CORRUPTION RESEARCH CENTER BUDAPEST

Three indicators of institutionalised grand corruption using administrative data

Explanatory note for the U4 - Proxy Workshop, Bergen, Norway, 4/2/2014

Mihály Fazekas: University of Cambridge and Corruption Research Center Budapest, mf436@cam.ac.uk [*corresponding author*]

István János Tóth: Hungarian Academy of Sciences and Corruption Research Center Budapest

23/1/2014.

The Corruption Research Center Budapest was created in November 2013 in response to the growing need for independent research on corruption and quality of government in Hungary. The central aim of the Center is to systematically explore the causes, characteristics, and consequences of low quality of government, corruption, and regulatory failure using an inter-disciplinary approach. In addition, the Center also aims to help citizens to hold governments accountable through the use of robust evidence. Our unique research approach combines qualitative and quantitative methods to analyse micro-level actor behaviour and generates novel hard data on the phenomena under scrutiny.

Corruption Research Center Budapest: http://www.crcb.eu/

This explanatory note provides background for the Proxy Workshop presentation of the team from the Corruption Research Center Budapest (Hungary) and the University of Cambridge (UK). It is intended to provide a snapshot of the ongoing work and to guide the reader to the more thorough discussion of measurement and results.

1. What is measured exactly?

In order to harness the large amounts of previously unexploited data as well as to reflect the large monetary value and its crucial importance in the functioning of governments, corruption is measured in the domain of public procurement. While public procurement corruption can manifest in a diversity of forms, the empirical analysis concentrates only on one form in order to focus attention on corruption which most likely has the widest ramification for democracy, provision of public goods, and development: institutionalised grand corruption. This type of corruption results from the joint action of public institutions and private firms, typically to the detriment of citizens financing public spending. The working definition adopted is the following:

> *institutionalised grand corruption in public procurement refers to the particularistic allocation and performance of public procurement contracts by bending universalistic rules and principles of good public procurement in order to benefit a group of individuals while denying access to all others.*

Without providing a critical discussion of relevant scholarship, it must be clear that this definition steers clear from the often-quoted definition of corruption - *the misuse of public office for private gain* (Rose-Ackerman, 1978) – by stressing the importance of open access, universalism, and impartiality in public spending (for related approaches to corruption see: Mungiu-Pippidi, 2006; North, Wallis, & Weingast, 2009; Rothstein & Teorell, 2008).

2. Core elements of a unique measurement approach

The starting point is that neither surveys of corruption, nor detailed audits and case studies are general and reliable enough for policy purposes and testing scientific theories (Sequeira, 2012). There are also a number of promising 'objective' indicators which are nevertheless either too narrow in focus or too expensive to replicate (for an overview see: Fazekas, et al., 2013a). While these can be part of a wider measurement strategy, harnessing Big Data - the immensely increasing speed and amount of data created which covers virtually the whole spectrum of social life - holds the promise of providing the much sought after new indicators.

By implication, the measurement approach seeks to provide indicators which

- are available on a real-time basis from electronic sources,
- solely derive from 'objective' administrative data describing actor behaviour,
- are defined on the micro level such as individual transactions,
- allow for consistent comparisons across countries, organisations, and time, and
- rest on a thorough understanding of the corrupt rent extraction process in its context.

This approach requires a combination of qualitative and quantitative methods where the two works closely together.

While corruption is clandestine, it must leave traces in official records of public procurement, company ownership, and firm financial information. As open access, fair competition, and transparency are prescribed by legal frameworks across every developed and in many less developed countries; corruption, that is particularistic limitations on open access, has to pretend that it is fully legal. This characteristic of institutionalised grand corruption in public procurement creates the opportunity for an *indirect measurement approach following from anomalies of open market competition*. In addition, the *competition between corrupt groups* and especially the change of power between them (e.g. which predatory elite group forms government) create a unique opportunity to separate what is open competition and what is only a pretence of it.

The proposed measurement approach is general as long as the underlying data is available and sufficient understanding of each country's context is warranted. The data presented here only come from Hungary which could be treated as a pilot country for a wider measurement exercise. Pointing at the use in comparative research, the approach has already been tried out on Czech and Slovakian data (Fazekas, et al., 2013d).

3. Data sources and feasibility

The heart of the database derives from Hungarian *public procurement* announcements of 2009-2012. The data represent a complete database of all public procurement procedures conducted under Hungarian Public Procurement Law. It 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. The quantitative database was compiled by, first, capturing the text files of the announcements from the official online source[1]; second, applying a complex automatic and manual text mining strategy, leading to variables with clear meaning and well-defined categories. As long as data is available electronically, data collection and analysis can be done from any location. Variables of interest are, for example, the name of winner companies, the value of contracts, the deadline of submitting bids. For a full discussion see: Fazekas, et al. (2013a).

Comparable data sets can be constructed with the same methodology from national public records for example in all EU countries, the US, Russia, Brazil, or Chile for the last 6-8 years (Table 1). Development agencies such as the EuropeAid[2] or UKaid[3] have similar systems of announcing public procurement tenders which warrant the feasibility of our approach in developing country contexts. Multiple development agencies' tenders can even be accessed at a single site, Development Business[4].

[1] Hungarian Public Procurement Bulletin: http://www.kozbeszerzes.hu/nid/KE (in Hungarian)
[2] http://ec.europa.eu/europeaid/work/funding/index_en.htm
[3] https://supplierportal.dfid.gov.uk/selfservice/pages/public/publicBulletinSearch.cmd
[4] http://www.devbusiness.com/Default.aspx

A crucial difference between data collection from such sources and say survey measurement, is that once the data capture and database structuring algorithms are in place, real-time and ongoing data collection is very cheap. In addition, the scope of data availability depends on the regulation of public procurement allowing for national governments and donor agencies to extend the scope of measurement at a low cost.

Table 1. 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)[5]	Period
Brazil	Portal da Transparencia Governo Federal	http://transparencia.gov.br/	4,730[6]	2004-2012
Czech Republic	Ministerstvo pro místní rozvoj ČR	http://www.isvzus.cz/usisvz/	39,000	2006-2012
Chile	Mercado Público	https://www.mercadopublico.cl/Portal/login.aspx	30,300[7]	2003-2012
EU	Tenders Electronic Daily	http://ted.europa.eu/	130,000	2005-2012
Hungary	Közbeszerzési Értesítő	http://www.kozbeszerzes.hu/	27,300	2005-2012
Romania	eLicitatie	http://www.e-licitatie.ro/	30,000	2007-2012
Russia	Goszakupki	www.zakupki.gov.ru	2,500	2006-2012[8]
Slovakia	Úrad pre verejné obstarávanie	http://tender.sme.sk/en/	30,000	2005-2012
UK	UK Contracts Finder	http://www.contractsfinder.businesslink.gov.uk/	11,600	2000-2012
US	Federal Procurement Data System - Next Generation	https://www.fpds.gov/fpdsng_cms/	1,850	2004-2012

While such public procurement data typically cover a large portion of public spending as well as GDP (for example, 3-9% of annual GDP in Czech Republic, Hungary, and Slovakia), they still only contain information on procedures under national procurement laws. These laws often exclude small contracts and tenders of special importance such as defence. By implication, a typical public procurement database is a biased sample of total public procurement, containing only the larger and more heavily regulated cases. This bias makes the data well suited for studying more costly and more high-stakes corruption where coverage is close to complete. Although, as removing contracts from the remit of the national procurement law can in itself be part of corrupt strategies, there remains some non-random bias in the data which nevertheless can be estimated using alternative data sources (Fazekas, et al., 2013b).

[5] National currencies are converted into EUR using official exchange rates of 5/2/2013 of the European Central Bank.
[6] 2010, using exchange rate of 1 BRL=0.32 EUR
[7] Approximate value of 1000 UTM (monthly tributary units).
[8] 2006-2010 only for some regions.

There are further administrative databases available in a wide range of countries which can greatly increase the usefulness of public procurement data. These describe institutional and individual actors taking part in public procurement:

1) political officeholder data: name, office/position, and party affiliation. Data comes from official list of elected officials and appointed office holders.
2) company financial and registry data: annual turnover, annual profit, data of incorporation. Data comes from official company registry and annual financial statements submitted by companies to the government.
3) company ownership and management data: name, position, and stake. Data comes from official company registry.

In some countries such as Hungary, the publication of the list of final owners and annual financial figures are preconditions for bidding in public procurement. This suggests that even in countries without reliable company registries, such information can be obtained if regulation prescribes it.

4. Overview of the proposed indicators

Institutionalised grand corruption in public procurement manifests in multiple forms and leaves many marks in official records; hence, many indicators can be built following the same indicator building logic. The proposed indicators aim to capture different aspects of the rent extraction process:

1. Corruption Risk Index (CRI) captures corruption in the generation and allocation of rents;
2. Political Influence Indicator (PII) gauges the influence of the political group in power on companies' public procurement success;
3. Political Control Indicator (PCI) marks the direct political control of public procurement contractors.

As indirect indicators of corruption are only approximate and contain a degree of error, using multiple indicators to characterise the same country, organisation, or transaction is likely to increase precision. In addition, different measurements allow for a more profound understanding of the underlying social structures maintaining corruption and keeping those in power who benefit from it.

The three indicators have different data requirements even though all three of them require public procurement data (Table 2). This implies that data collection can be prioritised depending on which indicator is more valuable.

Table 2. Data sources needed for indicator building

Corruption indicator	Procurement data	Political officeholder list	Company financial data	Company ownership data
Corruption Risk Index (CRI)	X			
Political Influence Indicator (PII)	X		X	
Political Control Indicator (PCI)	X	X		X

This is work in progress, not all indicators are fully documented and finalised, while there are further indicators 'in the making'.

4.1 Corruption Risk Index (CRI)

CRI measures the probability that the principle of open access is violated in the process of awarding and performing public procurement contracts in order to serve corrupt rent extraction by a select few. In other words, it expresses the probability of tender issuers pretending that tenders are competitive as prescribed by law while restricting competition to award contract to a well-connected bidder recurrently. CRI is a composite index of elementary corruption risk indicators capturing 'corruption techniques' such as tailoring eligibility criteria to fit a single company or using exceptional procedure types to limit openness of competition. It reflects a corrupt rent extraction logic where elementary corruption techniques are systematically used for restricting access and recurrently benefiting the same winner.

CRI is constructed in a three steps:

1) A long list of elementary corruption indicators is identified (30+ indicators) which are proven to indicate corruption in *some* cases, using qualitative methods like review of international academic literature, media content analysis, review of court judgements, and key informant interviews (Fazekas, et al., 2013b).
2) Those indicators are selected from the long list which prove to be systematically linked to restricted access as captured by single bidder contracts *as well as* to recurrent contract award to the same company as captured by winner contract share over 12 months. Regression analysis controlling for alternative explanations such as market specificities and low state capacity is used for identifying such indicators (Fazekas, et al., 2013a). In practical terms, corruption indicators that are significant and substantial in *both* regression models are selected[9].
3) CRI is calculated as the weighted sum of selected elementary corruption risk indicators where each elementary indicator is weighted to reflect its strength in predicting lack of competition and recurrent contract award. In addition, CRI is normed in order to fall in the 0-1 band. For the list of components and their weights see Table 3.

The resulting CRI can take any value between 0 and 1, where 0 means minimal or no corruption risk and 1 means maximal corruption risk observed.

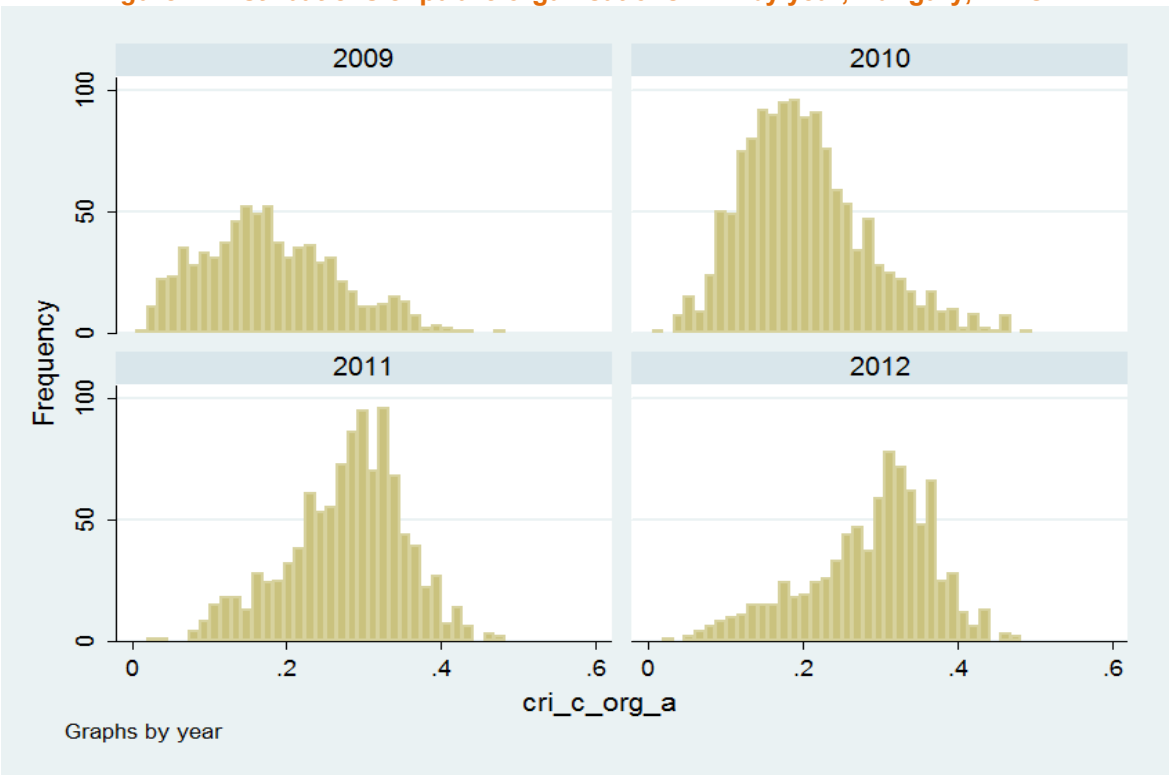
[9] Each of the two generic regressions analyses are run with multiple specifications in order to check for robustness.

Table 3. Component weights of CRI reflecting variable and category impact on corruption outcomes

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
length of eligibility criteria (deviation from market average)	
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	0.000
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	0.096
weight of non-price evaluation criteria	0.000
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 contract share	0.096

While CRI is defined on the level of individual public procurement tenders it can also be aggregated to characterise organisations, markets, or countries over time. For example, Figure 1 demonstrates the changing distributions of organisation-level CRI in Hungary throughout 2009-2012 pointing at an increasing average corruption risk.

Figure 1. Distributions of public organisations' CRI by year, Hungary, N=2841



4.2 Political Influence Indicator (PII)

PII indicates whether a company's success on the public procurement market depends on the political group in power at the national or local level. Such companies are identified by the change in total company contract volume due to government change. Those companies are designated as politically connected companies whose change in contract volume cannot be explained by standard economic explanations of market success such as main market or prior investment *and* whose deviation from the standard economic explanatory model is very large.

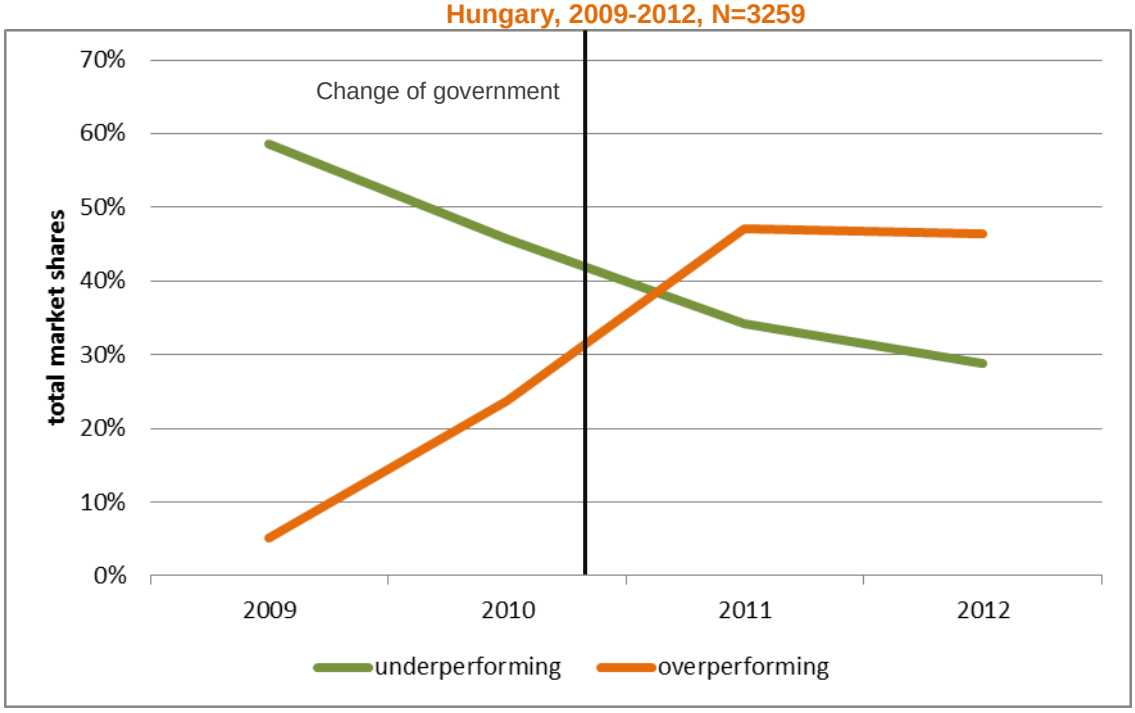
In order to separate favouritism from other influences, PII is created in three steps:

1) A range of baseline regressions are run to explain contract volume at the firm-level controlling for standard economic factors influencing market success. First, the simplest of these regressions is when total contract value difference from before to after national government change is analysed (Fazekas, et al., 2013c). Second, companies' contract volume is regressed over time looking for a structural break in the time series falling at the year of national government change. Third, municipal-level diversity can be exploited by regressing company contract volume per municipality on the sample of municipalities with change of local government.
2) Benchmarks are generated by running the same regressions, but this time for years different than the change of national government and for municipalities where there was no change of local government.
3) Those companies are marked for which the differences between baseline and benchmark regressions are significant and substantial for *every* regression.

As a result, PII takes the value of 0 *or* 1, where 0 means that the company's public procurement market success does not depend on which government is in power, and 1 means that it does.

While PII indicates companies' indirect political connections, it can also be used for characterising markets, public organisations, or whole countries. For example, Figure 2 depicts combined market shares of companies depending on how well their behaviour fits a standard economic model (for details of regressions see: Fazekas, et al., 2013c). This suggests that more than half of the Hungarian public procurement market may be controlled by firms who systematically deviate from a standard economic logic reflecting company size, main market, profitability, and prior investment. These results are only indicative, work is in progress.

Figure 2. Combined market shares of companies more successful than model prediction (overperforming) and less successful than model prediction (underperforming), Hungary, 2009-2012, N=3259



4.3 Political Control Indicator (PCI)

PCI measures whether a public procurement winning company has direct political connections. Political connections are defined as the winning firm's owners or managers holding a political office, where political office is broadly defined as elected national and local representatives and high-level appointed public officials such as supreme court judges or heads of national police force.

PCI is created in three steps (Fazekas et al., 2013a):

1) The list of names of all the owners and managers of companies winning public procurement contracts is identified and biographical information such as date of birth is obtained from official registries.

2) The list of names of political officeholders is created and additional biographical data is collected using official sources such as the electoral register.
3) Winning companies are designated as having political connections if at least one of their owners or managers has held or still holds a political office (i.e. appear in both name lists with the corresponding biographical information being identical). Where no additional biographic information is available for verifying the identity of individuals, matching likelihood is estimated using name frequency and organisation seat geographical distance.

As a result, PCI can take the value of 0 *or* 1, where 0 means that the company has no direct political connection and 1 means that it has at least one such connection.

While foreign owned companies may be less likely to have owners with national political office, they may well appoint top managers and board members in their national subsidiaries who are politically involved to secure local knowledge and useful ties.

Similarly to CRI and PII, PCI can also be used for characterising markets, public organisations, or whole countries.

5. The issue of validity

Clearly, none of the three proposed measures indicate institutionalised grand corruption directly. However, they are designed so that they signal the likelihood of institutionalised grand corruption to occur. Unfortunately, in systematically corrupt environments, we cannot rely on courts to validate our indicators: first, because they are highly likely to fail to uncover and prosecute most of the corruption cases (i.e. problem of false negatives); second, because they are also likely to hand out biased judgements serving political purposes (i.e. problem of false positives). By implication, the validation of indicators has to rely on alternative methods. We have conducted numerous validation tests with positive results (further work in progress). These fall in four broad categories:

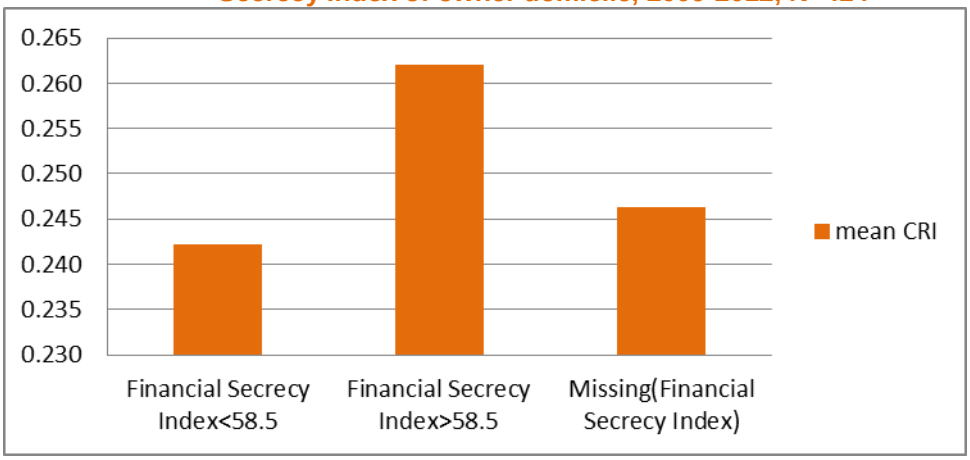
1) **Internal validity** confirmed by the set-up of indicators themselves such as the clarity of indicator building logic, the richness of qualitative evidence supporting components of each indicator, and the quality or regression models used for singling out corruption from other factors such as state capacity or market specificities.
2) **External validity established by the co-variation between the three lead indicators** which is expected to be only moderately strong as they indeed capture different aspects of the corrupt rent extraction process in public procurement with their individual flaws (for details see: Fazekas, et al., 2013a). Nevertheless, no co-variation would be a strong argument against validity. For example, the significantly and considerably higher CRI of politically connected firms compared to firms without political connections delivers a strong argument for indicator validity (Table 4). In a similar vein, companies with political connections (PCI=1) are 6% more likely to have over or under-performed compared to our economic model (PII=1).

Table 4. Comparisons of mean CRI of politically connected and not connected firms, Hungary, 2009-2012

Group	N	Mean CRI	Std. Err.	Std. Dev.	95% Conf.Interval	
PCI=0 (no political connection)	2687	0.254	0.002	0.113	0.250	0.258
PCI=1 (politically connected)	1318	0.264	0.003	0.112	0.258	0.270
combined	4005	0.257	0.002	0.113	0.254	0.261
difference [CRI(PCI=1)-CRI(PCI=0)]		***0.010******	0.004		0.017	0.003

3) **External validity indicated by further indicators** such as company profitability (Cole & Tran, 2011) or institutional integrity measures (Szente, 2011). For example, Fazekas et al. (2013a) finds that high CRI companies are 25% more profitable than low CRI companies. Or companies whose owners are registered in 'tax heaves' as measured by the Financial Secrecy Index (Tax Justice Network, 2013) are of significantly and considerably higher corruption risk (Figure 3). While it is possible to check the proposed 'objective' indicators of institutionalised grand corruption against widely used survey measures such as government favouritism (World Economic Forum, 2010), it is expected that the difference in indicator scope and quality may lead to little to no correlation.

Figure 3. Mean CRI of companies with owners outside Hungary according to the Financial Secrecy Index of owner domicile, 2009-2012, N=414



4) **External validity demonstrated by well-documented cases** is a tempting route to validation; however, due to weaknesses of courts in systematically corrupt environments, case selection may render such an exercise very difficult or impossible. However, contrasting organisations of very high corruption risks with those of very low risks based on rich qualitative scientific analysis may deliver a valuable validity test (work in progress).

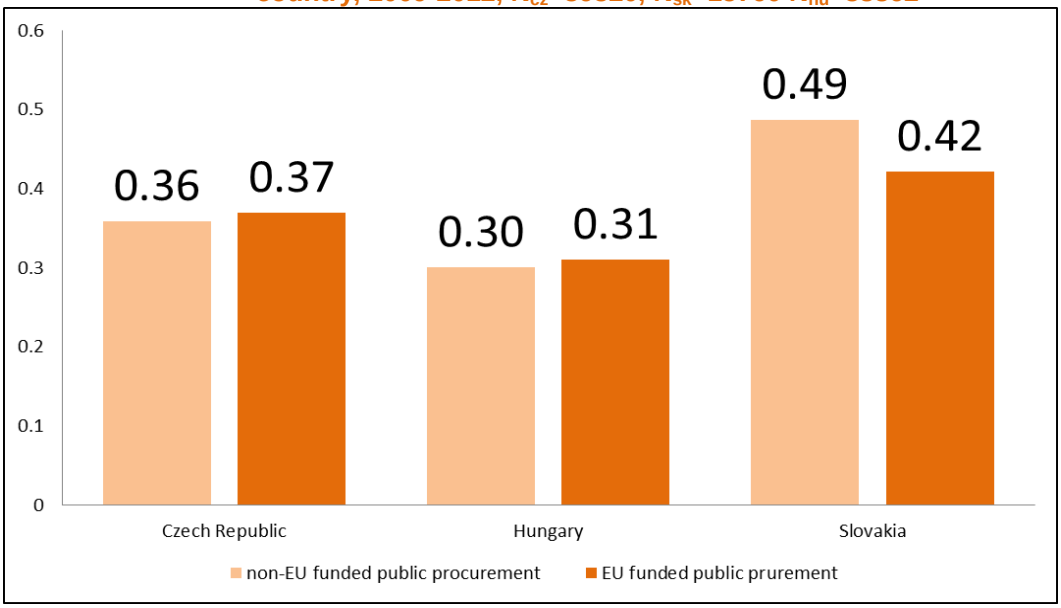
The proposed validity tests must be understood only as indications of retrospective validity because the problem of reflexivity is particularly troubling in corruption research. This means that validity can be established in retrospect, but corrupt groups are likely to respond to changes in monitoring technology and detection probabilities. Hence, as soon as any of these indicators is used for monitoring corruption at large, indicator validity is expected to deteriorate due to efforts of corrupt actors to better hide their actions. This means that further refinements, also including the incorporation of further variables (role of regulator to constantly increase transparency!), are necessary for the indicators to remain valid.

6. Potential for policy evaluation

As the indicators of institutionalised grand corruption are defined at the micro level and available over long time periods for multiple countries, the possibilities for policy evaluation are manifold and highly sophisticated. Without being exhaustive, the indicator set can be used in four different levels:

- **Evaluating countries** against each other or the same country over time. For example, it is possible to track the overall value of public procurement spending 'touched by' corruption and link it to countries' institutional characteristics.
- **Evaluating large funding programmes** or spending lines such as EU structural funds spending in Central and Eastern Europe. This can be done, for example by comparing similar public procurement tenders which only differ in the source of financing (EU vs national) in terms of their corruption risks. Figure 4 depicts average CRI scores for EU and non-EU funded tenders throughout 2009-2012 in Czech Republic, Hungary, and Slovakia, suggesting that EU funding *increases* corruption risks in Czech Republic and Hungary (Fazekas, et al., 2013d).
- **Evaluating single regulatory changes** on procurement systems' corruption risks, such as the effect of loosening transparency regulations (Fazekas, et al., 2013a) or tightening requirements on the minimum number of bidding firms[10].
- **Evaluating organisational reform** by, for example comparing similar public organisations some of which implemented organisation-level changes such as introducing an integrity management system.

Figure 4. Average CRI scores of EU and non-EU funded public procurement procedures, by country, 2009-2012, N_{cz}=39320, N_{sk}=15760 N_{hu}=38862



[10] For a Czech example see: http://blog.aktualne.centrum.cz/blogy/jiri-skuhrovec.php?itemid=21197

Bibliography

Cole, S., & Tran, A. (2011). 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.

Fazekas, M., Tóth, I. J., & King, L. P. (2013a). Anatomy of grand corruption: A composite corruption risk index based on objective data. CRC-WP/2013:02, Budapest: Corruption Research Centre.

Fazekas, M., Tóth, I. J., & King, L. P. (2013b). Corruption manual for beginners: Inventory of elementary “corruption techniques” in public procurement using the case of Hungary. CRC-WP/2013:01,Corruption Research Centre, Budapest.

Fazekas, M., Tóth, I. J., & King, L. P. (2013c). Hidden Depths.The Case of Hungary. In A. Mungiu-Pippidi (Ed.), *Controlling Corruption in Europe vol. 1* (pp. 74–82). Berlin: Barbara Budrich Publishers.

Fazekas, M., Chvalkovská, J., Skuhrovec, J., Tóth, I. J., & King, L. P. (2013d). Are EU funds a corruption risk? The impact of EU funds on grand corruption in Central and Eastern Europe. CRC-WP/2013:03, Corruption Research Centre, Budapest.

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

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.

Rose-Ackerman, S. (1978). *Corruption: A Study in Political Economy*. New York: Academic Press.

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

Sequeira, S. (2012). Advances in Measuring Corruption in the Field. In D. Serra & L. Wantchekon (Eds.), *New Advances in Experimental Research on Corruption* (pp. 145–176). Bingley, UK: Emerald Group Publishing Limited.

Szente, Z. (Ed.). (2011). *Elemzések a Közszféra korrupciós kockázatairól*. Budapest: Állami Számvevőszék.

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

World Economic Forum. (2010). *The Global Competitiveness Report 2010–2011*. Geneva, Switzerland: World Economic Forum.