Organisation for Economic Co-operation and Development

DAF/COMP/WP3/WD(2022)36

Unclassified

English - Or. English

21 November 2022

DIRECTORATE FOR FINANCIAL AND ENTERPRISE AFFAIRS
COMPETITION COMMITTEE

Cancels & replaces the same document of 8 November 2022

Working Party No. 3 on Co-operation and Enforcement

Data Screening Tools for Competition Investigations – Note by Brazil

28 November 2022

This document reproduces a written contribution from Brazil submitted for Item 3 of the 136th OECD Working Party 3 meeting on 28 November 2022.

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

Ms Despina PACHNOU
[Email: Despina.PACHNOU@oecd.org]

JT03508096

Brazil[1]

1. Project Cérebro: original design and primary goals

1. Launched in 2014, Project Cérebro was designed to be used by CADE's internal team in developing new cartel investigation tools[2].

2. The broad scope of the project (that is, developing, testing, and implementing data mining and screening tools) demanded overcoming the following challenges:

- hiring and training civil servants with expertise in the required fields of knowledge (antitrust, IT, statistics, and data science);
- accessing databases with the quality and granularity necessary to test and implement the tools;
- creating (and maintaining) a favourable environment for innovation where decision-makers could test, fail and correct[3].

3. Therefore, we created a team led by a civil servant with data science qualifications and two consultants from the United Nations Development Programme[4].

4. Once formed, the team mapped the databases necessary for implementing the investigative tools. The overall strategy aimed to collect as much data as possible via institutional partnerships[5] or web scraping[6]. The high ambition of such a strategy—i.e. to have a unified repository of public procurement from across the country to find patterns in competitors' interactions—delayed the beginning of the primary scope of the project. In short, the lean team and the time spent in obtaining data (including data processing, cleaning, and documentation) postponed the development and testing of data mining and screening techniques.

5. Thus, the team decided to define a reference database to enable start testing the techniques developed so far to assess their accuracy: federal government tenders that had the best documentation and most granular data, regardless of the poor quality of some key

[1] This paper was written by Felipe Leitão Valadares Roquete. It was translated from Portuguese into English by Arianne Mesquita, Ariel Menezes and Bruna Assunção, in-house translators at CADE's International Unit.

[2] The adoption of such tools, although with different levels of maturity and different goals, has become an inevitable reality for antitrust authorities. See: Schrepel, T.; Groza, T. The Adoption of Computational Antitrust by Agencies: 2021 Report (21 June, 2022). 2 Stanford Computational Antitrust, 78 (2022), VU University Amsterdam Legal Studies Paper Series, Available at SSRN: https://ssrn.com/abstract=4142225. https://ssrn.com/abstract=4142225

[3] Chen, J.; Walker, R. M.; Sawhney, M. (2020) Public service innovation: a typology, Public Management Review, 22:11, 1674-1695, DOI: 10.1080/14719037.2019.1645874.

[4] The two consultants—one expert in IT and the other in econometrics—worked during the first two years of the project providing input to the development and testing of the data mining and data screening tools.

[5] The partnership with the Ministry of Economy provided access to the database of federal public procurement from as of 1994, which totalled more than one million tenders.

[6] Especially for obtaining data on government tenders held by Brazilian states and municipalities.

variables. The decision to start testing with the best available database (but not a perfect database) was essential to produce the first results and have the institutional support needed for such an initiative.

6. As a project based on innovation, CADE's leaders continuously supported it as they understood that maturing the techniques developed could result in undesirable outcomes (requiring more time to fix or modify them[7]) and possibly delayed results. Therefore, institutional support[8] was vital for reaching the first results.

2. Key guidelines for the project

7. After overcoming the initial challenges, we developed a few ideas to guide our staff:

- Focus: We developed new data mining tools and screening models and quickly started testing the available databases (even if they lacked information or were imperfect) to verify which techniques were accurate to use in actual cases.
- Transparency: We documented the development and testing of the techniques and models, which allowed us to share and enhance them later and to pinpoint their limitations and results. Next, we "translated" this information to make it easier to comprehend for non-specialists who would later assess it, such as judges and attorneys.
- Partnerships: We shared our techniques and models with bodies of criminal prosecution and external control[9] that carry criminal and administrative investigations into cartels in procurement. It broadens the reach and deterrence of the tools and reduces development and testing costs. These partnerships may also include sharing technical information and access to data from tenders held by states and municipalities. Similarly, we conducted meetings with antitrust authorities from several other jurisdictions[10], in which we shared experiences and technical documents of reference.

8. In the context of Project Cérebro, CADE entered into technical cooperation agreements in Brazil with prosecution services from every state, and prominent external control agencies (like the courts of accounts from several states and towns). By doing so, it managed to access local public procurement data and conduct probes in partnership with these bodies.

[7] In this regard, it is important the leaders are mindful of the possibility of errors in the process of innovation (and, consequently, are prepared for possible delays)—something not so common for the Brazilian government.

[8] We stress that, since the implementation of Cérebro, four different Superintendent Generals (i.e. leaders of CADE's investigative arm, responsible for the project) have supported its continuation.

[9] In 2019, CADE published guidelines on the implementation of technical cooperation with bodies of persecution and control that, amongst other recommendations, introduced procedures for joint actions and information sharing in the development and use of mining and screening techniques. The document is available in Portuguese at: https://cdn.cade.gov.br/Portal/centrais-de-conteudo/publicacoes/guias-e-manuais-administrativos-e-procedimentais/guia-para-operacionalizacao-de-cooperacao-tecnica-com-orgaos-de-persecucao-e-controle.pdf.

[10] From its start, the project had meetings with 14 antitrust agencies on the use of data analysis tools applied to investigations.

3. Description of the case

9. In 2016, a CADE's procurement team member identified a red flag in a tender for outsourced services and reported it to the investigative unit.

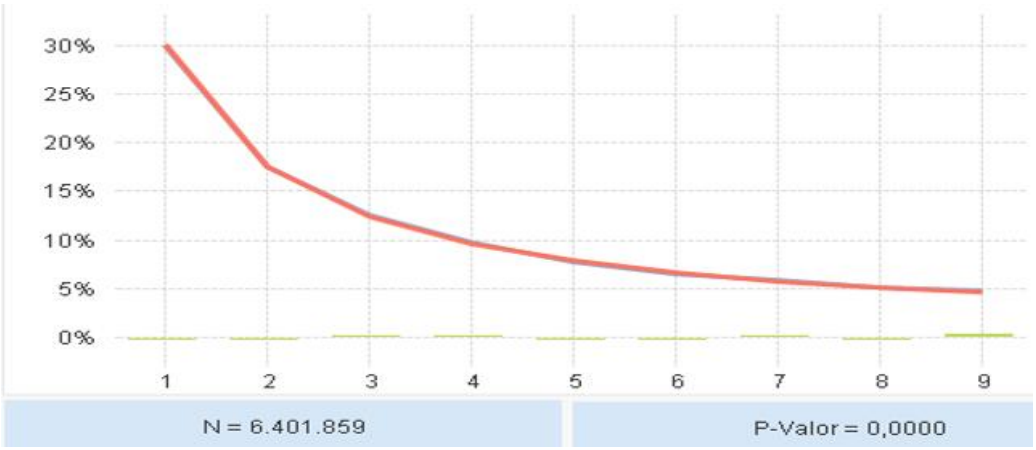
10. In summary, there were signs that two bidders of the same business group had either engaged in concerted behaviour or shared competitively sensitive information or both.

11. From these red flags, CADE (i) identified 55 companies that either acted in the same manner or were connected to the two initial companies or both and (ii) picked federal government tenders in which the 55 companies recently participated, accounting for 4,400 tenders.

12. The authority then used two screening models developed together with a UNDP consultant and validated by CADE's technical team[11].

13. The first model, based on the Newcomb-Benford law[12], aimed to identify whether there was cover bidding in the tenders. In broad terms, the authority analysed the frequency of the first significant figure in sealed bids[13] of 4,400 tenders and compared it to the expected frequency. Our hypothesis predicted that cover bids—made to lose the tender and, thus, disconnected from the actual costs and profit margins of companies—would not follow the Newcomb-Benford law:

Figure 1. Frequency of the first significant figure in sealed bids from Brazilian federal government tenders (2015–2020).[14]



Source: Administrative Council for Economic Defense

[11] The validation process included experts from a unit in charge of investigating cartels in procurement and the Department of Economic Studies.

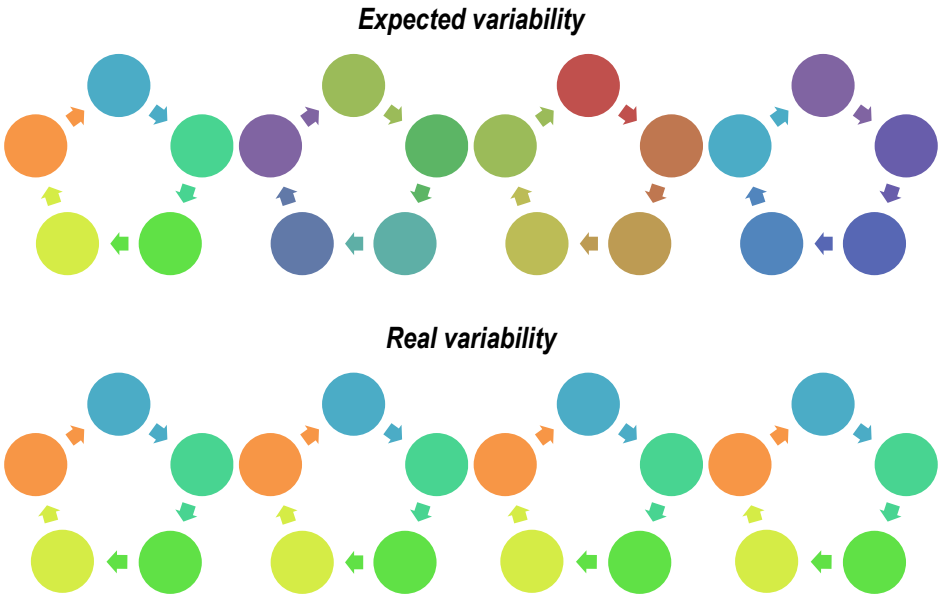
[12] Tödter, K-H. (2009). Benford's Law as an Indicator of Fraud in Economics. German Economic Review. 10. 339-351. 10.1111/j.1468-0475.2009.00475.x.

[13] Nearly every tender analysed consisted in electronic reverse auctions, where bidders would submit sealed bids (uploading them in the platform) and, later, make gradually smaller open bids in the platform. We examined the first significant figure of sealed bids, not the open bids.

[14] The figure shows that, from 6,401,859 sealed bids submitted at the time, the first significant figure followed the Newcomb-Benford law, indicating that sets of bids with different frequency patterns could indicate cover bidding.

14. The second model, called Entropy[15], aimed to identify bid rotation. In sum, we checked whether companies' positions in the tenders showed any variance, looking for frequent losers and winners.

Figure 2. Real and expected variability of bidders' positions[16]



Source: Administrative Council for Economic Defense

15. The two models raised red flags for 16 companies that consistently showed typical cover bidding and bid rotation behaviour[17].

16. Nonetheless, considering the applied techniques were very recent and the investigative strategy pointed to the necessity of a search and seizure to look for direct evidence of communication between the companies, CADE decided to search for additional evidence to corroborate the models' results.

17. Based on CADE's previous experience, the investigators suggested checking whether any of the 16 companies had engaged in two other atypical behaviours: two companies presenting, in the same tender, sealed bids with the same value that differs from the reserve price, and two or more companies submitting open bids from the same IP address. With the support of the UNDP consultant, we developed mining techniques in the federal government tender database, which allowed us to look for the evidence and prove the 16 companies were involved in cover bidding and bid rotation.

[15] Bajari, P., & Ye, L. (2003). Deciding between competition and collusion. Review of Economics and Statistics, 85(4), 971–989. MIT Press.

[16] The model aimed to check if the real variability, found in the suspicious tenders, was similar to the expected variability.

[17] Using two or more filters at a time is the best strategy to conduct probes used to launching administrative proceedings as it reduces the chance of Type I errors (false positives).

18. We shared these results with the Federal Police Department, which carried out additional measures and launched Operation Meeting Point on 31 October 2018, serving search warrants at the head office of 14 companies[18].

19. Finally, from the analysis of the seized electronic evidence, on 14 September 2021, CADE initiated an administrative proceeding against 42 companies and 43 individuals involved in a large cartel in public and private procurement for outsourced services[19].

20. This case provided us with the first results of the tools, the opportunity to validate their accuracy empirically and, most importantly, allowed our staff to develop strategies to communicate these findings to non-specialists. This effort was significant in formulating the investigative hypothesis (in partnership with the Federal Police), communicating with the judge responsible for analysing the requests for search warrants and, finally, preparing the papers that indicated the need for an administrative proceeding.

21. Although CADE's Administrative Tribunal has yet to issue its final decision on this case—which is still under probe—it is worth mentioning there has been no legal questioning of the quality of the techniques that gave reasons for the search and seizure.

4. Old problems, new challenges

22. Project Cérebro has enabled its team to find some points that can require particular attention in the coming years:

1. a) Development: The Brazilian institutional scenario requires an equilibrium between designing new filters and adapting those already tested and validated by experts. Investment in human and organisational resources is necessary to develop new techniques and adapt techniques that have been tested and used by other organisations since firms may adapt their behaviour to data mining techniques and screening tools the authority makes public[20]. Such development calls for a team with a high degree of expertise and the capability to use data mining and screening techniques. At the same time, it demands adapting the existing models to different institutions and their realities[21]. And these are no ordinary requirements.
2. b) The Cérebro team: In the creation and management of teams with the expertise necessary for the project, there is internal competition—from public agencies interested in professionals with similar qualifications—and external competition—as the growing remote work trend in the IT field intensified competition with the

[18] The police conducted only 14 dawn raids because out of the 16 companies, one was already closed, and two had the same address.

[19] A public version of the document that initiated the administrative proceeding, describing the adopted screening models and data mining techniques, is available in Portuguese at: https://sei.cade.gov.br/sei/modulos/pesquisa/md_pesq_documento_consulta_externa.php?HJ7F4wnIPj2Y8B7Bj80h1lskjh7ohC8yMfhLoDBLddY_LGcsxfX2HCQpNx7m0B1UF9RWscvGxs64vdOcfQxclLWkamZuuKFhT0ozapmEwrIBqmd4mq0X4xYuNG8X857Y

[20] Ortner, J.; Chassang, S.; Kawai, K.; Nakabayashi, J. Screening Adaptive Cartels (July 2022). NBER Working Paper No. w30219, Available at SSRN: https://ssrn.com/abstract=4153112. https://ssrn.com/abstract=4153112

[21] In other words, compatibility between the institutional reality of the country for which it was designed (that is, the types of procurement and the antitrust regulations) and the institutional reality of the country attempting to adapt it.

private market. These raise a few issues relating to professional recruiting and, most importantly, employee retention[22].

3. c) Governance: Our institution provided the necessary support to conduct the project, with the participation of our leaders in the conception of the project and transparency about the risks entailed in developing the employed techniques (especially the potential time lapse between the outset of the project and its first results). In this context, the project management team had the opportunity to discuss with other units of CADE about mapping the real needs of the areas responsible for investigating cases and gathering input to develop Cérebro[23]. Therefore, it was possible to have different units involved in the project and reach solutions to cross-cutting issues[24].

4. d) Data: Having access to data promptly and with the required levels of quality and granularity is a constant challenge[25]. It is also essential to find a balance between efforts to obtain new databases (through partnerships or AI initiatives); maintain, update, and process databases already in use[26]; map the possible effects (on data access and data processing) of personal data protection regulations and AI-based solutions.

5. e) Boundaries: Recently, more antitrust authorities have been adopting AI-based solutions to boost the efficiency of investigations or improve the detection of antitrust violations[27]. And they all face the same challenges in implementing these tools. Thus, it is essential to formulate strategies to prevent database biases from affecting the quality of results[28], which undermines the confidence in their investigative use.

23. Such challenges can be an opportunity for agencies that have yet to carry out similar projects and a call to those that have already implemented them. In short, they inspire antitrust agencies to create an open environment that allows for sharing experiences

[22] The team currently consists of two civil servants: one holds a PhD in Electrical Engineering and the other a M.Sc. in Economics. Formerly, the team had members with academic backgrounds in Law, Economics, and Computer Science.

[23] The discussions have allowed the Cérebro team to get input to develop techniques—such as the main strategies used in collusive agreements observed by investigators—and, especially, to consider investigators' real needs so as to develop solutions for concrete end-user problems.

[24] For instance, solutions concerning the analysis of shareholder structure—a common need of units responsible for investigating mergers and cartels.

[25] Recently, CADE obtained authorisation to access government procurement data with more granularity, which will enable the testing of models already validated by the literature and the development of new techniques.

[26] Although acquiring new databases should be a permanent goal for CADE, the authority must take into account the human resources and time expended in ETL processes (data extract, transform, and load) as well as for documenting and/or getting familiar with the variables of each new database.

[27] See Gryz, J. & Rojszczak, M. (2021). Black box algorithms and the rights of individuals: no easy solution to the "explainability" problem. Internet Policy Review, 10(2). https://doi.org/10.14763/2021.2.1564.

[28] Almost every database has been developed for different purposes and not for the antitrust authority's specific uses.

(effective and ineffective ones), techniques, tools, and methods to foster development and produce better results.

24. While it is true that data mining, screening tools, and artificial intelligence methods are no remedy for anticompetitive behaviour, they are, at least, essential for antitrust authorities to produce more dissuasive effects.