc = Missing	0.00 %	-	0.24 %	0.25 %	0.00 %
Submission period (red flag = corr_submp)	Mean	11.25	61.59122	11.82429	42.51561	48.18659
	Standard Deviation	7.11	68.9772	9.875408	36.14713	44.30766
	Min.	2	2	1	1	1
	Max.	88	497	183	183	183
	% corr_submp = 0	38.87 %	26.44 %	27.41 %	81.23 %	13.17 %
	% corr_submp = 1	23.92 %	6.18 %	51.18 %	3.29 %	9.87 %
	% corr_submp = 2	35.80 %	-	20.70 %	4.49 %	10.05 %
	% corr_submp = Missing	1.41 %	67.38 %	0.71 %	10.99 %	66.90 %
Decision Period (red flag = corr_decp)	Mean	19.01334	154.9743	12.22504	46.41831	7.520236
	Standard Deviation	13.16	162.1121	9.91122	35.01085	12.4544
	Min.	0	0	1	0	0
	Max.	183	1,705	360	183	183
	% corr_decp = 0	37.02 %	25.84 %	54.59 %	21.37 %	17.04 %
	% corr_decp = 1	62.83 %	3.83 %	30.87 %	43.48 %	29.59 %
	% corr_decp = 2	-	1.35 %	14.06 %	26.31 %	-
	% corr_decp = Missing	0.15 %	68.97 %	0.48 %	8.85 %	53.37 %
No CFT	% nocft = 0	98.59 %	75.86 %	99.50 %	93.39 %	45.41 %
	% nocft = 1	1.41 %	24.14 %	0.50 %	6.61 %	54.59 %
Tax haven	Foreign Supplier not in a tax haven	0.19 %	0.60 %			
	Foreign Supplier in a tax haven	0.04 %	0.03 %			
	Local Supplier	99.77 %	99.36 %			
Contract Share (w_ycsh/proa_ycsh)	Mean	0.3869508	0.3451785	0.6691212	.0436906	0.6236105
	Standard Deviation	0.3888428	0.3980906	0.3633201	0.1055271	0.417391
	10 th Percentile	0.009532	0.000582	0.1144647	0.0002355	0.0171891
	90 th Percentile	1	1	1	0.1179277	1
	Missing rate (%)	0.11 %	0.624 %	1.78 %	6.92 %	0.39 %



Annex B. The Corruption Risk Index

This appendix describes the Corruption Risk Index (CRI), each of its components (Section B.1) and also the micro-level regression based on which their validity and parameters are established (Section B.2).

Section B.1 CRI definitions

TABLE B.1: CRI COMPONENTS BY COUNTRY

Indicator group	Red flag	GE	PY	ID	RO	UG
Tendering risk	Non-open procedure type	x	x	x	x	x
Tendering risk	Lack of call for tender publication	x	x	x	x	x
Tendering risk	Short bid submission period	x	x	x	x	x
Tendering risk	Length of decision period	x	x	x	x	x
Tendering risk	Single bidder contract	x	x	x	x	x
Supplier risk	Supplier registered in tax haven	x			x	
Supplier risk	Spending concentration (by organisation, by year)	x	x	x	x	x

TABLE B.2: CRI BY COUNTRY

CRI	Mean	Standard deviation	10 th Percentile	90 th Percentile
Uganda	0.47	0.20	0.20	0.71
Georgia	0.39	0.17	0.17	0.60
Indonesia	0.30	0.13	0.10	0.50
Paraguay	0.27	0.16	0.09	0.43
Romania	0.19	0.22	0	0.52

FIGURE B.1: MEAN CRI AND STANDARD DEVIATION BY COUNTRY

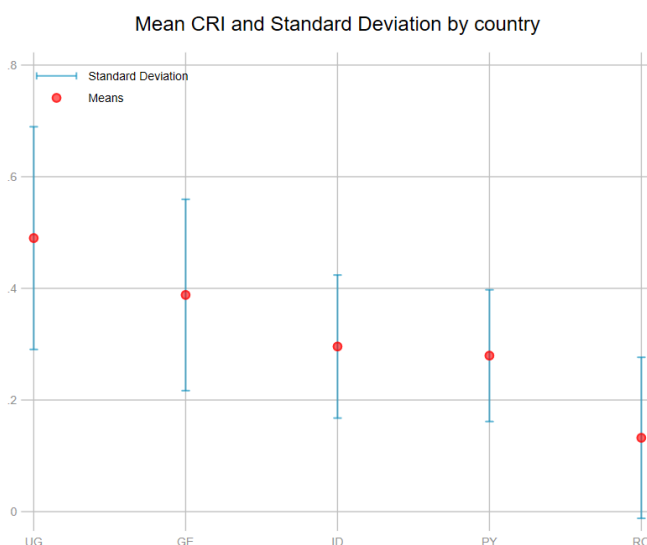




TABLE B.2.: NO CALL FOR TENDER IS PUBLISHED

Country	is_red flag
GE	YES
UG	YES
RO	YES
PY	YES
ID	YES

TABLE B.3: SUBMISSION PERIOD THRESHOLD RED FLAGS BY COUNTRY

Country	Red flag level 1	Red flag level 2	Not a red flag	Is missing
GE	Less than 13 days	Less than 6 days	More than 13 days	Yes
UG	17 to 41 days	Less than 17 days	More than 41 days	Yes
RO	30 to 33 days if procedure type is open or negotiated with publication & 9 to 14, and 65 to 378 days for the rest	-	Less than 30 days and more than 33 days for open and negotiated with publication procedure types and Less than 9 and more than 14 for the rest.	Yes
PY	13 to 30 days	Less than 13 days or 31 to 47 days	16 to 20 days or more than 47 days	Yes
ID	8 to 14 days	0 to 7 days	More than 14 days	Yes

TABLE B.4: DECISION PERIOD THRESHOLD RED FLAG BY COUNTRY.

Country	Red flag level 1	Red flag level 2	Not a red flag	Is missing
GE	Less than 14 days or more than 25 days	-	14 to 25 days	Yes
UG	1 day or more than 14 days	-	2 to 14 days	Yes
RO	33 to 53 days	Less than 32 days	More than 50 days	Yes
PY	23 to 64 days	0 to 22 days	More than 64 days	Yes
ID	5 to 11 days or more than 25 days	Less than 4 days	11 to 25 days	Yes

TABLE B.5.: NON-OPEN PROCEDURE TYPE RED FLAG BY COUNTRY

Country	Red flag level 1	Red flag level 2	Not a red flag	Is missing
GE	1. Electronic Tender (SPA) 2. Electronic Tender (SPA) via price list 3. Simplified Electronic Tender (SPA) 4. Simplified Electronic Tender (SPA) via price list.	1. e-Procurement Procedure (GEO) 2. e-Procurement Procedure (GEO) via price list	1. Donor electronic procurement procedure (DEP) 2. Electronic Tender (DAP) 3. Electronic Tender Without Reverse Auction (NAT)	Yes



			4. Electronic Tender Without Reverse Auction (NAT) via price list 5. Simplified Electronic Tender Without Reverse Auction (NAT) 6. Simplified Electronic Tender Without Reverse Auction (NAT) via price list 7. Simplified Electronic Tender (DAP) 8. Simplified Two Stage Electronic Tender (MEP) 9. Two Stage Electronic Tender (MEP) 10. Two Stage Electronic Tender (MEP) via price list	
UG	1. Approaching Bidders	1. Restricted	1. Open 2. Negotiated 3. Negotiated without publication 4. Negotiated with publication	Yes
RO	1. Negotiated 2. Negotiated without publication	-	1. Open 2. Approaching bidders 3. Competitive dialog 4. Negotiated with publication 5. Restricted	No
PY	1. Open within threshold	1. Direct contracting 2. Other	1. Open auction 2. Limited	Yes
ID	1. e-Lelang Sederhana, e-Lelang Umum 2. e-Seleksi Umum 3. Lelang Sederhana - Pascakualifikasi Satu File - Harga Terendah Sistem Gugur 4. Lelang Sederhana - Prakuifikasi Dua File - Kualitas dan Biaya 5. Lelang Sederhana - Prakuifikasi Dua File - Sistem Nilai 6. Lelang Sederhana - Prakuifikasi Satu File - Biaya Terendah 7. Lelang Umum - Pascakualifikasi Dua File - Sistem Nilai 8. Lelang Umum - Pascakualifikasi Dua File - Sistem Umum Ekonomis, Lelang Umum - Pascakualifikasi Satu File - Harga Terendah Sistem Gugur 9. Lelang Umum - Prakuifikasi Dua File - Kualitas, Lelang Umum - Prakuifikasi Dua File - Kualitas dan Biaya 10. Lelang Umum - Prakuifikasi Dua File - Sistem Nilai 11. Lelang Umum - Prakuifikasi Dua Tahap - Harga Terendah Sistem Gugur 12. Lelang Umum - Prakuifikasi Dua Tahap - Sistem Nilai 13. Lelang Umum - Prakuifikasi Satu File - Biaya Terendah 14. Lelang Umum - Prakuifikasi Satu File - Harga Terendah Sistem Gugur 15. Lelang Umum - Prakuifikasi Satu File - Pagu Anggaran 16. Seleksi Umum - Pascakualifikasi Satu File - Harga Terendah Sistem Gugur 17. Seleksi Umum - Pascakualifikasi Satu File - Kualitas 18. Seleksi Umum - Prakuifikasi Dua File - Kualitas	1. e-Lelang Pemilihan Langsung 2. e-Penunjukan Langsung 3. e-Seleksi Langsung 4. Lelang Pemilihan Langsung - Pascakualifikasi Satu File - Harga Terendah Sistem Gugur	1. e-Lelang Terbatas 2. e-Seleksi Sederhana 3. Lelang Terbatas - Pascakualifikasi Satu File - Harga Terendah Sistem Gugur 4. Lelang Terbatas - Prakuifikasi Dua File - Sistem Nilai 5. Lelang Terbatas - Prakuifikasi Dua Tahap - Harga Terendah Sistem Gugur 6. Lelang Terbatas - Prakuifikasi Satu File - Harga Terendah Sistem Gugur 7. Seleksi Sederhana - Pascakualifikasi Satu File - Biaya Terendah 8. Seleksi Sederhana - Pascakualifikasi Satu File - Harga Terendah Sistem Gugur 9. Seleksi Sederhana - Pascakualifikasi Satu File - Kualitas 10. Seleksi Sederhana - Pascakualifikasi Satu File - Pagu Anggaran, Seleksi Sederhana - Prakuifikasi Dua File - Kualitas 11. Seleksi Sederhana - Prakuifikasi Satu File - Biaya Terendah 12. Seleksi Sederhana - Prakuifikasi Satu File - Pagu Anggaran	Yes



	19. Seleksi Umum - Prakuafifikasi Dua File - Kualitas dan Biaya 20. Seleksi Umum - Prakuafifikasi Dua File - Pagu Anggaran 21. Seleksi Umum - Prakuafifikasi Dua File - Sistem Nilai 22. Seleksi Umum - Prakuafifikasi Satu File - Biaya Terendah 23. Seleksi Umum - Prakuafifikasi Satu File - Pagu Anggaran			
--	---	--	--	--

Section B.2 CRI validation regressions

In this section we present the results of our red flag validity regressions in each country. We use two regressions to validate each of our CRI components. In the first set of regressions (Tables B.6, B.8, B.10, B.12, B.14), we run a logit model where single bidding is the dependent variable and the rest of the CRI components as the independent variables. A positive significant result supports the validity of each CRI component, following the methodology outlined in prior academic literature¹⁰. The regressions control for contract values, markets (based on assigned product codes), buyer types, and tender year.

$$\text{logit } P(\text{single bidding}) = \alpha + \sum \beta_i X_i + \sum \beta_j \text{ controls},$$

where X_i includes the list of available red flags

Secondly, to validate the CRI components in relation to supplier contract share, we run an OLS regression (Tables B.7, B.9, B.11, B.13, B.15) where supplier contract share is the main dependent variable and the remaining CRI components as the independent variables. Similarly, a positive and significant coefficient offers evidence for the validity of the CRI component in question. The regressions are limited to suppliers with more than 4 contracts per year. The regressions controls for contract values, markets, buyer types, and tender year.

$$\text{Contract Share} = \alpha + \sum \beta_i X_i + \sum \beta_j \text{ controls},$$

where X_i includes the list of available red flags

¹⁰ e.g. Fazekas-Kocsis (2020): Uncovering High-Level Corruption: Cross-National Objective Corruption Risk Indicators Using Public Procurement Data. British Journal of Political Science, Volume 50, Issue 1, pp. 155-164 <https://doi.org/10.1017/S0007123417000461>



TABLE B.6: GEORGIA - DEPENDENT VARIABLE: SINGLE BIDDING

CRI components	Component categories	Coefficient (Std. error)
Procedure type (base: open procedure)	Non-open procedure (type 1)	0.128*** (0.0159)
	Non-open procedure (type 2)	0.542*** (0.0516)
	Missing procedure type	1.442* (0.829)
Submission period (base: more than 13 days)	Less than 13 days	0.109*** (0.0153)
	Less than 6 days	0.145*** (0.0121)
	Missing submission period	-0.543*** (0.0441)
Decision period (base: more than 13 days)	Less than 14 days or more than 25 days	0.0709*** (0.00972)
	Missing decision period	0.407*** (0.124)
Tax haven (base: foreign supplier not located in a tax haven)	Foreign supplier in tax haven	-0.0408 (0.255)
	Local supplier	-0.831*** (0.110)
Observations		200,403
Pseudo-R ²		0.051

Regression includes controls for contract values, buyer type, market, and tender year.

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1



TABLE B.7: GEORGIA - DEPENDENT VARIABLE: SUPPLIER CONTRACT SHARE (> 4 CONTRACTS PER YEAR)

CRI components	Component categories	Coefficient (Std. error)
Single bidding (base: singleb=0)	Single bidding = 1	0.0693*** (0.00143)
Procedure type (base: open procedure)	Non-open procedure (type 1)	0.00781*** (0.00240)
	Non-open procedure (type 2)	0.237*** (0.00857)
	Missing procedure type	0.104 (0.184)
Submission period (base: more than 13 days)	Less than 13 days	-0.0558*** (0.00232)
	Less than 6 days	-0.0621*** (0.00182)
	Missing submission period	-0.0336*** (0.00618)
Decision period (base: more than 13 days)	Less than 14 days or more than 25 days	0.00791*** (0.00146)
	Missing decision period	0.00479 (0.0200)
Tax haven (base: foreign supplier not located in a tax haven)	Foreign supplier in tax haven	-0.688*** (0.0874)
	Local supplier	-0.716*** (0.0389)
Observations		139,891
R ²		0.232

Regression includes controls for contract values, buyer type, market, and tender year.

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1



TABLE B.8: PARAGUAY - DEPENDENT VARIABLE: SINGLE BIDDING

CRI components	Component categories	Coefficient (Std. error)
Procedure type (base: open and other procedure)	Non-open procedure (out-right award)	0.952*** (0.0367)
	Non-open procedure (restricted and missing)	1.373*** (0.0398)
	Missing procedure type	1.815*** (0.184)
Submission period (base: 16 to 20 days or more than 47 days)	Category 1: 13 to 30 days	0.138** (0.0597)
	Category 2: Less than 13 days or 31 to 47 days	0.406*** (0.0502)
	Missing submission period	-0.0853** (0.0356)
Decision period (base: more than 64 days)	Category 1: 23 to 64 days	0.469*** (0.0230)
	Category 2: 0 to 22 days	0.875*** (0.0259)
	Missing decision period	-0.254*** (0.0650)
Call for tender (base: published)	Call for tender not published	1.453*** (0.0920)
Observations		80,643
Pseudo-R ²		0.129

Regression includes controls for contract values, buyer type, market, and tender year.

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1



TABLE B.9: PARAGUAY DEPENDENT VARIABLE: PROCUREMENT AUTHORITY CONTRACT SHARE (> 4 CONTRACTS PER YEAR)

CRI components	Component categories	Coefficient (Std. error)
Single bidding (base: singleb=0)	Single bidding = 1	0.00609*** (0.000723)
Procedure type (base: open and other procedure)	Non-open procedure (out-right award)	0.00795*** (0.00131)
	Non-open procedure (restricted and missing)	0.00796*** (0.00138)
	Missing procedure type	0.00334 (0.00772)
Submission period (base: 16 to 20 days or more than 47 days)	Category 1: 13 to 30 days	0.00690*** (0.00202)
	Category 2: Less than 13 days or 31 to 47 days	-0.00574*** (0.00174)
	Missing submission period	0.00286** (0.00132)
Decision period (base: more than 64 days)	Category 1: 23 to 64 days	0.00866*** (0.000805)
	Category 2: 0 to 22 days	0.0216*** (0.000974)
	Missing decision period	-0.00591*** (0.00210)
Call for tender (base: published)	Call for tender not published	0.00658 (0.0135)
Observations		75,762
R ²		0.246

Regression includes controls for contract values, buyer type, market, and tender year.

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1



TABLE B.10: UGANDA - DEPENDENT VARIABLE: SINGLE BIDDING

CRI components	Component categories	Coefficient (Std. error)
Procedure type (base: open procedure)	Non-open procedure (type 1)	0.316*** (0.114)
	Non-open procedure (type 2)	1.662*** (0.100)
	Missing procedure type	
Submission period (base: more than 30 days)	Category 1: 14 to 30 days	1.465*** (0.0540)
	Category 2: Less than 13 days	1.785*** (0.0536)
	Missing submission period	-1.868*** (0.0695)
Decision period (base: 1 to 15 days)	Category 1: 1 day or more than 15 days	0.275*** (0.0416)
	Missing decision period	0.835*** (0.0869)
Call for tender (base: published)	Call for tender not published	4.226*** (0.102)
Observations		46,669
Pseudo-R ²		0.4658

Regression includes controls for contract values, buyer type, market, and tender year.

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1



TABLE B.11.: UGANDA DEPENDENT VARIABLE: SUPPLIER CONTRACT SHARE (> 4 CONTRACTS PER YEAR)

CRI components	Component categories	Coefficient (Std. error)
		-0.0524*** (0.00704)
Single bidding (base: singleb=0)	Single bidding = 1	
Procedure type (base: open procedure)	Non-open procedure (type 1)	0.0620*** (0.0220)
	Non-open procedure (type 2)	0.183*** (0.0335)
	Missing procedure type	
Submission period (base: more than 30 days)	Category 1: 14 to 30 days	0.0476*** (0.0114)
	Category 2: Less than 13 days	0.115*** (0.0109)
	Missing submission period	0.0593*** (0.00968)
Decision period (base: 1 to 15 days)	Category 1: 1 day or more than 15 days	-0.0656*** (0.00751)
	Missing decision period	-0.222*** (0.0160)
Call for tender (base: published)	Call for tender not published	0.184*** (0.0166)
Observations		28,994
R ²		0.083

Regression includes controls for contract values, buyer type, market, and tender year.

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1



TABLE B.12.: ROMANIA - DEPENDENT VARIABLE: SINGLE BIDDING

CRI components	Component categories	Coefficient (Std. error)
Procedure type (base: open procedure)	Non-open procedure (other and outright award)	1.081*** (0.0179)
Submission period (base: more than 33 days)	Less than 33 days	0.205*** (0.0167)
Decision period (base: more than 50 days)	From 32 to 53 days	0.344*** (0.0181)
	Less than 32 days	0.8402*** (0.0265)
	Missing decision period	-0.6145*** (0.0182)
Call for tender (base: published)	Call for tender not published	0.3943*** (0.0183)
Tax haven (base: foreign supplier not located in a tax haven)	Foreign supplier in tax haven	-0.6345*** (0.1571)
	Local supplier	-0.7660*** (0.0391)
Observations		544,419
Pseudo-R ²		0.1151

Regression includes controls for contract values, buyer type, contract type, market, and tender year.

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1



TABLE B.13.: ROMANIA - DEPENDENT VARIABLE: SUPPLIER CONTRACT SHARE (> 4 CONTRACTS PER YEAR)

CRI components	Component categories	Coefficient (Std. error)
Single bidding (base: singleb=0)	Single bidding = 1	0.0141*** (0.00099)
Procedure type (base: open procedure)	Non-open procedure (other and outright award)	-0.004* (0.00243)
Submission period (base: more than 33 days)	Less than 33 days	-0.0174*** (0.0020)
Decision period (base: more than 50 days)	From 32 to 50 days	-0.0141*** (0.0021)
	Less than 32 days	0.0259*** (0.00333)
	Missing decision period	0.2322*** (0.00227)
Call for tender (base: published)	Call for tender not published	0.0045* (0.00246)
Observations		394,770
R ²		0.319

Regression includes controls for contract values, buyer type, market, contract type and tender year.

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1



TABLE B.14:. INDONESIA - DEPENDENT VARIABLE: SINGLE BIDDING

CRI components	Component categories	Coefficient (Std. error)
Procedure type (base: open procedure)	Non-open procedure (type 1)	0.418** (0.191)
	Non-open procedure (type 2)	0.582** (0.290)
	Missing procedure type	-
Submission period (base: more than 14 days)	Category 1: 8 to 14 days	0.865*** (0.168)
	Category 2: 0 to 7 days	1.490*** (0.170)
	Missing submission period	-11.52 (1,027)
Decision period (base: 11 to 25 days)	Category 1: 5 to 11 days or more than 25 days	0.855*** (0.124)
	Category 2: Less than 4 days	1.537*** (0.126)
	Missing decision period	1.330 (1.031)
Call for tender (base: published)	Call for tender not published	12.61 (1,027)
Observations		647,401
Pseudo-R ²		0.0963

Regression includes controls for contract values, buyer type, contract type, market, and tender year.

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1



TABLE B.15: INDONESIA - DEPENDENT VARIABLE: SUPPLIER CONTRACT SHARE (> 4 CONTRACTS PER YEAR)

CRI components	Component categories	Coefficient (Std. error)
Single bidding (base: singleb=0)	Single bidding = 1	0.122*** (0.0217)
Procedure type (base: open procedure)	Non-open procedure (type 1)	0.0678*** (0.00388)
	Non-open procedure (type 2)	0.173*** (0.00472)
	Missing procedure type	0.0841*** (0.0238)
Submission period (base: more than 14 days)	Category 1: 8 to 14 days	0.0141*** (0.00168)
	Category 2: 0 to 7 days	0.0128*** (0.00211)
	Missing submission period	0.00405 (0.0136)
Decision period (base: 11 to 25 days)	Category 1: 5 to 11 days or more than 25 days	-0.0206*** (0.00144)
	Category 2: Less than 4 days	-0.0117*** (0.00189)
	Missing decision period	0.0342*** (0.0126)
Call for tender (base: published)	Call for tender not published	0.109*** (0.0189)
Observations		251,986
R ²		0.186

Regression includes controls for contract values, buyer type, contract type, market, and tender year.
Standard errors in parentheses*** p<0.01, ** p<0.05, * p<0.1



Annex C. Price regressions

This section presents the impact of CRI and specific red flags on relative prices following prior academic literature¹¹. Relative prices are calculated as actual contract values divided by estimated contract values (or through savings, if available directly in the dataset). Two main tables are presented for each country.

First, we present the results from using CRI in different models. Model 1 has CRI as the only independent variable. Model 2 further includes controls for contract value, year, market, contract type (if available). Model 3 additionally controls for buyer characteristics such as buyer type, and location. Relative price is restricted to be between 0.5 and 1.5 in Model 1-3. In model 4 and 5, relative price is further restricted to be between 0.5 and 1. Finally, model 5 allows for a quadratic specification for CRI to capture non-linearities in the data. Model 4 is chosen as the main prediction model for all countries as it presents the highest impact and explanatory power.

$$\text{Relative Price} = \alpha + \beta \text{CRI} + \sum \beta_j \text{controls}, \text{ for Model 1-4 (Tables C.1, C.3, C.5, C.7, C.9)}$$

$$\text{Relative Price} = \alpha + \beta_1 \text{CRI} + \beta_2 \text{CRI}^2 + \sum \beta_j \text{controls}, \text{ for Model 5 (Tables C.1, C.3, C.5, C.7, C.9)}$$

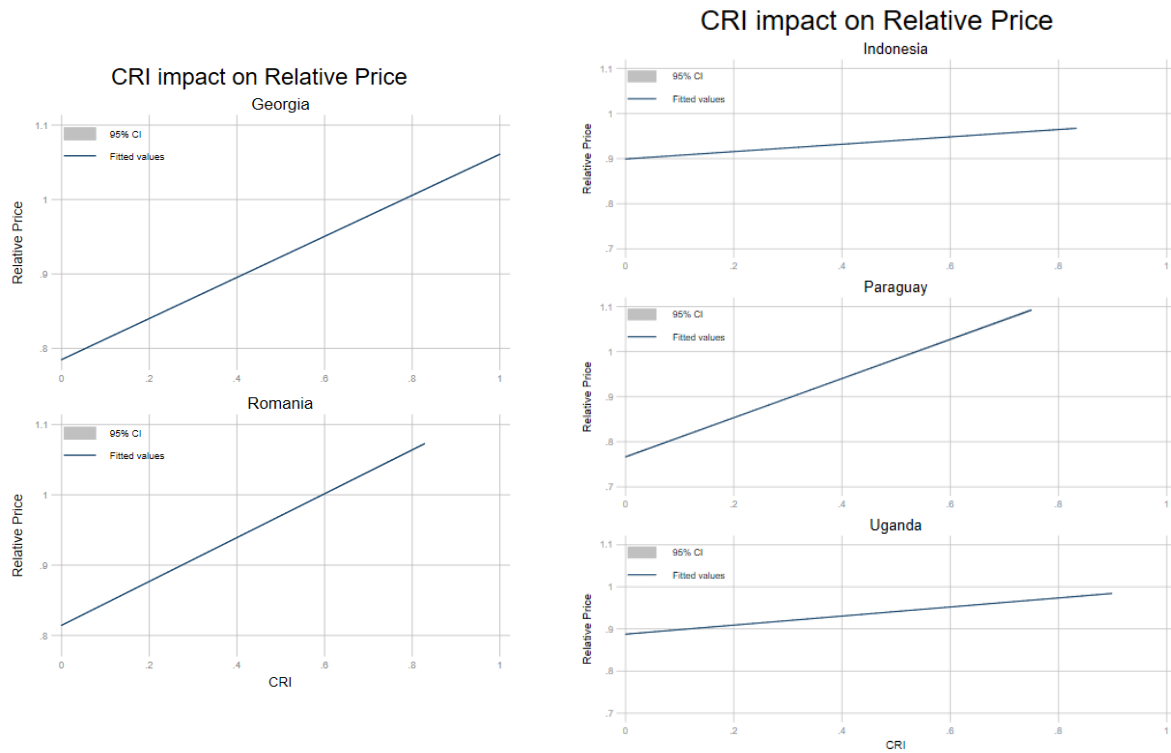
The second table in each country section, shows the impact of specific CRI components by including each component as the main independent variable in subsequent regressions. Model 1 includes single bidding, model 2 uses the 'call for tender not published' red flag, and model 3 uses contract share as the main independent variable. Positive coefficients demonstrate positive correlations between the red flag and relative price.

$$\text{Relative Price} = \alpha + \beta \text{Red flag} + \sum \beta_j \text{controls}, \text{ for Models 1-3 (Tables C.2, C.4, C.6, C.8, C.10)}$$

¹¹ Fazekas, Mihály and Tóth, Bence, (2018), The extent and cost of corruption in transport infrastructure. New evidence from Europe. Transportation Research Part A: Policy and Practice, 113, July 2018. <https://doi.org/10.1016/j.tra.2018.03.021>



FIGURE C.1: PRICE IMPACT OF CRI BY COUNTRY, MARGINAL EFFECTS PLOTS





Georgia

**TABLE C.1: GEORGIA – MAIN EFFECTS – DEPENDENT VARIABLE: RELATIVE PRICE
(FROM CONTRACT VALUE AND ESTIMATED VALUE)**

VARIABLES	(1) (0.5<relprice<1.5)	(2) (0.5<relprice<1.5)	(3) (0.5<relprice<1.5)	(4) (0.5<relprice<=1)	(5) (0.5<relprice<=1)
CRI	0.276*** (0.00390)	0.315*** (0.00440)	0.312*** (0.00428)	0.312*** (0.00428)	0.222*** (0.01167)
(CRI) ²					0.116*** (0.01557)
Year controls		✓	✓	✓	✓
Contract Value (100 quantiles)		✓	✓	✓	✓
CPV division		✓	✓	✓	✓
Buyer type			✓	✓	✓
Buyer location			✓	✓	✓
Observations	188,472	188,472	188,472	188,414	188,414
R-squared	0.148	0.203	0.206	0.206	0.207

Robust standard errors in parentheses. Clustered over buyers.

*** p<0.01, ** p<0.05, * p<0.1

**TABLE C.2: GEORGIA - ALTERNATIVE SPECIFICATION: DEPENDENT VARIABLE: RELATIVE PRICE
(FROM CONTRACT VALUE AND ESTIMATED VALUE)**

VARIABLES	(1) (0.5<relprice<=1)	(2) (0.5<relprice<=1)	(3) (0.5<relprice<=1) (> 4 contracts per year)
1.singleb	0.149*** (0.00111)		
1.nocft		-0.0192*** (0.00311)	
w_ycsh4			0.0567*** (0.00364)
Observations	188,414	188,414	130,722
R-squared	0.395	0.057	0.069

Regression includes controls for contract values, buyer type, buyer location, market, and tender year.

Robust standard errors in parentheses. Clustered over buyers.

*** p<0.01, ** p<0.05, * p<0.1



Paraguay

TABLE C.3: PARAGUAY – MAIN EFFECTS – DEPENDENT VARIABLE: RELATIVE PRICE (FROM SAVINGS)

VARIABLES	(1) (0.5<relprice <1.5)	(2) (0.5<relprice<1. 5)	(3) (0.5<relprice<1. 5)	(4) (0.5<relprice<=1)	(5) (0.5<relprice<=1)
CRI	0.378*** (0.0245)	0.380*** (0.0201)	0.371*** (0.0191)	0.386*** (0.0162)	0.554*** (0.0563)
(CRI) ²					-0.306*** (0.0958)
Year controls		✓	✓	✓	✓
Contract (100 quantiles)	Value	✓	✓	✓	✓
CPV division		✓	✓	✓	✓
Buyer type			✓	✓	✓
Buyer location			✓	✓	✓
Observations	25,597	25,597	25,597	23,551	23,551
R-squared	0.090	0.164	0.168	0.242	0.243

Robust standard errors in parentheses. Clustered over buyers.

*** p<0.01, ** p<0.05, * p<0.1

TABLE C.4: PARAGUAY - ALTERNATIVE SPECIFICATION: DEPENDENT VARIABLE: RELATIVE PRICE (FROM SAVINGS)

VARIABLES	(1) (0.5<relprice<=1)	(2) (0.5<relprice<=1)	(3) (0.5<relprice<=1) (> 4 contracts per year)
1.singleb	0.0851*** (0.00375)		
1.nocft		0.0229** (0.00935)	
proa_ycsh4			0.0573*** (0.0145)
Observations	23,398	23,551	22,859
R-squared	0.244	0.167	0.171

Regression includes controls for contract values, buyer type, buyer location, market, and tender year.

Robust standard errors in parentheses. Clustered over buyers.

*** p<0.01, ** p<0.05, * p<0.1



Uganda

TABLE C.5. UGANDA – MAIN EFFECTS – DEPENDENT VARIABLE: RELATIVE PRICE (FROM CONTRACT VALUE AND ESTIMATED VALUE)

VARIABLES	(1) (0.5<relprice<1.3)	(2) (0.5<relprice<1.3)	(3) (0.5<relprice<1.3)	(4) (0.5<relprice<=1)	(5) (0.5<relprice<=1)
CRI	0.0845*** (0.0240)	0.0863*** (0.0224)	0.0913*** (0.0218)	0.0996*** (0.0191)	0.0703 (0.0475)
(CRI) ²					0.0347 (0.0527)
Year controls		✓	✓	✓	✓
Contract Value (100 quantiles)		✓	✓	✓	✓
CPV division		✓	✓	✓	✓
Buyer type			✓	✓	✓
Buyer location			✓	✓	✓
Observations	41,394	41,394	41,394	35,793	35,793
R-squared	0.020	0.041	0.066	0.103	0.104

Robust standard errors in parentheses. Clustered over buyers.

*** p<0.01, ** p<0.05, * p<0.1

TABLE C.6: UGANDA - ALTERNATIVE SPECIFICATION: DEPENDENT VARIABLE: RELATIVE PRICE (FROM CONTRACT VALUE AND ESTIMATED VALUE)

VARIABLES	(1) (0.5<relprice<=1)	(2) (0.5<relprice<=1)	(3) (0.5<relprice<=1) (> 4 contracts per year)
1.singleb	0.0330*** (0.0109)		
1.nocft		0.0524*** (0.00957)	
w_ycsh4			0.0148** (0.00618)
Observations	35,793	35,793	22,334
R-squared	0.091	0.113	0.090

Regression includes controls for contract values, buyer type, buyer location, market, and tender year. Robust standard errors in parentheses. Clustered over buyers.

*** p<0.01, ** p<0.05, * p<0.1



Romania

TABLE C.7: ROMANIA – MAIN EFFECTS – DEPENDENT VARIABLE: RELATIVE PRICE (FROM CONTRACT VALUE AND ESTIMATED VALUE)

VARIABLES	(1) (0.5<relprice<1.5)	(2) (0.5<relprice<1.5)	(3) (0.5<relprice<1.5)	(4) (0.5<relprice<=1)	(5) (0.5<relprice<=1)
CRI	0.325*** (0.0423)	0.312*** (0.0348)	0.311*** (0.0327)	0.307*** (0.0331)	0.491*** (0.0155)
(CRI) ²					-0.437*** (0.0323)
Year controls		✓	✓	✓	✓
Contract (100 quantiles)	Value	✓	✓	✓	✓
Contract type		✓	✓	✓	✓
CPV division		✓	✓	✓	✓
Buyer type			✓	✓	✓
Buyer location			✓	✓	✓
Observations	247,750	247,750	247,750	233,946	233,946
R-squared	0.0898	0.131	0.139	0.159	0.167

Robust standard errors in parentheses. Clustered over buyers.

*** p<0.01, ** p<0.05, * p<0.1

TABLE C.8: ROMANIA - ALTERNATIVE SPECIFICATION: DEPENDENT VARIABLE: RELATIVE PRICE (FROM CONTRACT VALUE AND ESTIMATED VALUE)

VARIABLES	(1) (0.5<relprice<=1)	(2) (0.5<relprice<=1)	(3) (0.5<relprice<=1) (> 4 contracts per year)
1.singleb	0.111*** (0.00380)		
99.singleb	0.0557*** (0.00772)		
1.nocft		0.0218*** (0.00375)	
w_ycsh4			-0.00293 (0.00603)
Observations	233,946	233,946	159,139
R-squared	0.203	0.085	0.093

Regression includes controls for contract values, contract type, buyer type, buyer location, market, and tender year.

Robust standard errors in parentheses. Clustered over buyers.

*** p<0.01, ** p<0.05, * p<0.1



Indonesia

TABLE C.9: INDONESIA - MAIN EFFECTS - DEPENDENT VARIABLE: RELATIVE PRICE (FROM CONTRACT VALUE AND ESTIMATED VALUE)

VARIABLES	(1) (0.5<relprice<1.5)	(2) (0.5<relprice<1.5)	(3) (0.5<relprice<1.5)	(4) (0.5<relprice<=1)	(5) (0.5<relprice<=1)
CRI	0.0800*** (0.00546)	0.101*** (0.00721)	0.0699*** (0.00546)	0.0700*** (0.00546)	-0.0632*** (0.01390)
(CRI) ²					0.2079*** (0.01867)
Year controls		✓	✓	✓	✓
Contract Value (100 quantiles)		✓	✓	✓	✓
Contract type		✓	✓	✓	✓
CPV division		✓	✓	✓	✓
Buyer type			✓	✓	✓
Buyer location			✓	✓	✓
Observations	655,861	654,590	654,590	654,262	654,262
R-squared	0.014	0.058	0.142	0.142	0.145

Robust standard errors in parentheses. Clustered over buyers.

*** p<0.01, ** p<0.05, * p<0.1

TABLE C.10: INDONESIA - ALTERNATIVE SPECIFICATION: DEPENDENT VARIABLE: RELATIVE PRICE (FROM CONTRACT VALUE AND ESTIMATED VALUE)

VARIABLES	(1) (0.5<relprice<=1)	(2) (0.5<relprice<=1)	(3) (0.5<relprice<=1)	(4) (0.5<relprice<=1) (> 4 contracts per year)
1.singleb	0.0256*** (0.00555)			
1.corr_bid		0.0388*** (0.00134)		
2.corr_bid		0.0546*** (0.00191)		
1.nocft			0.0149** (0.00681)	
w_ycsh4				0.0222*** (0.00246)
Observations	654,262	654,262	654,262	244,579
R-squared	0.135	0.191	0.135	0.144

Regression includes controls for contract values, contract type, buyer type, buyer location, market, and tender year. Model 2 shows an alternative specification to the bidding structure in Indonesia, instead of single bidding we define cut-offs based on the distribution of the bidding behaviour such as 1.corr_bid corresponds to 12 to 22 bidders and 2.corr_bid corresponds to 1 to 11 bidders.

Robust standard errors in parentheses. Clustered over buyers.

*** p<0.01, ** p<0.05, * p<0.1