

# Mining public procurement data for corruption

**Daniel Tanis**

University of Cambridge and DIGIWHIST

[dt444@cam.ac.uk](mailto:dt444@cam.ac.uk)

**Mihaly Fazekas**

University of Cambridge and DIGIWHIST

[mf436@cam.ac.uk](mailto:mf436@cam.ac.uk)

# Agenda

- DIGIWHIST
- Building corruption indicators
- Exploring corruption indicators: selected examples
- Exploring collusion indicators: selected examples

# Agenda

- **DIGIWHIST**
  - Building corruption indicators
  - Exploring corruption indicators: selected examples
  - Exploring collusion indicators: selected examples

# What is DIGISWHIST?

## Institutional background

- 3-year project, Horizon2020 funding (EU). **2015-2018**
- Main output: [opentender.eu](https://opentender.eu): a big data on procurement transparency in Europe.
- Interdisciplinary consortium with 6 partners (with Cambridge University leading)
- Ultimate objective: increase trust in governments and improve public spending.

## Data Collection

- We map national sources of procurement information in 35 jurisdictions and have made a thorough assessment of data availability (highly heterogeneous)
- We map relevant variables on individual contracts in each country and program the automated extraction of the information
- We compile, clean and structure a comparative database

## Data analysis

- We automatically analyse each contract announcement using expert-informed 'red flags'
- And we produce validated indicators of procurement corruption risks, transparency and administrative capacity: historical and real-time

## Sustainability

- After its running life (February 2018) the project foresees the sustainability of all efforts by:
  - a. Establishing partnerships with local civil society organisations who could take over the national portals.
  - b. Full open license to data, indicators, and watchdog tools.
  - c. "keeping the lights on" for 3 years after the project ends, by facilitating the tools to continue the efforts.

## Dissemination and citizen engagement

- A single portal for all 34 countries:
  - a. With a comment section to help verify the data
  - b. Link to FOIA portals
  - c. Link to whistleblower portals to help and prevent corruption

# Data analysis: Tackling corruption in public procurement

## 1 What is corruption?

Broader understanding of corruption: beyond bribery and legality

In public procurement, the aim of corruption is to steer the contract to the favored bidder without detection. This is done in a number of ways, including:

- *Avoiding competition* through, *e.g.*, unjustified sole sourcing or direct contract awards.
- *Favoring a certain bidder* by tailoring specifications, sharing inside information, *etc*\*.

The company benefits from a "temporary monopoly" in a given transaction to extract rents

## 3 How can we measure it?

### A. Tendering Risk Indicators (TRI)

Using tender data

### B. Contracting Body Risk Indicators (CBRI)

Treasury account of public organisations

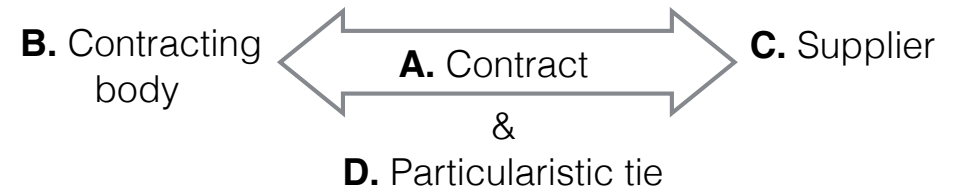
### C. Supplier Risk Indicators (SRI)

Company data: financials, registry and ownership

### D. Political Connections Indicators

Political office holder data

## 2 How is it organised?



## 4 How can it be used?

- Identifying corrupt networks
- Identifying collusion
- Increase resolution of current corruption indicators
- Compare countries, policies, years

# Agenda

- DIGIWHIST
- **Building corruption indicators**
- Exploring corruption indicators: selected examples
- Exploring collusion indicators: selected examples

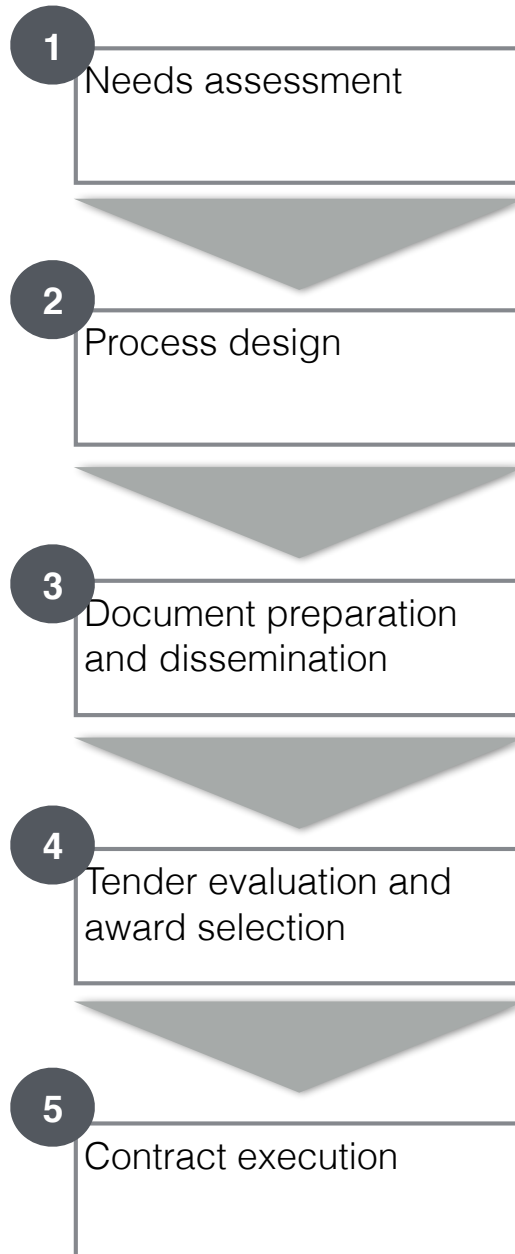
## Indicator building assumptions

Corruption is deviation from the norm (*de jure/de facto*)

Expert knowledge is necessary, but not sufficient

Corruption is categorical/non-linear

# Understanding the process: procurement phases



**Each of these phases is vulnerable to specific techniques of corruption**



# Understanding corrupt strategies: selected examples

- T.1.1** Defining unnecessary needs
- T.1.2** Defining needs to benefit a particular supplier
- T.2.1** Tinkering with threshold and exceptions
- T.2.2** Tailoring eligibility criteria
- T.2.3** Abusing formal and administrative requirements
- T.2.4** Tailoring evaluation criteria
- T.2.5** Using long term complex contracts
- T.2.6** Tinkering with submission period
- T.3.1** Selective information provision
- T.3.2** Avoiding publication of call for tenders
- T.3.3** Strategically modifying call for tenders
- T.3.4** Excessively pricey documents, difficult access to documents
- T.3.5** Deliberate errors in document publication
- T.4.1** Strategically annulling procedures
- T.4.2** Repeated violation of public procurement rules
- T.4.3** Unfair scoring
- T.4.4** Abusing unit prices in the contract
- T.5.1** Modifying contracts strategically
- T.5.2** Abusing add-on contracts
- T.5.3** Performance violating contract

# Validated Tendering Risk Indicators

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=non-open procedure
	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 product description	number of characters of product description MINUS average number of characters in the given market
	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
assessment	Call for tenders modification	0=call for tenders NOT modified 1=call for tenders modified
	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	Unit price	% deviation in the price of standardized unit compared to private market price or lowest public procurement price
	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)

# Agenda

- DIGIWHIST
- Building corruption indicators
- **Exploring corruption indicators: selected examples**
- Exploring collusion indicators: selected examples

# Understanding non-linearity

Road reconstruction between two medium-sized cities, 25 km two-lane road, delivery in the summer

How long should the advertisement period be?

5 calendar days

10 calendar days

20 calendar days

30 calendar days

50 calendar days


Statistical analysis is needed to move from expert-suggested indicators to validated indicators (non-linearity, selection, etc.)

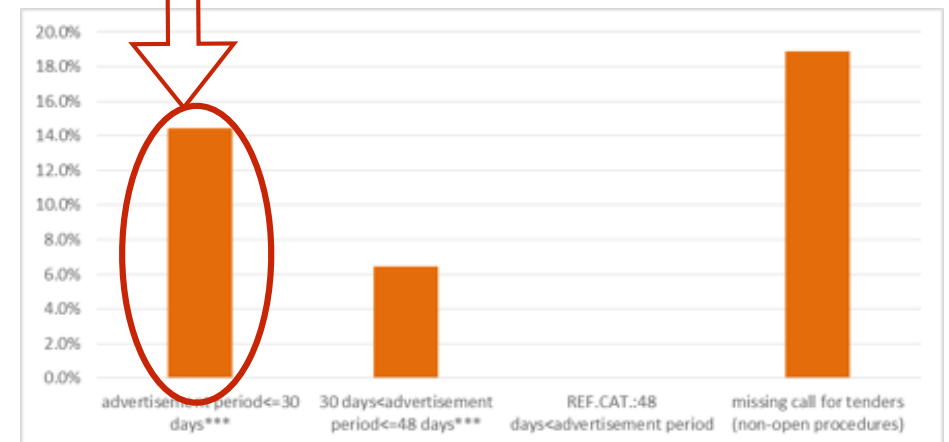
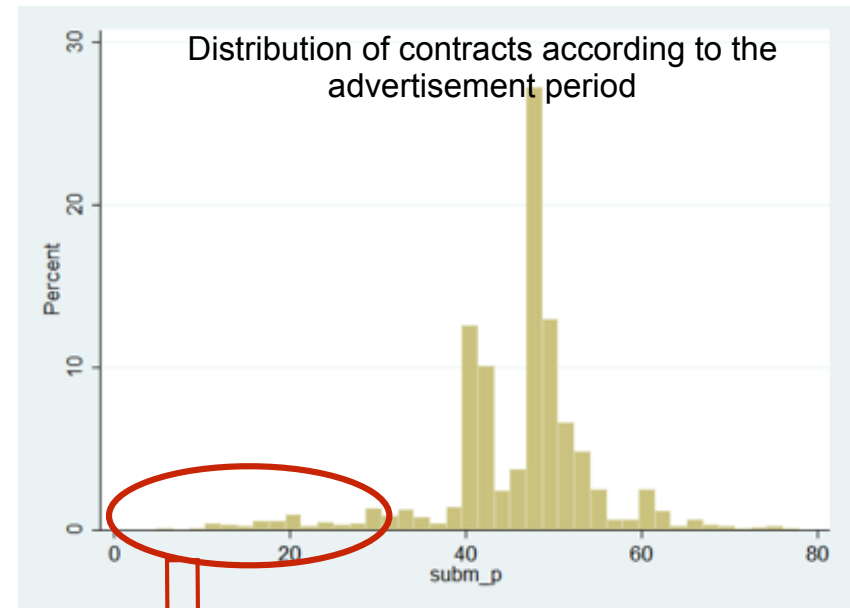


# And tenders with single bidders are often the result of specific tender conditions

There are several strategies companies and public officials can use to reduce tender competition:

- Reducing the advertisement period
- Over-specifying technical requirements
- ...

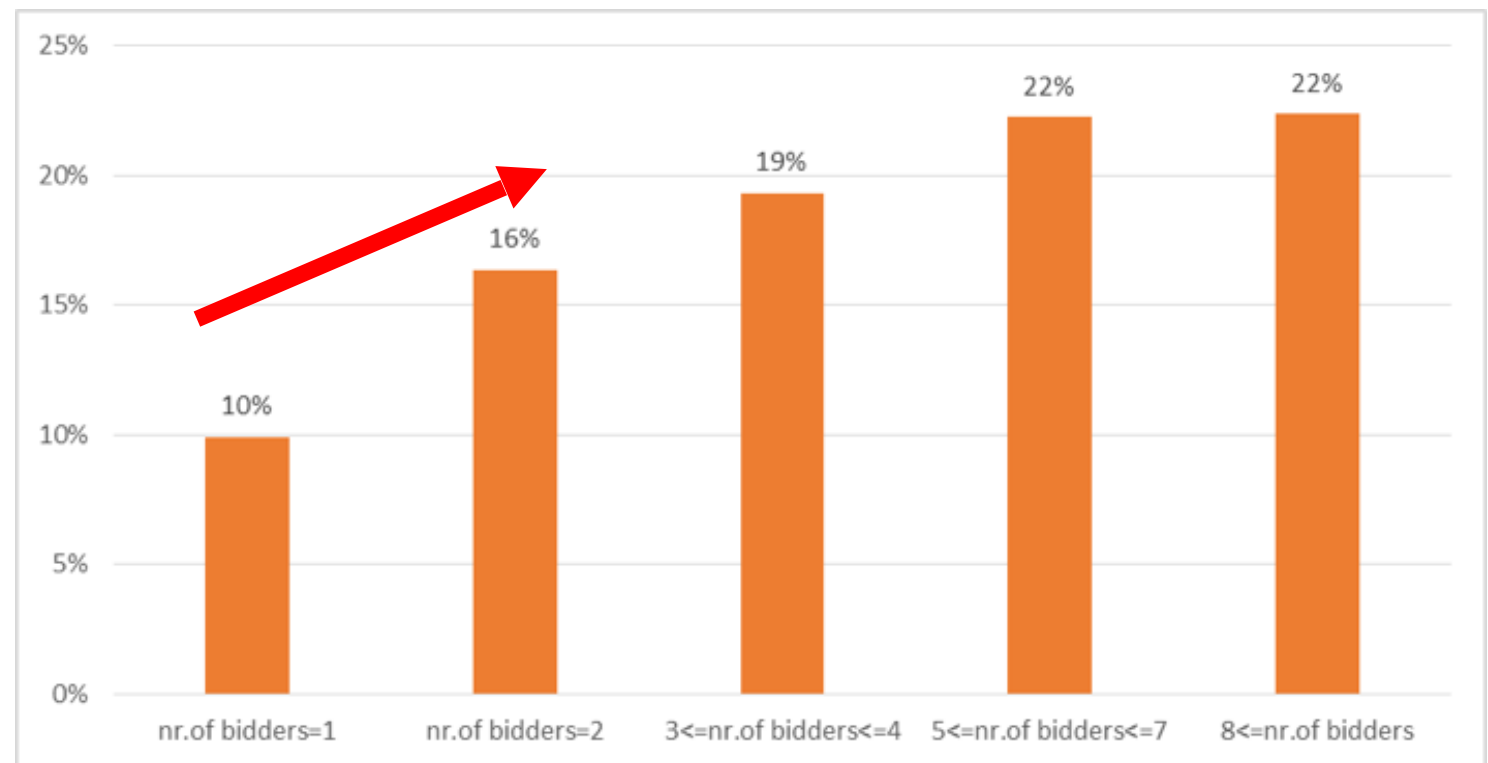
Even though not all single bid tenders are the result of corrupt interactions, many of its currencies can be directly linked to specific decisions taken somewhere along the tender process



## Number of bidders and price

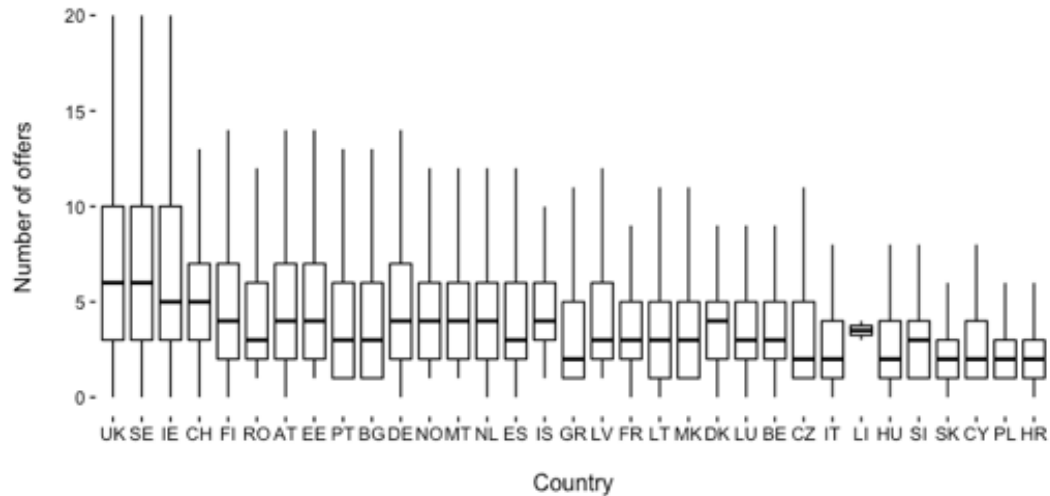
- Price savings by the number of bidders

543,705 contracts, EU27, 2009-2014

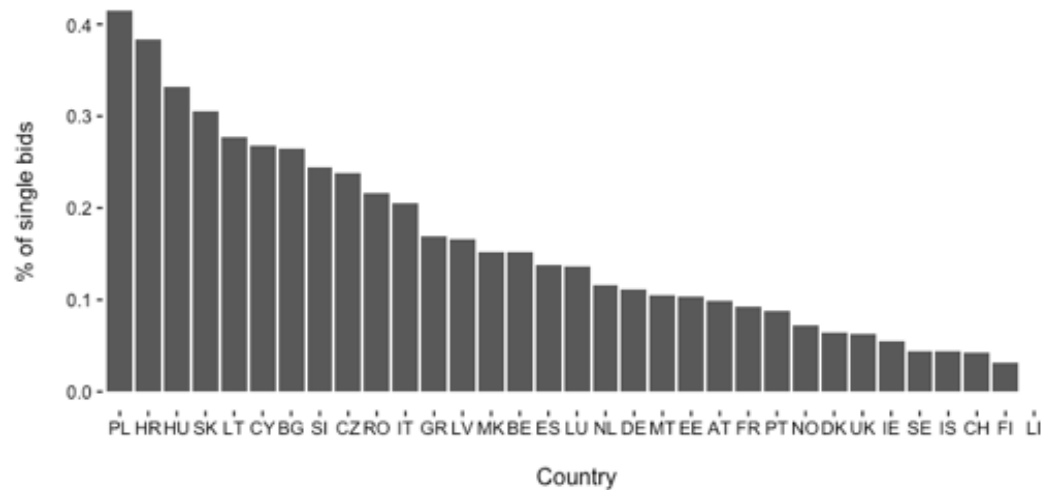


# Tenders with single bidders have an uneven distribution across Europe

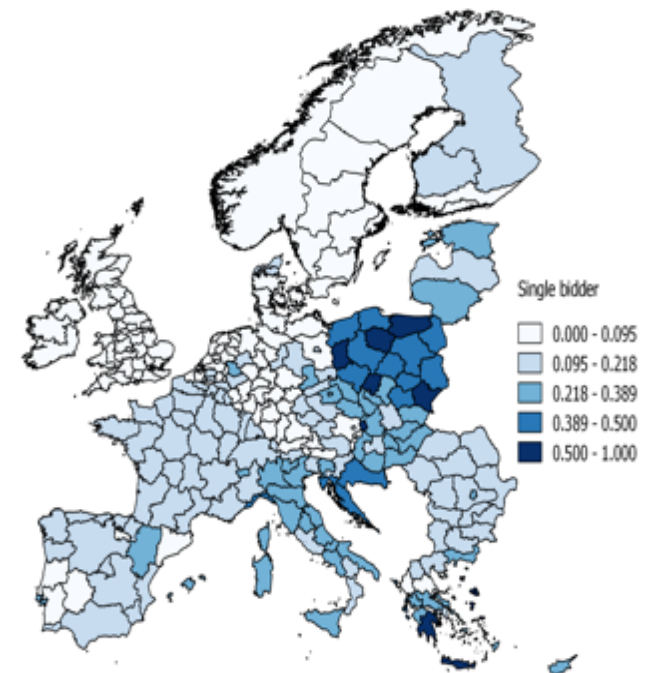
Distribution of number of bidders 2015 data



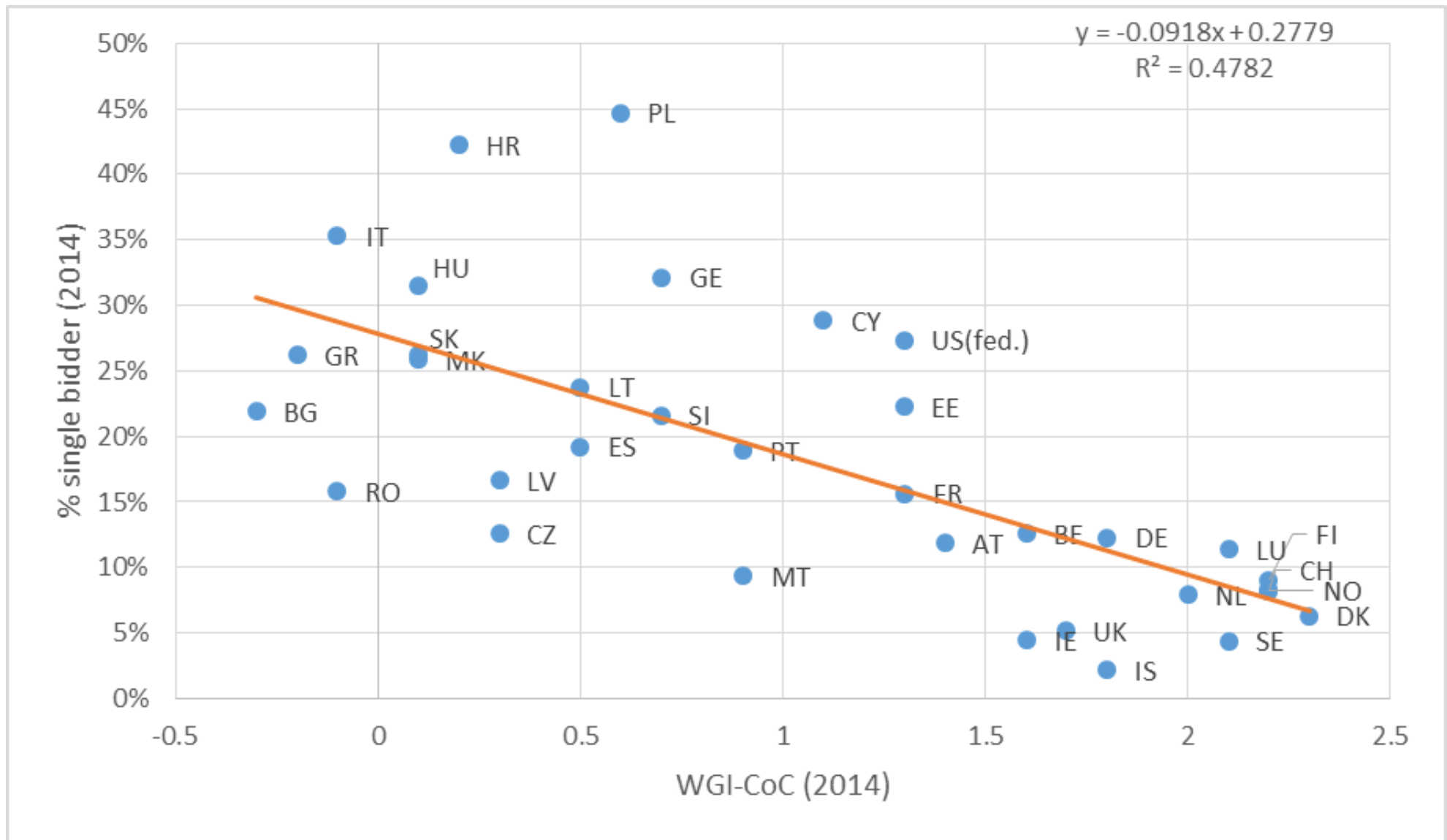
Single bidders 2015 data



Single bidder ratio by region, 2009-2014



## And this trend correlates well with subjective indicators of corruption

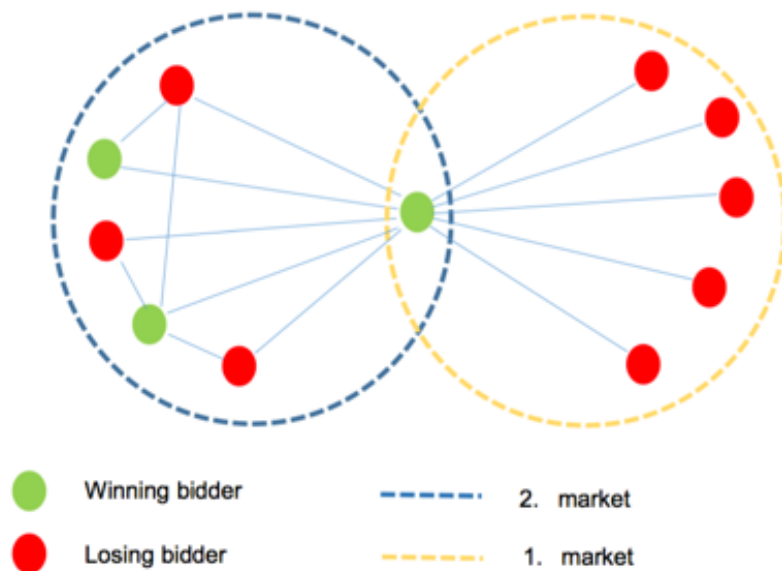




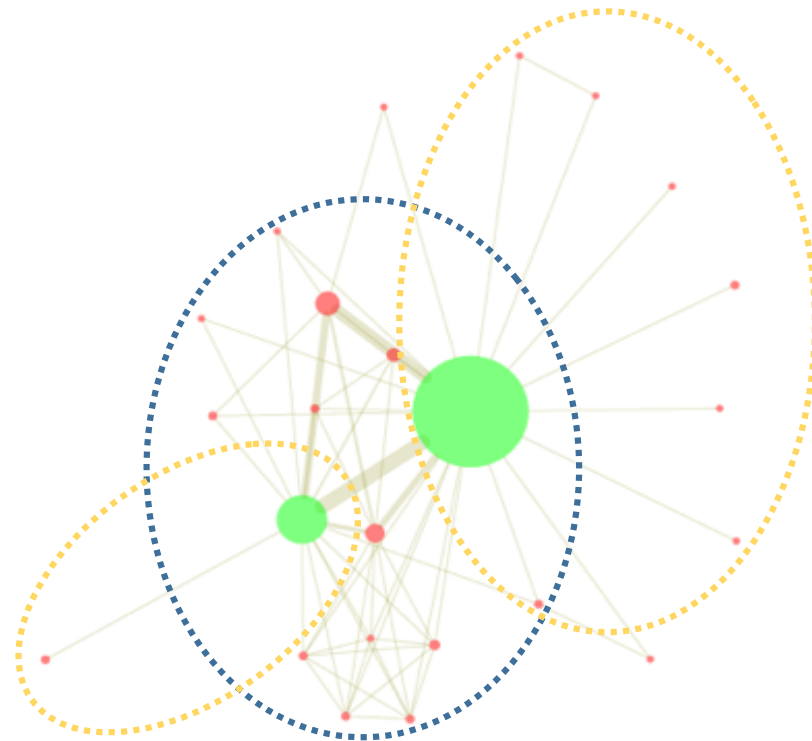
# But there are sophisticated collusive strategies to avoid single bid tenders

## Co-bidding networks

### Theory - cut points

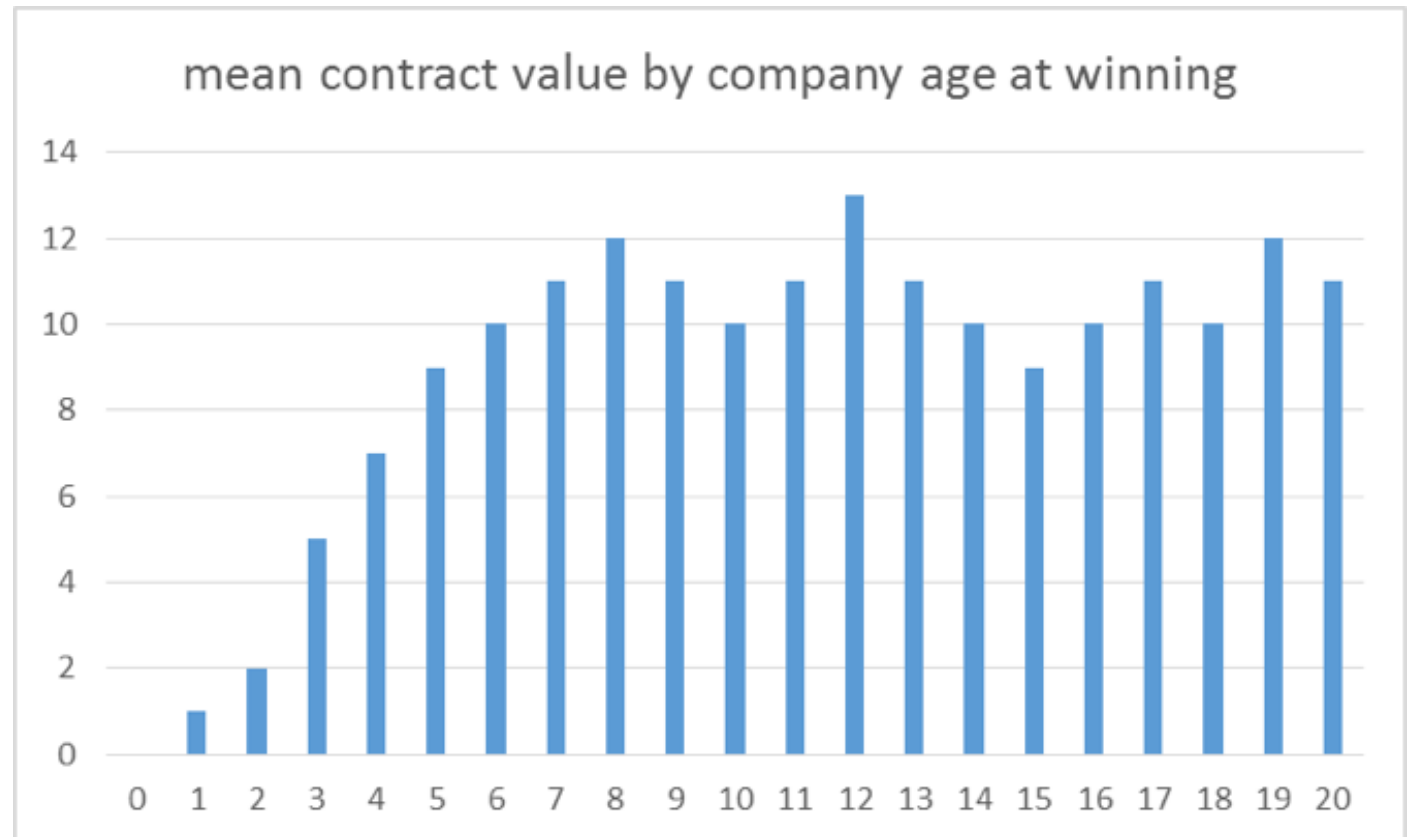


### Example: Sweden 2009-2014



# Supplier Risk Indicators: Expected success of companies by age

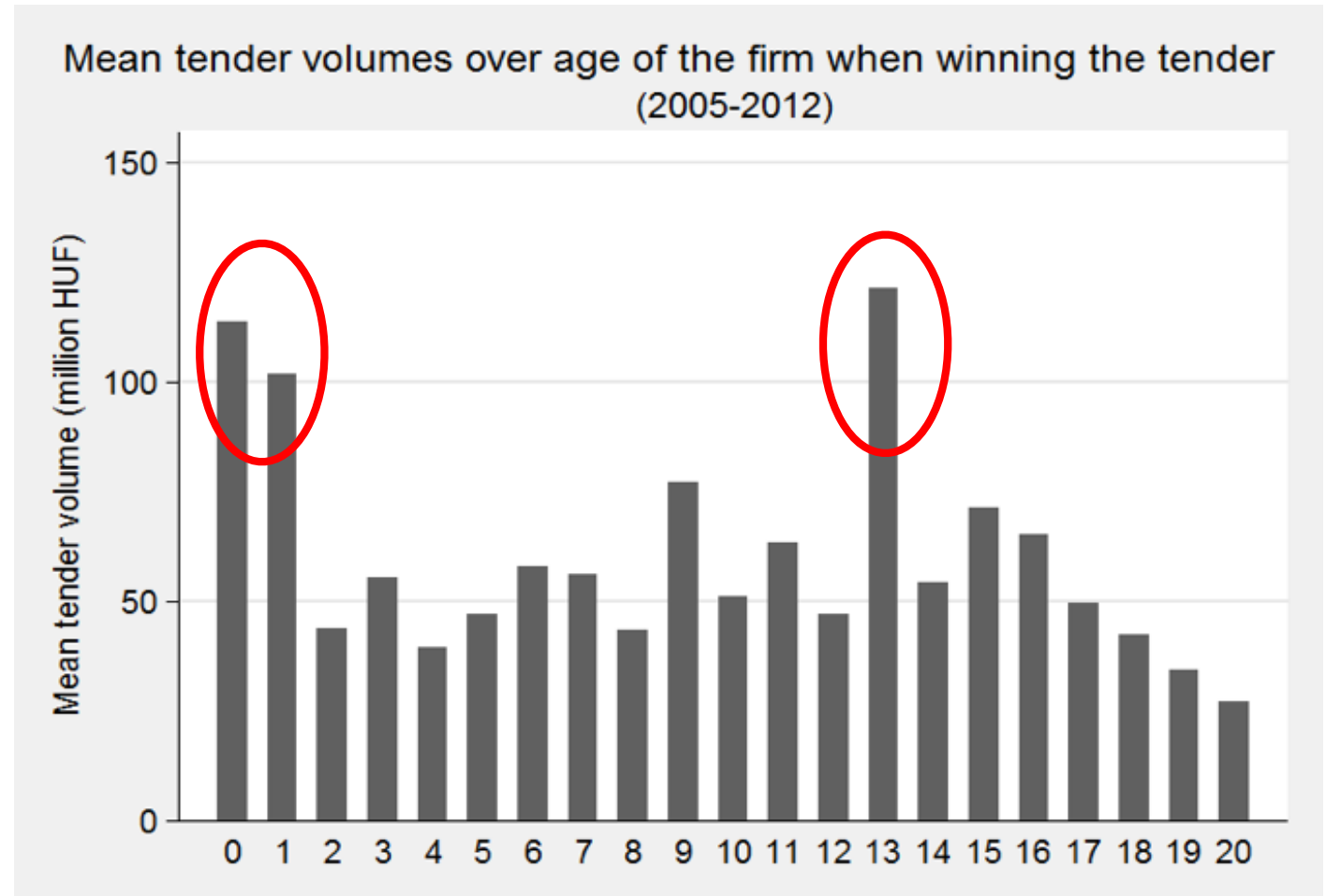
1. Gradual build-up of contracts
2. Natural fluctuation over time



# Supplier Risk Indicators: Observed success of companies by age

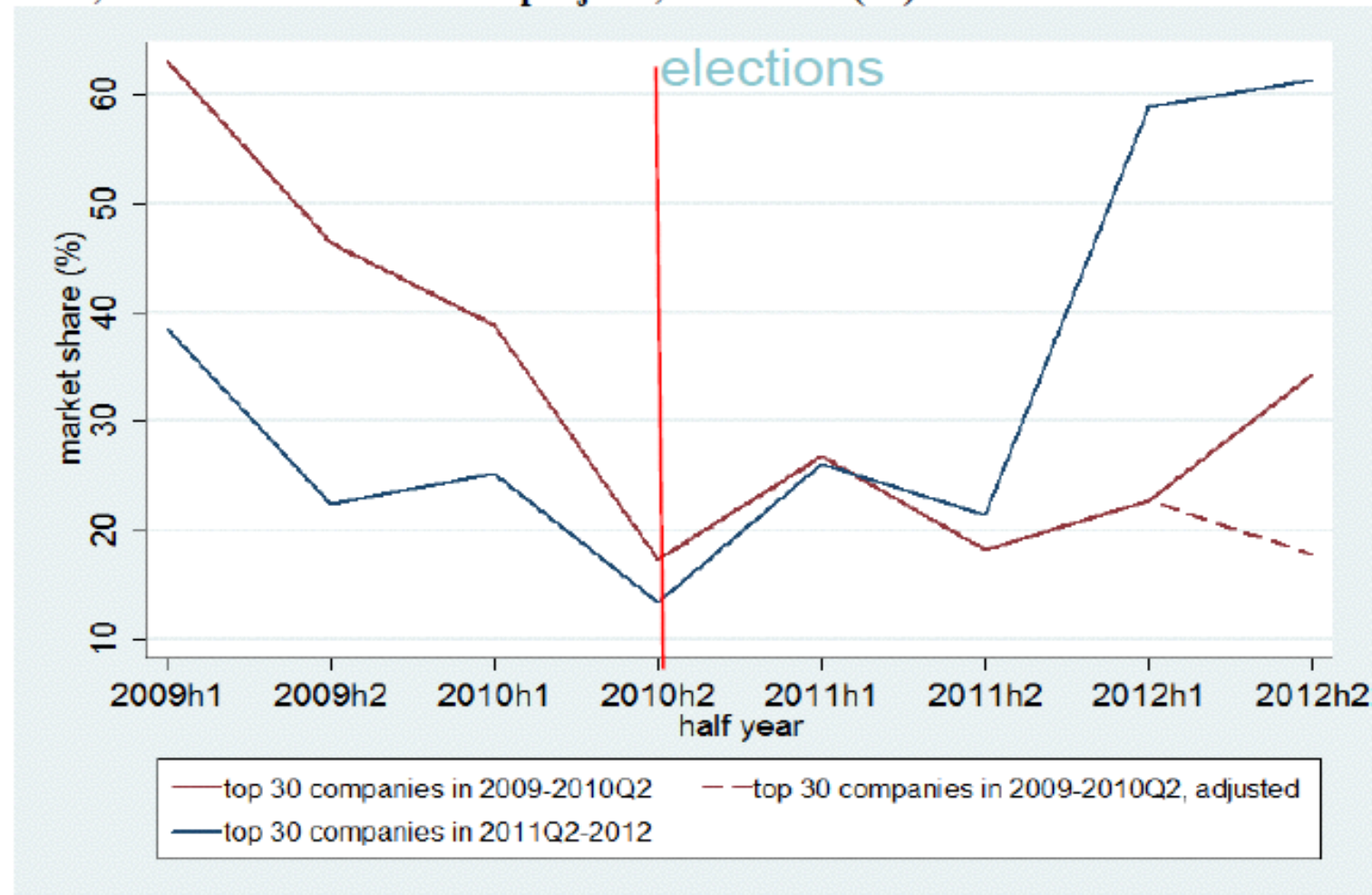
1. “Just founded” companies
2. Companies founded under party last in power

In Hungary



# Supplier Risk Indicators: Company success driven by political power

**Figure 1. Changes in market shares of the top 30 firms of 2009-2010H1 and of 2010H2-2012, EU funded construction projects, 2009-2012 (%)<sup>5</sup>**

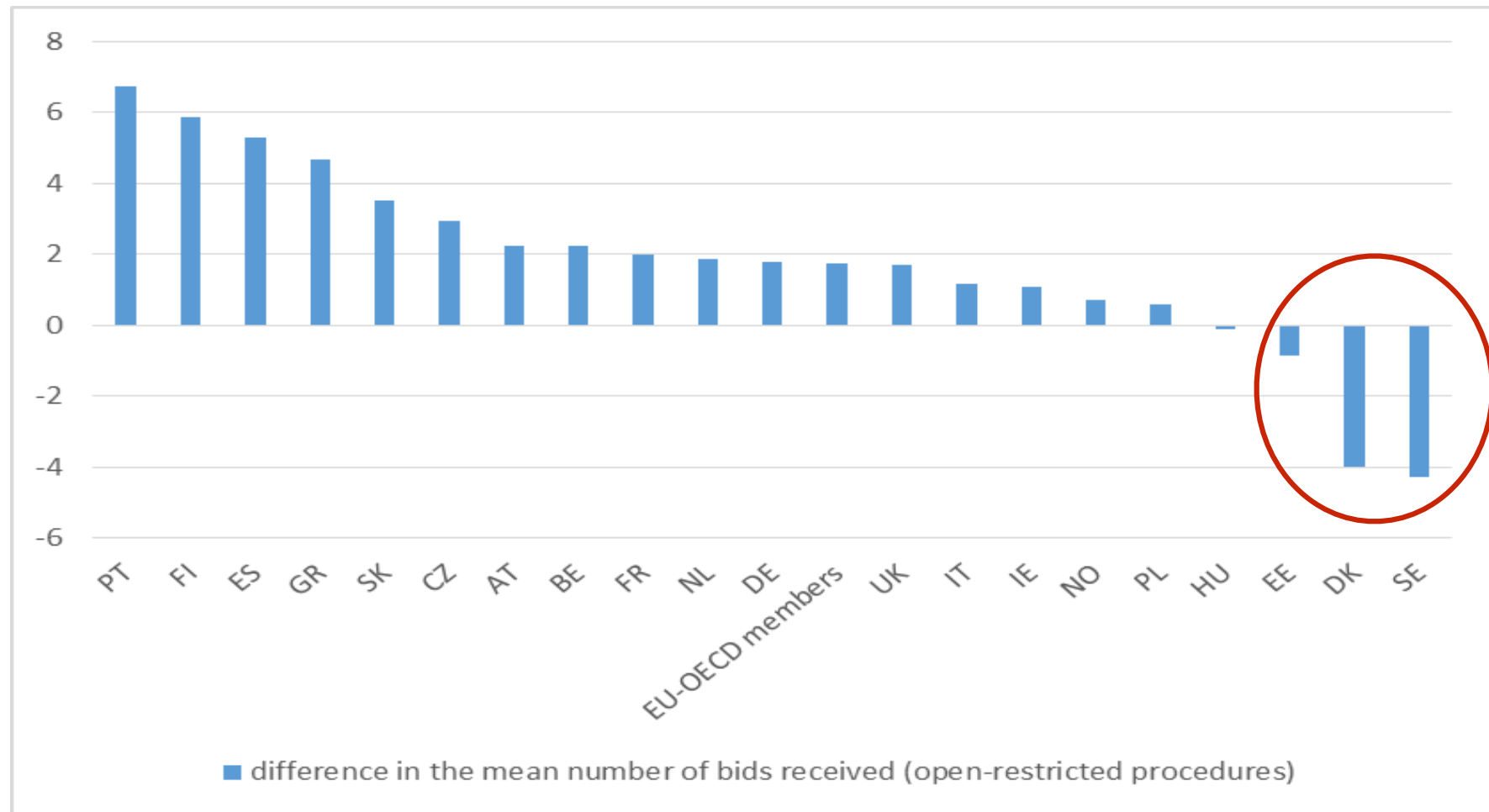


Source: MaKAB

Note: market share of company  $i$  in time  $t$  = total value of contracts won by company  $i$  in EU funded construction in time  $t$  / total value of contracts won in EU funded construction in time  $t$

# It is possible to compare data across countries, but context is still relevant

Differences in the average number of bids submitted between non-restricted and restricted procedures, OECD-Europe, 2013, c.value>58k EUR



# Agenda

- DIGIWHIST
- Building corruption indicators
- Exploring corruption indicators: selected examples
- **Exploring collusion indicators: selected examples**

# Type of collusive behaviour

## Elementary collusion technique

- Withheld bids
- Non-competitive-bidding
- Joint Bidding

## Form of rent-sharing

- Sub-contracting
- Consortia/joint ownership
- Coordinated Bidding
- Informal

## Resulting market structure

- Concentrated
- False competitive

Market structure	Elementary collusion technique	Forms of rent sharing			
		Sub-contractor	Consortia/joint ownership	Coordinated bidding	Informal side-payment
Concentrated	Withheld bids	A	-	-	Partial
	Losing bids	B	-	-	Partial
	Joint bids	-	C	-	-
False competitive	Withheld bids	D	-	F	-
	Losing bids	E	-	G	-
	Joint bids	-	-	-	-

# Defining the market

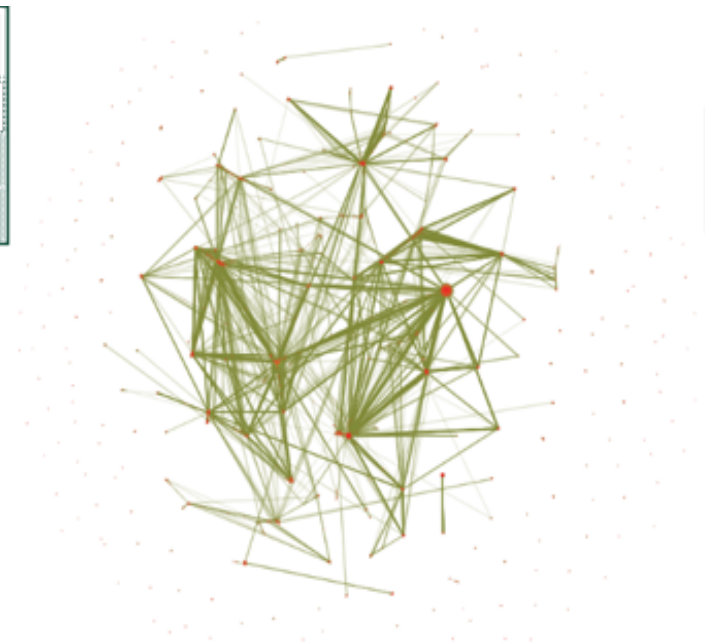
3 generic approaches exist, building on

1. Statistical nomenclatures
2. Co-bidding networks
3. Hybrid (combining 1+2)

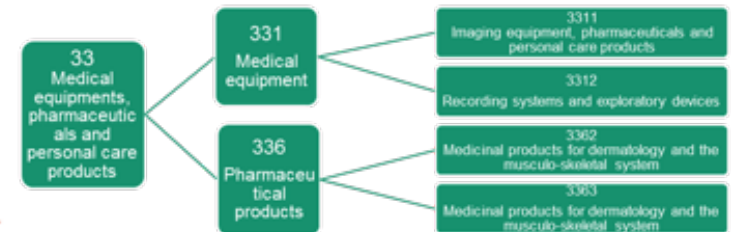
## Nomenclatures

Product \ Geography	CPV code, below threshold	CPV code, above threshold
National	Country-wide	Country-wide
Regional	NUTS1	
Local	NUTS3	

## Co-bidding networks



## Hybrid





# Developing relevant indicators for collusion

Analysing the data available

Defining the conceptual framework

Proposing a set of indicators

Testing internal validity

Testing external validity

Indicator group	Indicator name	Indicator definition
Prices	Bidders having the same bid price	Tenders with at least two equal submitted bid prices are marked.
	Difference between lowest and second lowest bid prices	Relative difference between lowest and second lowest bid price per tender
	Relative standard deviation of bid prices	Relative standard deviation of bid prices per tender
	Bid price range	Range of submitted bid prices per tender
	Benford's law	Whether submitted bid prices follow Benford-law
Bidding patterns	Winning probability	Nr. of tenders won / Nr. of bids submitted by company in period
	Cyclical winning	Bidders have autocorrelated winning patterns
	Missing bidders	a. Change in the average number of bids in a given market b. Change in the number of NUTS regions a company bids in
	Superfluous losing bidders – network analysis	Superfluous losing bidders are those bidders which only submit losing bids in the presence of one dominant company extracting the rents of collusion
Market structure	Concentrated market structure	Indication of increasing market concentration by measures of market structure
	Stable market structure	Standard deviation between time periods of measures of market structure

## Basic indicator validity - if there is no corruption, there should be no substantial financial impact on companies

Indicator name	Collusion proxy formulation	Before tax profit rate	Before tax profit rate growth	Before tax profit growth	Level of aggregation	Control group
Bidders having the same bid price	binary	<b>15.6%**</b>	0.0	0.0	NUTS-3 market	Same product in different regions (without equal bid prices)
Difference between the lowest and second lowest bid	outliers: bottom 5%	-5.25%	<b>15.5%***</b>	<b>16.0%***</b>	Company	Companies without outlier bid price difference
Relative standard deviation of prices	outliers: bottom 5%	- 11.8%	<b>-60%***</b>	- 10.9%	Company	Companies without outlier average relative standard deviation
Bid price range	outliers: bottom 25%	-2.9%	<b>17%***</b>	<b>13.7%***</b>	Company	Companies without outlier price range
Benford's law				data quality is insufficient		
Winning probability	outliers: top 25%	<b>37%***</b>	<b>18.5%***</b>	9.5%*	Company	Companies without outlier winning probability
Bid rotation				data quality is insufficient		
Missing bidders I (avg. number of bid per tender)	outliers: bottom 25%	<b>12.6%***</b>	- 1.3%	4.50%	NUTS-3 market	NUTS-3 regions without outlier or relatively high bidder per tender decrease
Missing bidders II (avg. number of NUTS regions)	outliers: companies with decreasing NUTS scope	-6.5%	15.9%*	9.6%*	Company	Companies with constant/increasing NUTS scope
Superfluous losing bidders (cut-points)	binary	<b>-39.3%**</b>	41.4%	40%	Company	Companies not in a cut-point position
Superfluous losing bidders (nearest neighbours)	outliers: top 5%	<b>-50.8%**</b>	4.2%	14%	Company	Companies without outlier nearest neighbour degree
Concentrated market structure				data quality is insufficient		
Stable market structure				data quality is insufficient		

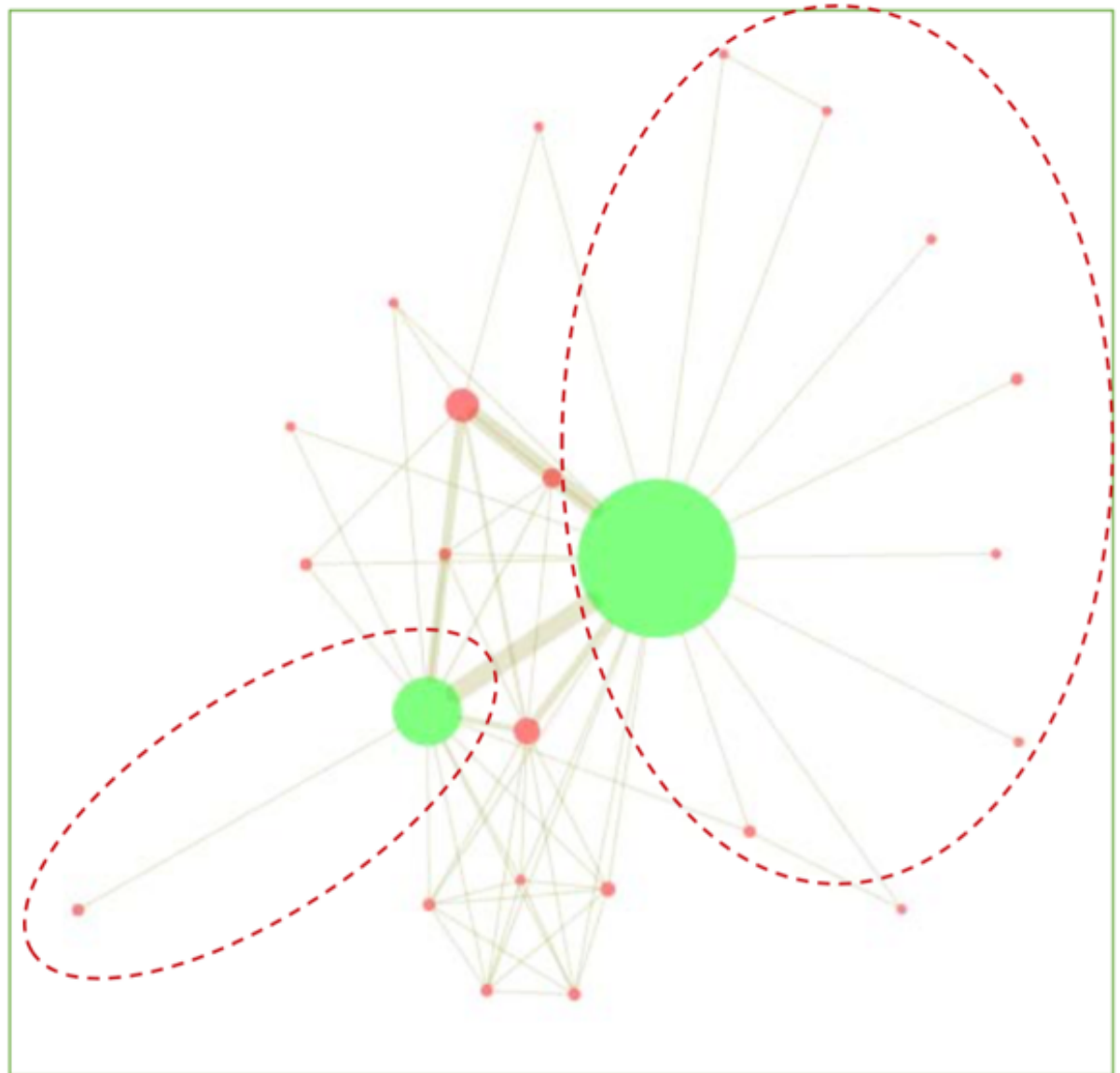
# Elementary indicators

An example co-bidding network

2009-2014

Sweden

Cut-points in green



# Proposed set of indicators

- High risk tenders have significantly smaller relative price range (-13.7%) than low risk tenders
- Average profit rate of companies active in high risk markets is considerably higher: 11.3 % vs 3.6 % (significant and 1%)

Collusion type	Indicator
A	<ul style="list-style-type: none"> <li>• Missing bidders</li> <li>• Concentrated market structure</li> <li>• Winning probability</li> </ul>
B	<ul style="list-style-type: none"> <li>• Superfluous losing bidders</li> <li>• Concentrated market structure</li> <li>• Winning probability</li> <li>• Diff between lowest and second lowest bids</li> <li>• Range of offer prices</li> <li>• Benford's law</li> </ul>
C	<ul style="list-style-type: none"> <li>• Superfluous losing bidders</li> <li>• Concentrated market structure</li> <li>• Winning probability</li> <li>• Diff between lowest and second lowest bids</li> <li>• Range of offer prices</li> </ul>
D	<ul style="list-style-type: none"> <li>• Missing bidders</li> <li>• Stable market structure</li> </ul>
E	<ul style="list-style-type: none"> <li>• Superfluous losing bid</li> <li>• Stable market structure</li> <li>• Difference between lowest and second lowest bids</li> <li>• Relative range of offer</li> <li>• Benford's law</li> </ul>
F	<ul style="list-style-type: none"> <li>• Missing bidders</li> <li>• Stable market structure</li> <li>• Cyclical winning</li> </ul>
G	<ul style="list-style-type: none"> <li>• Superfluous losing bid</li> <li>• Stable market structure</li> <li>• Difference between lowest and second lowest bids</li> <li>• Relative range of offer prices</li> </ul>