

Assessing randomness in case assignment: the case study of the Brazilian Supreme Court

DIEGO MARCONDES[†]

*Department of Applied Mathematics, Institute of Mathematics and Statistics,
University of São Paulo, Brazil*

CLÁUDIA PEIXOTO

Department of Statistics, Institute of Mathematics and Statistics, University of São Paulo, Brazil

AND

JULIO MICHAEL STERN

*Department of Applied Mathematics, Institute of Mathematics and Statistics,
University of São Paulo, Brazil*

[Received on 18 October 2018; revised on 23 January 2019; accepted on 31 January 2019]

Sortition, i.e. random appointment for public duty, has been employed by societies throughout the years as a firewall designated to prevent illegitimate interference between parties in a legal case and agents of the legal system. In judicial systems of modern western countries, random procedures are mainly employed to select the jury, the court and/or the judge in charge of judging a legal case. Therefore, these random procedures play an important role in the course of a case, and should comply with some principles, such as transparency and complete auditability. Nevertheless, these principles are neglected by random procedures in some judicial systems, which are performed in secrecy and are not auditable by the involved parties. The assignment of cases in the Brazilian Supreme Court is an example of such a procedure, for it is performed using procedures unknown to the parties involved in the judicial cases. This article presents a review of how sortition has been historically employed by societies and discusses how Mathematical Statistics may be applied to random procedures of the judicial system, as it has been applied for almost a century on clinical trials, for example. A statistical model for assessing randomness in case assignment is proposed and applied to the Brazilian Supreme Court. As final remarks, guidelines for the development of good randomization procedures are outlined.

Keywords: sortition, case assignment, Brazilian Supreme Court, statistical modelling, randomness.

κληρω νυν πεπαλασθε διαμπερες ος κε λαχησιω:

Let the lot be shaken for all of you, and see who is chosen
Iliad, VII, 171.

מְרִיגִים יִשְׁבֵּית הַעֲרֵל וּבֵן עֲצוּמִים יִפָּרֵד

Casting the dice puts judgement quarrels to rest and keeps powerful parts separated
Proverbs 18: 18.

[†]Corresponding author. Email: dmarcondes@ime.usp.br

1. Introduction

The judicial systems of modern western states make use of randomization procedures. Important decisions in the course of a judicial case are made by rolling dice. This may be perceived as an unnatural or counter-intuitive expedient. Why should the flip of a coin take part in casting the fate of a legal case? Why should we abdicate making (best-informed) deterministic decisions? The next three sections address these questions, after a brief historical review. The following two sections present a statistical analysis designed to test the effective randomness of case assignment in the Brazilian Supreme Court. The last section gives our final remarks.

1.1 *Running democracy and rolling the dice*

The city-state of Athens was the undisputed champion in the use of sortition (random appointment) for managing its public affairs, especially in the period between the democratic reforms of Kleisthenes, in 507 BCE, and the fall of Athenian democracy, in 322 BCE; see Dowlen (2009), Hansen (1999), Headlam (1933) and Staveley (1972). While some specialized officers were elected based on specific expertise and experience, like military chiefs, water supply and treasury managers, religious priests etc., most of the public duties in Athens were carried out by citizens of the polis, that held short term positions appointed by a system based on district representation, sortition and frequent rotation.

Sortition was also used to constitute *dikasteria*, citizen's courts where *dike* (fair justice) was served. From a total of approximately 30 000 men recognized as qualified citizens (excluding women, slaves, foreigners and their descendants and other non-represented categories), 6000 district representatives would be selected to the general assembly and, out of those, 501 individuals would be randomly chosen to constitute a *dikasterion*. Some specially important cases would require a larger court of justice, with up to 2501 members! Moreover, in contrast to contemporary jurors, *dikastai* swore by the Heliastic Oath *not* to discuss the case in judgement among each other. In contrast, each *dikastes* should vote, by secret ballot, according to his individual and independent opinion, expressing his own best knowledge of the law and concern for the public good; see Hansen (1999, p. 182).

Implementation of a democratic system based on the Heliastic oath required a corresponding reform of the Athenian social organization and representation system. In the oligarchic society, predating Kleisthenes reforms of 507 BCE, civil duties were performed by representatives sent by their *phratia* (brotherhood)—a sub-group within traditional *phyle* (clans or tribes). After the reforms, representatives were elected in their *demos*—a geographically defined district or neighborhood, see Hansen (1999).

The Hebrew bible uses randomization devices for several purposes, including ritualistic divination and judgement, and also as a practical tool for appeasement or fair distribution of goods and services; see for example, Leviticus 16: 8–10, Numbers 26: 56, 1Chronicles 26: 13, Proverbs 16: 33 and Jonah 1: 7. Biblical use of randomization was a strong argument for its practical adoption for political and legal procedures in Christian Europe, notwithstanding moral objections to gambling and philosophical concerns about voluntarily relaxing the grip of reason and relinquishing important decisions to fate.

Late Medieval and Renaissance Italy (12th to 17th century) saw many political experiments in the use of sortition to randomly appoint citizens for legislative and executive positions or legal functions, the two most important prototypes being the Florentine *scrutiny* and the Venetian *brevia*; see Najemy (1982) and Wolfson (1899). In England and the USA sortition was (and still is) mostly used for jury duty appointment, with mandatory random selection officially documented since the 17th century; see Dowlen (2009). Nowadays, judicial systems of many western countries use randomization procedures

to select the jury and/or the court in charge of judging a legal case. Randomization procedures can also be used to choose who gets drafted into the army, see Fienberg (1971).

All of the aforementioned historical examples include the appointment of jurors and judges by complex procedures involving a mixture of the following aspects: (1) Pre-selection of a pool of able or qualified candidates; (2) Voting or election by secret ballot; (3) Proportional representation constraints (by district, tribe or family, corporate guild, professional specialty, etc.); and, finally, at some point(s) in the appointment process, (4) Sortition, i.e. random selection among possible candidates.

Based on surviving original manuscripts and secondary literature, the references in this section give detailed accounts of *how* sortition was used in several historical examples. However, the original sources give us very little explanation of *why* sortition was used. Archaic religious texts see the ritual use of randomization devices as a doorway for the manifestation of divine will, but the aforementioned civilizations quickly realized that sortition also brought important practical advantages. Nevertheless, the articulation of good rational arguments for the use of sortition and randomization had to wait the development of Mathematical Statistics in the 20th century. This is the topic of our next section.

1.2 *Decoupling, separation and haphazardness*

Why do we randomize? Legal systems of modern societies use sortition as a firewall, a technological barrier designed to prevent spurious communication of vested interests or illegitimate interference between parties in a legal case and agents of the legal system. Its purpose is to warrant impartial or non-partisan justice, literally, assure that no party (person, group of persons or social organizations), directly or indirectly involved or interested in a legal case, could illegally intervene or manipulate the due legal process.

Judges and jurors do not come to court as blank slates. They had a life history full of experiences that formed individual opinions, prejudices and idiosyncrasies. Hence, the opportunity given to a party to divert the case to a sympathetic judge or to select favourable jurors would constitute a virulent manipulation mechanism.

A similar situation is faced by statisticians and physicians when planning and conducting clinical trials, i.e. studies designed to test the performance of a newly proposed or alternative treatment for a given disease. For the sake of treatment comparison, patients taking part in the clinical trial have to be divided into a *control-group*, receiving an old treatment, a placebo, or no treatment whatsoever; and a *treatment-group*, receiving the newly proposed treatment. Giving the opportunity, participating patients would manifest their preferences, and try by all means possible to get their preferred choice of treatment. Wealthy and well-educated patients would be better informed, would make better choices, would be in better position to get their preferred treatment, but would also probably have better chances to overcome the disease and get well anyway. Hence, such *confounding effects* would ruin the very investigative purpose of the clinical trial; see Pearl (2004) and Stern (2008). Truly randomized experiments in empirical legal studies are far less frequent, requiring sophisticated statistical design and legal planning to reconcile scientific, moral and ethical constraints; for interesting examples, see Stern (2008).

As well known by contemporary scientists, randomization is the key instrument used in statistics to overcome the aforementioned conundrums related to avoid confounding effects and partial judgments. In this context, *decoupling* refers to randomization techniques aiming to eliminate systematic vulnerabilities to uncontrolled influences received from or exerted by participating agents; see Laureto *et al.* (2012, p. 195). Charles Sanders Peirce (1839–1914), Joseph Jastrow (1863–1944) and Ronald Aylmer Fisher (1890–1962) introduced the key concept of randomization in Mathematical Statistics, as

succinctly expressed by Judea Pearl in the following quotation; see also, Peirce and Jastrow (1884), Fisher (1926), Fisher (1935) and Stern (2008).

Fisher's great insight was that connecting the new link to a random coin flip 'guarantees' that the link we wish to break is actually broken. The reason is that a random coin is assumed to be unaffected by anything we can measure on macroscopic level - including, of course, a patient's socioeconomic background.¹

Pearl (2000) further develops the axiomatic study of related issues, and encodes the concepts of decoupling by randomization in the language of abstract inference diagrams, giving the all-important formal definition of *d-separation*. Furthermore, Dennis Lindley (1923–2013) uses the word *haphazardness* to distinguish clearly between the desired effect of *decoupling*, from randomness, the tool used to achieve it; see next quotation.

The Role of Randomization in Inference: We describe a possible allocation that the experimenter judges to be free of covariate interference as haphazard. Randomization may be a convenient way of producing a haphazard design. We argue that it is the haphazard nature, and not the randomization, that is important.²

Notwithstanding the important conceptual distinction highlighted by Lindley, randomization remains the most basic and fundamental tool used to achieve *decoupling* in modern statistics. Nevertheless, this conceptual distinction opens the possibility for improvement, by development of more efficient techniques used to render well-decoupled designs; see for example, Stern (2008), Laureto *et al.* (2012), Fossaluzza *et al.* (2015) and Lauretto *et al.* (2017). Finally, the consistent use of randomization techniques as well as the compliance with the implied protocols can be investigated and audited, to some extent, with tools provided by Mathematical Statistics and computational data analysis; this is the goal of Section 2.

1.3 *Managing randomization procedures: traps and pitfalls*

In light of the former discussion, the advantages and the necessity of randomized case assignment procedures should be evident. In the Brazilian judicial system, the benefits of randomized procedures are nowadays broadly accepted, but there are also many doubts concerning best implementation practices. There are misgivings about the possibility of abuse and corruption of the current procedures, but subjective perceptions and opinions vary concerning the widespread or systemic versus the punctual or isolated nature of possible malfunctions or anomalies. These and related topics have been an area of continuous research in the international literature; see Aney *et al.* (2017), Anderson (2016), Chilton and Levy (2015), Eisenberg *et al.* (2012), Levy (2017), Macfarlane (2014). In contrast, the same topics have rarely been the object of academic studies in Brazil, some possible reasons for this contrast are explained below.

Security through obscurity is a fallacy of paramount importance in the discussions concerning randomization procedures currently taking place at the Brazilian judicial system. This fallacy is recognized as the source of many false arguments justifying the use of secret designs and proprietary codes in the implementation of sensitive procedures (as those related to the subject of this article). In

¹ (Pearl, 2000, p. 348) also quoted in Stern (2008, p. 59).

² (Lindley, 1982, pp. 438–439), also quoted in Fossaluzza *et al.* (2015, p. 173).

contrast, *Security by design* relies on transparency and auditability as powerful aids for achieving robust and reliable security properties.

Security through obscurity and *Security by design* are standard expressions used in the software industry, but the issue predates the computer age and its interest extends far beyond the particularities of any specific area of application. At the London Great Exhibition of 1851, the locksmith Alfred Charles Hobbs gave sensational demonstrations on how to pick poorly designed locks, generating a heated controversy concerning the consequences of exposing the weaknesses of existing security systems. Nevertheless, his actions forced manufacturers to improve their designs, and only a few years later he was awarded a medal by the Institution of Civil Engineers for his contributions to the field. Sadly, mankind has a tendency to forget hard-learned lessons; see Stross (2006) for examples of absurd (or tragicomic) legal and police action by the Transportation Security Administration and the FBI against those who dare to expose existing flaws in airport security systems.

Already in 1883, the Dutch linguist Auguste Kerckhoffs published his basic principles for good cryptographic systems, including the following requirements: (1) *The system must be substantially, if not mathematically, undecipherable*; (2) *The system design must not require secrecy and it can be stolen by the enemy without causing troubles*; see Kerckhoffs (1883). These are considered the first clear formulation of modern principles of: (1) Security by design; and (2) No reliance on security by obscurity. Moreover, on one hand, transparency facilitates the early detection of design flaws or code errors, an effect known in computer science as *Linus's law*. On the other hand, security by obscurity requires system developers, managers and operators with special privileges, and those are often the entry point of malicious interventions. A system designed using the *principle of least privilege* eliminates such unnecessary vulnerabilities. For the principles discussed in this paragraph, see Pfleeger and Pfleeger (2015), Raymond (1999), and Wattenhofer (2017).

For well-documented case-studies concerning potential security hazards posed by internal malicious intervention in information systems that are obscurely specified see Courtois (2009), where the author concludes:

We must realize that the secrecy of a product specification poses a threat of a very large-scale electronic subversion. We need to have the courage to examine these questions and stop pretending that research in security is about discovering vulnerabilities that are always not intentional. . . . developers are also potential attackers for all such systems and trade/ industrial secrets should always be regarded as, potentially, a very major security breach.

Randomization procedures currently in use at the Brazilian judicial system are diverse, often kept under tight secrecy and, in some circumstances, even rely on manual or mechanical devices. Although such devices could in principle provide effective randomization, they are intrinsically opaque, hard to audit and inefficient for expeditious or real time processing.

As should be expected, all matters related to the use and implementation of randomization procedures in the judicial system are highly sensitive, and the consequences of changing established procedures are difficult to control: On the one hand, improvements of such procedures are often motivated by rather technical aspects concerning statistical soundness, automated tracing and auditability, computational efficiency, etc. On the other hand, the motivations, intentions and consequences of proposed improvements can be easily misunderstood. Any procedural improvement, change of functionality or revocation of privilege also has the potential of casting doubts or false impression about the potential or actual weaknesses and vulnerabilities of procedures currently in use, and such misunderstanding can generate further disruptions and liabilities; see Maurushat (2013) and Haber and Hibbert (2018).

Given the preceding considerations, a judicial system must manage such matters with great care, taking precautionary steps to avoid unnecessary misunderstandings that could delay or even halt the implementation of beneficial improvements. This is one of the goals of the research reported in this article. We seek to dissipate the fear of widespread systemic biases, and to point the limits of what can be accomplished by post-hoc statistical analyses, hence paving the way for subsequent steps in research and system reengineering.

2. Statistical modelling of case assignments

A model to assess randomness in case assignment may be applied to the following abstract scenario. Suppose there is a courthouse constituted of n judges and that new cases arrive daily and must be assigned to one of them. Also, assume that it is of interest to study the mechanism of case assignment of the courthouse, in order to examine the possibility of a tendency towards assigning cases to a specific judge, under particular circumstances. To meet this goal, the probability of a new case being assigned to each judge, under given circumstances, may be modelled.

In order to fit such a model, a sample of case assignments is needed. The sample must contain the number of cases that were assigned to each judge daily for a fixed period of time. Furthermore, it must also present the circumstances under which such case assignments occurred. These circumstances are represented by variables that summarize the daily conditions of the courthouse regarding possible sources of bias in the assignment process.

The caseload of the judges, the nature of the cases that have arrived (their class³ and their relation to cases already being handled by the court) and the availability of each judge to receive new cases are known mechanism that may alter the assignment probabilities, which can be added to the model. Judges with a greater caseload shall have a smaller probability of being assigned a new case. Cases that are related to a case already assigned to a judge may not partake in any random assignment process and be assigned directly to the judge handling its related case (related-case rule). Moreover, if a judge is not available to receive new cases, his or her probability of being assigned a case must be set to zero. Such conditions take particular forms at each court, and cannot be precisely described by any known general formulation.

A reliable characterization of the daily conditions of the courthouse is essential, for it is desirable to model any tendencies in the case assignment process, so that any bias still present on the fitted model is due to randomness or an unknown source. It is important that the circumstances under which a case was assigned be presented in the sample so that it may be incorporated to the model, in order to reliably model mechanisms that may alter the respective assignment probabilities.

The model chosen for the case under study is a Multinomial Logistic Regression (Hilbe, 2009), which is presented in details in the Appendix. This regression models the probability of a case being assigned to a judge as a non-linear function of numerical variables that represent the conditions under which the case is assigned. Other kinds of known sources of bias, which may not be represented by numerical variables, may also be incorporated to the model.

2.1 Case study of the Brazilian Supreme Court

The assignment of cases in the Brazilian Supreme Court (*Supremo Tribunal Federal*) has recently become a matter of public interest due to the publicity of the corruption cases related to the Car Wash

³ The cases a court is eligible to receive may be divided into classes, according to their purpose. For example, *Habeas Corpus* may be a class containing all habeas petitions received by the court.

operation⁴ (*Operação Lava Jato*) heard by the court. Various media outlets have made inquiries to the court about the assignment process in order to fully understand how the assignment of these important cases is performed, especially after the death of the rapporteur⁵ of the Car Wash operation in the court, which automatically received all cases related to the operation. However, the court has not fully attended these inquiries, limiting itself to say that the assignment process is *randomly* performed by a closed source algorithm taking into account the related-case rule, the caseloads and the number of cases assigned to a justice chair⁶ in the past. Therefore, the assignment process of the court is unknown to the public, and is not auditable by the involved parties.

The interest of the parties on the judge that receives a case is justified by the fact that he or she becomes its rapporteur and may take three actions regarding its trial. The rapporteur may reach a monocratic verdict; take the case to trial in front of a pre-defined judge panel⁷; or take the case to the plenary-court (*en banc*), which is formed by all 11 judges of the court. If the assigned judge chooses to take the case to trial, he or she must present a report detailing his or her decision about the merits of the case, on which the other judges may agree or dissent. The court decision about the case is made upon the majority agreeing or dissenting on the rapporteur opinion. Thus, the fact that a group of judges may be more sympathetic to some causes, and the willingness of the assigned judge to make a monocratic decision, or take the case to the judge panel or *en banc* trial, add to the importance of the assignment process, so it plays a role in the course of a case.

In view of this, we fit the abstract model outlined in the preceding section to a sample of case assignments of the Brazilian Supreme Court. The sample contains the number of cases of each class assigned to each justice chair of the court between 28 February 2008 and 10 July 2017. At the considered period of time, cases of 35 different classes were assigned to the 11 chairs of the court, although only 14 classes of cases are considered, as there were less than a thousand cases of each one of the other classes assigned over the 10-year period. Considering only the assignments of the top 14 classes, the sample amounts to 22 720 assignments.⁸

Table 1 presents the number of cases of each class that were assigned during the considered period of time to each justice chair, and Figure 1 presents the proportion of the cases of the class Special Appeal with Aggravation, which were assigned daily for each chair. The cases of this class are special appeals, which, as set forth in the constitution, should be heard by the Supreme Court. Furthermore, this is also the class with more cases assigned in the considered period.

Although thousands of cases are assigned to the court each year, they are in general not immediately tried and the court may take years to reach a decision. Indeed, there are cases from decades ago awaiting decision from the court. The reason for the great number of cases assigned to the court is the

⁴ A massive operation carried out by Brazilian Federal Police, which started targeting money laundering, and in the last four years has uncovered a variety of corruption schemes throughout the government and state-owned companies.

⁵ The rapporteur is the judge of the court responsible for analysing the case and presenting a report expressing his or her opinion about the matter at hand to the court, so that it can reach a decision about it. The rapporteur may also choose to reach a monocratic decision in some circumstances.

⁶ The Brazilian Supreme Court is formed by justice chairs, each one having a minister (judge). When a minister retires, his or her caseload is transferred to the newly appointed minister of his or her chair. Therefore, in order to follow the caseload of each minister over time, it is convenient to consider that the cases are assigned to the chair, not to the minister.

⁷ The Brazilian Supreme Court is formed by two fixed judge panels, of five judges each, which may rule about the cases assigned to its members. The court president is not part of any panel.

⁸ One for each day and class. If in a day there were assigned cases of m different classes, then there were m random samples in this day. Note that multiple cases are assigned in each assignment, so in the considered period of time 628 819 cases were assigned to the chairs. Each assignment may be modelled as a multinomial distribution with size in general greater than one, so that not necessarily all cases of a given class received in a day are assigned to a same judge.

TABLE 1 Number of cases assigned by class to each justice chair during the considered period of time⁹

Class	Justice Chair											Total
	1	2	3	4	5	6	7	8	9	10	11	
Precautionary Action (AC)	271 (10.5)	268 (10.4)	234 (9.1)	255 (9.9)	198 (7.7)	231 (9)	231 (9)	248 (9.7)	220 (8.6)	200 (7.8)	213 (8.3)	2569
Originating Civil Action (ACO)	214 (10.7)	172 (8.6)	204 (10.2)	213 (10.6)	143 (7.1)	180 (9)	156 (7.8)	189 (9.4)	146 (7.3)	190 (9.5)	196 (9.8)	2003
Direct Unconstitutionality Action (ADI)	212 (11.2)	159 (8.4)	147 (7.7)	230 (12.1)	157 (8.3)	160 (8.4)	134 (7.1)	197 (10.4)	171 (9)	156 (8.2)	177 (9.3)	1900
Bill of Review (AD)	16, 229 (10.3)	18, 353 (11.6)	14, 172 (9)	14, 644 (9.3)	9, 452 (6)	16, 226 (10.3)	17, 664 (11.2)	14, 605 (9.2)	14, 941 (9.5)	11, 445 (7.2)	10, 256 (6.5)	157 987
Special Appeal with Aggravation (ARE)	37, 605 (16.4)	16, 848 (7.4)	20, 278 (8.9)	20, 184 (8.8)	19, 936 (8.7)	17, 789 (7.8)	9, 569 (4.2)	19, 988 (8.7)	13, 063 (5.7)	33, 778 (14.7)	19, 998 (8.7)	229 036
Habeas Corpus (HC)	6, 535 (13.6)	3, 854 (8)	4, 529 (9.4)	4, 630 (9.6)	3, 436 (7.2)	4, 056 (8.4)	3, 349 (7)	4, 447 (9.3)	3, 708 (7.7)	5, 249 (10.9)	4, 249 (8.8)	48 042
Inquiry (Inq)	216 (10.8)	166 (8.3)	183 (9.2)	200 (10)	172 (8.6)	194 (9.7)	153 (7.7)	198 (9.9)	216 (10.8)	189 (9.5)	106 (5.3)	1993
Injunction Order (MI)	672 (11.3)	601 (10.1)	592 (10)	607 (10.2)	398 (6.7)	485 (8.2)	553 (9.3)	573 (9.6)	451 (7.6)	449 (7.6)	560 (9.4)	5941
Writ of Mandamus (MS)	696 (9.4)	597 (8)	670 (9)	693 (9.3)	638 (8.6)	1, 347 (18.2)	426 (5.7)	661 (8.9)	671 (9)	379 (5.1)	642 (8.7)	7420
Petition (Pet)	198 (8.6)	169 (7.3)	187 (8.1)	181 (7.8)	367 (15.9)	127 (5.5)	137 (5.9)	214 (9.3)	276 (11.9)	331 (14.3)	126 (5.4)	2313
Extraordinary Appeal (RE)	17, 021 (12.5)	13, 352 (9.8)	11, 518 (8.5)	12, 240 (9)	8, 891 (6.5)	13, 730 (10.1)	11, 181 (8.2)	12, 031 (8.8)	10, 759 (7.9)	17, 891 (13.1)	7, 565 (5.6)	136 179
Habeas Corpus Appeal (RHC)	599 (12.6)	398 (8.4)	472 (9.9)	451 (9.5)	393 (8.3)	403 (8.5)	326 (6.9)	459 (9.6)	276 (5.8)	543 (11.4)	439 (9.2)	4759
Writ of Mandamus Appeal (RMS)	169 (11.1)	145 (9.5)	168 (11.1)	153 (10.1)	117 (7.7)	125 (8.2)	137 (9)	161 (10.6)	96 (6.3)	101 (6.6)	147 (9.7)	1519
Complaint (Rcl)	3, 198 (13.4)	2, 007 (8.4)	2, 039 (8.6)	2, 537 (10.6)	1, 789 (7.5)	2, 018 (8.5)	1, 454 (6.1)	2, 190 (9.2)	1, 784 (7.5)	2, 702 (11.3)	2, 119 (8.9)	23 837
Other	288 (8.7)	419 (12.6)	295 (8.9)	362 (10.9)	271 (8.2)	321 (9.7)	296 (8.9)	300 (9)	242 (7.3)	246 (7.4)	281 (8.5)	3321
Total	84, 123 (13.4)	57, 508 (9.1)	55, 688 (8.9)	57, 580 (9.2)	46, 358 (7.4)	57, 392 (9.1)	45, 766 (7.3)	56, 461 (9)	47, 020 (7.5)	73, 849 (11.7)	47, 074 (7.5)	628 819

The percentage of cases assigned to each justice chair is presented within parentheses.

⁹ The name of the classes was translated from Portuguese. Their abbreviations are kept as they are in Portuguese.

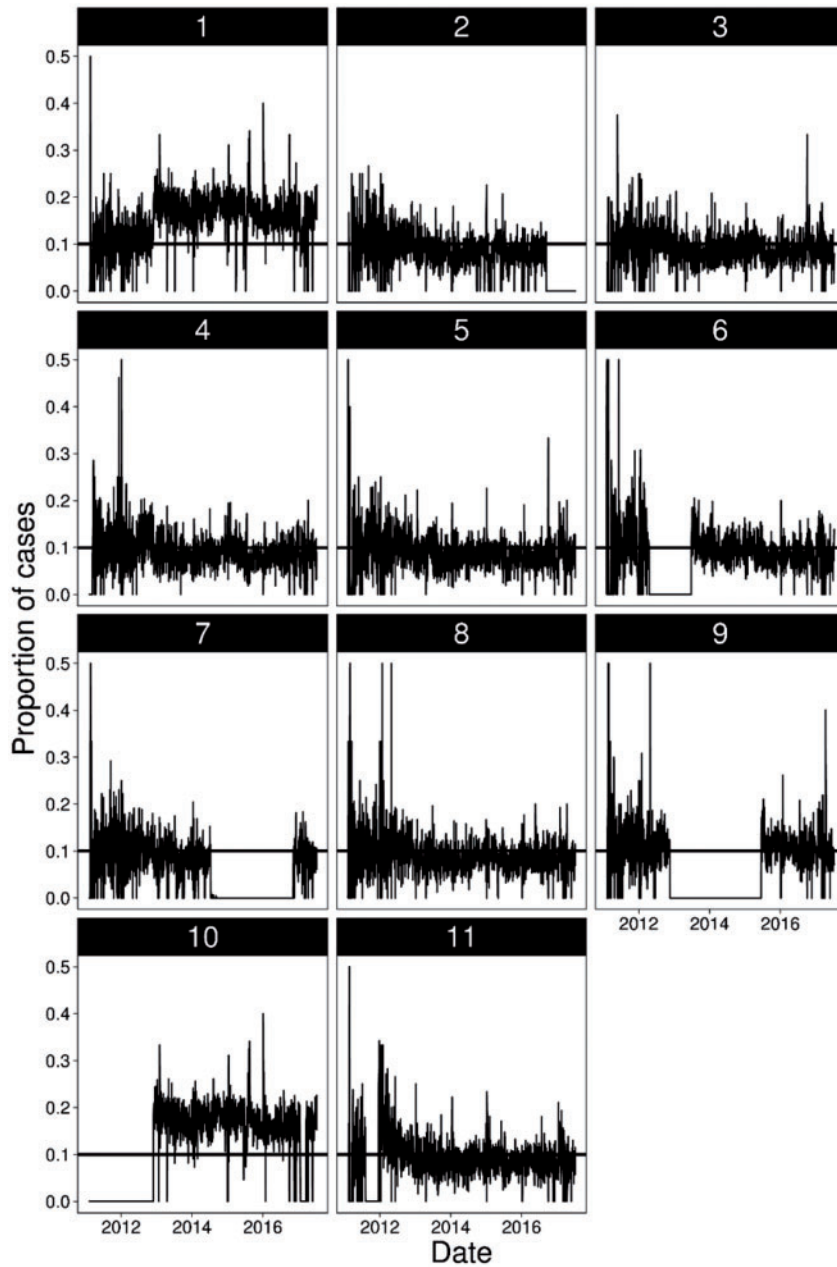


FIG. 1. Proportion of cases of the class Special Appeal with Aggravation assigned daily to each chair.¹⁰

¹⁰ The dotted line represents the 0.10 proportion, which is the proportion of cases that is expected to be assigned to each chair if the assignment process is performed randomly and uniformly among the chairs, i.e. each chair having the same probability to receive a new case, as no case of this class is assigned to the court president. This class of case was created in 2011 and the period of time in which a chair systematically receives no cases refers to when the chair was empty.

large range of cases under its jurisdiction, which contrasts with U.S. Supreme Court, for example, which has a narrower range of cases under its jurisdiction; grants plenary review to about 80 cases; and disposes of about a hundred cases without plenary review, each year.¹¹

It can be seen in Figure 1 that there are some justice chairs that systematically receive a greater proportion of cases of class Special Appeal with Aggravation for a period, although this may be explained by some kind of known mechanism and may not be evidence of lack of randomness in the case assignment process. In addition, we see in Table 1 that there are some chairs that have received a greater percentage of the cases, with special attention to chair 1 that has received 13.4% of all cases. However, such fluctuations do not directly indicate the existence of systematic bias because, at different times and circumstances, some chairs may not be available to receive a given case. The availability to receive cases is a circumstance of each assignment, which could be regarded by a regression model as a source of bias.

The case assignment process of the Brazilian Supreme Court has some known mechanisms that may change the assignment probabilities, which must be controlled. Indeed, although 11 justice chairs form the court, at least one of them is unavailable for receiving new cases, as one of the ministers is the court president and, therefore, occupies the presidency chair and does not receive new cases from the random assignment process.¹² Furthermore, there may be other empty chairs, for when a minister retires, his or her chair remains empty for a period of time, until a new minister is appointed by the President of the Republic and approved by the Senate. Indeed, these are known mechanism that may cause a chair to receive more or less cases, which are included in the model as *structural zeros* and not as numerical variables that describe the circumstances of the case assignment.

In each assignment, there are two types of *zeros* when no case is assigned to a chair: when it is not available to receive cases and when it is available, but receives no cases nonetheless. This fact is taken into account by the model, so that it becomes a Multinomial Logistic Regression Model with Structural Zeros, for which more details are given in the Appendix.

Another mechanism that may alter the assignment probabilities is the related-case rule, by which some cases are not randomly assigned, but are directly assigned to a specific chair that is handling related cases. The available sample does not distinguish the cases that were randomly assigned from the cases that were not and, therefore, this mechanism cannot be incorporated to the model. However, it should not affect all classes of cases. On the one hand, cases from classes as Habeas Corpus and Habeas Corpus Appeal are *prone* to be related to a case already being handled by the court. On the other hand, cases from the class Direct Unconstitutionality Action are believed to not be in general related to a case being handled by the court. Furthermore, the caseload of each chair at the moment of the assignment, which may alter the assignment probabilities (chairs with lesser caseload may have a greater probability of receiving a case), is another mechanism that may alter probabilities which cannot be added to the model, as it is not available at the sample.

The model for the Brazilian Supreme Court considers only one variable describing the circumstances under which each case assignment was performed. The variable considered in our analysis is the proportion of cases of each class that have been assigned to each chair since 2001 to the day of the assignment, as the court claims (without presenting any evidence) that, together with other quantities not available in the sample, as the caseloads of the chairs, the cases are assigned taking into account the

¹¹ See <https://www.supremecourt.gov/about/justicecaseload.aspx>.

¹² The court president receives cases that are exclusive to the presidency chair. Therefore, the president minister does not partake on the random assignment process.

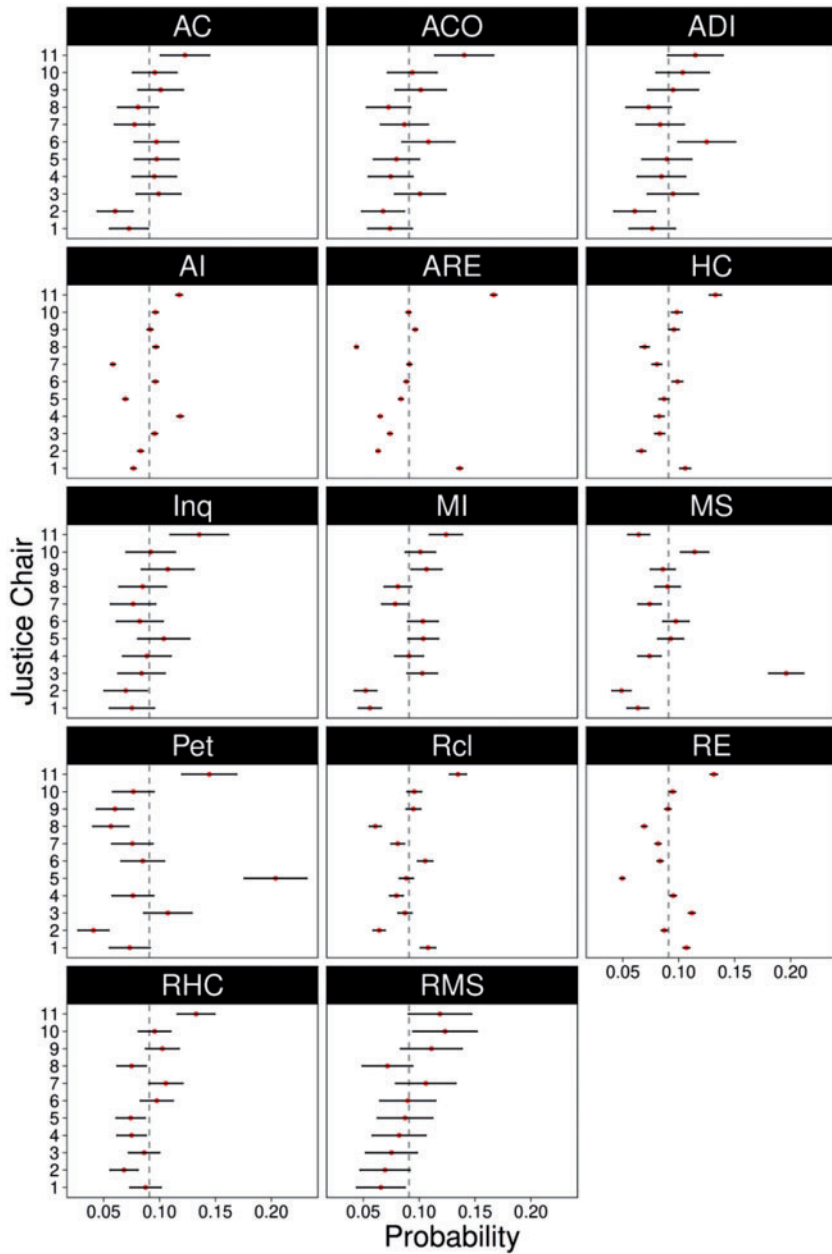


FIG. 2. Confidence intervals for the probabilities of a new case of a given class being assigned to each justice chair.¹³

¹³ The probabilities were estimated by the first Multinomial Logistic Regression model, which supposes that they depend only on the class of the case. The confidence of the intervals is 99% and they are corrected by the Bonferroni method (Neter et al., 1990, p. 164).

number of cases assigned to each chair in the past. However, the court does not disclose how the assignment algorithm takes into account this proportion when assigning cases, being very secretive about the procedures used for allocation and randomization.

In order to circumvent this missing information, the probability of a chair receiving a new case of a given class is modelled as a non-linear function of the proportion of cases of the class assigned to the chair since 2001 to the day of the assignment. The proportion assigned since 2001 is considered for this is the earliest date on which we could gather information about the number of cases assigned yearly to each chair. This proportion is updated daily according to the cases assigned at the previous day.

In our first model, we suppose that the probability of a case being assigned to a chair depends only on the class of the case. Figure 2 presents the confidence intervals for the probabilities, estimated by this model, of a new case of a given class being assigned to each chair. These probabilities are estimated taking into account the *structural zeros*, so they should differ from the percentages in Table 1. These probabilities should be compared with $1/11$ as, even though there has not been any assignment in the considered period of time in which all the 11 chairs were available to receive new cases, the proportion $1/11$ must be considered as the equilibrium, as the unavailable chairs have not been always the same on all assignments. Indeed, eleven different ministers from nine different chairs have occupied the presidency chair since 2001. Therefore, with this rotation of president ministers, we expect that, under the assumption that the cases are distributed craving to balance off the number of cases of each class assigned to the chairs, each chair should have received $1/11$ of the cases, and not $1/10$.

Furthermore, as the objective of the model is to detect the existence of systematic tendencies towards assigning cases of a specific class to a specific chair, it is necessary to suppose that all chairs are available to receive new cases, for otherwise we would have to choose *ad hoc* a chair to be unavailable to receive cases. Moreover, the mathematical properties of the fitted statistical model, i.e. the incorporation of the *structural zeros*, allow for the estimation of the probability of all 11 chairs to receive a new case of a given class, even though at most ten chairs are available for receiving cases at each assignment.

The confidence intervals of the model under consideration were estimated by the Multinomial Logistic Regression detailed in the Appendix. From the estimated intervals, we see that, for all classes, the probability of a chair receiving new cases is most often less than 0.15, with just a few probabilities greater than this threshold. Hence, there is *no* strong evidence indicating *large* biases on the assignment process. Nevertheless, for all classes, we reject that the probability is the same for all chairs, i.e. the difference between the probabilities of assignment are statistically significant, as not all confidence intervals contain the equilibrium $1/11$, for all classes. However, the difference is not *practically* significant on most of the classes, as the probabilities are homogeneous, though different.

In our second model, we suppose that the probability of a case being assigned to a chair depends on the class of the case and on the proportion of cases of the class assigned to the chair since 2001 to the day of the assignment. The probabilities estimated by this model, supposing that each chair received $1/11$ of the cases of each class assigned from 2001 to the day of the assignment, are presented in Figure 3, where they may be compared with $1/11$. We see that, apart from class AC, in which all intervals contain¹⁴ $1/11$, for all other classes we reject that the probability is the same for all chairs. However, yet again, the difference is not *practically* significant on most of the classes, as the probabilities are homogeneous, though different (see classes AI, Inq and RHC, for example).

¹⁴ So that we do not reject the hypothesis that the probability of a new case being assigned to a chair is equal for all chairs.

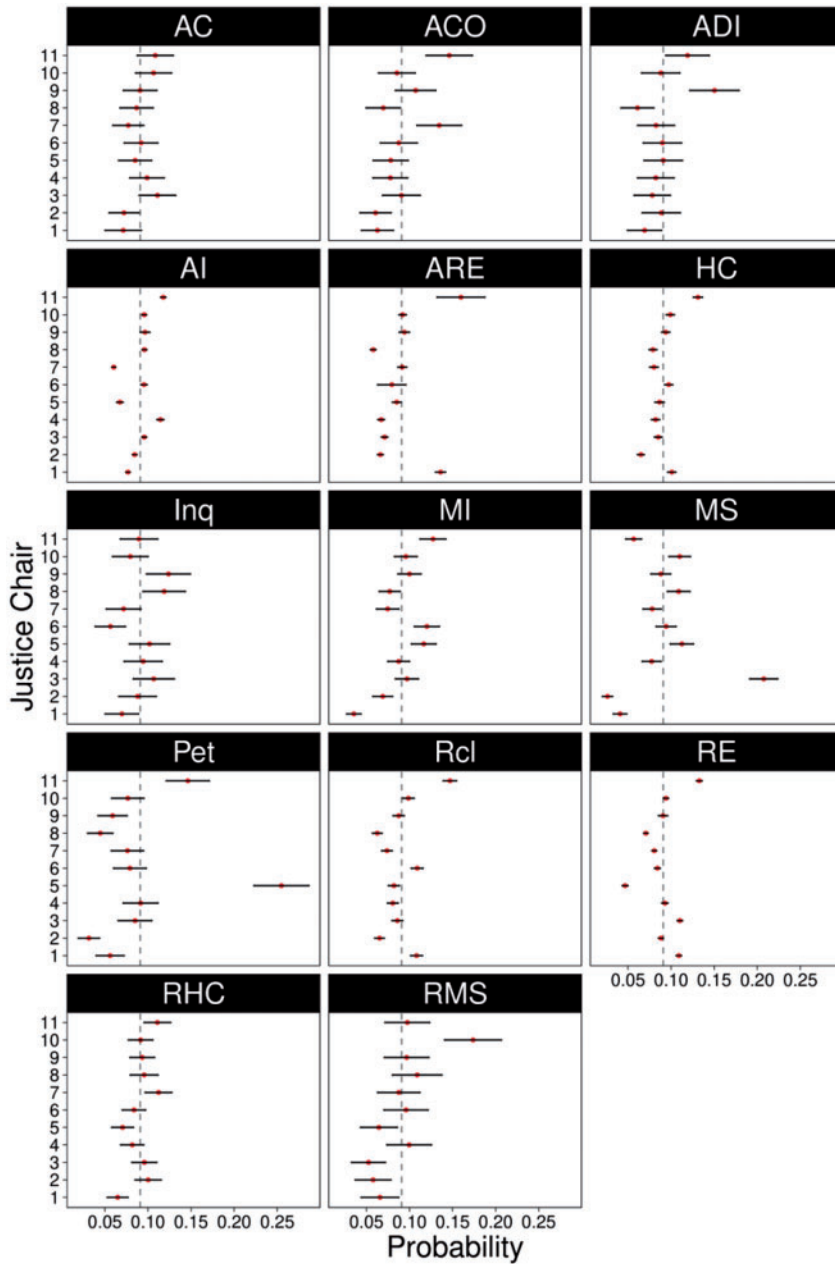


FIG. 3. Confidence intervals for the probabilities of a new case of a given class being assigned to each justice chair, supposing that each chair received 1/11 of the cases of the class assigned since 2001 to the day of the assignment.¹⁵

¹⁵ The probabilities were estimated by the second Multinomial Logistic Regression model, which supposes that they depend on the class of the cases and the proportion of cases of the class assigned to a chair since 2001 to the day of the assignment. The confidence of the intervals is 99% and they are also corrected by the Bonferroni method.

Nevertheless, there are classes in which some chairs have a higher probability of receiving a new case (see classes ADI, ARE, MS, Pet and RMS for example). In these classes, even though all chairs have been assigned the same proportion of cases, some of them have a higher probability of receiving a new case, what points to the existence of some tendency on the assignment process. This tendency may be either intrinsic to the assignment process, i.e. a flaw of the process, or not be part of the assignment process itself, but be rather caused by some variable that is not being taken into account by our model (as the related-case rule and the caseload of each chair, for example). Therefore, although we see that the probability of some chairs receiving a new case deviates from their expected values, we *cannot* conclude that the process itself is biased (at least not without further studies taking into account other factors that may alter the probabilities of assignment).

Furthermore, hard conclusions about the fairness of the assignment processes cannot be reached by post-hoc statistical analyses. A thorough analysis of the randomization and assignment procedures requires appropriate means and methods conceived at the design stage and built in the system as pre-defined functionalities, as discussed in Section 1.3. Only under such conditions would it be possible to determine for sure the existence (or not) of specific biases and flaws in the processes under scrutiny.

3. Final remarks and future research

The statistical analyses in the preceding sections give us a global assessment of the allocation process. Although we detect deviations from expected statistics, indicating the existence of systematic biases on the random allocation of judicial cases, we cannot make any conclusions about the fairness or appropriate randomization process of any individual judicial case. Such a conclusion cannot be reached by post-hoc statistical analysis of historical data. Furthermore, the systematic biases may be created by sources that, though unknown on the dataset at hand, are known to the court (as for example, the cases that fit the related-case rule and the caseloads of each chair).

Finally, given the juridical and social importance of the themata under scrutiny, we believe that it is important to develop randomization procedures in full compliance with the following desiderata: (a) Statistical soundness, see Hammersley and Handscomb (1964, Ch. 3), Haramoto *et al.* (2008), Knuth (1997, Ch. 3) and Ripley (1987, Ch. 2); (b) Procedural, cryptographical and computational security, see Boyar (1989), L'Ecuyer (2012) and Wattenhofer (2017); (c) Complete auditability, see Haber and Stornetta (1991); (d) Open-source programming; (e) Multiple hardware platform and operating system implementation; (f) User friendliness and transparency, see Parikh and Pauly (2012); (g) Flexibility and adaptability for the needs and requirements of multiple application areas (like, for example, clinical trials, selection of jury or judges in legal proceedings, and draft lotteries). These are important topics of current and further research.

Acknowledgements

We are grateful for the support received from FAPESP—the State of São Paulo Research Foundation (grants CEPID-CeMEAI 2013/07375-0 and CEPID-Shell-RCGI 2014/50279-4); and CNPq—the Brazilian National Council of Technological and Scientific Development (grant PQ 301206/2011-2). We thank for the support received from IME-USP, the Institute of Mathematics and Statistics of the University of São Paulo; and ABJ, the Brazilian Jurimetrics Association. We are also grateful for advice and comments received from Adilson Simonis, Rafael Bassi Stern and Marcelo Guedes Nunes.

We also thank the *Supremo Tribunal Federal* (Brazilian Supreme Court) which made available the data analysed in this article.

REFERENCES

- ANDERSON, J. (2016) “Judge Shopping in the Eastern District of Texas.” *Loyal U. Chi. LJ*, **48**, 539–560.
- ANEY, M. S., DAM, S., and KO, G. (2017) *Jobs for Justice(s): Corruption in the Supreme Court of India*. Technical Report, Singapore Management University, School of Economics.
- BOYAR, J. (1989) “Inferring Sequences Produced by Pseudo-Random Number Generators.” *Journal of the ACM*, **36**, 129–141.
- CHILTON, A. S., and LEVY, M. K. (2015) “Challenging the Randomness of Panel Assignment in the Federal Courts of Appeal.” *Cornell L. Rev.*, **101**, 1–52.
- COURTOIS, N. T. (2009) “The Dark Side of Security by Obscurity and Cloning Mifare Classic Rail and Building Passes, Anywhere, Anytime.” *4th International Conference on Security and Cryptography (SECRYPT 2009)*. INSTICC Press, 331–338.
- DOWLEN, O. (2009) *The Political Potential of Sortition: A Study of the Random Selection of Citizens for Public Office*. Charlottesville, VA: Imprint Academic.
- EISENBERG, T., FISHER, T., and ROSEN-ZVI, I. (2012) “Does the Judge Matter? Exploiting Random Assignment on a Court of Last Resort to Assess Judge and Case Selection Effects.” *J. Emp. Legal Stud.*, **9**, 246–290.
- FIENBERG, S. (1971) “Randomization and Social Affairs: The 1970 Draft Lottery.” *Science*, **171**, 255–261.
- FISHER, R. A. (1926) “Introduction to ‘The Arrangement of Field Experiments’”. *J. Minist. Agric. G B*, **33**, 503–513.
- FISHER, R. A. (1935) *The Design of Experiments*. Edinburgh: Oliver and Boyd.
- FOSSALUZA, V., DE SOUZA LAURETTO, M., DE BRAGANÇA PEREIRA, C. A., and STERN, J. M. (2015) “Combining Optimization and Randomization Approaches for the Design of Clinical Trials.” In *Interdisciplinary Bayesian Statistics*, 173–184. Cham: Springer.
- HABER, M. J., and HIBBERT, B. (2018) *Asset Attack Vectors: Building Effective Vulnerability Management Strategies to Protect Organizations*. Apress.
- HABER, S., and STORNETTA, W. (1991) “How to Time-stamp a Digital Document”. *J. Cryptol.*, **3**, 99–111.
- HAMMERSLEY, J. M., and HANDSCOMB, D. C. (1964) *Monte Carlo Methods*. London: Chapman and Hall.
- HANSEN, M. H. (1999) *The Athenian Democracy in the Age of Demosthenes*. London: Bristol Classical Press.
- HARAMOTO, H., MATSUMOTO, M., NISHIMURA, T., PANNETON, F., and L’ECUYER, P. (2008) “Efficient Jump Ahead for F2-Linear Random Number Generators”. *INFORMS J. Comput.*, **20**, 290–298.
- HEADLAM, J. W. (1933) *Election by Lot at Athens*. Cambridge: Cambridge University Press.
- HILBE, J. (2009) *Logistic Regression Models*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press.
- KERCKHOFFS, A. (1883) “La Cryptographie Militaire”. *J. des Sci Militaires*, vol. IX, part I: 5–38; part II: 161–191.
- KNUTH, D. E. (1997) *The Art of Computer Programming, vol. 2: Seminumerical Algorithms*. 3rd ed. Reading, MA: Addison-Wesley Longman Publ. Co., Inc.
- LAURETO, M. S., NAKANO, PEREIRA, F. C. A. B. and STERN, J. M. (2012) “Intentional Sampling by Goal Optimization with Decoupling by Stochastic Perturbation.” *AIP Conference Proceedings*. AIP, 189–201.
- LAURETTO, M. S., STERN, R. B., MORGAN, K. L., CLARK, M. H., and STERN, J. M. (2017) “Haphazard Intentional Allocation and Rerandomization to Improve Covariate Balance in Experiments.” *AIP Conference Proceedings*. AIP, 1–8.
- L’ECUYER, P. (2012) “Random Number Generation”. In *Handbook of Computational Statistics*, 35–71. Berlin, Heidelberg: Springer.
- LEVY, M. K. (2017) “Panel Assignment in the Federal Courts of Appeals.” *Cornell L. Rev.*, **103**, 65–118.
- LINDLEY, D. V. (1982) “The Role of Randomization in Inference.” In *PSA: Proceedings of the Biennial meeting of the philosophy of science association* (Vol. 1982, No. 2, pp. 431–446). Philosophy of Science Association.

- MACFARLANE, K. (2014) “The Danger of Nonrandom Case Assignment: How the SDNY’s Related Cases’ Rule Has Shaped the Evolution of the Stop-and-Frisk Law.” *Mich. J. Race Law*, **19**, 199–247.
- MAURUSHAT, A. (2013) *Disclosure of Security Vulnerabilities: Legal and Ethical Issues*. London: Springer Briefs in Cybersecurity.
- NAJEMY, J. M. (1982) *Corporatism and Consensus in Florentine Electoral Politics, 1280–1400*. Chapel Hill: Univ. North Carolina Press.
- NETER, J., WASSEMAN, W., and KUTNER, M. (1990) *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs*. Burr Ridge: R. D. Irwin.
- PARIKH, R., and PAULY, M. (2012) “What Is Social Software?” In van Eijck, *et al.*, 3–13. LNCS 7010, Berlin: Springer.
- PEARL, J. (2000) *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
- PEARL, J. (2004) “Simpson’s Paradox: An Anatomy.” Technical Report, Univ. of California at Los Angeles.
- PEIRCE, C. S., and JASTROW, J. (1884) “On Small Differences in Sensation.” *Mem Natl Acad Sci*, **3**, 75–83.
- PFLIEGER, C. P., and PFLIEGER, S. L. (2015) *Security in Computing*. Upper Saddle River, NJ: Prentice Hall.
- RAO, C. R. (1973) *Linear Statistical Inference and its Applications*. 2. New York: Wiley.
- RAYMOND, E. S. (1999) *The Cathedral and the Bazaar*. Sebastopol, CA: O’Reilly Media.
- RIPLEY, B. D. (1987) *Stochastic Simulation*. NY: Wiley.
- STAVELEY, E. S. (1972) *Greek and Roman Voting and Elections*. Ithaca, NY: Cornell University Press.
- STERN, J. M. (2008) “Decoupling, Sparsity, Randomization, and Objective Bayesian Inference.” *Cybernetics & Human Knowing*, **15**, 49–68.
- STERN, J. M. (2018) “Verstehen (causal/interpretative understanding), Erklären (law-governed description/prediction), and Empirical Legal Studies.” *Journal of Institutional and Theoretical Economics (JITE)*, **174**, 105–114.
- STROSS, R. (2006) Theater of the Absurd at the T.S.A. *The New York Times*, 17 December.
- WATTENHOFER, R. (2017) *Distributed Ledger Technology: The Science of Blockchain*. San Bernardino, CA: Inverted Forest.
- WOLFSON, A. M. (1899) “The Ballot and Other Forms of Voting in the Italian Communes.” *The American Historical Review*, **5**, 1–21.

Appendix: Technical details

In this Appendix, we present the details of the Multinomial Logistic Regression that is applied to the Brazilian Supreme Court. This regression model aims to model the logarithm of the odds in favour of a case being assigned to a justice chair over a so-called reference chair under given circumstances. The mathematical structure of the model imposes that a reference chair is fixed so that the probability of a case being assigned to a chair is compared with the probability of this same case being assigned to the reference chair through the respective odds. The logarithm of the odds is then modelled as a linear function of the numerical variables that represent the courthouse circumstances at the moment of the assignment. For more details about Logistic Regression models see Hilbe (2009).

Let J be the variable that indicates to which justice chair a given case is to be assigned. This variable takes values in $\{1, \dots, n\}$, as each one of the n chairs may be represented by a number. Also, let $(x_{j,1}, \dots, x_{j,k})$ be a set of numerical variables that represent the circumstances of chair j when the assignment occurred, and define (v_1, \dots, v_n) as $v_j = 1$ if chair $j \in \{1, \dots, n\}$ is available for receiving cases, and zero otherwise. Note that v_j introduces a structural zeros mechanism into the model, i.e. a mechanism that causes a chair to be empty and the number of cases assigned to it to be zero. Considering that the reference chair is the number 1 and that $v_1 = 1$, i.e. the reference chair

is always available for receiving cases, the Multinomial Logistic Regression with Structural Zeros may be written as

$$\begin{cases} \log\left(\frac{P(J = j|(x_{j,1}, \dots, x_{j,k}), (1, v_2, \dots, v_n))}{P(J = 1|(x_{j,1}, \dots, x_{j,k}), (1, v_2, \dots, v_n))}\right) = \beta_0^{(j)} + \beta_1^{(j)}x_{j,1} + \dots + \beta_k^{(j)}x_{j,k} & \text{if } v_j = 1 \\ P(J = j(x_{j,1}, \dots, x_{j,k}), (1, v_2, \dots, v_n)) = 0 & \text{if } v_j = 0 \end{cases} \quad (1)$$

for $j = 2, \dots, n$, in which P represents probability. Observe that in model (1) each chair has its own parameters. In addition, for simplification, we consider that all chairs have the same number k of parameters, which is not necessary.

In order to estimate the parameters $\beta_0^{(j)}, \dots, \beta_k^{(j)}, j \in \{2, \dots, n\}$, of the model by the Maximum Likelihood Method it is necessary to have a sample $(\mathbf{Y}, \mathbf{V}, \mathbf{X})$ of size m given by

$$\mathbf{V} = ((v_{1,1}, \dots, v_{1,n}), \dots, (v_{m,1}, \dots, v_{m,n}))$$

$$\mathbf{Y} = ((y_{1,1}, \dots, y_{1,n}), \dots, (y_{m,1}, \dots, y_{m,n}))$$

$$\mathbf{X} = ((\mathbf{x}_{1,1}, \dots, \mathbf{x}_{1,n}), \dots, (\mathbf{x}_{m,1}, \dots, \mathbf{x}_{m,n}))$$

$$\mathbf{x}_{s,j} = (x_{s,j,1}, \dots, x_{s,j,k}); s = 1, \dots, m; j = 1, \dots, n$$

in which $y_{s,j}$ is the number of cases assigned to chair j at assignment s ; $v_{s,j}$ equals one if chair j was available for receiving cases at assignment s and zero otherwise; and $x_{s,j}$ are the numerical variables that describe the circumstances of chair j at assignment s . The likelihood of the model may be written as

$$L(\boldsymbol{\beta}) \propto \prod_{s=1}^m \prod_{j=1}^n \left(\frac{v_{s,j} \exp(\beta_0^{(j)} + \beta_1^{(j)}x_{s,j,1} + \dots + \beta_k^{(j)}x_{s,j,k})}{1 + \sum_{l=2}^n v_{s,l} \exp(\beta_0^{(l)} + \beta_1^{(l)}x_{s,l,1} + \dots + \beta_k^{(l)}x_{s,l,k})} \right)^{y_{s,j} \times v_{s,j}}$$

with the convention that $\beta_0^{(1)} = \beta_1^{(1)} = \dots = \beta_k^{(1)} = 0$. The estimates $\widehat{\beta}_0^{(j)}, \dots, \widehat{\beta}_k^{(j)}, j \in \{2, \dots, n\}$, of the parameters are obtained by maximizing the likelihood presented above, and the probability of a new case being assigned to chair j , under new circumstances $\mathbf{X}_{m+1} = (\mathbf{x}_{m+1,1}, \dots, \mathbf{x}_{m+1,n})$ and $\mathbf{V}_{m+1} = (1, v_{m+1,2}, \dots, v_{m+1,n})$ is estimated by

$$P(J = j|\mathbf{X}_{m+1}, \mathbf{V}_{m+1}) = \frac{v_{m+1,j} \exp\left(\widehat{\beta}_0^{(j)} + \widehat{\beta}_1^{(j)}x_{m+1,j,1} + \dots + \widehat{\beta}_k^{(j)}x_{m+1,j,k}\right)}{1 + \sum_{l=2}^n v_{m+1,l} \exp\left(\widehat{\beta}_0^{(l)} + \widehat{\beta}_1^{(l)}x_{m+1,l,1} + \dots + \widehat{\beta}_k^{(l)}x_{m+1,l,k}\right)}$$

again with the convention that $\widehat{\beta}_0^{(1)} = \widehat{\beta}_1^{(1)} = \dots = \widehat{\beta}_k^{(1)} = 0$. Confidence intervals for these probabilities may also be obtained applying the properties of Maximum Likelihood Estimators (Rao, 1973, Chapter 5).

We will consider two distinct models for the Brazilian Supreme Court that are particular forms of (1). In the first model, we suppose that the probability depends only on the class of the case:

$$\begin{cases} \log\left(\frac{P(J = j | (x_{j,1}, \dots, x_{j,14}), (1, v_2, \dots, v_{11}))}{P(J = 1 | (x_{j,1}, \dots, x_{j,14}), (1, v_2, \dots, v_{11}))}\right) = \beta_1^{(j)}x_{j,1} + \dots + \beta_{14}^{(j)}x_{j,14} & \text{if } v_j = 1 \\ P(J = j | (x_{j,1}, \dots, x_{j,14}), (1, v_2, \dots, v_{11})) = 0 & \text{if } v_j = 0 \end{cases}$$

for $j = 2, \dots, 11$, in which $x_{j,i} = 1, i \in \{1, \dots, 14\}$,¹⁶ if the class of the process is the i -th and zero otherwise. This model has 140 parameters, one for each class and chair, but the first.

In the second model, we suppose that the logit (1) depends on the class of the case and on the proportion of cases assigned to the chair since 2001:

$$\begin{cases} \log\left(\frac{P(J = j | (x_{j,1}, \dots, x_{j,15}), (1, v_2, \dots, v_{11}))}{P(J = 1 | (x_{j,1}, \dots, x_{j,15}), (1, v_2, \dots, v_{11}))}\right) = \beta_1^{(j)}x_{j,1} + \dots + \beta_{15}^{(j)}x_{j,15} & \text{if } v_j = 1 \\ P(J = j | (x_{j,1}, \dots, x_{j,15}), (1, v_2, \dots, v_{11})) = 0 & \text{if } v_j = 0 \end{cases}$$

for $j = 2, \dots, 11$, in which $x_{j,i} = 1, i \in \{1, \dots, 14\}$, if the class of the process is the i -th and zero otherwise, and $x_{j,15}$ is the proportion of cases of the class assigned to chair j since 2001. There are 150 parameters in this model, one for each class and chair, but the first, adding up to 140 parameters, and one for each chair that refers to the proportion.

In both models, the *structural zeros* are differentiated from the *real zeros* so that we may estimate the probability of each one of the 11 chairs receiving a new case, even though there are never 11 chairs available to receive cases. Therefore, we should compare the estimated probabilities with $1/11$, as there are 11 of them and we seek to establish if they are equal.

¹⁶ We represent the classes by numbers as it makes the notation clearer. We may number the classes in an arbitrary order, without loss of generality.