Selecting Top Bureaucrats: Admission Exams and Performance in Brazil[*]

Ricardo Dahis [†]   Laura Schiavon [‡]   Thiago Scot [§]

June 7, 2021

Abstract

In the absence of strong incentives, public service delivery crucially depends on bureaucrat selection. Despite wide adoption by governments, it is unclear whether civil service examinations reliably select for job performance. We investigate this question focusing on state judges in Brazil. Exploring monthly data on judicial output and cross-court movement, we estimate that judges account for at least 23% of the observed variation in number of cases disposed. With novel data on admission examinations, we show that judges with higher grades perform better than lower-ranked peers. Our results suggest competitive examinations can be an effective way to screen candidates.

JEL Codes: D73, J45, M50

[*]We thank Ernesto Dal Bó, Alessandra Fenizia, Fred Finan, Guilherme Lambais, Guo Xu, Carlos Henrique Corseuil, and audiences at SBE, Berkeley Development Lunch and SIOE for helpful comments.

[†]Northwestern University. rdahis@u.northwestern.edu

[‡]Universidade Federal de Juiz de Fora. lauracschiavon@gmail.com

[§]UC Berkeley Haas School of Business. Corresponding author: thiago_scot@berkeley.edu

1 Introduction

Public employees play a key role in designing and delivering essential public services to development worldwide (Finan, Olken, & Pande, 2017). Recent studies have focused on the role that incentives and monitoring can play to improve the performance of government employees, particularly of frontline providers (Ashraf, Bandiera, & Jack, 2014; Khan, Khwaja, & Olken, 2019; Lavy, 2009). However, the impact of such tools is limited for the typical bureaucrat in developing countries whose career is often characterized by tenure benefits, absence of performance pay, and promotion based on seniority (Bertrand, Burgess, Chawla, & Xu, 2020). In the face of such low-powered incentives once hired, the issue of how to select bureaucrats becomes essential.

One widely used selection mechanism, particularly in some of the largest developing countries like Brazil, China and India, is competitive examinations. These may reduce corruption and patronage in hiring by political leaders (Colonnelli, Prem, & Teso, 2020; Brollo, Forquesato, & Gozzi, 2017; Weaver, 2021), but potentially at the expense of assessing candidates' soft and noncognitive skills (Hoffman & Tadelis, 2018; Hanna & Wang, 2017).[1] Further, it is an open empirical question, and a highly policy-relevant one, as to whether examinations reliably select the candidates who perform better on the job.

In this paper we study the role of exams in the selection of an important group of public sector employees in Brazil: state judges. Similar to the majority of civil servants in the country, judges are selected through highly competitive and mostly impersonal examinations, comprised of written and oral exams.

[1]These characteristics of public sector recruitment differ markedly from what is observed in the private sector, where managers and human resources officers have wide discretion in selecting employees and subjective assessments plays an important role through interviews, for example (Hoffman, Kahn, & Li, 2018).

Candidates are ranked based on their grades and top performers are offered jobs based on pre-specified number of available positions. Our estimates suggest that, within selected candidates, those ranking higher in exams are also high performers on the job as judges. In terms of magnitudes, we show that candidates that rank in the top quintile in their admission exam cohort dispose of approximately 20% more cases on a monthly basis than those in the bottom quintile.

The first step of our analysis is to estimate judge-level measures of performance. To do so, we leverage novel administrative data to construct a panel of judicial productivity at the judge-court-month level, covering the universe of state judges working in Brazil from 2009 through 2015. Across the 76 months encompassed by our data, judges often work in several different courts. This high level of mobility allows us to estimate a two-way fixed-effects model akin to those in the labor literature decomposing wage variation between worker and firm fixed-effects. We separately estimate judge and court fixed-effects, and show that judges are important in explaining the observed variation in output: using bias-adjusted estimates, individuals' fixed effects account for at least 23% of the variation in the number of cases disposed.

We focus on the number of cases disposed for two reasons. First, timely delivery of judicial decisions is critical in developing countries. At the current pace of case disposition, it would take Brazilian courts three years to clear the backlog, assuming no additional cases were initiated (Conselho Nacional de Justiça, 2018). Ponticelli & Alencar (2016) show that judicial timeliness matters for important economic outcomes. They explore differences in court congestion across Brazilian municipalities to show that a bankruptcy reform has larger effects on investment and financial access of firms located in district with more efficient courts. Second, the speed with which judges dispose of cases is considered an important indicator of performance by the judicial branch in Brazil and is used, along with other considerations, to define promotions throughout the career of judges.

Yet, theory suggests that if the quantity of cases disposed is easily observable but quality is not, judges might divert efforts into the observable dimension of performance (e.g. speed), possibly to the detriment of quality (Holmstrom & Milgrom, 1991). One implication of that hypothesis is that fast judges might sacrifice important inputs in the "production process" of case disposition in order to increase output quantity. We leverage our detailed microdata to show that this is not the case for one important input for court decisions: the number of hearings held by judges. We re-estimate our two-way fixed-effects model using hearings as the dependent variable and show a strong, positive correlation between judges fixed-effects in both models. This show that faster judges are not decreasing the number of hearings, one important input for high-quality case decision.[2]

Next, we examine whether judges highly ranked on entrance exam actually perform better on the job. We collect novel data on admission exams for over 25% of all state judges working in Brazil during the period covered by our productivity dataset, including their final rankings and grades. Our results suggest a positive and strong correlation between admission exam and on-the-job performances: we estimate that, within cohorts, being ranked in the top quintile of one's admission examination is correlated with a 0.2 s.d. increase

[2]As we will discuss in Section 3, one natural measure of quality of case disposition would be the likelihood of case reversals in higher courts. However, there is no such systematized data for Brazil in the time period we study. Another common measure of judicial quality, particularly in common law countries, are citations of judicial decisions (Landes, Lessig, & Solimine, 1998). These are not as commmon in civil law countries, such as Brazil, particularly in first instances courts such as the ones we study.

in estimated FE when compared to the bottom quintile. This result is robust to different measures of performance both in exams and on the job; the results are also robust when excluding the top and bottom candidates in each cohort.

Taken as a whole, our results suggest that admission exams are able to rank candidates in a way consistent with their future performance on the job.[3] In order to make progress in understanding which dimensions of the exams are most relevant for future performance, we restrict our sample to a subset of judges for which we can break-down final grades in each of the recruitment phases and consider whether achievement in any of the specific exams is particularly predictive of performance on the bench. Across different specifications, grades on the Judicial Decision Writing exam, where candidates are given a hypothetical case and asked to produce a decision, are the strongest predictors of performance. While these correlations should be interpreted with caution, they suggest that the use of impersonal examinations to screen candidates might be particularly efficient if focused on "practical" exams that mimic the situations faced by employees on the job.

Our paper makes contributions to three strands of literature. First, we add to the recent literature on the selection of workers in the public sector, mostly focused on the role of wages and career benefits in selection (Dal Bó, Finan, & Rossi, 2013; Deserranno, 2019; Ashraf, Bandiera, Davenport, & Lee, 2020). We study the role of impersonal admission examinations, aimed precisely at avoiding the kind of patronage documented in contexts as different as colonial

[3]The implications of sidelining any subjective assessments of candidates' qualities for job performance are not obvious. If knowledge about objective exam content is the crucial requirement to perform well, or if subjective traits that predict exam performance are also correlated with service delivery capacity, then objective recruitment strategies might be simultaneously effective and impartial. If certain subjective characteristics are very relevant to perform well on the job but hard to capture on objective admission examinations, nonetheless, these recruitment strategies are maintaining impartiality at the expense of accuracy.

governors in the British Empire (Xu, 2018) and public officials hired at the discretion of newly elected politicians in Brazil (Colonnelli, Prem, & Teso, 2020). In the context of high courts in Pakistan, Mehmood (2020) also documents that politically appointed judges are more likely to rule in favor of the government. However, the use of discretion when selecting officials need not lead to negative selection of providers. In an extreme example, Weaver (2021) shows that the selection of supervisors of community health workers by outright bribery leads to high quality workers being hired, since wealth and performance are strongly positively correlated. To the best of our knowledge, our paper is the first to show that competitive examinations successfully select the most efficient magistrates and to quantify the importance of judges to court efficiency. We contribute to a nascent literature documenting that performance on impersonal admission exams, the dominant screening process for public sector employees in several countries, is informative about performance on the job.[4]

Our research also adds to efforts of measuring the role of bureaucrats in determining public sector performance (Finan, Olken, & Pande, 2017). While the relevance of front-line service providers like teachers (Chetty, Friedman, & Rockoff, 2014; Muralidharan & Sundararaman, 2011; Duflo, Dupas, & Kremer, 2015; Jacob, Rockoff, Taylor, Lindy, & Rosen, 2018) and community health workers (Deserranno, 2019; Ashraf, Bandiera, Davenport, & Lee, 2020; Weaver, 2021; Dal Bó, Finan, & Rossi, 2013) have been extensively discussed, the role of other decision-makers in the public sector bureaucracy has only recently garnered more attention. Our empirical strategy, exploring the movement of judges between courts to identify individual fixed-effects, is particularly related

[4] Aman-Rana (2020) documents that public officials ranked at the top 10% of their admission cohorts in Punjab, Pakistan, also collect more taxes. Bertrand, Burgess, Chawla, & Xu (2020) documents a positive correlation between admission exam rankings and performance measured by 360 degree evaluations of IAS officers in India.

to the work of Best, Hjort, & Szakonyi (2019) on the role of procurement officers in Russia in explaining price dispersion in public purchases, and of Fenizia (2020) on how managers of Social Security offices in Italy explain variation in productivity.

Lastly, we contribute with new evidence about the determinants of judicial efficiency in the developing world. Research in Pakistan (Chemin, 2009), Senegal (Kondylis & Stein, 2021) and Mexico (Sadka, Seira, & Woodruff, 2018) has shown that judicial reforms aimed at simplifying procedures and speeding up the disposition of cases can be effective. Kondylis & Stein (2021), in particular, collect rich data at the case-level and show that higher speed in commercial courts in Senegal does not seem to affect the quality of decisions. The effects of judicial reforms more broadly also depend on the capacity of courts to deliver timely decisions, as shown in Ponticelli & Alencar (2016) and Rao (2020). To the best of our knowledge, our paper is the first to perform a two-way fixed effects decomposition to quantify the importance of judges to court efficiency and, further, to show that competitive examinations successfully selects the most efficient magistrates.

The remainder of this paper is organized as follows. Section 2 describes the structure of Brazilian courts and the admission process for judges. Section 3 presents the data used, provides descriptive statistics and explains how we obtain the sample used when estimating the two-way fixed-effects model. Section 4 describes our empirical model, identification and estimation procedures. The main results are presented in Section 5, while Section 6 concludes and discusses avenues for future research.

2 Institutional Context

2.1 Brazilian Courts

The Brazilian Judiciary is comprised of five branches: State, Federal, Electoral, Labor and Military Courts. This paper uses data exclusively from State courts, which cover all cases that are not specifically under the competency of the other branches (that is, State courts have residual judicial competency). The majority of criminal and civil cases fall under the competence of State courts: in 2017, over 60% of all cases in the Judiciary were allocated to the first instance of these courts (Conselho Nacional de Justiça, 2018).

Each of the 27 Brazilian federative units (26 states plus the federal district) is responsible for establishing and organizing the state courts. Within each state, the main administrative unit of the state justice are the judicial districts (*comarcas*), which encompass one or more municipalities. Judicial districts are mainly divided in three administrative levels, related to the underlying demand for judicial services: *first level* districts are located in rural or less urbanized municipalities and contain a single court of general competency (i.e. it covers all types of cases); *second level* districts are located in municipalities with smaller cities and encompass specialized courts, often separate Civil and Criminal courts; while *third level* districts encompass the state capital and possibly other large cities, and include several specialized courts.

Court congestion is considered a serious impediment to the efficient application of justice in Brazil (Ponticelli & Alencar, 2016): at the state level, there were over 60 million cases allocated to courts in 2017. If no more cases entered the justice system and current levels of productivity were held constant, it would take almost three years to clear the backlog (Conselho Nacional de

Justiça, 2018). While overall congestion is very high, there exists a large dispersion among judicial districts not fully explained by simple regional differences: Schiavon (2017) shows that the dispersion of several congestion and performance measures is larger within states than between states, highlighting the relevance of local determinants in explaining variation in performance.

The importance of timely decisions by courts and the challenges faced by the Brazilian Judiciary in that regard have not escaped the attention of policy-makers and legislators. For example, the 2004 Constitutional Amendment that created the National Justice Council, among several other sweeping changes to the organization of the Judiciary, also included specific language requiring that the promotion of judges take into account "objective criteria of productivity".[5] During the launch of the Open Justice System, in 2008, a Supreme Court Justice praised the tool as a way to "improve the management of justice and decrease the slowness of decisions".[6]

2.2 Selection of judges through competitive examinations

The broad rules for recruitment of judges are determined by Article 93 of the Brazilian Constitution. It states that all judges should be selected through public examinations (*Concursos Públicos*); since 2004, a Constitutional Amendment also institutes the requirement of three years of professional judicial experience. Judgeship admission exams are highly competitive (the ratio of candidates per position often exceeds 100), not only due to the prestige of the position but also likely because it is among the highest paid in the public sector.[7]

[5] Constitutional Amendment n.45, December 30th 2004.

[6] Available at https://www.estadao.com.br/noticias/geral,para-stf-criticas-ao-justica-aberta-sao-infundadas,195051. Accessed 08/10/2020.

[7] While the Constitution establishes that wages in the public sector should not surpass those of Supreme Court Justices, set at BRL 33,763 (approximately USD 8,500) per month until

Until 2009, federal law did not detail the content or structure of these examinations, which were left to the discretion of State courts. Since then, the structure of exams, including minimum content, qualification thresholds in each phase and weights for final ranking were harmonized.[8]

In practice, nonetheless, the overall structure of these examinations was already rather similar across states. Upon deciding to hire new judges, courts publicly announce the beginning of a *Concurso* through a call for applications, informing how many positions are available and details about the timeline, content and structure of examinations. Potential candidates must enroll online and pay a fee[9] in order to be considered eligible for the position.

Most examinations are comprised of four phases: Multiple Choice, Written, Judicial Decision Writing and Oral Exams. The first phase is often a *Multiple Choice Exam* covering a wide range of topics: constitutional, civil, criminal, commercial, administrative and family law are among the themes covered. Like the other three phases, this exam is both qualifying, meaning that candidates with performance below a certain threshold are immediately eliminated, and classifying, since the grade received is a component of the weighted average that determines the final ranking of candidates. Those approved in the Multiple Choice phase are invited to take a *Written Examination* that encompass the same topics mentioned before and also topics such as the sociology and philosophy of law, and ethics. The following phase is a *Judicial Decision Writ-*

2018, the vast majority of judges receive total compensation significantly higher than that due to fringe benefits not included in the above mentioned rule. In fact, in Table A1 we compare average nominal wages for judges and various other occupational categories between 2003 and 2019. We find that in 2019 judges' wages were significantly higher than federal government (257%), private sector (1502%), and other groups. The only comparable category is attorneys, with an average monthly wage of BRL 36,768 in 2019.

[8] National Justice Council Resolution 75 05/12/2009

[9] Resolution 75 determines that the fee can be no greater than 1% of the gross monthly salary for the position, which amounts to about BRL 300, or USD 75.

ing, also called a "practical exam", where candidates are given a hypothetical case and asked to write a judicial decision. In most cases this phase includes two decisions, one in criminal and another in civil law. The last qualifying phase is the *Oral Exam*. Candidates are randomly assigned a topic from a predetermined list 24 hours before their examination, and are then expected to answer questions from a committee composed of other judges and attorneys.

Candidates approved in the *Oral Exam* are eligible to be in the final ranking that defines hiring. Other than the grades in each of the previous phases, the final score also includes the so called *Titles Exam* (*Exame de Títulos*), additional points for career and academic achievements, such as previous judgeship, professorship or advanced degree in Law, and publications in Law journals. Since 2009, the weights that define the final score are the following: 10% Multiple Choice, 30% Written Exam, 30% Judicial Decision Writing, 20% Oral Exam and 10% Title Exam. Candidates are ranked according to their final grades and the top performers are offered jobs according to the number of vacancies available.

It is worth briefly mentioning that these recruitment processes are considered transparent and free from undue influence of judges or politicians, unlike the hiring for other public sector positions which are heavily influenced by patronage practices (Colonnelli, Prem, & Teso, 2020; Brollo, Forquesato, & Gozzi, 2017; Barbosa & Ferreira, 2019). First, every step of the process is highly publicized: grades and lists of approved candidates in each phase are made public, as are the content of each exam. The composition of the committee writing exams and participating in the Oral tests is also made public at the beginning of the *Concurso*, and candidates can appeal for the exclusion of members (e.g. due to family ties of members to any candidate).[10] Second, any deviation from

[10]Graders are blind to the identity of exam-takers in the Multiple Choice, Written and Judicial Decision Writing phases. In the Oral exam candidates present in front of a committee

the stipulated rules regarding exams often leads candidates to sue and annul specific phases or even the entire recruitment process. In 2014, for example, candidates in the state of Para successfully sued to have their Oral exams annulled after being asked only three questions during the evaluation, while the call for applications determined four questions.[11] In that sense, the selection process of judges is believed to be broadly free from corruption and reflect the performance of candidates.[12]

2.3 Judges' careers and allocation of cases

Once hired, judges are considered "substitute judges" for a period of two years, a probational stage before gaining tenure protection[13]. After this period judges can only be dismissed if convicted of crimes or found guilty of administrative infractions. In practice, this is very rare: between 2005 and 2017, only 82 judges in the entire Judicial branch were punished by the National Justice Council, and 53 of those received "mandatory retirement", meaning they were excluded from judgeship but kept receiving salaries.[14]

As previously discussed, judicial districts are divide in three levels: first, second and third. This administrative division is directly linked to judges' ca-

---

and therefore identities are known to graders.

[11] Available at: http://cnj.jus.br/noticias/cnj/61524-cnj-anula-prova-oral-de-concurso-para-ingresso-na-magistratura-do-tjpa. Accessed 08/10/2020.

[12] Exceptions do exist. In 2010 the Supreme Court ruled in favor of candidates asking for the annulment of a *Concurso* in the state of Minas Gerais, arguing that more candidates were accepted to the second phase of the process than initially announced. Two daughters of an appellate judge from that state were benefited (Available at: https://www1.folha.uol.com.br/fsp/poder/po2606201029.htm. Accessed at 08/10/2020)

[13] There are no aggregate statistics on the share of judges dismissed in the probational stage, but conversations with members of the judiciary suggest these are extremely rare: very few judges nationwide are denied tenure.

[14] Available at: https://g1.globo.com/politica/noticia/cnj-puniu-82-juizes-no-brasil-desde-2005-53-deles-continuam-recebendo-salario.ghtml. Accessed 08/10/2020

reers. Substitute judges are often allocated to first level districts, where they work in general courts, dealing with all types of judicial cases. Promotion means being reallocated to a higher level district, which comes with wage increases. After achieving third level status, judges can be promoted to appellate courts, meaning they leave the first instance (and our database).

The allocation of magistrates to judicial districts is governed by the Constitution. One of the core principles considered is that of the *immovability* of judges, meaning that judges cannot be transferred from their assigned district without their consent.[15] This should make clear that in no way we argue that the movement of judges between courts is quasi-random: judges must assent to being transferred between districts.[16] The identification of judges' fixed-effect, therefore, does not rely on exogenous allocation of judges to courts; our model allows for rich patterns of endogenous matching between judges and courts, and as discussed in detail below only rules out specific types of matches.

Finally, it is important to note that the distribution of cases among judges is as good as random. In judicial districts where there is only one court, cases will be randomly assigned to one judge in that court. For larger districts that encompass specialized courts, cases will be assigned to the proper court depending on their topics or, in the case where more than one relevant court exists, randomly assigned to one of the courts and a judge.[17] That should allay concerns that, within courts, different judges will have distinct composition of

[15]The principle is supposed to protect the public against the undue influence of politicians who might want to exclude a judge from judging a case in which they have interest, for example, but it is also a clear benefit to judges who are only reassigned if they so decide.

[16]In our empirical exercises we explore the movement of judges between courts, which can occur in courts within a same district or between different districts. The immovability principle applies to the latter.

[17]The method used to implement the random allocation is described as an electronic platform that randomly distributes cases. Unfortunately, we have no access to case-level data to check whether process characteristics are similar across judges assigned to the same court.

cases, making it harder to interpret the number of cases disposed.

3 Data and descriptive statistics

3.1 Data sources

This paper uses three main data sources: information on monthly output of judges and courts provided by the Open Justice System, admission exam's rankings collected from several different sources and administrative data on formal employment (*RAIS*).

All data on judicial performance come from the Open Justice System (*Sistema Justiça Aberta*), an online platform maintained by the National Justice Council (*Conselho Nacional de Justiça – CNJ*).[18] The Open Justice System provides monthly information, supplied by courts, on a range of quantitative outcomes at both the court and judge levels, including the number of cases disposed, hearings and intermediary decisions.

We construct a panel at the judge-court-month level: each observation is a vector of quantitative outcomes related to a judge working on a given court in a specific month. The dataset covers the universe of state judges working on first instance courts (i.e. excluding appeal level) from January 2009 through April 2015,[19] and we construct unique IDs using judges' full names to track the movement of magistrates between courts over time.

Our preferred measure of judges' performance is the number of *cases dis-*

[18]The National Justice Council was created in 2004, through a Constitutional Amendment, with the goals of improving the efficiency and transparency of the Brazilian judiciary. Among other tasks, the Council receives complains from citizens against members of the judiciary, promotes tools to improve the efficient functioning of the courts and publishes data on judicial efficiency.

[19]The Open Justice System was extinguished in 2015, and replaced by a new system later that year. The new dataset, nonetheless, is not strictly comparable to the data we use.

posed on merits in a given court and month. This refers to the number of cases for which the judge has issued a final decision based on the merits of the process, i.e., it excludes any cases terminated for procedural reasons or by a decision of one of the parts to withdraw. The decision to exclude cases decided for other reason rather than on the merits is an attempt to reduce the possible noise introduced by considering cases that are concluded for reasons unrelated to the judges' efforts.

Figure 1 presents preliminary evidence on the dispersion of judges' output. We plot the histogram of average monthly number of cases disposed at the judge level, across the entire panel. There is remarkable dispersion: judges on the 10th percentile of the distribution dispose of 11 cases on the merits on average, while judges on the 90th percentile dispose of 8 times as many. This dispersion reflects several forces, including potentially judges' efforts and capacity to make the court function efficiently, but also levels of demand in different courts.[20] We will attempt to disentangle these determinants with our empirical model.

The data on admission examinations (*Concursos*) was collected from a variety of sources. Results of *Concursos* are mandated to be public and are often published in PDF format either on the website of the State courts hiring or by the private institutions hired by the state to manage and implement the recruitment process. We scraped these document and constructed a database of candidates' exam performance. We have collected data for 79 recruitment waves for the selection of Judges from 24 different states in the period 2000-2013. For all these examinations the final ranking of approved candidates is available; for a subset of them, we also collect the final grade and the indi-

[20]Moreover, it is likely not driven by variation in backlogs across courts because the judicial system as a whole faces large excess demand in cases (Ponticelli & Alencar, 2016).

vidual grades in all phases of the exam.[21] We then match judges' grades with performance using full names and state of judgeship.[22] We are able to match over 2,800 judges observed in the productivity dataset to their admission examination performance, covering over 25% of all state judges working at some point between 2009 and 2015.

One additional data source used to recover information from judges' careers is administrative matched employer-employee data from *RAIS* (*Relação Anual de Informações Sociais*) for the period 1995-2017. We use unique individual identification numbers (*CPF – Cadastro de Pessoa Física*) to follow individuals over the years, and then match workers at RAIS to the judge productivity database using full names. We are able to uniquely match approximately 9,400 judges between the two datasets, or 80% of all judges observed in the productivity dataset in the period 2009-2015. We use RAIS data to obtain information on judges' gender, education, formal labor market experience, experience as judges and wages (prior to and during judgeship).

3.2 Sample Selection and Descriptive Statistics

The complete productivity dataset comprises close to one million observations at the judge-court-month level. Here we briefly describe the steps to obtain the sample used to estimate the two-way fixed effects model.

Despite the efforts by CNJ to assure quality of the performance data reported, there are clear instances of incorrect entries, such as hundreds of thousands of cases disposed by a single judge in a month. We therefore trim all

[21] Recent recruitment processes always include results for all the phases of the examinations. As we go back in time, nonetheless, the information available online becomes scanter. The minimal information we require to include an examination in the dataset is the nominal list of approved candidates and their final rankings.

[22] We benefit from the fact that Brazilians often hold several last names, which makes precise matches on names feasible.

performance measures at the 99th percentile.[23] We also observe a very high frequency of "mobility" in the raw data, as presented in Column (1) of Table 1: on average judges work in 11 different courts throughout the period. Yet, a large proportion of these judge-court matches is clearly transitory: for over half of the judge-court pairs the duration of the match is a single month.[24] In our estimates, we drop any judge-court *spells* with a duration of less than three months. Our final sample includes approximately 730,000 observations[25].

Table 1, Column (1) presents descriptive statistics for the full panel, while Column (2) refers to the sample used to estimate the two-way fixed-effects model[26]. There are 10,479 different judges and 9,048 courts in the estimating sample. Unlike other settings where there is limited mobility explored to estimate two-way fixed-effects models, that is clearly not a problem in our context: almost 80% of judges work in at least two different court throughout the period, and only in about 10% of courts we observe a single judge in the entire period[27].

The first panel of Table 1 characterizes judges in the sample. While the panel covers a 76-month period, the median judge is observed working on any court in 56 months. Very few judges work in one single court throughout these five years: on average judges work in four different courts. While judges might work in more than one court on a given month, that is the exception rather than the rule: for over half of judge-month observations, magistrates are working in a single court. Once we drop short-lived judge-court matches, the average

[23]For case disposition, the 99th percentile is 350 cases disposed by a judge in a single month.
[24]Informal conversations with judges suggest that it is common for judges work in different courts when colleagues are on vacation or sick leave.
[25]In Appendix A we present results using alternative sample definitions.
[26]Detailed descriptive statistics for the baseline sample are presented in Table ??.
[27]Using matched employer-employee data from Italy, Kline et al. (2020) report that in their largest connected set 21% of workers are movers.

number of months for any match is over 16 months and the median 9 months, meaning that we have several repeated observations of output for each pair, reducing the noise inherent in a measure like the number of cases disposed.

Details about courts are presented in panel B of Table 1. While in any given month most courts are likely to be staffed by a single judge, their rotation means that, throughout the period, the average number of different judges working in a court is almost five, or one per year. We also present the breakdown of courts by category, according to the type of cases they hear. General courts, located in first level districts and handling all types of cases, comprise around 20% of the sample. The remaining courts are specialized on specific cases, such as Civil (22%), Criminal (16%), Small-stakes (18%) and Family Law (10%). As one might expect, courts dealing with different topics present systematic differences in the number of cases disposed on a monthly basis. Figure 2 presents the average monthly number of cases disposed by judges, in each type of court. On one extreme, judges in criminal courts typically dispose of only 20 cases per month, while judges in small-stakes courts, which deal exclusively with less severe criminal cases or low-value civil cases, dispose of almost 50 cases. This highlights why simple comparisons of performance between judges working in different courts might be misleading, and the need to condition on court fixed-effects when estimating judge-level performance.

Descriptive statistics on judicial performance are presented in panel C of Table 1. The average number of case disposed on the merit per month is 40, but the distribution has a long right tail (maximum number is 350) and a non-negligible number of zeros: in 13% of judge-court-month observations the number of cases disposed was zero. As discussed below, this motivates our main specification using the inverse hyperbolic sine of cases disposed as the main explanatory variable. The Table also shows that the average number of hearings is 35 (median = 17).

The assessment of the predictive power of admission exams about judge performance relies on a smaller subsample of individuals matched between the two datasets. We present descriptive statistics for that matched sample in Column (3) of Table 1. We are able to match 2,881 judges in the productivity sample to their admission exam ranking, or 28% of judges observed in the estimation sample. Judges in the matched sample are observed for less months (45 vs. 50 months in non-matched sample), work in more courts (5.9 vs. 4.3) and have slightly lower monthly output of cases disposed on the merit (36 vs. 40). It is important to note that candidates in the matched sample are not a random sample of the universe of judges. In particular, Figure 3 highlights the share of judges we are able to match to recruitment exams by state: whereas in some states like Minas Gerais (MG), Paraná (PR) and Acre (AC) we obtain grades for almost half of judges, three states are not represented at all (Amapá (AP), Rio Grande do Norte (RN) and Rondonia (RO)).

4 Empirical strategy and identification

4.1 Empirical Model

In order to estimate the permanent component of performance for judges, our main challenge is to separate the individual contribution of judges from the effects of courts they work in: courts in larger district might have inherently more demand, or even within districts there might be systematic differences in length of cases between courts, so we cannot simply compare the performance of judges working in different courts. In order to do that, we borrow from the labor literature and estimate a two-way fixed effects model.

We model the number of cases disposed as follows. For a given judge j working on court c on month-year m, we model (the inverse hyperbolic sine of) the number of cases disposed as:

$$y_{jcm} = \theta_j + \gamma_c + \alpha_{s(jc)} + \boldsymbol{X}'_{jcm}\boldsymbol{\beta} + \epsilon_{jcm} \tag{1}$$

where θ_j refers to the permanent component of judge effect; γ_c refers to permanent component of court effect; and X_{jcm} is a vector of time-varying controls. In our baseline specification X_{jcm} includes month-year indicators, the number of courts a Judge work in on a single month and the number of judges working in each single court.[28] Note that we also include an intercept for each connected set, α_s. As previously mentioned, the number of cases disposed is zero in approximately 13% of observations in our dataset. To deal with this, we use the inverse hyperbolic sine transformation (Bellemare & Wichman, 2020) of the number of cases disposed, which, unlike the log transformation, does not drop observations with zero cases disposed.

The separate identification of judge and court fixed-effects in the model above, as shown by Abowd, Creecy, & Kramarz (2002) in the context of workers and firms, is only possible within connected sets – groups of individuals and organizations connected by movers, individuals who work on different organizations throughout the period. Formally, within each connected set g with C_g organizations and J_g individuals, we can identify at most $C_g + J_g - 2$ effects.

The vast majority of judges work in several courts during the period, and

[28]Both the number of judges working in a court and the number of courts a judge works on are computed in the full sample, and not in the estimating sample. While we do not use the variation coming from short judge-court matches, our estimates take into account that, for any given month, judges might be "moonlighting" in other courts and thus have lower performance.

even in more than one court in the same month, meaning that connected sets *within states* are very large: in the majority of states the largest connected set comprises over 95% of judge-court-month observations, and only one state it comprises less than 90%.[29] Within each state, we lose very few observations by restricting our sample to the largest connected sets, providing us with 27 connected sets in our estimating sample.

As previously discussed in Section 2.2, however, judges are selected to work in a specific state, and never work in courts of different states. That means each state is a separate connected set, and we cannot compare court or judge fixed effects across states. While that is not an impediment to our analysis of the predictive power of admission exams, since we only compare individuals in the same exam cohort (and therefore same connected set), adjustments are needed in order to perform the variance decomposition exercise.

We follow Best, Hjort, & Szakonyi (2019) in estimating the variance components with several connected sets. When estimating equation (1), we impose the additional restrictions that both court and judge fixed-effects have mean zero in each connected set. If we define $\tilde{\theta}_j$ and $\tilde{\gamma}_c$ to be the true judge and court fixed-effects, respectively, what we can identify in equation (1) are $\theta_j = \tilde{\theta}_j - \overline{\theta}_g$ and $\gamma_c = \tilde{\gamma}_c - \overline{\gamma}_g$, the deviations of the true effects from the connected set means. We can then write the variance of number of cases disposed as:

$$\begin{aligned} Var(y_{jcm}) =& Var(\theta_j) + Var(\gamma_c) + 2Cov(\theta_j, \gamma_c) + Var(\alpha_s) + \\ & Var(\boldsymbol{X}'_{jcm}\boldsymbol{\beta}) + 2Cov(\alpha_s, \boldsymbol{X}'_{jcm}\boldsymbol{\beta}) + \\ & 2Cov(\theta_j + \gamma_c, \alpha_s + \boldsymbol{X}'_{jcm}\boldsymbol{\beta}) + Var(\epsilon_{jcm}) \end{aligned} \tag{2}$$

Best, Hjort, & Szakonyi (2019) show that, since we can only estimate within

[29] In the small state of Sergipe (SE), the largest connected set comprises only 65% of observations.

connected sets variances, the estimates recovered are lower bounds of the total variance of both judges and courts fixed-effects. The total variance attributable jointly to judges and courts, nonetheless, can be recovered using the variance of the connected sets effects: $Var(\tilde{\theta}_j + \tilde{\gamma}_c) = Var(\theta_j + \gamma_c) + Var(\alpha_s)$.

4.2 Identification and estimation

As discussed in detail in Card, Heining, & Kline (2013), Card, Cardoso, & Kline (2016) and Card, Cardoso, Heining, & Kline (2017), identification in the two-way fixed-effects model does not require random allocation of workers (judges) across firms (courts). The structure of the model allows for rich patterns of sorting, including for judges that dispose of more cases to select into better courts, or for judges to specialize in certain courts where their output is higher. In other words, our identification assumption of exogenous mobility is that judges do not sort on the error term in Equation (1).

Here we focus on assessing whether two particular issues affect the identification of our model. First, we model judge and court fixed-effects as additive and linearly separable. If that is not the case and there exists a judge-court match effect (i.e. more productive judges are particularly efficient in productive courts), then our estimates of judge effect might be biased. Figure 5, panel A, presents a heatmap where we break down residuals of our model by vingtiles of judge and court fixed effects, and graph the average residuals in each cell. To interpret these results, consider Figure 5, panel B, where we simulate a model in which there exists judge-court match effects, but we erroneously estimate a linearly separable mode. The residuals then are systematically large/small in cells with extreme fixed-effects, reflecting the incapacity of the model to capture the matching effect. Going back to panel A, the actual heatmap, we do not observe the same pronounced pattern as in the simulation, suggesting that even if match effects are real (our model seems to be unable to match the outcomes at the very top cell in terms of both judge and court fixed-effects), they are not large enough to severely affect our estimates.

The second issue we consider is whether judges are moving into courts systematically due to *trends in court productivity*. While the selection of judges into courts due to levels of productivity does not affect our estimates, the same is not true if judges can select into courts because they are improving/decreasing their performance. To consider whether that seems to be the case, we perform an event study that assess how the number of cases disposed by judges evolve around the time judges make clear transitions between judicial districts (i.e. judges working in a given court for at least three months prior to transition and at least three months after).[30] Figure 7 reports the coefficients of the event-study, in which we consider the indicator for 6 months before the transition as the omitted category. Three things stand out from these results. First, productivity starts falling in the last two months before a judge moves: knowing they will change courts, they might put in less effort to dispose of more cases or transfer their cases to other magistrates. Second, the fall in performance persists for at least three to fourth months after the transition, but six months after there is no distinguishable effect on performance. Finally, and most important for the model, there seems to be no selection in trends: judges do not seem to be on a trend to be more or less productive, either before or after the movement between judicial districts. These results suggest that selection

[30]It is much harder to create such event-study when judges start working in different courts in the same district, because they often do not clearly leave one court for another, but keep a "connection" to their old appointment. For that reason we restrict our analysis to clear changes of court when judges move from one district to another.

on trends do not seem to be a threat to identification in this context.[31]

Consistent estimation of individual fixed-effects require not only that the number of observations in a panel is large enough, but also that the number of periods in the panel grows to infinity. Since our dataset encompass around 70 months, finite sample bias will lead to excess dispersion in our estimates of both judge and court fixed-effects, inflating the estimated share of total variance explained (Best, Hjort, & Szakonyi, 2019; Silver, 2020). We deal with that issue by using a non-parametric, split-sample correction method that shrink our variance estimates (Finkelstein, Gentzkow, & Williams, 2016)[32].

We randomly split our sample in two, stratifying at the judge-court level, so that we preserve the number of judge-court pairs in both samples. We then proceed to estimate the two-way fixed effect model separately in each sample and obtain separate judge and court fixed-effect estimates. While FEs are noisily estimated in each sample, the errors should be uncorrelated due to the random split. Formally, if in each sample $s = \{1, 2\}$ the estimated judge fixed effect can be written as $\hat{\theta}_{(j,s)} = \theta_j + e_{j,s}$, where θ_j is the true FE for individual *j* and $e_{j,s}$ the error term, with $Cov(e_{(j,1)}, e_{(j,2)}) = 0$, then it holds that $Cov(\hat{\theta}_{(j,1)}, \hat{\theta}_{(j,2)}) = Cov(\theta_j, \theta_j) = Var(\theta_j)$. That is, we can recover the true variance of FEs by separately estimating variances in the random samples and calculating their covariance.

[31]The Figure also shows confidence intervals growing in width after the transition. This happens because we only require judges to reappear in the sample post-transition three times.

[32]Kline, Saggio, & Sølvsten (2020) propose a leave-one-out estimator for the variance of fixed-effects in similar models and show their estimates differ substantially from "naive" estimators that do not take into account limited mobility bias. Their estimator was developed for a single connected set, nonetheless, while in our application we estimate variances from several connected sets.

5 Results

5.1 Judges role in explaining variation in output

Before presenting the results decomposing the variance of total output, we present preliminary evidence that judge fixed-effects matter in explaining courts' output. Table 2, Columns (1) and (2), present goodness-of-fit measures when estimating Equation (1) excluding and including judge fixed-effects, respectively. The inclusion of judge fixed-effects increases the adjusted R-squared of the model by 8 p.p. and reduces the residual standard error (RSE) from 1.43 to 1.34. This is evidence that judges matter in explaining the variation in output observed across courts.

We present the results of formal variance decomposition in Table 3.[33] Column (1) presents the raw variance estimates, with no finite-sample corrections, while Column (2) present corrected variance estimates using split-sample strategy, and Column (3) presents the share of total variance explained by each component using the split-sample estimates. The finite-sample corrected variance of judges' FE is very similar to the raw estimates, on the range of 0.74-0.80, suggesting that judges explain at least 23% of the total variance of output. To put that magnitude in context, it is significantly larger than the estimate of Fenizia (2020) on the share of social security offices' productivity in Italy explained by managers (9%), but very similar to those of Best, Hjort, & Szakonyi (2019) on the share of public procurement prices explained by procurement officers in Russia[34]. The estimates for share of total variance explained by courts

[33]Due to the high dimensionality of fixed-effects, we cannot simply invert matrices to obtain OLS estimates. We then estimate the parameters using the *-lfe-* command in R, also used by Best, Hjort, & Szakonyi (2019).

[34]In Table ??, we present similar variance decompositions using alternartive sample restrictions. The lower bound of total variance explained by judges ranges from 12% using the entire sample (including very short matches) to 29% using a minimum of 4-month spells.

fixed effects, on the other hand, is more sensitive to the correction method, varying from 34-48%. Estimates for the variance explained by the sum of judge and court FEs range from 35-51%: since the sum of explained variance independently explained by judge and courts is much larger than that, it means the covariance of these fixed effects is large and negative, meaning that judges with higher FE are observed matched with courts of low FE, and vice-versa.

While the previous estimates show that judges are important in explaining the quantity of cases disposed and provided individual measures of judge performance, one might worry that judges that dispose of more cases are prioritizing quantity over quality. If that is the case, judges with higher fixed-effects in our model might actually be those that cut back on the inputs necessary to arrive at "good decisions", hastening the process to increase their case disposition number. We test whether this is a plausible explanation in our context by investigating one important input for case decision: the number of hearings that judges hold each month. To assess if "high fixed-effect" judges are conducting systematically less hearings than their peers with lower fixed-effects, we follow Silver (2020) and re-estimate the two-way fixed-effects model using the number of hearings as dependent variable, thus obtaining a new fixed-effect estimate for each judge. If judges are severely cutting back on hearings in order to increase their case disposition, we might expect a weak or even negative correlation between the fixed-effects in both models. Figure 6 shows that this is not the case: fixed-effects from the two models are strongly positive correlated, suggesting that judges who dispose of more cases are also those that hold more hearings. While we are not able to assess whether the use of other inputs, including length or quality of hearings, this alleviates concerns that judges who dispose of more cases are systematically sacrificing on quality.

5.2 Correlates of judges' and courts' fixed-effects

In this Section we briefly describe whether estimated fixed-effects of courts and judges are systematically correlated with observable characteristics. We start by presenting results for courts' FEs in Table 4. The first panel shows that courts' have higher fixed effects when located in judicial districts outside the state capital, with larger populations and higher urbanization rates. Conditional on time and judges' fixed-effects, this suggests that the number of cases disposed is particularly high in poorer, large urban districts outside the largest urban center of states. There are several possible explanations for that finding. If relative demand for judicial services is higher in these poorer areas, relative to supply, courts in those areas might present higher case disposition, possibly in detriment of decision quality. It is also possible that the composition of cases in these areas are different, and the higher number of cases in poorer areas reflect the fact that cases are easier to dispose. All those factors might co-exist, and will be picked up by courts' fixed-effects in our model. The results in Table 4 also shed light on how fixed-effects differ by the nature of cases assigned to each court. Similarly to what we observed in the simple descriptive statistics of Figure 2, criminal courts and those dealing with other topics such as commercial law (pooled with "others" here) have particular low level of case disposition when compared to general courts.

We now turn to describe how judges' fixed-effects correlate with observable characteristics. Here we rely on the sample matched to RAIS, the employer-employee database of formal workers, in order to construct judges' work history and obtain individual traits such as gender and age. Results are presented in Table 5. In Column (1) we present results for all judges that are matched to RAIS, and in Column (2) we restrict to judges that are observed at least once working outside of the judiciary, in order to include wages prior to judgeship as a correlate. All estimates include connected-sets (State) fixed-effects. Results in Column (1) suggest that individual traits explain very little of the estimated effects: gender, education and experience, both in general and in the judiciary, are not significant predictors of judge fixed-effects. Age is correlated with the estimated effect, with a positive and concave relationship: older judges dispose of more cases, but the effect is diminishing in age. These results, however, are not very robust: when we restrict the sample to those observed working outside the judiciary since 1995, we no longer observe age as a significant predictor, but overall experience does seem positively correlated with case disposition. The coefficient on (log) average yearly wage received before joining the judiciary, which we interpret as potential earnings outside of judgeship, is small in magnitude and not statistically different from zero.

5.3 Entrance exams are predictive of performance

Results in the previous sections are strong evidence that the identity of judges matters for the timely delivery of justice. While we are unable to explain the reasons why some judges are more effective in disposing of cases than others, the fact that we observe such differences in judge output suggests that the screening of judges might be one tool in improving judicial efficiency. We now turn to the question of how candidates performance in the admission exams is related to their performance on the job. In all the exercises that follow we use the sample for which we can match judges' admission exam performance.

We start by presenting "reduced-form" evidence that entrance exam ranks are correlated with the number of cases disposed on merits, once we control for court and month fixed-effects. That is, in here we do not use estimated judges fixed-effects, but simply present OLS regressions of the form

$$y_{jcm} = \boldsymbol{\beta' ExamRankQuintile_j} + \gamma_c + \delta_{w(j)} + \boldsymbol{X'_{jcm}\theta} + \epsilon_{jcm} \quad (3)$$

where y_{jcm} is the IHS transformation of cases disposed, $ExamRankQuintile_j$ are indicators for quintile of exam performance of judge *j* in their exam cohort and δ_w are indicators for each cohort of candidates, since we can only meaningfully compare ranking among candidates sitting the same examination[35]. Standard-errors are clustered at the judge-level.

Results are presented in Table 6, where the omitted category for exam quintile is the bottom 20%. Column (1) presents estimates for a regression that only includes cohort fixed-effects, while in Columns (2) and (3) we add Court and Month fixed-effects, respectively. Focusing on Column (3), the results suggest that, when compared to judges ranking in the bottom quintile of their cohorts, those in the top 20% dispose of approximately 21% more cases. The estimated effect is smaller but statistically significant and economically meaningful for judges with ranks in the second to fourth quintiles, and we can reject that the coefficient for the top 20% is identical to those on the second and third quintiles. In Column (4) we present a much more stringent exercise: we include court-by-month fixed effects, meaning that the only variation used comes from different judges working in the same court on the same month (hence the large drop in sample size, since observations for courts with a single judge in any given month are dropped). The estimated coefficients are slightly larger in absolute value, but broadly consistent with previous estimates suggesting that better ranking in entrance exams are correlated with higher case disposition on the

[35] All our analyses consider perform conditional on being selected for the job, so rankings are computed within judges in each cohort and not across all applicants for the job.

job.

We now present results using the estimated judges' fixed-effects obtained in the previous section. Figure 4 presents non-parametric evidence of the correlation between (residualized) ranks in admission exams and standardized FE.[36] The strong positive correlation between performance measures suggests that judges who perform well in the admission exams are also among the ones with highest FE in their cohorts.

In Table 7 we present this same evidence in regression form. We estimate simple OLS regressions at the judge-level, using measures of on-the-job performance (standardized fixed-effect) as dependent variables and quintiles of performance in the recruitment exam as the main explanatory variable. Column (1) presents results from an OLS regression including cohort fixed-effects. Consistent with the findings in the reduced form regression, our results suggest that being ranked in the top 20% in the admission exam is correlated with a 0.2 s.d. increase in judge's performance (estimated fixed-effects) in comparison to those in the bottom quintile. Those ranking in lower quintiles are also estimated to perform 0.1-0.15 s.d. higher when compared to those at the very bottom. In Column (2) we replace the quintile ranking in the admission exam with the standardized final grade used to construct ranking.[37] The coefficient on grade is significant and indicates that an increase of 1 standard deviation in final grade is correlated with a 0.07 s.d. increase in performance (measured by judges' fixed-effects). Taken together with the results from the reduced-form model, this suggests that, among the candidates selected in the admissions exam, those that rank higher do perform better on the job than those

[36]Since we only compare judges entering in the same cohort, we first regress each rank on cohort indicators and use residuals to construct the binned scatter plot

[37]We could not collect final grades for some of the cohorts, therefore the smaller sample size in Column (2)

ranking lower.

While we believe documenting that the overall ranking is informative about job performance is an important result, it does not shed light on exactly which dimension of the screening process is leading to this positive correlation. It is possible, for example, that the Titles Exam, that takes into account previous work and academic accomplishments, is the most predictive component of the overall ranking. Or that the Oral Exam, in which there exists some degree of discretion by the selection committee, would be more informative. We attempt to provide evidence on that question by re-estimating the previous equation using grades in each exam phase as dependent variables and assessing which of those are more predictive about job performance. As previously discussed, we restrict our analysis to 20 examinations and 619 judges for whom we observe six separate grades: Objective Exam, Written Exam, Civil and Criminal case decisions, Oral Exams and "Titles" Exam.

Results are reported in Columns (3) and (4) of Table 7. We first report in Column (3) the results of estimating the equivalent equation of Column (2) in the subsample for which we have detailed grade information. The result is very similar to that obtained in the full sample, suggesting that candidates with higher final grades also perform better on the job. We then estimate the model including standardized grades in each of the exams separately, and report results in Column (4). The only coefficient that is significant, and also the largest in magnitude, is that of grades on the Judicial Decision Writing on civil cases (the coefficient on criminal cases is less than half the size in magnitude and not statistically different from zero). As shown below, this result is robust to other specifications of the estimation equation, suggesting that the civil case admission is indeed the most predictive component of the admission exam.

Recall that, since the harmonization of admission exams in 2009, each of the Judicial Decision Writing exams has weight 15% for the final ranking, so grades in the civil case decision contribute less to the final selection than grades in the written exam (30% weight) or Oral exam (20%). Our results suggest, in contrast, that if the goal is to select candidates who will increase the speed of case disposition, exams should overweight results in the civil case decision.

5.4 Results are robust to alternative specifications

We conduct several exercises to assess the robustness of our results. First, Table 8 presents regressions in which we drop top and bottom performers in each cohort, evaluating whether results are fully driven by the very best (or very worst) candidates. Column (1) reproduces our main specification, while the remaining Columns restrict the sample by dropping only the top 3 performers in each cohort (Column 2); the top 5% candidates in each cohort (Column 3); the bottom 5% candidates in each cohort (Column 4); and both the top and bottom 5% candidates in each cohort (Column 5). Estimates of the correlation between exam rank and FE are very stable, and we cannot reject they are statistically indistinguishable from our main specification[38].

In Table 9 we re-estimate the results of our main specification but use the rank of judges' FE as dependent variable instead of the standardized FE. Column (1) presents the coefficients on admission exam quintiles: among judges entering in the same cohort, those in the top quintile rank, on average, four

[38]We also assess whether the correlation between performance on entrance exam and fixed-effects are robust to using alternative samples in the two-way fixed-effects model. In Table ?? we present that correlation using the same samples in Table ??. Using the full sample yields almost identical results, while restricing the sample further decreases the magnitude of coefficients and make estimates noisier. In all specifications, nonetheless, we estimate that performing in the top quintile of the the entrance exam is significantly correlated with higher disposal of cases.

positions higher than those in the bottom 20% (coefficients are negative since a better rank equals a lower rank number). Those in the second quintile rank 2.6 positions higher, on average, and those in the third and fourth quintile between 1.3-2 positions. Overall, the results confirm our main specification findings that candidates ranking better in the admission exam also perform better on the job. Column (2) uses the final grade as explanatory variable, showing that a unit s.d. increase in final grade is correlated with a 1.5 higher position in FE ranking. When we restrict the sample to those observations with detailed grades, in Column (3), the coefficient on Final Grade is very similar in magnitude to that on the full sample. Finally, when we include separate grades by phases as predictors of judge FE we again find that the largest coefficient in magnitude and significant is that associated with civil case exam: an increase in 1 s.d. in the civil case grade is correlated with a 1.4 better ranking in performance.

We also perform a randomization inference exercise to assess the robustness of our findings of the positive correlation between admission exam and job performance ranks. Within each cohort of judges, we randomly assign exam rankings, re-compute quintiles and then estimate the baseline model presented in Column (1) of Table 7. Figure 8 presents the histogram of these 1,000 simulated beta-coefficients for the top 20% performance indicator, and the solid line marks the true coefficient of 0.227. 95% of estimated coefficients are on the interval [-.118, 0.118], and none of the estimates is larger in magnitude than the true estimate. These results suggest that it is very unlikely that we would obtain a coefficient of this magnitude purely by chance.

6 Conclusion

This paper provides evidence that states can effectively design impersonal exams that are able to screen good candidates for top public service positions, even when recruitment practices are constrained by fears of political influence. We explore rich data on judges and courts in Brazil to show that judges are relevant in explaining the observed variation in output, and estimate judge-level measures of performance in case disposition – an important indicator in a judicial system with high levels of court congestion. We then link these measures to the performance of judges in admission exams, and show that within cohorts of hired judges those with higher grades also dispose of more cases. In particular, it seems that not all phases of the admission exams are equally likely to predict job performance: across different specifications, grades on the civil case exam is the only statistically significant predictor.

Our results have meaningful implications for policy makers. First, it adds to recent evidence that not only frontline providers matter for the delivery of public service: managers and other officials working across the state bureaucracy can have significant impact on service provision (Best, Hjort, & Szakonyi, 2019; Fenizia, 2020; Aman-Rana, 2020). Carefully designing systems that select and incentivize these individuals is therefore very important. Secondly, it is also relevant for the debate about rules and discretion in hiring (Hoffman, Kahn, & Li, 2018). We show that an admission process with little discretion by the selecting agency is able to rank individuals in a way that meaningfully predicts job performance. In particular, by breaking down exam performance into its components, we find evidence that an examination that approximates the kind of task faced by candidates on the job (the writing of sentences by judges) is especially predictive about their future performance.

Data limitations do not allow us to further explore three mechanisms we believe are relevant for future research. The first is what makes for an efficient judge. Judges do not work in isolation writing decisions, but, rather, manage complex organizations staffed by several workers and in close contact with other state actors (Pinheiro, 2003; Oliveira Gomes, 2014). A more efficient judge might be one that simply puts longer hours and more effort to increase case disposition, but might just as well be one that is able to put in place an well-oiled machine where every staffer is pulling their weight and ensuring smooth handling of cases.[39] Management practices have shown to be very relevant in explaining productivity in both the private (Bloom & Van Reenen, 2007; Bloom, Eifert, Mahajan, McKenzie, & Roberts, 2013) and the public sector (Rasul & Rogger, 2018; Leaver, Lemos, & Dillenburg Scur, 2019; Bloom, Lemos, Sadun, & Reenen, 2015), so gaining better understanding of working practices in the judicial sector might shed light on the determinants of judge effectiveness.

Second, while we find a strong and robust positive correlation between grades in the admission test and performance, and consider this a relevant parameter for policy-makers designing screening processes, it is unclear exactly what is the force driving this correlation. One possibility is that exams are indeed effective in screening candidates with specific knowledge that is also useful for the tasks performed by a judge – the fact that grades in the civil case examination are the only ones with independent predictive power suggest this might be the case. Another possibility, however, is that competitiveness and difficulty of the exams screen candidates with high general ability and/or high motivation to be a judge, which implies that the congruence between test con-

[39]Fenizia (2020) finds that the mechanism through which managers in social security offices are able to increase output per worker is by letting go of workers while maintaining total output stable.

tent and requirements of the job is less important. We think this is an relevant distinction, particularly in light of the theory and evidence that highlight the role of intrinsic motivation in driving performance when high-powered incentives are limited (Deserranno, 2019; Ashraf & Bandiera, 2018; Prendergast, 2008).

Finally, given that we only observe judicial productivity for those who pass entrance examinations, we cannot make claims about the remaining pool of candidates. In particular, we cannot directly test if examinations are screening for the most productive candidates overall or not, or if exams should be made more selective or less. Future research with complete characteristics of candidates, their career paths and more examinations could evaluate the overall efficiency of the Brazilian selection mechanism into the public sector.

References

Abowd, J. M., Creecy, R. H., & Kramarz, F. (2002). Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data. *U.S. Census Bureau*. 19

Aman-Rana, S. (2020). In Self Interest? Meritocratic Promotions in a Bureaucracy through Discretion of Seniors. *Working Paper*. 5, 33

Ashraf, N., & Bandiera, O. (2018). Social Incentives in Organizations. *Annual Review of Economics*, *10*(1), 439–463. _eprint: https://doi.org/10.1146/annurev-economics-063016-104324. 35

Ashraf, N., Bandiera, O., Davenport, E., & Lee, S. S. (2020). Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services. *American Economic Review*, *110*(5), 1355–94. 4, 5

Ashraf, N., Bandiera, O., & Jack, B. K. (2014). No margin, no mission? A field experiment on incentives for public service delivery. *Journal of Public Economics*, *120*, 1–17. 1

Barbosa, K., & Ferreira, F. V. (2019). Occupy Government: Democracy and the Dynamics of Personnel Decisions and Public Sector Performance. Tech. Rep. 25501, National Bureau of Economic Research, Inc. Publication Title: NBER Working Papers. 10

Bellemare, M. F., & Wichman, C. J. (2020). Elasticities and the Inverse Hyperbolic Sine Transformation. *Oxford Bulletin of Economics and Statistics*, *82*(1), 50–61. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/obes.12325. 19

Bertrand, M., Burgess, R., Chawla, A., & Xu, G. (2020). The Glittering Prizes: Career Incentives and Bureaucrat Performance. *The Review of Economic Studies*, *87*(2), 626–655. 1, 5

Best, M. C., Hjort, J., & Szakonyi, D. (2019). Individuals and Organizations as Sources of State Effectiveness. Tech. rep. 6, 20, 23, 24, 33

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., & Roberts, J. (2013). Does Management Matter? Evidence from India. *The Quarterly Journal of Economics*, *128*(1), 1–51. 34

Bloom, N., Lemos, R., Sadun, R., & Reenen, J. V. (2015). Does Management Matter in schools? *The Economic Journal*, *125*(584), 647–674. 34

Bloom, N., & Van Reenen, J. (2007). Measuring and Explaining Management Practices across Firms and Countries. *The Quarterly Journal of Economics*, *122*(4), 1351–1408. Publisher: Oxford University Press. 34

Brollo, F., Forquesato, P., & Gozzi, J. C. (2017). To the Victor Belongs the Spoils? Party Membership and Public Sector Employment in Brazil. *SSRN Electronic Journal*. 1, 10

Card, D., Cardoso, A. R., Heining, J., & Kline, P. (2017). Firms and Labor Market Inequality: Evidence and Some Theory. *Journal of Labor Economics*, *36*(S1), S13–S70. Publisher: The University of Chicago Press. 21

Card, D., Cardoso, A. R., & Kline, P. (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *Quarterly Journal of Economics*. 21

Card, D., Heining, J., & Kline, P. (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *The Quarterly Journal of Economics*, *128*(3), 967–1015. 21

Chemin, M. (2009). Do judiciaries matter for development? Evidence from India. *Journal of Comparative Economics*, *37*, 230–250. 6

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, *104*(9), 2593–2632. 5

Colonnelli, E., Prem, M., & Teso, E. (2020). Patronage and Selection in Public Sector Organizations. *American Economic Review*, *110*(10), 3071–3099. 1, 5, 10

Conselho Nacional de Justiça (2018). Justiça em Números. Tech. rep., Brasília. 2, 7

Dal Bó, E., Finan, F., & Rossi, M. A. (2013). Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service. *Quarterly Journal of Economics*, *128*(3), 1169–1218. 4, 5

Deserranno, E. (2019). Financial Incentives as Signals: Experimental Evidence from the Recruitment of Village Promoters in Uganda. *American Economic Journal: Applied Economics*, *11*(1), 277–317. 4, 5, 35

Duflo, E., Dupas, P., & Kremer, M. (2015). School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools. *Journal of public Economics*, *123*, 92–110. Publisher: Elsevier. 5

Fenizia, A. (2020). Managers and Productivity in the Public Sector. *Working Paper*. 6, 24, 33, 34

Finan, F., Olken, B. A., & Pande, R. (2017). The Personnel Economics of the Developing State. In A. V. Banerjee, & E. Duflo (Eds.) *Handbook of Economic Field Experiments*, vol. 2 of *Handbook of Economic Field Experiments*, (pp. 467–514). North-Holland. 1, 5

Finkelstein, A., Gentzkow, M., & Williams, H. (2016). Sources of geographic variation in health care: Evidence from patient migration. *Quarterly Journal of Economics*. 23

Hanna, R., & Wang, S.-Y. (2017). Dishonesty and selection into public service: Evidence from India. *American Economic Journal: Economic Policy*, *9*(3), 262–90. 1

Hoffman, M., Kahn, L. B., & Li, D. (2018). Discretion in hiring. *Quarterly Journal of Economics*, *133*(2), 765–800. 1, 33

Hoffman, M., & Tadelis, S. (2018). People Management Skills, Employee Attrition, and Manager Rewards: An Empirical Analysis. Tech. Rep. w24360, National Bureau of Economic Research, Cambridge, MA. 1

Holmstrom, B., & Milgrom, P. (1991). Multitask Principal-Agent Analyses : Incentive. *Journal of Law, Economics, and Organization*. 3

Jacob, B. A., Rockoff, J. E., Taylor, E. S., Lindy, B., & Rosen, R. (2018). Teacher applicant hiring and teacher performance: Evidence from DC public schools. *Journal of Public Economics*, *166*, 81–97. Publisher: Elsevier. 5

Khan, A. Q., Khwaja, A. I., & Olken, B. A. (2019). Making moves matter: Experimental evidence on incentivizing bureaucrats through performance-based postings. *American Economic Review*, *109*(1), 237–70. 1

Kline, P., Saggio, R., & Sølvsten, M. (2020). Leave-Out Estimation of Variance Components. *Econometrica*, *88*(5), 1859–1898. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA16410.
URL `https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA16410` 16, 23

Kondylis, F., & Stein, M. (2021). The Speed of Justice. *Review of Economics and Statistics, Forthcoming*. 6

Landes, W., Lessig, L., & Solimine, M. (1998). Judicial Influence: A Citation Analysis of Federal Courts of Appeals Judges. *The Journal of Legal Studies*, *27*(2), 271–332. Publisher: The University of Chicago Press. 3

Lavy, V. (2009). Performance pay and teachers' effort, productivity, and grading ethics. *American Economic Review*, *99*(5), 1979–2011. 1

Leaver, C., Lemos, R. F., & Dillenburg Scur, D. (2019). Measuring and Explaining Management in Schools : New Approaches Using Public Data. Tech. Rep. WPS9053, The World Bank. 34

Mehmood, S. (2020). Judicial Independence and Development: Evidence from Pakistan. *Working Paper*. 5

Muralidharan, K., & Sundararaman, V. (2011). Teacher Performance Pay: Experimental Evidence from India. *Journal of Political Economy*, *119*(1), 39–77. Publisher: The University of Chicago Press. 5

Oliveira Gomes, A. (2014). Estudos sobre desempenho da Justiça Estadual de primeira instância no Brasil. *PhD Thesis*. 34

Pinheiro, A. C. (2003). Judiciary, Reform and Economics: The Judges' Viewpoint. SSRN Scholarly Paper ID 482801, Social Science Research Network, Rochester, NY. 34

Ponticelli, J., & Alencar, L. (2016). Court Enforcement, Bank Loans and Firm Investment: Evidence from a Bankruptcy Reform in Brazil. *Quarterly Journal of Economics*, *131 (3)*, 1365–1413. 2, 6, 7, 14

Prendergast, C. (2008). Intrinsic Motivation and Incentives. *American Economic Review*, *98*(2), 201–205. 35

Rao, M. (2020). Judges, Lenders, and the Bottom Line: Court-ing Firm Growth in India. *Working Paper*. 6

Rasul, I., & Rogger, D. (2018). Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service. *The Economic Journal*, *128*(608), 413–446. 34

Sadka, J., Seira, E., & Woodruff, C. (2018). Information and Bargaining Through Agents: Experimental Evidence from Mexico's Labor Courts. *NBER Working Paper Series*, *25137*. 6

Schiavon, L. (2017). *Essays on crime and justice*. PhD Thesis, PUC-Rio. 8

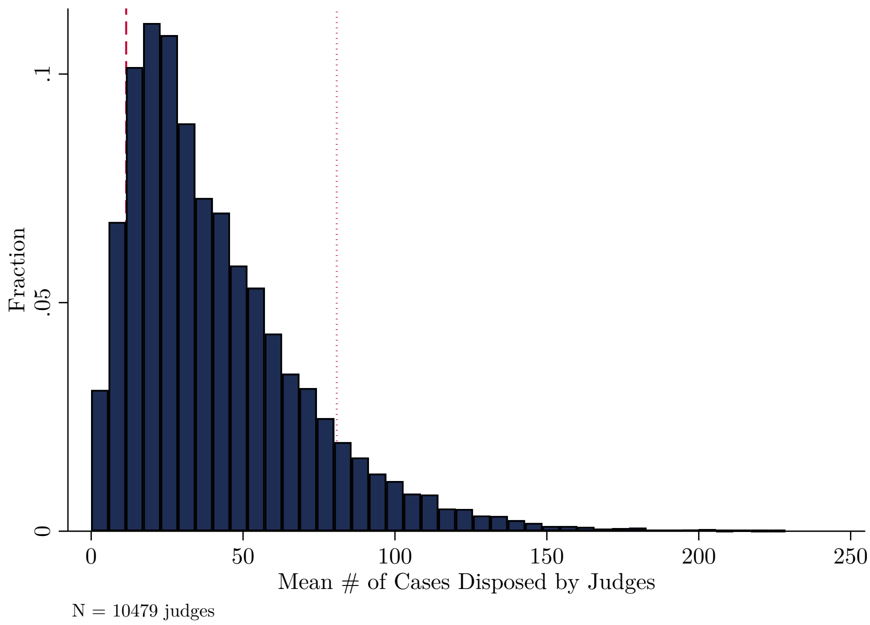
Silver, D. (2020). Haste or Waste? Peer Pressure and Productivity in the Emergency Department. SSRN Scholarly Paper ID 3588769, Social Science Research Network, Rochester, NY. 23, 25

Weaver, J. (2021). Jobs for Sale: Corruption and Misallocation in Hiring. *American Economic Review*, *Forthcoming*. 1, 5

Xu, G. (2018). The Costs of Patronage: Evidence from the British Empire. *American Economic Review*, *108*(11), 3170–3198. 5

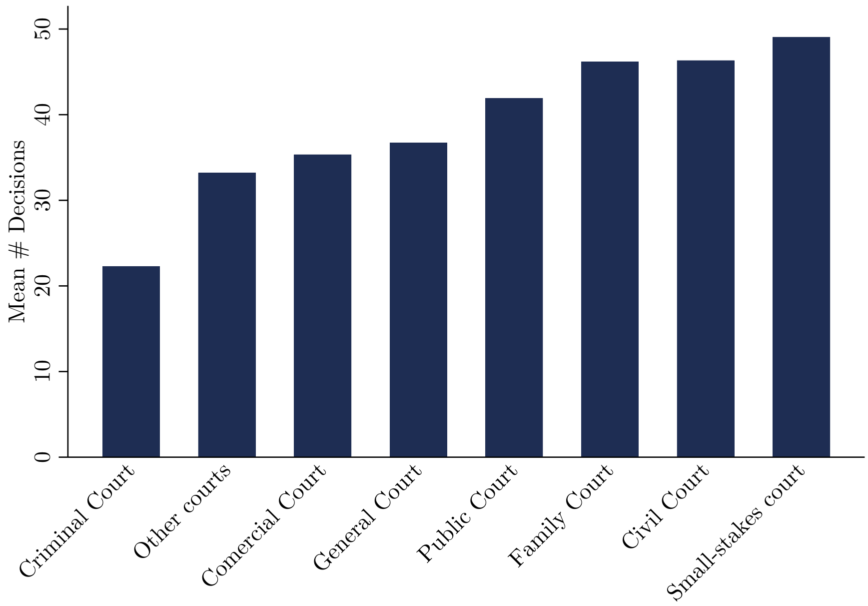
7 Figures and Tables

Figure 1: Histogram of mean number of cases disposed on the merits by judge.



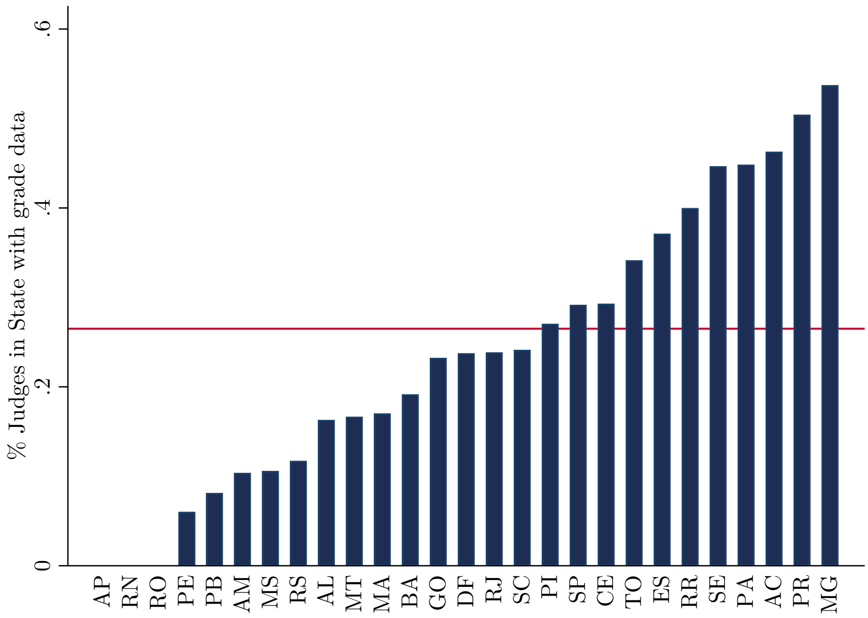
Note: The histogram presents the average monthly number of cases disposed by judges. Average number of cases is calculated in the sample used to estimate the two-way fixed-effects model, where outcome variables are trimmed at the top 1%, judge-court spells shorter than three months are dropped and only observations in the largest connected sets within each state are used. The dashed and dotted lines mark the 10th and 90th percentile of the distribution, respectively. The figure documents the vast dispersion in number of case disposition across judges: those in the 90th percentile of the distribution dispose of eight times as many cases as those in the 10th percentile.

Figure 2: Average number of cases disposed, by type of court



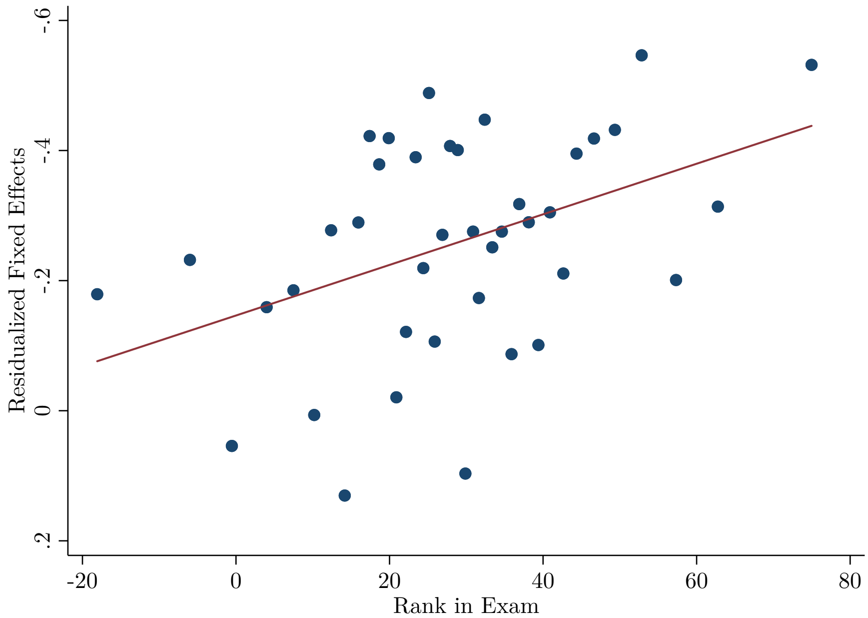
Note: This figure presents the average monthly number of cases disposed by judges, in each type of court. Number of cases is calculated in the sample used to estimate the two-way fixed-effects model, where outcome variables are trimmed at the top 1%, judge-court spells shorter than three months are dropped and only observations in the largest connected sets within each state are used. The figure documents systematic differences in number of case disposition across courts: judges in criminal courts dispose od twenty cases, on average, every month, while judges in small-stakes courts dispose of almost 50 cases.

Figure 3: Share of judges matched by State.



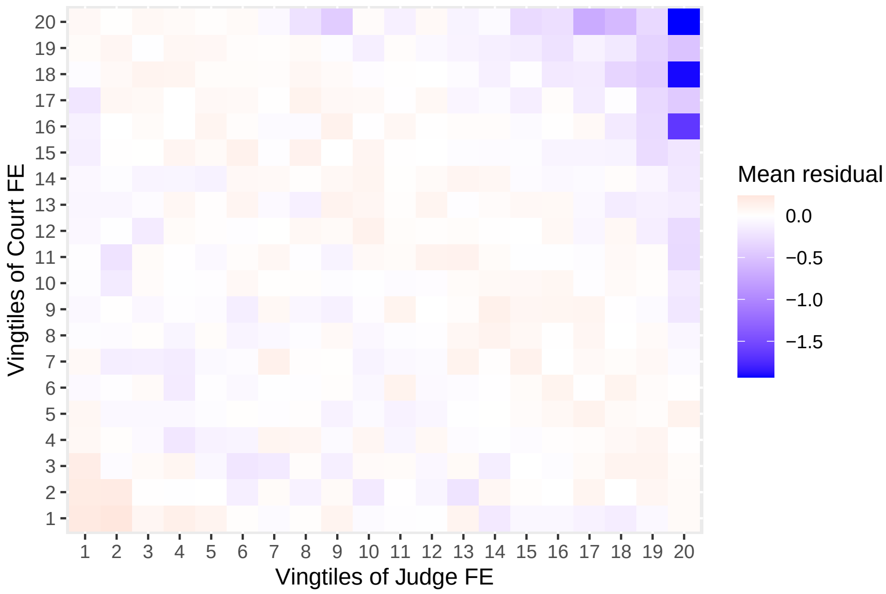
Note: This figure presents the share of judge in the estimation sample that are matched to their admission exam , by State. The red line mark the overall share of judges matched (28%).

Figure 4: Binscatter of residual ranks, conditioning on Concurso FE

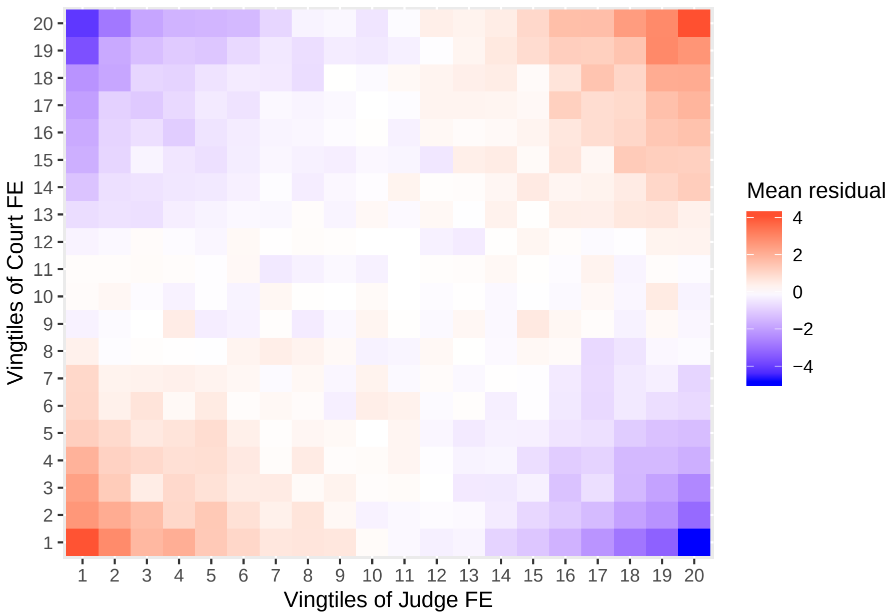


Note: The graph presents a binned scatter plot of residualized rank in fixed-effects obtained by estimating Equation (1) and residualized ranks in admission exams, at the judge level. Residues are obtained by regressing each of the variables on *Concurso* fixed effects.

Figure 5: Residuals heatmap from two-way fixed-effects model



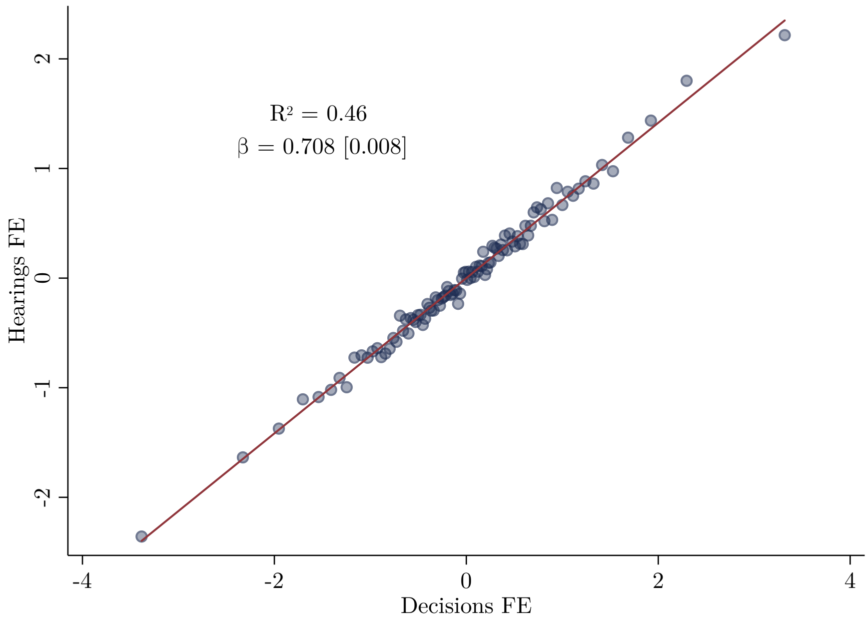
(a) Actual from data



(b) Simulated from misspecified model

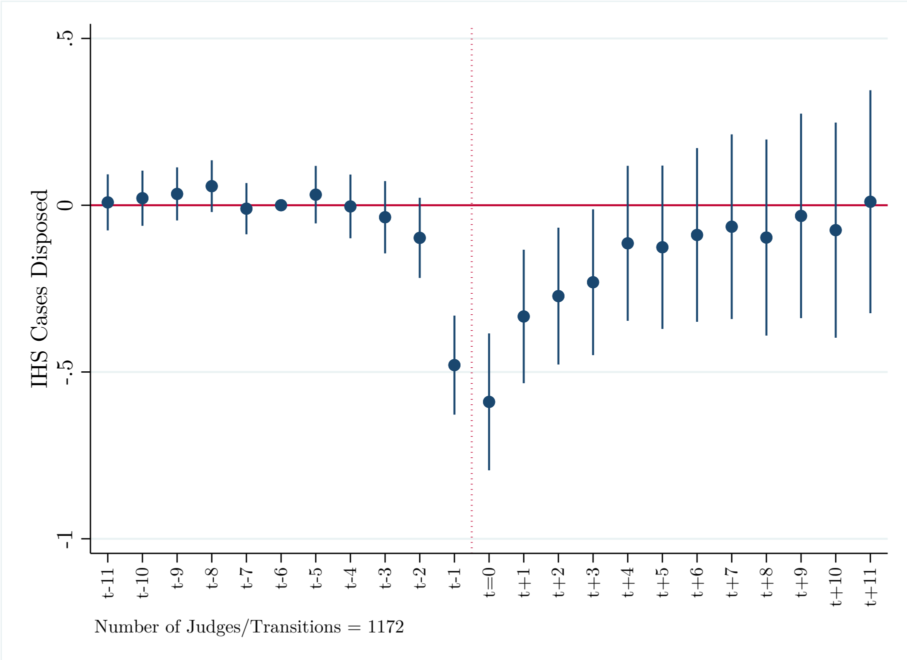
Note: These figures present heatmaps of average residuals from a two-way fixed-effects model. Panel A presents results from actual data estimated using equation (1). Panel B presents results from a simulated model with 10,000 worker-firm observations (200 workers in 50 firms) containing match effects between workers and firms, but estimated using a misspecified model in equation (1). Darker blue cells represent large negative residuals, while darker red cells represent large positive residuals. Judges and courts are binned into vingtiles of estimated fixed-effects.

Figure 6: Binscatter of hearings and case disposition fixed-effects



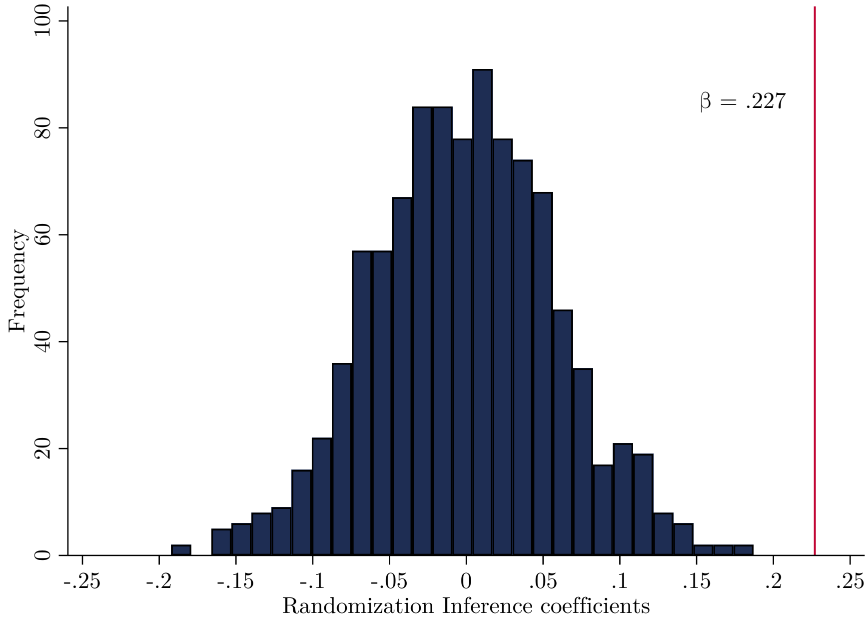
Note: This graph presents a binned scatter plot of residualized judge fixed-effects obtained by estimating equation (1) using hearings and case disposition separately. Residuals are obtained by regressing both FEs on connected set dummies. The R-squared and coefficients presented refer to a regression of hearing FE on case disposition FE including connected set dummies.

Figure 7: Event-study around judicial district movement



Note: This figure reports point estimates and 95% CI for coefficients on an event-study regression of the form $y_{jm} = \sum_t \beta_t \boldsymbol{RelativePeriod_t} + \gamma_j + \delta_m + \epsilon_{jm}$, where y_{jm} is the IHS of cases disposed, γ_j and δ_m are judge and month fixed-effects, respectively, and β_t are the event-study coefficients. The omitted category is the indicator for six months before the movement. The sample is restricted only to clear moves across judicial districts, as detailed in Section 4.2. Standard errors are clustered at the transition-level.

Figure 8: Histogram of simulated beta-coefficients



Note: The figure presents the histogram of 1,000 simulated coefficients for the top 20% indicator using our main specification, equivalent to the results presented in Column (1) of Table 7, where we randomly assign final admissions ranking within each cohort. The true coefficient is marked by the solid red line

Table 1: Descriptive statistics

	Full Sample	Estimation Sample	Exam matched Sample
Panel A: Judges			
Share male judges	0.61	0.60	0.61
Mean # courts by judge	10.52	4.28	5.87
Mean # months by judge	50.97	49.99	44.91
Mean # courts at judge-month level	1.70	1.39	1.55
Mean # judicial districts at judge-month level	3.72	2.28	3.29
Mean # months per judge-court pair	8.22	16.23	11.83
Panel B: Courts			
Mean # of judges by court	12.64	4.96	2.99
Mean # judges at court-month level	1.67	1.38	1.20
Share civil courts	0.22	0.22	0.23
Share general courts	0.20	0.20	0.24
Share small-stakes courts	0.18	0.18	0.16
Share criminal courts	0.16	0.16	0.16
Share family court	0.10	0.10	0.09
Share other courts	0.14	0.13	0.11
Panel C: Output measures			
Cases Disposed (on merit)	33.82	40.13	36.10
Total Hearings (presided or held)	29.32	34.88	35.85
Number of judges	11,462	10,479	2,881
Number judges ever working in multiple courts	10,378	8,500	2,653
Number of courts	9,540	9,048	5,667
Number of courts with multiple judges	9,201	8,152	3,925
Number of judge-court pairs	120,642	44,850	16,918
Number of judge-court spells	273,074	77,799	24,089
Number of connected sets	68	27	24
Number of judge-court-month observations	991,324	727,784	200,212

Note: This table reports descriptive statistics for key variables. Column (1) refers to the full original panel. Column (2) refers to the sample used to estimate the two-way fixed-effects model, where outcome variables are trimmed at the top 1%, judge-court spells shorter than three months are dropped and only observations in the largest connected sets within each state are used. Column (3) refers to the sample matched to admission exams, i.e., it only retains judge-court-month observations for which judges were matched to their admission exams ranking. This is the sample used in both the "reduced-form" exercises presented in Table 6 and the main results on the correlation between admission ranking and performance in Table 7.

Table 2: Goodness of fit measures

	(1)	(2)	(3)
R-squared	0.379	0.464	0.619
Adjusted R-squared	0.371	0.45	0.593
Residual Standard Error (RSE)	1.434	1.341	1.153
Observations	727784	727784	727784
Judge FE	No	Yes	No
Judge-by-Court FE	No	No	Yes

Note: This table presents goodness-of-fit measures for several different models using the two-way fixed-effects estimation sample. Column (1) presents results from a model that does not include judge fixed-effects; Column (2) is our main specification from equation (1), including judge fixed-effects; while Column (3) includes judge-by-court fixed effects.

Table 3: Variance decomposition

	Raw Variance	Split Sample Variance	Split sample Var - % Total
Cases disposed (IHS)	3.27	3.27	1.00
Judge FE	0.80	0.74	0.23
Court FE	1.16	1.11	0.34
Connected Set FE	0.24	0.24	0.07
Judge+Court FE	1.23	1.00	0.31

Note: This table presents the variance decomposition exercise using estimates from the two-way fixed effects model in equation (1). Column (1) presents the variance estimates without adjustment, while Column (2) presents variance estimates corrected for finite-sample bias using the spit-sample technique. Column (3) presents the finite-sample corrected variance estimates as a share of total variance.

Table 4: Correlation between courts' fixed-effects and courts' characteristics

	(1)
Judicial district characteristics	
State Capital	-0.122**
	(0.0475)
Log population (2010)	0.0982***
	(0.0154)
Log GDP per capita (2016)	-0.0730***
	(0.0278)
Share urban households (2010)	0.291***
	(0.0926)
Second level	0.268***
	(0.0419)
Third level	0.122**
	(0.0542)
Special level	0.0169
	(0.0894)
Type of courts	
Criminal court	-0.920***
	(0.0478)
Civil court	-0.198***
	(0.0464)
Family court	-0.316***
	(0.0529)
Small-stakes court	-0.181***
	(0.0439)
Other courts	-0.453***
	(0.0512)
Observations	9,047
R-Squared	0.073
Number Connected Sets	27
CS fixed-effect?	Yes

Note: This table reports regressions using the estimated courts' FE (standardized to have unit standard deviation within connected sets) as dependent variable. State capital is a dummy variable indicating whether the judicial district where the court is located is a state's capital; Log population is from the 2010 Census and Log GDP per capita is from the 2016 national accounts published by the Brazilian Institute of Geography and Statistics (IBGE). Robust standard errors in parentheses (* p<0.1, ** p<0.05, *** p <0.01)

Table 5: Correlation between judges' fixed-effects and individual characteristics

	(1)	(2)
Male	-0.00721	-0.0224
	(0.0220)	(0.0341)
Age in 2015	0.0516***	0.0232
	(0.0122)	(0.0203)
Age (squared)	-0.000537***	-0.000272
	(0.000119)	(0.000203)
Graduate degree	0.0754*	0.121**
	(0.0433)	(0.0602)
Formal labor experience in 2015	-0.000124	0.0540*
	(0.0142)	(0.0312)
Formal experience (squared)	-0.000161	-0.00194*
	(0.000526)	(0.00109)
Formal judicial experience in 2015	-0.00262	-0.00140
	(0.0113)	(0.0161)
Judicial experience (squared)	0.000964**	0.00103
	(0.000466)	(0.000743)
Formal experience outside judicial sector	-0.0119	
	(0.0344)	
Log average wage before judiciary (2017 prices)		0.0175
		(0.0158)
Observations	8,597	2,827
R-Squared	0.019	0.046
Number Connected Sets	26	26
CS fixed-effect?	Yes	Yes

Note: This table reports regressions using the estimated judges' FE (standardized to have unit standard deviation within connected sets) as dependent variable. Independent variables are obtained from matching judges' in performance dataset to RAIS, a matched employer-employee administrative dataset. Data from RAIS covers the period 1995-2017, so measures of experience in the formal sector and in the judiciary in 2015 are capped at 20 years. Robust standard errors in parentheses (* p<0.1, ** p<0.05, *** p <0.01)

Table 6: Reduced form regressions: output and admission exam performance

	(1)	(2)	(3)	(4)
Top quintile (β_1)	0.113**	0.216***	0.207***	0.334***
	(0.0546)	(0.0403)	(0.0402)	(0.0820)
4th quintile (β_2)	0.102**	0.180***	0.176***	0.199**
	(0.0508)	(0.0388)	(0.0387)	(0.0809)
3rd quintile (β_3)	0.0854*	0.137***	0.133***	0.242***
	(0.0503)	(0.0371)	(0.0368)	(0.0795)
2nd quintile (β_4)	0.0920*	0.143***	0.141***	0.246***
	(0.0511)	(0.0375)	(0.0373)	(0.0783)
Observations	200,206	200,206	200,206	59,795
R-Squared	0.11	0.42	0.43	0.53
Concurso FE	Yes	Yes	Yes	Yes
Court FE	No	Yes	Yes	No
Month FE	No	No	Yes	No
Court-by-Month FE	No	No	No	Yes
$\beta_1 = \beta_2$	0.82	0.35	0.40	0.08
$\beta_1 = \beta_3$	0.55	0.03	0.03	0.20
$\beta_1 = \beta_4$	0.66	0.05	0.07	0.27

Note: This table reports results from estimating equation (3): $y_{jcm} = \boldsymbol{\beta ExamRankQuintile_j} + \gamma_c + \delta_{w(j)} + \boldsymbol{X'_{jcm}\theta} + \epsilon_{jcm}$, where y_{jcm} is the inverse hyperbolic sine of cases disposed. All specifications include examination cohort (*Concurso*) fixed-effects. Columns (1) through (3) use the exam matched sample, observations used in the two-way fixed-effects models for which judge admission exams are available. Column (4) uses a subset of that sample that excludes all observations for which only one judge is working in any given court on a month. Standard-errors are clustered at the Judge level (* p < 0.1, ** p < 0.05, *** p < 0.01).

Table 7: Main results – correlation between admission grades and performance

	(1)	(2)	(3)	(4)
Top quintile	0.227***			
	(0.0587)			
4th quintile	0.147**			
	(0.0599)			
3rd quintile	0.107*			
	(0.0575)			
2nd quintile	0.152**			
	(0.0589)			
Final Grade (standardized)		0.0675***	0.0692*	
		(0.0224)	(0.0389)	
Objective Grade (standardized)				-0.0151
				(0.0417)
Written Exam Grade (standardized)				0.0141
				(0.0392)
Civil Essay Grade (standardized)				0.105**
				(0.0408)
Penal Essay Grade (standardized)				0.0362
				(0.0385)
Oral Grade (standardized)				0.0123
				(0.0422)
Titles Grade (standardized)				-0.0127
				(0.0429)
Observations	2878	2142	619	619
R-Squared	0.253	0.269	0.274	0.280
Number Admission Cohorts	78	65	20	20
Concurso Fixed-Effect	Yes	Yes	Yes	Yes

Note: This table reports results from estimating equations of the form: $\mathtt{JudgeFE}_j = \mathtt{ExamOutcome}_j'\beta + \delta_{w(j)} + \epsilon_j$, where $\delta_{w(j)}$ are admission cohorts (*Concurso*) fixed-effects and $\mathtt{ExamOutcome}_j$ are the the independent variables of interest in each model in Columns (1) though (4). The dependent variable is Judge FEs, standardized to have unitary standard deviation within exam cohorts. All grades are standardized to have unitary standard deviation within each admission cohort. Robust standard errors in parentheses (* p<0.1, ** p<0.05, *** p <0.01)

Table 8: Robustness – excluding top and bottom performers

	(1)	(2)	(3)	(4)	(5)
Top quintile	0.227***	0.287***	0.229***	0.221***	0.222***
	(0.0587)	(0.0678)	(0.0640)	(0.0668)	(0.0713)
4th quintile	0.147**	0.155**	0.147**	0.141**	0.138**
	(0.0599)	(0.0605)	(0.0600)	(0.0682)	(0.0683)
3rd quintile	0.107*	0.107*	0.107*	0.101	0.0990
	(0.0575)	(0.0575)	(0.0574)	(0.0658)	(0.0657)
2nd quintile	0.152**	0.151**	0.152***	0.146**	0.143**
	(0.0589)	(0.0589)	(0.0589)	(0.0671)	(0.0671)
Observations	2,878	2,644	2,731	2,653	2,506
R-Squared	0.253	0.255	0.256	0.248	0.251
Number Admission Cohorts	78	77	78	78	78
Drop Top 3	No	Yes	No	No	No
Drop Top 5%	No	No	Yes	No	Yes
Drop Bottom 5%	No	No	No	Yes	Yes

Note: This table reports results from estimating equations of the form: $\mathtt{JudgeFE}_j = \boldsymbol{\beta ExamRankQuintile_j} + \delta_{w(j)} + \epsilon_j$, where $\delta_{w(j)}$ are admission cohorts (*Concurso*) fixed-effects. Column (1) reproduces the main result from the Table 7, while Columns (2) through (5) re-estimate the model in subsamples that exclude top and/or bottom contenders, as specified above. Robust standard errors in parentheses (* p<0.1, ** p<0.05, *** p <0.01)

Table 9: Robustness – correlation between admission grades and performance

	(1)	(2)	(3)	(4)
Top quintile	-4.705***			
	(1.082)			
4th quintile	-2.609**			
	(1.092)			
3rd quintile	-1.303			
	(1.073)			
2nd quintile	-2.042*			
	(1.089)			
Final Grade (standardized)		-1.451***	-1.783***	
		(0.350)	(0.629)	
Objective Grade (standardized)				-0.366
				(0.678)
Written Exam Grade (standardized)				-0.568
				(0.625)
Civil Essay Grade (standardized)				-1.311**
				(0.650)
Penal Essay Grade (standardized)				-0.845
				(0.656)
Oral Grade (standardized)				-0.626
				(0.667)
Titles Grade (standardized)				-0.227
				(0.756)
Observations	2879	2143	620	620
R-Squared	0.425	0.393	0.421	0.423
Number Admission Cohorts	79	66	21	21
Concurso Fixed-Effect	Yes	Yes	Yes	Yes

Note: This table reports results from estimating equations of the form: $\mathtt{RankFE}_j = \mathtt{ExamOutcome}_j'\beta + \delta_{w(j)} + \epsilon_j$, where $\delta_{w(j)}$ are admission cohorts (*Concurso*) fixed-effects and $\mathtt{ExamOutcome}_j$ are the the independent variables of interest in each model in Columns (1) though (4). All grades are standardized to have unitary standard deviation within each admission cohort. Robust standard errors in parentheses (* p<0.1, ** p<0.05, *** p <0.01)

For Online Publication

A Appendix figures and tables

Table A1: Nominal Monthly Wages in BRL for Judges and Other Occupation Categories

Year	Judge (1)	Public Federal (2)	Private (3)	Lawyer (4)	Attorney (5)	Defender (6)	Teacher (7)	Health (8)
2003	14022.05	2705.25	797.69	3739.90	13020.26	4861.35	832.61	1948.77
2004	14746.32	2927.70	865.33	3847.76	13302.02	5253.32	873.49	2139.42
2005	16626.96	3234.74	918.03	4026.98	14839.64	5733.74	957.81	2321.79
2006	20315.95	3753.44	975.94	4316.30	18774.73	7388.01	1055.26	2534.81
2007	20874.43	3767.66	1025.27	4389.33	20980.27	7985.20	1114.25	2742.50
2008	21866.22	4509.74	1104.17	4643.52	21682.48	9776.80	1229.63	2983.96
2009	22358.53	5135.16	1196.97	4896.28	21666.93	15589.55	1386.66	3276.08
2010	22820.27	4886.85	1268.28	4823.15	23035.83	16635.17	1478.91	2649.55
2011	22974.93	5946.65	1381.07	5192.09	23443.62	18492.54	1628.27	2901.54
2012	23218.68	6284.55	1517.16	5523.90	24208.90	18549.19	1918.28	3130.53
2013	24497.50	6237.62	1658.07	5890.48	25877.93	18942.78	2012.77	3347.08
2014	26504.97	6691.68	1776.56	6222.14	26462.71	20923.63	2288.11	3598.71
2015	30403.51	7422.67	1928.85	6723.52	30493.29	23818.92	2487.00	3898.43
2016	30767.15	7436.03	2098.78	7071.63	30415.53	24150.48	2702.13	4190.56
2017	30822.91	8357.96	2231.52	7346.37	30939.93	25297.52	2822.94	4385.94
2018	31508.46	8795.70	2273.77	7523.34	31352.18	26167.90	2938.03	4470.61
2019	35910.21	10034.71	2241.54	7479.57	36768.85	28696.86	3026.65	4396.10

Note: This Table reports average nominal wages for judges and various other occupational categories for Brazil between 2003 and 2019 sourced from RAIS.

Table A2: Variance decomposition - alternative samples

	Baseline	Full sample	4-month spell	6-month total
Cases disposed (IHS)	3.27	3.65	3.13	3.54
Judge FE	0.23	0.12	0.29	0.23
Court FE	0.34	0.21	0.41	0.33
Connected Set FE	0.07	0.07	0.08	0.07
Judge+Court FE	0.31	0.20	0.38	0.30
Adju R-squared	0.45	0.41	0.46	0.43
Observations	727,835	988,160	650,998	795,669
Number Judges	10,479	11,273	10,000	10,092
Share movers	0.81	0.92	0.78	0.80

Note: This table presents the variance decomposition exercise using estimates from the two-way fixed effects model in equation (1). Column (1) presents our baseline sample restriction, while the following columns consider alternative samples indicated in each column. All variance estimates are obtained using the split-sample bias-correction method. Share of movers refers to the share of judges in each sample that were observed working in two or more courts throughout the period.

Table A3: Main results – correlation between admission grades and performance

	(1)	(2)	(3)	(4)
Top quintile	0.227^{***}	0.218^{***}	0.184^{***}	0.116^{*}
	(0.0587)	(0.0595)	(0.0591)	(0.0605)
4th quintile	0.147^{**}	0.165^{***}	0.113^{*}	0.0847
	(0.0599)	(0.0601)	(0.0598)	(0.0607)
3rd quintile	0.107^{*}	0.115^{**}	0.0815	0.0457
	(0.0575)	(0.0580)	(0.0587)	(0.0592)
2nd quintile	0.152^{**}	0.173^{***}	0.121^{**}	0.128^{**}
	(0.0589)	(0.0603)	(0.0601)	(0.0608)
Observations	2,878	2,878	2,782	2,726
R-Squared	0.253	0.381	0.228	0.283
Number Admission Cohorts	78	78	76	74
Concurso Fixed-Effect	Yes	Yes	Yes	Yes
Sample restriction	Baseline	Full sample	4-month spells	6-month total

Note: This table reports results from estimating equations of the form: $\mathtt{JudgeFE}_j = \boldsymbol{\beta ExamRankQuintile_j} + \delta_{w(j)} + \epsilon_j$, where $\delta_{w(j)}$ are admission cohorts (*Concurso*) fixed-effects. The dependent variable is Judge FEs, standardized to have unitary standard deviation within exam cohorts. Each column refers to a different sample restriction used to estimate Judge FEs using the two-way fixed-effects model. Robust standard errors in parentheses (* p<0.1, ** p<0.05, *** p <0.01).

Table A4: Detailed descriptive statistics in estimation sample

	Mean	SD	Median	N
Panel A - Judges				
Male	0.60	0.49	1	10,218
# Courts by Judge	4.28	3.56	3	10,479
Number of months Judge is observed	49.99	21.05	56	10,479
# of Courts at Judge-Month level	1.39	0.83	1	523,813
Number Municipalities Judge ever works in	2.28	1.69	2	10,479
Unique number of months per judge-court pair	16.23	17.15	9	44,850
Panel B - Courts				
# Judges by Court	4.96	3.60	4	9,048
# of Judges at Court-Month level	1.38	0.87	1	528,483
Civil Court	0.22	0.42	0	9,048
General Court	0.20	0.40	0	9,048
Small-stakes Court	0.18	0.39	0	9,048
Criminal Court	0.16	0.37	0	9,048
Family Court	0.10	0.30	0	9,048
Other Courts	0.13	0.34	0	9,048
Panel C - Output measures				
Cases Disposed (on merit)	40.13	50.09	22	727,784
Total Hearings (presided or held)	34.88	46.39	17	716,736

Note: This table reports descriptive statistics for key variables in the sample used to estimate the two-way fixed-effects model, where outcome variables are trimmed at the top 1%, judge-court spells shorter than three months are dropped and only observations in the largest connected sets within each state are used.

Table A5: Pairwise correlations between performance in admission exam phases

	Objective	Written	Civil case	Criminal case	Oral	Titles
Objective	1					
Written	0.0283	1				
Civil case	0.0127	0.0144	1			
Criminal case	0.0276	0.0886**	0.0604	1		
Oral	0.0352	0.103**	0.112***	0.0656	1	
Titles	-0.0403	0.0825**	0.122***	0.0265	0.154***	1

Note: This table reports pairwise correlations between residualized grades in each one of the six phases of admission examinations. Residues are obtained by regressing grades on admission exam fixed-effects so all grades are represented as deviations from exam average. Sample is restricted to exams with available grades for all exams (N = 619).

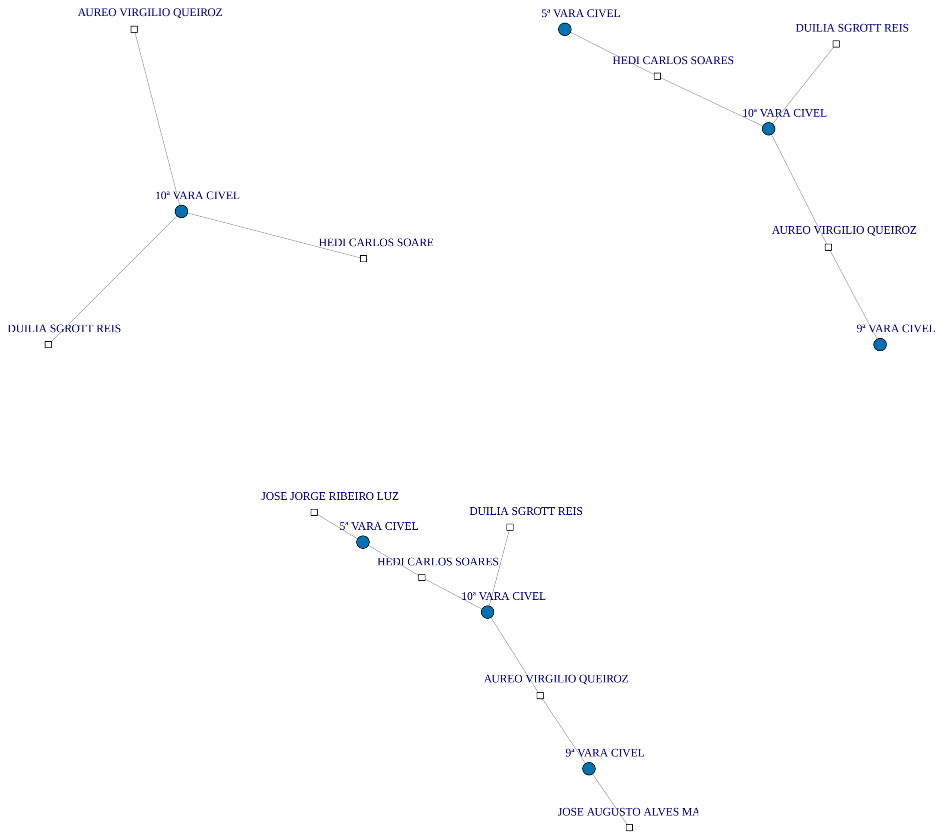
B Connected sets in the data

Connected sets are defined as as groups of organizations (courts) and individuals (judges) connected by "movers", workers who are observed in more than one organization. In our context, there are two sources of variation that allow us to construct connected sets. First, judges are often observed working in more than one court in the same month, allowing us to create connections even within a single period (month). Figure A1 below illustrates this fact. The top-right figure show three judges observed working in the 10th Civil Court of Porto Velho, in the state of Rondonia, during the month of May 2013. As we can see in the top-right figure, two of these judges also worked in additional courts in that same month – in the 5th Civil Court and the 9th Civil court. These two courts, and all the judges working in them on that same month, are also part of the original connected set – the bottom figure shows that two additional judges were working in these courts in May, and our connected set has expanded.

This within month connections is only one source of variation used to build connected sets. Since we have a panel that covers 76 months, we can build all connections that happened at any point in that period. Figure A2 takes a broader view and presents all connections in the states of Rondonia and Amapa, two small states that allow for better visualization of the judge-court networks. The top two figures and the bottom-left one shows all connections for the states of Rondonia in three periods: 2009, 2009-2010 and 2009-2011. Note that when only connections in 2009 are considered, connected sets are large but multiple: clusters of judges and courts are often not connected to other parts of the network. When we explore judges' movements across several years, on the other hand, the network becomes more densely connected: if we consider the entire 2009-2015 period, all judges and courts *within each state* belong to a single connected set, as shown in bottom-right figure[40]. Since judges are hired to work in a specific state, nonetheless, the figure also shows that each state is a separate connected set: judges in the state of Rondonia, for example, are never observed matched to courts in Amapa, and vice-versa.

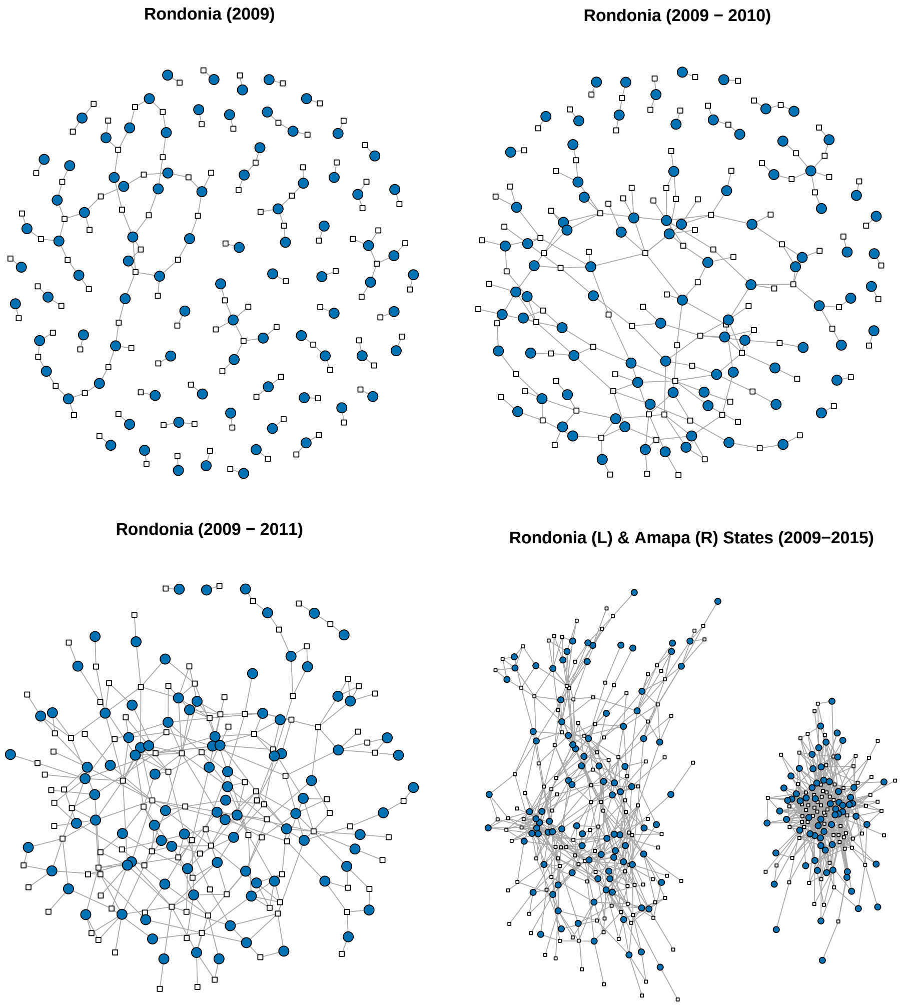
[40] All graphs represent connections in the sample used to estimate the two-way fixed-effects model, and therefore include a single connected-set by construction. As discussed above, nonetheless, the largest connected set within each state often includes over 95% of all observations.

Figure A1: Construction of connected sets in the data (Rondonia – May 2013)



Note: These graphs present the a selected network of judges (white squares) and courts (blue dots) in the state of Rondonia. Connections between dots and squares represent judges being observed working in a court in the month of May 2013. Starting from the top-left and moving clockwise, the graph expands the connected set by adding courts and judges observed paired in that month. All graphs use data from the sample used to estimate the two-way fixed-effects model.

Figure A2: Visualizing connected sets in the data



Note: These graphs present the networks of judges (white squares) and courts (blue dots) for the states of Rondonia and Amapa. Connections between dots and squares represent judges being observed working in a court in the referred period. The top-left figure presents connections observed in the state of Rondonia in 2009; the top-right includes connection observed in 2009 and 2010, while the bottom left presents connections in the period 2009-2011. The bottom right figure presents the universe of connections observed in in the entire panel for the states of Rondonia and Amapa. It highlights that there are no connection across states, since judges from one state are never observed working in a different state. All graphs use data from the sample used to estimate the two-way fixed-effects model, and therefore within each state all observations are connected by construction.