Title
Assessing the potential for detecting fraud in the Brazilian public procurement using Latent Class Analysis

Author
Alencar, Eduardo

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Assessing the potential for detecting fraud in the Brazilian public procurement using Latent Class Analysis

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Social Ecology

by

Eduardo Carvalho Nepomuceno Alencar

Dissertation Committee:
Dr. John Hipp, Committee Chair
Criminology, Law & Society
University of California, Irvine

Dr. Kirk Williams
Director, Ph.D. Core Program in Social Ecology
Criminology, Law & Society

Dr. Richard McCleary
Criminology, Law & Society
University of California, Irvine

2021

DEDICATION

To my spouse, Ludmilla Alencar,

in recognition of her unwavering love, support, and willing sacrifice

To my parents and friends

in recognition of their worth

"The ultimate measure of a man is not where he stands in moments of comfort and convenience, but where he stands at times of challenge and controversy."

Dr. Martin Luther King, Jr.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AIC	Akaike's Information Criterion
Alice	System for the Analysis of Bids and Tenders
BIC	Bayesian Information Criterion
BLRT	Bootstrap Likelihood Ratio Test
CAIC	Consistent AIC
CGU	Office of the Comptroller General
GAMs	Generalized Additive Models
GDP	Gross Domestic Product
IC	Information Criteria
ICT	Information and Communication Technology
LCA	Latent Class Analysis
LMR	Likelihood Ratio Test
LPA	Latent Profile Analysis
MB	Managing Body
NPB	Non-participating Body
OECD	The Organization for Economic Co-operation and Development
PB	Participating Body
PMI	Project Management Institute
PR	Price Registry
PRS	Price Registration System
PTE	Pre-Tender Estimate
SABIC	Sample Adjusted BIC
SAI	Supreme Audit Institution
SIASG	General Service Management System
TCU	Brazilian Federal Court of Accounts
UASG	Service Administrative Unit
UIF	Upper Inner Fence

ACKNOWLEDGEMENTS

I want to express my most profound appreciation to my committee chair, Professor John Hipp. I am grateful for the guidance and continuous support during this project. His view about data analysis and science inspired me to move forward on my academic journey. A special thanks to my committee members, Professors Kirk Williams, Richard McCleary, Alexander Ihler, and Gabriella Montinola, for my research's insightful and objective contributions.

I am thankful to my wife, Ludmilla Alencar, for embarking on this challenging adventure to live abroad with our three kids. I thank you for your endless support, patience, and love. I'm sure these last few months haven't been easy for anyone, but in moments of high pressure, I can see who I can count on and the value and importance of family in the central moments of life. I also would like to mention my kids, Eduarda, Catarina, and Miguel. You were so brave to have the ability to adapt to a different culture in the US and make my journey more pleasant.

I want to thank my parents, Eduardo and Jussara, for the life values they transmitted to me, making me always try to be correct and honest. You deserve all the best.

I also couldn't fail to mention my in-laws, Possidonio and Mariza, who supported us from the very first moment.

Harm thanks to my friend, Fabrizio Almeida Marodin, and his family. Having met you was a blessing. Thanks to Caio Batista for being humble and open to help. Also, I want to express my gratitude to Jose Saramago for sharing his views and experience on navigating the UCI programs.

It is my life mission to be a correct person who positively impacts people's lives. I hope that this research can somehow improve government efficiency and fight corruption.

VITA

Eduardo Carvalho Nepomuceno Alencar

2002-03	Economic Consultant, Institute for Applied Economic Research (Ipea), Brazil
2002	Graduate Certificate in Business, Fundação Getúlio Vargas, Brasília, DF, Brazil
2003	B. S. in Economic Sciences, Universidade de Brasília, Brasília, DF, Brazil
2005–present	Economic Specialist in Regulation of Telecommunication Services, National Agency of Telecommunications (Anatel), Brazil
2006-09	Advisor, Anatel, Brazil
2007	International Executive MBA, Fundação Getúlio Vargas, Brasília, DF, Brazil
2009-13	Talents Development Manager, Anatel, Brazil
2013-14	Project Manager, Secretariat of Strategic Affairs at the Presidency of the Republic, Brazil
2014	Lecturer, UPIS - Faculdades Integradas, Brasília, DF, Brazil
2014-15	Advisor, Cabinet of the Presidency of the Republic, Brazil
2015-16	Advisor, Office of the Comptroller General (CGU), Brazil
2016	Advisor, Ministry of Planning, Development, and Management, Brazil
2016-17	Chief of Staff, Ministry of Foreign Affairs, Brazil
2016-17	Member of Administration Board, Geological Survey of Brazil (CPRM)
2018-20	Teaching Assistant, School of Social Ecology, University of California, Irvine
2020	M.A. in Social Ecology, University of California, Irvine
2021-present	Chief of Governance and Strategic Management, National Institute of Educational Research Anisio Teixeira (Inep), Brazil

FIELD OF STUDY

Public Policy, Regulation, Structural Equation Modeling, Synthetic Control Modeling, Latent Class Analysis, Machine Learning, Fighting Corruption, Public Procurement

PUBLICATIONS

Acknowledgments in "What Drives Water Utility Selection of Pricing Methods? Evidence from California". Allaire, M., Dinar, A. Water Resour Manage (2021). https://doi.org/10.1007/s11269-021-03018-8

"Applying Synthetic Control Method to estimate the impact of the Lava Jato operation on the Worldwide Governance Indicators in Brazil", Journal of White-Collar and Corporate Crime, Alencar, E. C. N., & Jackson-Green, B., 2021. https://doi.org/10.1177/2631309X211017874

"Applying Structural Equation Modeling to Analyze the Mutual Causality between Corruption and Trust and the Impact of Lava Jato: Evidence from Brazil and Mexico", University of California, Irvine, 2020.

"The Impact of the Lava Jato on Corruption and Trust: Evidence from Brazil", Policy Brief, 2020.

"Trends in Water Rate Structures to Encourage Conservation: Evidence from California" with Maura Allaire. American Geophysical Union, Fall Meeting, Washington, DC, 2018.

"Gestão de pessoas aplicada ao coaching" in Coaching e Mentoring - Foco na Excelência, Sermais, 2014 "Caderno Vozes da Classe Média: Classe Média e Emprego Assalariado", SAE, 2013.

"Projeto Micro Recarga no Serviço Móvel Pessoal – SMP". Anatel, 2006.

"Política industrial, acompanhamento e análise", 24, IPEA, 2004.

"Política industrial, acompanhamento e análise", 23, IPEA, 2004.

"Política industrial, acompanhamento e análise", 22, IPEA, 2004.

"Política industrial, acompanhamento e análise", 21, IPEA, 2003.

Endnote in "A Classroom Guide to the Equilibrium Exchange Rate Model", Economic Issues, 2002.

ABSTRACT OF THE DISSERTATION

Assessing the potential for detecting fraud in the Brazilian public procurement using Latent Class Analysis

by

Eduardo Carvalho Nepomuceno Alencar

Doctor of Philosophy in Social Ecology

University of California, Irvine, 2021

Professor John Hipp, Chair

Based on 2019 procurement data composed of almost 1.5 million observations, this study employs Latent Class Analysis (LCA) to investigate how latent profiles are defined to represent behavior in the Price Registration minutes based on a set of selected variables. LCA approach brought the possibility of identifying some characteristics that can suggest or direct the efforts of the control bodies in identifying illegal cases. From the seventeen-profile model with binary variables, we identified a latent class (*N*=2589) that could be the object of a more detailed analysis by the control bodies. From the LCA with binary and continuous variables, the seven-class solution provided the optimal fit for our data, with the lowest BIC value and a high entropy value. The seven relevant factors that increase the risks of renting price registration drafts are the: (i) number of Non-participant Bodies (NPB) per tender; (ii) relative item value; (iii) amount of NPB per supplier per tender; (iv) distribution of winners; (v) proportion of approved quantities per item for NPB per tender; and (vi) tender with only Managing Body (MB) and NPB. The latent class with a high NPB presence with few suppliers (*N*=749) was the most salient to be more favorable for the sale of price registration minutes, intended to be carried out in future studies. Our results highlight that such an organization-centered approach can provide insights not apparent in the more common variable-centered practice. Although our goal was to unpack further insights within Price Registration System, future work may wish to include more of the measures from the models in the clustering routine.

INTRODUCTION

The literature has shown how corruption affects nations' economic development, reducing market efficiency, eroding public services' quality, and contributing to poverty and social inequality (Kessing et al., 2007; Mauro, 1995). Corrupt activities can lead to public money leakages, reducing governments' capacity to collect tax revenues and artificially increasing the values for goods, services, or investment projects. According to the International Monetary Fund report (2019), distortions in spending priorities under-mine the State's ability to promote sustainable and inclusive growth. Although it is challenging to measure corruption's exact cost due to its hidden nature, corruption costs $1 trillion in tax revenue globally (IMF, 2019). According to the United Nations, those illicit transactions cost at least 5% of GDP, around USD 3.6 trillion[1], every year. However, the cost of corruption is higher than the sum of the lost money.

Consumers worldwide expect the benefits of open, fair, and equitable competition in the market, resulting in having the best products and services at the lowest prices (Vadász et al., 2016). In a genuinely free market, companies compete to attract and meet their customers' purchasing needs. This efficient process can only succeed when competitors set their prices independently. When there is collusion between competitors, prices go up, quality often drops, and the population loses. In all developed countries globally, price-fixing, bid-rigging, and other forms of cartels are illegal and can be prosecuted.

Public procurement refers to all tenders carried out by authorities to meet the economy's needs, from tenders to brochures and construction works. This covers the entire process from start to finish, which means that many counterparties are involved, and the risk of corruption is high. The World Bank distinguishes between two different categories of corruption that can be applied to public procurement (Bhargava, 2005). The first category is state capture, which refers to the actions of individuals, groups, or organizations to influence the design of public order through the illegal transfer of private benefits to public officials. The second refers to

[1] https://www.weforum.org/agenda/2018/12/the-global-economy-loses-3-6-trillion-to-corruption-each-year-says-u-n

administrative corruption, a similar corruption, and bribery used by the same actors to disrupt the proper application of laws, rules, and regulations.

We can see that corruption in the public sector is at stake around the world. Besides, about two-thirds of the identified foreign bribery cases occurred in the extractive, construction, transportation, and information technology sectors (OECD Foreign Bribery Report, 2014). Public procurement is "one of the government activities most vulnerable to corruption" (OECD, 2016, p. 6) and represented in 2014 an essential share of the Brazilian economy, corresponding to approximately 8.2% of GDP (OECD, 2017a). One way to discourage collusion and anti-competitive behavior by firms in federal procurement processes would be to sustain free and fair competition.

The work is developed in two stages. Thus, after introducing the object of the work and its context, Part I explains the Brazilian public procurement process and reviews the relevant literature on corruption, cartel, and machine learning. Part II presents the research. We apply Latent Class Analysis (LCA) to signal price registry drafts potentially used as rent instruments. We investigate what factors would most characterize those rented price registry drafts.

Research constraints

According to project management's best practices, it is essential to identify the triple restrictions (scope, time, cost) to verify its viability (PMI, 2017). In this sense, the primary constraint for this project was time. I am an employee of the Brazilian government, licensed to pursue my doctoral research abroad in up to 4 years. Thus, according to current Brazilian legislation, I must return to Brazil by August 2021. I made efforts to accelerate my learning process and complete the program in 4 years instead of 6 (average duration). In this sense, I used the strategy to narrow the research scope and prioritize public data without compromising the intellectual value delivered after the research.

Another constrain was data access. To overcome the need to access data using API from the Government website, I negotiated with the Brazilian Federal Court of Accounts (TCU) to access the data. TCU provided data for the dissertation. Although there is a perception by the control bodies of the relevance of this problem, the subject is relatively new and requires investigation. TCU seeks to analyze the topic and has access to the Price Registration system database. My interest in obtaining the data through the TCU aims to guarantee data extraction quality. The agency has technical excellence in understanding the data and standard extraction routines already used.

1. THE BRAZILIAN PROCUREMENT PROCESS

Brazil's public procurement computer system has several advantages. First, it has a reasonably well-implemented organizational identification system, a prerequisite for analyzing organizational behavior as collusion. Secondly, it registers bid prices for the winners and the losing bidders, necessary for in-depth analysis of cartels. However, there are some shortcomings in Brazil's public procurement computer system. First, the central platform (Comprasnet), which publishes all tenders regulated by national public procurement laws, does not have loaded all state and city tenders, increasing the cost of obtaining tender information if it is available. Second, Comprasnet does not include purchases made by wholly or partly state-owned companies such as Petrobrás, Banco do Brasil, and others.

Comprasnet contains high-quality data on several key variables, such as contracting authority information or key procedural characteristics (e.g., type of procedure, framework agreement). Structurally, the Brazilian public procurement database registers information at the procurement and per lot level. The entire database can be used as an actual transaction between buyers and sellers. In addition to Comprasnet, the federal public administration also has other tender portals such as Licitações-e, Petronect, and the Caixa Econômica Federal procurement portal. However, access to information from these portals is more complicated, making it difficult to use automated data extraction tools[2].

In public bidding, one of the main reasons for the diversion of public money is a cartel[3] formation, which consists of a group of suppliers that have made an explicit agreement to limit competition between them for their benefit. In the cartel, the decision consists of a voluntary association of decision-makers, each aware that their profit depends on the behavior of all industry participants. Therefore, it involves much more than a simple definition of the price level and maximum profit amount (Campos, 2008). Braga (2015) conceptualizes cartels as agreements between competitors to achieve some collective benefit to the competition's detriment. Common sense is that cartels are adjustments between competitors to fix prices above the

---

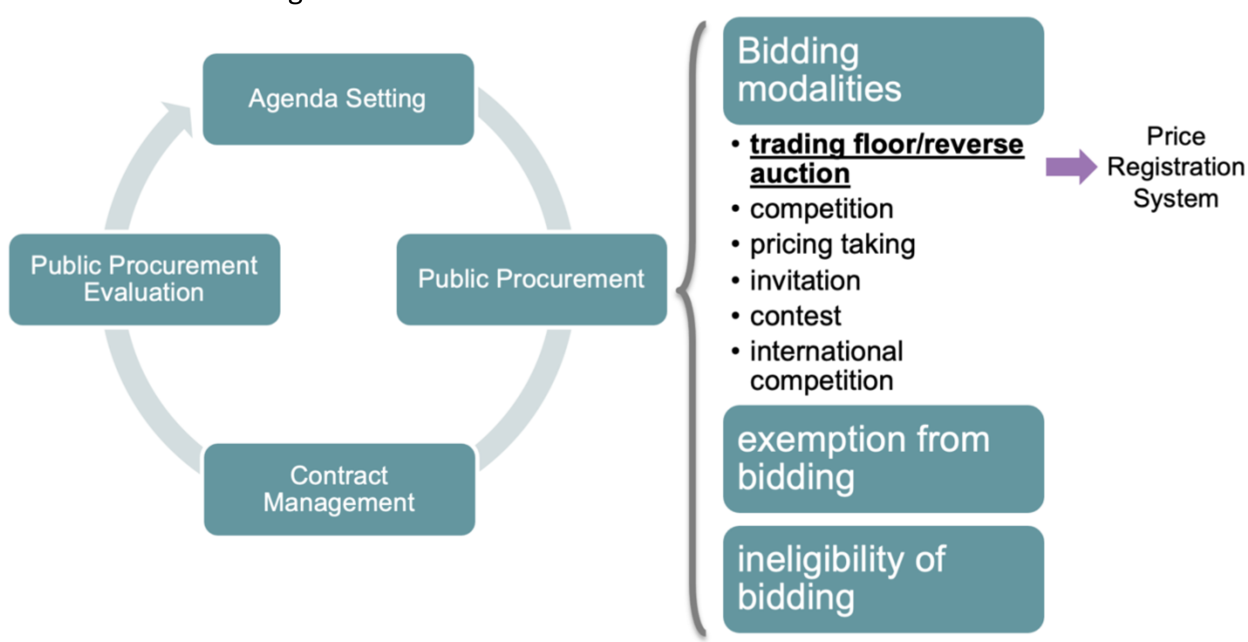
[2] Perception captured through meetings with TCU auditors.

[3] The terms collusion and cartel will be used interchangeably.

equilibrium price verified in the market. However, this myth must be immediately broken, as these companies do not seek gains solely through price-fixing.

The procurement process of products or services in the Brazilian federal government is based upon Law 8666/1993 (Brazil, 1993), called Procurement Law. It details the bidding process stages, the bidding types allowed, types of contracts, aspects of qualification of companies and defines administrative and criminal penalties to be applied to suppliers in non-compliance. The Brazilian public procurement macro process is presented in figure 1.1. The Brazilian public procurement legal framework provides six types of bidding and two possibilities of tender exemptions[4]. According to Laws 8.666 and 10.520 (Brazil, 1993, 2002), competition is the bidding modality between any interested parties that, in the initial preliminary qualification phase, prove to have the minimum qualification requirements required in the execution notice of its object. Price taking refers to interested parties duly registered or who meet all the conditions required for registration by the third day before the date of receipt of proposals, subject to the necessary qualification. The auction is the bidding modality aimed at the acquisition of common goods and services.

Figure 1.1: Brazilian Public Procurement Macro Process



Source: Elaborated by the author

[4] Articles 24 and 25 of the Public Procurement Law provide for cases that are exempted from the public tender procedure.

According to data of the Purchasing Panel of the Federal Government, between 2015 and 2019, the government's total value of contracting through bidding and tender exemptions was R$ 435 billion (US$ 108 billion). Furthermore, the government used the Reverse Auction (*Pregão*) modality to correspond to 48% of the value contracted. Additionally, considering the evolution of public purchases, more than 95% of the public purchases have been realized in the Reverse Auction modality or tender exemptions. The technology could help public managers identify corruption and prevent fair allocation of public resources.

In this context, two auditing institutions have a decisive role in improving government efficiency. The Office of the Comptroller General (CGU) is the Federal Government's internal control body responsible for carrying out activities related to public assets defense and increased management transparency through public actions audit, correction, prevention, and fight against corruption and ombudsman. Depending on the type of problem or risk identified in its contract audits, CGU may recommend improving governance, liability verification, replacement of assets and values, adjusting objects, or cessation (suspension of execution) of items. The Brazilian Federal Court of Accounts (TCU) is the Brazilian supreme audit institution (SAI), responsible for auditing government accounts to hold public officials accountable. TCU uses information on materiality, risk, and typologies identified by the Alice system to audit the contracts.

The development of the System for the Analysis of Bids and Tenders (Alice) by CGU in 2014 and its adoption by TCU in 2016 marked a new stage of how public agencies analyze data. The System aims to identify evidence of irregularities in tenders through artificial intelligence automatically. Alice reads bidding notices and price records published by the federal administration by collecting information from the Brazilian government's official journal (*Diário Oficial da União*) and Comprasnet (Desordi & Bona, 2020). This tool accesses the Comprasnet[5] and downloads files and data from all public tenders and minutes published to identify irregularities in bids and electronic auctions from the federal public administration based on the text of the public notice. This innovation has enabled the timely and automated evaluation of

[5] Comprasnet is the Federal Government's Purchasing Portal established to make available to society information related to bids and contracts carried out by the federal government.

bidding notices and trading minutes, identifying signs of irregularities, fraud, diversion, and waste of public resources, enabling more efficient and effective control actions (Panis, 2020).

1.1. Price Registration System (PRS)

Law 8,666/93 attempted to program a new figure within the scope of the Brazilian Public Administration's procurement system: the price registration system (PRS). It is not another type of bidding but a variant that can help and give a new public procurement speed (Da Silva, 2017). PRS is an instrument that facilitates the acquisition and contracting of shared services by the Public Administration in Brazil. It has several characteristics that untie a series of bureaucratic ties that hinder the efficiency and agility of Brazilian tenders. The PRS also allows for shared hiring between public agencies, including between different political entities. Decree 7,892/2013 regulates the PRS and establishes three types of public bodies:

- Managing body (MB): an entity of the federal public administration responsible for conducting procedures for registering prices and managing the resulting price registration drafts.
- Participating body (PB): a public administration entity that participates in the Price Registration System's initial procedures and price registration drafts.
- Non-participating body (NPB): an entity of the public administration that, having not participated in the initial bidding procedures, complying with the requirements of this standard, adheres to the price registration drafts.

Renting price registration drafts (*Barriga de Aluguel* in Portuguese) occurs when price registration minutes are generated with unnecessary or overestimated amounts solely to favor a particular supplier. Later, this supplier will try to "market" the items registered in other bodies and entities of the public administration for membership purposes. If not adequately motivated by the real needs of the participants, this practice artificially increases the volume tendered in the minutes, allowing a more significant number of adhesions, and allowing the winning company to freely market its products with the various bodies of the Public Administration without bidding[6].

Otherwise, the rent of price registration drafts occurs when the bid quantities are falsely increased. The resulting price registration record allows for a more significant number of adhesions and allows the winning company to market its products with various Administration bodies Public without bidding.

Purchases through PRS can be made through the adhesion in price registry (PR) drafts by non-participating bodies. Santana (2014) recalls that the Brazilian Federal Court of Accounts (TCU) recognizes a parallel market for acquisitions, even focusing on corruption and deviations from public and private interests due to renting price registration drafts. Besides, according to conversations with TCU auditors, the acceptance of its non-participating bodies opens a possibility of misuse of the PRS to obtain undue advantages on the part of the managed agency's public agent. In this case, the managing body may ask the supplier company for some advantage (in the form of monetary values or not) to authorize non-participating bodies to adhere to the price registration drafts. Consequently, TCU auditors highlighted that a line of investigation to identify suspicious behaviors in the SRP framework could generate an unprecedented saving of resources in public waiting.

With all this in mind, we plan to gain a better technical understanding of the behavior of firms in the public procurement process through this research, then make policy recommendations on how public policy should decrease the possibilities of fraud in the public biddings. The proposed dissertation seeks to circumvent the lack or scarcity of identified frauds. There is no information available on cases already identified and judged in the context of the sale of price registry minutes. The analysis strategy involves the use of screening and Latent Class Analysis, which will be detailed later.

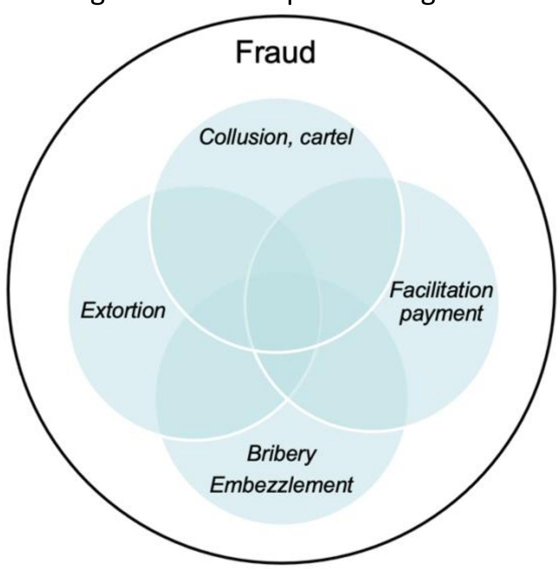
[6] http://www.controleexterno.com/jurisprudencia-comentada-barriga-de-aluguel/

2. REVIEW OF RELEVANT LITERATURE

2.1. Corruption

The literature widely associates corruption with abuse of power for personal gain (Bardhan, 1997; Jain, 2001; Svensson, 2005). The above definition leads to three types of corruption in public administration depending on the kind of decision-makers; a source of power that distracts the decision-maker; and the models used to explain corruption. The first type of "grand corruption" is the corruption of political elites in economic policy development. The second type of "bureaucratic corruption" is the corruption by bureaucracies in their relations with political elites or the public. The third type, "legislative corruption", is the extent to which interest groups can influence legislators' voting behavior (Jain, 2001). Rose-Ackerman and Palifka (2016) identify eight types of corruption: bribery, extortion, judicial fraud, accounting fraud, public service fraud, favors' exchange, cronyism, nepotism, embezzlement, kleptocracy, influence peddling, and conflicts of interest. The authors also describe petty corruption as small-scale corruption at the lower levels of society (between public officials and citizens). Menocal (2015) provides a summary of corruption in well-defined categories (Table 2.1). However, in reality, those categories are not independent of one another. They all overlap in some instances (Figure 2.1).

Figure 2.1: Corruption Categories



Elaborated by the author

Tanzi and Davoodi (1998) and Salih (2013) presented some indicators used to explain corruption. The World Bank published guidelines for ten significant fraud and corruption indicators in public procurement for bank-financed projects, confirmed by evidence for fraud and corruption investigations in the public and the private sector (Põlluste, 2019). Van Vlasselaer et al. (2015) provide additional theoretical sources, theoretical models, and expert opinions on the probable symptoms of corruption. Finally, Transparency International's research on public procurement corruption has combined corruption indicators in the overall public procurement cycle (Jensen & Anderson, 2016).

Table 2.1: Corruption Categories

Categories of corruption	Description
Bribery	Convince a person to act dishonestly on their behalf through payment or other incentives. Incentives can take the form of gifts, loans, fees, rewards, or other benefits (taxes, services, donations, etc.). The use of bribes (for example, inspectors report violations in exchange for bribes) or extortion (for example, bribes) can lead to corruption.
Embezzlement	To steal, misdirect or misappropriate funds or assets placed in one's trust or under one's control. From a legal point of view, embezzlement need not necessarily be or involve corruption.
Facilitation payment	A small payment, also known as "speed" or "grease" payment, is made to perform a routine, or required action to which the payer is legally or otherwise entitled.
Fraud	Intentional and dishonest fraud by a person to obtain an unfair or illegal advantage (economic, political, or otherwise)
Collusion, cartel	An agreement between two or more parties aimed at achieving an illegal purpose, including improper influence on the other party's actions
Extortion	The act of causing bodily injury or damage, or the threat of injury or damage, directly or indirectly, to any part or property of a party with the intent to wrongly influence the party's actions.
Patronage, clientelism, nepotism	In the government, this refers to the practice of direct appointment.

Source: Menocal (2015).

Corruption also involves different forms of collusion between the government and bidders. For example, the contracts can be awarded based on bribes, political supporters, or firms with a personnel interest. Such corruption, which can occur in the execution and award of contracts, may prevent authorities from achieving value for money in their acquisitions since corrupted contracts will not be awarded to the best firms (Tas et al., 2017). The principle of integrity is a decisive component of this discussion because procurement regulation should ensure probity, and improper collusion between government and particular suppliers should be eliminated (McCrudden & Gross, 2006).

2.2. Cartel

One primary type of corruption is collusion or cartel, an agreement between competing firms to control prices or exclude a new competitor's entry into a market (Vadász et al., 2016). Economists have a long-standing interest in detecting cartels using data on companies, bids, and prices over the years. An early example is the Oklahoma asphalt market in the late 1950s, where companies submitted identical bids (Funderburk, 1974). Bid-rigging scenarios and market allocation in various US school milk markets in the 1980s and 1990s led to articles on counter-behavior and pricing on collusion and competition (Lanzillotti, 1996; Porter & Zona, 1997, Pesendorfer, 2000; Scott, 2000). The professional and scientific literature contains many publications on healthy competition, cartels, collusion, and public procurement corruption. In particular, since 2005, several well-documented articles have been published in scientific journals (Cosnita-Langlais & Tropeano, 2013; Hinloopen & Martin, 2006; Marshall & Marx, 2007; Morgan, 2009; Padhi & Mohapatra, 2011; Werden et al., 2011), books and scientific articles (Abrantes-Metz, 2013; Harrington, 2006, 2008; Hüschelrath & Veith, 2011; Morozov & Podzolkina, 2013), as well as in official government documents (Danger & Capobianco, 2009a, b; OECD, 2012).

A detailed analysis of the economic consequences of cartel activities is outside the scope of this research. However, Ivaldi et al. (2014) provide an example of the cumulative economic impact of cartels in 12 developing countries (Brazil, Chile, Colombia, Indonesia, South Africa, Mexico, Pakistan, Peru, Russia, South Korea, Ukraine, and Zambia) in terms of sales related to Gross Domestic Product (GDP) taken in average during 2002. The analysis shows that a cartel can reduce production by about 15% in a specific market. For the case of Brazil, the impact was 1.86% on affected sales in 1999, and the actual damage reached 0.4% of GDP in terms of cartels' excess revenues.

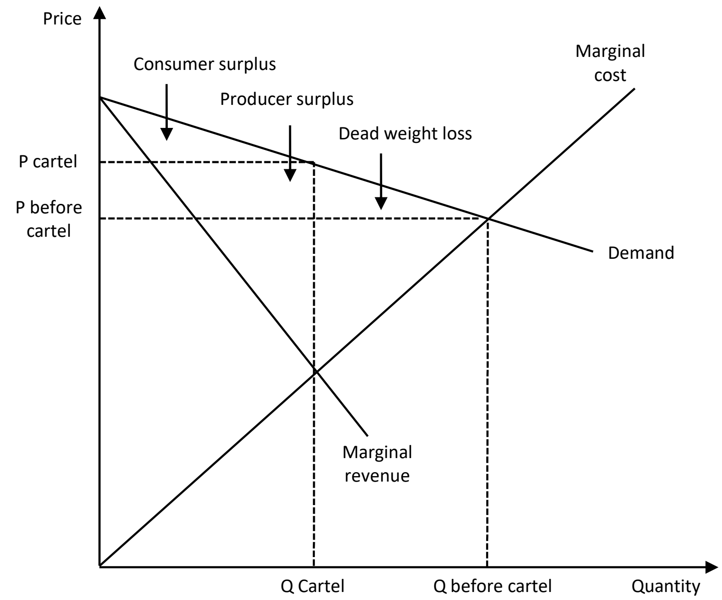
Developed countries devote about 15% to 20% of their GDP to purchasing products and services. On average, this percentage is higher in developing countries because they are potentially more exposed to corruption and cartelization phenomena and are still at the stage of significant spending on infrastructure projects. Considering an average overprice of 20% in bidding cartels, roughly speaking, the State pays an average of 3% to 4% more of its GDP when purchasing products and services in the presence of cartels (Martinez, 2014).

Fighting corruption and cartel had been a challenge also in Brazil, a country of continental dimensions. Its administration has three levels of government: federal, state, and municipal. The country comprises 26 states and the Federal District and at the municipal level by 5,570 municipalities. Furthermore, the country presents a tremendous economic and social disparity. In addition, cartels reduce public confidence in the tender process and the authorities responsible for providing goods and services (Marshall & Marx, 2009; Anderson & Cau, 2011). In this context, the challenges and benefits obtained from possible improvements in the fight against fraud in public procurement are enormous.

Cartels are associated with three types of economic inefficiencies: allocative, productive, and dynamic. The allocative inefficiency is related to the inefficient allocation of social resources due to increased prices and supply restrictions. Figure 2.2 shows that part of a consumer surplus is improperly transferred to the producer in the cartel's presence. The amount not appropriated by either the consumer or the producer destroys social wealth (the so-called "dead weight"). That is why financial penalties that recover the amounts wrongly transferred from the consumer to the producer do not return society to the state it was in the cartel's absence, since part of what would have been appropriated by the consumer has been destroyed for the conduct. Productive inefficiency is related to economic agents operating at higher costs than they would have without a collusive arrangement. In turn, dynamic inefficiency is associated with the loss of social well-being motivated by the reduction of incentives for innovation - the cartel reduces the incentives for market agents to improve their production processes and launch new and better products and services on the market. Besides, cartels' existence increases barriers to entry in a market, making these effects especially perverse (Martinez, 2014).

A cartel is an agreement between competing companies that aims to control prices or not bring a new competitor to the market. The formal organization of sellers or buyers agrees to set sales prices, buy prices, or reduce production with various measures (Bishop & Walker, 2010). The purpose of such a consensus is to increase the profits of individual members by reducing competition.

Figure 2.2: Cartel Dead Weight Loss



Source: Martinez (2014)

The most basic differentiation between cartels is the commercially sensitive variable that is the subject of the conduct. According to this criterion, cartels are classified into price/quantity cartels and market or customer groups (also referred to as market-allocation or market-sharing agreements). In the latter case, the intention is to replicate a monopoly environment in which each economic agent reserves an area (e.g., neighborhood, city, State of the Federation, or even a country) in an international cartel) or a customer profile for its performance. Bidding cartels can be considered a subspecies of market allocation cartels.

Another category of cartel classification refers to diffuse (soft) and classic (hard-core) cartels. The OECD defined traditional cartels in a recommendation to its members issued in 1998 as an anti-competitive agreement, a concerted anti-competitive practice or an anti-competitive arrangement between competitors to fix prices, defraud bids, establish restrictions on offers or quotas, or divide markets by allocating customers, suppliers, geographic area or lines of trade. According to the OECD, classic cartels do not include agreements, practices, or economically efficient arrangements (Martinez, 2014). There are several forms of a cartel, such as:

- Setting minimum or target prices (i.g., price-fixing);
- Reducing total industry output (i.g., fixing market shares);
- Allocating customers and territories;

- Bid-rigging;
- Illegally sharing information;
- Coordination of the conditions of sale; or
- A combination of the above

2.2.1. Cartel in Public Procurement

Cartels do not have a generally accepted definition. Still, they have generally considered instruments for analyzing observable economic data and information and flag markets that may have been affected by collusion or be more susceptible to collusion (OECD, 2013). In known bid-rigging cases, specific models and screens are usually present. OECD checklist provides a non-exhaustive list of areas for active vigilance (Bishop & Walker, 2010). Many papers on collusion exploit nuances of particular auction formats (i.g. average bid auctions, constrained bids) to highlight suspicious patterns (Chassang & Ortner, 2019; Conley & Decarolis, 2016; Kawai & Nakabayashi, 2014). More recently, fine-grained data also enables applying novel methods from network science (Morselli & Ouellet, 2018, Wachs & Kertesz, 2019) and machine learning (Huber & Imhof, 2019; Schwalbe, 2018; Vadász et al., 2016).

The characteristics of collusive behavior in public procurement markets are very similar to that of conventional markets. Companies coordinate their behavior in terms of price, quantity, quality, or geographical presence to raise market prices. Czibik et al. (2014) contend the long-term factors for the occurrence of this type of behavior are:

1. The coordination's ability;
2. Internal sustainability (reliable punishment system and detection of cheating);
3. External sustainability (ability to exclude new market entrants).

Besides, public procurement is more exposed to collusion in the light of the above characteristics than traditional markets. In public procurement, the outcome is determined by auction, implying no quantity adjustment as price changes. Ultimately, this leads to inelastic demand (Czibik et al., 2014).

Bidding (or collision bidding), a type of cartel, occurs when companies are expected to compete, raise prices or reduce the quality of the goods or services acquired through a tender process. Bid rigging can seriously damage public procurement. If the bidders agree to compete unfairly in the procurement process, the public will be denied a fair price. If suppliers cheat when they bid, the taxpayers' money is lost because governments pay more than the reasonable price (Danger & Capobianco, 2009a, b). Such collusion between bidders (winners and losers) is illegal in most countries (Vadász et al., 2016). It is a form of pricing and market allocation. It is often practiced when contracts are determined through competitive tendering, such as contracts for government buildings, the provision of information and communication technology (ICT) (Danger & Capobianco, 2009a).

Considering an average overprice of 20% in bidding cartels, the State pays an average of 3% to 4% more of its GDP when purchasing products and services in the presence of cartels. There is a peculiarity in this type of cartel: depending on the bidding format, there is no loss of well-being in the strict sense. The loss of well-being, also referred to as dead weight or static inefficiency, occurs when the consumer of goods or services fails to purchase the desired quantity due to the price increase resulting from market power. In many competitions, the bidding entity sets the purchased amount of the good or service, assuming a given price range. In such cases, as an eventual cartel's effect does not reduce the amount consumed, there is no dead weight. There is only one transfer of income from the bidding entity to colluding bidders. However, the absence of a direct loss of well-being does not make the effects of such a cartel necessarily less severe, especially concerning those who reach the public power. The transfer of income from the public authorities to the cartel agents indirectly decreases public services' quantity and quality. The grievous effects of cartels in tenders and the fact that they are often associated with corruption episodes make some countries criminalize only this type of collusive arrangement. Bidding cartels can cross-dress in a variety of ways, notably the allotment and the rotation. In the allotment, the cartels members divide the entities that promote bids or the items of the same bid (Martinez, 2014).

Some common characteristics help maintain collusion, such as a small number of companies; the smaller the number of suppliers, the easier it will be to agree on the proposals. The low or zero levels of market entry is another factor, when companies have recently entered the market or when it is unlikely to happen because market entry is not a simple process, companies that are already operating in the market are protected from competitive pressure potential new competitors, keeping cartels going (OECD, 2009). Market conditions are essential, as significant changes in demand or supply conditions tend to destabilize ongoing collusion agreements. A steady and predictable flow of demand from the public sector tends to increase the risk of collusion. Frequent bids help members of a cartel agreement to distribute contracts among themselves. The cartel members can also punish the company that does not comply with the agreement by staying with the proposals initially assigned (OECD, 2009).

A vital feature highlighting a possible cartel's existence is the low capacity utilization rate, producing less than they can and selling the product at higher prices. Participants combine this quantity produced and the products' prices (Tatsch, 2012). The cartel formation also becomes more common when few companies offer the same product or service because it is easier to monitor and inspect the possible awareness of all the entrepreneurs who supply the product. It is also less likely that some participants do not comply with the agreements in cartels, thus guaranteeing greater chances of success for the cartel (Carvalho, 2018). A relevant characteristic of a cartel is the barriers to the entry of new companies in the sector. The cartel, by nature, is unstable. If new companies access to the same product is constant or growing, it will be challenging to maintain that cartel (Carvalho, 2018).

The homogeneity of the products or services also facilitates the constitution and maintenance of the cartel. Having products with similar quality and production costs in a price-fixing will simplify defining a base value. However, if the products or services are heterogeneous, it does not mean that there is no cartel, as there are alternatives for this type of situation (Carvalho, 2018). When there is little or no technology on the market that produces a particular product or performs a service, or its technologies are already consolidated, with the application of existing technology being shared, it is easier for the cartel to appear and coordinate, considering that, in these cases, generally the products will be homogeneous (Carvalho, 2018).

Cartel identification is not a simple task to carry out. It can include many companies, and customers can rarely detect a cartel, especially in public bodies. The number of employees destined for that type of activity is usually not enough (Silva, 2020). As tenders involve public money, it is necessary to assist and supervise the formation of cartels by various available means and instruments.

2.2.2. Cartel Patterns

Various methods of bid-rigging can be models that can recognize illegal activities. Czibik et al. (2014) classify collusive market behavior according to three dimensions: methods of distorting competition or basic collusion techniques, forms of rent-sharing, and market structure. The indicators below, in combination with other registers, suggest a suspicious cartel case (Danger & Capobianco, 2009b; Hölzer 2014; Vadász et al., 2016):

- Winner steps back. The odds of falsifying bets are high if the winner withdraws his bet and lets the second-best (and much more expensive) participant get the deal.
- Invalid bids. Some or all bidders submit invalid bids, so the agreed entity wins.
- High prices. Most proposals are higher than previous procedures, above company list prices, or a reasonable cost estimate.
- Coordinated high prices. All but one of the offers is incredibly high and far exceeds the market standard.
- Market coordination based on geographical distribution. A company that, for some reason, such as a cost difference, offers significantly higher prices in several regions than others may indicate that it cooperates in tenders.
- Similar suspicious prices. If several companies have submitted bids with the same or seemingly strange prices, this may mean that they have agreed to split the contract.
- Suspected boycott. If an offer is not accepted, a coordinated boycott can affect the terms of the contract.
- The number of suspicious offers. Too many companies do not apply; he can represent a cartel with market sharing.
- Similar suspicious offers. If bids relate to industry agreements that affect price, companies may have agreed to, for example, apply shared price lists, penalties for late payments, or other terms of sale for the industry.

- Suspected subcontractor arrangements. If the winning company awards or transfers part of the contract to a competitor who submitted a more expensive bid in the same process, this may indicate a collusion cartel.
- Suspected joint tenders. A joint application from more companies than is necessary to complete the task may be illegal.
- Cyclical winning. A series of auctions were held in agreement with various participants.

Research publications and literature assume that individual companies or collectives belonging to the interest group that commits cartel fraud are independent and remain static. This is a misconception when individuals, companies, and organizations enter into less evident and convincing relationships based on procurement data. Other cartels suspicious indication can come from:

- The relationship of dependence on the participation of two companies in the same bidding processes (Castro et al., 2018).
- Purchases with total competitors are below average for that type of object (Tas, 2017).
- Perennial losing bidders: companies that give legitimate competitors the appearance without the real intention of receiving the award (Herrera Murillo, 2019).

Vadász et al. (2016) note that the study of hidden insight requires in-depth research from several sources other than the official public procurement bulletin. An obvious indication of an investigation may, for example, be where registers show that the same person owns several companies. In this case, the probability of violation is high. Vadász et al. (2016) present several typical scenarios that should be analyzed:

- Owners seem to be different, but one is related to the other.
- The owner's wife is another owner but is listed by her maiden name.
- Owners of different companies, but with the same address in an offshore trust.
- Companies do not appear to be related, but the same auditor or web pages register with the same supplier if they use the same bank branch.
- Common CEOs, board members, or other senior executives.
- Person related to the actual owner.

2.3. Machine Learning

The role of technology in facilitating public resources management, leveraging the capacity for analysis by control bodies, and providing greater transparency and participation of society is urgent. This section aims to present an overview, not exhaustive, of efforts in several countries to identify fraud and fight corruption and the basic machine learning concepts.

Machine learning is a subfield of Artificial Intelligence and designs algorithms that iteratively learn from data and experience (OECD, 2017b). Those algorithms allow computers to learn without the need to be explicitly programmed (Samuel, 1959). According to Anitha et al. (2014), machine learning algorithms can be classified into three broad categories, depending on their learning pattern. Supervised learning algorithms use a sample of labeled data to learn a general rule that maps inputs to outputs. Unsupervised learning algorithms aim to identify hidden structures and patterns from unlabeled data. Finally, reinforcement learning algorithms perform tasks in a dynamic environment and learn through trial and error.

Such algorithms have been used by government agencies and the private sector (Buchanan & Miller, 2017). One of the government's challenges is to signal collusion practices when companies use learning algorithms to move business decisions from humans to computers. Doing so, "managers do not only avoid any explicit communication during the initiation and implementation stages of collusion but are also released from the burden of creating any structures" (OECD, 2007c, p. 29). Furthermore, Mehra (2015) highlights that algorithms can amplify the so-called "oligopoly problem" and make tacit collusion an even more frequent market outcome.

Risk assessment often goes beyond manual analysis because there are often too many data sets and variables, even for experienced analysts. Data mining and analysis tools by creating large open groups and creating clever crossword puzzles can also help identify risky transactions. These tools make it possible to develop risk maps to gain insight into illegal financial flows and complex corruption networks (Argüello & Ziff, 2019). Fraud detection is an opportunity that companies, and governments can use to use their financial resources to save money efficiently. The main area of knowledge that has been intensively used to improve fraud detection is machine learning.

Machine learning is a field of knowledge that intersects with computer science, statistics, neuroscience, and other subjects. It consists of the process of teaching computers to learn. However, unlike traditional software development, machine learning is about "programming computers to teach themselves from data rather than instructing them to perform certain tasks in certain ways" (Buchanan & Miller, 2017, p. 5). The growing interest in using machine learning systems for decision-making in the public sector has raised questions about how these technologies can be designed, implemented, and managed responsibly (Veale, 2017). Machine learning can algorithmically model large amounts of data that are difficult for humans to handle. These models can then be used for tasks such as forecasts or data structuring. Following its successful application in business and several major public exhibitions, there has been significant interest in its application to solve public problems. In this context, many research gaps in the supply chain arise related to quality control in manufacturing, order picking and inventory control systems in warehousing, and demand shaping (Nguyen et al., 2018). Also, algorithmic methods provide potent opportunities to encourage discoveries about social phenomena by increasing access to large amounts of data or collections of mixed qualitative/quantitative data (Miller, 2019).

Most statistical learning problems can be divided into supervised and unsupervised (James et al., 2017). Supervised learning tries to adapt the response model to predictors to predict future observations (predictions) correctly or better understand the relationship between the response and the predictors (inference). For each observation of the predicted measure x_i, $i = 1, ..., n$, there is a corresponding response measurement y_i. Many classical statistical learning methods, such as linear regression, logistic regression, generalized additive models (GAMs), boosting, and support vector machines, work with supervised learning (James et al., 2017).

Vadász et al. (2016) argue that supervised learning can only detect a small fraction of fraudulent events, especially in the context of complex procedures for large modern companies. A supervised model is created based on several positive and negative scenarios that it provides. As a general critique of supervised fraud detection, the System can only detect known episodes of fraud. However, supervised machine learning has compelling and measurable methods, leading to an explosion of ever-improving methods.

James et al. (2017) clarify that unsupervised learning describes a slightly more complicated situation at a vector of xi for each observation $i = 1, ..., n$, but without the corresponding response y_i. The linear regression model cannot be adapted because there is no response variable to predict. In this context, it seems to be working blindly; the case is referred to as unsupervised because there are no response variables that can guide our analysis (James et al., 2017). Das (2016) defines unsupervised learning as the process of reorganizing and improving input to structure unlabeled data. Cluster analysis is a good example, which creates a set of objects with some attributes and divides the object space into sets or groups based on all objects' proximity attributes. It reorganizes and labels data with additional tags (in this case, the cluster number/name). Another unsupervised learning method is factor analysis.

In unsupervised learning, all events that deviate from standard activity patterns can be considered deceptive (Vadász et al., 2016). The main advantage is that it does not require a predefined set of known fraudulent events and can detect previously unknown malicious activity types. Unsupervised learning is a branch of artificial intelligence that can assimilate and grow its knowledge base without a predefined set of flagged events. Unsupervised methods include anomaly detection, a method class that examines all available data and identifies outliers who deviate from the dataset (Vadász et al., 2016).

Despite promises to detect unknown types of fraud, unsupervised methods have two main disadvantages limiting their use in practice (Vadász et al., 2016). First, it isn't easy to assess their effectiveness if labeling data are not available. In this case, there is no information to compare the performance of the two models. Second, the concept of an outlier cannot be precisely defined. A higher contract value or area with few procurement procedures can be considered an exception but not fraudulent.

Machine learning research has focused primarily on supervised models that enable people to interact by creating learning scenarios. "Active" learning (Tong & Koller, 2001) is a promising path where everyone in the cycle can help to gradually improve their classification by providing the most critical labeling options. However, sometimes the question of whether the analysis should be supervised or unsupervised is less clear (James et al., 2017). Suppose a set of n observations and for m observations $(m < n)$ with predictor and response measures. There are available predictor measurements for the rest of the observations ($n - m$) but no response measurement. In this framework, a statistical learning method can integrate observations for which response measurements are available and the $n - m$ observations for which there are none. Semi-supervised machine learning methods can address this situation.

2.3.1. Machine Learning and statistics to signal fraud and collusion

It is generally accepted in the field literature the use of various types of information to develop a method of detecting bid-rigging. The forms can be reactive or proactive (Vadász et al., 2016). Response techniques that lead to an investigation are an approach based on external incidents, such as receiving a complaint (from a competitor, client, agency, or employee), possession of external information (whistleblowers or informants), or information obtained from leniency applicants (Anti-cartel Enforcement Manual, 2015; Hüschelrath, 2010). The authorities responsible for protecting open and fair competition integrity heavily depend on response methods, especially leniency programs (Vadász et al., 2016). Relying on one single tool or technique is unwise and insufficient. The second proactive approach is to identify the nature of the initiative. Several proactive methods have been presented in the literature, including the analysis of previous cartel and competition scenarios; marketing, industry, press, and online reporting; cooperation with antitrust and other national and foreign investigative bodies; and quantitative units to identify suspicious patterns, among other strategic displays (Abrantes-Metz, 2013; Anti-cartel Enforcement Manual, 2015; Harrington, 2006; Hüschelrath, 2010; Hüschelrath & Veith, 2011; OECD, 2013).

Vadász et al. (2016) explain that screening methods can be divided into structural and behavioral groups. The first is structured based on industry data and identifies markets where a cartel is likely to be formed. On the other hand, behavioral methods use data on companies' behavior or actions, which indicates the formation of a cartel. Behavioral screens are usually based on publicly available data sets, such as public procurement data (Anti-Cartel Handbook, 2015; Harrington, 2006).

One way to identify cartels is through quantitative analysis by calculating complex indicators. These indicators can trigger and signal many types of collision behavior and, when used in conjunction with traditional investigative methods, they can use limited resources better (Vadász et al., 2016). Many publications (Abrantes-Metz, 2013; Harrington, 2006; Tóth et al., 2015; Wensinck et al., 2013) discussed those indicators to find different possible solutions. Toth et al. (2015) developed initial and complex indicators to sign collusion in public procurement. They identified and categorized the main types of collusive bidding and showed how to adapt those indicators to different market realities.

The approach to identifying and testing for bid-rigging in procurement auctions developed by Bajari and Ye (2003) is considered the most effective for detecting collusion (Signor et al., 2021). Their method produces a basic scenario based on the structural cost parameters for bidders provided by industry experts. In the basic case, the bids that competing participants must submit are described. Consequently, significant and systematic deviations from the base scenario may indicate collusion. Besides, Bajari and Ye (2003) tested their model in a set of highway repair contracts that included tender information for almost all public and private road construction projects in the Midwestern United States from 1994 to 1998. Using the reduced regression function with linear regression, they were able to identify the presence of a cartel and stressed the importance of the pre-tender estimate (PTE) as the closest price estimate possible to the market price.

Because the cartel, like all corruption, is carried out in secret, automatic detection of information can be crucial for law enforcement. Vadász et al. (2016) propose an automated text and data analysis method to save time and professional costs by allowing investigators to identify suspicious behaviors and encourage more personal investigation. These authors argue that open-source information is publicly available by definition. The availability of relevant information resources, whether paid or free, varies from country to country. However, even so-called "free" content may be subject to restrictions through mandatory identification and registration, password protection, or mandatory human challenge-response tests (e.g., "CAPTCHA") that prevent access and automatic indexing or back-distribution robot.

Huber and Imhof (2019) combine machine learning techniques with statistical screens calculated from the distribution of bids in the Swiss construction sector to predict collusion. The dataset contained 483 auctions comprising 300 collusive and 183 competitive (post-collusion).

The scholars evaluate the out-of-sample performance and estimate that it classifies over 80% of all tender processes as collision or non-collusion. They also examine trade-offs to reduce false positives versus false negatives.

Signor et al. (2020) developed a probabilistic method to signal collusion based on the Brazilian Federal Police experience on Operation Car Wash. They analyzed all bids from a single auction and identified abnormal dispersions and bid locations. Signor et al. (2021) later refined the collision detection method by calculating the percentage of order statistics with the lowest bid in the 101 auctions in the Car Wash Operation dataset. The scholars used actual auction data from the Car Wash Operation to validate their model. The data set contained 101 uncapped auctions of the Brazilian oil company between 2002 and 2013. They concluded that the lowest prices with a very high percentile (> 90%) are more likely to collude. Besides, type I and II errors are significant in the amount of data analyzed.

Extracting relevant data from publicly available data sets is an essential step in identifying cartels. Information on public procurement is mainly available in structured or semi-structured formats. Therefore, data extraction must be performed with specially programmed solutions to make an assessment (Czibik et al., 2014; Fóra, 2014). This topic's future should be based on the connected, open data paradigm, where competent authorities publish their available public data in a more convenient format (Vojtěch et al., 2014).

We argue that all of these strategies to signal collusion, though valuable, may tend to overfit their methods to specific cases. A foundation of modern machine learning practice is the evaluation of predictive algorithms on unobserved data. The scarcity of clean data on proven cartel cases makes applying cartel screening methods to multiple examples difficult. Exceptions are the works of Czibik et al. (2014) and Huber et al. (2020). For instance, ensembling or combining multiple cartel screens in a suite of indicators opens an opportunity to widen the scope of cartel screening to large and heterogeneous markets with varying data quality.

3. DETECTING FRAUD IN PRICE REGISTRATION SYSTEMS (PRS)

3.1. Research Question

The distortion of the purposes pursued by the PRS and the non-observance of mandatory requirements for the viability of non-participating bodies made TCU realize the need to reassess the rules established in Decree 7,892/2013, considering the limits adherence to the price registration drafts. However, the PRS does not reject the possibility of fraud or, at the very least, suspicious behavior aimed at serving the interests of the contracted companies and not the public interest. One example is the sale of price registration drafts, which uses the same drafts, or very similar drafts, by different management bodies. There may also be an effort by suppliers to convince management to make large contracts. Another abnormal behavior is the rent of price registration drafts. This occurs when the management body authorizes many non-participating bodies and contracts a negligible amount concerning the total amount of the respective price registration drafts. This practice can discourage the proper planning of acquisitions of the agencies and open the possibility of illicit acts, especially when the management body authorizes the Non-participating body (NPB) and when the supplier accepts NPB adhesion. To date, no study in the literature quantifies the existence of rent of price registration drafts or indicates the degree of risk of this type of behavior.

The contracted amount for 14,635 current price registry drafts is R$ 3.7 billion (data from the Government Purchasing Panel obtained on 03/06/2021[7]). There is no information on the estimated number of the rent of price registration drafts. It should be noted that Comprasnet does not hold the universe of price registration drafts, as it does not contain a relevant number of documents from States and municipalities. The rent of price registration drafts can generate possibilities of fraud and substantial damages to the treasury. Although there is a perception by the control bodies of the relevance of this problem, the subject is relatively new. It requires investigation to better understand the problem and prioritize human, financial, and technological resources for TCU's activities. In this sense, this study investigates whether patterns differentiate the rent of price registration drafts minutes. Thus, we consider the following research questions:

[7] http://paineldecompras.economia.gov.br/ata-precos

- How are latent profiles defined to represent behavior in the Price Registration minutes based on a set of selected variables?
- Is there a systematic way of signaling price registration drafts with the potential to be used as a rent instrument? If so, what factors would most characterize those rented price registration drafts?

Therefore, we will investigate the following hypotheses:

- H1: Some attributes signal price registration drafts with the potential to be rented.
- H2: It is possible to identify clusters of rented PR drafts significantly different from regular price registration drafts.

3.2. Methodology

We borrow the approaches used to detect fraud and collusion when there are few or no tagged fraud cases. In those situations, when we do not have a priori information about the investigated behavior, screening methods and unsupervised machine learning approaches could be used in the analysis.

In general, fraudulent bid detection contains two main elements: structural screening (which deals with market and trade features) and behavioral screening (which deals with issues related to tenderers' behavior) (Busu & Busu, 2021). The role of structural screening is to act as a filter that can identify these proposals in sectors or areas that are more likely to collude. Behavioral screening is performed ex-post during an auction to examine certain aspects that may indicate agreements between the auction participants (Busu & Busu, 2021). Thus, behavioral screening can capture specific behavioral patterns of bidders. In our case, behavioral screening can be applied to suppliers and public agencies that adhere to the price registration drafts. Unlike structural screening, behavioral screening can lead to identifying tenders with a high probability of fraud. Initially, the indicators considered may trigger initial investigations and, subsequently, even evidence. One type of fraud analysis in price registration that still requires a more detailed investigation comprises the rent of drafts to non-participating bodies (NPB).

Thus, it is assumed that there are different clusters of registration price minutes based on various features. The presumption is that one of them (or maybe even two or three) are potentially fraudulent ones. Our study considers fraudulent the price registration minutes that were loaned to non-participating bodies. Different techniques can be used to composition the weights for each candidate attribute, including feature engineering. The detection of possible rented minutes will be carried out in the following steps:

1. Performing a cluster analysis. This will be done by dividing the data into many groups based on the joint analysis of the candidate attributes. The number of groups is typically determined empirically; however, choosing the adequate clustering method in heterogeneous data sets is challenging. We applied Latent Class Analysis to our investigation.

3.3. Data

Data for this study comes from all the pooled procurements performed in 2019 with at least one NPB. There are 5,554 unique tenders in the database refereed as price registries, with 2,288 distinct General Service Administrative Unit (UASG), the General Service Management System (SIASG), established by Decree 1,094/1994. In practice, each UASG receives goods and services that are auctioned off via Comprasnet. The data also has 59,264 different products, 566 different services, and 9,543 suppliers.

We used information from 10 tables from Comprasnet, which names are presented below (Table 3.1). TCU provided the tables, and we considered just tenders in with there are at least one NPB. After data cleaning, the result is a table with 1,458,938 observations, in which each observation corresponds to a UASG per tender per item. The large sample provides some advantages, as statistical power. However, it also has the drawback of needing computational power for the estimations.

The initial strategy would be to use all continuous variables in the LPA model estimation. Usually, a model with continuous variables provides a more detailed analysis than models that use only binary variables. However, estimating models using continuous variables takes longer, and this was confirmed in the present study. Our estimations using just continuous variables became too time-consuming due to our limited computational capacity[8]. In this way, we initially used a model with only binary variables. We then replaced some binary variables in the model

[8] 3.1 GHz Quad-Core Intel Core i7, 16 GB 2133 MHz LPDDR3

with continuous variables, selected through attribute selection and substantive analysis with the TCU experts.

Table 3.1: Comprasnet tables used in the research

Dataframe Name	Content Description
ATASRP_FORNECEDOR	Suppliers of an item
ATASRP_ITEMATASRP	Items in the minutes
ATASRP_ITEMSOLICITACAOADESAO	Requests from non-participating bodies authorized
ATASRP_LICITACAOSRP	Information of the Price Registry System
ATASRP_OCORRENCIAITEMSOLADESAO	Items' information
ATASRP_OCORRENCIASOLICITACAOADESAO	Items' information
ATASRP_RESUMOFORNECEDORITEM	Item and supplier information
ATASRP_RESUMOUASGITEM	A view of acquisitions (items per UASG)
ATASRP_SOLICITACAOADESAO	Quantity requests per item
IRP_IRPITEM	Items' information

Source: Elaborated by the author

Latent Class Analysis (LCA) and Latent Profile Analysis (LPA) reconstruct hidden groups from observational data. They are a type of mixture model (Oberski, 2016). Whether in individual classes or personal profiles, these groups can be helpful when considered as unobserved latent mixture components (Ferguson et al., 2020). LPA attempts to restore hidden groups based on continuously observed variables. LCA does the same for categorical variables. Therefore, the more general term is LCA, and both methods assume no correlation between variables that once depended on group membership (Kubrin et al., 2021). Applying LCA using the candidate variables was extensively time-consuming, even using Mplus. So, I created item-level binary indicators for simplicity (i.e., reduce complexity and increase model convergence). Table 3.2 presents the descriptive statistics of the variables used in this research. A compelling feature is that LCA provides a log-likelihood solution for calculating Bayesian Information Criterion (BIC) values. Researchers can choose the best model with the lowest BIC value. BIC offers penalties for additional parameters (Kubrin et al., 2021). Second, the challenge of clustering algorithms is the arbitrariness inherent in the strategy, which begins with a random selection of seed cases to form the cluster (Steinley, 2003). With the LCA approach, selecting an "initial case" is unnecessary to avoid this problem (Kubrin et al., 2021). Ultimately, the challenge is getting the best solution locally, not globally (Hipp & Bauer, 2006). This step can be performed by estimating the model several times with different random starting values. This approach is now software-automated (Muthen & Muthen, 1998–2017). Finally, the LCA approach is conceptually attractive because it empirically determines potential classes with similarities based on the measure of interest (Kubrin et al., 2021).

Table 3.2: Candidate Variables Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
number of NPB per tender	1,458,938	24.235	19.431	0	83
proportion of approved quantity per item per NPB	1,458,938	0.094	2.298	0	1700
proportion of the approved quantity per item for NPBs per tender	1,458,938	1.463	8.551	0	1700
Item relative value	1,458,938	0.021	0.070	0	1
tender without MB	1,458,938	0.005	0.070	0	1
tender with only MB and NPB	1,458,938	0.330	0.470	0	1
tender without quantity to MB and with NPB	1,458,938	0.004	0.066	0	1
amount of NPB per supplier	1,458,938	131.012	153.516	0	948
amount of NPB per supplier per tender	1,458,938	16.327	18.295	0	83
number of tenders per auctioneer	1,458,938	8.840	10.865	1	77
tender amount related to total commitment	1,458,938	2.283	7.644	0	171.68
total approved amount related to total commitment	1,458,938	8.596	18.197	0	171.68
distribution of winners	1,458,938	0.202	0.175	0.001	2
f1: number of NPB per tender	1,458,938	0.023	0.149	0	1
f2: proportion of approved quantity per item per NPB	1,458,938	0.047	0.212	0	1
f3: proportion of the approved quantity per item for NPBs per tender	1,458,938	0.099	0.298	0	1
f4: Item relative value	1,458,938	0.131	0.337	0	1
f5: tender without MB	1,458,938	0.005	0.070	0	1
f6: tender with only MB and NPB	1,458,938	0.330	0.470	0	1
f7: tender without quantity to MB and with NPB	1,458,938	0.004	0.066	0	1
f8: amount of NPB per supplier	1,458,938	0.045	0.206	0	1
f9: amount of NPB per supplier per tender	1,458,938	0.053	0.224	0	1
f10: number of tenders per auctioneer	1,458,938	0.081	0.272	0	1
f11: tender amount related to total commitment	1,458,938	0.108	0.311	0	1
f12: total approved amount related to total commitment	1,458,938	0.158	0.364	0	1
f13: distribution of winners	1,458,938	0.046	0.210	0	1

The distinction between LCA and LPA is that the groups are defined based on the observed variables. In LCA, the observed variables are discrete and like the binomial model (Masyn, 2013, and Nylund-Gibson & Choi, 2018). In LPA, the observed variables are continuous and select the

Gaussian model (Oberski, 2016). LCA is commonly used when it is necessary to identify latent classes from a sample, or a "gold rule" of classifying people is not readily available. The classes in LCA are latent (unobserved) and are established on the individuals' responses to the indicators. Accordingly, LCA is a "person-centered" approach to creating empirically derived typologies, differing from the dominant "variable-centered" tradition that typically requires arbitrary cutoffs for classifying among individual cases (Nylund, Bellmore, Nishina, & Graham, 2007). Our study is coherent to apply LPA. However, instead of analyzing persons, we focus on analyze public entities. Besides, there is no identified rule to classify the fraud of selling participation in Price Registration Minutes.

As the primary inferential analysis method for the study, LCA was conducted using Mplus 8 (Muthén & Muthén, 1998-2017) with maximum likelihood estimation. LCA was used to explore latent profiles of public agencies. The research question used to inform the study was: how are latent profiles defined to represent behavior in the Price Registration minutes based on a set of selected variables?

Steps of LCA

In our study, we intend to classify the public institutions and not the variables. There are general steps in the LCA process that we will detail in this study. Step 1, as with all analyses, the data should be cleaned for analysis and checked for standard statistical assumptions (Osborne, 2012). Because the raw variables are skewed, with many zero values, I created binary indicators presented in Table 3.3.

It is worth explaining the implications of variable *f6* because it is customary to exist tenders with only MB and NPBs. A typical example of this case is government purchasing centers. The Ministry of Economy has a purchasing center that carries out centralized bids for the entire Federal Government, providing economies of scale and savings in acquiring products and services. For instance, government purchasing centers a licit and should be removed from the analysis of suspicious behaviors.

Table 3.4 presents additional statistics for the binary variables. It is possible to identify that most variables are skewed, implicating problems in estimating the models. However, we will discuss an alternative to this limitation.

Table 3.3: Item-level Binary Indicators

Variable	Variable	Meaning	Intuition
f1	number of non-participating bodies (NPB) per tender	1 if upper inner fence (UIF) and 0 otherwise	bids with many NPBs are more likely to have sale fraud from adherence to price registration drafts. Bids without NPB do not have this type of fraud.
f2	proportion of approved quantity per item per NPB	1 if upper inner fence (UIF) and 0 otherwise	amount cannot exceed 50% according to art 22 §3 of decree 7892. Tenders with a high amount approved for NPB have a greater probability of fraud
f3	proportion of the approved quantity per item for NPBs per tender	1 if upper inner fence (UIF) and 0 otherwise	amount cannot be more than double (2) by art 22 §4 of decree 7892
f4	Item relative value	1 if upper inner fence (UIF) and 0 otherwise	ratio of the item value and the tender value. It can be treated as a proxy for how expensive the item became. Higher values indicate higher collusion risks
f5	tender without managing body (MB)	1 if tender without managing body (MB) and 0 otherwise	tenders without managing body (MB) have higher chances of fraud
f6	tender with only MB and NPB	1 if tender with only MB and NPB and 0 otherwise	tenders with only MB and NPB have higher chances of fraud
f7	tender without quantity to MB and with NPB	1 if tender without quantity to MB and with NPB and 0 otherwise	tenders without quantity to MB and with NPB have higher chances of fraud
f8	amount of NPB per supplier	1 if upper inner fence (UIF) and 0 otherwise	suppliers with many NPBs are more likely to have sale fraud from adherence to price registration drafts.
f9	amount of NPB per supplier per tender	1 if upper inner fence (UIF) and 0 otherwise	suppliers with many NPBs are more likely to have sale fraud from adherence to price registration drafts.
f10	number of tenders per auctioneer	1 if upper inner fence (UIF) and 0 otherwise	greater number of tenders per auctioneer could signal an outlier
f11	tender amount related to total commitment	1 if upper inner fence (UIF) and 0 otherwise	high value could signal an outlier
f12	total approved amount related to total commitment	1 if upper inner fence (UIF) and 0 otherwise	high value could signal an outlier
f13	distribution of winners	1 if upper inner fence (UIF) and 0 otherwise	characterizes the distribution of items among suppliers. The smaller the relationship between the number of different suppliers (*nV*) and the number of items (*nI*) in a tender, the greater the market concentration and the possibility of favoring a given supplier.

Table 3.4: Item-level Binary Indicators Descriptive Statistics

Variable/ Sample Size	Mean/ Variance	Skewness/ Kurtosis	Minimum/ Maximum	Median
f1: number of NPB per tender	0.023	6.417	0	0
1458938	0.022	39.175	1	
f2: proportion of approved quantity per item per NPB	0.047	4.26	0	0
1458938	0.045	16.15	1	
f3: proportion of the approved quantity per item for NPBs per tender	0.099	2.69	0	0
1458938	0.089	5.236	1	
f4: item relative value	0.131	2.188	0	0
1458938	0.114	2.789	1	
f5: tender without MB	0.005	14.186	0	0
1458938	0.005	199.252	1	
f6: tender with only MB and NPB	0.33	0.723	0	0
1458938	0.221	-1.478	1	
f7: tender without quantity to MB and with NPB	0.004	15.096	0	0
1458938	0.004	225.886	1	
f8: amount of NPB per supplier	0.045	4.411	0	0
1458938	0.043	17.455	1	
f9: amount of NPB per supplier per tender	0.053	4	0	0
1458938	0.05	14	1	
f10: number of tenders per auctioneer	0.081	3.08	0	0
1458938	0.074	7.486	1	
f11: tender amount related to total commitment	0.108	2.519	0	0
1458938	0.097	4.346	1	
f12: total approved amount related to total commitment	0.158	1.879	0	0
1458938	0.133	1.532	1	
f13: distribution of winners	0.046	4.315	0	0
1458938	0.044	16.617	1	

Step 2 evaluates several hypothetical iterative LCA models, starting with a model with one latent class and usually ending with a model that estimates an optimum number of latent classes. In this study, we used the LCA in the research question to identify potential classes that represent the behavior of renting price registration drafts. Model 1 was estimated with only one latent class, model 2 with two latent classes, up to model 18 with eighteen profiles. The variables f5, f7, f11, and f12 are highly correlated with other variables[9]. So, I excluded f7 and f12 from the analysis.

Step 3 involves evaluating models to identify model fit and interpretability. The challenge for LPA is determining the number of profiles in the data (Morin et al., 2015). Hayton et al. (2004) emphasize that too few factors can distort factor loadings, leading to a combined solution of common factors, obscuring the correct factor's solution. If you extract too many factors, you can focus on being small, unimportant, challenging to interpret, and unlikely to duplicate factors. A delicate balance between integrity and savings is needed to reach an interpretable solution.

Statistical tests and indicators are available to facilitate this decision (McLachlan & Peel, 2000): Akaike's Information Criterion (AIC). Bayesian Information Criterion (BIC); Consistent AIC (CAIC); Sample Adjusted BIC (SABIC); Likelihood Ratio Test (LMR) by Lo, Mendell, and Rubin (2001); and bootstrap likelihood ratio test (BLRT). Overall, the lower the AIC, CAIC, BIC, and SABIC values, the better the model fit. Both LMR and BLRT are tests that compare the k-profile model with the *k-1* profile model. A significant p-value indicates that the *k-1* profile model should be rejected, and the *k* profile model should be preferred. Simulation studies show that CAIC, BIC, SABIC, and BLRT are particularly effective in choosing a model that best reflects the true parameters of the sample, but AIC does not and tends to over-extract (Henson, Reise, & Kim, 2007; McLachlan & Peel, 2000; Nylund, Asparouhov, & Muthén, 2007; Peugh & Fan, 2013; Tein, Coxe, & Cham, 2013; Tofighi & Enders, 2008; Tolvanen, 2007). For this reason, AIC has been reported only to ensure full disclosure of results and has not been used in this study. Morin et al. (2015) found that BIC and CAIC tend to underestimate the true number of profiles and SABIC and BLRT tend to overestimate them if the recommended indicators do not identify the optimal model.

Since all these tests are variations of tests of statistical significance, the class enumeration process can be significantly affected by sample size (Marsh et al., 2009). That is, even when given a sufficiently large sample, these indicators add latent profiles to the model without reaching the minimum point (CAIC, BIC, SABIC) or statistical significance (LMR, BLRT). In these cases, the information criterion should be graphically represented by an "elbow chart" showing the gains

[9] See Appendix for Correlation Table

associated with the additional profile (Morin, Maiano, et al., 2011; Morin & Marsh, 2015; Petras & Masyn, 2010).

In these graphs, the point where the slope flattens indicates the optimal number of profiles in the data. Two additional critical criteria used in this decision are (a) content-related meaning and theoretical agreement of the profile (Marsh et al., 2009; B. O. Muthén, 2003) and (b) statistical solution. It's validity. (Example: Negative variance estimates; Bauer & Curran, 2004). Finally, I also examined entropy, which should not be used to determine the optimal number of profiles (Lubke & Muthén, 2007), but it does provide a valuable summary of classification accuracy. The entropy varies from 0 to 1, with higher values indicating fewer classification errors.

3.4. Results

To identify the optimal number of classes for our data empirically, we estimated a series of LCA models with sequentially more significant numbers of specified classes (e.g., three class solution, four class, five class, etc.) and compared model BIC values, as well as the substantive characteristics of the solution. The seventeen-class solution provided the optimal fit for our data. It had a low BIC value and a high entropy value (implying the model is relatively good at placing each unit into a particular latent class). AIC and SABIC values also provide evidence of a good fit. However, the Lo-Mendell Ruben (LMR) and bootstrap likelihood ratio (BLRT) tests indicate different. Table 3.5 shows the model fit statistics.

Currently, applied researchers use a combination of criteria to guide their decisions on the number of classes in mixture modeling (Nylund et al., 2007). These criteria include a variety of statistical information criteria (IC) such as AIC (Akaike, 1987) and BIC (Schwartz, 1978) in conjunction with substantive theory. Nylund et al. (2007) state that scholars suggest the adjusted BIC (Sclove, 1987) is superior to other IC statistics for LCA models, and several scholars recommend using BIC as a good indicator for class enumeration over the rest (Collins, Fidler, Wugalter & Long, 1993; Hagenaars & McCutcheon, 2002; Magidson & Vermunt, 2004). For those reasons, we will consider BIC as the decision parameter.

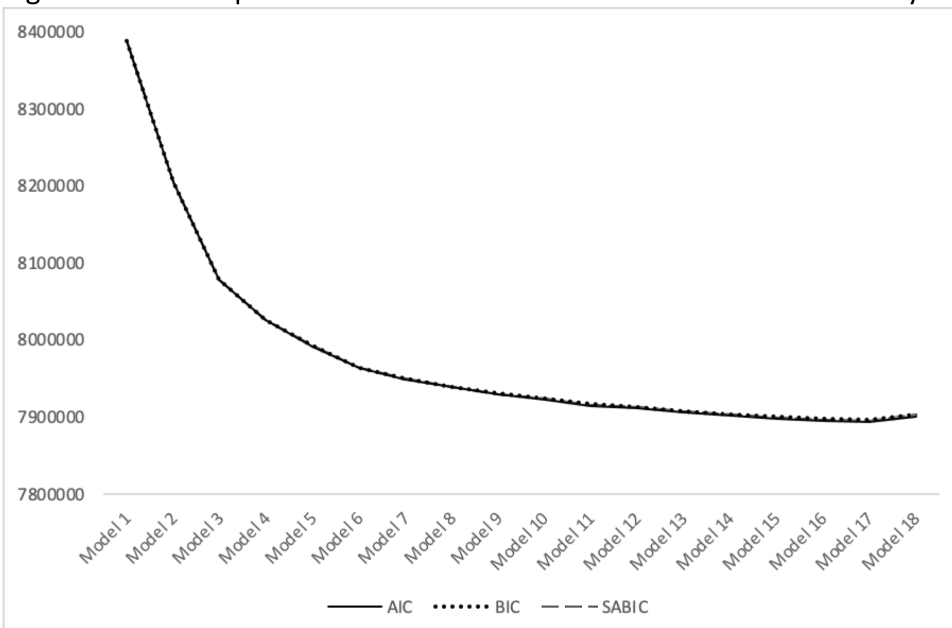
Table 3.5: LCA Model Fit Summary for Research Question

Model	Log-likelihood	AIC	BIC	SABIC	Entropy	Smallest class
1	-4194118.8	8388259.63	8388393.76	8388358.8		
2	-4102196.4	8204438.7	8204719.15	8204646.05	0.41	0.284
3	-4038846	8077762.08	8078188.84	8078077.61	0.61	0.037
4	-4012546.3	8025186.56	8025759.64	8025610.27	0.67	0.032
5	-3995786.3	7991690.51	7992409.91	7992222.41	0.73	0.026
6	-3981673.6	7963489.15	7964354.87	7964129.22	0.76	0.002
7	-3974293.4	7948752.84	7949764.87	7949501.09	0.76	0.002
8	-3969154.8	7938499.69	7939658.05	7939356.13	0.71	0.002
9	-3964419	7929052.07	7930356.75	7930016.7	0.76	0.002
10	-3961124	7922486.09	7923937.09	7923558.9	0.77	0.002
11	-3957272.5	7914807.08	7916404.4	7915988.07	0.78	0.002
12	-3955433	7911151.94	7912895.58	7912441.11	0.68	0.002
13	-3952624.8	7905559.57	7907449.52	7906956.92	0.66	0.002
14	-3950408.6	7901151.11	7903187.38	7902656.65	0.79	0.002
15	-3948865.2	7898088.34	7900270.93	7899702.06	0.75	0.002
16	-3947206.1	7894794.25	7897123.15	7896516.14	0.67	0.002
17	-3946500.7	7893407.41	7895882.64	7895237.49	0.74	0.002
18	-3950211.6	7900853.27	7903474.81	7903474.81	0.78	<0.001

Model	LMR p-value	LMR meaning	BLRT p-value	BLRT meaning	Elapsed time
1	-	-	-	-	0:01:02
2	<0.001	2>1	<0.001	2>1	0:01:28
3	<0.001	3>2	<0.001	3>2	0:02:11
4	<0.001	4>3	<0.001	4>3	0:02:08
5	<0.001	5>4	<0.001	5>4	0:02:25
6	<0.001	6>5	<0.001	6>5	0:02:44
7	<0.001	7>6	<0.001	7>6	0:02:41
8	<0.001	8>7	<0.001	8>7	0:03:18
9	<0.001	9>8	<0.001	9>8	0:03:26
10	<0.001	10>9	<0.001	10>9	0:03:27
11	<0.001	11>10	<0.001	11>10	0:03:38
12	<0.001	12>11	<0.001	12>11	0:04:03
13	<0.001	13>12	<0.001	13>12	0:04:45
14	0.0649	14<13	0.0655	14<13	0:05:35
15	0.002	15>14	0.018	15>14	0:05:44
16	<0.001	16>15	<0.001	16>15	0:08:15
17	<0.001	17>16	<0.001	17>16	0:39:15
18	<0.001	18>17	<0.001	18>17	0:07:59

In Step 4 of the LCA, the retained model is interpreted by examining patterns of the profiles and weights of included variables in each profile. The seventeen-profile model had the best fit, which is confirmed graphically by figure 3.1.

Figure 3.1: Elbow plot for the different models considered in this study



AIC = Akaïke Information Criterion; BIC = Bayesian Information Criterion; SABIC = sample-size Adjusted BIC.

Latent Class 1: public agencies in latent class 1 are more likely to adhere to price registry (PR) draft with several *NPB per tender* (f1), a high *proportion of the approved quantity per item for NPBs per tender* (f3), *tenders with only MB and NPB* (f6), several *NPB per supplier per tender* (f9), several *tenders per auctioneer* (f10), high *tender amount related to the total amount* (f11) and low *distribution of winners* (f13). This class has a small sample size (0.2%). From the conversation with the TCU technicians, the candidate indicators somehow represent risk levels. Class 1 has a high number of the binary red flags proposed in this study. We start from the premise that the more indicators appear in a class, the greater the risk level of that class. In this way, class 1 is most likely to be problematic and would be the main one to be analyzed as it presents the most significant number of flagged indicators. Table 3.6 details Class 1.

A more detailed description of the classes is provided in Table 3.6, which presents the results in a conditional probability scale for class 1. So, suppose the UASG belongs to Class 1. In that case, you have a 67.1% probability of participating in Price Registration tenders with a high *number of NPB per tender* and an 86.8% probability of participating in Price Registration tenders with a high *relative item value*. Similarly, Class 1 has a 98.6% probability of tenders with a high *NPB per supplier per tender*. In contrast, the UASGs from Class 1 has a 44.3% probability of being in tenders with a considerable *NPB per supplier.*

Table 3.6: Results in Probability Scale for Class 1 (category 2)

Variable	Estimate	SE.	Est./S.E.	Two-Tailed P-Value
f1: number of NPB per tender	0.671	0.012	53.711	0
f2: proportion of approved quantity per item per NPB	0.191	0.009	20.084	0
f3: prop. of the approved qt. per item for NPBs per tender	1	0	0	1
f4: item relative value	0.868	0.015	58.712	0
f5: tender without MB	0	0	0	1
f6: tender with only MB and NPB	0	0	0	1
f8: amount of NPB per supplier	0.443	0.013	35.305	0
f9: amount of NPB per supplier per tender	0.986	0.012	85.256	0
f10: number of tenders per auctioneer	0	0	0	1
f11: tender amount related to total commitment	0.212	0.009	22.672	0
f13: distribution of winners	0.585	0.013	44.64	0

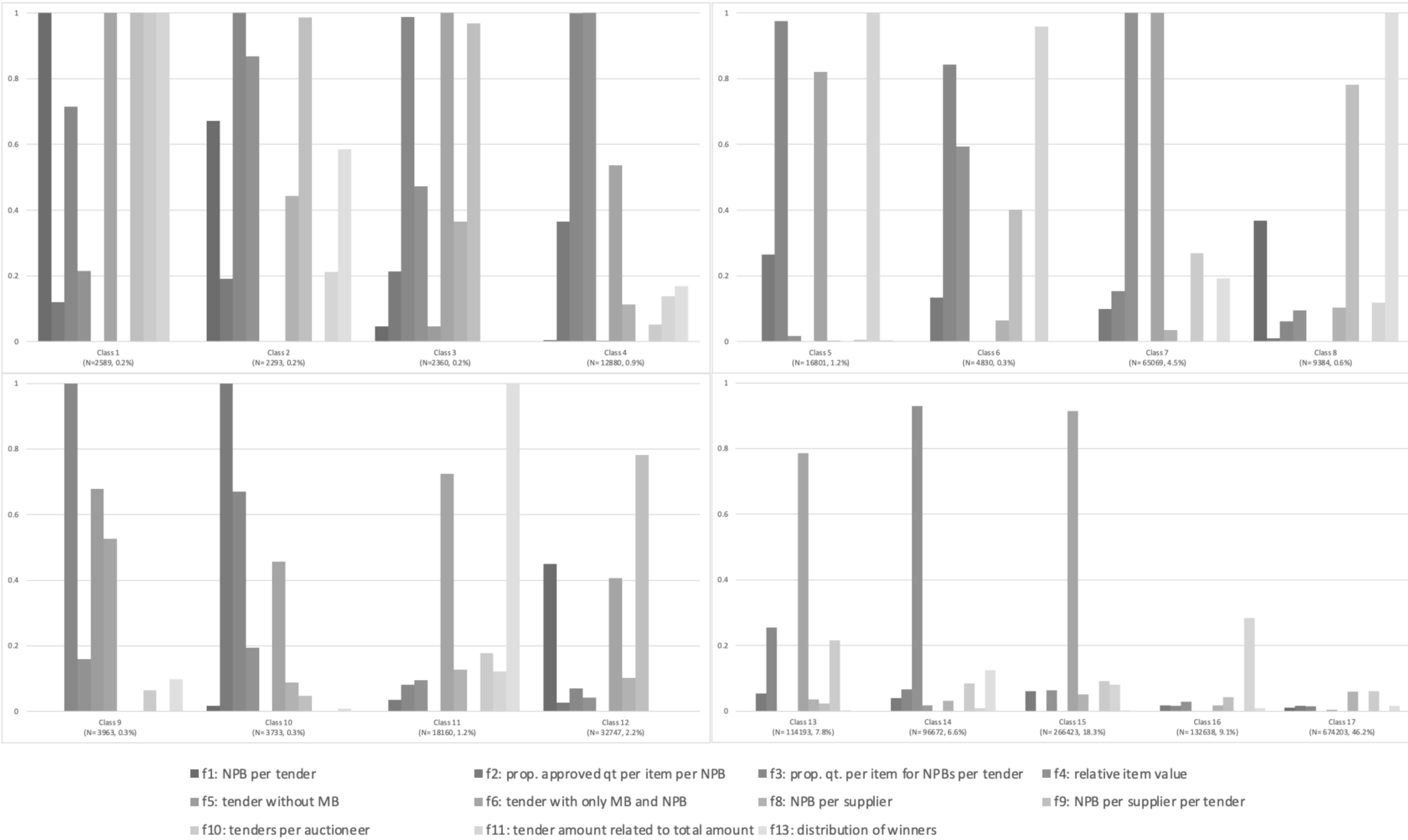
Note: results retaining 17 Class Model

Latent Class 2: comprises units that are more likely to adhere to tenders with several *NPB per tender* (f1), a high *proportion of the approved quantity per item for NPBs per tender* (f3), *tenders with only MB and NPB* (f6), and several *NPB per supplier per tender* (f9). This class has few flagged indicators, but it is still a fair amount and has a small sample size (0.2%). Those characteristics suggest that this cluster could also be a problematic one.

Latent Class 3: responds to participation in tenders with a high *proportion of the approved quantity per item for NPBs per tender* (f3), *tenders with only MB and NPB* (f6), several *NPB per supplier per tender* (f9), and low *distribution of winners* (f13). This would be the third class of interest to investigate the chance of renting price registration drafts.

Latent Class 4: represents the participation in tenders with a high *proportion of the approved quantity per item for NPBs per tender* (f3) and high *relative item value* (f4). The high relative importance of the items suggests that this group participates in bids that have a smaller number of items offered, but they are more expensive. This class may not be a priority in the investigation to find suspicious renting price registration drafts activities.

Figure 3.2: Estimated Probabilities for Category 2



Latent Class 5: represents the participation in tenders with a high *proportion of the approved quantity per item for NPBs per tender* (f3), *tenders with only MB and NPB* (f6), high *tender amount related to the total amount* (f11), and low *distribution of winners* (f13). Based on the opinions of TCU specialists, indicator f6 is considered one of the most critical if accompanied with other flagged indicators. A tender that only includes the managing body and non-participating bodies would have a greater chance of risking questionable practices. Thus, this is another group that could be analyzed in more detail.

Latent Class 6: comprises units that participate in tenders with a high *proportion of the approved quantity per item for NPBs per tender* (f3), and high *tender amount related to the total amount* (f11) and low *distribution of winners* (f13). These three indicators by themselves characterize a moderate risk that warrants further investigation.

Latent Class 7: comprises units that participate in tenders with high *relative item value* (f4) and *tenders with only MB and NPB* (f6). This group is another case in which the indicator f6 in conjunction with bids of high relative value in the item does not seem to be significant indicators for further analysis.

Latent Class 8: the only red flag is f9, several *NPB per supplier per tender*. Remember that the interpretation of the variable f13 (*distribution of winners*) is reverse. So, a value of one for f13 means a high level of distribution of winners, which is a good sign. From the literature on cartel formation, a low value for this indicator would signal a point of attention, which is not the case for this class. So, this group is not part of the future investigation focus.

Latent Class 9: this group has a high *proportion of the approved quantity per item for NPBs per tender* (f3) and participation in *tenders without MB* (f5). Indicator f5 by itself already characterizes an anomaly, as it is not expected that a managing body does not exist in a tender. For this group, an analysis should be carried out to verify possible errors in registration or identify the cause of such bids without a manager. However, these problems are not the focus of this study.

Latent Class 10: this group has a high *proportion of approved quantity per item per NPB* (f2), the *proportion of the amount approved per item for NPBs per tender* (f3), and low *distribution of winners* (f13). This group is characterized by a high number of items approved for price registration minutes and a high concentration of suppliers. These characteristics could indicate an environment conducive to selling price registration minutes. This could be another group to be considered in more detailed analyses.

Latent Class 11: this group has *tenders with only MB and NPB* (f6) and low *distribution of winners* (f13). Although there are two relevant indicators, their presence alone would not justify further investigation considering that other groups have a higher perceived risk of illegal practices.

Latent Class 12: the red flags are f9 (several *NPB per supplier per tender)* and low *distribution of winners* (f13), not signing that is most likely to be problematic. This group could be of greater investigative interest if the purpose of needing you here were to identify the practice of collusion, which is not the case.

Latent Class 13: represents participation in *tenders with only MB and NPB* (f6) and low *distribution of winners* (f13). This group is not problematic and is not characterized by the requirements of analysis of suspected cases. As mentioned before, it is customary to exist tenders with only MB and NPBs.

Latent Class 14: the only indicator is high *relative item value* (f4), which by itself does not represent a relevant analysis factor.

Latent Class 15: has a high *relative item value* (f4) and low *distribution of winners* (f13). This is yet another case in which only two indicators do not justify a high-risk perception for analysis.

Latent Class 16 and 17: As shown in figure 3.2, these two classes do not have indicators other than f13 (low *distribution of winners)*. A few suppliers for certain good service products do not represent a point of attention at first. In addition, it is noteworthy that class 17 means 46% of the sample. This makes us believe that class 17 represents the standard behavior for price registration minutes: few suppliers offer products and services in the same minutes.

Figure 3.2 corresponds to the table in the Mplus output entitled "Results in Probability Scale", reporting just category 2 (variable = 1)[10]. Remember that a value of 1 on these variables

[10] The sum of the probabilities of the two categories is always 1 (ex: for class 1, Category 1 = 0.329 and Category 2 =0.671)

indicates that the public body in this tender item was the upper inner fence. The 17^{th} classes model exhibited varied patterns across the Price Registry spectrum. We ordered the latent classes from the most problematic ones. Thus, the first classes present a greater risk of selling price registration minutes based on the indicators used in this study. In this line of reasoning, the last latent classes offer a lower risk of suspicious activities targeted by this study. It is also worth mentioning that due to the high number of classes identified, the labeling of each class is challenging to interpret and provide practical results. Latent Class 1 (*N*=2589), 2 (*N*= 2293), and 3 (*N*=2360) have a higher probability of being on the upper inner fence.

It may help report LCA to provide names or labels for the profiles based on the observed differences in included variables. However, researchers should be cautious in giving names to avoid a naming fallacy, suggesting the label assigned to a profile is correct and clearly understood, or a reification error offering the tag represents a real construct (Kline, 2011; Masyn, 2013). From the results of the model with seventeen latent classes, the task of labeling is challenging.

Because we do not have tagged cases of suspicious behaviors in the SRP framework based on the candidate indicators and seventeen latent classes make the interpretation challenging, we applied feature selection using information gain (using Weka) on my binary variables. Weka is a powerful data mining software with multiple clustering and feature selection algorithms. However, it does not support large datasets because it only supports sequential execution on a single node. Hence, the size of datasets and processing tasks that Weka can handle within its existing environment is limited by the amount of memory in a single node and sequential execution (Parmar et al., 2017). For instance, we subsampled 1.5% of the data to apply feature selection and got a shortlist of 3 variables to use: *number of non-participating bodies per tender* (f1); the *proportion of the approved quantity per item for NPBs per tender* (f3); and *amount of NPB per supplier per tender* (f9). Then, I applied substantive analysis and included the variables *item relative value* (f4), *tender with only MB and NPB* (f6), and *distribution of winners* (f13). As a result, the following analysis is focused on LCA with binary and continuous variables. The binary variables are f3 and f6.

To balance the benefits of a large sample with the computational power limitations in this research, I created standardized z-scores[11] for the variables *number of the non-participating body per tender*, *amount of NPB per supplier per tender*, *item relative value*, and *distribution of winners*. Using standardized z-scores eases interpretation. A value of zero represents the average value across the measures. In contrast, positive values indicate values higher than the grand mean, and negative values indicate values lower than the grand mean. Table 3.7 provides the descriptive statistics.

Table 3.7: Z-scored variables Descriptive Statistics

Variable/ Sample Size	Mean/ Variance	Skewness/ Kurtosis	Minimum/ Maximum	% with Min/Max	20% /60%	Percentiles 40%/80%	Median
number of NPB per tender	0	0.978	-1.247	0.0012	-0.887	-0.475	-0.269
14566	1.013	0.16	3.024	0.0028	-0.064	0.863	
relative item value	-0.004	0.893	-1.276	0.0172	-0.936	-0.473	-0.203
14566	0.996	0.367	3.69	0.0008	0.091	0.82	
amount of NPB per supplier per tender	-0.016	1.42	-0.892	0.202	-0.892	-0.564	-0.4
14566	1.004	1.397	3.644	0.0008	-0.127	0.693	
distribution of winners	0.001	1.454	-1.15	0.001	-0.796	-0.494	-0.296
14566	1.005	2.594	10.291	0.0001	-0.1	0.752	

LCA suggests the seven-class solution provided the optimal fit for our data (Table 3.8). It has the lowest BIC value and a high entropy value. AIC and SABIC values also provide evidence of a good fit. LMR and BLRT suggest a six-class solution; however, as mentioned before, most scholars use the BIC value as the primary fit statistic test, and we will follow this strategy. The model with an eight-class solution didn't converge, and it wasn't reported[12]. A more detailed description of the classes is provided in Table 3.9, which presents the z-scores mean for each class sub-sample and, for comparative purposes, the overall sample. The standardized z-scores give a relative sense of each measure. As shown in Figure 3.3, which accompanies Table 3.9, the seven classes exhibited varied patterns across the Price Registry spectrum.

[11] We compute z-scores using the formula: [(class mean(var) - grand mean(var))/ grand standard deviation(var)]

[12] Mplus reported the warning: 281 perturbed starting value run(s) did not converge in the initial stage optimizations. Final stage loglikelihood values at local maxima, seeds, and initial stage start numbers: 200 perturbed starting value run(s) did not converge or were rejected in the third stage. The model estimation did not terminate normally due to an error in the computation. Change your model and/or starting values.

Table 3.8: Mixed LCA Model Fit Summary for Research Question

M	Log-likelihood	AIC	BIC	SABIC	Entropy	Smallest class
1	-96707.703	193435.405	193511.27	193479.49	-	-
2	-85534.01	171110.02	171269.335	171202.60	0.81	0.474
3	-81191.326	162446.652	162689.418	162587.73	0.83	0.260
4	-77842.622	155771.244	156097.461	155960.81	0.82	0.200
5	-76689.97	153487.939	153897.607	153726.00	0.81	0.106
6	-75583.516	151297.032	151790.151	151583.59	0.80	0.062
7	-74988.198	150128.396	150704.966	150463.45	0.81	0.051

M	LMR p-value	LMR meaning	BLRT p-value	BLRT meaning	Elapsed time
1	-	-	-	-	0:00:03
2	<0.001	2>1	<0.001	2>1	0:02:35
3	0.015	3>2	0.0154	3>2	0:03:38
4	<0.001	4>3	<0.001	4>3	0:04:39
5	0.004	5>4	0.004	5>4	0:09:06
6	<0.001	6>5	<0.001	6>5	0:09:43
7	0.4361	7<6	0.4384	7<6	0:09:26

After applying the clustering algorithms to the indicators, we consulted with TCU experts to analyze the results. This analysis was carried out subjectively based on the aspects that may lead to an environment more favorable to suspicious renting price registration drafts. Below are descriptions for each of the groupings.

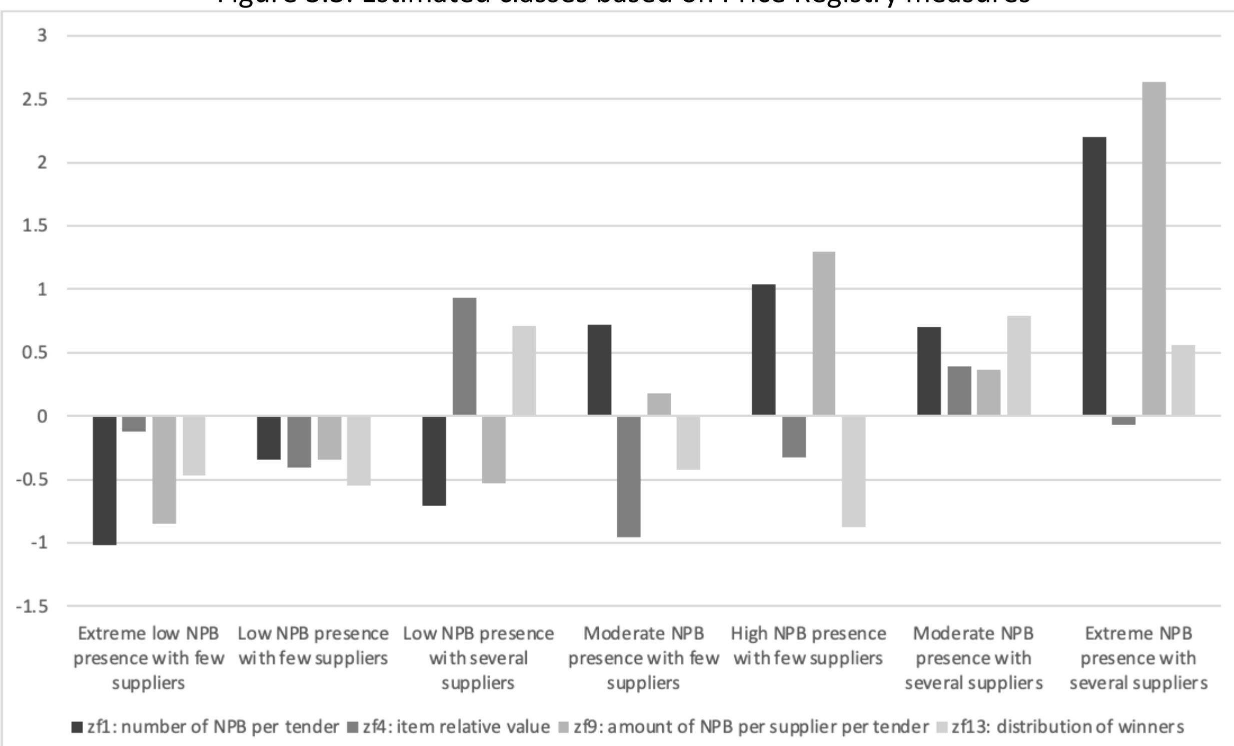
Extreme low NPB presence with few suppliers (*N*=2718) is characterized by government agencies that participated in Price Registration Minutes with a low *number of NPB* (zf1), low *amount of NPB per supplier* (zf9), and low *distribution of winners* (zf13). Remember that a low distribution of winners is equivalent to a high concentration of them. In this sense, the zf13 indicator signals market concentration. This class has a *relative item value* (zf4) around the overall mean and doesn't seem to be a suspicious class scope of this research.

Low NPB presence with few suppliers (*N*=4106) is somewhat like the previous class in terms of the polarity of the variables but with less extreme values. Although this class is composed of public agencies that deal with few suppliers, it also has few NPB, not characterizing an environment more propitious to price registration sales.

Table 3.9: Z-score means of LCA model

Variables	Extreme low NPB presence with few suppliers	Low NPB presence with few suppliers	Low NPB presence with several suppliers	Moderate NPB presence with few suppliers	High NPB presence with few suppliers	Moderate NPB presence with several suppliers	Extreme NPB presence with several suppliers
zf1: number of NPB per tender	-1.014	-0.345	-0.71	0.724	1.042	0.702	2.2
zf4: item relative value	-0.124	-0.407	0.931	-0.957	-0.326	0.396	-0.065
zf9: amount of NPB per supplier per tender	-0.847	-0.34	-0.531	0.183	1.294	0.369	2.631
zf13: distribution of winners	-0.468	-0.543	0.715	-0.42	-0.873	0.791	0.565
Sample size	*N*=2718 18.7%	*N*=4106 28.2%	*N*=2230 15.3%	*N*=1222 8.4%	*N*=749 5.1%	*N*=2647 18.2%	*N*=894 6.1%

Figure 3.3: Estimated classes based on Price Registry measures



Note. Y-axis presented in standard deviation units.

Low NPB presence with several suppliers (*N*=2230) has a low *number of NPB* (zf1) and a low *amount of NPB per supplier* (zf9). The high *relative value of the item* (zf4) may indicate Minutes with a lower quantity of bid items. This group also has an extensive distribution of suppliers. These characteristics lead us to consider this group as participants in bids with a more competitive nature.

Moderate NPB presence with few suppliers (*N*=1222) represents agencies that adhere to price registration minutes that have a high *amount of NPB* (zf1) and low *distribution of winners* (zf13), which means a high concentration of suppliers. Furthermore, the *relative value of the item* (zf4) is also the lowest among the groups. This seems to us to be one of the groups with the potential for suspicious behavior.

High NPB presence with few suppliers (*N*=749) has a high *amount of NPB* (zf1) and *NPB per supplier* (zf9) and is the group with the highest *concentration of suppliers* (zf13). The *relative item value* (zf4) is low. The composition of the indicators in this group leads us to believe that it is more favorable for the sale of price registration minutes. It is an environment with several NPB with few suppliers and in purchases with many items. This class is the more prominent to be considered for further analysis of suspicious activities.

Moderate NPB presence with several suppliers (*N*=2647) has all indicators with positive polarity, indicating that it is composed of organizations that participate in bids with the relevant presence of NPB, relatively few items, and a diversification of suppliers. These elements lead us to dismiss this group as the focus of our analysis of the most favorable environment for the sale of price registration minutes.

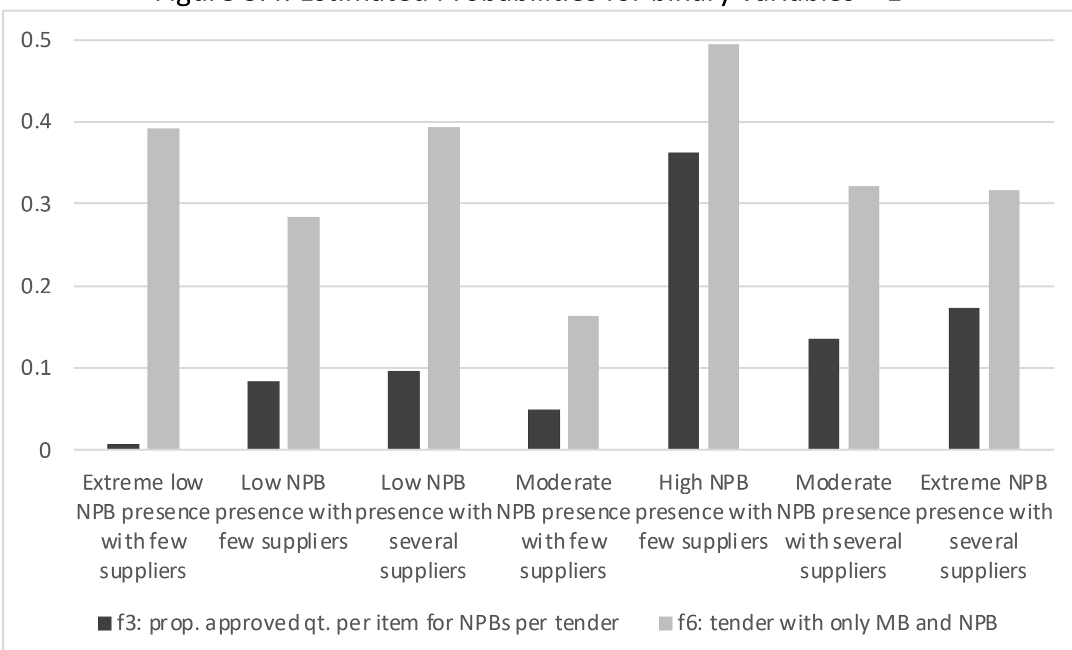
Finally, for *Extreme high NPB presence with several suppliers* (*N*=894), there is increased participation of NPB, with an increased number of NPB per supplier. However, this is somewhat offset by the greater distribution of suppliers. At first, this group would not be the first on the list to be analyzed in suspected cases.

Considering the binary variables in the model (f3: *proportion of approved quantities per item for NPBs per tender* and f6: *tender with only MB and NPB*), the goal is to focus on the relative comparison (figure 3.4).

The class *High NPB presence with few suppliers has a* much higher probability of both variables than all seven classes. This result supports the identification of these indicators as being important in the analysis of suspected cases. Indicator *f3* reflects non-compliance with the provisions of Article 22. § 4º, Decree 7892/2013[13]:

> *The summoning instrument will provide that the quantity resulting from adhesions to the price registration minutes may not exceed, in total, twice the amount of each item registered in the price registration minutes for the managing agency and the participating agencies, regardless of the number of non-participating bodies that join.*

Figure 3.4: Estimated Probabilities for binary variables = 1



Having *tenders with only MB and NPB* (f6), as mentioned before, is the main element that characterizes an environment conducive to the practice of illegally selling price registration minutes. Thus, this indicator further reinforces the degree of risk of this class.

[13] http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2013/decreto/d7892.htm

4. LIMITATIONS, CONCLUSIONS AND POLICY IMPLICATIONS

Results must be interpreted within the context of study limitations. The main limitation of this study is the lack of an outcome variable. To date, TCU has not identified a case that rent of price registration drafts occurred. However, TCU experts assert that this fraudulent practice can result in millions of losses to public coffers. Also, a high opportunity cost of not using these resources in public programs and policies that meet the population`s critical needs, such as health, education, and public safety.

One possibility of improving the model to allow regression analysis would be identifying proxies for minutes rental. A possible proxy would be detecting the rental of minutes held at the state or municipal level. In 2019, the media reported that the municipality of Caravelas copied a bid from the city of Marataízes to serve as a surrogate for a company linked to Espírito Santo politicians[14]. Another possibility would be to identify bids that had the participation of free riders and showed some irregularity related to the contracted prices or the collusion practice, as the market concentration and high prices can encourage the rental of minutes. Suppose we had a proxy as an outcome. In that case, we could estimate binomial regression models, including the latent classes as covariates in these models. We are interested in the extent to which membership in these various classes explains the rental of minutes.

The second limitation is that the data are cross-sectional, limiting our understanding of how these UASGs evolve based on these structural features and suggesting essential directions for future research. Second, we were determined to study tenders in with there is at least one NPB. Even with this scope limitation, the data comprised 1,458,938 observations, in which each observation corresponds to a UASG per tender per item. The large sample provides some advantages, as statistical power. However, it also has the drawback of needing computational power for the estimations. Future studies must utilize a UASG-centered approach in other settings where the long-term trajectory of price registration participation may yield differences in the latent classes detected.

[14] https://www.zerohoranews.com.br/noticia/3025/barriga-de-aluguel-prefeitura-de-caravelas-copiou-licitacao-da-cidade-de-marataizes-para-favorecer-empresarios-e-politicos-capixaba-em-esquema-de-venda-de-atas

Third, correct classification is not guaranteed. In addition, class assignments are probability-based, so it is not possible to determine the exact number or percentage of sample members in each class. In addition, researchers typically assign names to identified classes. Due to the complexity of the class, it is possible to mistakenly engage in a “naming fallacy” where the name of the class does not accurately reflect the membership of the class (Weller et al., 2020).

Fourth, there is no guarantee that different strategies will produce similar results if the available clustering algorithms are different. We have used a popular technique with desirable properties, but we must keep in mind the relative ambiguity of all clustering methods. Finally, the lack of theory and research of price registry sales led us to use TCU experience and indirectly correlated theories (like cartel formation) to build the model.

This type of research has a potential high public impact because it could be used in other sectors in Brazil and could be adapted to other countries, allowing governments to avoid wasting public resources allocated to other projects. Instead of just publishing information on public procurement in an open format, political decision-makers can take a step towards adopting international standards for public procurement and implement a fully digital and automated e-procurement portal that works in South Korea, the Philippines, and Singapore, among others (Argüello & Ziff, 2019). It seems that the Brazilian government is following this strategy.

The public procurement system contains vital procedures, standard documents, rules, and the corresponding legal framework (Dello & Yoshida, 2017). Herrera Murillo (2019) advocates achieving a uniform adoption of terms, languages, and public procurement phases. We could see that the documents for the procurement processes in Brazil are not fully standardized. This inputs additional challenges to detect fraud.

Another significant extension of the System is the contract management module, which is crucial as it allows you to control the work's quality and productivity. The generated historical data can then improve future suppliers' choices (Dello & Yoshida, 2017). In this case, the Brazilian government needs to strengthen its System to incorporate, for example, all the tender proposals in the data frame.

It is also clear that there is much room for better cross-sectoral cooperation, more accurate data, and better use of available data between countries and between the three levels of government in each country. Argüello and Ziff (2019) claim that these types of partnerships can accelerate the scaling of effective testing programs. Furthermore, civilian technology applications and central platforms often publish government data transparently but on a limited scale. A significant challenge is to create deeper links between civil society organizations that collect, extract, and publish government data to detect corruption and government bodies - regulatory, legal, or judicial - to support exchanging information and best practices (Argüello & Ziff, 2019).

Consistent cross-referenced data sets can be identified. However, government data quality still varies between data sets and agencies and is often incomplete, unverifiable, illegible, unaggregated, or out of date. In general, there are many gaps in the regional information from beneficial owners about the ownership or profit of specific companies, making it almost impossible to get a complete picture of the business and ensure that it does not contribute to profit generation based on illegal transactions. Even when public procurement information becomes publicly available, it can be spread across multiple databases, making it difficult to cross-reference and detect corruption or crime signs. The lack of transparency in public procurement is complemented by the fact that "accessibility buyers" or additions changes included during the renegotiation of contracts can open vulnerabilities for corruption. Furthermore, in many countries, these changes are not subject to the same requirements for transparency as the original procurement process and contracts (Argüello & Ziff, 2019).

Another opportunity for improvement is data publishing with a clear goal against corruption. The communication strategy and the design of public portals should be developed, considering society's need for an easy way to access, interpret, and evaluate information on public procurement. Government agencies can also have an exact point of contact for reporting technology indicators and promoting transparency and clean business practices.

The goal of this study has been to challenge the variable-centered approach to study suspicious activities of price registration sales in the context of public procurement in Brazil. Despite the absence of concrete cases prosecuted by TCU, the control agency states that a line of investigation to identify suspicious behaviors in the SRP framework could generate an unprecedented saving of resources in public waiting.

Here we adopt a "UASG-centered" approach. Collectively, findings reveal a broad array of UASG types with nuanced differences in price registration structure, which underscores the importance of treating public tenders holistically. First, we estimated a latent class model using 11 binary indicators selected from the joint analysis of TCU experts and theoretical framework related to cartel formation. This strategy was adopted considering the absence of specific studies on renting price registration drafts (*Barriga de Aluguel* in Portuguese). It was identified that the model with 17 clusters has the best adherence to the database. We identified five groups with a higher risk of questionable practices in the sale of price registration minutes.

In the second approach, we applied feature selection to identify the most relevant variables. Then, a mixed model with four continuous and two binary indicators was applied. LCA suggested the seven-class solution provided the optimal fit for our data. The results identify seven classes across the Price Registration landscape: *Extreme low NPB presence with few suppliers, Low NPB presence with few suppliers, Low NPB presence with several suppliers, Moderate NPB presence with few suppliers, High NPB presence with few suppliers, Moderate NPB presence with several suppliers,* and *Extreme NPB presence with several suppliers. T*hese classes can be differentiated using the number of NPB per tender, NPB per supplier per tender, the relative item value, and the distribution of winners. The findings underscore broader conclusions that serve as more prominent take-away points from the study.

High NPB presence with few suppliers (*N*=749) has a high *amount of NPB* (zf1) and *NPB per supplier* (zf9) and is the group with the highest *concentration of suppliers* (zf13). The *relative item value* (zf4) is low. The composition of the indicators in this group leads us to believe that it is more favorable for the sale of price registration minutes. It is an environment with several NPB with few suppliers and in purchases with many items. This class is the more prominent to be considered for further analysis of suspicious activities.

It is important to consider is distinguishing between false-positive and false-negative prediction errors. The false-positive result means that the algorithm marks a tender as problematic even though no rent of price registry occurred. From an SAI perspective, this can lead to unjustified investigations and look like the worst prediction error. In contrast, a false-negative means that the method will not mark the tender as fraudulent, even if a draft price record rental occurs. Techniques with too many false-negative results do not seem to be worth implementing due to their lack of statistical ability to detect this type of fraud, so false-negative is not desirable. Therefore, an attractive method for the SAI should have a well-balanced false-positive and false-positive performance with acceptable overall out-of-sample performance.

Considering the difficulty in obtaining identified and judged cases of renting price registration drafts and the lack of specific studies in this area, the use of the LCA brought the possibility of identifying some characteristics that can suggest or direct the efforts of the control bodies in identifying illegal cases. Using the LCA model, we identified 749 observations with suspicious attributes for further analysis, intended to be carried out in future studies.

It was also possible to identify errors in the databases reported to TCU for correction during the research. It was also possible to build a dynamic panel based on binary indicators used by TCU. Another practical result of this research is TCU's interest in incorporating part of the indicators established in this project, particularly the *proportion of approved quantity per item per NPB* (f2) and the *proportion of the amount approved per item for NPBs per tender* (f3), in the tracks of the Public Notice and Bid Analysis System (Alice). This artificial intelligence system scans public notices for bidding in the search for irregularities.

This study aims to use statistical and machine learning techniques to flag minutes with the potential for fraud. We provided to TCU a detailed list of suspicious tenders with potential to be problematic. This list is not presented in this study to guarantee the information confidentiality. We expect that the results of this research would benefit public agencies. Considering the partnership with TCU, we expect that the institution will benefit from the realization of this project through the identification of:

- preventive ways for the TCU to act;
- groups with more significant potential for selling minutes for TCU action;
- patterns of occurrence of the rent of minutes by agency, supplier, product, and service;
- systematic behavior of selling minutes;
- improvements in the quality of public procurement procedures;
- new perceptions on public procurement;

- new typologies and classifications specific to the Alice system;
- new databases to be added to the TCU systems; and
- opportunities for improvement in the allocation of financial, technological, and human resources of TCU.

In closing, we acknowledge that the models presented are not perfect, but they provide a direction on which tenders to investigate. Although it is difficult to manage large amounts of data, data analysis shows promise for trial practice and research (Singh et al., 2019). Data analysis can test transactions, maintain audit quality, detect fraud more effectively (Earley, 2015) and create new forms of audit evidence. The fraud system is constantly changing and dynamic. Therefore, new rules should continuously be developed to monitor further fraud attempts. In such scenarios, a classification-based approach is essential for identifying anomalous transactions. It is not always possible to rely on a rule-based approach. Therefore, a classification-based approach is urgently needed to track abnormal or fraudulent transactions without rules (Singh et al., 2019).

Future research should focus on applying the UASG-centered approach to start a new research agenda that can help public agencies to improve their surveillance of suspicious activities of price registry rentals. Our results highlight that such an approach can provide insights not apparent in the more common variable-centered course. Although our goal was to unpack further insights within Price Registration System, future work may wish to include more of the measures from the models in the clustering routine.

REFERENCES

Abrantes-Metz, R. M. (2013). Proactive vs. reactive anti-cartel policy: the role of empirical screens. Social Science Research Network. http://ssrn.com/abstract=2284740

Acock, A. (2013). Discovering Structural Equation Modeling Using Stata. 1st Edition. STATA Press.

Acock, A. (2016). A Gentle Introduction to Stata. 5th Edition. STATA Press.

Akaike, H. (1987). Factor analysis and AIC. Psychometrika, 52, 317–332.

Anderson, E. J., & Cau, T. D. H. (2011). Implicit collusion and individual market power in electricity markets," Eur. J. Oper. Res., vol. 211, no. 2, pp. 403– 414.

Anitha, P., Krithka, G., & Choudhry, M. D. (2014). International Journal of Advanced Research in Computer Engineering & Technology, Vol. 3, No. 12, pp. 4324-4331, http://ijarcet.org/wp-content/uploads/IJARCET-VOL-3-ISSUE-12-4324-4331.pdf.

Anti-cartel Enforcement Manual. (2015). International competition network. http://www.internationalcompetitionnetwork.org/working-groups/current/cartel/manual.aspx

Argüello, M. F. P., & Ziff, T. (2019). Hacking Corruption: Tech Tools to Fight Graft in the Americas. Washington, D.C.: Inter-American Dialogue and the Atlantic Council of the United States.

Bajari, P., & Ye, L. (2003). Deciding between competition and collusion. Rev. Econ. Statist., vol. 85, pp. 971–989.

Ballesteros-Pérez, P., & Skitmore, R. M. (2017). On the distribution of bids for construction contract auctions. Construction Manage. Econ., vol. 35, n. 3, pp. 106–121.

Bardhan, P. (1997). Corruption and Development: A Review of Issues. Journal of Economic Literature, Vol XXXV: 1320-1346.

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. Psychological Methods, 9, 3-29.

Bhargava, V. (2005). World Bank Global Issues Seminar Series - The Cancer of Corruption. pp. 2-10.

Bishop, S., & Walker, M. (2010). The economics of EC competition law. Sweet and Maxwell, Andover, UK, p 171.

Braga, T. C. A. (2015). CADE, Cartéis e Licitações: um Novo Nicho da Política Antitruste Brasileira. Revista de Defesa da Concorrência, Brasília, v. 3, n. 1, p. 108-132, May.

Brannman, L. , Klein, J. D. , & Weiss, L. W. (1987). The price effects of increased competition in auction markets. Rev. Econ. Statist., vol. 69, no. 1, pp. 24–32.

Brazil. (1993). Federal Law n. 8,666 of June 21, 1993 (in Portuguese) [Online]. The Federative Republic of Brazil. Available at http://www.planalto.gov.br/ccivil_03/leis/l8666cons.htm on January 2021.

Brazil. (2002). Federal Law n. 10.520 of July 17, 2002 (in Portuguese) [Online]. The Federative Republic of Brazil. Available at www.planalto.gov.br/ccivil_03/leis/2002/l10520.htm on January 2021.

Buchanan, B., & Miller, T. (2017). Machine Learning for Policymakers: What It Is and Why It Matters. The Cyber Security Project. Belfer Center for Science and International Affairs at Harvard Kennedy School.

Busu, M.; & Busu, C. (2021). Detecting Bid-Rigging in Public Procurement. A Cluster Analysis Approach. Administrative Sciences 11: 13. https://doi.org/ 10.3390/admsci11010013.

Campos, F. (2008). As práticas de conluio nas licitações públicas à luz da teoria dos jogos. Revista Análise Econômica, Porto Alegre, v. 1, n. 50, p. 185-206, September.

Carvalho, L. M. S. (2018). O Cartel na Licitação Pública. 115 f. Dissertação (Mestrado) - Curso de Ciências Jurídico-políticas Com Menção em Direito Administrativo, Universidade de Coimbra, Coimbra.

Castro, A., Rogério, D., Márquez, R. & Lira, W. (2018). Identifying Behaviors and Irregularities in Public Procurement in the Brazilian Federal Government. Paper 2707-2018.

Chassang, S., & Ortner, J. (2019). Collusion in auctions with constrained bids: Theory and evidence from public procurement. Journal of Political Economy, 127(5), 2269-2300.

Collins, L. M., Fidler, P. L., Wugalter, S. E., & Long, J. D. (1993). Goodness-of-fit testing for latent class models. Multivariate Behavioral Research, 28, 375–389.

Conley, T. G., & Decarolis, F. (2016). Detecting bidders groups in collusive auctions. American Economic Journal: Microeconomics, 8(2), 1-38.

Cosnita-Langlais, A., & Tropeano, J. P. (2013) Fight cartels or control mergers? On the optimal allocation of enforcement efforts within competition policy. Int Rev Law Econ 34:34–40

Czibik, Á., Fazekas, M., Tóth, B, & János, T. (2014). Toolkit for detecting collusive bidding in public procurement. With examples from Hungary. Working Paper series: CRC-WP/2014:02, Version 2.0—Corruption Research Center, Budapest.

Da Silva, W. C. (2017). As aquisições na administração pública federal, através da adesão em ATAS de registro de preços. Pós-graduação lato sensu, Universidade Cândido Mendes, Rio de Janeiro. http://www.avm.edu.br/docpdf/monografias_publicadas/posdistancia/54706 .pdf on February 2021.

Danger, K. & Capobianco, A. (2009b). Detecting bid rigging in public procurement, OECD. www.oecd.org/competition/cartels/42594486.pdf

Danger, K., Capobianco, A. (2009a). Guidelines for fighting bid rigging in public procurement, OECD. https://www.oecd.org/competition/cartels/42851044.pdf

Das, S. R. (2016). Data Science: Theories, Models, Algorithms, and Analytics. S. R. Das. https://srdas.github.io/Papers/DSA_Book.pdf.

Dello, A., & Yoshida, C. (2017). Online tendering and evaluation for public procurement in Tanzania. 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Kanazawa, pp. 137-141, DOI: 10.1109/SNPD.2017.8022713.

Desordi, D., & Bona, C. D. (2020). A inteligência artificial e a eficiência na administração pública. Revista De Direito, 12(02), 01-22. https://doi.org/10.32361/202012029112

Earley, C. E. (2015). Data analytics in auditing: Opportunities and challenges. Business Horizons, 58, 493–500.

Fazekas, M., & Tóth, B. (2016). Assessing the potential for detecting collusion in Swedish public procurement. Stockholm: Swedish Competition Authority.

Ferguson, S. L., G. Moore, E. W., & Hull, D. M. (2020). Finding latent groups in observed data: A primer on latent profile analysis in Mplus for applied researchers. International Journal of Behavioral Development, 44(5), 458–468. https://doi.org/10.1177/0165025419881721

Fóra, G. (2014) Közbeszerzési eljárások dokumentumainak adatfeldolgozása 1998–2004 (Data extraction from public procurement documents 1998–2004). 8th Report. Korrupciókutató Központ, Budapest.

Funderburk, D. R. (1974). Price Fixing in the Liquid-Asphalt Industry: Economic Analysis versus the Hot Document. Antitrust L. & Econ. Rev., 7, 61.

Gómez-Lobo, A., &Szymanski, S. (2001). A law of large numbers: Bidding and compulsory competitive tendering for refuse collection contracts. Rev. Ind. Org., vol. 18, pp. 105–113.

Grega, M., & Nemec, J. (2015). Factors influencing final price of public procurement: Evidence from Slovakia. Procedia Econ. Finance, vol. 25, pp. 543–551.

Gupta, S. (2002). Competition and collusion in a government procurement auction market. Atlantic Econ. J., vol. 30, no. 1, pp. 13–25.

Goryunova, N., Baklanov, A., & Ianovski, E. (2021). Detecting corruption in single-bidder auctions via positive-unlabelled learning. HSE University, Russian Federation. February 11.

Hagenaars, J., & McCutcheon, A. (Eds.). (2002). Applied latent class analysis models. New York: Cambridge University Press.

Harrington Jr., J. E., & Chang, M. H. (2009). Modeling the birth and death of cartels with an application to evaluating competition policy. Journal of the European Economic Association, 7(6), 1400-1435.

Harrington, J. E. (2006). Behavioral screening and detection of cartels. In: Ehlermann CD, Atanasiu I (eds) European competition law annual 2006: enforcement of prohibition of cartels. Hart Publishing, Oxford, Portland.

Harrington, J. E. (2008). Detecting cartels. In: Buccirossi P (ed) Handbook in antitrust economy. MIT Press.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. Organizational Research Methods, 7(2), 191–205. https://doi.org/10.1177/1094428104263675.

Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. Structural Equation Modeling, 14, 202–226.

Herrera Murillo, D. J. (2019). Using social network analysis in open contracting data to detect corruption and collusion risks. Eindhoven University of Technology. September.

Hinloopen, J., & Martin, S. (2006) The economics of cartels, cartel policy, and collusion: introduction to the special issue. Int J Ind Organ 24(6):1079–1082.

Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. Psychological Methods, 11, 36–53.

Hölzer, H. (2014). Checklist of the Swedish Competition Authority in Harmonization of public procurement system in Ukraine with EU standards, Report on the development of diagnostic tools and Guidelines for the AMCU for identifying and preventing bid-rigging cases, Annex 3., Crown Agents Ltd., 2 Oct 2014.

Huber, M., & Imhof, D. (2019). Machine learning with screens for detecting bid-rigging cartels. International Journal of Industrial Organization, Elsevier, vol. 65(C), pages 277- 301.

Huber, M., Imhof, D., & Ishii, R. (2020). Transnational machine learning with screens for flagging bid-rigging cartels. Université de Fribourg.

Hüschelrath, K. (2010). How Are Cartels Detected? The Increasing Use of Proactive Methods to Establish Antitrust Infringements. Journal of European Competition Law & Practice, 1, 522-528.

Hüschelrath, K., & Veith, T. (2011). Cartel detection in procurement markets. Discussion Paper No. 11-066. Centre for European Economic Research, Mannheim.

International Monetary Fund (IMF). (2019). Fiscal Monitor. Washington, April. Chapter 2. Curbing Corruption reframe.

Ivaldi, M., Khimich, A., Jenny, F. (2014). Measuring the economic effects of cartels in developing countries, published in the framework of the CEPR PEDL program. https://unctad.org/system/files/official-document/ditcclpmisc2014d2_en.pdf on January, 2021.

Jain, A. (2001). Corruption: A Review. Journal of Economic Surveys 1 5(1): 71-121.

James, G. Witten, D., Hastie, T. & Tibshirani, R. (2017). An Introduction to Statistical Learning with Applications in R. Springer.

Jensen, J., & Anderson, T. (2016). A Qualitative Comparison of Anti-Corruption Measures in Guatemala and Brazil. Journal of Political Democracy, vol. 1, no. 2, pp. 1–20, 2016.

Kawai, K., & Nakabayashi, J. (2014). Detecting large-scale collusion in procurement auctions. Available at SSRN 2467175.

Kessing, S. G., Konrad, K. A., & Kotsogiannis, C. (2007). Foreign direct investment and the dark side of decentralization. Economic Policy, 22(49), 6-70.

Kubrin, C. E., Branic, N., & Hipp, J. R. (2021). (Re)conceptualizing Neighborhood Ecology in Social Disorganization Theory: From a Variable-Centered Approach to a Neighborhood-Centered Approach. Crime & Delinquency. https://doi.org/10.1177/00111287211041527

Kuhlman, J. M., & Johnson, S. R. (1983). The number of competitors and bid prices. Southern Econ. J., vol. 50, no. 1, pp. 213–220.

Lanzillotti, R. F. (1996). The great school milk conspiracies of the 1980s. Review of Industrial Organization, 11(4), 413-458.

Levenstein, M. C., & Suslow, V. Y. (2011). Breaking up is hard to do: Determinants of cartel duration. The Journal of Law and Economics, 54(2), 455-492.

Lima, M. C. (2010). Comparação de custos referenciais do DNIT e licitações bem sucedidas. XIII SINAOP – Simpósio Nacional de Auditoria de Obras Públicas - Porto Alegre – RS (in Portuguese).

Lo, Y., Mendell, N., & Rubin, D. (2001). Testing the number of components in a normal mixture. Biometrika, 88, 767–778.

Lubke, G., & Muthén, B. O. (2007). Performance of factor mixture models as a function of model size, criterion measure effects, and class-specific parameters. Structural Equation Modeling, 14, 26-47.

Magidson, J., & Vermunt, J. (2004). Latent class models. In D. Kaplan (Ed.), Handbook of quantitative methodology for the social sciences (pp. 175–198). Newbury Park, CA: Sage.

Marsh, H. W., Lu¨dtke, O., Trautwein, U., & Morin, A. J. S. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept. Structural Equation Modeling, 16, 191-225.

Marshall, R. C., & Marx, L. M. (2007). Bidder collusion. J Econ Theory 133(1):374–402

Marshall, R. C.; & Marx, L. (2009). The vulnerability of auctions to bidder collusion (Report). Quart. J. Econ., vol. 124, no. 2, pp. 883–910.

Martinez, A. P. (2014). Aplicação do Direito da Concorrência a Licitações Públicas: Cartéis. Available on http://www.seae.fazenda.gov.br/assuntos/advocacia-daconcorrencia/

apostilas/arquivos/2seae_aplicacao_direito_concorrencia_licitacoes_publicas_carteis-final.pdf on January, 2021.
Mauro, P. (1995). Corruption and growth. Quarterly Journal of Economics, 110, 681-712.
Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In T. Little (Eds.), The Oxford handbook of quantitative methods (pp. 551–611). New York, NY: Oxford University Press.
McCrudden, C. & Gross, S. G. (2006). WTO Government Procurement Rules and the Local Dynamics of Procurement Policies: A Malaysian Case Study, European Journal of International Law, Volume 17, Issue 1, February 2006, Pages 151–185, https://doi.org/10.1093/ejil/chi162
McLachlan, G., & Peel, D. (2000). Finite mixture models. New York: Wiley.
Menocal, A. (2015). Why corruption matters: understanding causes, effects, and how to address them. DFID Evidence Paper. London: DFID. February 2015.
Mehra, S. K. (2015). Antitrust and the Robo-Seller: Competition in the Time of Algorithms, Minnesota. Law Review, Vol. 100, www.minnesotalawreview.org/wp-content/uploads/2016/04/Mehra_ONLINEPDF1.pdf.
Miller, D. T. (2019). Three International Studies Computational Social Science Inquiries Examining Large Corpora of Natural Data. University of Washington, Jackson School of International Studies.
Morgan, E.J. (2009). Controlling cartels: implications of the EU policy reforms. European Management Journal, 27, 1-12.
Morin, A. J. S., Meyer, J. P., Creusier, J., & Biétry, F. (2016). Multiple-Group Analysis of Similarity in Latent Profile Solutions. Organizational Research Methods, 19(2), 231–254. https://doi.org/10.1177/1094428115621148
Morin, A. J. S., Maıano, C., Nagengast, B., Marsh, H. W., Morizot, J., & Janosz, M. (2011). General growth mixture analysis of adolescents' developmental trajectories of anxiety: The impact of untested invariance assumptions on interpretations. Structural Equation Modeling, 18, 613-648.
Morin, A. J. S., & Marsh, H. W. (2015). Disentangling shape from level effects in person-centered analyses: An illustration based on university teacher multidimensional profiles of effectiveness. Structural Equation Modeling, 22, 33-59.
Morozov, I., & Podkolzina, E. (2013). Collusion detection in procurement actions. Working Paper. National Research University, Higher School of Economics (Russia), Moscow.
Morselli, C., & Ouellet, M. (2018). Network similarity and collusion. Social Networks, 55, 21-30.
Muthén, B. O. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). Psychological Methods, 8, 369-377.
Muthén, L. K., & Muthén, B. O. (2017). Mplus user's guide (8th ed.). Los Angeles, CA: Muthén & Muthén. (Original work published 1998).
Nguyen, T., Zhou, L., Spiegler, V., Ieromonachou, P. & Lin, Y. (2018). Big data analytics in supply chain management: A state-of-the-art literature review. Computers and Operations Research, 98 (2018) 254–264.
Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. Translational Issues in Psychological Sciences, 4, 440–461. doi:10.1037/tps0000176

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Structural Equation Modeling, 14, 535–569.

Oberski, D. (2016). Mixture models: Latent profile and latent class analysis. In J. Robertson & M. Kaptein (Eds.), Modern statistical methods for HCI (pp. 275–287). Cham, Switzerland: Springer International Publishing.

OECD (2007). Roundtable on Facilitating Practices in Oligopolies. www.oecd.org/daf/competition/41472165.pdf.

OECD (2009). Diretrizes para combater o conluio entre concorrentes em contratações públicas. https://www.oecd.org/daf/competition/cartels/44162082.pdf

OECD (2012). Recommendation of the OECD council on fighting bid rigging in public procurement. OECD Council, Paris, 2012.

OECD (2013). Korea's submission to the Roundtable on Ex officio Cartel Investigations, www.oecd.org/daf/competition/exofficio-cartel-investigation-2013.pdf, Report to the OECD Council on the implementation of the Recommendation of the Council on Fighting Bid Rigging in Public Procurement.

OECD (2014). OECD Foreign Bribery Report: An Analysis of the Crime of Bribery of Foreign Public Officials, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264226616-en.

OECD (2016). Preventing Corruption in Public Procurement. http://www.oecd.org/gov/ethics/Corruption-in-Public-Procurement-Brochure.pdf.

OECD (2017a). Government at a Glance: Latin America and the Caribbean 2017. https: //www.oecd.org/ gov/lac-brazil.pdf.

OECD (2017b). Algorithms and Collusion: Competition Policy in the Digital Age www.oecd.org/competition/algorithms-collusion-competition-policy-in-the-digital-age.htm.

OECD (2017c). Algorithms and Collusion - Background Note by the Secretariat. DAF/COMP(2017)4.

Osborne, J. W. (2012). Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data. Thousand Oaks, CA: Sage.

Padhi, S. S., & Mohapatra, P. K. J. (2011) Detection of collusion in government procurement auctions. J Purchasing Supply Manage 17(4):207–221

Panis, A. C. (2020). Inovação em compras públicas: estudo de caso do robô Alice da Controladoria-Geral da União (CGU). 2020. 116 f., il. Dissertação (Mestrado em Administração) - Universidade de Brasília, Brasília, 2020.

Pereira, G. P. C. (2002). O mercado da construção civil para obras públicas como instrumento de auditoria: Uma abordagem probabilística (in Portuguese), Mestrado Dissertação, Universidade Federal de Pernambuco, Recife, Brazil.

Pesendorfer, M. (2000). A study of collusion in first-price auctions. The Review of Economic Studies, 67(3), 381-411.

Petras, H., & Masyn, K. (2010). General growth mixture analysis with antecedents and consequences of change. In A. R. Piquero & D. Weisburd (Eds.), Handbook of quantitative criminology (pp. 69-100). New York, NY: Springer.

Peugh, J., & Fan, X. (2015). Enumeration index performance in generalized growth mixture models: A Monte Carlo test of Muthén's (2003) hypothesis. Structural Equation Modeling, 22, 115–131. DOI: 10.1080/10705511.2014.919823

PMI, (2017). A Guide to the Project Management Body of Knowledge (PMBOK® Guide). Sixth Edition, Project Management Institute.

Põlluste, M.K. (2019). Detecting Corruption in Public Procurement through Open Data Analysis.

Porter, R. H., & Zona, J. D. (1997). Ohio school milk markets: An analysis of bidding (No. w6037). National Bureau of Economic Research.

Rose-Ackerman, S. & Palifka, B. (2016) Corruption and Government: Causes, Consequences, and Reform. Cambridge, UK: Cambridge University Press.

Salih, M. A. R. (2013). The Determinants of Economic Corruption: A Probabilistic Approach. Adv. Manag. Appl. Econ., vol. 3, no. 3, pp. 155–169.

Samuel, A. (1959). Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development 3 (3): 210–29.

Santana, J. E. (2014). Pregão presencial e eletrônico: sistema de registro de preços: manual de implantação, operacionalização e controle. 4. ed. Belo Horizonte: Fórum, 2014.

Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6, 461–464.

Schwalbe, U. (2018). Algorithms, machine learning, and collusion. Journal of Competition Law & Economics, 14(4), 568-607.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. Psychometrika, 52, 333–343.

Scott, F. A. (2000). Great school milk conspiracies revisited. Review of Industrial Organization, 17(3), 325-341.

Selten, R. (1973). A simple model of imperfect competition, where 4 are few and 6 are many. Int. J. Game Theory, vol. 2, no. 1, pp. 141–201.

Signor, R., Ballesteros-Pérez, P., & Love, P. E. D. (2021). Collusion Detection in Infrastructure Procurement: A Modified Order Statistic Method for Uncapped Auctions. IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT.

Signor, R., Love, P. E. D., Belarmino, A. T. N., & Olatunji, O. (2020). Detection of collusive tenders in infrastructure projects: Learning from operation car wash. J. Construction Eng. Manage., vol. 146, no. 1, 2020, Art. nº. 05019015.

Silva, M. A. S. (2020). Descoberta de conhecimento na análise de licitações no Estado de Goiás. PUC Goiás. https://repositorio.pucgoias.edu.br/jspui/handle/123456789/1273

Singh, N., Lai, K., Vejvar, M., & Cheng, T.C. (2019). Data-driven auditing: A predictive modeling approach to fraud detection and classification. Journal of Corporate Accounting & Finance.

Steinley, D. (2003). Local optima in K-Means clustering: What you don't know may hurt you. Psychological Methods, 8, 294–304.

Svensson, J. (2005). Eight Questions about Corruption. The Journal of Economic Perspectives, 19(3):19-42.

Tanzi, V., & Davoodi, H. (1998). The Welfare State, Public Investment, and Growth: Selected Papers from the 53rd Congress of the International Institute of Public Finance, pp. 41–60.

Tas, B. K. O., Dawar, K., Holmes, P. M., & Togan, S. (2017). Does the World Trade Organization Government Procurement Agreement Deliver What It Promises? (March 16). Available at SSRN: https://ssrn.com/abstract=2934127 or http://dx.doi.org/10.2139/ssrn.2934127

Tatsch, D. R. D. (2012). Caracterização dos Cartéis no Brasil. 36 f. Monografia (Especialização) - Curso de Economia, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.

Tein, J. Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. Structural Equation Modeling, 20, 640–657. doi:10.1080/10705511.2013.824781

Tofighi, D., & Enders, C. (2008). Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), Advances in latent variable mixture models (pp. 317-341). Charlotte, NC: Information Age Publishing.

Tolvanen, A. (2007). Latent growth mixture modeling: A simulation study (Unpublished doctoral dissertation). Department of Mathematics, University of Jyva¨skyla¨, Jyva¨skyla¨, Finland.

Tong, S., & Koller, D. (2001). Support vector machine active learning with applications to text classification. J Mach Learn Res 2:45–66, November.

Vadász, P., Benczúr, A., Füzesi, G., and Munk, S. (2016) Identifying Illegal Cartel Activities from Open Sources. In Akhgar et al. (eds.), Open Source Intelligence Investigation, from Strategy to Implementation. Springer International Publishing.

Van Vlasselaer, V., Akoglu, L., Eliassi-Rad, T., Snoeck, M. & Baesens, B. (2015). Guilt-by-constellation: Fraud detection by suspicious clique memberships. Proc. Annu. Hawaii Int. Conf. Syst. Sci., vol. 2015-March, pp. 918–927.

Veale, M. (2017). Connecting diverse public sector values with the procurement of machine learning systems. Presented at the Data for Policy 2016 - Frontiers of Data Science for Government: Ideas, Practices and Projections (Data for Policy), Cambridge, United Kingdom: Zenodo. http://doi.org/10.5281/zenodo.571786.

Vojtěch, S., Mynarz, J., Węcel, K., Klímek, J., Knap, T., & Nečaský, M. (2014). Linked open data for public procurement. In: Auer S, Bryl V, Tramp S (eds) Linked open data. Creating knowledge out of interlinked data. Lecture Notes in Computer Science 8661. Springer, Heidelberg, pp. 196–213.

Wachs, J., & Kertész, J. (2019). A network approach to cartel detection in public auction markets. Scientific reports, 9(1), 1-10.

Weller, B. E., Bowen, N. K., & Faubert, S. J. (2020). Latent Class Analysis: A Guide to Best Practice. Journal of Black Psychology, 46(4), 287–311.

Werden, G. J., Hammond, S. D., & Barnett, B. A. (2011). Deterrence and detection of cartels: using all the tools and sanctions. Antitrust Bull 56(2):207–234

APPENDIX

Table A: Correlation Table

	f1	f2	f3	f4	f5	f6	f7
f1	1						
f2	0	1					
f3	0.0359*	0.1764*	1				
f4	-0.0139*	0.0945*	0.1526*	1			
f5	-0.0107*	-0.0082*	0.1126*	0.0353*	1		
f6	0.0016	0.1088*	0.1928*	0.0539*	-0.0036*	1	
f7	-0.0100*	-0.0068*	0.0946*	0.0400*	0.9403*	-0.0155*	1
f8	0.0424*	0.0170*	0.0168*	0.0072*	-0.0140*	0.0214*	-0.0130*
f9	0.4546*	-0.0015	0.0618*	-0.0171*	-0.0112*	-0.0054*	-0.0098*
f10	-0.0013	0.0094*	0.0628*	0.0580*	-0.0011	0.1914*	-0.0016
f11	-0.0013	0.0171*	0.0786*	-0.0539*	-0.0236*	-0.0877*	-0.0221*
f12	0.1634*	0.0081*	0.0463*	-0.0610*	-0.0179*	-0.0322*	-0.0165*
f13	0.0750*	0.0079*	0.0366*	0.1822*	0.0183*	0.0372*	0.0083*

	f8	f9	f10	f11	f12	f13
f8	1					
f9	0.0592*	1				
f10	0.0016	-0.0266*	1			
f11	-0.0372*	0.0280*	-0.0613*	1		
f12	-0.0334*	0.1137*	-0.0105*	0.5258*	1	
f13	0.0531*	0.0921*	0.0297*	-0.0128*	-0.0039*	1