

Unclassified

English - Or. English

24 May 2023

**DIRECTORATE FOR FINANCIAL AND ENTERPRISE AFFAIRS
COMPETITION COMMITTEE**

Working Party No. 3 on Co-operation and Enforcement

Summary of Discussion of the Roundtable on Data Screening Tools for Competition Investigations

Annex to the Summary Record of the 136th meeting of Working Party 3

28 November 2022

This document prepared by the OECD Secretariat is a detailed summary of the Roundtable on Data Screening Tools for Competition Investigations, held by Working Party 3 on 28 November 2022.

More documents related to this discussion can be found at
<https://www.oecd.org/daf/competition/data-screening-tools-for-competition-investigations.htm>

Please contact Mr Antonio CAPOBIANCO if you have questions about this document.
Email: Antonio.CAPOBIANCO@oecd.org

JT03519569

Summary of Discussion of the Roundtable on Data Screening Tools for Competition Investigations

1. Introduction

On 28 November 2022, Working Party No. 3 held a discussion on data screening tools for competition investigations chaired by Jonathan Kanter. This was an agency-only session. There were no external speakers.

The Chair noted that this would be a topical discussion, particularly looking forward at how new technologies are likely to cause a changing landscape for competition authorities. There were 17 written contributions for this session. The Chair asked the Secretariat to briefly summarise the finding of its background paper.

The Secretariat explained that the background paper focused on the use of screening tools to detect cartels, as this is the most common use of data screening tools by competition authorities. Thus, in this context, data screening tools are empirical methods to identify price or bid patterns to detect collusion. The Secretariat noted that researching the practical use of these tools by competition authorities can be difficult as authorities often choose to keep much of their activity, at least the specific details, confidential.

There has been increasing interest in, and seemingly use of, cartel screening tools, both by competition authorities and now also private companies. There are three main reasons why this will likely continue to increase: (1) increasing quality and quantity of data, particularly procurement data; (2) a proliferation of new methods in the academic literature, with some machine learning methods leading to increasing accuracy; and (3) a decline in leniency applications suggesting the need for authorities to engage more with proactive cartel detection methods.

There is not a ‘one size fits all’ approach to cartel screening. Cartel screens can broadly be classified as structural and behavioural, with most of the recent academic literature focusing on behavioural screens. A broad classification of behavioural screens is: (1) collusive markers; (2) structural breaks; and (3) anomalies. A competition authority looking to start a cartel screening programme with relatively limited resources could use a mix of collusive markers and structural breaks applied to procurement markets.

Despite progress developing procurement databases, data continues to be a major stumbling block for systematic cartel screening. Several jurisdictions are still in the process of collecting data from procurement agencies. It can require considerable work to aggregate and clean the data. It can be useful to provide procurement agencies with an example template of the data necessary to perform cartel screening. Furthermore, it is also important to have the knowledge and skills to perform cartel screening. Many competition authorities are now hiring data scientists and technologists to work on their cartel screening efforts. These profiles can complement the existing skills of economists and lawyers. Ultimately, cartel screening results are indirect evidence of collusion. The results are typically used to open an investigation or to prioritise cases. Direct evidence is normally required to close a case.

The Chair thanked the Secretariat for the informative summary. The Chair emphasised the importance of the discussion, particularly in the light of declining leniency applications. The ways in which companies engage in collusion is changing. Bids are now often electronic and there is more scope for the digitisation and standardisation of procurement

data. These combined issues mean that competition authorities will face dramatic changes in the coming years, both in terms of new challenges as well as improving capabilities. With this in mind, the chair asked Spain to open the discussion on data.

2. Data

Spain (CNMC) highlighted the creation of its procurement database. The database has existed since 2012, although it contained some typographical errors and incomplete information. It was also missing information on non-winning bids. Since 2014 the CNMC has been working to make it a clean and complete database. For example, they developed an automated way to clean the data, such as standardising company names. They have supplemented the centralised database with additional information from regional and local sources. They scraped regional procurement data platforms for additional information. They also developed an algorithm to download and read all attached documents (such as PDF, Word, Excel, etc.) and recognise if there is information on losing bids in the document, and if so, convert that information into a structured dataset.

The CNMC now has the most complete procurement dataset in Spain, as it combines all central, regional and local data. It contains mostly clean and complete data for almost 33 million contracts. Spain is now beginning the process of developing cartel screens using supervised and unsupervised machine-learning. Responding to a question from the Chair, **Spain** clarified that the database is frequently updated, downloading the latest data from the centralised procurement platform twice per week, with the new data automatically cleaned and harmonised.

Switzerland relies on publicly available data for its cartel screening tools, for two main reasons: (1) they do not want to alert companies under investigation, in case they destroy any direct evidence; and (2) they want to use common cartel screening methods, thus it is best to have a common database on which to apply these methods. Sometimes bid information is openly available on the internet, and if not, COMCO can ask the procurement agency for the data directly. In the past, the Swiss Cartel Act has been the basis for information requests for data from procurement agencies. However, the recent revision of the public procurement law included a specific provision that allows them to ask for bidding data to use in investigations. The drawback of public data is that it typically only contains limited information, including the identity of the bidders, submitted bids, location of bidders and a short description of the project. IT typically does not contain information on engineer estimates of the cost of the project or details of any subcontractors. Therefore, there are some limits to the possible cartel screening methods that COMCO can apply. Switzerland is also training supervised machine-learning models using data from past cartel investigations.

Colombia has developed two digital screening tools: Sherlock and Inspector. The Sherlock tool aims to detect collusion in procurement data. It has two main steps. First, it converts public information into a clean and structured database that investigators can use. Second, it automatically reviews the data to identify any patterns that may indicate anticompetitive conduct. The tool generates alerts based on red flags. The Inspector tool identifies new regulatory projects based on a review of regulator web pages. This tool allows the SIC Competition Advocacy Working Group to review these new regulatory projects.

The SIC developed these tools to harness increasingly available data. In Colombia, access to public information is a constitutional right. It has resulted in several public data portals. Colombia has been investing in improving the quality of public procurement data in a

platform called SECOP. This contains data from 1100 entities. The platform consolidates data from several sources into a clean and structured format.

Portugal uses webscraping to collect new data. It does not need webscraping to collect procurement data, as it already has access to a centralised government procurement database. E-procurement has been mandatory in Portugal for over a decade now. Rather, webscraping can gather additional relevant price information that is helping to detect anti-competitive behaviour in digital markets, notably algorithmic-led collusion and resale price maintenance.

The digital task force performs this webscraping. It allows them to rapidly assess the relevance of complaints. For example, if they receive a complaint alleging resale price maintenance in e-commerce, they determine a first assessment within hours. Web-scraping reduces the amount of time that it takes to collect the relevant data, allowing them to allocate resources more efficiently.

Webscraping involves extracting pricing data in an automated fashion from online stores, marketplaces or price comparison websites and converting it into readable formats. The team currently uses the Python programming language. They select the product categories or brands to be scraped based on previous intelligence (this can be a complaint, investigation in another jurisdiction, or some other source).

The webscraping itself may take a few hours, as they do not wish to strain the websites from which they are collecting data. They can usually collect information on about 500 products per hour. The webscraping tends to be a one-off when it is based on a tip off, although they may scrape a few more times, for example, one or two weeks later to check that the results are persistent.

After collecting the pricing data, they apply a screening technique to confirm whether the results are consistent with an antitrust theory of harm. This is also done automatically and in a matter of seconds. The code provides a list of the top products and their associated price alignment. The products are ranked according to a score by using criteria such as the percentage of online retailers or sellers whose price is aligned, or the difference between the alignment price and the lowest price for a specific product in the market, or the duration of the price alignment, etc.

Thus, if they find enough price alignment, they may deem a complaint to have high potential and deserve further analysis. So far, by using this tool, they have obtained a search warrant and successfully concluded an investigation on resale price maintenance in the pharmaceutical sector. They also have other ongoing investigations in e-commerce. This is just an example of the proactive methods of detection that they have been using throughout the years, this year they received a record number of leniency applications, so successful proactive approaches appear to feed more reactive methods of detection.

Responding to a query from the Chair, **Portugal** clarified that they have not yet found evidence of algorithmic-led collusion that warranted further investigation. So far, there has only been evidence of resale price maintenance that allowed them to open and successfully close cases.

3. Screening tools, methods and techniques

Korea has a system for detecting bid-rigging in public procurement called BRIAS (Bid-Rigging Indicator Analysis System). The BRIAS tool improved following amendments in 2015 as part of the Advancement Project.

Prior to improvements in 2015, BRIAS had too many positive cases (around 1500 per year) and thus it was under-utilised by investigators. It only resulted in 3 successful cases during this period. Since the improvements, the number of flagged cases reduced to 297 cases per year, allowing greater focus and efficacy. The KFTC opened 26 ex-officio bid-rigging cases between 2018 and 2021, with 9 resulting in sanctions and 4 closed with the issuance of a warning notice.

There were several changes to BRIAS due to the Advancement Project. First, the number of procurement contracts included in the database increased due to a lowering of the minimum monetary threshold, as the previous minimum amounts were so high that BRIAS had blind spots in its monitoring activity. Second, the reliability of the screening output increased. The KFTC examined its previous bid-rigging decisions for the last five years to tailor the parameters it uses in BRIAS. Finally, it is now possible to group bids of the same product or business entity. This allows investigators to access more condensed information. Currently, BRIAS compiles data from 16 government agencies.

BRIAS still faces some challenges, however, the KFTC will continue to improve BRIAS as it learns from the experience of using it.

Singapore has two tools: (1) the bid-rigging detection tool ('BRDT'); and (2) the document similarity tool ('DST'). Upon request from the Chair, Singapore focused on the DST.

The DST provides a score indicating the degree of similarity between pairs of documents. The tool determines the score using natural language processing algorithms. The inputs to the model are documents in any format (e.g., PDF, WhatsApp chat, or even a paper document that has been scanned and converted to text using optical character recognition technology). The DST creates a matrix of all document pairs and provides a document similarity score for each pair. Thus, the enforcer can prioritise the manual review of documents with a high document similarity score.

The DST was developed semi-in-house. Within CCCS, there is a data management and analytics unit, where data scientists develop tools and algorithms. However, to develop the DST, CCCS collaborated with GovTech, which is a centralised technology agency, which pools together a group of technologists and data scientists for the benefit of all government agencies in Singapore.

The DST is complementary to the BRDT. In a first step, CCCS can screen all bidding data for suspicious bids using the BRDT. Then, in a second step, CCCS can check the similarity of documents for those suspicious bids using the DST.

CCCS has successfully applied the DST in cases, but it is not possible to disclose any further details. Infringement decisions do not explicitly mention the use of the DST.

CCCS is considering future enhancements to the DST. Currently, all elements in a document are treated as text. In the future, the DST will better be able to process numbers. For example, 199 and 200 are currently considered as completely different text, but these will eventually be recognised as numbers.

Finally, CCCS welcome international collaboration, particularly the prospect of sharing code. Since the roundtable discussion was public, they didn't go into much depth when discussing the functioning of the tools but are happy to share more details through bilateral conference calls with different agencies. Several calls have already taken place.

Croatia considered these useful tools and welcomed the offer of international collaboration. Croatia have completed a first step towards screening, which is a bilateral agreement between the competition authority and the Ministry of Economy, such that the competition authority has access to the electronic public procurement database. Croatia is

now taking a second step, which is developing an IT tool to detect bid-rigging in public procurement markets. Croatia is already receiving assistance from Spain. Croatia was happy to hear that other competition authorities may also be able to help.

Ecuador described its ‘a priori’ algorithm which identifies signs of collusion in public procurement data. The algorithm was developed following a market study of the Ecuador national public procurement system, which identified that the government lacked a tool to detect collusion.

The algorithm identifies sets of frequent elements. This has been used previously in the context of supermarket purchases, identifying products that are more likely to be purchased together. For example, establishing the probability that a customer buys product Y if they buy product X. In the context of cartel screening, the algorithm is used to identify frequent combinations of procurement participants and winners. The algorithm can uncover anti-competitive market sharing practices.

The algorithm was applied to two public procurement databases. First, the total procurement database for the period 2015 to 2020. Second, to a specific database that related to a cartel investigation that resulted in a sanction. When applied to the second database, the algorithm correctly identified some suspicious behaviour. The tools will now be applied to specific subsets of data. The results will be considered in the context of knowledge of the sector in case they can be explained by industry-specific market structure or conditions.

4. Publicising cartel screens

The Chair asked whether any jurisdictions had considered the extent to which cartel screening tools should be publicised, and how transparent authorities should be. On one hand, the more firms are aware of such tools, the more they may be deterred from engaging in new anticompetitive conduct, and they may also be more inclined to engage with leniency programmes regarding any existing anticompetitive conduct. On the other hand, if authorities publicise too much, it can create opportunities for evasion and companies to develop anti-competitive behaviour that can dupe the screening tools.

The Chair considered that authorities will benefit from companies understanding that authorities are innovating to keep pace with market realities and deploying new expertise and tools, even if this falls short of disclosing the actual algorithms that authorities use.

Australia was reluctant to publicise its tool until it is more developed. They have been investing heavily in the software and collaborating with other agencies, both bilaterally and in the context of the ICN (notably, the ICN Cartel workshop). Australia is tackling issues regarding access to procurement data and the quality of that data. The ACCC has struggled to gather good quality data on which to apply cartel screening tools. The ACCC doesn’t want to undermine its credibility by claiming it can do more than it can.

The ACCC has had more success with two other proactive techniques. First, an anonymous whistle-blower tool, which has resulted in some ongoing investigations. The tool allows the ACCC to continue engaging with the whistle-blower on an anonymous basis throughout the investigation. They found this was a real advantage compared to receiving a single piece of information and not being able to follow-up.

Second, post-merger monitoring and auditing, which led to the detection and successful prosecution of a cartel. Quite often mergers in Australia are allowed to proceed subject to certain conditions. One of the conditions has been that there be an independent monitor to prepare a report on the implementation of the conditions related to the merger. The monitor

includes conduct that they're concerned about in the report. In one case the report detailed conduct which ultimately turned out to be cartel conduct. The ACCC investigated this conduct and found significant cartel conduct, which they prosecuted, obtaining guilty pleas.

5. Resources and staffing

Mexico (COFECE) has a unit focused on detecting potential anti-competitive conduct and other competition issues in the Mexican market. The unit uses several tools and methods, including digital screening tools.

The unit has staff with different experience and educational backgrounds. For example, currently the unit has statisticians, mathematicians, as well as computer and data scientists, and these experts work with lawyers and economists to assess the extent to which a finding indicates potential anti-competitive conduct.

The growing use of large datasets and artificial intelligence by private companies creates the need for competition authorities to have staff with strong technological expertise. However, it can be difficult to hire or develop these staff internally, given the relatively limited competition authority budgets, particularly when compared the salaries offered by private companies. Well-resourced companies can often attract and retain specialised staff, as well as poach staff that have been internally developed at a competition authority.

The authority cannot compete with the salaries and benefits offered by private companies. Rather, the authority hopes to retain its specialised staff through attractive non-monetary incentives such as training on the latest digital screening techniques and tools.

6. Successful use of screens in enforcement

Italy provided a couple of examples of successful cases resulting from the use of cartel screening methods.

As in many other jurisdictions, Italy tried to use data screening tools in a more general and systematic way. However, they encountered the usual data issues described by the Secretariat. The authority instead focused on a smaller number of tenders, in cooperation with CONSIP (the Italian central procurement agency acquiring goods and services for the public administration).

The competition authority advised on the design of several tenders before they were launched. The competition authority could then obtain data on these specific tenders. They performed simple statistical tests on these tenders. It was not sophisticated analysis. However, this served as a screen to prioritise cases for further investigation. In two cases they identified suspicious patterns.

One case concerned school cleaning services. The authority noticed that the firms never overlapped in their bids. The screen was just the initial step, which was followed by dawn raids, from which the authority obtained direct evidence of a collusive agreement.

Another case concerned consultancy services on the use of European structural funds. Again, one of the suspicious flags was that the offers didn't overlap. They also found that the firms offered discounts that seemed to be too low. And again, following these red flags, the authority launched an investigation and found direct evidence of a collusive agreement.

These are not examples of systematic or sophisticated cartel screening tools but show that authorities can be successful even using more simple methods.

The Chair closed the session thanking all the participants for their excellent contributions and noting that this was still only the start of an ongoing discussion about the use of data screening tools. There are already so many different innovative approaches, in what is still a relatively nascent market. The OECD is a useful forum for authorities to share best practice and details about the state-of-the-art tools that are deploying in their respective jurisdictions around the world.