Assessing the potential for detecting collusion in Swedish public procurement

Av Mihály Fazekas och Bence Tóth på uppdrag av Konkurrensverket

UPPDRAGSFORSKNINGSRAPPORT 2016:3

KONKURRENSVERKET
Swedish Competition Authority

Konkurrensverkets uppdragsforskningsrapport oktober 2016
Utredare: Mihály Fazekas och Bence Tóth
ISSN-nr 1652-8069
Foto: Matton Images

Preface

I Konkurrensverkets uppdrag ingår att främja forskning på konkurrens- och upphandlingsområdet.

Konkurrensverket har gett Dr. Mihály Fazekas samt Dr. Bence Tóth vid Universitetet i Cambridge i uppdrag att, inom ramen för Konkurrensverkets uppdragsforskning, analysera databaser med uppgifter om annonserade offentliga upphandlingar i Sverige och EU i syfte att hitta indikatorer på otillåtna samarbeten. Utöver detta har en viktig uppgift varit att analysera vilka uppgifter som finns tillgängliga för detta ändamål i olika databaser.

Till projektet har knutits en referensgrupp bestående av professor Giancarlo Spagnolo från Handelshögskolan i Stockholm – SITE. Från Konkurrensverket har Li Feng, Martina Hansson, Stefan Jönsson, Göran Karreskog, Diego Gomez, Arvid Fredenberg samt Joakim Wallenklint deltagit.

Författarna ansvarar själva för alla slutsatser och bedömningar i rapporten.

Stockholm, oktober 2016

Dan Sjöblom
Generaldirektör

Contents

Summary

This research report provides a detailed discussion of three fundamental topics relevant for building a public procurement system in Sweden which supports both government accountability, and monitoring the risks of collusion. First, it offers a comparison of the current Swedish data system to a set of European best practices in terms of public procurement data scope, depth, and quality using data collected by the EU-funded DIGIWHIST research project. Second, it provides an assessment of the public procurement database currently available in Sweden, and puts forward recommendations for improvements supporting better collusion screens. Third, it develops and validates a set of collusion risk indicators based on prior experience with such indicators globally and statistical analysis of Swedish public procurement data.

The Swedish public procurement data system has a number of key strengths. First, it purports a relatively well-implemented system of organizational IDs which is a precondition for analysing organisational behaviour such as inter-bidder collusion. Second, it records bid prices not only for winners, but also for losing bidders which is a comparatively rare, but important practice for in-depth collusion detection analysis. There are however several weaknesses of the Swedish public procurement data system. First and foremost, there is no central platform publishing all tender announcements which are regulated by national public procurement law; availability of which would drastically decrease the cost of obtaining bidding information. Second, contrary to the usual practice across Europe, there is little public control over data capture processes and data quality leading to one of the most fragmentary and incomplete public procurement datasets in the EU.

The Swedish public procurement database contains high quality data content for a range of key variables such as information on the contracting authority, or essential procedural characteristics (e.g. procedure type, framework agreements). However, a number of variables of central importance for investigating collusion risks have a very high missing rate such as name of winning bidder or bid prices. The categorisation used for identifying the products procured (CPV codes) are merely listed without any indication as to which product represents higher or lower importance in the whole contract, making market definitions problematic for bundled products. On a structural level, the Swedish public procurement database records information on the tender level rather than per lot, which limits the use of the whole database as the actual transaction between the buyer and seller is not precisely identified.

The analysis calculated and tested 13 collusion risk indicators each of which is tied to different forms of collusive schemes such as those splitting up markets by abstaining from bidding or mimicking competition through submitting losing bids. Most of these indicators reveal sufficient variability across tenders, bidders, and markets to use them for discriminating between high and low risk activities. In

addition, many of them also pass basic validity tests, such as crossing company collusion risk scores with profitability measures: indicators of bid price distributions such as bidders submitting the same prices or prices really close to each other, bidders' winning probability, missing bidders, and superfluous losing bidders. Unfortunately, some key indicators related to market structure such as market concentration cannot be reliably calculated due to missing contract value information. A demonstrative example of one high risk market is discussed in detail in order to show how individual collusion risk indicators can be used in combination resulting in a simple 'traffic light system': only those markets are considered as high risk which display multiple and consistent risk patterns at the same time.

The report concludes with two sets of policy recommendations, one outlining which reforms could lead to a public procurement data system better equipped to support government accountability and quantitative cartel detection needs; while the other discusses how to make best use of the current data system for collusion detection purposes.

Sammanfattning

Denna forskningsstudie har haft till syfte att ta fram indikatorer som kan identifiera anbudskarteller i databaser med uppgifter om annonserade offentliga upphandlingar i Sverige och EU. Det har även varit viktigt att identifiera vilka uppgifter som finns och inte finns tillgängliga för detta ändamål i databaserna.

Rapporten innehåller en jämförelse av hur upphandlingsdata samlas in i Sverige jämfört med den datainsamling som existerar inom ramen för det EU-finansierade projektet DIGIWHIST. Utifrån detta görs en övergripande bedömning av det svenska systemet för att samla in data om offentlig upphandling och tar fram förslag på förbättringar som skulle underlätta upptäckten av otillåtna samarbeten. Framförallt har det handlat om att utveckla och kvalitetssäkra ett flertal indikatorer som kan identifiera en ökad risk för anbudskarteller. Dessa indikatorer är framtagna utifrån internationella erfarenheter av liknande indikatorer samt en analys av insamlad data som finns tillgänglig om svensk offentlig upphandling.

Det svenska systemet för insamling av data om offentlig upphandling har i jämförelse med andra länder ett antal viktiga fördelar. Genom att ange organisationsnummer för företagen som deltar i anbuden skapar man en möjlighet att kunna identifiera de deltagande aktörerna, vilket är en grundläggande förutsättning för att identifiera otillåtna samarbeten. För det andra registreras anbudspriser inte bara för vinnaren utan även för anbudsgivare som förlorat, vilket jämförelsevis är väldigt sällsynt men viktigt då detta möjliggör för en djupare och mer utökad analys av anbudsdata.

Det finns även ett flertal nackdelar med hur man samlar in upphandlingsdata i Sverige. Framförallt handlar det om att det saknas en central plattform som offentliggör alla annonseringar av offentliga upphandlingar, vilket drastiskt skulle minska kostnaden för inhämtning av anbudsinformation. Dessutom har man i Sverige, till skillnad från normal praxis i Europa, en relativt liten statlig kontroll över processerna för datainsamling och önskvärd datakvalitet. Detta är de viktigaste faktorerna till att Sverige har en mer fragmentarisk och ofullständig tillgång till data om offentliga upphandlingar i jämförelse med resten av EU.

Analysen av de svenska databaserna över offentliga upphandlingar visar att det finns en hög kvalitet i det data som samlats in om upphandlande myndigheter samt om väsentliga egenskaper som t.ex. typen av förfarande, ramavtal etc. Det finns emellertid ett antal variabler som är av central betydelse för identifiering av riskerna för otillåtna samarbeten som ofta saknas, t.ex. namn på vinnande anbudsgivare och anbudspriser. Det är även så att den kategorisering som används för att klassificera de upphandlade produkterna (CPV-koder) listas utan någon uppgift om vilken kategorikod som står för huvuddelen av kontraktet, vilket försvårar marknadsdefinitioner för kombinationsprodukter. I Sverige registrerar man dessutom information på upphandlingsnivå snarare än per delkontrakt, vilket

begränsar användningen av hela databasen eftersom den faktiska transaktionen mellan köpare och säljare inte kan identifieras till sin helhet.

Denna studie har tagit fram och testat tretton indikatorer möjliga att använda för att identifiera otillåtna samarbeten. Dessa indikatorer är var och en knutna till olika former av otillåtna samarbeten, t ex uppdelning av marknader genom att någon avstår från att lämna anbud eller ge sken av fungerande konkurrens genom att någon lämnar in förlorande anbud. De flesta av dessa indikatorer uppvisar tillräckligt stor variation mellan olika anbud, anbudsgivare och marknader för att de ska kunna användas för att skilja mellan verksamheter med låg eller hög risk för otillåtna samarbeten. De flesta av dessa indikatorer klarar även grundläggande validitetstester som t ex jämförelser mellan gradering av ett företags risk att ägna sig åt otillåtna samarbeten och eventuella vinstmått. Tyvärr kan vissa nyckelindikatorer kopplade till marknadsstruktur, t ex marknadskoncentration, inte beräknas på ett tillförlitligt sätt på grund av att information om kontraktsvärde saknas.

Rapporten lyfter även upp ett beskrivande exempel på en högriskmarknad som diskuteras i detalj för att visa hur enskilda indikatorer kan användas i kombination för att skapa ett enkelt ”trafikljussystem” för att identifiera otillåtna samarbeten. De marknader som betraktas som högriskmarknader är de som uppvisar flera och konsekventa riskmönster samtidigt.

Rapporten avslutas med ett par policyrekommendationer: den ena beskriver vilka reformer som behövs för att möjliggöra ett system för datainsamling om offentlig upphandling som är bättre rustat för att stödja regeringens redovisningsskyldighet och behovet av kvantitativ kartelldetektion, medan den andra diskuterar hur det nuvarande systemet på bästa sätt skulle kunna användas för att upptäcka anbudskarteller.

Introduction

Government spending through public procurement tenders comprises a considerable part of overall public expenditure. The exact amount varies by country (29% of overall government expenditures in case of OECD counties); Sweden spends approximately 31% of public money through public procurement tenders.[1] This is approximately 16.5% of the overall GDP[2], adding up a rather significant part of aggregated demand. Therefore, it is of utmost importance to have a good understanding of how these markets work in terms of competition. However, in order to understand market dynamics, high quality data is crucial.

Public procurement markets can be regarded as more vulnerable to collusive behaviour compared to ordinary markets (see e.g. Tóth et al. 2014). First, as there is often no quantity adjustment on these markets – i.e. the demand side is more inelastic – the potential gains of coordinated behaviour can be significantly higher. Second, most of the public procurement markets are transparent to a great extent, which significantly lowers the cost of monitoring collusive deals. Third, the transparency effect is further amplified in case of markets with frequent contracting, which makes punishment and deterrence significantly easier. Consequently, the expected risks of sustainable collusive rings are higher on public procurement markets.

Despite, the obvious collusion risks in public procurement markets, using and systematically analysing administrative data is still in its infancy. Countries often simply lack good enough data for analysing bidding patterns, or comprehensive measurement approaches for identifying collusion risks. Motivated by the above explicated reasons, this research report aims to give a detailed discussion of the following three topics:

1) Comparison of the current Swedish data system to a set of international best practices of public procurement data scope, depth, and quality by using data collected by the EU-funded DIGIWHIST research project.
2) Assessment of the public procurement database currently available in Sweden and recommendation for potential improvements supporting better collusion screens.
3) Developing and analysing collusion risk indicators on the Swedish public procurement database which involves: i) defining indicators, ii) exploring the proposed indicators and iii) conducting elementary validity tests for each indicator.

The structure of this report follows the above outlined main topics. This research is based on previous research conducted by the Government Transparency Institute

---

[1] In case of year 2013, see OECD's Government at Glance 2015 (https://www.oecd.org/gov/Sweden.pdf)

[2] Base on OECD's Government at Glance 2015 (https://www.oecd.org/gov/Sweden.pdf)

on public procurement collusion indicators in Hungary, the EU-funded DIGIWHIST project, which collects and analyse public procurement data in 35 European countries, and the Swedish contract-level public procurement data provided by Visma Opic for years 2009-2014.

1 International best practices: data scope, depth, and quality

Chapter 1 contains a comparative analysis of public procurement data availability across European countries in terms of database scope, depth, and quality. It focuses on transparency requirements of national public procurement regimes across Europe and how these requirements are realized in practice.

First, an internationally recognised best practice is briefly described by invoking the Open Contracting Data Standard which has emerged as a globally reliable standard to which an increasing number of national governments sign up. In addition, the DIGIWHIST data template is discussed more in detail as it represents an adaptation of the global standard to the European context.

Second, a comparative analysis of European public procurement data systems is carried out by measuring country systems against the best practice identified, with a particular focus on the Swedish case. Dimensions of international comparisons include for example, national thresholds for announcing public procurement tenders in national public procurement portals, amount of information reported (e.g. reporting bidders rather than the winner only), and the amount of mandatory information missing from published procurement announcements.

Third, based on the international comparisons a small set of key strengths and weaknesses of the Swedish public procurement system will be selected and discussed in detail. Using the available international evidence and theoretical claims the likely costs and benefits of these features will be outlined.

1.1 International standards

The digital age implies at least two novel opportunities for pursuing good governance and safeguarding effective competition. First, by digitalizing and disclosing how the public sector operates in different areas, including public procurement, governments can become more accountable. Second, the availability of big data allows for gaining better insights and giving more effective responses to understanding problems such as hidden collusive rings. However, one needs to have the appropriate data framework in order to be able to exploit the opportunities offered by big data. Therefore, this section aims to give a concise overview on how governmental data can be stored, by discussing the Open Contracting Data Standard (OCDS) and DIGIWHIST[3] data templates.

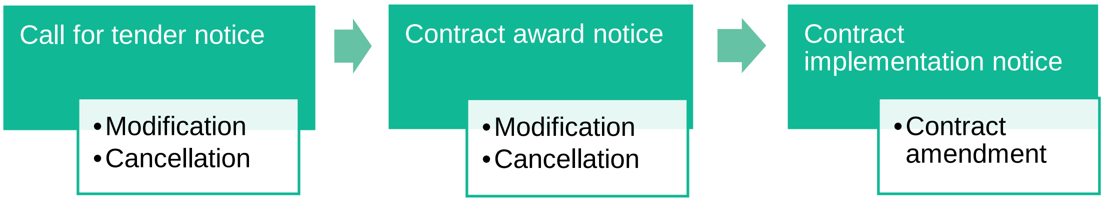
---

[3] The Digital Whistleblower: Fiscal Transparency, Risk Assessment and Impact of Good Governance Policies Assessed (http://digiwhist.eu/)

Before going into the details of what international data templates contain, it is important to have a quick look at how public procurement procedures are run in general. The academic and policy literature identifies four major phases of the public procurement process spanning from the identification of organisational needs up until the implementation and conclusion of the contract (Byatt 2001; Piga 2011; Thai 2009; Transparency International 2006): i) needs assessment; ii) process design and document preparation; iii) tender evaluation and award decision; and iv) contract implementation and management. While building indications of collusive behaviour uses data predominantly from the tender evaluation and award decision phase, it is important to emphasize, that data can (and should) be collected on every other phase of the public procurement process in order to fully support evidence-based policy making, and gain a comprehensive overview of policy problems and impacts.

In practice, the scope of public announcements covering a procurement procedure's different phases is limited to three key stages by and large corresponding to the following i) document preparation (and publication), ii) award decision and iii) contract implementation. Each of these phases have a unique set of public announcements: call for tender notice, contract award notice, and contract implementation notice (see Figure 1). While call for tender and contract award announcements are available in all EU countries, announcements on contract implementation are only rarely available.[4] Linked to each stage of public procurement procedures, announcements on tender modification, cancellation or contract amendment are also published in most EU countries (Cingolani et al. 2015). As all analysed data comes from these public documents describing the tendering process, their availability, content and quality determines the list of collusion risk indicators as well as the potential for validity test.

Figure 1 Key public announcements characterising the tendering process according to tender phase (standard documents are highlighted in green, while changes to them are in the white boxes)



The collection and organization of such administrative information is still in its infancy even in developed countries, partly because there is a lack of widely useable frameworks for storing tender information. To plug this gap, the OCDS[5] template aims to enable governments to disclose contracting data in a common model, which covers all stages of the contracting process. Its main building blocks

[4] For a more detailed discussion see 1.2.

[5] http://standard.open-contracting.org/latest/en/

are planning, tender, award, contract, implementation, and amendments. All these blocks consist of the most important aspects of the given stage of government contracting. Planning contains data on the given project's budget, needs assessment or feasibility. The tender object holds information that is usually contained in the 'call for tender' documents (i.e. the announcement commencing a government contract) such as product related items (description, product code), procedure related items (procurement method, award criteria, eligibility, submission details). The award consists of information on the supplier(s), the final price[6], and the contract period. Contract contains data on the exact contract that was concluded between the public organization and the winner company, such as the final price, the implementation period, signature date, contract status. After the contracting phase, implementation and amendment holds information on the actual contract performance stage.

Another internationally applicable data template which is more directly geared towards European public procurement systems is the DIGIWHIST data template. It directly builds on the OCDS framework in order to strengthen the global data movement, but it also incorporates a range of specific European features. There are two important differences between OCDS and DIGIWHIST data templates. First, DIGIWHIST uses a more tailor-made variable list that captures the European characteristics of disclosing information.[7] DIGIWHIST uses more detailed variables on eligibility criteria, details on bidders (e.g. number of valid/invalid bids), while less detailed regarding the planning phase information (it is typically not available publicly) or some technical details (e.g. document descriptions or tender submission method). Second, DIGIWHIST also links information on public procurement actors. That is, both contracting bodies and bidding companies are appended to public procurement data in order to give a more complete picture of all aspects of the contracting process.

Below, we briefly discuss the main features of the DIGIWHIST data standard by variable group.[8] Table 1 gives an overview of the most important variable groups by defining their scope and providing a few examples.[9] While many of these variables can be accessed in several European countries, unfortunately, some of the more important variables are not as widely collected, or disclosed. For example, while contract awards usually contain the winning price, in most of the cases no connection can be made between individual government contracts and actual disbursement data.

---

[6] Final price refers to the winner company's bid price and can be different from the price that is included in the concluded contract, or the actual payments made by the contracting authorities.

[7] As there is considerable EU level public procurement legislation, the scope of collected and disclosed data is rather similar among EU countries even within the national legislation.

[8] For the full datastandard see: http://digiwhist.eu/wp-content/uploads/2016/01/DIGIWHIST_D2-3_submitted.pdf

[9] Note, that while in Table 1.1 variable groups represent themes, data storage follows a somewhat different logic (i.e. relational database structure) in order to optimize data storage and retrieval. Details are outlined in Appendix A.

Beyond the scope and depth of the different variable groups, data quality also hinges upon certain technical elements of public procurement announcements. Most importantly, the usage of unique IDs is essential for building a reliable dataset. First, unique tender and announcement IDs are paramount to tracing information on the same tender as it progresses from the bidding phase into contract award and implementation. Second, using unique identifiers on public procurement actors (companies and contracting authorities) is essential for calculating many crucial indicators such as market shares.

Table 1 Main variable groups (tables) included in the DIGIWHIST database structure by thematic

Variable groups	Description and Variable scope
Tender	**Scope** These variables describe the procured goods/services/works and the procedure at the most general level. Note, that one tender can contain multiple lots, and can lead to multiple contracts. **Key variables** tender/contract ID, title, procedure type, description, product type, bid deadline, documentation (price, deadline, location), product codes, reasons for using restricted/negotiated procedures, whether variants are accepted, personal/economic/technical requirements, approached bidders
Dates	**Scope** These variables cover all important date items across the whole tendering process, so that the sequence of tender events can be fully mapped. **Key variables** Announcement publication dates, deadline for obtaining documentation, bid deadline date, estimated start and completion date, award decision date, cancellation date, contract signature date.
Requirements	**Scope** These variables contain information on requirements for bidders to participate. **Key variables** personal, technical and financial requirements
Documentation	**Scope** These variables contain information on the accessibility of detailed tender documentation. **Key variables** whether the documentation is payable, documentation price, documentation location
Funding	**Scope** These variables contain all funding related information. **Key variables** funding source, funding programme, funding amount/proportion per source
Buyer	**Scope** These variables contain information on buyer organizations (contracting authorities). Optimally, contracting authorities can be also connected to other administrative datasets with a unique ID, hence more detailed data can be assembled (e.g. budgetary and employment information). **Key variables** buyer's name and address, buyer ID, buyer's type and main activity, whether awarded by a group of buyers, whether purchased by a central purchasing authority, purchase for another authority, whether a public buyer, whether sectoral buyer, whether subsidized buyer

Bidder	**Scope** These variables contain information on individual bidder companies. Optimally, companies can be also connected to company registers (i.e. tax IDs), which makes it possible to supplement the data with more detailed administrative, financial and ownership information. **Key variables** bidders' name and address, bidder's ID, whether is it a winner, estimated number of winners (in case of a framework agreement), whether the bidder is part of a consortium, and information on subcontractors.
Price	**Scope** Consists of all relevant price items related to the procurement procedure from offer prices to final contract value after completion. **Key variables** Offer prices of bidders, value of the awarded contract, VAT rate, currency, minimum/maximum amounts (e.g. estimated contract price is often given as a range), unit prices, disbursements (actual payments), and final contract value at completion.
Cancellation/ correction	**Scope** These variables contain detailed cancellation and modification-related information. **Key variables** Cancellation reason, modified text/part, modified tender text – especially product codes and dates
Other information	**Scope** These variables contain variables that describe administrative and legal contact points and events. **Key variables** Variables on tender administrator, supervisor, specifications creator, court proceedings and interventions, details of the appeal body and mediation body, court proceedings
Publication	**Scope** These variables contain information on the source of information. It makes the data traceable and reliable such as allowing for tracking all the modifications made to a signed contract. **Key variables** Publication ID (can be multiple), publication source, publication date, date of the last update

The last point to be made on data standards is related to a somewhat technical, but nevertheless important structural question: the unit of observation. While in practice, one public procurement announcement usually contains information on a single tender, the said tender can refer to multiple lots, leading to multiple contracts. A typical example is when a hospital purchases different drugs; the process is done through one tender (i.e. one published call for tenders and contract award announcement), whereas there are different lots corresponding to the different types of drugs that are purchased from various suppliers. In this case bidding companies can submit bids to each of the lots. Preferably, the unit of observation should correspond to the level of bidding activities rather than any higher level of aggregation; that is the ideal unit of observation is the individual lot also corresponding to the individual contract.

1.2 International data availability and quality assessment

In this section, we discuss the availability and quality of public procurement data across Europe. Both aspects of public procurement data vary greatly across European countries allowing for comparison so that we can draw lessons for policy development, but also making it difficult to directly compare performance. Our analysis of data availability encompasses the i) format of data publication, ii) the scope of data reported, and iii) the depth of public records in terms of capturing all key aspects of procurement tenders. The analysis of data quality across Europe is restricted to comparing the degree of missing mandatory information as defined by EU Directives. Assessment of European countries along these dimensions has been carried out by DIGIWHIST, full results can be accessed at digiwhist.eu, in particular Deliverable 1.1 and EuroPAM.

Data publication format and location

While there is a lot of publicly available information on public procurement tenders, there are only a few readily available structured and analysable public procurement databases across Europe. Data on tenders regulated by the EU Public Procurement Directives (i.e. contracts above the EU thresholds)[10] are available as a downloadable database; however, national public procurement databases are not widely available. Unfortunately, most national public procurement systems publish tender announcements as online texts rather than as a database. This implies that first a database has to be constructed from public records requiring a laborious and lengthy programming work, only then can analysis of public procurement markets commence. The difficulty of database building depends on whether data is available in a central platform and whether data is machine-readable. Two aspects of public procurement data systems we discuss in detail.

Publishing public procurement announcements in a central platform drastically reduces transaction costs for bidding companies and it also makes it easier to construct a public procurement database (Coviello and Mariniello 2014). While tenders falling under the EU Directives are available in one centralized website[11], there are significant differences among national level public procurement portals. Although, most of the European countries do have a centralized platform where all national tender announcements appear, there are several exceptions.[12] First, many of Europe's biggest economies have multiple regional platforms, e.g. Germany, Italy, Spain, France, Belgium, among which some contains overlapping information. Second, there is no official, publicly available platform for national level public procurement tenders in countries such as Sweden or Austria. In Sweden, national public procurement legislation is very open-ended being largely at odds with all national legislative frameworks: the only requirement is the

---

[10] https://open-data.europa.eu/en/data/dataset/ted-1

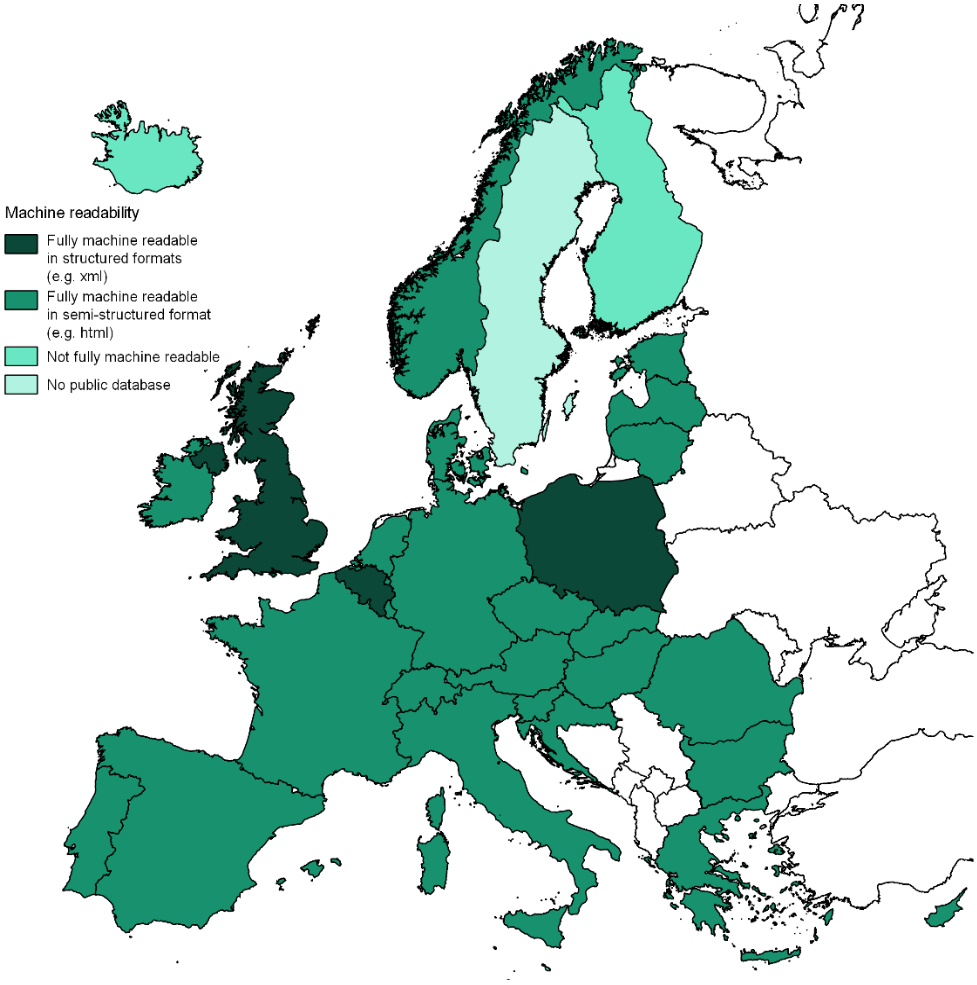
[11] http://ted.europa.eu/TED/main/HomePage.do

[12] See Cingolani et al. (2015).

publication ”in an electronic publication database that is publicly accessible, or publish the procurement in another manner that facilitates effective competition”(pp. 24, SCA 2015). This effectively privatises every national public procurement publication activities in a largely uncontrolled manner: without mandating data quality standards, controls on data reporting processes, requirements of data storage, and ultimately severely limiting the government’s and the public’s ability to gather reliable intelligence about public procurement spending. While a private company, Visma Opic, compiles an extensive database on public procurement tenders following its own data collection standards, its quality is problematic in many respects such as missing information suggesting that rules are not followed sufficiently (for more on data quality see below).

Besides the existence of a central website, where all public procurement announcements are published, the format or ’machine-readability’ of information is also essential for the quantitative analysis of public procurement data. While accessing the data through an application programming interface (API) or direct database download would be the best mode of access, countries mostly disclose structured xml or html/pdf files (Figure 2). While xml files can be simply turned into an analysable database, information from html and pdf files is often hard to extract.

Figure 2 Machine-readability information



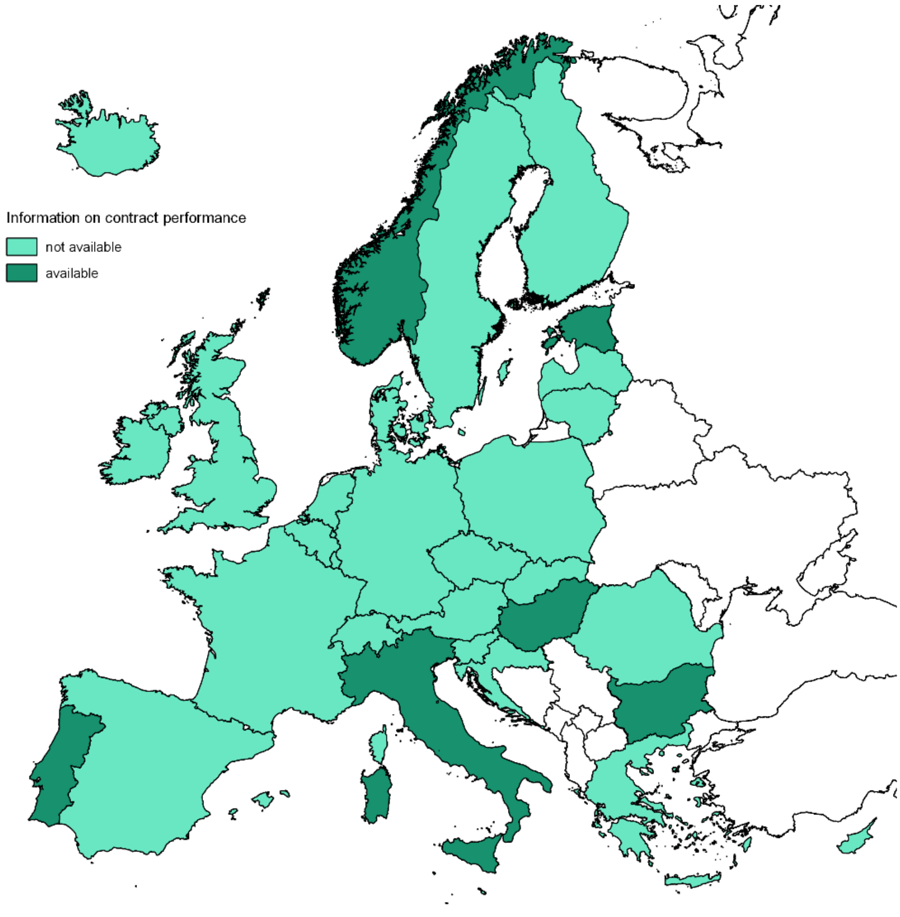
Source: Cingolani et al. (2015)

Data reporting scope

A sufficiently wide database scope is indispensable for meaningful research and for a comprehensive analysis of the whole procurement process. Unfortunately, most European public procurement databases capture only the bidding and bid evaluation phases without any information on contract implementation and often not even on major contract modifications. Furthermore, the portion of centrally reported public procurement activities compared to total public procurement spending is very low in many countries as the mandatory reporting thresholds are very high and historical data is available only for a few years.

Possibly the most important question regarding the administration of government contracting is whether the whole tendering process is described by public announcements or not. Evidence suggests, that increased transparency can lead to higher quality implementation (Lewis-Faupel et al. 2014) and stronger competition as measured by the average bidder number (Center for Global Development 2014). Ideally, every major stage of the public procurement process – i.e. planning, tendering, awarding/contracting and the contract implementation stages - should be publicly announced in order to have a clear picture on how public money is spent. In most European countries, publicity requirements only cover procurement stages up until the contract award announcement. While modifications to public announcements throughout the tendering and contract evaluation/award phase are typically reported, there are only a handful of countries publishing information on the eventually signed contract, its implementation or modifications (see Figure 3). Consequently, no information is available on the signed contract, or on contract completion (i.e. delays, cost-overrun, and quality) in most EU countries (including the above EU level tenders).[13] Announcement publication is more problematic in Sweden than most EU countries: while there is a standardized template used for call for tenders, there is no template for publishing information on contract awards leading to the loss of critical information and awareness.

Figure 3 Information on contract performance



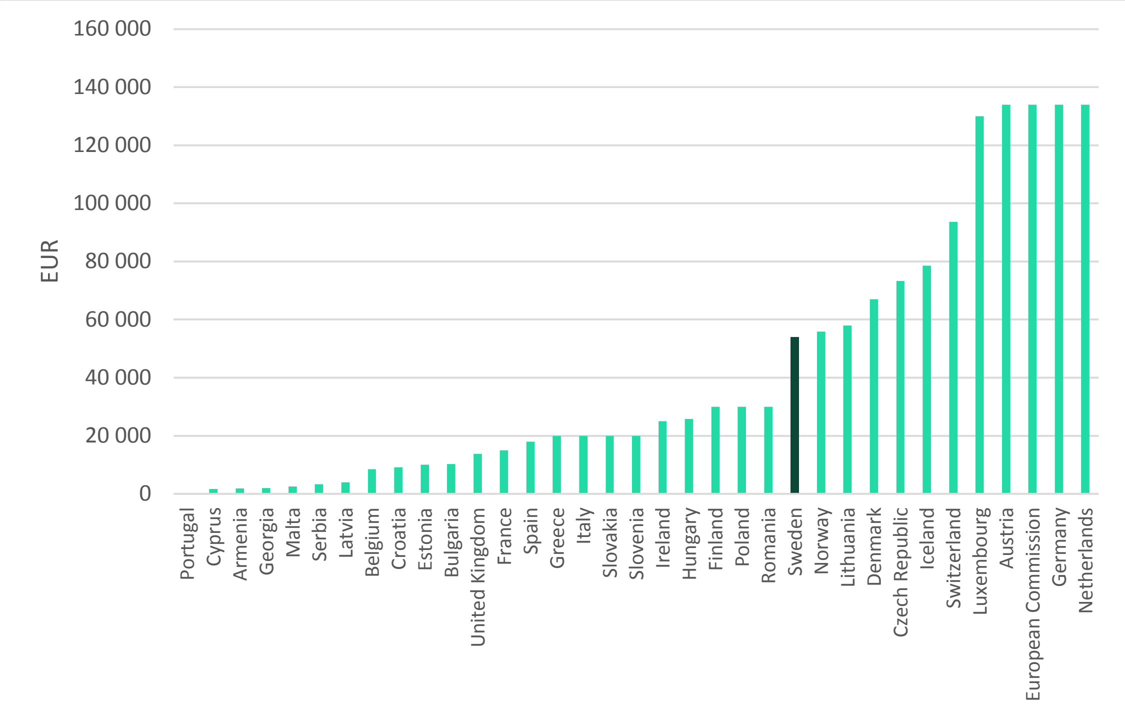
Source: Cingolani et al. (2015). This graph only represents the 'best performance' of each country, for example the publication of contract performance information is no longer mandatory in Hungary or contract implementation data is only available for the largest contracts in Italy.

[13] A related technical question is the number of announcement templates used, as it significantly changes the costs of data extraction (i.e. using several templates complicates both data extraction and cleaning).

A further fundamental aspect of data availability is the minimum national threshold above which no direct award can be made and public announcements are mandatory. While transaction costs pose an obvious rationale for using a certain threshold, enhanced transparency typically increases competition even for lower value tenders (Coviello and Mariniello 2014). Suggesting contacting authorities' desire to avoid additional administrative and transparency requirements posed by the EU Procurement Directives, there is an observed concentration of tenders just below the EU threshold in Sweden (Bobilev et al. 2015).

Figure 4 shows the minimum thresholds above which transparency and a regulated procedures fostering competition apply for goods and services. It is apparent, that there are huge differences among European regulatory systems: many countries do not even have a separate threshold for below EU threshold contracting: Austria, Germany and the Netherlands. Sweden also falls into the relatively higher threshold countries[14]. The minimum thresholds applied for works are significantly higher in most of the cases, but generally follow a similar pattern.

Figure 4 Minimum threshold for publication supplies and services (EUR)



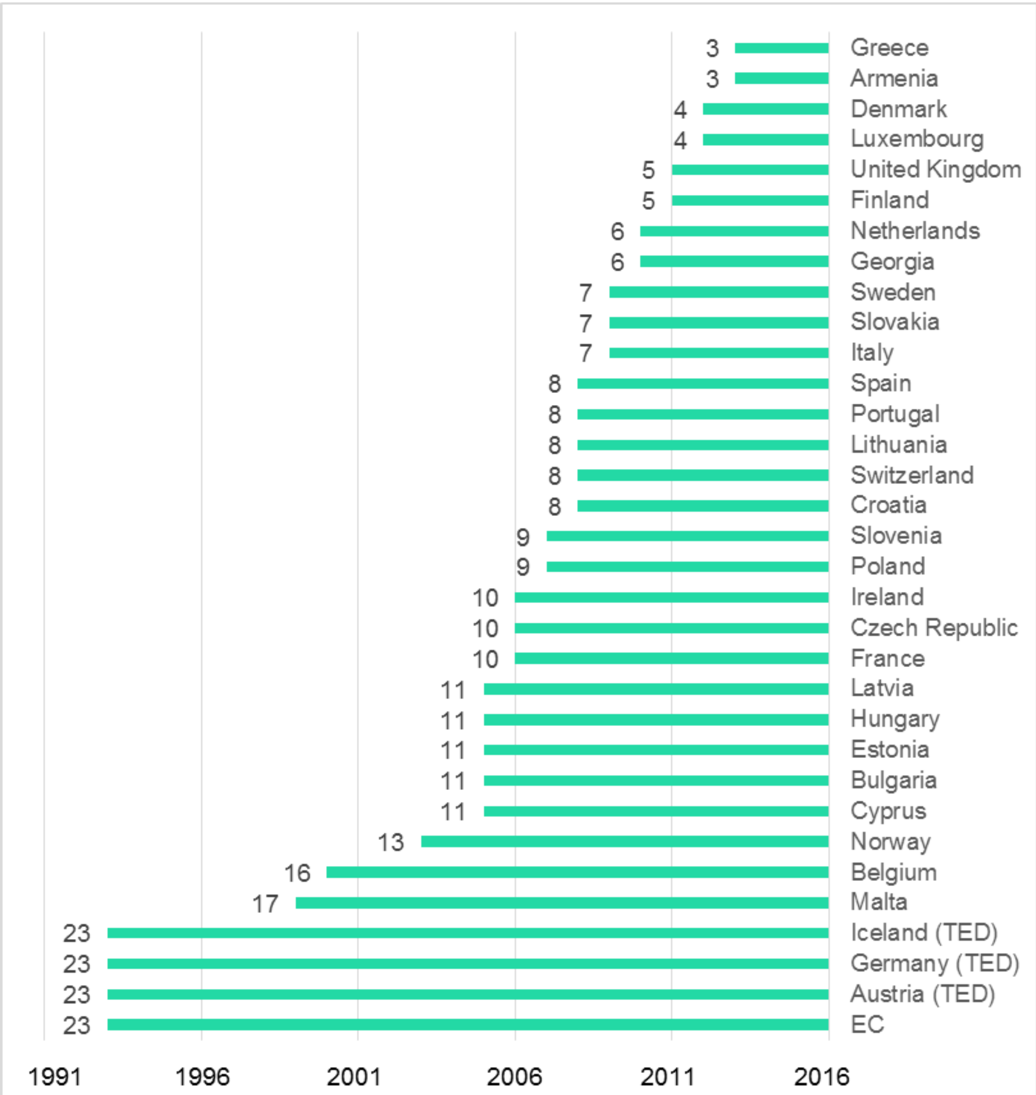
Source: Europam (2016).

In addition to the scope of public procurement data as of 2015, countries also differ significantly in terms of historical data availability. Historical data of at least a couple of years is typically necessary for identifying collusive behaviour as it allows for tracking the risk indicators associated with cartel formation (Tóth et al. 2014). Figure 5 shows that the availability of past data ranges from 3 to 23 years, with the

[14] The maximum threshold for direct awards was raised significantly in 2014 (SCA 2015a) to ca. 505000 SEK.

most extensive historical data often held by less developed Eastern European countries. In this respect, the Swedish data availability (7 years) is slightly below median.[15] Although it must also be noted that in countries with long historical datasets, data quality is often of particularly low quality in earlier years with gradual improvement over time, hence time series amenable for collusion screen calculation are often shorter than depicted. In addition, some countries implemented sweeping changes to their procurement systems and/or the way data is reported further decreasing the availability of time series for detecting collusive rings.

Figure 5 Historical data availability across countries[16]



[15] Although the database received from Visma Opic only covers 2009-2014, we assume that more recent data is also available.

[16] Historical data availability is based on the assessment of the official public procurement portals (central or local), as assessed in Cingolani et al. (2015). For Sweden, historical data availability is based on the Visma Opic database that we use for analysis. For Iceland, there is no historical data published, hence it is not depicted in Figure 5.

Data reporting depth

While assessing the scope of the available public procurement datasets can provide an overall understanding of which market segments and procurement process can be analysed. A similarly important characteristic of public procurement datasets is the depth of information that is stored and published. While countries publish more and more public procurement data, a thorough analysis of bidding patterns requires detailed information both about the procured goods and the bidding processes. If some of the key variables are missing from a procurement data system, it can rule-out the possibility for qualitative collusion risk analysis altogether.

While national procurement data templates tend to follow the templates defined by the EU Procurement Directives, there is still a considerable amount of variation across countries in the types and amounts of information reported (for details see Table 16[17]). The variable availability mapping done by the DIGIWHIST team entailed checking more than 80 variables across each EU jurisdiction (for detailed variable level mapping, see Cingolani et al., 2015). This mapping weighs every variable in the same way, thus for example having winner ID has the same weight as bid disqualification reason. As announcements are not available publicly in Sweden, our assessment is based on Visma Opic's database.[18] Furthermore, we also assessed the additional information that is available from the call for tender documents published at Visma's platform (see an example in Table 17)

Comparing Swedish procurement data to EU countries reveals that it contains fewer variables than the EU average, and in most variable categories. The only exception is the bidding price related information, as individual bid prices are not available in most European countries. Overall, there is ample space for broadening the scope of collected and published public procurement information.

First, while excellent data is collected on bid prices, the estimated price of the tenders (or lots) is not available, which makes it harder to measure the strength of competition (e.g. the value of discounts companies offer) and hence to validate collusion risk measures.

Second, there is no information on whether companies cooperate explicitly in the bidding and implementation periods, i.e. no information on subcontracting and

[17] Table 6.1 is based on an initial assessment of the data scope of national level public procurement announcement templates used on official platforms across Europe from Cingolani et al. (2015). In case of Sweden, the evaluation of variable availability was based on the i) Visma Opic database and ii) examples of calls for tenders published by Visma Opic. Note, that the variable categories slightly differ from the the ones introduced in Table 1.1, the detailed variable list assessment can be found in Cingolani et al. (2015).

[18] Note, the variables that are the basis for this assessment are only a subgroup of the variables included in the OCDS or DIGIWHIST templates, however they should give a fair overview on what kind of data is available in different templates. Another important methodological note is that while in case of Sweden we only used Visma Opic's database and an exemplary call for tender announcement for assessing data availability, in case of other countries, the procurement notices were assessed directly.

consortia. This information is directly relevant for identifying collusion risks, as both can serve as a device for sharing undue profits (see e.g. OECD, 2014).

Third, the lack of information on company location (both winner and participating companies) makes it impossible to analyse the geographical characteristics of competition or the lack of it (see e.g. OECD, 2014).

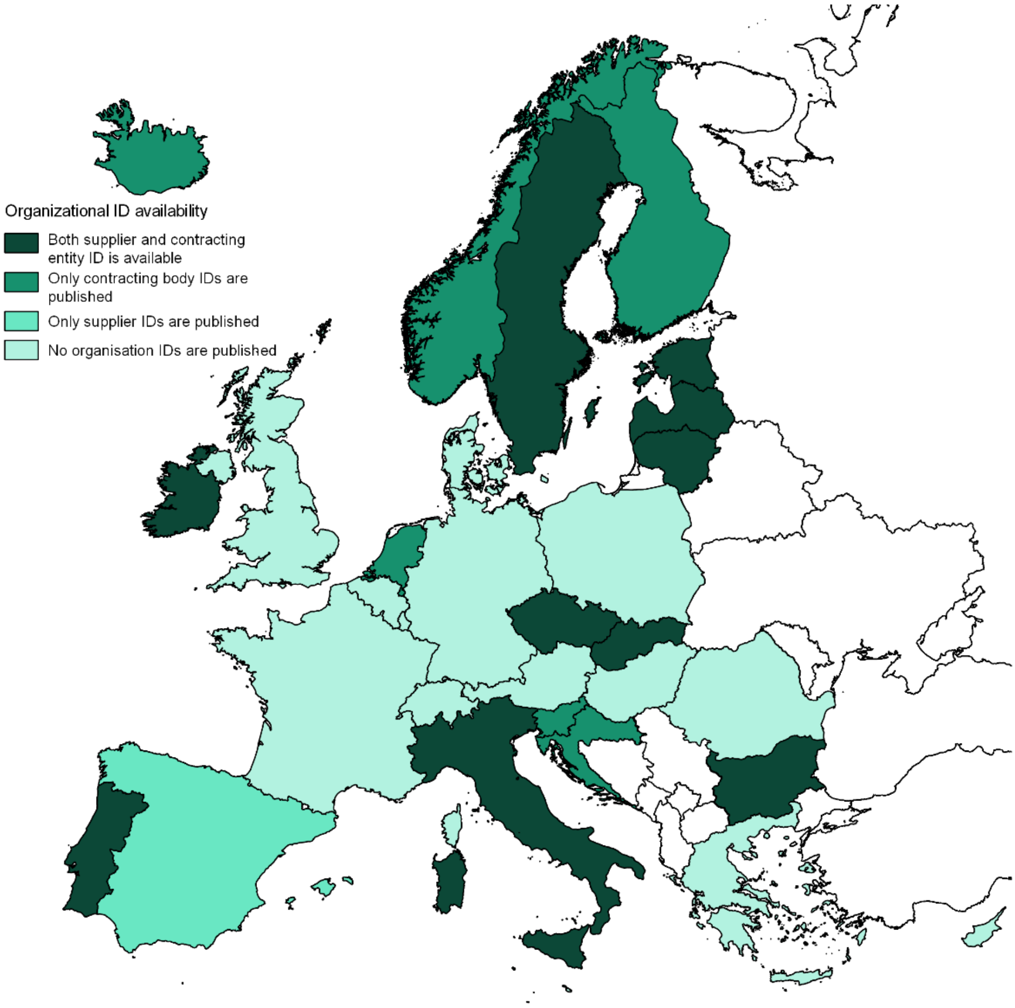
Fourth, several 'supplementary' variables are completely missing such as requirements for participation, funding related information (funding source, programme etc.), tender modifications, cancellation reason, or tender administrators.

Uniquely identifying bidders, suppliers and procuring entities is the precondition for even the most essential market analytics, for instance, calculating market shares. Unfortunately, the majority of European countries do not use unique identifiers, which makes reliably assigning contracts to individual suppliers impossible (Figure 6). For the Swedish public procurement database, Visma Opic recorded bidders and their addresses as input into free text fields, while carrying out the laborious task of matching them to the company register ex post (where it is possible it also used continuous company registration to maintain consistency of records). Such matching is naturally prone to error; for example, the author's prior experience in Hungary has shown just how troublesome free text input can be for the data quality of such a database. In that case, the research team followed a similar approach of matching the often erroneous company names and company addresses to the official company registry (Czibik, Tóth, and Fazekas, 2015). However, many company names and addresses are ambiguous or simply incorrect so that for about 15% of contracts were not assignable to any existing company registry ID.

Closely related to the issue of company IDs is the concern of identifying the final or beneficial owners of bidders. This is crucial for example for assessing whether different legal entities genuinely compete against each other.[19] Even if the relationship between contracts and companies can be perfectly established, the true owners behind companies are often hard to assess. Based on data from 2015, there are only a handful of countries, where beneficial ownership has to be published when placing a bid in a public procurement tender (Hungary, Lithuania, Romania, Slovakia). Although, using beneficial ownership in market analysis is still in its infancy, on the long run this kind of information can be a crucial supplement in assessing market performance, and especially collusion.

[19] For evidence on the the importance of identifying beneficial owners in fraudulent business conduct see e.g. Riccardi and Savona (2013) or de Willebois et al. (2011).

Figure 6 Organizational identifiers



Source: Cingolani et al. (2015) for official public procurement platforms, while Visma Opic's database in case of Sweden.

Data quality: missing information

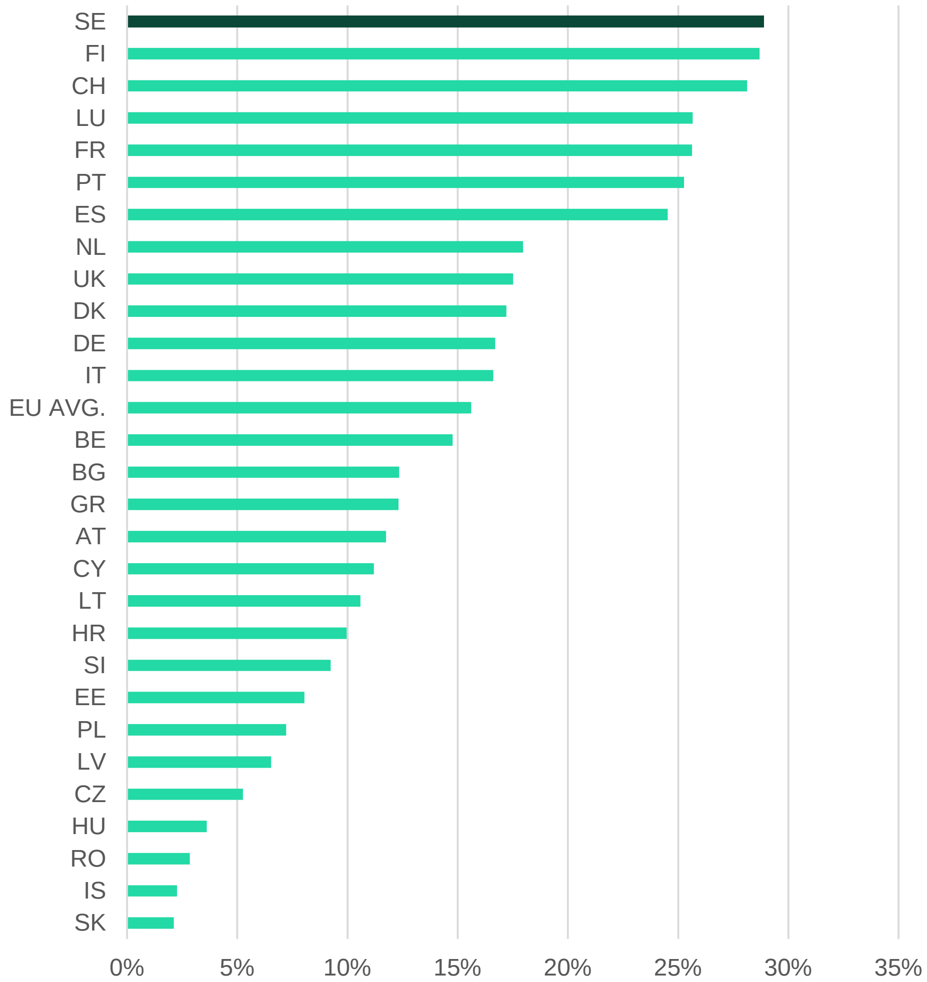
Although the announcement templates can give a general overview about what kind of data is expected to be available, in reality, the high prevalence of missing or erroneous information can deprecate the quality of any analysis, if not render it unusable. Unfortunately, missing data is one of the main and rather extensive problems in public procurement datasets. While in several countries there are comprehensive templates for data reporting, due to poorly designed online platforms and the lack of enforcement, the quality of the resulting administrative data is low (see e.g. Czibik, Tóth, and Fazekas, 2015).

The only dataset which allows for comparing EU countries' public procurement reporting quality is the TED database containing tenders regulated by the EU Procurement Directives. Figure 7 shows the average rate of missing information for above EU threshold tenders published in TED. The information content tracked by our indicator only includes the following fields which are mandatory to include in all contract award announcements: contracting authority's name, address, city, postcode; winner's name, address, town, postcode and country; contract award date, procedure type, contract type (e.g. works, supplies), CPV codes, NUTS codes, whether financed by the EU, total number of bids, award criteria, estimated contract value (lot or tender level), final contract value, and the use of subcontracting.

In addition to surprising variation in administrative error among EU member states, it is also apparent, that more developed and older EU member states tend to have 5-7 times higher rates of missing and erroneous data than new member states Figure 7. Sweden is no exception: missing data represents almost 30% of mandatory fields between 2009 and 2015 which is the highest ratio in Europe. Fields with the highest missing rate in Sweden are estimated contract value (90% missing), total final value of the whole tender (85% missing), final contract value (83% missing), whether subcontractors are used (81% missing). The importance of missing estimated and final values was already highlighted in Bobilev et al. (2015).

While this is only regarding the larger contracts regulated by the EU Procurement Directives and no reliable assessment can be given on below threshold tenders due to the lack of widely available data, it nevertheless provides tentative evidence that the Swedish public procurement data in comparison to other EU countries is of lower quality.

Figure 7 Average administrative error 2009-2015 (missing or erroneous data, based on TED data)



Source: own calculations based on TED data (https://open-data.europa.eu/en/data/dataset/ted-1)

1.3 Key strengths and weaknesses of the Swedish public procurement system

Based on the review of key characteristics of public procurement data systems across Europe, this section identifies the key strengths and weaknesses of the Swedish public procurement data system (Table 2).

The Swedish public procurement data system has a number of key strengths. First, it purports a relatively well-implemented system of organizational IDs which is a precondition for analysing organisational behaviour such as inter-bidder collusion. Second, it records bid prices not only for winners, but also for losing bidders which is a comparatively rare, but important practice for in-depth collusion detection analysis.

However, it is apparent from the above discussion that there are also a number of weaknesses of the Swedish public procurement data system. First and foremost, there is no central platform publishing all tender announcements which are regulated by national public procurement law. Having one central website containing all relevant information on public tenders increases market transparency which has a positive effect on competition. The pro-competitive effects of increased transparency are amply evidenced by academic research (Coviello and Mariniello 2014; Lewis-Faupel et al. 2014). In addition, public information accessible at a low transaction costs also facilitates accountability, that is easily available public procurement data encourages civil society monitoring further supporting good governance of public procurement (Center for Global Development 2014).

Second, as a usual practice across Europe, both in richer and poorer EU member states, central publication websites also serve as vehicles for ensuring uniform data capture processes and safeguard data quality for example publishing guidelines on how to fill in standard forms and how to publish them on the website. As information flows through a single portal (regardless of public or private ownership), quality of submitted data can be checked before it is publicly released supporting reliable public information provision.

Third, the lack of standardized data templates (only calls for tenders have a standard template) makes accurate data collection difficult. One result of this is that the data on the contract award and signature is scattered. Fourth, there is no data on contract implementation and completion, making it hard for the government as well as the public to get a complete picture of public procurement spending and competition.

Fifth, due to the recent increase of the maximum threshold of direct awards, Sweden is in the upper third of European countries, where high value contracts can be awarded without competition and open publication. Sixth, the proportion of missing data in TED is the largest in Sweden among all European countries at least partially reflecting the lack of effective data quality controls and supportive guidelines.

Table 2 Strengths and weaknesses of the Swedish public procurement data

Strengths	Weaknesses
▪ Organization ID's are used both for buyers and suppliers/participants. ▪ In case of available bidding information, the scope is wide: both winners' and participants' bid prices are available in certain cases.	▪ Very limited scope of publicly available data: no official, central website exists. ▪ Lack of uniform data recording and publication standards and no check on data validity. ▪ There is only one template for call for tenders, whereas no other types of information are collected systematically. ▪ Lack of information on contract implementation and completion. ▪ Comparatively high threshold for direct awards. ▪ High percentage of missing data. In case of above threshold data, it is the highest in Europe.

2 Data quality assessment and data cleaning

Chapter 2 provides a comprehensive assessment of the currently available data amenable for analysing collusion risks and puts forward recommendations for future policy development based on the assessment. First, it reviews the quality of the available data, in particular missing and incoherent data. A discussion of feasible and infeasible data cleaning procedures is also offered. Problems regarding database structure and variable storage are set out too. Second, it enumerates the variables which are not collected currently in Sweden, but are useful for detecting collusion. This list is designed to provide insights for future policy development in this area.

2.1 Data quality and cleaning: key variables

As data collection methods applied by Visma Opic define the whole database and the range of potential analyses, we discuss briefly what has been communicated to us by Visma Opic as to data collection and sources. The database covers two types of sources: i) tenders with value between national and EU thresholds and ii) tenders with value above the EU thresholds. The below EU threshold tenders are either published by Visma Opic directly or another local public database from which Visma Opic collects the information. Based on the correspondence with Visma Opic, its database contains all below EU threshold tenders irrespective of where they were published (based on correspondence and SCA 2015). As there is no publication requirement for direct awards below the national threshold, the database only contains such low value tenders if they were voluntarily published.

Part of the data is published and recorded in the templates used for below threshold call for tenders, while due to the lack of a common contract award template for below threshold tenders, Visma collects information on an individual basis on the bid evaluation and contract award phase. As a result, the information available on this phase is more limited. Contracts with value above the EU thresholds are published in the standard forms defined by the EU Directives capturing a wide set of variables both for the call for tenders and the contract award announcements.

In total, there are 116,318 unique tenders in the database between 2009 and 2015 (for 2015, the data is incomplete), roughly 70% belonging to the national regime, and 30% to the EU regime. Furthermore, filtering out cancelled tenders leaves 106,582 successful tenders which form the basis for all subsequent analysis.

While in depth discussion of selected variables can be found below, we first provide a summary of all the variables in the database received from Visma Opic along with our assessment of the variables' suitability for collusion risk analysis (Table 3). Precise figures on the prevalence of missing values can be found in Appendix D.

Table 3 Summary of variable availability

Variable name	Description	Assessment
Buyer's name	Name of the contracting authority	OK
Buyer's department	Department within the contracting authority that implements the tender process	OK
Buyer's ID	Unique ID of the buyer	OK (based on Visma Opic's own procedure)
Buyer's address	Address	Several missing
Buyer's type	Municipality/state agency	OK
Procedure type	Specifies the type of the awarding procedure	OK
Framework	Specifies whether it is a framework agreement or not	OK
Award criteria	Specifies whether the contract is awarded based on lowest price or price+quality	Several missing
Threshold	Specifies whether the tender falls under the EU regulation or not	OK
CPV code	Product code of the given tender	OK (but problems with precision)
NUTS[20] codes	Regional code of contract completion	OK
Document type	Type of document	OK
Status	Specifies the current stage of the tender	OK
Call for tender publication date	Publication date of the call for tender announcement	Only year and month are given in the DB.
Bidding deadline	The final date, until tenders can be submitted	Only year is included
Contract length	Length of the contract implementation period	Majority is missing
Bidder's name	Companies' name	Several missing
Bidder's ID	Unique company IDs	Several missing (based on a closed company list)
Number of bidders[21]	Number of received bids	Several missing
Bid price	Price	Majority is missing
Bid type	Specifies whether the bid was a winning bid or not	Several missing
Bid price unit	Specifies the unit of the bid price	Several missing

Note: "OK" is assigned to a variable if it has negligible proportion of missing data, "Several missing" refers to cases when the missing ratio affects tens of percentages, "Majority missing" refers to cases, where data covers less than 50% of the sample.

Our analysis reveals that data availability follows a divergent pattern variable by variable: while certain types of information are available almost completely, some of the variables have rather high missing ratio or they are not detailed enough for analysis. Almost complete information is available for many key variables

[20] http://ec.europa.eu/eurostat/web/nuts/overview

[21] Information on the number of bidders can be acquired through two channels. There is a separate variable in the database referring directly to the number of total bids, while it is also possible to calculate the total number of different company names (or IDs) by tenders. Comparing the two reveals, that in the overall majority of the cases, the two values are the same. However, in case of ca. 5700 tenders, there is a difference between the two, among which ca. 3900 cases mean a 1-3 difference.

describing the essential contract details: region, publication date (year and month), status of the tender, contracting authority name and type (with ID), procedure type, whether the contract is under or above the EU threshold and CPV codes. As these variables are available for almost all the tenders in the database, we don't discuss them in detail. Below those variables are reviewed in greater depth which are problematic from a data availability perspective. Nevertheless, it is important to note that having no to little missing values in a variable is no guarantee that the value accurately corresponds to the actually values contained in contracts and other legally binding documents. It may well be that the some information is entered and appears to be correct, but there is a typo, say a zero is missing from the contract value; or it is also possible that the contract has been modified during contract implementation, but there was no publication announcing the change rendering the publicly available information outdated.

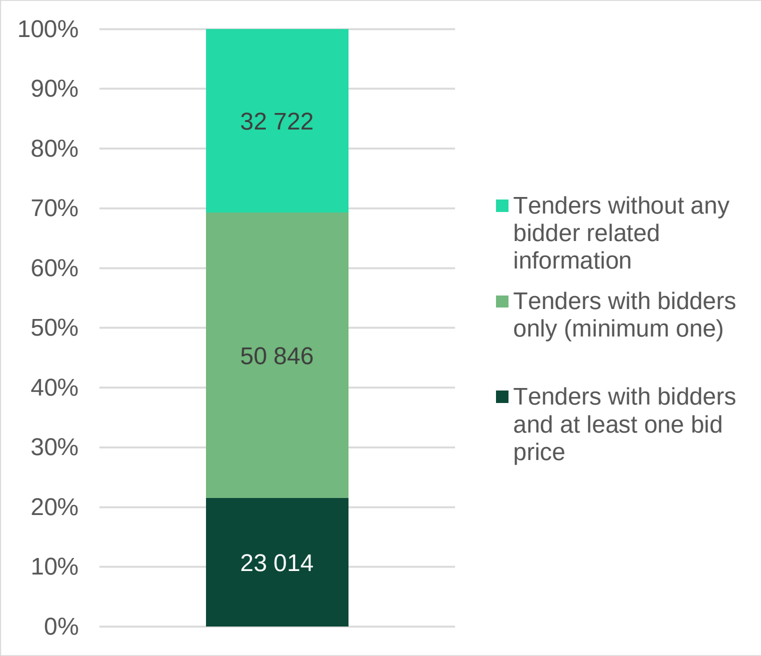
Bidder and price information

First, we discuss the availability of information on bidders, as it is one of the most important types of data for collusion risk analysis. Figure 8 shows that bidder information is only available for slightly less than 70% of the tenders, while price information is only available in around 20% of cases. Tenders with price information are understood as those tenders where the price refers to the total price (there are ca. 5000 additional tenders having unit prices). Tenders with bidder information are understood as those tenders where bidders have a unique ID (there are ca. 3000 tenders having a company name, but no company ID). Therefore, while co-bidding patterns can be investigated relatively widely, suspicious bid pricing can only be analysed to a limited degree.

As section 1.2 discussed, the rate of missing information in Sweden is among the highest in Europe based on TED data. Further reinforcing the picture of weak public procurement publication discipline, the Visma Opic database contains price information on individual bids for approximately 18% of tenders (19,556 tenders have at least one winner and participant bid).[22] While this feature is unique across Europe, as no other country collects individual bid prices, the analysis of bidding price patterns is possible only to a limited degree. The available data allows for demonstrating the power of such information for collusion detection in markets where the data is available to a sufficient degree.

[22] There is a separate variable containing the number of bids. Comparing the number of bidder names and the number of bids variable reveals a difference in 8% of the cases. However, the difference between the two measures of bidder number is only greater the 3 in 2% of the cases.

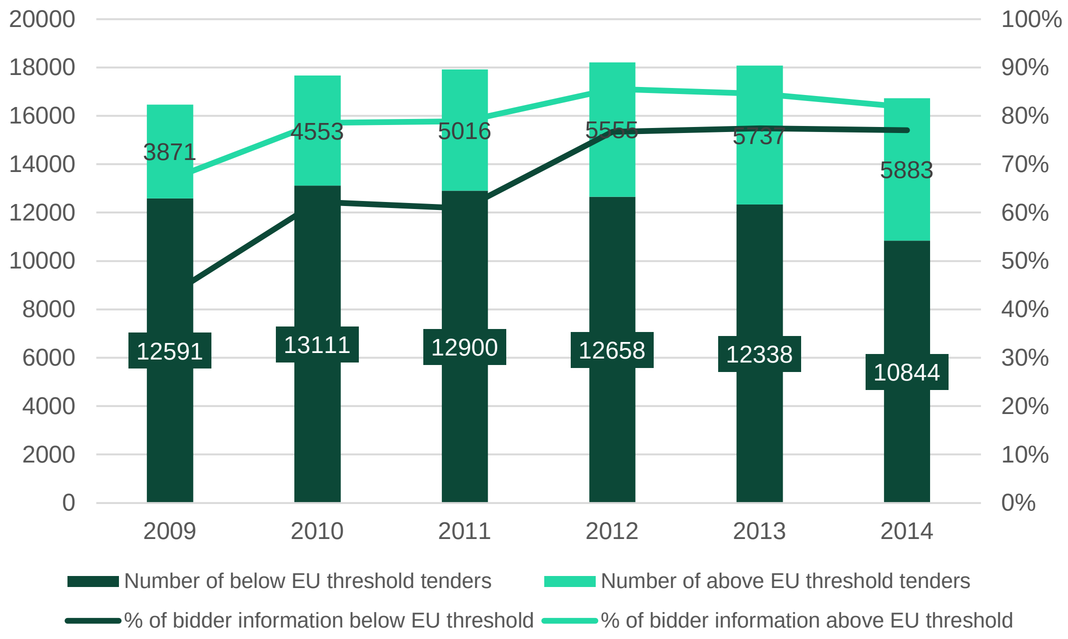
Figure 8 Number of tenders with bidder and price information (N=106,582)



Note: Tenders with price information refer to those tenders, where the price refers to the total price. Tender with bidder information refers to those tenders, where bidders have a unique ID.

Turning to the trend of data volumes, Figure 9 shows that the number of tenders increases somewhat during the investigated period with a marked drop in 2014. This is most likely due to the increased threshold for direct awards (SCA 2015a). As there is no publication requirement for direct awards, transparency significantly decreased in the affected markets (about 2000 tenders became largely intransparent). Nevertheless, the increase of above EU threshold tenders (about 10%), improves overall data quality as they have a somewhat lower ratio of missing values and more extensive set of variables captured. Furthermore, there is also a significant increase (around 30 percentage points) in the availability of bidder information during the period for below EU threshold tenders. Nevertheless, the graph shows that despite the recent developments, there is still a lot space for improving data quality both below and above EU thresholds.

Figure 9 Number of tenders and % of bidder information (n=105,057)

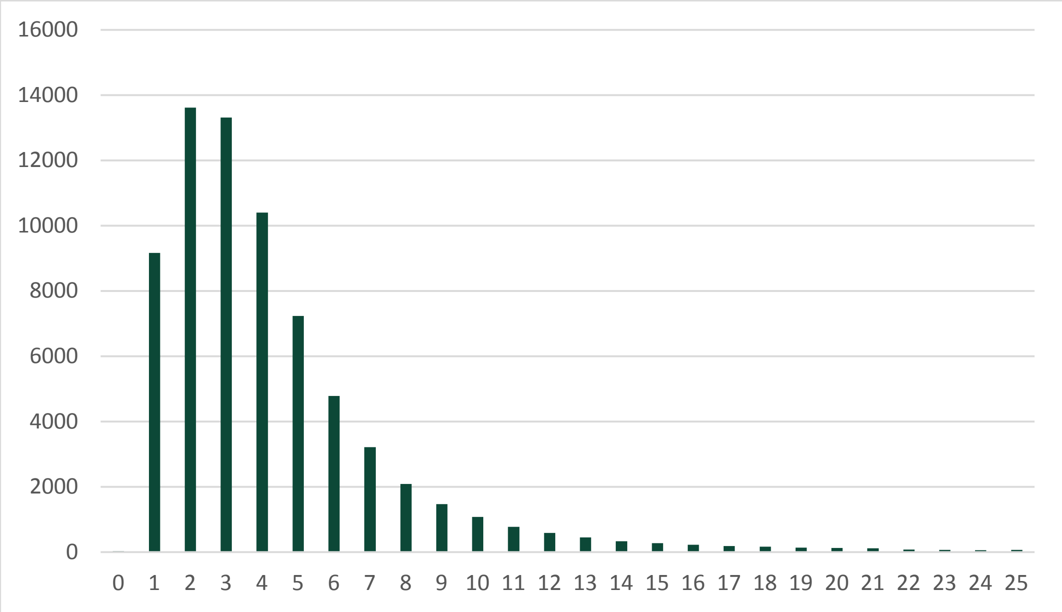


Note: Bidder information refers to the availability of company ID.

Figure 8 and Figure 9 point out that 35-40% of the winning company names and IDs are missing, and winner and bid prices are missing for about 80% of tenders. Unfortunately, no cleaning technique can improve data quality in the case of so extensive missing information. Furthermore, as the database lacks any relevant product quantity information, unit prices cannot be converted into 'total' contract values either. Another potential way of improving data quality would be imputation based on non-price observed contract characteristics such as the main market of the contract. However, based on TED data Varela-Irimia (2014) shows, that this method is not sufficiently reliable.

The number of bids submitted is a key variable for estimating collusion risks, hence it deserves deeper analysis. Figure 10 shows, that slightly more than half the tenders receive between 1 to 3 bids (51%) including a considerable share of tenders with only 1 submitted bid (13%) suggesting a key collusion risk factor already. Unfortunately, this variable is also missing to a considerable degree: 35,799 or 33.6% of published tenders contain no information on the number of submitted bids. Following patterns observed in other European countries, the proportion of tenders with more than 3 bids quickly decreases with tenders having more than 15 bids representing less than 3% of the data represented.

Figure 10 Number of tenders according to the number of bids submitted (N=70,783)

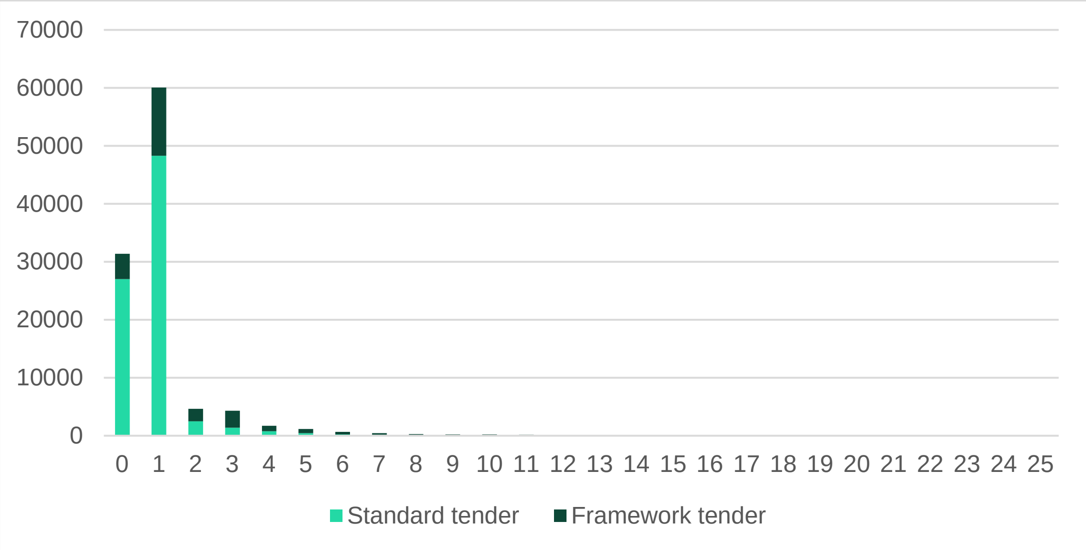


Note: For the sake of visibility, we only included a truncated the sample, tenders with more than 25 submitted bids are not displayed. Number of tenders is based on the separate „number of submitted tenders" variable.

Looking at the number of winner companies per tender reveals a fundamental flaw of the database recorded by Visma Opic (Figure 11). While a standard procurement procedure, not belonging to a framework agreement should have only one winner and one contract associated with it. However, the Visma Opic database has a considerable number of non-framework agreement tenders with multiple winner companies. This indicates that there is a considerable number of tenders with multiple lots without the database precisely recording lots, e.g. separate descriptions of the procured products or number of submitted bids. With insufficient precision in recording lots, collusion risk measurement is rendered less precise for multi-lot contracts. In addition, there is a considerable number of tenders, slightly more than 40,000) without any recorded winning bidder which most likely indicates missing information, but can also imply that the tender was unsuccessful[23].

[23] Of this roughly 40,000 tenders, 9,500 are marked as unsuccessful in the database, still it is unclear whether the remaining tenders without winner company are truly successful or they were recorded in an incomplete manner.

Figure 11 Distribution of tenders according to the number of winning companies



Note: For the sake of visibility, we only included truncated a sample, tenders with more than 25 winner companies are not displayed.

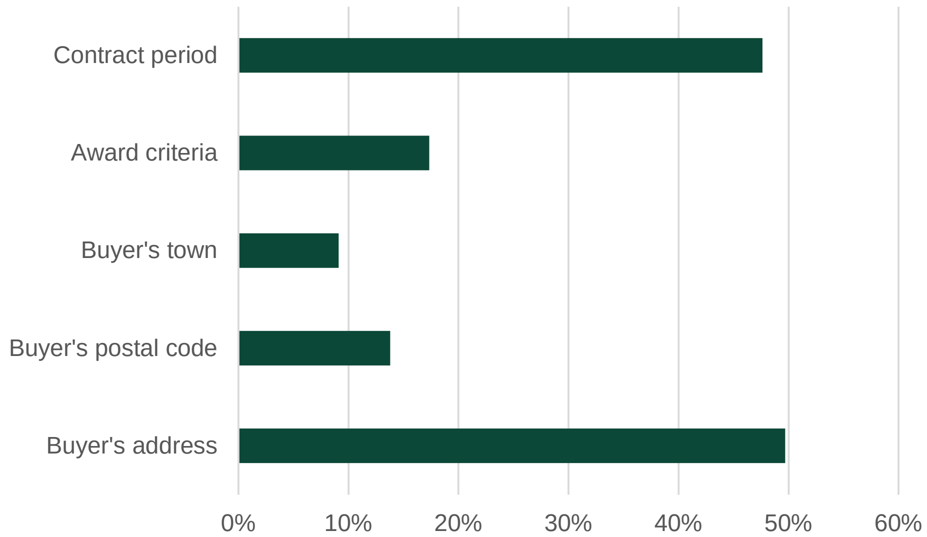
Tender data

As it was already mentioned above, the majority of the basic tender-related variables is almost perfectly available. Most problems are related to the storage and precision of the information which is discussed in section 2.2. Nevertheless, there are two types of variables where significant deficiencies could be identified. First, the contract period, that is the length of planned contract performance between contract signature and contract completion, is only available in 52% of the cases (Figure 12). This is problematic, as the sequencing of contract winning is an important parameter of collusive schemes (Ishii 2009). While this problem can be mitigated by substituting with publication dates of call for tender documents, contract length is still an important aspect when it comes to collusive contract distribution.

Second, information on contracting authorities' location is missing for a substantial portion of tenders (Figure 12): buyer's town and postal code is missing for around 10% of tenders, while street level address is missing for around 50%. The exact location of contracting authorities is a relevant variable more broadly, for example when assessing the regional distribution of public spending. Furthermore, contracting authority location can also be useful for pinpointing contract performance location more precisely when the NUTS code of contract performance is either missing or too broad to be relevant (e.g. NUTS-0 level referring to the whole country). For more information on NUTS codes of contract performance location see section 2.2. For the purposes of detecting collusive bidding, precisely identifying the market where economic operators compete is fundamental. The market definition largely depends on the geographical location of contract performance, especially for smaller contracts.

Third, other variables which are only moderately relevant for collusion risk analysis have also significant missing rates. For example, when the award criteria include price as well as quality elements, the organisation of collusion typically needs to be more complex in order to fully determine the ranking of bids. Therefore missing information on award criteria partially limits our ability to gauge the complexity and types of collusive behaviours.

Figure 12 Percentage of tenders with missing information by variable (N=106,582)



Comparing the Visma and TED datasets

For above threshold tenders in the Visma database, it is possible to compare Visma records to the data available from TED (as extracted by DIGIWHIST). [24] The scope of available information in the two sources differs due to i) information is stored in lots in TED, but not in the Visma database, i.e. greater precision of information in TED (see section "Tender versus lot levels")[25], and ii) TED templates require the publication of a wider set of variables. Some of these additional variables can be

[24] Matching the original European public procurement dataset (TED) and Visma dataset revealed that there are only a few hundred potentially missing tenders between 2010-2013 (based on tender IDs provided by Visma the number of non-matched tenders by each year are the following: 2010: 193, 2011: 89, 2012: 147, 2013: 121). These matching rates should be considered as estimates, missing IDs can account for the discrepancies. In case of year 2009 and 2014, the missing ratio of corresponding IDs were significant. Note, that DIGIWHIST will disclose a somewhat more detailed data on above threshold public procurement tenders compared to the existing TED dataset.

[25] An important note is that comparing above threshold data contained in the Visma database and TED reveals that the number of separate tenders is significantly lower in the official TED database (Table 5). A potential explanation is that some of the lots under a given tender are stored as separate tenders rather than separate lots.

used for more precise collusion risk estimation such as relative price[26], subcontracting or joint bidding[27]. While additional information in TED announcements is conducive to more precise collusion risk estimation, the quality of additional information is similarly low compared to the below EU threshold Visma data which poses problems for the analysis (for an overview Table 4)[28]. In the subsample of matched TED-Visma tenders, only 11% of tenders contain information on estimated price, and 18% on subcontracting. Number of bidders is available with more accuracy; close to 80% of tenders have non-missing values.

Table 4 Number of tenders and contracts and the % availability of different variables in the Swedish TED database, by year ($N_{contracts}$=51,607)

Year	2009	2010	2011	2012	2013	2014
Number of tenders	3127	3431	3592	3916	4199	3926
Number of contracts	7115	7586	7887	9140	9661	10218
Estimated price	19%	15%	14%	12%	7%	4%
Final price	37%	24%	19%	12%	8%	6%
Relative price	15%	10%	9%	5%	3%	2%
Number of bidders	81%	74%	76%	80%	83%	83%
Main CPV code	100%	100%	100%	100%	100%	100%
NUTS	80%	82%	82%	90%	96%	96%
Subcontracting	34%	24%	22%	17%	10%	9%
Award criterion	91%	91%	76%	80%	72%	60%
Procedure type	100%	100%	100%	100%	100%	100%

Note: Percentage values reflect data availability (i.e. ratio of non-missing data) calculated over the total number of contracts in each year.

In terms of discrepancies between the TED and Visma databases, the first apparent difference is that certain non-cancelled tenders are seemingly missing from TED. While most of the Visma tenders have a corresponding ID to the TED announcements (row 3 of Table 5), many of these are not referring to contract award notices (the unique tender ID used in the TED database), but call for tender announcements only. Therefore, it seems that while there are 30,615 above threshold (non-cancelled) tenders in SE between 2009-2014, only 17,238 could be unambiguously connected to the TED database. The difference can be explained by i) missing TED references in the Visma database (especially in case of year 2014) and ii) the actual non-publication of contract award announcements in TED. In case of the second explanation, it is not clear whether the announcements are missing because the EC does not publish them or it doesn't receive them from the Swedish contracting authorities in the first place.

---

[26] The availability of estimated value makes it possible to calculated relative prices (i.e. the ratio between the final price and the estimated price), a simple indicator of measuring whether competition.

[27] Note, that the newly introduced templates include information on whether tenders were awarded to a group of companies (see section V.2.2. at http://simap.ted.europa.eu/documents/10184/99173/EN_F03.pdf)

[28] Similar calculations with narrower data scope are included in Bobilev et al. (2015), smaller differences are due to updates implemented on the source data.

Table 5 Number of tenders in Visma and TED databases, per year

	2009	2010	2011	2012	2013	2014	Total
Total number of tenders in Visma database	4167	4935	5575	6247	6443	6642	34009
Total number of tenders in Visma database (excluding cancelled tenders)	3871	4553	5016	5555	5737	5883	30615
Number of tenders with TED announcement reference (CFT or CA) in Visma database	3,865	4,543	5,001	5,516	5,694	1,126	25,745
Number of tenders without TED announcement reference (CFT or CA) in Visma database	6	10	15	39	43	4,757	4,870
Total number of tenders in TED	3127	3431	3592	3916	4199	3926	22191
Total number of tenders in the matched Visma-TED database	1516	3331	3535	3819	4107	930	17,238

Note: the higher number of tenders in row 3 compared to row 5 indicates that some of the tenders in the Visma database are stored as 'lots' in the TED data. The 4757 tenders without TED reference in 2014 is most probably due to missing data, that would be available at a later stage, hence it does not reflect systematically missing references.

When looking at the variables which are supposedly the same in the Visma and TED datasets, we can compare missing rates as well as the actual values reported. First, it seems that there are no large differences in terms of missing data (Table 6). Although, in most cases, there is less missing data in the Visma database, these differences are most probably due to the extra data collection efforts conducted after publishing information on TED. In terms of the concrete values, while there are some differences, these could be also caused by ex post data improvements rather than systematic differences (recall that Visma manually enriches its public procurement database). Second, comparing the values reported in TED and Visma databases for the same variables and tenders, we find relatively little deviation. For example, number of bids values is the same for 9170 tenders that are most observations in the matched database, while there are a few hundred tenders with different values for award criterion and procedure type. For contract value, the two data sources use different currencies; hence cross-checking is unreliable.

Table 6 Number of tenders with non-missing information in TED and VISMA (N= 11,715)

Variables	TED	VISMA
Number of bids	82%	94%
Contract value	21%	24%
Procedure type	100%	100%
Award criterion	82%	95%

Note: only tenders with one lot (i.e. one winner in case of the VISMA data) are considered.

2.2 Data quality and cleaning: database structure

This section briefly discusses how the current way of storing data can make it difficult to assess both individual tenders and markets in general.

Tender versus lot levels

First, the most important structural problem is that instead of storing information on individual lots and awarded contracts, the level of observation is the tender. This problem was already highlighted previously (e.g. Bobilev et al., 2015). As it was explained in section 1.1, the optimal way of storing public procurement information, is the lot/contract level, as it is the level where bidding is done and contractual relationships are established. Currently, neither the different products, nor the submitted bids can be assigned to individual lots/contracts; in addition, in tenders with multiple lots, bidders and winners cannot be separated either.

This is a crucial problem from the perspective of developing collusion risk indicators as it confounds the information on bidding behaviour. Furthermore, when different lots within one tender correspond to different products or contract values, the identification of a collusive market also becomes harder.[29] As it is discussed in Tóth et al. (2014), in order to identify collusive markets, accurate information is needed on contract location, value, and product type. These enable the approximate definition of markets which confine the activity of collusive rings.[30]

Variable level

First, the most important problem of variable-level data storage concerns how individual bid information is recorded. As it was already mentioned in section 2.1, total price is contained in the same variable as unit prices and there is no separate variable expressing the quantities of procured products. In the absence of such information on quantities, it is not possible to estimate the total contract value. This means that the number of useable bid prices shrinks by roughly 5000 observations (tenders).[31]

Second, a further important variable-level problem is related to CPV coding. As it was already mentioned earlier, good quality product coding is essential for identifying collusive markets. Unfortunately, the structure and accuracy of CPV codes are problematic in the Europe-wide TED dataset: previous research found that more than 20% of the codes are used incorrectly, and 8% of the codes are too general (pp. 11 Commission 2012). Nevertheless, CPV codes remain the only widely usable input for identifying product markets. In addition, information on the type of purchase - services, goods or public works – can also be used to check the validity of CPV codes, albeit on a very high-level of aggregation only.

---

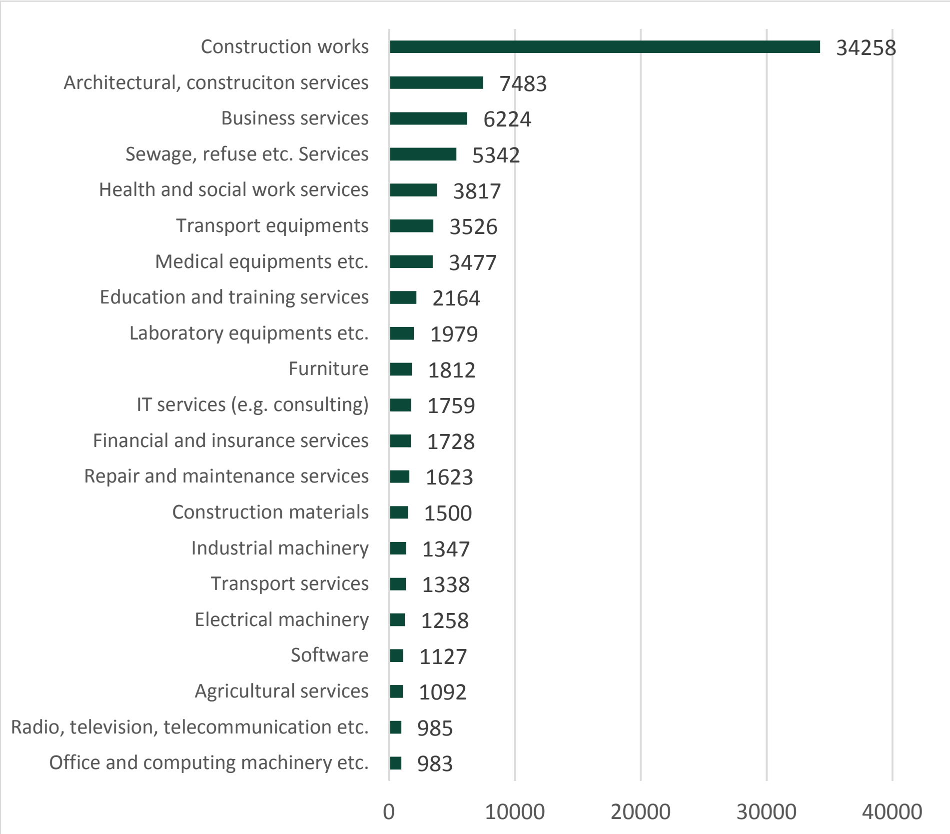
[29] For example if a tender includes both a larger tender on a building refurbishment work, and a smaller on maintenance, it is very probable, that these are two different markets. In case of no lot level information, these different markets cannot be separated with their different lists of bidders etc., which makes the identification of suspicious bidding patterns even more problematic.

[30] It is important to note, that the main goal is to identify markets that are relevant for a particular collusive scheme, which may or may not equal to the market definition used in competition policy.

[31] There are ca. 27000 tender with price information, while only ca. 23000 refers to the total amount.

Currently, the Visma Opic database contains only 'non-hierarchical' CPV codes: no distinction is made between the main product type and the related, additional purchases. Seemingly, more detailed CPV codes should be available, at least a random sample of announcements[32] from Visma Opic's publicly available platform shows, that the distinction of "main" and "additional" codes can be done, similarly to the practice in other European countries. Based on the number of tenders by main 2-digit CPV code categories, it is apparent from Figure 12, that the distribution of tenders are rather skewed: more than 30% of the tenders are related to construction works.

Figure 13 Number of tenders per 2-digit CPV code categories



Note: Only categories with more than 1000 observations were included. The graph uses the discussed 'greatest common divisor' methodology, however, only the most common 2-letter CPV code is used for this graph.

Taking into account both the expected erroneous use of CPV codes and the lack of distinction between main and additional codes, we could still define a suitable approximation of the main product market for each tender. We used a simple 'greatest common divisor' approach to improve data quality in this respect. This implies that we identify the common parts of all CPV codes assigned to each tender. This calculation is repeated for 2-, 3- and 4-digits, hence the most common CPV code can be identified on different levels of product code accuracy. The simplest case is when we use 9091 as the best common code for the following array

[32] See Appendix C: An example announcement.

of CPV codes: 909111; 909112; 909192; 909111; 909112. For CPV codes of different precision, we considered non-conflicting coding as a sufficient condition for identifying the best common CPV code; for the CPV code array: 72; 724 and 7242, the bets common CPV code is 7242. As this method cannot always classify all tenders unambiguously, i.e. there is at least one CPV code conflicting the other codes, we had to resort to using the frequency of CPV codes with non-conflicting common codes. For example, if a tender has 9 CPV codes and 8 can be classified into the 9091 category while only 1 deviates, then we assign CPV code 9091 to the whole tender. In practical terms, for more than 60% of the tenders, the assigned 2-digit CPV codes are the same, i.e. they can be perfectly assigned to a unique (but broad) product market. In the subsequent analysis, we do several types of groupings including 2, 3 and 4-digit CPV codes.

Third, similar to CPV codes, the Visma Opic database stores the contract performance location's NUTS codes only as 'non-hierarchical' data with multiple values describing the same contract. In most cases this reflects different levels of aggregation (recall NUTS codes have 3 hierarchical levels each corresponding to different geographical unit sizes), in many other cases it implies that the purchase delivery takes place in multiple locations or in a location crossing regional borders. In order to define the location of contract performance flexibly, each contract was assigned to the lowest non-conflicting level of NUTS code.

2.3 Data requirements of high-quality collusion risk assessment framework

This section gives a brief overview of which additional variables can be collected to improve the current Swedish public procurement data system supporting higher quality collusion risk assessment. Table 7 gives an overview of those additional indicators which have been successfully used in prior academic work. However, we also want to highlight that improving data quality, especially lowering the rate of missing information, would already greatly increase the scope and accuracy needed for better collusion detection.

First, bidder and bid price-related information are essential for inspecting anti-competitive bidding patterns. First, as collusion often aims to eliminate competition on local markets, hence increasing price levels until the point when the next outside competitor would enter, it is important to know how bidders are allocated geographically. A change in the distribution (e.g. more concentration) of bids in a geographical area, can serve as a tentative evidence for cooperation (SCA 2015b). Furthermore, beneficial ownership can reveal whether there is any kind of relationship conducive to collusion between companies (e.g. Riccardi and Savona (2013) explains a bogus 'Chinese-box' ownership structure hiding the same owner behind companies covering a significant part of a regional public works market in Italy). While information on the actually considered bids are indeed important, submitting erroneous bids is a widely used collusion technique in public procurement markets, hence data on excluded bids would allow for better tracking some specific collusion risks (Detecting Bid Rigging in Public Procurement n.d.).

Second, the availability of price information is crucially important. In order to analyse bidding patterns, it is of utmost importance to know final contract values rather than unit prices only. The importance of investigating bidding patterns in collusive schemes was amply demonstrated by academic research (Abrantes-Metz et al. 2006).

Third, in addition to information on contract values, accurate data on the contract award phase and contract implementation can also deliver valuable insights (Ishii 2009). Academic research underlines the importance of collecting and analysing data on bidding deadlines, award decision dates and contract signature dates.

Fourth, a rather technical, nevertheless important variable is the CPV coding of purchased products. As every collusive scheme aims to eliminate competition on a particular market, without being able to test the indicators on precisely defined markets, it is hard to identify collusive schemes (Tóth et al. 2014). Despite the limitations of using CPV codes (see section 2.2), it is still the most accurate measure of product categories, and supplementing it with information on the type of the procurement (i.e. service/supply/work), and taking into account contract size and region, it is the best way of defining local and national product markets.

Lastly, both the elimination of competition and rent-reallocation are intimately linked to the use of joint-bidding schemes and the prevalence of subcontracting. While both subcontracting and co-bidding can have a positive effect on competition, as joining supplementary knowledge is an important way for companies to become more efficient (Albano, Spagnolo, and Zanza 2009; Estache and Iimi 2008); joint-bidding can also mitigate several operational problems of collusion, such as the adherence to the collusive deal or rent re-allocation among colluding firms as the collusive scheme is enforced by a formal contract. Subcontracting (often through fake subcontracts) is also a widely used technique identified by several court cases[33], and policy guides (Detecting Bid Rigging in Public Procurement n.d.). The investigation of these factors over time, especially their covariation with other risk indicators can reveal patterns that are at odds with open competition.

[33] Examples can be found on different markets, including flour, railway construction and accounting software (Hungarian Competition Authority 2004, 2007, 2008).

Table 7 Additional variables supporting high quality collusion risk indicators

Variable group	Variable	Rationale
Bidder/Bids	Bidder's address	Bidder's address can be used to show whether bid patterns change geographically
	Beneficial owners	Beneficial ownership can reveal that seemingly independent companies are intertwined[34]
	Number of excluded bids	Submitting deliberately erroneous bids is a widely used indication of bid rigging
Price	Final price vs. Unit price	Final price of the contract is essential in order to be able to assess how contract allocation changes in different collusive schemes
	Estimated price	Estimated price can serve as a useful indicator for gauging discounts bidders offer, hence the strength of competition
	Exact time of bid submission	Collusive bidding often involves coordination of bid submission or even multiple submissions by the same person, hence analysing the timing of bid submissions can reveal risks of collusion
Dates	Publication date of contract award	For collusive schemes, the exact timing of contract award can reveal the mechanism of splitting up the market
	Bid submission deadline	Bidding deadlines are indispensible for detecting collusive schemes where timing of bid submission is an important factor
Product	Main CPV code	Separating main and supplementary CPV codes are crucial for identifying product markets.
	Procurement type (service, supply, work)	The type of the purchase can be also an important input for separating different product markets.
Subcontracting	Subcontractor's name and ID	Subcontracting is a commonly used instrument for rent sharing among colluding firms
	Subcontractor's share	Increased share or amount of subcontracts can reveal a structural change of a given market, that can be connected to collusive schemes
Consortium	Consortium members' name and ID Consortium members' share	Bidding and winning in a consortium is a widely used instrument for sharing rents earned from collusion. It is an especially useful enforcement mechanism, as rent sharing is agreed in a formal contract.

[34] Although beneficial ownership can lead to the investigation of dominant position.

3 Proposed collusion risk indicators

3.1 Conceptual frame and methodology

3.1.1 Collusion in public procurement markets

The characteristics of collusive behaviour[35] in public procurement markets is very similar to that of conventional markets: companies coordinate their behaviour regarding price, quantity, quality or geographical presence in order to increase market prices. The essential long term determinants of the prevalence of this kind of (mis-)conduct are 1) the ability of coordinating bidding behaviour, 2) internal sustainability (credible punishment system, effective detection of cheating), 3) external sustainability (ability to exclude new market entrants); and 4) ability to avoid detection and punishment (e.g. competition authority fine).

Public procurement markets are more vulnerable to coordinated gaming in light of the above features than traditional markets. In these markets, the outcome is determined by an auction mechanism, implying that there is no quantity adjustment as price changes, at least not on the short term (think for example about annual procurement plans mandatory in many countries). Ultimately this leads to an inelastic demand side.

Relatively large contracts in markets where tenders are announced infrequently can incentivize companies to bid aggressively for the first tenders, where they can win lasting market power. This market power is supported by switching costs (i.e. switching to a new supplier during delivery or in between two related contracts) and the high cost of market entry which is related to economies of scale. These features of procurement markets can further facilitate collusion (Klemperer 2007). Furthermore, Heimler (2012) argues that in such a 'winner takes all' system –with its resistances to quality or quantity adjustments driven by inelasticity of demand – the gains of collusive conduct are higher.

Beyond issues raised by the structure of public procurement markets, one of the most significant problems from the viewpoint of deterring collusion is transparency (Stigler 1964) which is conducive to government accountability more broadly. Since the results of public procurement tenders are public, the monitoring of the collusive agreements is generally costless, which facilitates the agreement adherence. In the case of procurement markets with many contracts, this monitoring effect is further strengthened, as immediate punishment is possible. Hence detection and the cost of punishment can be significantly lower compared to traditional markets, if tenders are frequent enough.

[35] Throughout this paper, the terms 'collusion' or 'collusive ring' are used to capture all forms of anti-competitive behaviours as defined above. Other authors in the literature often use by and large overlapping terms such as cartel which we don't use here to avoid confusion.

These characteristics of public procurement markets make the likelihood of creating and maintaining collusion more likely, which can be evidenced by the low number of public procurement collusion cases revealed and the relative ineffectiveness of leniency programmes across Europe. The need for insights other than from insiders and purchasing bodies combined with the availability of complex micro-level data together motivate the development of advanced quantitative collusion risk indicators.

3.1.2 Types of collusive behavior

Most if not all collusive behaviours can be categorized according to three dimensions describing the whole spectrum of activities from the implementation of competition restriction to the sharing of rents earned:

1. means of competition distortion or elementary collusion techniques,
2. forms of rent-sharing, and
3. resulting market structure.

When it comes to the means of competition distortion or the so-called elementary collusion techniques, three strategies can be identified which jointly describe the whole field of available strategies: (i) withheld bids, (ii) non-competitive bidding, and (iii) joint bidding. For the first strategy, one or more companies withhold their bids, so that there is less competitive pressure on the remaining firm(s), raising the price. For the second strategy, the parties mimic competition. Losing companies either bid a higher price than the competitor(s), their submitted bids are weaker in quality or they simply submit erroneous bids. This is considered to be the most common form of public procurement collusion by expert practitioners (Detecting Bid Rigging in Public Procurement n.d.). The third strategy involves companies biding jointly or in a consortium. This is a special form of collusion as it also determines and formalizes the method of rent allocation among the winning parties[36].

The second dimension of collusion categorization is based on the profit or rent allocation mechanisms used by colluding firms. The most general distinction regarding rent allocation can be based on whether the companies are active or passive members of the public procurement tenders (Pesendorfer, 2000). If they are active, then one of the most trivial profit allocation mechanisms is given by the formation of a consortium. Geographical, market-segmentation or time-based coordinated allocation of tenders also yield straightforward profit allocation mechanisms. When companies are passive, i.e. are not directly present, then a

[36] Bidding in a consortium is very similar to the case of horizontal mergers, hence the number of competitors decreases, which ultimately can lead to higher consumer prices. Furthermore, this can also be exacerbated by coordinated effects, as information sharing among fewer players becomes easier (Albano et al., 2009). Although, based on a signalling model, a company initiating a consortia would indicate high costs (Estache and Iimi, 2008), economies of scale can also explain such cooperation (Albano et al., 2009).

common ownership network (somewhat similar to the presence of consortia) or the use of subcontracts can solve allocation problems. Another straightforward form of redistribution is to simply use informal side-payments, which is the hardest mechanism to be detected empirically (some indicators common in organised crime research targeting cash use and financial accounts' anomalies might be adequate in this context, but they are not discussed at length here).[37]

The above elementary collusion techniques and rent allocation mechanisms can lead to different market structures each reflecting collusion rather than genuine competition. As competition distortion results from geographic, product market-wise, or temporal coordination using active or passive participation of colluding firms (i.e. present or not in the auction), collusion can result in two types of non-competitive market structures. On the one hand, it is possible that a concentrated market structure is created by collusive bidders, hence there is explicit market division with relatively high market shares (Levenstein and Suslow 2006; Pesendorfer 2000). On the other hand, a competitive market structure can also be imitated by colluding bidders ((Athey, Bagwell, and Sanchirico 2004; Mena-Labarthe 2012; Pesendorfer 2000; World Bank 2011). In this case, concentration can occur by time, as companies agree on a given winning order in a specific market which results in an artificially stable market structure. But concentration can also happen by geographical or product submarket which implies a more concentrated market but only on those narrowly defined markets.

Means of market distortion, rent allocation mechanisms, and the resulting market structures can be combined in a number of different ways, each of which is compatible with the logic of rent extraction from collusion. Table 8 succinctly summarizes these three dimensions, their main values, and the collusion types they define. However, not every combination is conceptually meaningful and empirically relevant. In addition, the table also depicts which theoretically conceivable collusion type can be measured by the proposed indicator framework using widely available public data.

---

[37] Different rent allocation techniques are likely to have varying efficiency implications. The use of informal or formal side-payments can lead to a more efficient collusive ring, as rents are maximized by the most effective firm serving the market. When all colluding parties are active in bid-rigging, efficiency losses can be considerable due to less efficient firms production (see Pesendorfer, 2000).

Table 8 Main characteristics of collusion types and the availability of indicators

Resulting market structure	Elementary collusion technique	Form of rent sharing			
		Sub-contractor	Consortia/ joint ownership	Coordinated bidding[38]	Informal side-payments
Concentrated market structure	Withheld bids	A			
	Losing bids	B			
	Joint bids		C		
False competitive market structure	Withheld bids	D		F	
	Losing bids	E		G	
	Joint bids				

Notes: every dimension is measured, some dimensions are measured, conceptually non-existent type

3.1.3 Measurement approach

The proposed measurement approach is highly ambitious as it aims at generating a generally applicable toolkit relevant across time, markets, and regulatory regimes. This approach is expected to be feasible because most collusive rings are understood as being essentially similar in their goals and strategies on the micro-level, following the previous section's discussion.

Such a broad remit also implies that it can only use data widely available as well as comparable across markets. Unfortunately, we have to neglect an emerging and very insightful literature using data specific to given contexts and collected to fit the purposes of the analysis only. With the increased wide availability of more detailed public procurement data, we hope to be able to incorporate and generalise these specific insights too. Such a specific study is for example done by Andreyanov, Davidson, and Korovkin (2016) who investigate the exact timing of bid

[38] While coordinated bidding typically creates a concentrated market structure on sub-markets, taking markets as defined by 'normal' competitive environments as the unit of analysis only allows for false competition to arise rather than monopoly. Hence, the theoretical impossibility of coordinated bidding and monopolistic market structure.

submissions.[39] Furthermore, Bajari and Ye (2003) shows a parametric collusion test applied to a very specific case with detailed bid prices and industry specific information to fit small-sample cases.

Public procurement covers practically the whole spectrum of economic activity from the construction of nuclear power plants to the provision of school meals in all sorts of local and global economic environments. The authors cannot hope to understand the detailed complexity of all cases. Instead quantitative data analysis should be deployed to define 'healthy' and collusive public procurement competition across a range of dimensions including prices or number of bidders. Any quantitative claim should naturally be further investigated and eventually verified by investigators knowledgeable of the given market as part of a mixed methods approach. Nevertheless, as such investigations and verifications are part of the traditional methods employed by competition authorities the subsequent discussion focuses on quantitative indicators and the procedures for developing them.

No silver bullet is offered in this paper; rather a five-step fine-tuned process is described, leading to an appropriate measurement framework fit for the local context.

1. **Market definition**: Colluding firms typically target markets defined by product, geography, and time. Hence, any collusion detection framework has to reliably identify markets for indicator development. There are no universally applicable and stable market definitions, rather key dimensions of market differentiation can and should be used for identifying alternative market definitions each of which can be used as testing the robustness of risk scores.

2. **Elementary indicators**: A broad set of elementary collusion risk indicators is defined covering as many types of collusive behaviours as (Table 12). These indicators are expected to signal collusive bidding, albeit they may as well be associated with confounding factors. For example, monopolistic market structure may be the result of markets divided up between colluding firms just as well as severe economic contraction bankrupting all but one competitor on a market. The literature has proposed many elementary indicators, many of which have already been applied to specific cases. Here, only those are discussed which can be calculated on widely available public procurement datasets, such as the Swedish public procurement database.

[39] Such cases can also include rather interesting scenes, such as two representatives submitting all bids personally for participating companies. See a recent Hungarian case: http://index.hu/gazdasag/2016/05/09/igy_nezett_volna_ki_ha_dumb_es_dumber_kozbeszerzesi_kartellt_szervez/

3. **Benchmarks**: public procurement markets and the proposed elementary collusion indicators vary a great deal due to diverse reasons entirely unrelated to collusive bidding; for example, economic growth or regulatory framework idiosyncrasies. In order to identify those values of elementary indicators which are more likely to indicate collusion, collusive markets have to be compared with 'healthy' markets. Ideally, collusive markets are defined using court judgements. In the absence of such judgements and the presumed specificity of proven cases, quantitative comparisons have to be carried out for elementary indicators. Comparisons can exploit exogenous variation in terms of time, geography, or product market. For example, the same product market may behave similarly across major regions, but starting from a given time point one of the geographical sub-markets may deviate from the others (for Hungarian examples see (Tóth et al. 2014)). Probably the simplest benchmark, albeit a debatable one, is the whole public procurement market other than the potentially risky market under investigation, which assumes that overall colluding bidders represent a minority. While any such 'deviation from the established competitive norm' may also be due to a range of alternative explanations, it is the first necessary step in identifying collusion risks and defining which changes in indicator values amount to substantial variation, and which changes amount to noise.

4. **Validity tests:** While indicators suggested by the academic literature or proven cases represent the starting point for a valid and reliable collusion warning system, each indicator needs to be thoroughly tested on the public procurement data it is applied to because markets may differ in substantial ways making some indicators valid in some contexts, while not in others. Using benchmarks as described above, indicator validity can be established by exploring: i) indicator co-variation in line with theoretically defined collusion types; ii) collusion risk indicators' correlation with prices; and iii) collusion risk indicators' correlation with company financial performance such as profitability. This implies that some elementary indicators are expected to co-vary. For example in markets where monopolistic market structure has arisen, prices should increase, and previously active competitors should abstain from the market, in contracts to benchmark markets. At the same time, elementary collusion risk indicators which signal a different type of collusive behaviour could move in the opposite direction or appear completely unrelated. For example, when the risk of monopolistic market distortion is high, indicators of fake competition should be low. Furthermore, a company which gains a monopoly position in public

procurement should also improve its profitability compared to benchmark companies, assuming that public procurement represents a substantial source of its income.

5. **Intelligent warning system**: A validated set of elementary collusion risk indicators give rise to an indicator system which makes the best use of each indicator, considering their interdependencies, strengths and weaknesses. This may imply developing one or more composite scores or operating a web of elementary indicators retaining the high degree of granularity in risk assessment. In any case, indicators can be transformed into broad categories or retained as continuous variables. They can also be preserved to characterise the lowest level of aggregation, typically an individual tender, but they can also be aggregated to characterise companies or markets over a certain period. The choice of indicator system should reflect user needs as well as indicator reliability (e.g. if 2-3 indicators are deemed jointly valid, but rather unreliable individually, a composite score might be a suitable choice).

In spite of the wide range of elementary risk indicators and innovative ways of combining them, no such approach can hope to indicate the presence of collusion with high precision, hence the reference to collusion risks rather than collusion *per se*. In addition, sophisticated collusive rings can learn the specificities of the measurement methodology and develop strategies to avoid detection. This necessitates a dynamic monitoring framework where emergent forms of collusion in public procurement are incorporated in the detection framework on a continuous basis.

3.1.4 Identifying markets

Colluding companies tend to concentrate on specific, well-defined markets where entry of non-colluding firms is less likely and a limited number of companies can jointly achieve market power. Hence, identifying a collusive ring crucially hinges upon adequately defining the market where collusion takes or could take place. Defining markets used as a basic unit of analysis has a crucial impact on results. If we draw the boundaries of markets too narrowly or broadly, the existing patterns in procurement contracts that would indicate the presence of a collusive ring could become unnoticeable.

Generally, relevant markets can be defined in a bottom-up or top-down approach. The bottom-up market definition builds on the bidding patterns of companies on the whole public procurement market without recourse to widely used product or geographical classifications. It presumes that when two companies bid for the same tender they are likely to belong to the same market, hence, company co-bidding defines markets. Formally, this can be represented as clustering of companies in a co-bidding network over a period. Markets are defined by the companies who are very likely to co-bid, and the tenders they actually participate in jointly. Market boundaries are defined by the low probability, or complete lack of cobidding among dense co-bidding clusters.[40] The strength of this approach is that it does not superimpose any statistical classification which may or may not be accurate in a given context, while also allowing for a formal identification of different degrees of product market similarity. However, a major shortcoming of this approach is that it is highly sensitive to the presence of collusive rings. If collusion is taking place on a market, it is likely to define markets incorrectly (e.g. companies divide up the market and withstand co-bidding on each other's submarkets). As our fundamental goal is to define markets supporting collusion detection, we cannot implement this approach.

The top-down market definition follows from theoretical considerations evoking standard demand and supply side factors such as product substitutability or geographical range of suppliers. Using official classifications, markets can be defined by these factors prior to exploring collusion risks which makes market definitions independent of collusion risks hence non-sensitive to anti-competitive bidding patterns. Nevertheless, the so-identified market definitions depend on the accuracy of official classifications in defining products and geographical units both of which may change rapidly rendering at least some market definitions imprecise. Given independence of the top-down market definition approach from collusive behaviours, we will explore this avenue in detail.

In order to reflect the inherent uncertainty of official classifications and the flexibility of company bidding behaviour, we first discuss dimensions (variables) that could be used for defining markets in flexible ways (e.g. using higher or lower geographical resolution) and provide a feasible simple partitioning of the procurement market which will be used for the below calculations. We defined public procurement markets using the following three dimensions:

- the type of the product or service, which can be defined based on the CPV (Common Procurement Vocabulary)[41] codes in the contract award announcements;
- the location of the performance of the contract, which can be identified based on the NUTS codes[42] in the contract award announcements; and
- value of the goods and services procured.

---

[40] This market definition approach is analogous to how internet companies devise music or video recommendations by simply observing co-occurences of music or video files. On the general method see: http://tdunning.blogspot.co.uk/2008/03/surprise-and-coincidence.html?m=1

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

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

While categorizing tenders based on these three dimensions seem straightforward, due to data quality issues we needed to implement several data improvement procedures. First, both CPV and NUTS codes have been improved to increase their precision. Second, due to the structural inconsistencies of the CPV categorisation (i.e. categories on the same level of hierarchy have vastly different detail), some unrealistically small markets have been joined-up.

Dimensions of market definitions

CPV codes

The discussion of this section builds on 2.2, while also explaining the CPV code adjustments in more detail. As it was explained above, the Visma Opic database contains only 'non-hierarchical' CPV codes which make product market identification problematic. Furthermore, the CPV categorization also suffers from structural inconsistencies in terms of product categorisation precision at any given level of the coding hierarchy (Commission 2012). This is best demonstrated by very broad CPV groups like 452 'Works for complete or part construction and civil engineering work' in contrast with much narrower CPV groups such as 723 'Data services' which are similarly at the 3-digit precision level, so in principle should be equally specific. While the former can contain tenders from simple building construction to pipeline construction, the latter contains relatively similar services, such as data-processing or data management.

Therefore, as a first step we assign each tender to the most detailed CPV code category. Eventually, this allows us to separate smaller sub-markets while aggregating those being overly detailed ex post (see below). While in the most ideal scenario of this should be done by using 4-digit CPV codes, due to the lack of detailed enough coding, we construct the most accurate 4-, 3- and 2-digit CPV codes for each tender in the following three steps.

First, as it was already discussed in section 2.2, the simplest way to assign one CPV code for a tender is a 'greatest common divisor' approach. Here, we echo the two straightforward cases discussed in section 2.2 as well. There can be a non-conflicting CPV code array, such as example A in Table 9, whereas 7242 is the most detailed 4-digit CPV code, that does not contradict any other assigned CPV code. Example B is a case, where '9091' is contained in each CPV code, hence an unambiguous 4-digit categorization is possible. This categorization logic was also applied for 3- and 2-digit CPV codes. We refer to this first case of categorization as 'non-conflicting CPV codes'.

Second, we calculated the highest relative occurrence of each 4-, 3- and 2-digit CPV codes per tender. For each tender without non-conflicting CPV codes, we assigned the dominant CPV code, that covers either i) more than 50% of the CPV codes, or ii) exactly 50% of the tender level CPV codes, but without any equally dominant CPV codes (i.e. there is no other CPV code with 50% occurrence). As in example C in

Table 9, 67% of the assigned CPV codes are within the 3234 CPV code category. We refer to this second case of categorization as 'dominant CPV codes'.

Third, if no dominant tender level CPV codes can be identified, the tenders are assigned the highest average value CPV code. We refer to this third type of categorization as 'highest value CPV codes'. For calculating the average tender value within 4-, 3-, and 2-digit CPV code categories, we use that subsample of the Visma database which includes CPV codes of relatively high precision.[43] Furthermore, we only use those categories where we have at least 4 observations per a 4-, 3- or 2-digit CPV category. Note that if calculating the average tender price for a 4-digit CPV category due to insufficient number of observations is not possible, the 3- or 2-digit CPV categories are used. Obviously, this estimation holds only if both the availability of contract values and unambiguous CPV coding are independent. Although, this is probably not the case, there is no better way to assign one CPV code to each tender.[44] In example D, the 4-digit assigned CPV code is based on the highest value CPV 4 category: among 4531, 5071 and 453 CPV coded tenders the 4531 CPV code has the highest average contract value. Note, that there are dominant 3- and 2-digit CPV codes (50%+ incidence ratio) for the same tender (i.e. 3-digit CPV code 453 and 2-digit CPV code 45). In case of example E, it was not possible to assign a 4-digit CPV code due to the insufficient number of tenders clearly assigned to 6371 or 6311, which prevents the assignment of well-established average tender size. Therefore, only 3-digit CPV categorization was possible.

[43] We only include contract from the 'non-conflicting' and 'dominant' CPV categories, if more than 80% of the CPV codes are the same.

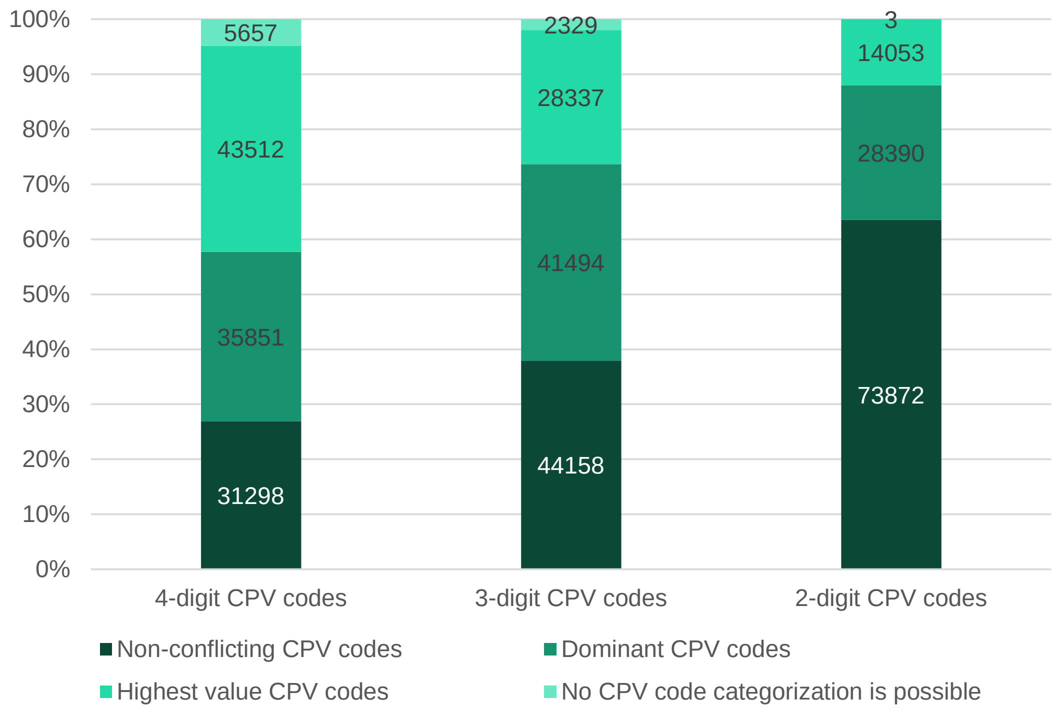
[44] As the basis for calculating average tender prices is tenders with non-conflicting or dominant CPV codes, it is plausible that these tenders are among the simpler contracts. Larger and more complex contracts are expected to have more CPV codes assigned. Also, these more complex projects are not independent of CPV codes. Therefore, the CPV code assignment might be biased in these cases.

Table 9 CPV coding examples

Example ID	Assigned CPV code	4 digit CPV code	3 digit CPV code	2 digit CPV code	Ratio of most commonly used 4-digit CPV code	Ratio of most commonly used 3-digit CPV code	Ratio of most commonly used 2-digit CPV code	Assigned 4 digit CPV code	Assigned 3 digit CPV code	Assigned 2 digit CPV code
A	72	72	724	72	-			7242	724	72
	724	724	724	72	-			7242	724	72
	7242	7242	724	72	-			7242	724	72
B	909111	9091	909	90	100%	100%	100%	9091	909	90
	909112	9091	909	90	100%	100%	100%	9091	909	90
	909192	9091	909	90	100%	100%	100%	9091	909	90
	909111	9091	909	90	100%	100%	100%	9091	909	90
	909112	9091	909	90	100%	100%	100%	9091	909	90
C	32233	3223	322	32	67%	67%	100%	3234	323	32
	32343	3234	323	32	67%	67%	100%	3234	323	32
	3234421	3234	323	32	67%	67%	100%	3234	323	32
D	4531	4531	453	45	33%	67%	67%	4531	453	45
	50711	5071	507	50	33%	67%	67%	4531	453	45
	453	453	453	45	33%	67%	67%	4531	453	45
E	637121	6371	637	63	50%			-	637	63
	6311	6311	631	63	50%			-	637	63
F	454531	4545	454	45	13%	50%	88%	4521	454	45
	45421141	4542	454	45	13%	50%	88%	4521	454	45
	50	50	50	50	13%	50%	88%	4521	454	45
	452311	4523	452	45	13%	50%	88%	4521	454	45
	45262321	4526	452	45	13%	50%	88%	4521	454	45
	454423	4544	454	45	13%	50%	88%	4521	454	45
	452125	4521	452	45	13%	50%	88%	4521	454	45
	4543	4543	454	45	13%	50%	88%	4521	454	45

As a result of the above three-step procedure, we generated unique 4-, 3- and 2-digit CPV codes for each tender representing our best estimation of the main product purchased (i.e. product with highest value in the bundle of products). Figure 14 outlines the quality of assigning CPV codes to each tender on the 4, 3 and 2 digit levels. Generally, the less precise the CPV category the more straightforward it is to assign a unique main CPV code. In particular, the 4-digit CPV codes are the most problematic: only 31,298 tenders, about one third of the sample, could be unambiguously assigned to a 4-digit CPV code, while an additional 35,851 tenders have a dominant 4-digit CPV code, and another 43,512 tenders can be assigned a main CPV code based on the average CPV category contract values.

Figure 14 CPV code categorization (N=116318)



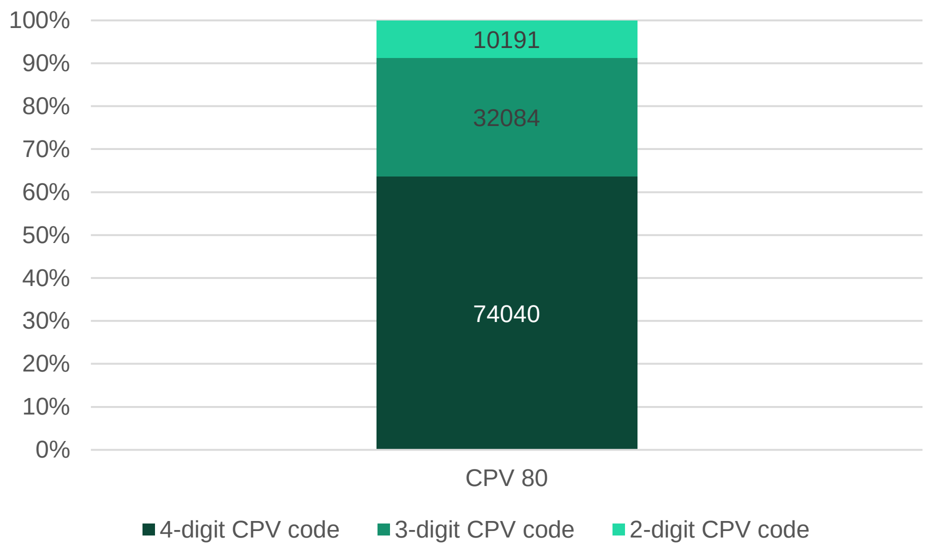
It follows from the above discussion that there is a trade-off between using the 'dominant' and 'highest value' CPV code categorization, i.e. trusting the accuracy based on relative CPV code occurrence and the average tender size by each CPV categories. While choosing the 'dominant' CPV code might be useful, in the case of a tender has one high value component accompanied by smaller parts, it is often the case that the relatively infrequent CPV codes of the tender will dominate product market classification in reality. Therefore, the actually competing companies – hence the tender itself – should be assigned to another product market.

Assigning main CPV codes to tenders is further complicated by potential conflicts between CPV levels. For example, while the 4-digit classification can be based on the highest value 4-digit CPV code, it might contradict with the 3-digit classification that is made based on the unambiguously dominant CPV code. In order to systematically resolve such conflicts in the coding and assign a unique highest-precision, still reliable CPV code to each tender, we followed the below sequence:

- First, we accept all 4-digit CPV codes that are either non-conflicting, or there is a dominant CPV code with at least 80% congruence.
- Second, we accept 3-digit CPV codes with at least 80% congruence
- Third, we accept the 'highest value' 4-digit CPV codes, if they fall under the same 3-digit and 2-digit CPV codes
- Fourth, we accept the 'highest value' 3-digit CPV codes, if they fall under the same 2-digit CPV codes
- Fifth, we accept the 2-digit CPV codes for the remaining unclassified tenders

For the sake of simplicity, we will refer to this CPV code classification as CPV80. By applying these rules, we have managed to come up with relatively detailed CPV categorizations, so that the top-down market definition contains relatively detailed CPV codes, that we can adjust by inspecting bidding patterns (see the next section).

Figure 15 Number of tenders according to the 4-, 3- and 2-digit CPV code classification in case of CPV80 market categorization



NUTS codes

Geographical location of contract performance represents the second major dimension of defining markets. In this respect, we can rely on the standardized NUTS nomenclature used in procurement notices in Sweden, but also Europe-wide. We assumed *a priori* that national, regional and local markets may exist separately in Sweden even for the same product. This means that e.g. the school building renovation market of a certain settlement could be interpreted as a separate market which is not equivalent to the school building renovation market of another settlement. However, in other cases the location of the project does not separate markets from each other, because the companies in the market are able to deliver goods or offer services in the whole country. The highway construction market is a good example of this case. We have to determine which markets are country-wide and what geographical size the local markets have.

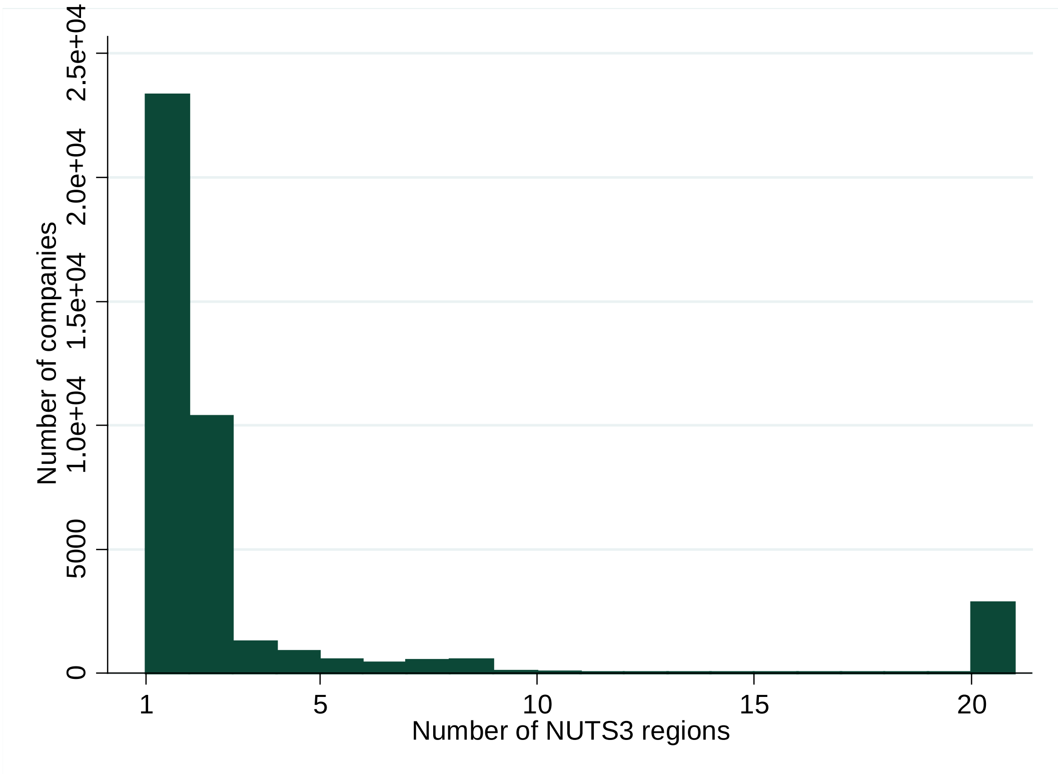
In practice, as it was already discussed in section 2.2, each tender was assigned the lowest non-conflicting level of NUTS code. However, in order to be able to decide on the most likely regional scope of different markets, we need an overall idea on the geographic scope of companies' bidding behaviour.[45] Therefore, we calculated a weighted measure of the geographical scope of bidding. Although, the NUTS coding preserves some bias regarding the size of certain regions, we regarded it as

[45] As it is shown by Table 20, while there is some variance of the availability of company IDs between different regions, there is no viable bias according to the number of tenders within different regions.

a good proxy for categorising companies according to the number of different regions they are active in.

We calculated the number of different NUTS3 regions where each bidder submitted a bid.[46] This metrics reveal that most companies bid only in a very small geographic area: 23,317 companies bid only within one NUTS3 area, and 10,370 in two NUTS3 areas (Figure 16). Whereas, there are only a few thousand companies bidding in a wider geographical region such as a NUTS 2 or NUTS 1 region, with the majority of them actually bidding nationally.[47] Furthermore, companies bidding between different NUTS 2 markets tend to bid between different NUTS 1 markets as well (Table 19).[48] By implication, we will only use NUTS 3, NUTS 1 and NUTS 0 (national) level geographical divisions for defining markets

Figure 16 Geographic scope of company bidding behaviour (N=41046)



[46] An important note here, is that as in many cases, the NUTS codes are given at the NUTS1, NUTS2 or NUTS0 level, we needed to weight them accordingly. E.g. if a company was bidding in SE22, then its geographical scope is „2", as there are two NUTS3 codes within the SE22 region: SE221 and SE224. Also in case of a company submitting a bid to a tender, that is country level, its scope becomes „21", as there are 21 separate NUTS3 regions in Sweden.

[47] Note, that it is possible, that tenders with a country-wide scope are inappropriate, i.e. their geographical scope is miss-specified.

[48] For example, there are 5,587 companies bidding within two NUTS 2 regions, whereas in 3218 cases, these companies bid in two different NUTS 1 regions as well. This implies that once a company is bidding outside a narrow NUTS 3 or NUTS 2 region, they tend to compete across NUTS 1 regions as well. Note, in case of strong 'mid-regional' competition, we should observe relatively more companies bidding in several NUTS 2 regions within one NUTS 1 region.

Contract value

The last dimension we use for defining markets is tender size. Obviously, companies of different sizes cannot compete effectively for each types of contracts. Consequently, we have to separate tenders that are relevant for different groups of companies: i.e. smaller road refurbishment works vs. new motorway construction. In the best case scenario, this categorization should be done by using either the estimated or final price of the tenders in question. However, as it was already discussed in 2.1, price related information is rather scarce in the Swedish database. As it was shown, only ca. 23,500 tenders have price information, representing about 22% of the tenders. Therefore, instead of relying on prices, we used information on whether the given tender is below or above the EU threshold.[49] First, this information is available for most of the tenders. Second, the threshold values can be considered to be good proxies for differentiating large and small contracts.[50] Third, being above the EU threshold, hence following the EU-wide regulatory regime and transparency requirements, also indicates that the possibility of foreign entry is higher. Tenders are not only published in Swedish procurement portals, but also on the EU's Tenders Electronic Daily, making tenders open to the public, and potentially increasing competition.

Market definition adjustments and final market definitions

The above defined 3 dimensions lead to several thousand unique markets, among which there are many with very few tenders (Table 10).

Table 10 Descriptive statistics on the markets defined, 2009-2014

CPV coding	Maximum accuracy of the market	Number of distinct markets	Number of markets with less than 5 tenders	Average number of tenders per market	Median number of tenders per market
CPV80	NUTS 3 level	10206	7420	11,4	2
	NUTS 1 level	3551	1927	32,8	4
	NUTS 0 level	1881	823	61,8	8

Very small markets are problematic for two reasons. First, due to the above discussed structural deficiency of the CPV classification (i.e. categories on the same level of the hierarchy denoting product groups of very different detail), small markets are likely due to the too high level of CPV code specificity rather than because they represent products of a unique market. Second, in order to analyse market-level behaviour, more observations are needed for each market which

[49] There are 80235 tenders below the threshold (69%), 34009 below the threshold (29%), while this information is missing for 2074 tenders (0.2%).

[50] The EU threshold differs for supplies, services and works. While in case of the former two categories, the EU regulation applies above 135000 EUR (for central governments) and 209000 EUR (for other contracting authorities), in case of work contracts it is 5225000 EUR for both types of contracting body. Note, that these are the current threshold, however, these thresholds are relatively stable over time, only a few adjustments were applied in the investigated time period.

reveal sufficient information about interactions among bidders. Therefore, we re-adjust overly narrow market definitions along either CPV or NUTS code dimensions.

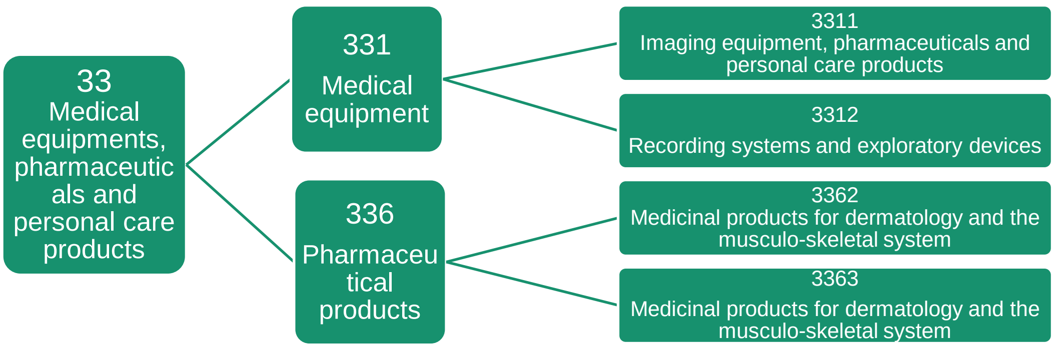
Market adjustment 1: product categories

After having classified all tenders according to the two procedures, minor product market adjustments are still need to be implemented. As it is shown in Figure 18, there are a few hundred cases, where the CPV80 categorization leads to very few observations. For example the 4-digit CPV code '3363' referring to 'Medicinal products for dermatology and the muscular-skeletal system', or the CPV '3917' referring to 'Shop furniture' (e.g. counters or display cases). There are two possible explanations. One is that there are products that are procured very rarely, and they should be treated individually indeed. On the other hand, CPV codes might be overly detailed, therefore the analysed product market could have a broader scope.

In order to address these concerns, we investigate whether companies bidding in smaller sub-markets are also participating in other, but related ones. Therefore, we can regroup extremely small markets based on bidding between different product markets. Two types of improvements were implemented. First, we investigated whether products of 4-digit CPV codes within the same 3-digit CPV code categories can be connected based on bidding behaviour. For example, whether companies bidding for the '3363' CPV code are also bid for '3362', which is 'Medicinal products for dermatology and the muscular-skeletal system'.

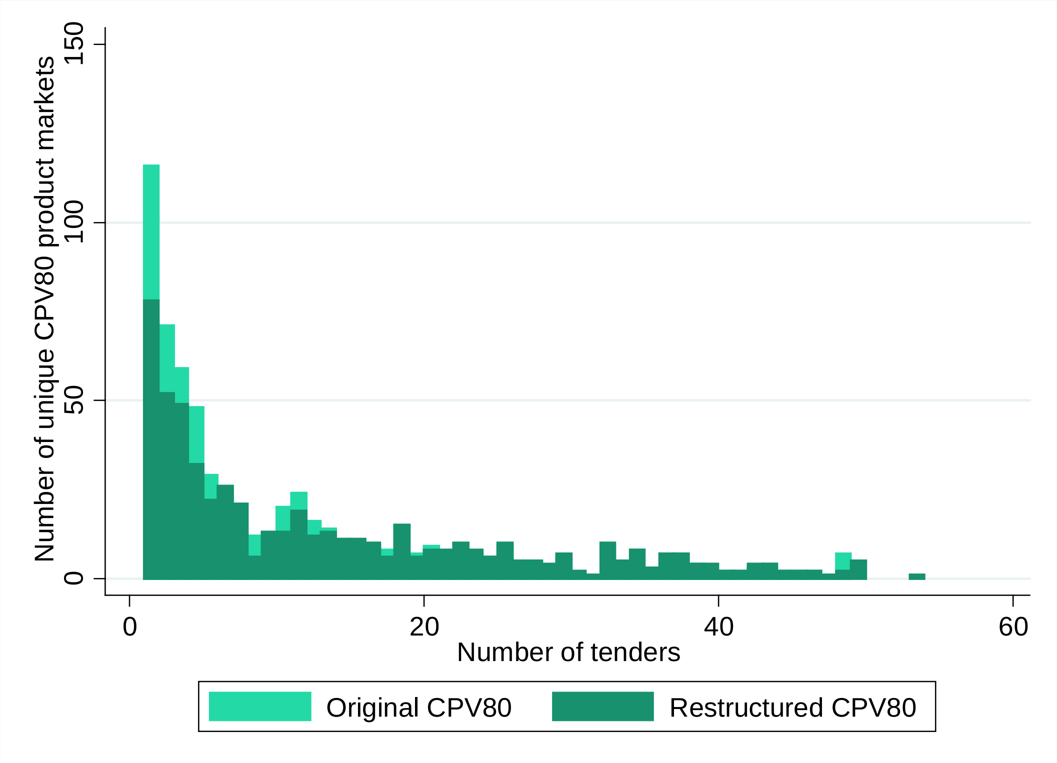
In technical terms, we calculate a cross-bidding measure based on 4-digit CPV codes within 3-digit CPV categories for each company (i.e. the sum of 4-digit CPV codes within a given 3-digit CPV code where a company placed at least one bid divided by the number of all occurring 4-digit CPV codes within a given 3-digit CPV code). As a second step we calculate 3-digit CPV code level averages of these measures and investigate whether cross-bidding between the 4-digit CPV codes is prevalent. As a third step, we regroup those 4-digit CPV codes into 3-digit CPV codes, where there are only very few tenders within 4-digit CPV categories (less than 10 for the whole period) and there are greater than median cross-bidding indicator.

Figure 17 CPV coding example



As a result, we regrouped 1300 tenders of CPV80 categories, i.e. 1300 4-digit CPV codes were regrouped either by joining 2 or more 4-digit CPV codes within the same 3-digit CPV code, according to the above explained rules. Figure 18 shows that while CPV categories with very few tenders remained, some of the very small product markets could be aggregated to an analysable size.

Figure 18 Number of unique product markets (based on CPV80 categorization) according to the number of tenders (if number of tenders is less than 50)



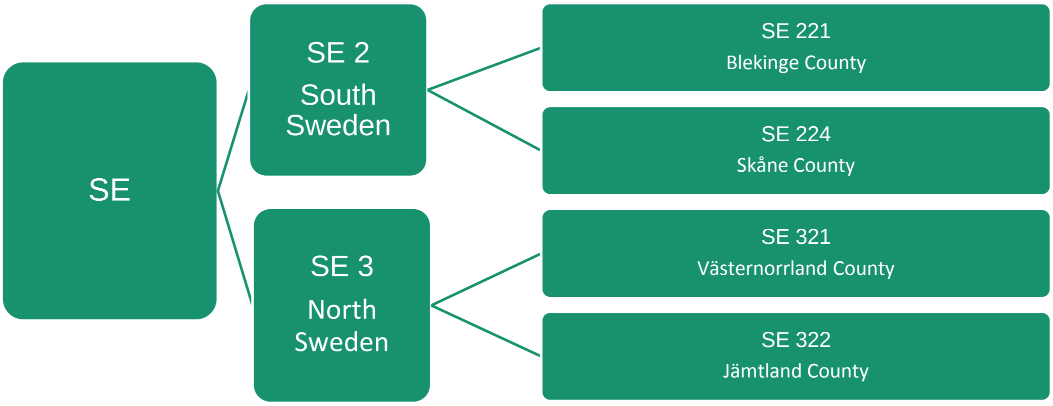
Market adjustment 2: geographical scope

While in the previous section we implemented an aggregation based on product codes, the geographical scope of markets can also be too narrow, hence underestimating market size. The problem to be tackled in this section is the same as in the previous one: certain markets remained extremely narrow for analysing collusion risks, hence we want to aggregate them in case of sufficient cross-bidding is observed. For example, it is trivial that certain goods are procured only occasionally across the whole country, hence only 30-50 tenders constitute the whole market for the investigated period. However, the companies active in these markets are probably bidding for the majority of these tenders; therefore aggregating them allows us to calculate the risk indicators with a better matching market definition.

Therefore, we apply a similar logic as in case of the product codes. First, we investigate whether companies bid across NUTS 3 regions within the same NUTS 1 region. The cross-bidding measure is calculated based on NUTS-3 codes within each NUTS-1 region for each company (i.e. the sum of NUTS-3 regions where a company placed at least one bid within each NUTS-1 region are calculated and divided by the sum of all NUTS-3 regions within the given NUTS-1 region).[51] Second, we calculated NUTS-1 level averages of these measures and investigate the prevalence of cross-bidding between NUTS-3 regions. [52] Third, we regroup those NUTS-3 level markets into NUTS-1 regions, where there are less than 6 tenders for the whole period (less than 1 observation per year), and inter NUTS-3 level bidding is greater than the national median.

For example, if for a given product market (based on CPV code) the number of tenders located in the SE-221 region is less than 6, and the companies bidding on these tenders are also relatively active on the SE-224 market as well, we aggregate these markets in the SE-2 geographical region.

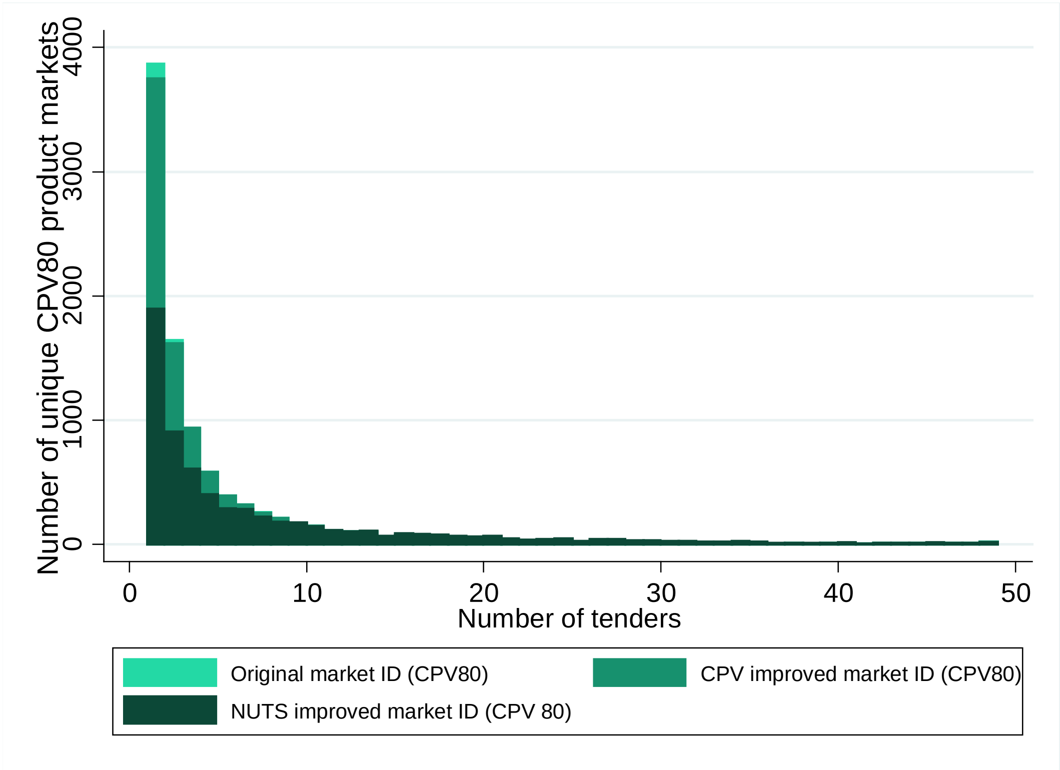
Figure 19 NUTS code example



[51] For example, a company bidding on the 'Technical inspection and testing services' market ('7163' CPV code) has a value of 0.5 when the company is placing a bid on separate tenders implemented in SE110, SE121, SE122, SE123 markets, as there are 8 different NUTS 3 or NUTS 2 regions.

[52] Note, that only companies with more than one bid were included in the sample, i.e. the mean cross-bidding is not distorted by temporary participants.

Figure 20 Number of unique product markets (based on CPV80 categorization) according to the number of tenders (if number of tenders is less than 50)



Final market definition

Below we present some descriptive statistics of the final market definition used for calculating the elementary indicators (Table 11).

Table 11 Descriptive statistics on the markets defined

CPV coding	Maximum accuracy of the market	Number of distinct markets	Number of markets with less than 5 tenders	Average number of tenders per market	Median number of tenders per market
CPV80	NUTS 3 level	7029	4104	16.5	4
	NUTS 1 level	2801	1141	41.5	9
	NUTS 0 level	1533	515	75.8	12

3.2 Defining collusion risk indicators

3.2.1 Indicator overview

Before discussing each indicator in depth, we provide a brief overview of collusion proxies available in the Swedish public procurement database (Table 12). They are grouped according to the nature of information they capture: i) prices; ii) bidding patterns; and iii) market structure. This indicator list is not complete, surely further indicators could be thought of. As noted above, we restricted the analysis to those indicators which have received more extensive academic discussion and which are based on public procurement data widely available rather than requiring extensive collection of additional information. Indicators which have to be excluded such as tracking the IP addresses of bid submissions, comparing the e-identity of document creators, tracing social media-based connections between bidders, etc. represent very exciting future possibilities which will have to remain unexplored for the time being.

Table 12 Summary of elementary collusion risk indicators feasible on Swedish public procurement data

No.	Indicator group	Indicator name	Indicator definition
1	Prices	Bidders having the same bid price	Tenders with at least two equal submitted bid prices are marked.
2		Difference between lowest and second lowest bid prices	Relative difference between lowest and second lowest bid price per tender
3		Relative standard deviation of bid prices	Relative standard deviation of bid prices per tender
4		Bid price range	Range of submitted bid prices per tender
5		Benford's law	Whether submitted bid prices follow Benford-law
6	Bidding patterns	Winning probability	Nr. of tenders won / Nr. of bids submitted by company in period
7		Cyclical winning	Bidders have auto-correlated winning patterns
10		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
11		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
12	Market structure	Concentrated market structure	Indication of increasing market concentration by measures of market structure
13		Stable market structure	Standard deviation between time periods of measures of market structure

There are a few collusion risk indicators which cannot be calculated on the Swedish data, but can add value to an effective collusion detection framework with relatively little additional administrative burden of data collection (consider for example that they are part of many public procurement databases in Europe):

- **Prevalence of subcontracting**: It shows whether subcontractors are involved in contract delivery. It is typically calculated on market level and used in conjunction with bidder numbers to capture whether previous competitors became subcontractors.
- **Prevalence of consortia**: It indicates whether winning bids were submitted jointly or not. It is often calculated on the market level and also considers whether consortium members were competitors prior to consortium forming.
- **Prevalence of faulty bids**: It is calculated as the ratio of faulty bids to total submitted bids. On markets where the prevalence of faulty bids suddenly increases, it can be suspected that companies concluded a collusive deal.

In addition, using cash or informal payments as a mechanism of rent-sharing among colluding companies is hard to track using publicly available datasets, hence it represents the only major aspect of collusion types which is insufficiently covered by indicators. Nevertheless, it is suggested that indicators of cash use as developed for example in the organized crime literature, could be adapted to the collusion case (not the least because organized crime has been found to organize and police collusive rings in procurement in Italy) such as the cash-ratio indicator.

Neither of these indicators is discussed in detail below, for more information see: (Tóth et al. 2014).

3.2.2 Indicators exploiting bid price distributions

A key characteristic of competition in public procurement markets is the distribution of offered prices. Variance, range and skewness can each signal a behaviour that is at odds with genuinely competitive behaviour. Unfortunately, empirical results have mixed conclusions regarding which patterns of bid prices signal collusive behaviour. On the one hand, low price variance and bid prices very close to the winning price can signal collusion. For example, Abrantes-Metz et al. (2006) find that collusion leads to decreased variance in prices on the retail gasoline market[53], which is also consistent with theoretical considerations, as lower price variance makes monitoring easier. On the other hand, such behaviour of decreasing price variance might not be useful in public procurement markets as tenders must be lost for maintaining the collusive ring. Hence, for non-competitive bidding, using artificially high prices may represent a safe option assuring the 'right'

---

[53] The investigation of Abrantes-Metz et al. (2006) was not in a public procurement market.

company wins the tender[54]. For example, the skewness of bid prices is used as a collusion indicator in Padhi and Mohapatra (2011), and also recommended by Oxera (2013). According to these studies, positive skewness indicates artificially high priced losing bids. It is not possible and even not desirable to decide which types of suspicious bid prices are more likely to indicate collusion, instead it is advocated that either an artificially high or low variance may signal collusion depending on the strategy the colluding companies adopt.

At any rate, a number of characteristics of bid price distributions can be used to gauge collusion risks, each of which follow a similar theoretical reasoning while being formulated in different ways:

- Bidders having the same bid price
- Relative difference between lowest and second lowest bid prices
- Relative standard deviation of bid prices
- Relative bid price range

Each of these is discussed in detail below.

Bidders having the same bid price

One of the most straightforward symptoms of collusive agreements is when the lack of price competition leads to very similar or equal offer prices, unrelated to the cost and demand characteristics of the market. While equal offer prices can also characterize markets with standardized products and dispersed demand or producers with virtually equivalent production functions, collusive schemes in public procurement can also involve bidding with equal offer prices (SCA 2015b).

By implication, a straightforward collusion proxy signals whether equal prices (either total or unit prices) were submitted to a given tender. Therefore, a categorical variable can be defined as the following:

$$Bidders\ having\ the\ same\ bid\ price_i = \begin{cases} 1\ if\ the\ tender\ has\ at\ least\ two\ equal\ bid\ prices \\ 0\ if\ the\ tender\ has\ only\ different\ bid\ prices \end{cases}$$

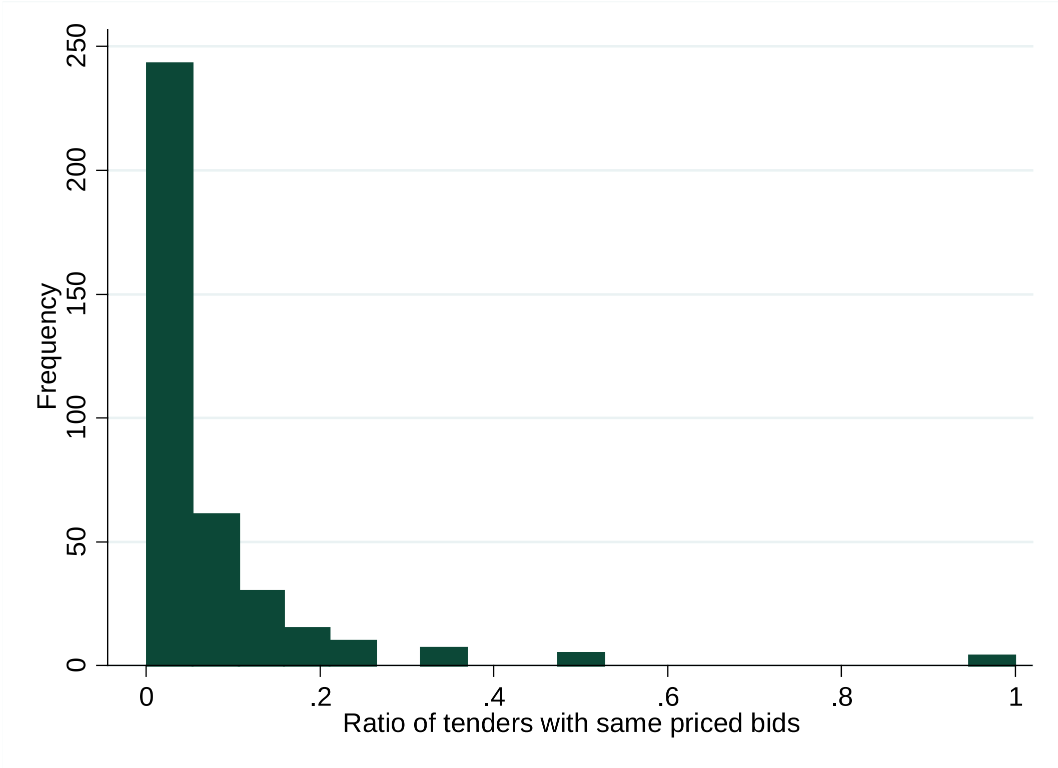
While a tender can have equal bid prices, especially if it is a standard commodity, can be due to random price variations or data errors[55]; systematic market level patterns make these explanations much more unrealistic. Therefore, we calculated the relative prevalence of tenders having equal offer prices at NUTS-3 markets (Figure 21). While some of the higher ratios are likely due to the small number of

[54] While in traditional markets, there is continuous quantity adjustment, in public procurement there is no continuous transition, hence loosing bids have to be strictly higher than non-loosing bids. Of course, if there is more sophisticated decision making scoring system, the implications for pricing can be mixed.

[55] For example, certain bids seem to be accounted twice whereas belonging probably to the same company.

tenders in the market, there are larger markets with reappearing equally priced bids throughout the whole period, across geographical locations.

Figure 21 Ratio of tenders having at least two equal bid prices by NUTS-3 level markets (sub-sample: markets where there were at least such a tender, N=373)



Difference between lowest and second lowest bid prices

In well-established competitive markets where companies regularly bid for similar contracts, the winning bidder and its close competitor(s) tend to offer prices following a certain distribution. Due to the central importance of the lowest (winning) and second lowest bid (best losing) prices for the outcome of a tender, their difference is of particular importance for collusion measurement. Following Abrantes-Metz et al. (2006), Oxera (2013), Padhi and Mohapatra (2011), both extremely small and large differences[56] between the lowest and second lowest offer prices can signal collusive behaviour, depending on the collusion mechanisms used. Constant (relative) differences over time can also suggest a coordinated pricing behaviour.[57]

Collusion of only a subset of bidders in a given market can also have detrimental efficiency effects. Therefore analysing the bidding patterns of the first and second

[56] Effective competition under similar cost structures should lead to very similar offer prices, hence this indicator is most applicable to markets with highly competitive companies of similar production technologies.

[57] In the case of competition, the effective cost levels, hence the bid offers of a company should vary over time, hence constant differences between first and second offers are likely artificial.

bid prices is an inaccurate attempt to make inferences regarding the whole bid price distribution. For example, through the collusion of the two most efficient firms, the effective competition constraint on pricing will be the marginal cost of the third lowest cost firm.

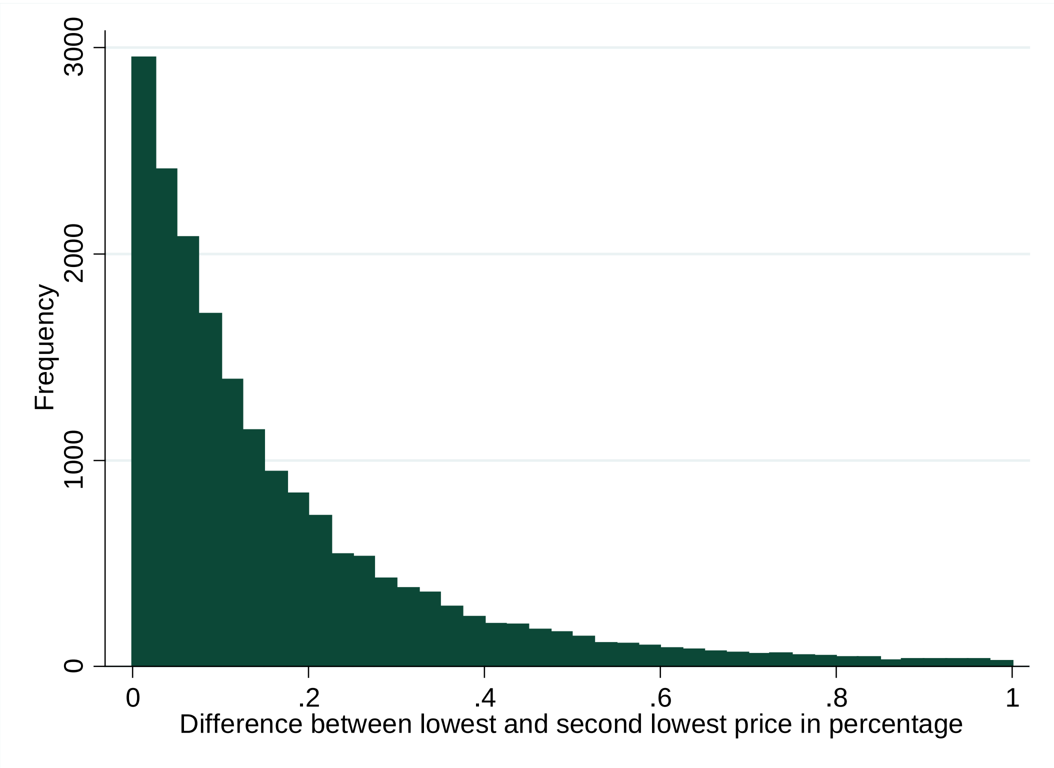
Following the above logic, the difference between first and second relative offer prices can be calculated for any tender *i* as the following:

$$Lowest\ and\ second\ lowest\ bid\ price\ difference_i = \frac{Second\ lowest\ bid\ price_i - Lowest\ bid\ price_i}{Lowest\ bid\ price_i}$$

While this indicator can be calculated on the tender level to preserve the highest level of granularity of analysis, aggregating to company or market levels can reveal systematic competitive behaviour and, if there is any, point at colluding firms.

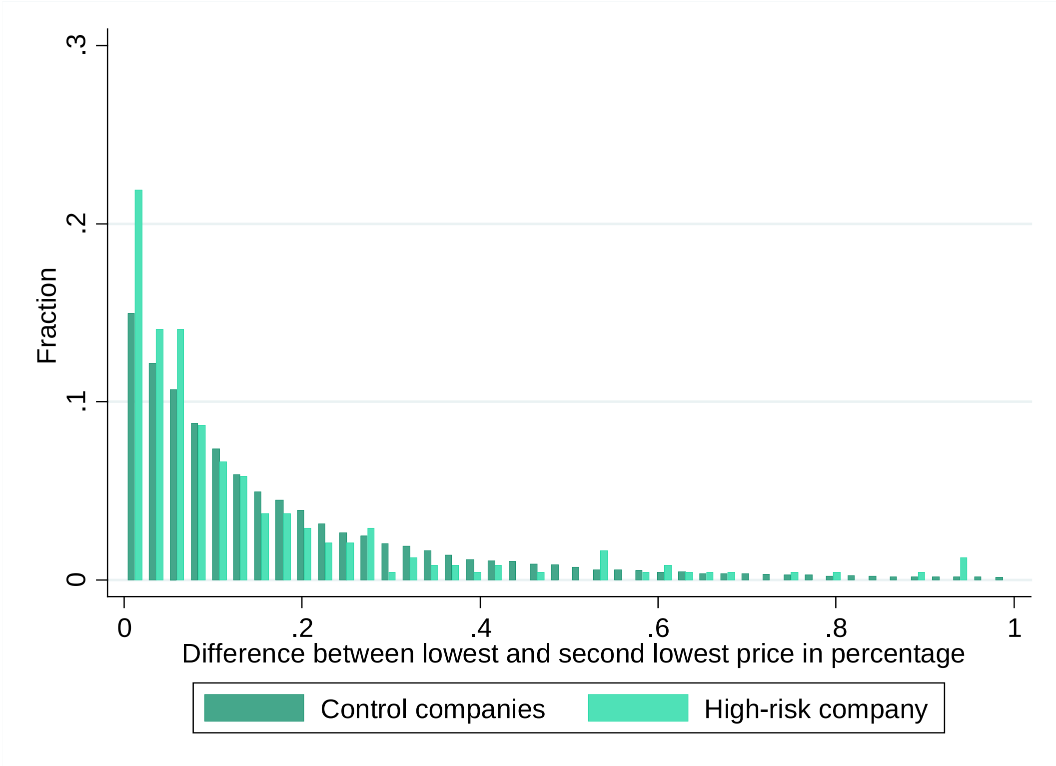
Looking at the relative difference between the lowest and second lowest bid prices, it can be seen that the difference is volatile: many tenders having minor differences between the two lowest bids, while others having rather major ones (Figure 22). While the median relative difference is around 10% and the mean 16.6%, both the extremely low and high average differences can signal anti-competitive bidding.

Figure 22 Distribution of the difference between lowest and second lowest bid price divided by the lowest bid price (N = 19170, whole period)



To further demonstrate the kind of risk differences revealed by the lowest and second lowest price difference, we zoom in on a high risk company and compare it to the rest of the public procurement market. This company has a 0.16 average difference compared to 0.19 on the total market and quite interestingly it participated in a lot more tenders with very small price difference than other companies.

Figure 23 Distribution of the difference between lowest and second lowest bid price divided by the lowest bid price for all tenders and a high risk company (N_{all}= 19170, $N_{high\ risk\ company}$= 272, 2009-2014)



Relative standard deviation of bid prices

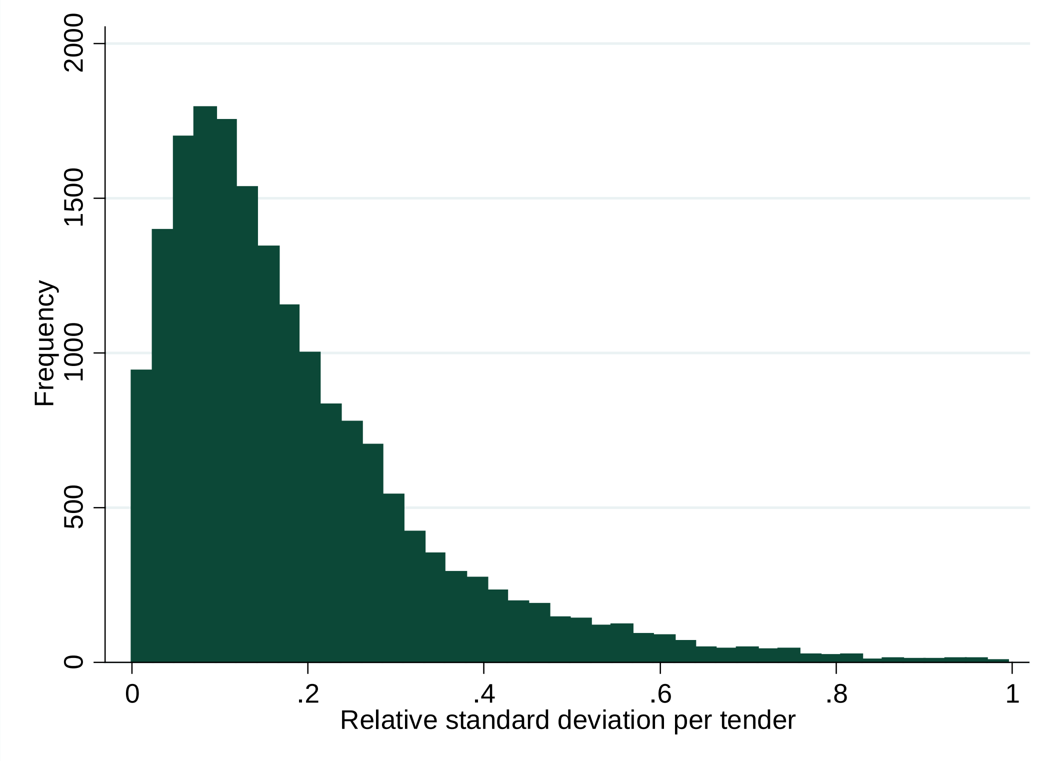
As it was already highlighted in the previous section, bid prices should follow a certain distribution under non-collusive competitive conditions. While scrutinizing the two lowest bids focuses attention on the potential collusion between the two lowest cost companies, more elaborate collusive schemes can be detected by analysing the whole distribution of bids. One of the comparable measures that can be used for analysing bid price distributions is the relative standard deviation (i.e. the standard deviation of bid prices divided by the mean of bid prices), that can be calculated for each tender. Similarly to the previous indicator, extreme or unusual offer price distributions are found to signal collusion by academic literature Abrantes-Metz et al. (2006), Oxera (2013), Padhi and Mohapatra (2011).

Using a standard formulation this collusion risk indicator can be calculated for tender i in the following way:

$$Relative\ standard\ deviation\ of\ bid\ prices_i = \frac{Standard\ deviation\ of\ bid\ prices_i}{Average\ of\ bid\ price_i}$$

As expected, the relative standard deviation of prices is around 10-20%, reflecting the actual marginal cost differences of companies (Figure 24Figure 22). However, it is also apparent that there are several tenders with extremely low and extremely large bid price differences. Although, these outlier values might be driven by fierce competition or special features of the tendering process (e.g. multiple winning criteria including quality or environmental considerations inflating certain bids leading to wide distributions), persistently low or high relative standard deviations on the market level may indicate a stable collusive scheme.

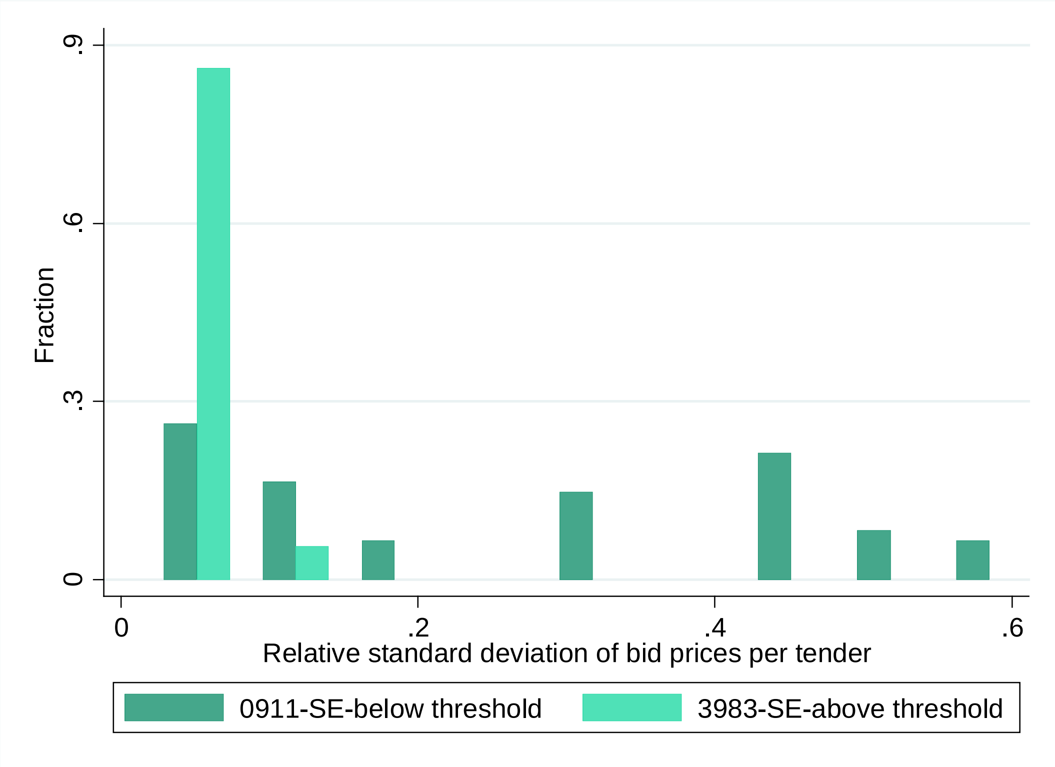
Figure 24 Distribution of relative standard deviation of bid prices per tender (N= 18,583, whole period)



Note: Tenders included only if total price was non-missing, and relative standard deviation was less than 1.

To demonstrate, to what degree consistently low relative standard deviation markets can differ from average or supposedly non-collusive markets, we depict the distribution of tenders within two markets: one of high risk (3983_SE_above), another of low to moderate risk (0911_SE_above) (Figure 25).

Figure 25 Relative standard deviation of tenders for two markets ($N_{0911\text{-}SE\text{-}a}$ = 9, $N_{3983\text{-}SE\text{-}a}$ = 22 whole period)



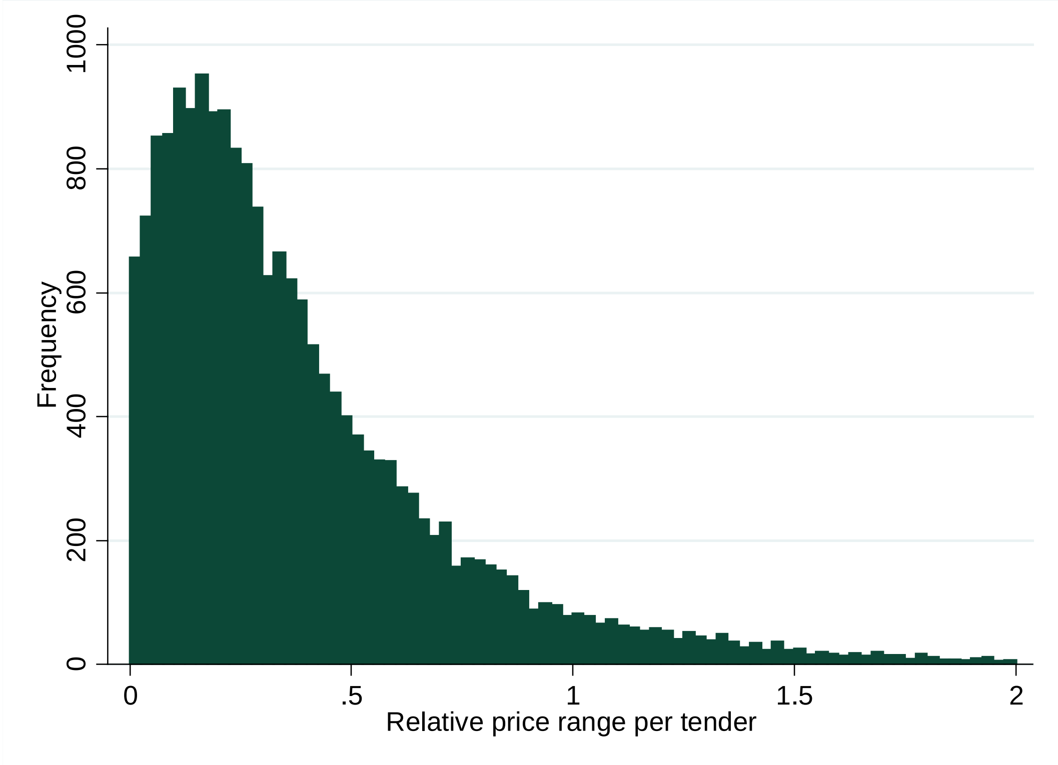
Bid price range

A further aspect of potentially risky bid price distributions is the range of offered prices. Following the above outlined logic, both very low and relatively high price ranges can be indicators of non-competitive bidding behaviour – depending on the game played by the collusive parties. Therefore a straightforward indicator for collusion risk is the range of bid prices for tender *i*:

$$Bid\ Price\ Range_i = (highest\ bid_i - lowest\ bid_i)/average\ of\ bid\ prices_i$$

Figure 26 depicts the relative offer price ranges in Sweden between 2009 and 2014 on the level of contracts awarded. It makes it clear that very small price ranges and relatively high price ranges (i.e. those above 1) are also prevalent among public procurement tenders.

Figure 26 Distribution of relative price range of tender (N = 19,625, 2009-2014)



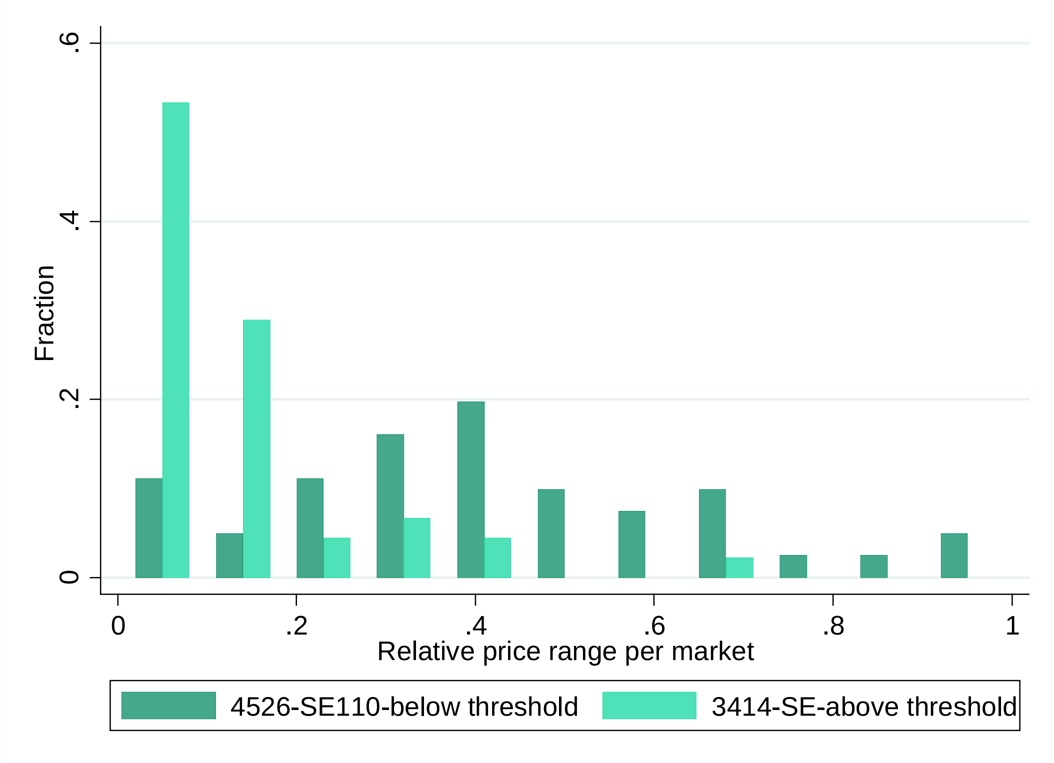
Notes: Excluding those tenders when the relative price range is 0 as these tenders are either marked by the "same bid price" indicator already or there is only 1 non-missing price value for the tender.

By aggregating the tender level relative price range indicator, it is also possible to identify markets of extremely low and extremely high bid price ranges which may be inconsistent with competitive bid price distributions. Once again, outlier values on the level of tenders might be due to a confounding factor unrelated to collusion, while a market level indicator is more likely to indicate systemic market behaviour less prone to random fluctuations and confounders. For example, the 'heavy-duty motor vehicles' market have disproportionately high share of tenders with less than 10% relative price range (Figure 27).[58]

---

[58] It is important to note, that the availability of estimated price can have a compressing effect on bid prices.

Figure 27 Distribution of relative range of tenders for two sample markets (N_{4526}=95, N_{3414}=49, 2009-2014)



Benford's law

Irregularities in bid price distributions can be also analysed by using the overall sample or sub-samples (i.e. markets) of bid prices of public procurement markets. An increasingly widely used statistical test is comparing bid prices to the so called Benford's law (or first digit law). Benford's law posits that the first digit of most naturally occurring sets of numerical data follows a particular pattern[59] Although, it was primarily used first to detect insurance fraud, it was also used for detecting irregular pricing in public procurement markets as well (CRCB, 2016).

Nevertheless, it is important to note, that any conclusion based on Benford's law should be interpreted with caution if the ratio of missing data is high. In case of the Swedish data, this problem can be particularly important as there is almost no product market with sufficiently high ratio of available final or bid prices. Furthermore, prices are often given as unit prices, that are meaningless for this test, as the market price of these products will be driving the prices[60]. On the top of that, government purchases are reported typically when contracts fulfil some pre-set criteria such as contract value or product group; it is easy to see if contracts below

[59] The proportion of 1, 2, 3 etc. numbers as first digit should be proportional to the logarithmic difference between them: $P(d) = \log_{10}(d) - \log_{10}(d+1)$.

[60] E.g. a product that costs 50 EUR/kg on average will lead to bid prices with a 4, 5 or 6 as first digits, hence it will never fit Benford's law.

say a 200,000 EUR threshold are not or only very sporadically reported, the resulting price distribution will be distorted even in the absence of collusion.

Using the Swedish public procurement data, we calculated chi-squared tests in order to test weather prices are following Benford's law or not. In order to identify high and low risk markets, we calculated these tests for each market separately. The first example shows a market following Benford's law, as the distribution of the first-digits (columns) follows the theoretical distribution (Figure 28). Conversely, the second example shows that number 3 is over-, while numbers 5, 6, 7 and 8 are underrepresented among the first-digit of the bid prices (Figure 29).

The surprising overall result is, however, that not only smaller markets (where the tests are less reliable), but larger markets also seem to have distorted prices according to this test. However, as only around 20% of the tenders have bid prices, this result is most probably driven by a non-random sample selection; hence no conclusion can be made by using this indicator.

Figure 28 Distribution of first-digits according to Benford's law and 4523 SE110 market (green columns: observed, red line: expected)

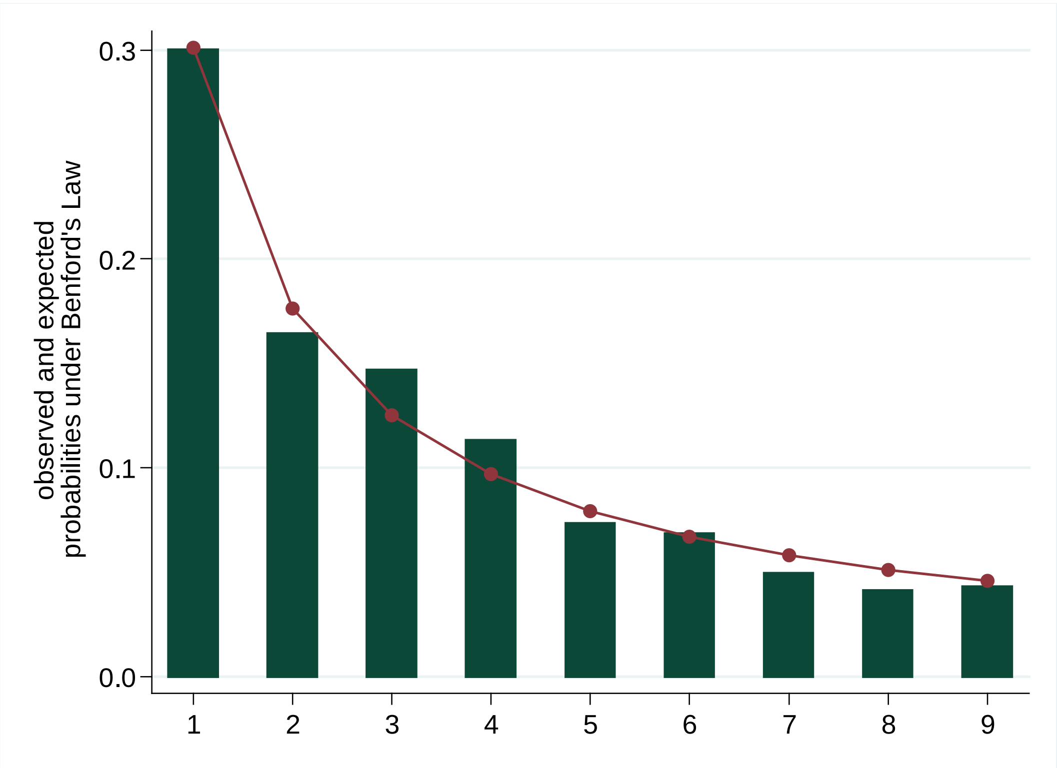
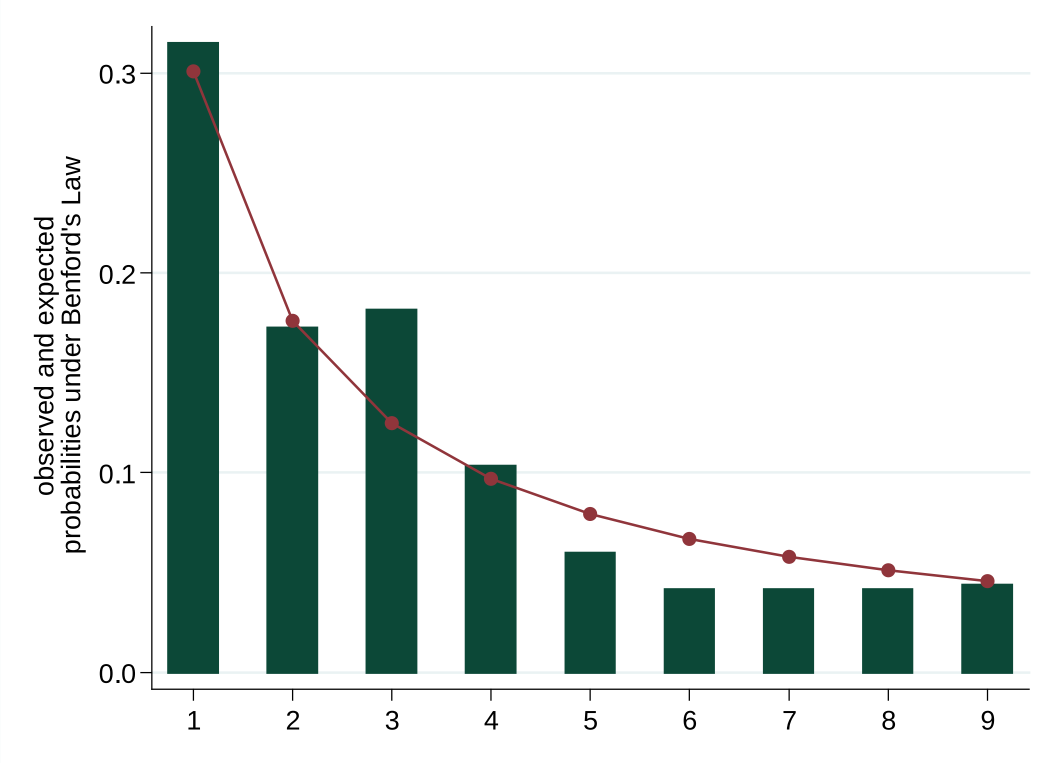


Figure 29 Distribution of first-digits according to Benford's law and 4523 SE211 market (green columns: observed, red line: expected)



3.2.3 Indicators exploiting bidding patterns

While in the previous section we reviewed indicators related to the pricing behavior of participating companies, in this section we discuss how bidding and winning patterns, that is who bids and wins in relation to whom, can be exploited for identifying high risk markets. Four indicators belong to this group:

- Winning probability;
- Cyclical winning;
- Missing bidders; and
- Superfluous losing bidders.

Winning probability

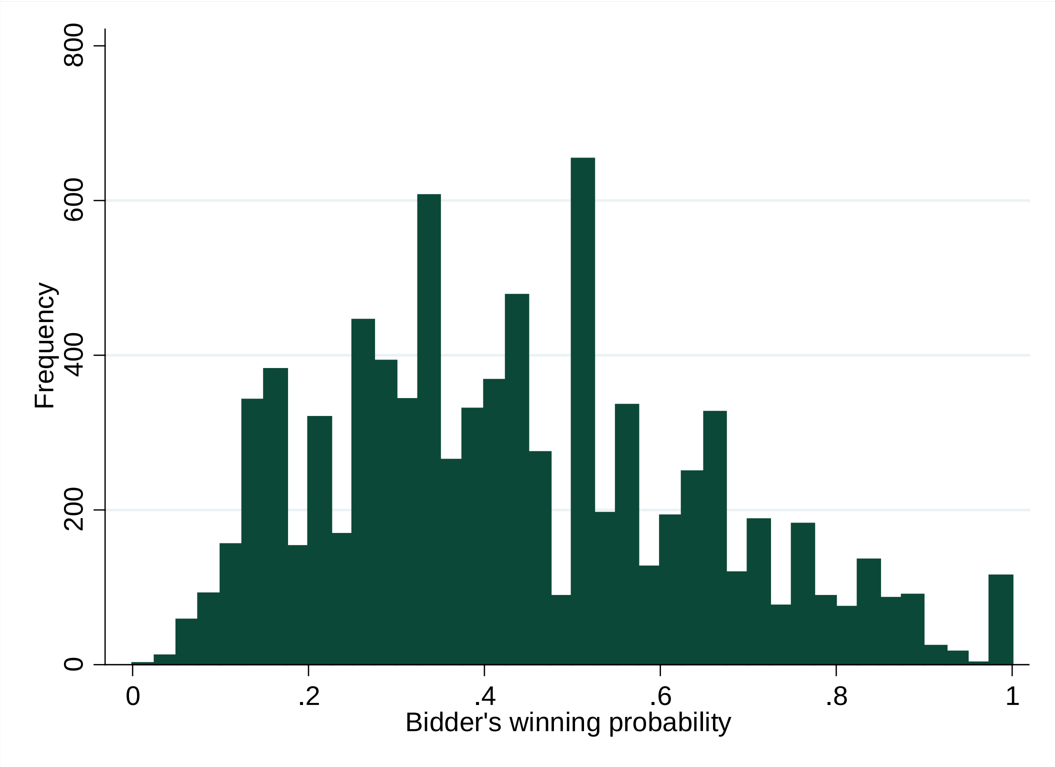
Most companies on genuinely competitive markets win some public procurement tenders while losing others with only very rarely managing a 100% success rate over a long period. However, a collusive ring which artificially allocates all the contracts of a submarket to one company will result in the pre-selected company winning all the contracts it bids for. Hence, a very simplistic indicator of collusion risk is bidders' winning probably with values close to or equal to 1 capturing the risky pattern while all lower values can be regarded as potentially produced by genuinely competitive market behavior.

Winning probability of company i during period t is calculated as

$$Winning\ probability_{it} = (number\ of\ contracts\ won_{it}\ /\ number\ of\ contracts\ bid\ for_{it}$$

As expected from a public procurement market where collusion is the exception rather than the norm, most bidders in Sweden have intermediate winning probability values, with a small minority of about 100 companies winning every contract they bid for over a longer period (Figure 30).

Figure 30 Distribution of companies according to their total winning probability in 2009-2014,whole Swedish public procurement market, only companies with at least 6 bids submitted ($N_{company}$ = 8564)



Cyclical winning

Cyclical winning indicates whether winning patterns are consistent with efficient competitive behaviour or rather with coordination. In the case of collusion free competition, winning and bidding patterns should be based on cost factors such as capacity constraints, or production costs. This implies that more experienced or bigger companies, who have excess capacity should have lower bid offers, that is any company's winning chances at a given tender is related to individual firm characteristics. One should not expect that any company's winning pattern is determined by its co-bidding history with specific companies.

Suspicious bidding/winning patterns represent a rather simple measure of a potentially anti-competitive bidding scheme which is also part of the OECD recommendations (Detecting Bid Rigging in Public Procurement n.d.) as an A,B,A,B winning pattern. This concept is formalized in Padhi and Mohapatra (2011), as they suggest that significant partial autocorrelation in the companies' winning patterns can indicate collusion[61] . Therefore, partial autocorrelation of winning for company c in market m can be calculated as the following:

$$Autocorrelated\ winning\ pattern_{mc} = PACF_{mc}$$

While calculating company level partial autocorrelations might seem straightforward, as contracts need to be sorted according to the "theoretical collusive allocation", its practical use can be problematic. In practice, both the publication date of the call for tenders and the date of actual contract start can be used to sort contracts; without any clear theoretical reason to use one or the other. However, the Swedish public procurement database lacks precise, daily and largely non-missing information on either of this making contract sorting arbitrary and the analysis swamped by noise (Table 3).

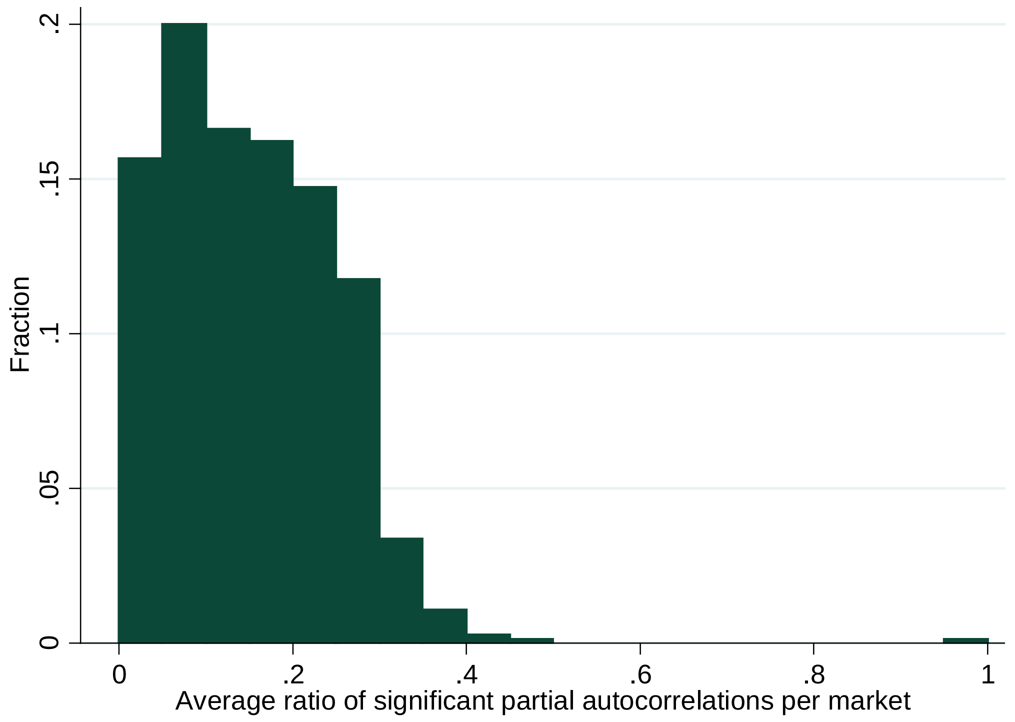
Due to the fact, that many companies only participate and win contracts once or very rarely, the resulting high partial autocorrelations is not surprising (i.e. 0 or close to -1). However, larger companies with an extensive bidding and winning history should not have auto correlated winning patterns throughout longer periods. In case of Sweden, even if we restrict the sample to markets with more than 30 contracts awarded, it is apparent, that there are markets with high average ratio of significant auto correlated winning patterns (Figure 31).[62]

Unfortunately, due to the high percentage of missing winning and bidding firm information, as well as the fact that we cannot be sure that missing information is fully random, this indicator cannot be further developed despite its high potential.

[61] Obviously, auto correlated winning patterns can be explained by switching costs etc. Therefore, this should be further validated by other indicators as well.

[62] For the autocorrelation related calculations, we restricted the sample to NUTS-3 level markets with at least 30 contracts awarded and calculated company level partial autocorrelations of a variable capturing company winning (1 if company *i* wins 0 if not). Contracts were sorted by call for tender publication date as a primary and contract start date as a secondary determinant (e.g. in case two tenders having the same call for tender publication date, the one with an earlier contract start date was listed as being earlier). Second, we counted and aggregated every significant partial autocorrelation up to 10 lags per company and calculated the their average value for each NUTS-3 regions.

Figure 31 Average ratio of significant partial autocorrelation of company winning patterns within the first 10 lags by NUTS-3 markets (N = 740, years: 2009-2014)



Missing bidders

Withholding bids from certain tenders is a straightforward way to restrict competition, hence missing bids of a previously active company at a given market can indicate collusive bidding. Despite the limited scope of empirical research on collusive schemes using missing bids, it should be considered, as this is a bidding strategy which leads to the same market outcome as using faulty or overpriced bids. Furthermore, this technique is a widely pointed out as being suspicious (OECD 2014; SCA 2015b).[63]

A simple indirect indicator for missing bidders is the prevalence of single bidder contracts that is when all, but one bidder refrains from bidding. As this indicator is widely thought to signal corruption (Charron et al. 2016), and that it would rather oversimplify the modus operandi of collusive schemes, we focus on the overall decrease in the number of bids and the geographic scope of companies' bidding activities.

[63] Legal considerations may also increase the prevalence of withheld bids as a collusion technique: sometimes, there are clausules in public procurement regulation that bidding (but loosing) companies cannot participate in contract implementation as sub-contractors. Therefore, in order allow for rent redistribution through sub-contracting withholding bids is often a necessity condition.

There are two straightforward indicators of withdrawn bids that can be calculated and compared widely across markets: yearly changes in average bidder numbers, and yearly changes in the geographical scope of company bidding. Both of these indicators point at the same malpractice, however from a different point of view: markets or companies. The market-focused view is indifferent to geographical divisions or product market-wise divisions used by colluding firms to split up markets, while it is sensitive to the appropriateness of the market definition applied. Whereas, the company-centered indicator only considers the geographical split of markets among colluding firms, but it is indifferent to the specific product market definition applied. Formulating these two collusion proxies in terms of change over time makes them geared towards detecting the inception or demise of collusive rings.

In order to have a simple measure for bidding activity we calculated the year-on-year differences in the average number of bidders per tender for each NUTS3 region market:

$$\begin{aligned} &Annual\ change\ of\ the\ average\ number\ of\ bids\ per\ tender_{mt} \\ &\quad = Average\ number\ of\ bids\ per\ tender_{mt} \\ &\quad - Average\ number\ of\ bids\ per\ tender_{mt-1} \end{aligned}$$

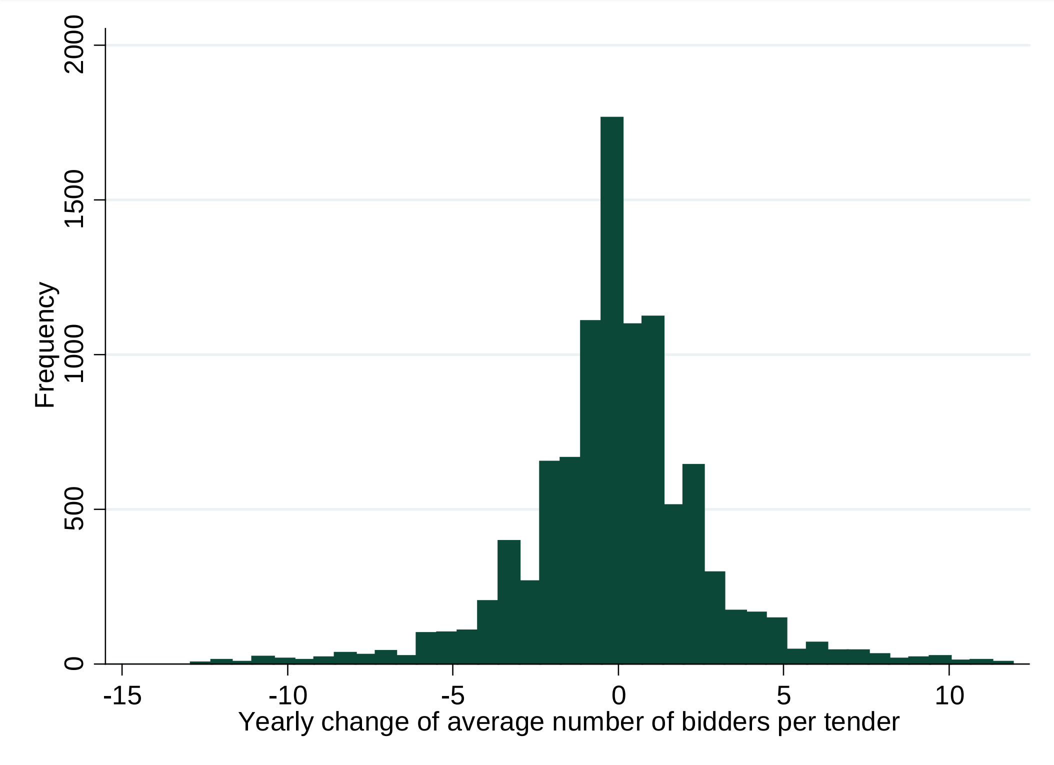
Extremely low values of this indicator suggest the creation of a collusive scheme while extremely high values its demise, with intermediary values deemed most likely due to natural fluctuations in bidding activity.

While the annual changes were in line with our expectations as most of the markets have only relatively small changes in the average number of bidders, it is apparent, that a small number of markets have rather significant changes – i.e. more than 5 bids less per tender etc.[64]

---

[64] Some these extreme values are partly driven by special tenders such as framework agreements etc., therefore, each market has to be scrutinized in detail to exclude such cases.

Figure 32 Annual change of the average number of bidders per tender per NUTS-3 markets ($N_{market\text{-}year}$ = 10,302, years: 2009-2014)



Following the very same logic outlined above, focusing on the geographical scope of the companies' bidding activity can reveal another crucial aspect of coordinated bidding. Coordinating the geographical scope of bidding is suggested by many to be an important dimension of coordination (Detecting Bid Rigging in Public Procurement n.d.; SCA 2015b).

By taking the year-on-year difference of the number of different NUTS-3 regions where a company submitted a bid, we can obtain a rough measure of the change in bidding companies' geographical scope.[65]

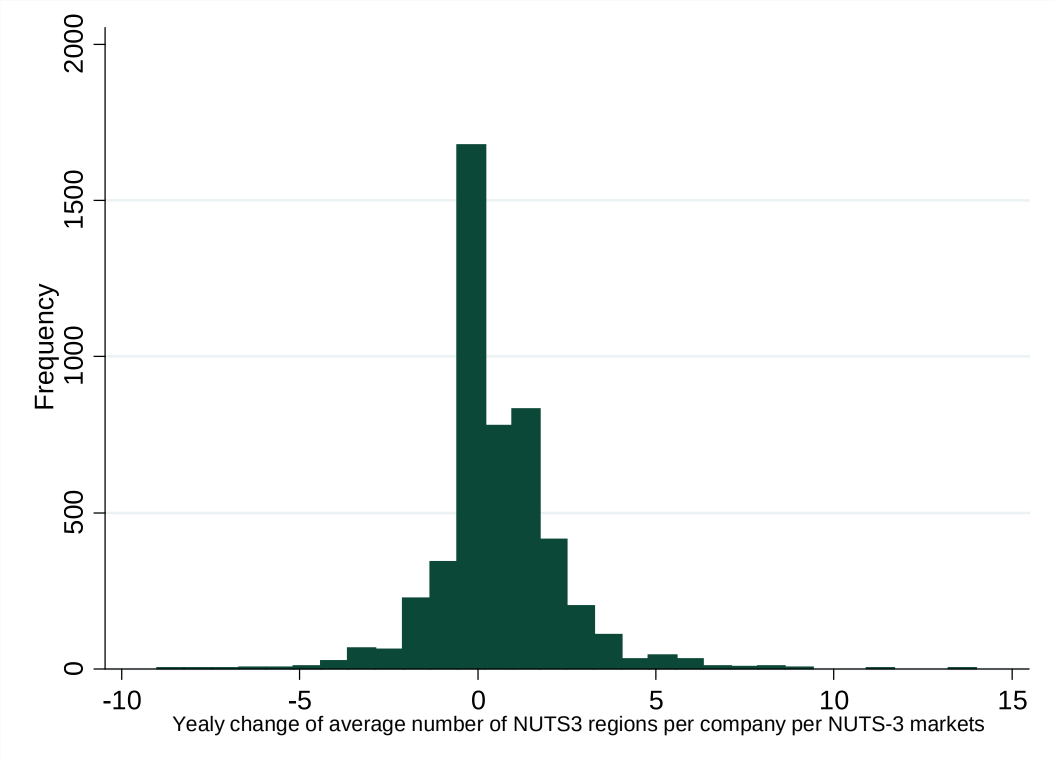
$$\begin{aligned} &Yearly\ change\ of\ the\ \#\ of\ NUTS3\ regions\ a\ company\ placed\ a\ bid_{it} \\ &\quad = Number\ of\ NUTS3\ region\ a\ company\ placed\ a\ bid_{it} \\ &\quad - Number\ of\ NUTS3\ region\ a\ company\ placed\ a\ bid_{it-1} \end{aligned}$$

As above, extremely small values indicate the creation of a collusive scheme, while high values its demise. While company level changes can already be informative on their own, summing up the yearly changes by NUTS-3 market make this indicator

[65] We calculate a weighted geographical scope based on the companies' bid coverage of NUTS-3 regions. The maximum yearly number of geographical coverage is 21, as there are 21 NUTS-3 regions in Sweden. NUTS-1 and NUTS-0 regions are weighted according the number of NUTS-3 regions. For example, if a company submits a bid in SE2 (South Sweden), its value is 8, as there are 8 NUTS-3 regions within SE2 region. If the same company bids only in SE221 and SE213 regions next year, its geographical scope will be 2. Based on the indicator definition, its yearly change will be -6.

directly commensurate to the previous proxy and it can reveal whether companies bidding on the same product markets start to bid less from one year to another. It is apparent, that there are several NUTS-3 regional markets with companies bidding on significantly less and less NUTS-3 regions overall (Figure 33).

Figure 33 Annual change of the average number of NUTS-3 regions where companies have submitted a bid per NUTS-3 markets ($N_{NUTS\text{-}3\ markets}$= 4,865 , years: 2009-2014)



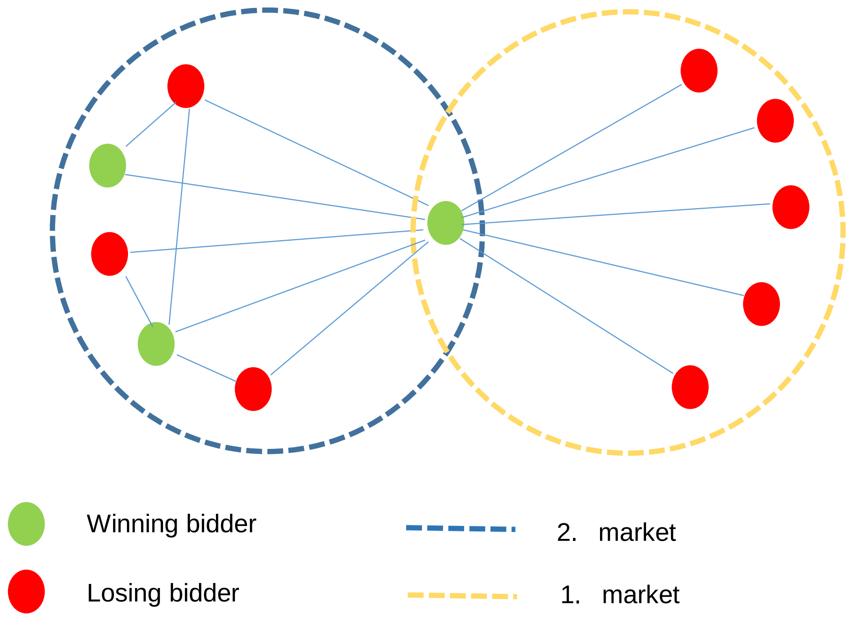
Superfluous losing bidders – network analysis

One of the most straightforward ways to mimic competition while in fact coordinating bidding and pre-determining who wins is when competitors recurrently submit losing bids making a pre-selected company a sure winner. This indicator captures such a situation.

In a truly competitive context more than one company is expected to win over time, while companies losing throughout a prolonged period are unlikely to keep bidding. Thus, observing a dominant company winning always or almost always while a set of 'competitors' bidding, but losing always or almost always may signal collusion. In such a scheme, bids can lose either because they contain deliberate errors disqualifying them, or because they deliberately achieve lower scores than the winning bid (i.e. higher price and/or lower quality). In an ideal database, both types of losing bids could be separately identified allowing for developing additional, but linked indicators such as the prevalence of faulty bids which however is not available for Sweden (for an example how this indicator is used see (Tóth et al. 2014)).

When colluding firms control the entire market and the company extracting rents (i.e. winning) is the same over a longer period, it is possible to identify distinct network formations underlying this kind of collusion. In a co-bidding network, that is a network of bidders where each tie represents a tender where two companies co-bid, the cleanest manifestation of this type of collusion would result in a so-called cut-point formation (Figure 34, right-hand side). In network terms, cut-points are vertices whose removal from the network would cut off other vertices; in other words, eliminating the cut-point would make the whole network falling into two sub-graphs (Wasserman 1994). In such a formation, it is the cut-point which is expected to win always or almost always, while those companies which are linked to the rest of the network solely through the cut-point are expected to lose always or almost always.

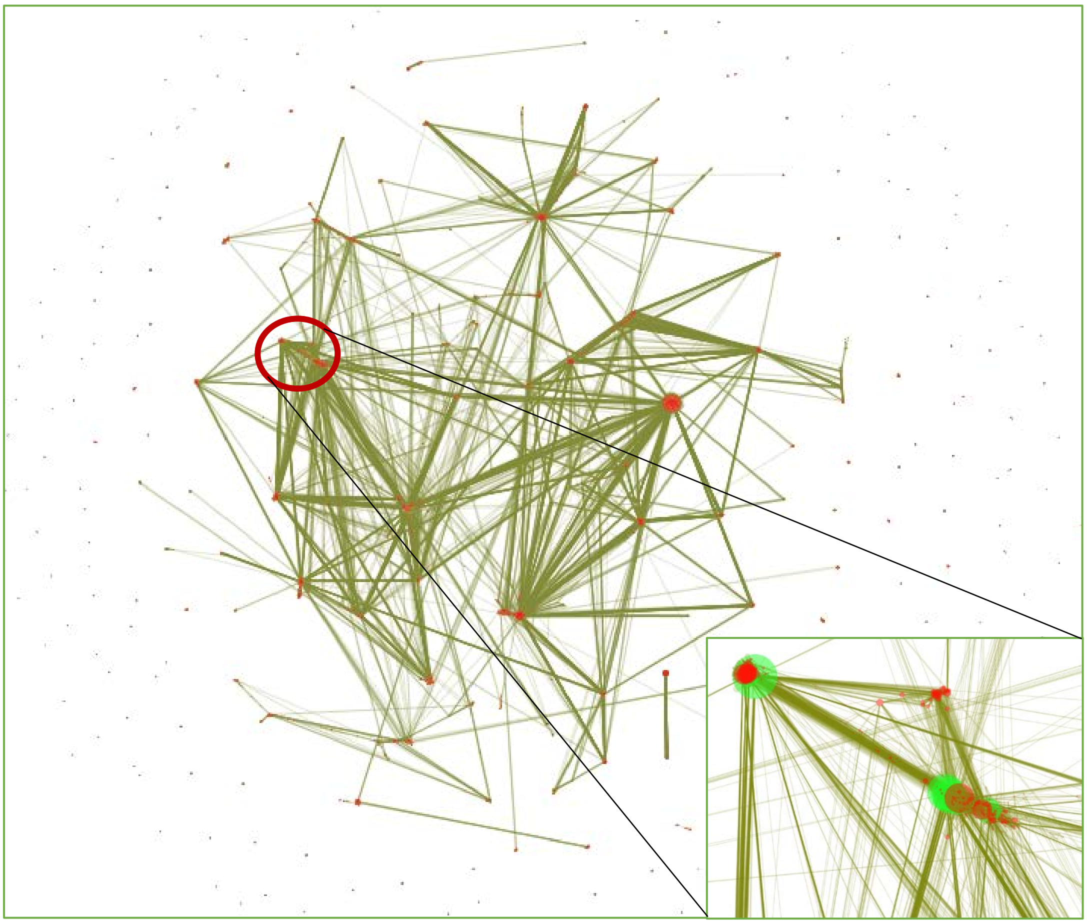
Figure 34 Schematic representation of a cut-point network formation



As Figure 34 highlights, each of the companies may be present at multiple markets allowing for colluding and genuinely competitive behaviour at the same time depending on the market, hence the identification of the cut-point network formation crucially depends on the appropriate definition of markets. In the absence of a robust understanding of the relevant market definition, cut-points can be sought by cycling through different market definitions from the entire public procurement market to very specific markets (Figure 35). By narrowing the market definition, the likelihood of cut-point formations increases as fewer genuinely competitive sub-markets are included. Cross-referencing to other collusion indicators should guard against arriving at an unreasonably high number of cut-points and a correspondingly too narrow market definition.

A major shortcoming of the Swedish public procurement dataset is that it certainly misses a number of bidder and winner company IDs both of which biases network statistics. Hence, the below analysis is only indicative of potential risks and serves as a demonstration of the method's potential.

Figure 35 Co-bidding network of the entire Swedish public procurement market in 2013, depicting bidders with at least 1 tie (participating in at least one tender with at least 2 bidders)



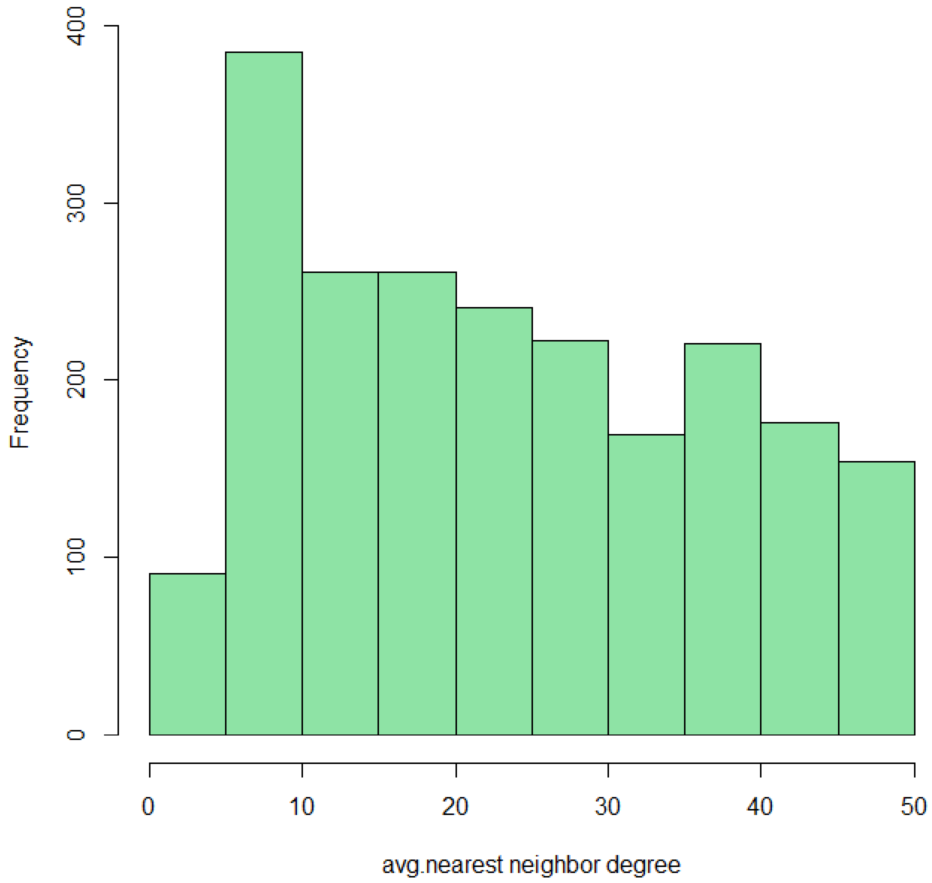
Note: the size of vertices show the number of times a company has won a contract. Green vertices are in a cut-point position and winning at least 3 contracts.

While the cut-point indicator represents a very clean and easy-to-interpret collusion proxy, it is vulnerable to mixed strategies, say colluding companies collude only on a particular, hard-to-detect sub-market while compete genuinely on other related markets (consider for example on Figure 34 a situation when more than one company of market 1 bids on market 2, still market 1 is dominated by colluding bidders). Such a scenario would lead to the cut-point indicator hugely underestimating the incidence of collusion. In order to mitigate this shortcoming, an alternative indicator is also developed following the same reasoning, but building on a more flexible approach. In network theoretical terms, the collusive ring depicted in Figure 34 (market 1) is such that the pre-selected company winning the contracts is central in the sub-graph while the companies submitting losing bids are peripheral with very few connections to each other or to the rest of the network. This concept can be captured by a range of network indices such as Burt's constraint, average nearest neighbour degree, or local clustering (it is sometimes called local transitivity too) (Barrat et al. 2004; Burt 2004). We will use average nearest neighbour density as it matches most closely our theoretical concept and it is the most straightforward we can develop:

$$k_{nn}(k_i) = \frac{1}{k_i}\sum_{j=1}^{N} A_{ij}k_j$$

Where $k_{nn}(k_i)$ denotes the average nearest neighbour value of vertex *i*; k_i expresses the number of ties vertex *i* has; k_j expresses the number of ties vertex *j* has; and A_{ij} marks if vertices *i* and *j* are direct neighbours or not (1 or 0). The distribution of this indicator across the whole public procurement market displays a marked drop at the lower end of the distribution which is collusion risks are expected to increase (Figure 36). These high risk nodes also partially overlap with cut-points as identified above while also pointing at a range of further risky companies.

Figure 36 Average nearest neighbour degree distribution of bidders with at least 5 ties and 3 won contracts (distribution truncated to 0-50 range), entire Swedish public procurement market in 2013



3.2.4 Indicators exploiting market structure information

As it was already discussed above, collusive bidding leads to inherently different pricing and bidding patterns compared to the competitive markets. This section adds to our understanding of collusive behaviour by discussing the third group of collusion proxies, which capture variation in market structure. Suspicious, seemingly non-competitive market structures can be twofold:

- Concentrated market structure; and
- (Artificially) stable market structure.

While both of these competitive outcomes can arise under non-collusive competitive conditions, they are much less likely to arise, and even more unlikely to persist over a long period.

These indicators are particularly sensitive to our prior market definition matching the market scope collusive companies control. Crucially, when collusion results in splitting up the market among the members of the ring, sub-markets have to be considered which are typically defined along geographic or product type dimensions. Without adequately identifying these sub-markets, suspicious patterns might not be identifiable.

Concentrated market structure

One of the typical results of collusive bidding is concentrated market structure as opposed to a competitive market with multiple players. Concentration in a public procurement market refers to a single or a few companies winning all the contracts while competing bidders are either entirely absent or only mimic participation. It can be taken as a sign of collusive behaviour if it takes place on an otherwise competitive market. Concentrated market structure can also be closely linked to rent sharing methods. As Pesendorfer (2000) shows, when tender completion is done by the most efficient companies (though not competitively) in order to reap the largest profits possible, market shares can rise. By implication, concentrated market structure should be defined with reference to an elevated market concentration compared to a competitive situation. A clear-cut situation when concentration signals collusion is when a particular market turns from competitive to a concentrated one in a short period of time without any apparent alternative explanation such as changing regulations, technology, or steep decline in total demand.

As for many other dimensions of collusive schemes, concentrated market structure can be measured by several indicators, such as the market share of the largest or combined market share of the four largest companies (C4), or the also widely used HHI index. We used the following formula:

$$C1_{mt} = \frac{Total\ contract\ value\ of\ the\ largest\ company_{mt}}{Total\ value\ of\ contracts\ awarded_{mt}}$$

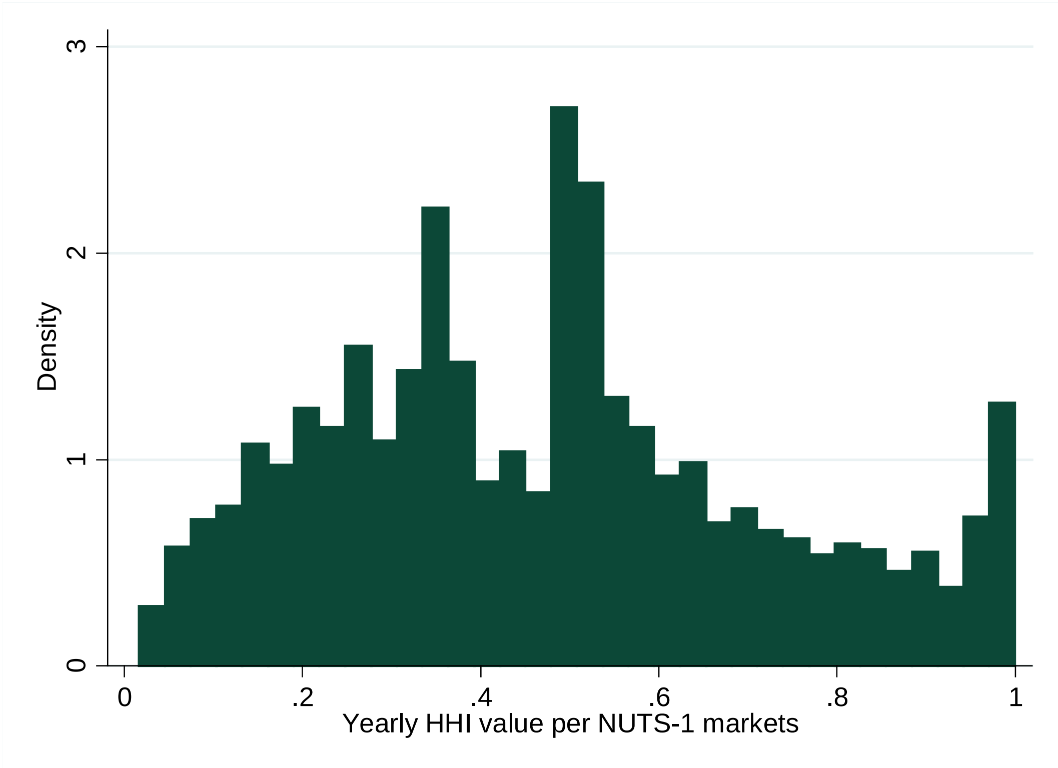
$$C4_{mt} = \sum_{i=1}^{4} S_{imt}$$

$$HHI_{mt} = \sum_{i=1}^{n} S_{imt}$$

Where S_{imt} denotes the market share of company *i* in market *m* and period *t*. Unfortunately, as it was discussed in section 2.1, final contract values are only available in around 20% of the tenders. Furthermore, as neither of the markets has disproportionately high ratio of available final prices, the calculation of yearly market shares have to rely on estimates.[66]

[66] The two anonimised markets provided by SCA often disclose unit prices only, which makes calculating market shares inaccurate.

Figure 37 Yearly HHI values per NUTS-1 market (N=2,631)[67]



Stable market structure

Stable market structure means that the variance in market shares in a market is artificially low, which is not consistent with natural, competitive market outcomes. Athey and Bagwell (2001) or Athey et al. (2004) shows that following a market share rule for allocating rent can be also an optimal conduct in collusion. Regarding empirical studies, (Pesendorfer 2000) shows that when instead of side-payments the rent reallocating mechanism is bid-rigging, relatively stable market shares can be observed. Mena-Labarthe (2012) also shows that in the collusive period, the market shares of the colluding parties were practically the same[68]. In Harrington (2008), two relevant collusion indicators are introduced based on market structure: highly stable market shares over time and highly stable market shares of a subset of firms.[69]

Building on the indicators of concentrated market structure, stable market shares can be identified by measuring the stability of the same market structure indicators (C1, C4, and HHI); for example, by calculating their year-on-year standard deviation. While we are facing the same problem as in the case of concentrated market structure – i.e. lack of detailed enough data – it is apparent that the stability

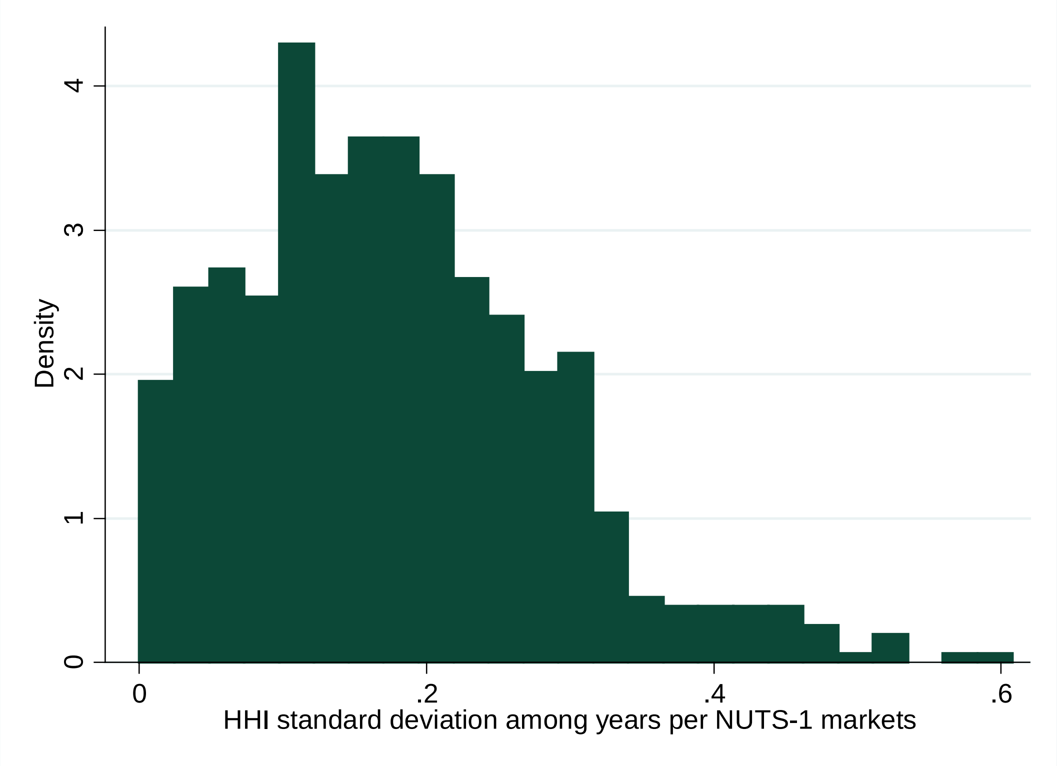
[67] Market-year pairs where HHI value is 0 or 1 are excluded as they are more probably due to insufficient price data.

[68] Mena-Labarthe (2012) investigates a Mexican pharmaceutical cartel.

[69] Also, Harrington (2006) argues that market shares are often fixed at the pre-collusion levels.

of market structure varies a lot by markets (Figure 38). It is apparent, that the "natural" state of each market is accompanied by a certain annual variation of market shares (i.e. around 0.1-0.2). However, there are markets with extremely limited market structure changes over time.

Figure 38 Standard deviation of NUTS1 level markets' HHI across years (N: 632, years: 2010-2014)



3.3 Validity of elementary collusion risk indicators

While each of the above indicators have been tested in at least some academic research papers and many of them have also been endorsed by practitioners and policy expert groups such as those of the OECD, their validity is far from warranted in general, but also in the specific Swedish case. Certainly, hard evidence on actual collusion could go a long way in underpinning indicator validity (e.g. Conley and Decarolis 2016), however such evidence is not available in the Swedish database and court sentences most likely only represent a biased sample of collusion cases making any interferences problematic.

In the absence of hard evidence on collusion cases, indicator validity has to rely on statistical relationships between

1. Collusion risk indicators and external indicators of validity; and
2. Collusion risk indicators among each other.

First, for conducting external validity tests, the most appropriate indicators are those which come from an independent source than the bidding data we use and are linked to collusion outcomes rather than the process of collusion itself to avoid merely correlating two sides of the same event. There are two sets of indicators which fit the bill, i) indicators of prices and ii) indicators of financial performance. Both of these are intimately tied to collusion outcomes which are expected to be higher prices and better company financial performance.

Indicators of prices are unfortunately not readily available in the Swedish database. In many procurement databases, however, relative contract value (awarded contract value / estimated contract value) or winning rebate (1 – [awarded contract value / estimated contract value]) can serve as a general indication of prices assuming that prior estimates are non-biased (Coviello and Mariniello 2014). Moreover, for some markets such as road construction unit prices are also available providing a much more reliable measure of over-priced contracts (Fazekas and Tóth 2017).

Indicators of financial performance are available in Sweden from the official company registry records of annual financial files. We merged the public procurement data to company records using bidder ID as assigned by Visma Opic. While both the company registry database and the procurement records have missing data (e.g. only 60% of tenders has valid bidder IDs), the amount of matched company financial and procurement information allow for a meaningful analysis. Among many indicators of financial performance, two are expected to vary closely with public procurement collusion when a company's public procurement income represents a considerable share of its total income: i) profit rate or profit rate growth, and ii) total profit growth. Depending on the type of rent sharing mechanism used either or both of these indicators could move together with collusion. If rents earned by a dominant firm are shared using subcontracts or side payments company profitability may not increase drastically rather its turnover and overall value of profits as subcontracting or side-payment costs also rise with procurement income. Whereas when rents are shared in a joint or coordinated bidding arrangement, each participant's profit rate is expected to increase, and most likely their overall profit too.

Second, for conducting validity tests by exploiting the expected co-variation among elementary collusion risk indicators, we need to rely on the theoretically defined collusion types. Each of these types implies a set of relevant collusion proxies with many further proxies either irrelevant or even moving in the opposite direction. For example, in collusion types which use coordinated bidding such as ABAB type cyclical winning patterns, indicators of increased market concentration are likely to go down or stay the same as under the previous competitive scenario.

Below, we carry out both of these tests of validity tests. For collusion proxies and external validity tests, we make use of each elementary collusion risk indicator as defined above (simple validity), while for the validity tests exploiting co-variation among elementary collusion risk indicators we only highlight selected examples (complex validity). The main reason for this is that there are too many data problems and missing indicators for meaningfully conducting a systematic analysis of indicator co-variation.

Our approach is conservative; we expect correlations between elementary collusion risk indicators and company financial performance as well as among selected collusion risk indicators. This approach assumes that each elementary indicator is sufficient on its own to indicate collusion albeit, with a certain degree of measurement error. An alternative and quite sensible approach is to treat each elementary collusion risk indicator as a necessary, but not sufficient indication of collusion and defining combinations of indicators, or composite scores, which together much more reliably mark collusion. Due to many missing data points and indicators this latter approach is not explored in detail, further work could certainly expand in this direction.

3.3.1 Validity tests of individual indicators

This section presents the results of simple correlational analysis between elementary collusion risk indicators and company financial performance, taking collusion risk indicators as categorical denoting extreme values as risky and all other values as non-risky. This set-up implies a simple comparison of group averages using three financial performance indicators:

- Before tax profit rate;
- Before tax profit rate growth; and
- Before tax profit growth.

Such a simple approach naturally oversimplifies the empirical relationships between collusion proxies and validating indicators (e.g. neglecting procurement income share in total company income), hence they are only indicative of validity. Nevertheless, these tests are relatively easy to conduct, and repeated later on, and due to their relative simplicity they are comparable across indicators and samples without recourse to regression analysis or matching algorithms.

Group average comparisons made below depend on defining adequate benchmark markets (assuming that the collusion risk indicator is precisely formulated). As discussed above, we can consider 3 different types of control groups which are considered relatively comparable:

- same product market, but different region
- different product market, but same region
- same product market and region, but different time period

Any of these control or comparison groups run the risk of also including colluding firms which were not adequately picked up by observed collusion risk indicators, as the collusive ring is likely to operate in similar markets, but collusion proxies might pick up more noise in some markets then others. By implication, the total procurement market can also be considered as an adequate control group. Naturally, the exact choice of control group influences the magnitude of group differences, fortunately in most cases without substantially altering the qualitative findings.

Before discussing the results of simple validity tests, a caveat applies: there are substantive data constraints such as i) data on prices is only available for 20% of tenders while many valid values are only given as unit prices and lots cannot be distinguished (on data problems see section 2); and bidder IDs are only available for about 60% of tenders. If such errors are non-systematic and uncorrelated with collusion, point estimates should still be unbiased; however, we cannot know if these conditions hold.

Overall, most elementary collusion risk indicators are supported by basic validity tests, except for the relative standard deviation of bid prices indicator where differences are either insignificant or negative, that is contrary to our expectations (Table 13). For the superfluous losing bidders indicators, group differences are not always significant, due to the very precise nature of these indicators marking only a handful companies. While in terms of yearly changes, the direction of profit and profit rate change is confirmatory, the profit rates of high-risk cut-point companies are significantly smaller. Unfortunately, 4 indicators couldn't be reliably tested even at this basic level due to insufficient data quality: 1) Benford's law, 2) Bid rotation, 3) concentrated market structure, and 4) stable market structure. The magnitude of group differences greatly vary and indeed partially depend on the control group chosen, still in general they are significant at conventional levels and of substantial size, ranging between 9.5 and 37 percent difference. As underlined earlier, depending on the rent sharing mechanism used, not all three financial indicators are expected to follow collusion risks, a pattern which we could see in the empirical results. This means that typically high collusion risk companies are either much more profitable (profit rate difference) or they increase their profitability greatly (profit or profit rate growth rate differences).

Table 13 Validity test summary: elementary collusion risk indicators and company financial performance (differences expressed in percent change compared to control group)

Indicator name	Collusion proxy formulation	Before tax profit rate difference	Before tax profit rate growth difference	Before tax profit growth difference	Level of aggregation	Control group
Bidders having the same bid price	binary	15.6%**	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 %	15.5%***	16.0%***	Company	Companies without outlier bid price difference
Relative standard deviation of prices	outliers: bottom 5%	- 11.8%	-60%***	- 10.9%	Company	Companies without outlier average relative standard deviation
Bid price range	outliers: bottom 25%	-2.9%	17%***	13.7%***	Company	Companies without outlier price range
Benford's law	data quality is insufficient					
Winning probability	outliers: top 25%	37%***	18.5%***	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%	12.6%***	- 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	-39.3%**	41.4%	41%	Company	Companies not in a cut-point position
Superfluous losing bidders (nearest neighbours)	outliers: top 10%	-50.8%**	4.2%	14%	Company	Companies without outlier nearest neighbour degree
Concentrated market structure	data quality is insufficient					
Stable market structure	data quality is insufficient					

Notes: Significance levels: ***1%, **5%, *10%. Validity tests for the indicators of superfluous losing bidders were based on 2013 data only. In case of winning probability and superfluous losing bidders (both cut-points and nearest neighbours) only companies with at least 5 bids are included.

3.3.2 Complex validity analysis

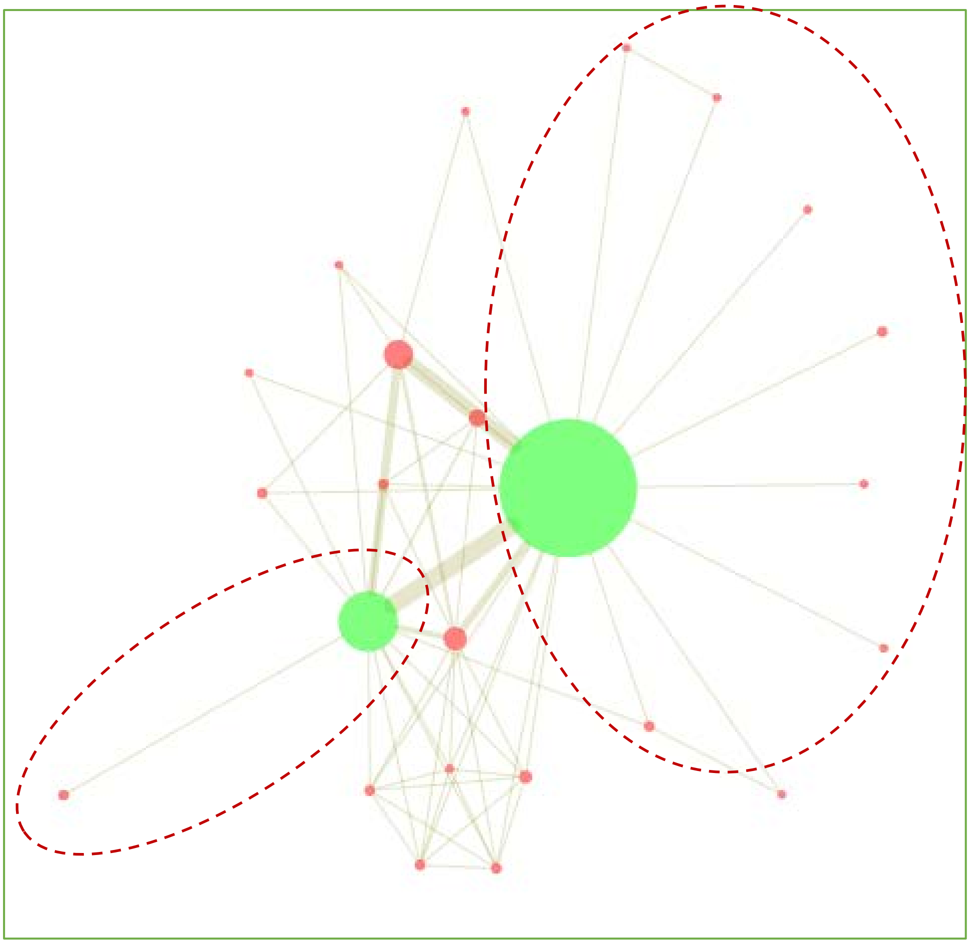
A full-blown analysis formally investigating the covariation between different collusion risk indicators as well as with company financial performance indices is not performed at this stage due to data quality issues. We envisaged using explorative Principal Component Analysis to formally identify suitable ways of combining indicators. For selected sub-markets with higher quality data some of this analysis might be possible, even though results would remain largely tentative.

Nevertheless, in order to outline the logic of analysing the co-variation among collusion risk indicators and precisely identify the indicators currently not collected in Sweden, we systematically discuss collusive scheme types and the corresponding indicators (Table 14). Solving the issues of missing variables and data quality would lead to a reasonably complete collusion detection framework already in the near term.

In order to demonstrate the power of combining collusion risk indicators, one high risk market is discussed in detail which has somewhat higher, albeit far from perfect data quality. The analysis starts off with identifying high risk bidding patterns, that is companies in a cut-point position (Figure 39). The analysis returns two high risk companies which potentially act as main contractor while rewarding the companies submitting losing bids through subcontracts or informal side payments. The companies in a cut-point position are also the most successful companies as indicated by the number of contracts won by them (around 3 quarters of all the contracts awarded on the market). Note the close matching of the empirical network below and the theoretical graph in Figure 34.

In order to gather further evidence of potential collusion on this market, indicators of price distributions are calculated for the supposedly colluding firms and the non-collusive part of the network (i.e. comparing the two large green nods' tenders when they bid with their supposedly losing rim or each other; with the rest of the network). This comparison yields a significant and substantial difference in line with our theoretical expectations: on high risk tenders, the relative price range is 50% smaller than in low risk tenders, that are 0.34 compared to 0.17 (significant at 10% level). A further piece of evidence could be added if company profitability data could be matched to the currently anonymised database on the company level. In the absence of company IDs, we simply compared profit rates of all the companies in the market with all the companies on the other chemical markets of Sweden in order to compare like with like. This arguably imperfect test suggests that companies on the high collusion risk market have a considerably higher profit rate: 11.3% versus 3.6% (the difference is significant at the 1% level) in line with our expectations.

Figure 39 Co-bidding network in market 1 (market data manually collected by SCA), 2009-2014. Cut-points and their linked periphery highlighted by red circles



Note: node=bidding firm; node size=number of contracts won; green node=cut-point position; red node=non-cut-point position; ties=joint bidding on a tender; tie width=number of times two companies co-bid on a tender

Table 14 Summary of collusive scheme types and their specific proxies

Collusion type	Market structure	Technique	Rent allocation	Indicators calculated	Indicators not calculated	Measurement problems
A	Concentrated	Withheld bids	Sub-contractor	Missing bidders Concentrated market structure Winning probability	Prevalence of subcontracting Relative contract value	-Missing contract values -Missing bidder IDs
B	Concentrated	Losing bid	Sub-contractor	Superfluous losing bidders Concentrated market structure Winning probability Relative difference between lowest and second lowest bid price Relative range of offer prices Benford's law	Prevalence of subcontracting Relative contract value	-Missing contract values -Missing bidder IDs -Reporting threshold truncates sample
C	Concentrated	Joint bid	Consortia	Missing bidders Concentrated market structure Winning probability Relative difference between lowest and second lowest bid price Relative range of offer prices	Prevalence of consortia Relative contract value	-Missing contract values -Missing bidder IDs

Summary of collusive scheme types and their specific proxies (continued from previous page)

Collusion type	Market structure	Technique	Rent allocation	Indicator calculated	Indicator not calculated	Measurement problems
D	Stable	Withheld bids	Sub-contractor	Missing bidders Stable market structure	Prevalence of subcontracting Relative contract value	-Missing contract values -Missing bidder IDs
E	Stable	Losing bid	Sub-contractor	Superfluous losing bidders Stable market structure Relative difference between lowest and second lowest bid price Relative range of offer prices Benford's law	Prevalence of subcontracting Relative contract value Ratio of faulty bids	-Missing contract values -Missing bidder IDs -Reporting threshold truncates sample
F	Stable	Withheld bid	Coordinated-bidding	Missing bidders Stable market structure Cyclical winning	Relative contract value	-Missing contract values -Missing bidder IDs
G	Stable	Losing bid	Coordinated-bidding	Superfluous losing bidders Stable market structure Relative difference between lowest and second lowest bid price Relative range of offer prices	Relative contract value Ratio of faulty bids	-Missing contract values -Missing bidder IDs

4 Policy recommendations

4.1 Improving the public procurement data environment in Sweden

Adequate, timely, and publicly available public procurement data is a fundamental precondition for controlling governments, efficient use of public resources, and tracking risks of collusion. Unfortunately, the Swedish system does not meet the basic requirements of a well-functioning public procurement data framework such as those standards the EU and OECD set for countries wishing to join the EU (OECD/Sigma 2014).

In order to support the capacity of the Swedish Competition Authority to detect collusion in public procurement and more broadly steer the Swedish data system towards European best practice, we identified the below areas of policy development. These recommendations represent an ideal, but achievable scenario, while implementing only some elements is already expected to produce positive effects. If the recommendations are implemented simultaneously, their positive effect is further strengthened through a number of synergies.

1. **Set up a central public procurement platform publishing all tender notices which are regulated by national public procurement law**. This is expected to increase market transparency, decrease transaction costs, and facilitate government accountability.

2. **Ensure a uniform data capture process and safeguard data quality, ideally using the central public procurement platform and under the supervision of a dedicated public agency.** Directly controlling information recording processes and checking data quality at the point of data entry gives sufficient control to the government over data content and quality to support reliable and timely public information provision. A dedicated public body such as a national public procurement agency is best placed to provide guidance and monitor the full data generation process and safeguard data quality as it can reap economies of scale in each of these tasks.

3. **Introduce standard forms precisely defining the minimally expected publication content, while preserving procuring entities' freedom to include additional data.** Defining the exact information content which is expected to be readily published further strengthens data comparability, quality, and depth with as little administrative burden as possible.

4. **Introduce the requirement to systematically collect essential information on contract implementation such as contract modification, final total contract value and actual completion date**. Producing reliable data on contract performance would allow the public as well as

the Swedish Competition Authority to monitor the full tender cycle, hence avoid major gaps in the data landscape. For the ideal variable list and standard forms see Table 15. All relevant contract implementation information is easiest to compile and publish in a contract completion announcement published shortly after the contract is complete.[70]

5. **Introduce the requirement to systematically collect additional key data points supporting collusion risk analysis such as data on consortia and subcontracting**. The inclusion of collusion proxies necessary to complete the measurement framework among the regularly reported data points would greatly enhance the Swedish Competition Authority's capacity to set up an effective early warning system. For the ideal variable list and standard forms see Table 15.

6. **Lower threshold for direct awards closer to the level in best practice countries in order to increase competition for public contracts**. While a lower monetary threshold for direct award would certainly increase administrative burden, scientific evidence is solid on the beneficial effects of increasing competition and transparency on bidder numbers and bidder composition.

7. **In order to decrease transaction costs for citizens and businesses create a central repository of contracts and bids proactively disclosing them.** Given that public access to contracts as well as submitted bids is already granted by law, a central repository of all such information would drastically decrease transaction costs and further facilitate the use of such information by the public as well as government agencies.

8. **On the short term, make sure that Swedish procuring entities fulfil their legal obligations to precisely report public procurement information in TED as set out in the EU Public Procurement Directives**. As currently Sweden produces one of the highest error rates in TED data, it fails to make full use of the existing central TED platform to the detriment of open and fair competition as well as government accountability.

9. **In support of future potential policy reform, estimate the opportunity cost of the current fragmented publication practice compared to a central portal.**

---

[70] For an example contract completion standard form from Hungary see: http://www.kozbeszerzes.hu/static/uploaded/document/6%20%20mell%C3%A9klet.doc. Consult further country examples as listed in (Cingolani et al. 2015)

Table 15 An ideal complete list of variables and their publication location in support of government accountability and cartel screening[71]

Variable group	Variable	Included in the announcement		
		CFT	CA	CC
Buyer	Buyer's name	●	●	●
	Buyer's department/office	●	●	●
	Buyer's unique ID	●	●	●
	Buyer's address	●	●	●
	Buyer's type	●	●	●
Bidder / bids	Bidder's name		●	●
	Bidder's unique ID/tax ID		●	●
	Bidder's address		●	●
	Number of bids submitted		●	
	Number of bids excluded		●	
	Bid price (details on total and unit prices)		●	●
	Exact time of bid submission		●	
	Bid type (winner/loser bid)		●	
	Beneficial owners		●	●
Tender / contract	Procedure type	●	●	
	Framework agreement (1st/2nd stage)	●	●	
	Award criteria	●	●	
	Threshold (below/above EU thresholds?)	●	●	
	Estimated price (details on total or unit prices)	●	●	
	Procurement type (service, supply, work)	●	●	●
	CPV codes (% contract value per product)	●	●	●
	NUTS code(s) of contract implementation	●	●	●
	Status (cancelled, pending, etc.)	●	●	●
Dates	Call for tender publication date	●	●	●
	Bid submission deadline	●		
	Contract start and end dates	●	●	●
	Publication date of contract award		●	
	Contract signature date			
	Publication date of contract completion			●
Subcontracting	Subcontractor's name and unique ID (tax ID)		●	●
	Subcontractor's share		●	●
Consortium	Consortium members' name and unique ID (tax ID)		●	●
	Consortium members' share		●	●
Contract performance	Contract performance end date			●
	Was performance according to the contract			●
	Explanation in case of deferring from contract			●
	Information on contract modification			●
	Information on performance quality			●

Note: CFT=call for tender, CA=contract award, CC=contract completion

[71] Example call for tender announcement template: http://simap.ted.europa.eu/documents/10184/99173/EN_F02.pdf; example contract award template: http://simap.ted.europa.eu/documents/10184/99173/EN_F03.pdf.

4.2 Short-term potential of a quantitative collusion detection framework

If the above major policy changes are introduced to the Swedish public procurement system and data framework, the possibilities for quantitative collusion detection would increase substantially. However, until then the existing database should be put to a maximum use. To this end we identified the following recommendations:

1. **Set up a continuous data pipeline from public procurement data providers to assure maximum database quality and timely data availability**. Currently, data is recorded on a continuous basis; however, generating a database which is ready for statistical analysis is laborious and requires extensive human input. Based on the initial work carried out by the authors, database building and data cleaning tasks can be largely automated cutting the costs of any further analysis.

2. **Collect information on missing data points on selected high risk markets to improve data quality and deploy selected indicators to their full potential**. Missing data, unfortunately, limits the possibility for using a number of key collusion proxies. First, identifying high risk markets based on indicators less sensitive to missing information can guide where to plug in data gaps and conduct further analysis.

3. **Collect and link additional relevant data to public procurement records, in particular company registry, financial and ownership information.** This report has shown that company annual financial statements can be used to gather further evidence for suspicious company behaviour. This work could be taken further by looking at company ownership for example.

4. **Conduct risk-based checks using the identified reliable indicators for markets with sufficient data quality**. This report identified a small number of promising indicators based on the relevant literature and statistical analysis of the Swedish public procurement data. These could be put to use.

5. **Start collecting public procurement collusion investigation information systematically as a key input for a continuously improving learning indicator system**. While the initial set of cartel screens have been identified by this report, investigation information should be able to provide further evidence for tailoring them to local contexts and establishing validity.

5 References

Abrantes-Metz, Rosa M., Luke M. Froeb, John Geweke, and Christopher T. Taylor. 2006. "A Variance Screen for Collusion." *International Journal of Industrial Organization* 24(3): 467–86. http://www.sciencedirect.com/science/article/pii/S016771870500158X (May 13, 2016).

Albano, G. L., G. Spagnolo, and M. Zanza. 2009. "REGULATING JOINT BIDDING IN PUBLIC PROCUREMENT." *Journal of Competition Law and Economics* 5(2): 335–60. http://jcle.oxfordjournals.org/content/5/2/335.short (May 13, 2016).

Andreyanov, Pasha, Alec Davidson, and Vasily Korovkin. 2016. *Corruption vs Collusion : Evidence from Russian Procurement Auctions ˚*. http://www.anderson.ucla.edu/Documents/areas/fac/gem/Workshop 2016/Corruption vs Collusion.pdf.

Athey, Susan, and Kyle Bagwell. 2001. "Optimal Collusion with Private Information." *RAND Journal of Economics* 32(3). http://ssrn.com/abstract=279768 (May 13, 2016).

Athey, Susan, Kyle Bagwell, and Chris Sanchirico. 2004. "Collusion and Price Rigidity." *The Review of Economic Studies* 71(2): 317–49. http://www.jstor.org/stable/3700628.

Bajari, Patrick, and Lixin Ye. 2003. "Deciding Between Competition and Collusion." *Review of Economics and Statistics* 85(4): 971–89. http://dx.doi.org/10.1162/003465303772815871.

Barrat, Alain, Marc Barthelemy, Romualdo Pastor-Satorras, and Alessandro Vespignani. 2004. "The Architecture of Complex Weighted Networks." *Proceedings of the National Academy of Sciences of the United States of America* 101(3747).

Bobilev, Roman, Andrea Guglielmo, Elena Paltseva, and Giancarlo Spagnolo. 2015. *Public Procurement Thresholds and Data in Sweden*.

Burt, R S. 2004. "Structural Holes and Good Ideas." *American Journal of Sociology* 110: 349–99.

Byatt, Ian Charles Rayner. 2001. *Delivering Better Services for Citizens. A Review of Local Government Procurement in England*. London.

Center for Global Development. 2014. *Publishing Government Contracts. Addressing Concerns and Easing Implementation*. Washington DC.

Charron, Nicholas, Carl Dahlström, Mihály Fazekas, and Victor Lapuente. 2016. "Careers, Connections, and Corruption Risks: Investigating the Impact of Bureaucratic Meritocracy on Public Procurement Processes." *Journal of Politics* in press.

Cingolani, Luciana, Roberto Martinez B. Kukutschka, Mihály Fazekas, and Bence Tóth. 2015. *Towards a Comprehensive Mapping of Information on Public Procurement Tendering and Its Actors across Europe*. http://digiwhist.eu/wp-content/uploads/2016/01/DIGIWHIST_D1_1-AccessToTenderInfo.pdf.

Commission, European. 2012. "Final Report Review of the Functioning of the Cpv Codes / System." (December).

Conley, Timothy G, and Francesco Decarolis. 2016. "Detecting Bidders Groups in Collusive Auctions." *American Economic Journal: Microeconomics* 8(2): 1–38. http://www.aeaweb.org/articles?id=10.1257/mic.20130254.

Coviello, Decio, and Mario Mariniello. 2014. "Publicity Requirements in Public Procurement: Evidence from a Regression Discontinuity Design." *Journal of Public Economics* 109: 76–100.

CRCB. 2016. "Versenyerősség És Korrupciós Kockázatok A Magyar Közbeszerzések Statisztikai Elemzése – 2009-2015." (2016 February). http://www.crcb.eu/wp-content/uploads/2016/03/hpp_2016_crcb_report_2016_hu_160303_.pdf.

Czibik, Ágnes, Bence Tóth, and Mihály Fazekas. 2015. *How to Construct a Public Procurement Database from Administrative Records? With Examples from the Hungarian Public Procurement System of 2009-2012*. http://www.govtransparency.eu/index.php/2015/12/21/constructing-public-procurement-databases-from-administrative-records-technical-report/.

"Detecting Bid Rigging in Public Procurement." http://www.oecd.org/competition/cartels/42594486.pdf (May 13, 2016).

Estache, Antonio, and Atsushi Iimi. 2008. "Joint Bidding in Infrastructure Procurement." http://papers.ssrn.com/abstract=1232760 (May 13, 2016).

Fazekas, Mihály, and Bence Tóth. 2017. "Infrastructure for Whom? Corruption Risks in Infrastructure Provision across Europe." In *Governance Report 2016. Governance of Infrastructure*, eds. Gerhard Hammerschmid, Genia Kostka, and Kai Wegrich. Oxford, UK: Oxford University Press.

Harrington, Joseph. 2008. "Detecting Cartels." In *Handbook of Antitrust Economics*, MIT Press, 796.

Heimler, Alberto. 2012. "Cartels in Public Procurement." *Journal of Competition Law and Economics* 8(4): 849–62.

Hungarian Competition Authority. 2004. *Decision Vj-162/2004/125*.

———. 2007. *Decision Vj-174/2007/332*.

———. 2008. *Decision Vj-134/2008/275*.

Ishii, Rieko. 2009. "Favor Exchange in Collusion: Empirical Study of Repeated Procurement Auctions in Japan." *International Journal of Industrial Organization* 27(2): 137–44. http://www.sciencedirect.com/science/article/pii/S0167718708000684 (April 25, 2016).

Klemperer, Paul. 2007. "Bidding Markets." *Journal of Competition Law and Economics* 3(1): 1–47.

Levenstein, Margaret C., and Valerie Y. Suslow. 2006. "What Determines Cartel Success?" *Journal of Economic Literature* 44(1): 43–95.

Lewis-Faupel, Sean, Yusuf Neggers, Benjamin a. Olken, and Rohini Pande. 2014. "Can Electronic Procurement Improve Infrastructure Provision? Evidence from Public Works in India and Indonesia."

Mena-Labarthe, Carlos. 2012. "Mexican Experience in Screens for Bid-Rigging." *Antitrust Chronicle* 3(1). https://works.bepress.com/carlos_mena-labarthe/6/ (May 13, 2016).

OECD. 2014. *Detecting Bid Rigging in Public Procurement*. http://www.oecd.org/competition/cartels/42594486.pdf (May 13, 2016).

OECD/Sigma. 2014. *The Principles of Public Administration*. Paris: OECD/Sigma.

Oxera. 2013. "Hide and Seek: The Effective Use of Cartel Screens." http://www.oxera.com/getmedia/210bc5bc-0cc9-40ea-8bc9-6c8b2406b485/Cartelscreens.pdf.aspx?ext=.pdf (August 29, 2016).

Padhi, Sidhartha S., and Pratap KJ Mohapatra. 2011. "Detection of Collusion in Government Procurement Auctions." *Journal of Purchasing and Supply Management* 17(4): 207–21.

Pesendorfer, Martin. 2000. "A Study of Collusion in First-Price Auctions." *Review of Economic Studies* 67(3): 381–411. http://restud.oxfordjournals.org/content/67/3/381.short (May 13, 2016).

Piga, Gustavo. 2011. "A Fighting Chance against Corruption in Public Procurement." In *International Handbook on the Economics of Corruption. Volume Two*, eds. Susan Rose-Ackerman and Tina Soreide. Cheltenham: UK: Edward Elgar, 141–84.

Riccardi, Michele, and Ernesto U Savona. 2013. *The Identification of Beneficial Owners in the Fight against Money Laundering—final Report of Project BOWNET*.

SCA. 2015a. *Facts and Figures on Public Procurement in Sweden*.

———. 2015b. *Honesty Pays: How to Detect and Prevent Bid-Rigging Cartels*.

Stigler, George J. 1964. "A Theory of Oligopoly." *Journal of Political Economy* 72(1): 44–61. http://dx.doi.org/10.1086/258853.

Thai, Khi V. 2009. "International Public Procurement: Concepts and Practices." In *International Handbook of Public Procurement*, ed. Khi V. Thai. London: CRC Press, 2–22.

Tóth, Bence, Mihály Fazekas, Ágnes Czibik, and István János Tóth. 2014. *Toolkit for Detecting Collusive Bidding in Public Procurement. With Examples from Hungary*. Budapest: Corruption Research Center Budapest.

Transparency International. 2006. *Handbook for Curbing Corruption in Public Procurement*. Berlin: Transparency International.

Varela-Irimia, Xosé-Luís. 2014. *Public Procurement Indicators 2012*. Brussel.

Wasserman, Stanley. 1994. *Social Network Analysis: Methods and Applications*. Cambridge, UK: Cambridge University Press.

de Willebois, Emile van der Does et al. 2011. *The Puppet Masters: How the Corrupt Use Legal Structures to Hide Stolen Assets and What to Do About It*. World Bank Publications.

World Bank. 2011. *Curbing Fraud, Corruption, and Collusion in the Roads Sector*. Washington DC.

6 Appendix A: Variable storage standard

Supplementing the variable list discussed in section 1.1, a very brief summary is outlined here discussing the main technical details of the most efficient data storage methods for public procurement data. Optimally, the database structure follows 'linked open data principles'[72], which is a useful way of storing data, for running diverse queries and linking different data sources. Both the OCDS standard and DIGIWHIST data schemes follow this structure. The main idea of linked datasets is that variables are organized into different variable groups or 'objects' that are linked by unique identifiers.

Unlike the thematic variable groups discussed in section 1.1, these objects are primarily organized according to the type/function of data they are storing. Therefore, certain objects are more or less the same as the thematic groups introduced in section 1.1, e.g. 'tender' object contains procedure type, region of performance etc. However, certain variables are grouped into new objects. For example, all price related variables – such as the originally estimated price, bid prices, final price, payments, documentation price etc. – are stored in a different 'price' object, that contain net amount, currency, VAT, and minimum and maximum values in case of intervals. Another example is the storage of organizational variables. Numerous organizations are involved in each tendering process: contracting authorities, bidders (including the winner), an appeal body, specification creator, administrator etc. As each has a name, ID, contact point, e-mail etc., these variables are contained in a 'body' object. Furthermore, within the body object, both the 'ID' (ID, type and scope of the ID can be multiple by organization) and 'address' (raw address, postal code, country, NUTS codes etc.) are stored in separate objects.

Figure 40 depicts an exemplary database structure where objects are nested within each other. The database schema follows an object-level hierarchy for the 'main' objects: Tender -> Lot -> Bid -> Bidder. Consequently, there is a clear distinction between the tender and lot(s) levels; hence data on different lots and the resulting contracts are stored separately.[73] Whereas 'main' objects can have multiple nested objects, for example one tender can include multiple lots and lots can have multiple bids. It is important to see, that the main structural frame of the database connects 'lots' with individual 'bids'[74] (or offers), as bids are unique for each lot[75], while bidders (i.e. individual companies) are connected to their bid(s).

[72] https://en.wikipedia.org/wiki/Linked_data

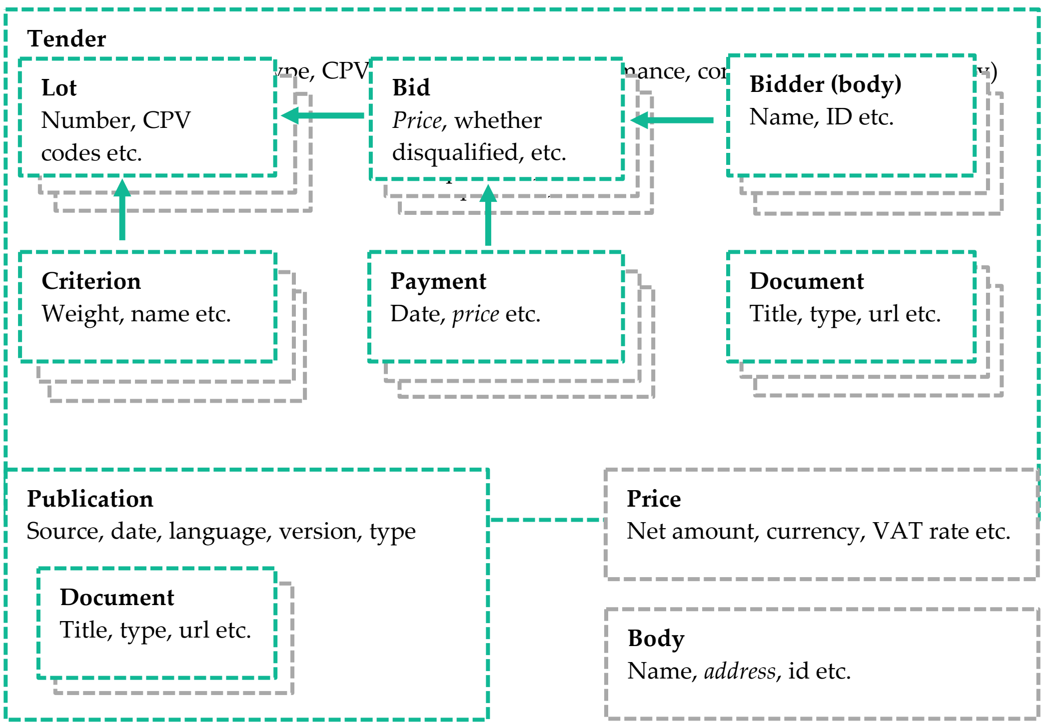
[73] Note, that the variables included in the 'lot' object are more or less the same as the ones under the 'tender' object. For examples, while procedure type tends to be the same in case of multiple lot tenders, it is usually the same for all lots, CPV codes are often different, hence given for each lot separately.

[74] 'Bid' object contains details on whether the bid is winning, disqualified etc., and stores the bid value in a connected 'price' object.

In addition to these main structural objects, others such as 'body', 'ID', 'address', 'prices', 'award criteria' or 'publication' are nested in each other (see some examples in the previous paragraph). For example tender A can have A1 and A2 lots, whereas A1 lot can have a1 and a2 bids, that are submitted by ac1 and ac2 companies. All these 'objects' are connected with a unique internal ID, and further 'sub-objects' are connected to them. For example company ac1 has an 'address' and a 'bid' object, bid a1 has a 'price' object, and lot A1 can have a separate 'award criteria' object.

In Figure 40, the turquoise boxes correspond to individual objects, their position within each other depicts the nested relationships. Grey boxes indicate that there might be multiple objects of the same type related to the same superior object, finally arrows show further 'nested' relations.

Figure 40 Schematic graph of data storage



[75] The reason is that one bid is unique by lots, while one company (or a consortia) can submit multiple bids if variants are accepted.

7 Appendix B: Variable availability

Table 16 Assessment of data coverage based on national public procurement announcement templates

Countries	Contract related items	Dates	Requirements	Documentation	Funding	Buyer	Bidder	Bid	Price	Cancellation /correction	Other
Austria	100%	80%	100%	100%	50%	64%	23%	20%	17%	100%	14%
Belgium	88%	80%	100%	50%	33%	64%	31%	40%	67%	67%	29%
Bulgaria	88%	70%	100%	75%	50%	64%	62%	20%	50%	67%	43%
Cyprus	81%	80%	100%	50%	33%	55%	15%	20%	33%	67%	14%
Croatia	69%	20%	0%	0%	0%	45%	0%	0%	0%	0%	0%
Czech Rep.	88%	80%	100%	50%	67%	45%	38%	40%	67%	67%	14%
Germany	88%	40%	100%	100%	50%	64%	23%	0%	17%	0%	14%
Denmark	63%	60%	0%	50%	17%	45%	0%	0%	33%	33%	14%
Estonia	81%	60%	100%	100%	50%	64%	46%	20%	33%	67%	14%
Spain	81%	80%	100%	50%	17%	55%	23%	20%	67%	100%	14%
Finland	56%	30%	0%	25%	0%	55%	0%	0%	0%	67%	0%
France	88%	50%	100%	0%	33%	45%	46%	40%	17%	67%	43%
Greece	75%	80%	100%	75%	67%	45%	15%	0%	33%	100%	14%
Hungary	94%	90%	100%	75%	33%	64%	62%	40%	100%	100%	86%
Ireland	31%	40%	0%	50%	0%	27%	15%	0%	0%	0%	0%
Italy	50%	40%	0%	0%	0%	18%	15%	0%	17%	67%	0%
Lithuania	88%	80%	100%	75%	50%	64%	54%	40%	67%	100%	29%
Luxembourg	81%	40%	100%	50%	33%	45%	31%	20%	33%	0%	29%

Countries	Contract related items	Dates	Requirements	Documentation	Funding	Buyer	Bidder	Bid	Price	Cancellation /correction	Other
Latvia	63%	80%	100%	75%	50%	64%	54%	20%	67%	100%	29%
Malta	56%	50%	33%	100%	17%	45%	38%	20%	67%	67%	14%
Netherlands	100%	80%	100%	75%	33%	55%	38%	40%	33%	67%	29%
Poland	75%	40%	100%	25%	33%	36%	23%	40%	67%	33%	0%
Portugal	94%	70%	100%	75%	50%	55%	31%	20%	50%	67%	43%
Romania	75%	70%	100%	75%	50%	55%	38%	20%	33%	33%	14%
Sweden	**69%**	**40%**	**0%**	**25%**	**0%**	**45%**	**31%**	**20%**	**67%**	**33%**	**71%**
Slovenia	88%	70%	100%	0%	33%	64%	31%	40%	33%	100%	0%
Slovakia	100%	80%	100%	50%	67%	45%	38%	20%	33%	67%	0%
United Kingdom	38%	50%	0%	0%	0%	18%	15%	0%	33%	0%	0%
Norway	88%	70%	100%	100%	33%	64%	23%	20%	67%	67%	14%
Switzerland	69%	70%	0%	25%	17%	36%	15%	0%	0%	100%	0%
Iceland	13%	20%	0%	100%	0%	27%	15%	20%	67%	67%	14%
Serbia	69%	50%	0%	50%	17%	55%	31%	20%	50%	67%	0%
Georgia	31%	70%	0%	100%	0%	36%	54%	40%	83%	33%	0%
Armenia	31%	30%	100%	75%	0%	9%	0%	0%	0%	0%	29%
EC	88%	80%	100%	75%	50%	45%	46%	40%	67%	67%	43%
Average	72%	60%	67%	57%	30%	48%	29%	20%	42%	58%	18%

For detailed, variable level mapping see Cingolani et al. (2015), where variable availability within each variable group is contained in Appendix A.
Note, that for SE, the mapping is based on the i) structured database provided by Visma, and ii) a random sample from Visma's public procurement platform. It is possible, that multiple templates are used, that we could not assess in detail. Therefore, this assessment should be only regarded as a good estimation of actual data coverage.

8 Appendix C: An example announcement

Table 17 Random sample of a published call for tender document

Title:	Laboratorietjänster 2017 (Ks 385/2016)
Document number:	000000
Final tender date:	15/08/2016
Text:	
Avsnitt I: Upphandlande myndighet	
I.1	Namn, adresser och kontaktpunkt(er)
Örebro kommun Att: Göran Gunnarsson Box 30070 SE-701 35 Örebro Sverige E-post: goran.gunnarsson@orebro.se	
I.1.2	Ytterligare upplysningar kan erhållas från
TendSign Sverige Internetadress https://tendsign.com/doc.aspx?ID=102677	
I.1.3	Förfrågningsunderlag och kompletterande handlingar (inklusive dokument för en konkurrenspräglad dialog eller ett dynamiskt inköpssystem) kan erhållas från
TendSign Sverige Internetadress https://tendsign.com/doc.aspx?ID=102677&Goto=Docs	
I.1.4	Anbud eller anbudsansökningar skall skickas till
TendSign Sverige Internetadress https://tendsign.com/doc.aspx?ID=102677&Goto=Tender	
I.2	Typ av upphandlande myndighet
Regional eller lokal myndighet	
I.3	Huvudsakliga verksamheter
Allmänna offentliga tjänster	
Avsnitt II: Upphandlingens föremål	
II.1	Beskrivning
II.1.1	Den upphandlande myndighetens benämning/rubrik på upphandlingen
Laboratorietjänster 2017	
II.1.2	Typ av upphandling och plats för byggentreprenad, leverans eller utförande
Tjänster	
II.1.2.7	Tjänstekategori
12A. Arkitekttjänster	
II.1.2.8	Huvudplats för utförandet
Plats: Örebro NUTS-kod: SE124, Örebro län	
II.1.3	Meddelandet gäller
Upprättande av ett ramavtal	
II.1.5	Kort beskrivning av upphandlingen/upphandlingarna
Denna upphandling avser ramavtal laboratorietjänster. Kommunerna/kommunalförbunden har inom sin verksamhet behov av kemiska och bakteriella laboratorietjänster avseende vatten, slam, livsmedel m.m. Uppdragets omfattning och utformning kan variera mellan kommunerna.	
II.1.6	Gemensam terminologi vid offentlig upphandling (CPV-referensnummer)
Huvudobjekt: 71900000 Laboratorietjänster Tilläggsobjekt: 71610000 Provning och analys av sammansättning och renhet	
II.1.8	Är kontraktet uppdelat i flera delar?
Nej	

II.3	Kontraktets löptid eller tidsfrist för slutförande
Datum för påbörjande	
2017-01-01 (åååå-mm-dd)	
Datum för slutförande	
2019-12-31 (åååå-mm-dd)	
Avsnitt III: Juridisk, ekonomisk, finansiell och teknisk information	
Avsnitt IV: Förfarande	
IV.1	Typ av förfarande
IV.1.1	Typ av förfarande
Öppet	
IV.2	Tilldelningskriterier
IV.2.2	En elektronisk auktion kommer att användas
Nej	
IV.3	Administrativ information
IV.3.1	Den upphandlande myndighetens referensnummer på ärendet
Ks 385/2016	
IV.3.4	Sista datum för mottagande av anbud eller anbudsansökningar
2016-08-15 (åååå-mm-dd)	
IV.3.6	Språk som får användas i anbud eller anbudsansökningar
Svenska	
IV.3.7	Minimiperiod under vilken anbudsgivaren är bunden av sitt anbud
T.o.m: 2016-12-31 (åååå-mm-dd)	
IV.3.8	Anbudsöppning
2016-08-17 00:00 (åååå-mm-dd)	
Avsnitt VI: Kompletterande upplysningar	
VI.3	Kompletterande information
Visma TendSign-annons: http://www.opic.com/notice.asp?req=dicvxlbb	
VI.5	Datum för införande av denna annons
2016-05-12 (åååå-mm-dd)	

This was the sample used for assessing additional variable availability (Table 16) on top of the information already included in the structured database from Visma Opic.

9 Appendix D: Variable list

Table 18 Variable list

Variable name	Included in the analysis?	Missing ratio
Database ID	Yes	0%
Call for tender publication year	Yes	0%
Bidding deadline year	Yes	1%
Status	Yes	0%
Contracting authority's ID	Yes	0%
Contracting authority's address	Yes	50%
Contracting authority's postal code	Yes	14%
Contracting authority's town	Yes	9%
Contracting authority's type	Yes	0%
Whether a framework agreement	Yes	0%
Procedure type	Yes	0%
Threshold level	Yes	0%
Number of tenderers	Yes	39%
Evaluation criteria	Yes	19%
CPV codes	Yes	0%
Region (NUTS)	Yes	0%
Contract start and end dates	Yes	48%
Bidders/Winners name	Yes	34%
Bidders/Winners ID	Yes	40%
Bid price	Yes	79%
Bid price unit	Yes	84%
Bid type (winner, participated, applied)	Yes	34%
Level of date verification	Yes	0%
Call for tender publication month	No	0%
Contracting authority's name	No	0%
Contracting authority's department	No	52%
Document type	No	0%
CPV name	No	0%
Time period of possible contract extension	No	62%
Order of bids	No	98%
Court case ID	No	0%
Basis of the case	No	100%
Plaintiff	No	0%
Procurement officer involved in the court case	No	0%
Court name	No	0%
Type of court ruling document	No	0%
Court decision	No	0%

10 Appendix F: Location of public procurement contracts

Table 19 Distribution of companies according to the number of unique NUTS 1 and NUTS2 regions[76]

Number of bids per unique NUTS 2 region	Number of bids per unique NUTS1 region: 1	2	3	4	Total
1	27183	0	0	0	27,183
2	2369	3218	0	0	5,587
3	265	1549	452	0	2,266
4	17	567	630	0	1,214
5	0	181	590	0	771
6	0	12	448	0	460
7	0	3	370	0	373
8	0	0	294	0	294
9	0	0	53	1	54
10	0	0	9	0	9
11	0	0	1	0	1
Total	29,834	5,530	2,847	1	38,212

Table 20 Number of tenders with and without company ID

Country	Number of unique tenders	Number of unique tenders with company ID	Ratio of unique tenders with company ID
DK	9	4	44%
FI	44	10	23%
MT	1	0	0%
SE	1433	861	60%
SE1	168	116	69%
SE110	20431	12145	59%
SE12	292	210	72%
SE121	4082	2820	69%
SE122	2935	2118	72%
SE123	5338	3327	62%
SE124	2322	1571	68%
SE125	2585	1667	64%
SE2	353	243	69%
SE21	131	92	70%
SE211	4148	2959	71%
SE212	2374	1558	66%
SE213	2777	1725	62%
SE214	717	428	60%

[76] Excluding companies bidding on tenders with a national scope.

SE22	6730	4086	61%
SE221	2031	1329	65%
SE224	12265	8108	66%
SE23	343	227	66%
SE231	3199	2164	68%
SE232	17619	11170	63%
SE3	248	173	70%
SE31	46	27	59%
SE311	2999	2147	72%
SE312	3455	2435	70%
SE313	2915	1813	62%
SE32	93	60	65%
SE321	2829	1815	64%
SE322	1802	1240	69%
SE33	111	82	74%
SE331	4707	2760	59%
SE332	4786	2987	62%
Sum	116318	74477	64%

KONKURRENSVERKET
Swedish Competition Authority

Adress 103 85 Stockholm
Telefon 08-700 16 00
Fax 08-24 55 43
konkurrensverket@kkv.se

www.konkurrensverket.se