

Manna from Heaven for Judges: Judges’ Reaction to a Quasi-Random Reduction in Caseload

*Christoph Engel** and *Keren Weinshall*

What is the impact of caseload on judicial decision making? Is increasing judicial staff effective in improving judicial services? To address these questions, we exploit a natural, near-randomized experiment in the Israeli judiciary. In 2012, six senior registrars were appointed in two of the six magistrate’s court districts. The choice of districts was motivated by reasons unrelated to judicial performance. In these two districts, the civil caseload per judge was substantially reduced. We find that the reduction had a significant impact on the process and outcomes of judicial decision making. Judges working in courts with reduced caseload invested more resources in resolving each case. The effect is mostly to the advantage of plaintiffs, who were more likely to win, recover a larger fraction of their claims, and be reimbursed for litigation costs. We discuss the implications for judicial management and theories about judicial decision making.

I. INTRODUCTION

Judiciaries worldwide claim to be facing increasing caseloads, presumably leading to inaccessible and delayed justice, reduced quality of judgments, judicial burnout, declining public confidence in the courts, and the vanishing trial phenomenon.

The perceived “caseload crisis” is interpreted in very different ways though, which we place on a spectrum between two distinct theoretical approaches: “judicial” versus “managerial.” According to the judicial approach, the crisis is mostly attributable to socio-

*Address correspondence to Christoph Engel, Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Straße 10, D 53113 Bonn, Germany; email: engel@coll.mpg.de. Weinshall is Edward S. Silver Chair in Civil Procedure, Hebrew University, Jerusalem

Professor Weinshall’s research was supported by the Israel Science Foundation (Grant No. 1404/17). Helpful comments by Jennifer Doleac, Stefanie Egidy, Jens Frankenreiter, Jacob Goldin, J. J. Prescott, Holger Spamann, Eyal Zamir, and audiences at the Workshop for Empirical Study of Public Law & Human Rights in Jerusalem (June 2019), Workshop on Judicial Decision-Making at the IVR Congress in Lucerne (July 2019), and the Conference on Empirical Legal Studies in Clermont (November 2019) are gratefully acknowledged.

© 2020 The Authors. *Journal of Empirical Legal Studies* published by Cornell Law School and Wiley Periodicals LLC. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

legal developments that generate an increasing volume and complexity of cases, outpacing a smaller increase in the number of judges. This line of thought advocates raising the supply side by appointing more judges, registrars, and judicial staff (Richman & Reynolds 2012; Levy 2013; Adler 2014; Stras & Pettigrew 2010). On the other hand, the managerial approach claims that clogged courts are, to a large degree, the result of mismanagement and inefficiency, outdated legal procedures, and judicial passivity (Castro & Guccio 2015; Dalton & Singer 2014; Mitsopoulos & Pelagidis 2010; Moffett et al. 2016). In this interpretation, increasing courts' budgets or appointing new judges is not perceived as an effective remedy (Agrast et al. 2011; Beenstock & Haitovsky 2004; Heaton & Helland 2011; Webber 2006).

The managerial approach corresponds with theories of judicial behavior that view judges as individuals motivated in part by leisure preferences (Posner 1993). Thus, when more judges are appointed to a court with a given caseload, the judges who maximize personal utility are predicted to react by working less and taking more time for leisure. On the other hand, the judicial approach asserts that judges are predominantly motivated to do a good job (Engel & Zhurakhovska 2017). If so, when judges have lighter caseloads they are expected to use the extra time to better resolve their remaining cases.

In this article, we study this question empirically: Is increasing judicial staff effective in improving judicial services? Does it impact the substantive outcomes of decisions? To address these questions, we exploit a natural, near-randomized experiment in the Israeli judiciary. In 2012, senior registrars were appointed for a pilot program to conduct evening hearings in two of the six magistrate's court districts. Assigning the registrars to two specific districts was motivated by the location of the court buildings in the city, and entirely unrelated to caseloads, work procedures, or judicial performance. These appointments constituted an increase of around 15 percent in judicial staff devoted to civil litigation in the two treated districts and to a decrease of an estimated 14 percent of the caseload for each civil judge active in these courts. Other magistrate's court districts did not experience changes in the number of judicial staff or caseload during the period studied. We test a series of indicators for resource investment, court performance, and outcomes of judicial decision making before and after the exogenous shock on caseload, compared to those in untreated districts.

The analysis is based on an original dataset of 2,085 small claims, randomly sampled from each of the six court districts, pre- and post-treatment. We chose to focus on small claims because they offer several methodological advantages. Chiefly, small claims are assigned in periodic rotations, with cases randomly distributed among judges and senior registrars. Additionally, disputes are fairly homogeneous, involving small amounts, mostly between private individual litigants, with no legal representation. The unrepresented "one-shooter" (Galanter 1974) petitioners in small claims are less likely to react strategically by filing more cases in the treated courts or reducing their claims to qualify for small claim procedure (as supported by the data, showing no displacement effects). The similarities between small claims further reduce concerns about unobservable variables causing differences in decisions before and after treatment in treated and untreated courts. As a further safeguard against potential imbalances in the sample of cases, we employ coarsened exact matching (CEM). Results after matching are consistent with results that use the raw data.

Prior to 2012, performance indicators and case outcome variables do not systematically differ between court districts. These indicators essentially remain constant after 2012 in the untreated courts. In the treated courts, however, we observe significant and strong changes. Our indicators suggest that judges in the treated courts invest the additional time in better resolving their assigned cases. For example, they use more laborious means of evidence (are more likely to hear witnesses), are less likely to write summary judgments, more likely to decide cases on the merits, and write more elaborate opinions. These changes are largely to the benefit of plaintiffs: plaintiffs are at least 18 percent more likely to obtain a full victory, to recover at least an additional 13 percent of the sum they claimed, and to have their litigation cost reimbursed by defendants.

The article proceeds as follows. Section II derives two competing hypotheses from the literature. Sections III and IV introduce the research design, identification strategy, dataset, and coding scheme. Section V reports our results. Section VI discusses the implications and possible mechanisms explaining the observed pro-plaintiff effect of a reduction in caseload. Section VII concludes with a discussion on the study's limitations, its possible generalizability, and topics for future research.

II. THEORY AND HYPOTHESES

A. Macro Court-Administration Level: Theoretical Approaches to the Study of Court Caseloads

We place the wide scope of literature on court caseloads on a spectrum between two distinct theoretical approaches that we classify as “judicial” versus “managerial.” We see these as ideal types in the sense of Weber ([1922] 1978). Hence, we use the two approaches as conceptual tools that facilitate our comparative analysis by presenting the purest form of theoretical approaches composed of all key features, though most studies and scholars adopt a more nuanced and balanced perspective.

We label the first approach “judicial” because it is most common among jurists, and particularly judges, who view the judiciary as a unique legal-democratic institution, responsible for providing justice—not as a mere commodity. In contrast, the competing managerial approach sees the judiciary as a service provider delivering justice, much like the health-care system provides health services. These different perspectives are reflected in a different understanding of what constitutes a “caseload crisis,” what causes judicial burdens, and how to alleviate overload.

The judicial approach developed mostly in the United States in the late 1960s, in response to what was later called a “litigation explosion” or “crisis of volume in the court’s docket” (Carrington 1969; Meador 1974). The expansion of judicial burdens is viewed as a direct result of supply and demand functions. Trends in the socio-legal environment have led to an increasing volume of cases, outpacing a smaller increase in the number of judges. The judicial approach focuses on what it perceives as nondiscretionary overload, caused by factors external to the courts. This understanding leads to obvious solutions for dealing with increased demand for justice: raising the supply side by appointing more judges, using senior or visiting judges (Adler 2014; Levy 2013; Richman & Reynolds 2012), or hiring additional staff (Stras & Pettigrew 2010). New and presumably efficient case management

techniques suggested by the managerial approach, such as actively encouraging settlements or developing procedural methods of swift case disposal (summary disposition, limiting oral arguments, issuing unpublished opinions, etc.), are often criticized by proponents of the judicial approach as “redefining sub silentio our standards of what constitutes rational, fair, and impartial adjudication” (Resnik 1982:980). Moreover, scholars writing in the judicial approach argue that some reform plans, more than the overload itself, are impairing the quality of justice, making it less accessible and less respectable, and transforming judges into bureaucrats working on an assembly line of justice.

On the other end of the spectrum, the managerial approach emerged in the late 1970s (Nagel et al. 1978; Posner 1985) and began to flourish in the new millennium, mostly among multidisciplinary scholars and social scientists. Scholars writing in the managerial approach often hold a different understanding of the present state of courts; rather than a crisis of high workloads, they describe a stagnation in caseloads over the past three decades (see for U.S. federal courts Habel & Scott 2014; Moore 2015). Other U.S. scholars adopting this approach assert that there is no longer a crisis in judicial caseload because gains in efficiency have mitigated concerns over higher case volumes (Posner 2006) or, alternatively, that judges were underworked in decades prior to the 1980s rather than being overburdened since then (Baker 2006). Moreover, they argue that if indeed a caseload crisis exists, it is not merely the result of an increase in nondiscretionary caseload, but also a consequence of mismanaged and inefficient courts, outdated legal procedures, and judicial passivity that allows strategic filings and procrastination (Castro & Guccio 2015; Dalton & Singer 2014; Mitsopoulos & Pelagidis 2010; Moffett et al. 2016). Accordingly, simply increasing the courts’ budgets (Webber 2006; Agrast et al. 2011; Heaton & Heland 2011) and appointing new judges are not the cures for overburdened courts.¹

In recent years, the managerial approach seems to be taking hold outside the United States, particularly in civil-law jurisdictions.² Bodies such as the European Commission for the Efficiency of Justice (CEPEJ—formed in the early 2000s) work to develop measures and empirical tools to analyze the functioning of judicial systems, recommend timeframes, and optimize judicial time management in European countries. Empirical studies conducted on court productivity in different judiciaries mostly support the managerial approach’s claim regarding the futility of increasing judicial manpower. In a widely cited paper, Beenstock and Haitovsky (2004) analyzed panel data on Israeli courts and found that judicial productivity was exogenous to the number of judges. The higher the caseload, the more cases each judge handled. Beenstock and Haitovsky thus suggested

¹Instead, the managerial approach recommends implementing previously mentioned procedural innovations to expedite or terminate cases without trial; applying new technologies (EI Bialy 2016) and case management systems (Coolson 2009); reforming attorney fee regimes to deter frivolous lawsuits (Miner 2012–2013); changing court culture by teaching judges to proactively manage cases (Bar-Niv Burnovski et al. 2010); and adopting performance measures to better manage courts (National Center for State Courts 2005; CEPEJ 2015; Maayan et al. 2012; Langbroek & Kleimann 2016).

²Perhaps because the civil-law tradition views the judicial role as more technical: to strictly apply codified law. See, for example, Lienhard and Kettiger (2011) (Switzerland), Mitsopoulos and Pelagidis (2010) (Greece), Rosales-López (2008) (Spain), and Castro and Guccio (2015) (Italy).

that a judge's productivity depends mainly on the existing demand. Several studies using observational court data found similar evidence, for example, in Brazilian courts (Yeung & Azevedo 2010), Slovenian courts (Dimitrova-Grajzl et al. 2014), and Spanish courts (Rosales-López 2008).

However, in recent years a small number of empirical studies support the opposing judicial approach's claim. Jonski and Mankowski (2014), for example, criticized the methodology and findings of Beenstock and Haitovsky, suggesting that their results may be attributed to a modeling flaw in neglecting time constraints faced by judges. Gomes et al. (2016) employed measuring strategies suggested by Jonski and Mankowski and found that enhanced judicial staff in the Brazilian state courts improved the courts' productivity and performance. We add a (quasi) natural experiment to this literature.

B. Micro Judicial-Behavior Level: What Do Judges Maximize?

These conflicting results call for a closer look at the micro level: What is it judges aim to achieve? Beenstock and Haitovsky built on the influential work of Posner (1993) and Cooter (1983), suggesting that judges are individuals seeking to maximize their expected utility from leisure and minimize the effort required for case adjudication. However, judges also draw negative utility from increasing backlogs that may diminish their prestige and prospects for promotion. Judicial behavior is explained by this tradeoff. When more judges are appointed to a court with a given caseload, judges maximizing personal utility are predicted to consume at least some of the additional freedom by working less and taking more time for leisure.

The competing approach asserts that judges are not mainly motivated by maximizing personal profit or leisure, but are motivated to do a good job (Engel & Zhurakhovska 2017). If so, judges enjoying a decrease in their caseloads can use the extra time to do a better job on the cases that remain with them. This could mean deciding cases faster or reducing the backlog of cases, as shown in the productivity estimates of Brazilian courts in Gomes et al. (2016). Another possibility is that the newly available time is utilized by motivated judges to invest more time and resources in the remaining cases. This would not result in higher court productivity or an increase in the quantity of resolved cases, but it might improve the quality of judicial services and output. For example, judges who are under less pressure could use more laborious means of evidence, could be less likely to dismiss borderline cases on procedural grounds, more likely to decide cases on the merits (rather than pursuing settlements), and could write more elaborate opinions or even invest time in professional learning unrelated to specific cases, such as reading about new case law and legislation.

Empirical studies associated with the competing theoretical approaches suggest that, indeed, caseloads affect judicial decisions in both form (disposition type) and substantive case outcomes. With regard to modes of disposition, Epstein et al. (2013) showed that high caseloads encourage judges to employ doctrines that allow early dismissal of cases, such as standing and ripeness, and deter judges from engaging in the time-consuming task of writing dissents (also see Epstein et al. 2011; Narayan & Smyth 2007). High caseloads were also found to reduce the volume of cases that go to full trial

(Galanter 2004). The driving force behind settlements and plea bargains is often an overburdened judge seeking to alleviate her load, though other stakeholders may also induce these outcomes. Yang (2016), for example, suggests that prosecutors respond to judicial scarcity by using their discretion to screen out cases, thus dismissing more cases during judicial vacancies and offering more favorable plea deals in order to obtain guilty pleas and avoid going to trial.

Caseloads seem to directly affect substantive outcomes even in fully adjudicated cases. In the criminal law setting, higher caseload pressures correlate with harsher trial penalties: defendants who insist on lengthy jury trials in clogged courts are penalized for wasting scarce resources and often receive a more severe sentence (Ulmer et al. 2009; Ulmer & Bradley 2006; Ulmer & Johnson 2004). Another example is the use of sentencing guidelines: Sisk et al. (1998:1484) found that judges with higher criminal caseloads are more likely to adopt sentencing guidelines, presumably because they view the guidelines as a labor-saving measure that can streamline and coordinate the sentencing process. Best and Tiede (2015) showed that, indeed, when federal district court judges have heavy caseloads, they save time by choosing not to depart downward from the guidelines in cases that may warrant departures, thus leading to longer imprisonment under heavy caseloads.³ In civil cases, caseload was found to influence not only outcomes and doctrinal development (de Mot et al. 2015; see, specifically, for bankruptcy cases Ponticelli & Alencar 2016; Iverson 2017), but also lateral decisions, rendering higher fees to attorneys as caseload rises (Eisenberg et al. 2014). Similarly, Helland and Klick (2007) found that when facing an increase in caseload, judges authorize higher attorney fees in class action suits, presumably in order to induce the rapid termination of these cases. Finally, time pressure can also reinforce a tendency to give greater deference to primary decisionmakers, thus decreasing the likelihood of winning administrative cases and civil or criminal appeals (Giles et al. 2015; Guthrie & George 2004; Huang 2011).

Judges are not only rational actors striving to optimize their use of time. Although most of the aforementioned studies focus on judges' strategic choices of the less time-consuming legal outcome, their decisions may also be affected by the physical and emotional fatigue, decline in cognitive performance, and elevated stress levels associated with high workloads.⁴ Research has shown that under time pressure, judges are more vulnerable to heuristics and biases. For example, rulings were found to be more inconsistent when judges face a high caseload (Norris 2018) and busy judges were found to expend less effort by according higher weight to non-legal cues, such as litigants' race or gender, to determine case outcomes (Guthrie et al. 2000, 2007; Rachlinski et al. 2008).

³On the other hand, Johnson (2006) did not find statistically significant correlations between judges' caseload pressure and sentencing outcomes in Pennsylvania.

⁴For studies showing that sleep-deprived judges and judges in a bad mood rule more harshly, see, for example, Cho et al. (2016) and Eren and Mocan (2018).

C. Hypotheses

We draw two contradicting hypotheses from the reviewed body of literature. The null hypothesis (H_0) is in line with the extreme versions of both the managerial approach and “maximizing utility judges” theories, according to which an increase in judicial staff will not significantly affect judicial services. Alternatively, the judicial approach or “motivated judges” hypothesis (H_1) predicts that the growth in judicial staff will lead to an increase in the resources invested in judicial decisions and improvement of services. The following resource and performance indicators are available in our dataset: Has a hearing been held before case resolution? Has a witness been heard? Has the case been decided on the merits, or by summary judgment, settlement, default judgment, or dismissal? Has the court delivered a written ruling? How long was the ruling, and how long has the case taken to conclude? All these indicators depend on the effort a judge exerts. In this mechanical sense, a high score on any of the indicators supports the judicial view. Some of these indicators also imply a higher risk of being overturned on appeal.⁵ This in particular holds for a decision on the merits, and the presence of a written ruling. Yet more effort and more scope for scrutiny from the appellate court does not, of course, automatically imply normatively more desirable outcomes. This is why we also consider substantive case outcomes. We have data on the following outcome indicators: Has plaintiff won fully, or at least partially? To which degree has plaintiff won (measured as the recovery amount/sum of claim)? Did the court shift cost, to either party, or to the plaintiff in particular?

III. THE QUASI-EXPERIMENT

We exploit a naturally occurring, near-randomized experiment to test the alternative hypotheses. In 2012, the Israeli Ministry of Finance, together with the Courts’ Directorate, launched the Pilot Program for Evening Court Hearings in Israeli Magistrate Courts. According to the official report evaluating the program’s effectiveness, the purpose of the pilot was to test the effectiveness of evening shifts in “reducing the backlog of civil cases in the courts, while improving the service rendered to court users. In order to achieve these goals, six senior registrars were appointed for two years to conduct hearings in second evening shifts (between 4–10 PM) in civil suits up to NIS 75,000” (Aviv & Galon 2015:4; also from the interim report, Weinshall-Margel & Galon 2013).

⁵We are unable to add appeal rate as an indicator for litigant satisfaction or appeal-outcome as a measure for accuracy because only two cases in our sample were appealed. According to official court reports, the rate of appeals submitted in all resolved small claims was 1.5 percent in 2011; 1.3 percent in 2012; 1.1 percent in 2013; 1.2 percent in 2014; 1.3 percent in 2015; and 1.2 percent in 2016. Unfortunately, the appeal-rate data are not separated according to the court district. Note that trends in appeal rates with regard to civil litigation at large correspond, from 2.9 percent in 2011; to 2.8 percent in 2012; 2.6 percent in 2013; 2.7 percent in 2014; and 2.8 percent in 2015 and 2016. See annual reports on the Israeli judiciary’s activity; links in note 11.

Israeli law grants senior registrars all judicial powers in small-scale civil cases.⁶ The six registrars were appointed to two of the Israeli judiciary's magistrate's court districts: three to the Central District magistrate's courts and three to the Jerusalem District magistrate's courts. These courts were chosen due to their geographical and physical attributes (Aviv & Galon 2015). The main court buildings in these two districts are situated in the centers of large cities, in friendly urban locations that can offer other services during evening hours (such as food, parking, shops). Moreover, the Jerusalem court is just across from the Jerusalem city hall and the Central district Rishon Le'zion court is placed in the city's government quarters. Therefore, security officers were already placed near these courts during the hours designated for the second shift. Critically for our approach, the choice of districts was not motivated by past comparative performance of the courts, by expectations about their future performance, or by the comparative effectiveness of deploying additional judicial personal.⁷

Appointments were made gradually, beginning in March 2012 and continuing until September 2012. Although the pilot was declared as a two-year trial, the six registrars are still employed in their assigned courts, though they now mostly work in morning shifts.⁸ The new registrars were generally assigned newly filed cases, each allotted a caseload of about 225 civil procedures per month, mostly fast-track procedure cases⁹ or small claims. Each newly appointed registrar was provided with a court typist, secretary, and legal aid (usually an intern) equivalent to those provided to other judges and senior registrars (Aviv & Galon 2015).

While the pilot was designed by the judiciary as an (imperfect) field experiment to test the effectiveness of evening shifts (compared to morning shifts), we utilize its design to study the effects of an increase in judicial staff and reduction in caseloads. The appointments constituted an increase of 11 percent in the number of judges and senior registrars ("judicial staff") devoted

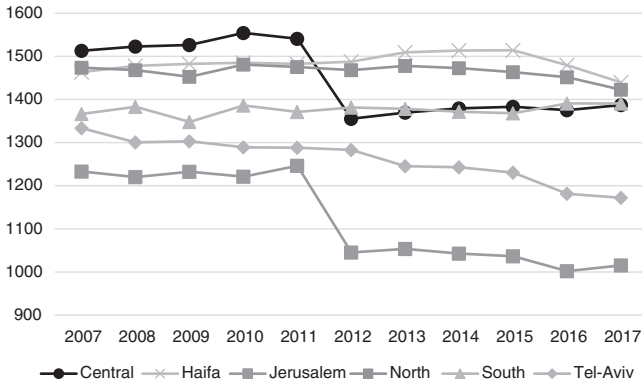
⁶Section 85A of the Courts Act, 1984 defines these small-scale cases as all claims up to NIS 75,000, with some exceptions such as bodily injuries and class action cases. As detailed in the next section, all cases submitted as a small claim are limited by law to NIS 33,500 and are thus well within the mandate of senior registrars.

⁷In fact, annual official reports on the Israeli judiciary's activity show that the two chosen districts are quite apart from one another in terms of their caseloads and performance. In the four years prior to the pilot, the average number of civil cases opened and resolved per civil judge/registrar was the highest in Central district and lowest in Jerusalem district. Note, however, that until the 2012 pilot, differences in caseload per staff were relatively minor in all districts, as the judiciaries' staffing policy was based on the number of cases in each district. See exact trends in Figure 1 and links to reports are in note 11.

⁸In practice, during the two-year pilot period, only four of the six senior registrars actually held hearings during the evening shift and the definition of "evening" gradually changed to the time slot between 2:30 and 7:00 pm. Two other registrars were transferred shortly after their appointments to morning shifts (Aviv & Galon 2015). Pilot results showed no differences in the productivity of senior registrars working in evening shifts compared to their counterparts in "regular" morning shifts. However, working in evening shifts was strenuous for judicial and secretarial staff employed in the late hours. See more in Aviv and Galon (2015) and Weinshall-Margel and Galon (2013).

⁹According to the Civil Procedure Regulations of 1984, fast-track procedures are used for civil cases being tried in the magistrate's courts with a financial value of less than NIS 75,000.

Figure 1: Annual civil filings per judge/senior registrar by district, 2007–2017.



to civil litigation¹⁰ in the Central district magistrate's courts and 19 percent in the Jerusalem district magistrate's courts. According to the judiciary's official reports, the judicial staff and case-loads remained almost constant in Israel's four other court districts during the pilot period.

The fact that the intervention was flagged out as a test of a different question (is trial by night a good idea?) is a helpful safeguard against a potential Hawthorne effect. Had the intervention been introduced as a test of a reduction in judicial workload, judges in the treated courts might have tried to convince the legislator, or the court administration for that matter, about the effectiveness of the program by putting in extra work while observed (and maybe replace it with a more comfortable life once the intervention is turned permanent). Yet no judge had reason to expect that the program would be used as a test of judicial effectiveness or the quality of judicial decision making.

Figure 1 illustrates the type and the magnitude of the intervention. It is based on data we extracted and analyzed from annual court reports.¹¹ The figure presents the yearly average number of opened civil cases assigned to a judge or senior registrar in each of the magistrate's court districts from 2007 to 2017. Until 2012, the graph conveys a common trend of relative stability in all districts. In 2012, after the appointment of the six senior registrars ("the

¹⁰Civil litigation is defined as including all civil disputes, including small claims, regular civil procedure, fast-track civil procedures, special civil procedure, and investigating civil motions, and excluding family cases and computerized claims.

¹¹See annual reports on the Israeli judiciary's activity under the Freedom of Information Act 1998 at https://www.gov.il/BlobFolder/reports/freedom_of_information_report2010.pdf; http://elyon1.court.gov.il/heb/haba/docho/doc/1-6_2011.pdf; https://www.gov.il/BlobFolder/reports/statistics_second_half2012/he/second2012.pdf; https://www.gov.il/BlobFolder/reports/statistics_annual_2013/he/annual2013.pdf; https://www.gov.il/BlobFolder/reports/statistics_annual_2014/he/annual2014.pdf; https://www.gov.il/BlobFolder/reports/statistics_annual_2015/he/annual2015.pdf; https://www.gov.il/BlobFolder/reports/statistics_annual_2016/he/annual2016.pdf; https://www.gov.il/BlobFolder/reports/statistics_annual_2017/he/annual2017.pdf.

treatment”), a sharp decline in caseloads is visible in the two treated (Central and Jerusalem) districts: caseload per judge decreases by 12 and 16 percent, respectively.¹² Caseload per judge did not change in the remaining (“untreated”) districts. Note that both pre and post intervention, and both in the treated and the untreated districts, average caseload per judge is flat. This speaks against a displacement effect: seemingly, after the intervention, cases have not been shifted away from untreated to treated districts.¹³ We also find no signs of displacement between ordinary claims (which do not benefit from additional manpower) and small claims (for which additional judicial personnel are deployed). In the treated courts, the frequency of small claims close to the statutory limit of NIS 33,500 did not increase compared with 2011.¹⁴ Additionally, the annual total number of small claims filed did not increase during the studied time period, and the share of small claims from all civil claims remained constant.¹⁵

IV. MICRO DATA AND IDENTIFICATION STRATEGY

We collect and analyze a representative random sample of 2,085¹⁶ small claims from 262 judges at all six court districts, pre- and post-treatment (note that we do not include cases decided by the six newly appointed registrars in the analysis). The sample consists of

¹² Cases assigned to the registrar are limited by the sum of claims and considered to be less complicated. Thus, the reduction in weighted caseload is a bit smaller: 11 percent of average weighted caseload per judicial staff in the Central district and 14 percent in the Jerusalem district. Case weights are calculated by comparing the varying amounts of judicial time needed to process different case types. Israel uses an advanced system of event-based case weighting. The weights are modeled based upon the interaction between the average frequency of all the events comprising the processing of a case in the courts and the average complexity of those events. For a detailed explanation on case weights in the Israeli judiciary, see Weinshall-Margel et al. 2015.

¹³ Thus, different from Yang’s (2016) findings, litigant decisions to file cases did not respond to the changes in judicial caseload. We suggest that this is also related to the nature of small claim plaintiffs, as unrepresented one-shooters in the court (see next section).

¹⁴ These data are available from the authors upon request.

¹⁵ According to annual reports on the Israeli judiciary (see note 11), 42,709 small claims were submitted to all court districts during 2010—constituting 14.4 percent of all civil claims submitted to magistrate courts; 42,106 in 2011—14.3 percent of all civil claims; 41,905 in 2012—14.5 percent of all civil claims; 41,618 in 2013—14.3 percent of all civil claims; 41,568 in 2014—14.4 percent of all civil claims; and 42,562 in 2015—14.5 percent of all civil claims. The annual data published on the number of small claims are, unfortunately, not separated according to the court district.

¹⁶ One-thousand of the cases were coded as part of the work of the Israeli Courts Research Division; see Gali Aviv & Inbal Galon, Small Claims Court Dataset, Israeli Courts Research Division (September 2016) <http://elyon1.court.gov.il/heb/Researchpercent20Division/doc/scc.xls>. We enlarged the sample by 1,085 cases and are happy to share the dataset upon request. The supplementary cases from 2014 were sampled using the same randomization techniques employed by Aviv and Galon: the “Net Hamishpat” search engine provided a report for all resolved small claims in each court district, organized by case ID number. The report is divided to webpages, covering 42 to 44 cases in each page. The first case in each search page was selected and sampled. For cases after treatment, if the selected case was resolved by a newly appointed registrar in the treated courts, the case below in the search page was also sampled (this is because we did not want to include cases decided by the new registrars in the analysis).

Table 1: Sample Composition and Dependent Variables

	Obs	Mean	SD
Sample			
Cases from affected districts (1 = yes)	2,085	0.320	0.467
Cases after treatment date (1 = yes)	1,925 ^a	0.502	0.500
Cases from affected districts after arrival of new personnel (treated) (1 = yes)	1,925	0.179	0.384
Dependent Variables			
Case disposition			
Adjudicated on the merits (1 = yes)	1,925	0.247	0.431
Summary judgment (1 = yes)	1,925	0.088	0.283
Settlement (1 = yes)	1,925	0.432	0.496
Default judgment (1 = yes)	1,925	0.076	0.265
Dismissal (due to inaction, failure to pay a fee, withdrawals, or lack of jurisdiction) (1 = yes)	1,925	0.154	0.361
Substantial case outcomes			
Plaintiff wins fully (1 = yes)	1,925	0.131	0.337
Plaintiff wins fully or partially (1 = yes)	1,925	0.718	0.450
Recovery amount (NIS)	1,669 ^b	3,418	4,975
Degree of plaintiff success (recovery amount/sum of claim)	1,669	0.366	0.367
Court orders shift of cost (to either party) (1 = yes)	1,669	0.306	0.461
Cost shifted to winning party (1 = yes)	1,669	0.262	0.440
Characteristics of procedure			
Hearing held (1 = yes)	1,919 ^c	0.716	0.451
Number of hearings	1,919 ^c	0.881	0.714
Witness heard (1 = yes)	1,918 ^d	0.584	0.493
Number of witnesses	1,918 ^d	10.288	10.349
Judgment written in chamber (1 = yes, 0 = oral judgment)	1,925	0.541	0.498
Written judgment has more than 1 page (1 = yes)	1,925	0.362	0.481
Number of pages in judgment	1,925	10.700	10.295
Length of proceedings (0 = below median [185], 1 = between median and 500, 2 = above 500)	1,925	0.532	0.563
Length of proceedings (in days)	1,925	210	127

^aTwenty of the 2,085 cases were filed before treatment and finally resolved in 2014, and 140 cases were assigned to the newly appointed six registrars in treated courts—these cases were not included in the analysis.

^bFifty-seven cases include nonmonetary relief. Other missing values are in cases resolved via out-of-court-settlement in which the parties did not inform the court of the settlement amount.

^cMissing values for six cases.

^dMissing values for seven cases.

cases resolved by all manners of disposition, including settlements, dismissals, default, and summary judgments. Case files were accessed using the Israeli judiciary's official computerized case routing and management system (Net Hamishpat) and the Israeli online legal database Nevo. Third- and fourth-year law students coded the cases under our supervision, after careful reading of all the documents in the case file. Data thus include extensive information encompassing the entire processing of each case, including case matter, pleadings, trial hearings, summations, and witnesses. A second tier of encoders randomly sampled 12 percent of the cases and checked them for accuracy and inner reliability. The additional raters found that the original coding was consistent with an accuracy of over 95 percent. Table 1 summarizes the data for the main variables analyzed in the next sections.

We focus on small claim cases because they offer several methodological advantages in our quasi-experimental design. Small claims are civil disputes limited to NIS 33,500.¹⁷ Only individual plaintiffs are allowed to file a small claim, and legal representation is not permitted.¹⁸ Jurisdiction rules limit the court district in which a plaintiff can file suit¹⁹ and litigants have no ability to impact the case assignment (to a judge, or to morning or evening shifts, in the treated courts). In the two treated districts (and two untreated districts), civil judges hold small claim hearings in periodic rotations, with cases randomly distributed according to the date of filing and the rotation period.²⁰

We highlight that there is no “prescreening” of small claim cases and that they are randomly assigned to all judicial personnel. These cases are considered to be “easy” and no special expertise is needed to resolve them. Thus, both judges and newly appointed registrars are included in the randomization of small claims, precluding concerns about a selection effect.²¹

Another important methodological advantage is the similarity between small claim cases. This mitigates concerns about unobservable factors related to case type, claim amounts, and litigation strategies. In addition, the nature of plaintiffs as “one-shooters” (Galanter 1974) makes it unlikely that they engage in forum shopping.

Finally, our design mitigates a further concern. Professional judges do not only hear small claims. The advent of senior registrars mechanically changes the composition of the portfolio of cases with which professional judges deal. In relative terms, they hear more cases not classified as small claims. It seems quite plausible that they devote more effort to these more important cases. This would also be normatively desirable, but when measuring the effect of additional judicial personnel on the handling of all cases, we could not disentangle the effect of a reduction in workload from an effect of a change in the composition of the portfolio of cases (e.g., since judges find more important cases

¹⁷Section 60 of the Israeli Court Law of 1984.

¹⁸Section 63 of the Israeli Court Law of 1984 determines that legal representation for small claim litigants can be allowed only in exceptional circumstances, and only with the court’s reasoned decision.

¹⁹Rule 2 of Small Claims Judgment Regulations (1976) determines that a plaintiff can file suit in the place of residence or place of business of the defendant, the place of contract creation or intended performance in contract disputes, the place of property delivery in property disputes, or the place of action/default for which claim is sought.

²⁰Note that other types of civil cases in the Israeli judiciary are not randomly assigned to the judicial staff, but distributed according to specialization. Thus, the impact of the decrease in equivalent number of assigned small claims to each judge in the treated courts might still differ between judges. For example, a judge burdened with higher caseloads might experience a more moderate decline in them compared to that of her fellows. Even so, we are less concerned with challenges of varied workload distribution, as the Israeli judicial management uses case weights to promote norms of equal distribution of workloads, especially among civil judges in magistrate’s courts (Weinshall-Margel et al. 2015).

²¹In our estimations, we consider only cases decided by professional judges, before and after the treatment. Had certain cases been set aside for the registrars, the pool of cases decided by professional judges would not be comparable in treated and untreated courts. Yet this is not the case. The only difference between treated and untreated courts is the caseload, not the type of cases.

also more interesting). This potential confound is removed by the fact that we exclusively consider how judges deal with (randomly assigned) small claims.²²

Each of the 2,085 data points specifies the dates of filing the case and of case disposition. It would be tempting to treat these as time series data, and to rely on regression discontinuity, or even difference in differences, for identification. Yet, unfortunately, neither approach is feasible.

For regression discontinuity, we would need data from as close as possible before and after treatment. Yet such data would not be meaningful. Although treatment was quasi random, it did not take place on one defined, uniform day, and it was not unexpected. The judicial staff in treated courts was informed of the expected appointments in early 2012 and could already anticipate their effect. The six senior registrars were gradually appointed in a time span that extends from March to September 2012. They were assigned new cases that had not yet been allocated to other judges. As judges' calendars are set around six months in advance, the reduction in caseload was actually felt by the judicial staff from late 2012 until early 2013. Furthermore, to test the full impact of treatment on process and outcomes, we need cases fully processed before or after the treatment—that is, those resolved by 2012 or filed after 2013. For these reasons, we decided to sample 2.4 percent of the small claims resolved in Israel's six jurisdictional districts in 2011 and 2014.²³ Filing dates in the whole sample range from 2008 to 2014, yet the vast majority of cases (close to 95 percent) were opened in 2010–2011 (pre-treatment) and 2013–2014 (post-treatment).

Unfortunately, nor can we employ difference-in-differences analysis. The main obstacle is a correlate of the concern with regression discontinuity. With diff-in-diff analysis, identification is through a difference in the time trend. Although trends of (later) treated and untreated cases must have been indistinguishable before treatment (they have a common trend), this trend continues for the untreated cases, while it changes for the treated case. Statistically, the critical effect is the interaction of treatment with time. Yet for this identification strategy to work, it must be possible to locate treatment in time. For the reasons spelled out for regression discontinuity, this is not possible. "Treatment" was a process that did not only take time; it also started and ended for different treated cases at different points in time.²⁴

²²Note, however, that we cannot exclude a positive spill-over effect: judges have more time for cases other than small claims; they use this time to handle these more important cases in ways they find more appropriate; the experience of doing a better job on high-profile cases trickles down to the way they deal with low-profile cases (small claims).

²³The first random draw of the cases resolved during 2014 yielded 978 small claims. Since 140 of the cases in the draw were resolved by newly appointed senior registrars, we sampled a supplementary sample of 140 cases decided by other judicial staff in the treated courts.

²⁴In the Appendix, we use simulation to explain a further obstacle. We have chosen sampling periods with the length of one year pre-, and one year post-intervention. Yet many (also small) cases stay considerably longer in the judiciary. This creates a mechanical time trend that is different from the true time trend. For the mere length of procedure, we could remove the effect with a control variable. However, we have no way to rule out similar sampling artifacts for our remaining dependent variables of interest, as for these other variables the potential effect is not mechanical. Sampling over a much longer period of time pre and post is also no solution. First, this would only attenuate, but not remove, sampling bias. The longer the sampling period, the bigger the concern that seeming treatment effects are actually driven by other unobserved changes.

For these reasons, we do not exploit the temporal structure of our (micro) data for identification. Our main analytic approach compares observations from the courts in the Central and Jerusalem districts on the one hand, and from the courts in the remaining districts, before and after the former courts have been “treated,” that is, have received additional personnel. We are chiefly interested in the interaction effect: Has easing the caseload led to normatively desirable outcomes? Our main reason to be confident in our findings is the quasi-random character of treatment.

To further increase confidence, we employ a series of safeguards. We run multivariate regressions, as all performance indicators originate from the same decision, and are therefore not independent. We use judge fixed effects, to remove any judge and court idiosyncrasies.²⁵ We cluster standard errors at the level of judges, as most judges have taken multiple decisions, and these decisions are dependent through the person of the decisionmaker. As Table 2 shows, cases from the unaffected and the affected districts are not perfectly balanced. Judges in the Central and Jerusalem districts are in particular considerably more senior.²⁶ Plaintiffs in the unaffected districts are somewhat more likely to file a motion. Disputes between clients and suppliers are a bit more frequent in the Central and Jerusalem districts. Plaintiffs in the unaffected districts are slightly more likely to sue a private party, and to sue multiple defendants. We react in two complementary ways. In one specification, we control for observed case characteristics.²⁷ In an alternative specification, we offer results that use coarsened exact matching on observed case characteristics.²⁸ These safeguards make us confident that we are not capturing a difference in the composition of cases across districts, or a change in the composition of the sample of cases in reaction to the advent of additional judicial personnel.

In our natural experiment, we observe the universe of the Israeli judiciary. Yet this universe is composed of only six districts, four untreated, and two treated. We have no reason to worry that we actually pick up idiosyncratic differences between court districts rather than the effect of the court receiving additional personnel because we only use data from judges who have not shifted from an untreated to a treated court, or vice versa. This is why judge fixed effects neutralize any court idiosyncrasies that might exist. Yet decisions from the same district might be more homogeneous than decisions across districts. Dependence at the level of districts might make standard errors unreliable that only take dependence at the level of judges into account.

²⁵As the syntax for multivariate regressions in Stata is not prepared for fixed effects estimation, we replace all variables with the difference from the mean of the variable for the judge in question.

²⁶This is mainly driven by the fact that in the Tel Aviv district, small claims are assigned mostly to newly appointed civil judges, who start their career with around three years of deciding small claims and then move on to other civil court divisions.

²⁷We do not control for court and district characteristics as they are collinear with the judge fixed effects.

²⁸For the underlying theory see Iacus et al. (2012). We use the implementation for Stata developed by Blackwell et al. (2009).

Table 2: Balance of Control Between Affected and Unaffected Court Districts

	<i>Unaffected</i>			<i>Affected</i>		
	<i>Haifa, North, South, Tel-Aviv</i>			<i>Central, Jerusalem</i>		
	Obs	Mean	SD	Obs	Mean	SD
Seniority						
Judge/registrar seniority	1,258	4.213	50.148	667	7.244	7.598
Seniority (0 = < 4 years, 1 = 4–10 years, 2 = > 10 years)	1,258	0.486	0.685	667	0.781	0.826
Claim						
Sum claimed (NIS)	1,257 ^a	10,309	8,958	662 ^b	10,822	9,289
Claim (0 = < 10,000 NIS, 1 = 10,000–20,000 NIS, 2 = > 20,000 NIS)	1,257	0.551	0.763	662	0.603	0.796
Plaintiff asks for restoration (1 = yes, 0 = only compensation)	1,258	0.089	0.281	667	0.091	0.288
Motions (0 = none, 1 = one, 2 = multiple)	1,252 ^c	0.919	0.828	667	0.733	0.817
More than one plaintiff (1 = yes)	1,258	0.154	0.361	667	0.145	0.353
Subject matter						
Client	1,258	0.425	0.495	667	0.492	0.500
Home	1,258	0.063	0.243	667	0.049	0.217
Motor	1,258	0.243	0.429	667	0.264	0.441
Other	1,258	0.269	0.443	667	0.195	0.396
Defendant Characteristics						
Defendant type (0 = private, 1 = commercial, 2 = government)	1,255 ^d	0.570	0.566	667	0.526	0.559
Defendant represents herself (1 = yes)	1,258	0.455	0.498	667	0.489	0.500
Defendant files a brief (1 = yes)	1,256 ^e	0.760	0.428	667	0.739	0.439
Multiple defendants (1 = yes)	1,258	0.314	0.464	667	0.255	0.436

^aMissing value for one case.

^bMissing values for five cases.

^cMissing values for six cases.

^dMissing values for three cases.

^eMissing values for two cases.

The natural safeguard is not available in our case. We cannot cluster standard errors at the highest potential level of dependence, that is, at the district level. The total number of districts (six) is way too small to meet the asymptotic requirements (Cameron et al. 2008). We react in two complementary ways. We report the test proposed by Ibragimov and Müller (2016), which tests the null hypothesis that there is no level of dependence above individual judges. This test is based on local regressions, confined to individual judges, of the dependent variable in question on a dummy variable that is 1 if the decision is taken at a point in time when the treated courts have been treated.²⁹ As this test turns out significant at conventional levels for some dependent variables, and as

²⁹Not all judges have been on the respective court before and after treated courts have been treated. This is why, for this specification test, we have a somewhat smaller sample. For comparison, we also offer “global” regressions based on this subsample.

the p values are not very high for other dependent variables, we also report results that use standard errors based on the wild bootstrap developed by Cameron et al. (2008).

V. RESULTS

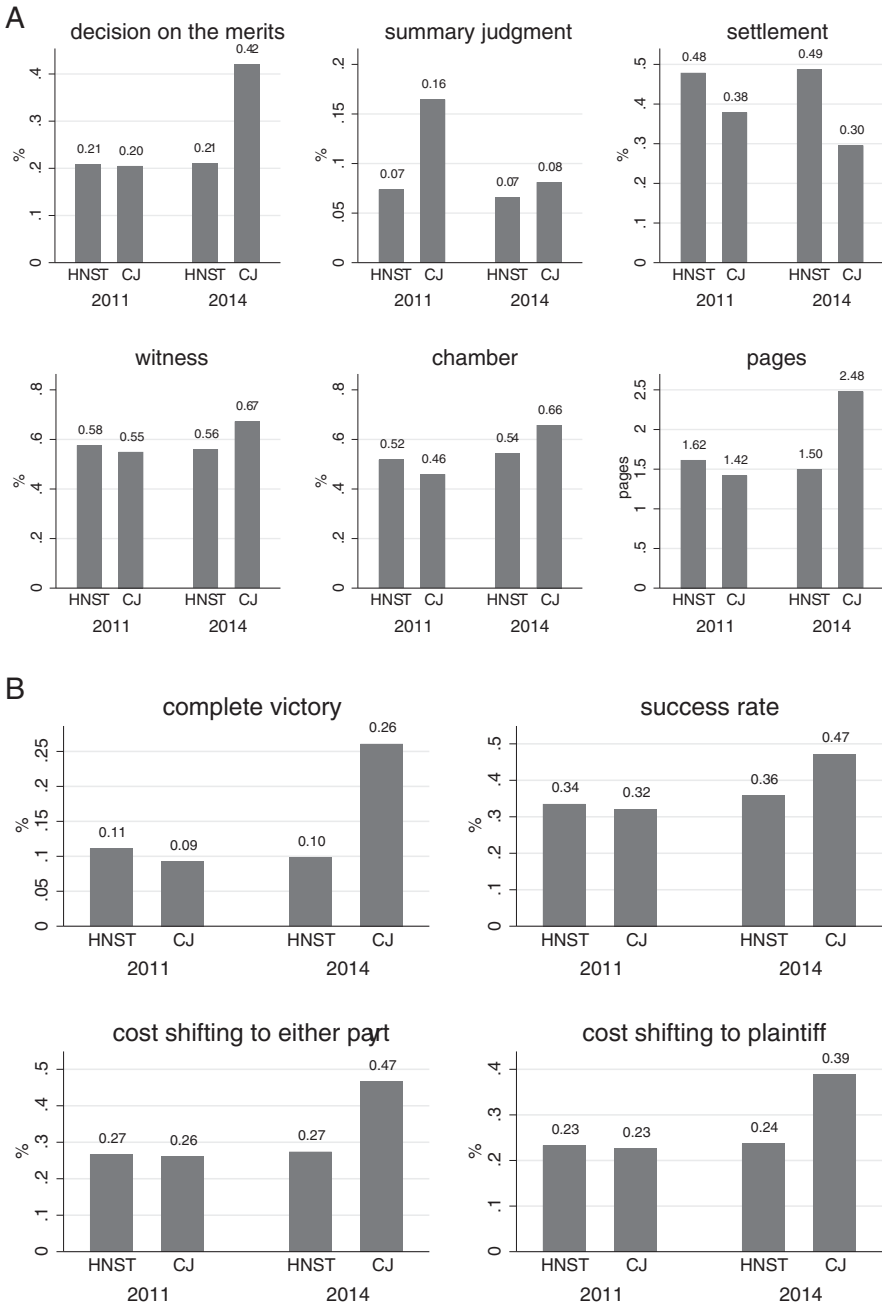
Descriptive statistics suggest a clear effect in support of the judicial approach (Figure 2).³⁰ In the treated districts, Central and Jerusalem, the fraction of cases that are *decided on the merits* jumps up from 20 to 42 percent, while it stays at 21 percent in the untreated districts of Haifa, North, South, and Tel-Aviv. In the treated districts, the fraction of cases that were concluded with *summary judgment* was high before the intervention (16 percent). After the increase in judicial manpower, this fraction goes down to 8 percent (which is approximately the same as in the untreated districts, where the fraction is 7 percent throughout). In the treated districts, the fraction of cases that end with *settlement* was lower in the first place (38 percent rather than 48 percent), but it goes down by another 8 percent, while it essentially stays constant in the untreated districts. In the Central and Jerusalem districts, the fraction of cases in which a *witness* is heard goes up from 55 to 67 percent, while it approximately stays constant in the untreated districts. Before the advent of new personnel, in the Central and Jerusalem districts small claims were predominantly concluded on the spot, with oral judgment. The fraction of cases in which the judge goes back to her *chambers* and drafts a written ruling increases from 46 to 66 percent, while it stays essentially constant in the untreated districts. Before treatment, judges in the Central and Jerusalem districts on average wrote a bit less text (1.42 *pages*) than their colleagues in the other districts (1.62 *pages*). The latter judges even reduce their writings a little over time (to 1.50 *pages*), while the average goes up to 2.48 *pages* with treatment.

The fact that judges work harder is to plaintiffs' benefit (lower panel of Figure 2). Before treatment, the chances for winning completely (*complete victory*) were at 9 percent in the Central and Jerusalem districts, and 11 percent elsewhere. This does not essentially change without treatment, but the probability of full success jumps up to 26 percent with treatment. An alternative way of considering the effect on plaintiffs is the *success rate*. It is calculated as the fraction between the sum claimed and the sum obtained. It was at 34 and 32 percent while courts had comparable personnel. Descriptively, there is a small increase to 36 percent in the untreated districts, but a substantial increase to 47 percent in the treated districts. The availability of additional personnel makes it also substantially more likely that the court shifts cost to either party.³¹ *Cost shifting* was at 27 and 26

³⁰We confine the analysis to indicators of judicial performance that turn out significant. The following indicators do not yield a significant treatment effect in any specification: default, dismissal, hearing, length, and full or partial success. These additional estimations are available from the authors upon request.

³¹The Israeli civil rules stipulate that at the end of every civil case judges may grant costs in favor of a litigant. Although the general expectation and guideline is that costs are normally to be shifted in favor of the prevailing litigant, judges have complete discretion in deciding whether to grant costs (see more in Eisenberg et al. 2014).

Figure 2: Descriptive statistics.



Notes: For definition of dependent variables see Table 1. Scale: % for whom dummy is 1; number of pages. HNST: unaffected districts (Haifa, North, South, Tel-Aviv). CJ: affected districts (Central, Jerusalem).

percent, respectively, before treatment. It stays perfectly constant without treatment, but goes up to 47 percent with treatment. Again this change is particularly beneficial for plaintiffs. While *cost shifting to plaintiffs* in all districts was at 23 percent before treatment, and essentially stays constant without treatment, it goes up to 39 percent in the Central and Jerusalem districts after the registrars take office.

With few exceptions, the effects are supported by statistical analysis.³² In Model 1 of Table 3, for each outcome we estimate:

$$y = \beta_0 + \beta_1 \text{ affected} + \beta_2 \text{ after} + \beta_3 \text{ affected*after},$$

where β_0 is the estimated outcome in a court district to which registrars have never been assigned (thus was not “affected” by the experiment) in 2011 (before the experiment). β_1 checks whether outcomes in the affected and the unaffected courts were already different in 2011. β_2 checks whether outcomes have changed between 2011 and 2014 in the districts that have not received additional judicial personnel. The critical coefficient is β_3 , the interaction between a district being affected by the intervention (i.e., the Central or the Jerusalem district) and the case being closed after these districts have received additional personnel (i.e., in 2014). If the interaction effect turns out significant, we learn that treatment has an effect. Deploying additional judicial personnel matters.

Models 2–6 of Table 3 add judge fixed effects. Our sample is confined to judges who have not moved from a treated to an untreated district, or vice versa. This is why judges are perfectly nested in districts. Consequently, in these models the effect of a district being affected is absorbed by the judge fixed effects.

As explained in Section IV, we use multiple safeguards to gauge the reliability of our findings. All models are multivariate, and therefore take into account that there is dependence resulting from the fact that one and the same judge, for one and the same case, has generated all dependent variables. The additional safeguards added in Models 2–6 are explained and motivated in Section IV. These safeguards sometimes matter. The following effects do not show up in all specifications. If we add controls, we no longer find that treated courts are more likely to hear a witness (Model 3). If we use the wild bootstrap with random draws at the level of districts, we no longer find an effect on the likelihood of settlement. Yet for this dependent variable, the Ibragimov Müller test is relatively far from significance ($p = 0.144$), which suggests that this additional safeguard is not necessary. This is different for cost shifting, and cost shifting for the plaintiff in particular. For both dependent variables, we no longer find an effect on the likelihood of settlement. Yet for this dependent variable, the Ibragimov Müller test is clearly

³²In Models 2–6, due to the judge fixed effects, the main effect of a court being affected (i.e., in the Central or Jerusalem district) drops out: the difference between the mean per judge and the concrete case is 0 by design. The interaction effect is nonetheless identified and interpretable, as the main effect of the court being affected is captured by the judge fixed effect.

Table 3: Statistics

	Model 1	Model 2 Judge FE	Model 3 Wild Bootstrap	Model 4 Controls	Model 5 Coarsened Exact Matching	Model 6 Only Judges Active Pre and Post
Merits						
Affected	-0.004 (0.044)	0.064 (0.051)	0.064 (0.124)	0.063 (0.048)	-0.007 (0.107)	0.064 (0.051)
After	0.002 (0.032)	0.332*** (0.068)	0.332*** (< 0.001)	0.280*** (0.064)	0.384** (0.120)	0.332*** (0.069)
Affected × After	NO	YES	YES	YES	YES	YES
Judge fixed effects	NO	NO	NO	NO	NO	NO
Controls	NO	NO	NO	NO	NO	NO
Matching	NO	NO	NO	NO	YES	NO
Cons	0.209*** (0.025)	-0.000 (0.000)	-0.000 (0.704)	-0.003** (0.001)	-0.013+ (0.007)	-0.001 (0.001)
N	1,925	1,925	1,925	1,908	1,126	722
						0.183
Ibragimov Müller test						
Summary Judgment						
Affected	0.091* (0.037)	0.016 (0.029)	0.016 (0.506)	0.020 (0.026)	0.016 (0.037)	0.016 (0.029)
After	-0.008 (0.021)	-0.110* (0.046)	-0.110+ (0.056)	-0.126** (0.047)	-0.143* (0.060)	-0.110* (0.047)
Affected × After	-0.076+ (0.041)	YES	YES	YES	YES	YES
Judge fixed effects	NO	NO	NO	NO	NO	NO
Controls	NO	NO	NO	NO	NO	NO
Matching	NO	NO	NO	NO	YES	NO
Cons	0.074*** (0.016)	-0.000 (0.000)	-0.000 (0.884)	0.000 (0.000)	0.008 (0.005)	0.000 (0.000)
N	1,925	1,925	1,925	1,908	1,126	722
						0.098
Ibragimov Müller test						
Settlement						
Affected	-0.100+ (0.056)	-0.018 (0.060)	-0.018 (0.700)	0.002 (0.064)	-0.018 (0.060)	-0.018 (0.060)
After	0.009 (0.042)	-0.162* (0.074)	-0.162 (0.132)	-0.145+ (0.076)	-0.162* (0.074)	-0.162* (0.074)
Affected × After	-0.092 (0.064)	YES	YES	YES	YES	YES
Judge fixed effects	NO	NO	NO	NO	NO	NO
Controls	NO	NO	NO	YES	NO	NO
Matching	NO	NO	NO	NO	NO	NO
Cons	0.479*** (0.036)	0.000 (0.000)	0.000 (0.996)	0.002+ (0.001)	0.000 (0.000)	0.000 (0.000)
N	1,925	1,925	1,925	1,908	1,126	722

Table 3: Continued

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
		Judge FE	Wild Bootstrap	Controls	Coarsened Exact Matching	Only Judges Active Pre and Post
Ibragimov Müller test						
Witness						0.144
Affected	-0.027 (0.044)					
After	-0.016 (0.030)	0.036 (0.050)	0.036 (0.496)	0.051 (0.069)	-0.064 (0.066)	0.036 (0.050)
Affected × After	0.139** (0.055)	0.158** (0.073)	0.158* (0.048)	0.091 (0.073)	0.238*** (0.079)	0.158* (0.073)
Judge fixed effects	NO	YES	YES	YES	YES	YES
Controls	NO	NO	NO	YES	NO	NO
Matching	NO	NO	NO	NO	YES	NO
Cons	0.576*** (0.023)	-0.000 (0.000)	-0.000 (0.834)	-0.000 (0.001)	0.008 (0.012)	-0.000 (0.000)
N	1,908	1,908	1,908	1,907	1,126	717
Ibragimov Müller test						0.065
Chamber						
Affected	-0.058 (0.040)					
After	0.025 (0.033)	-0.036 (0.049)	-0.036 (0.396)	-0.070 (0.054)	-0.048 (0.086)	-0.036 (0.049)
Affected × After	0.174** (0.053)	0.348*** (0.069)	0.348*** (< 0.001)	0.377*** (0.069)	0.352*** (0.103)	0.348*** (0.070)
Judge fixed effects	NO	YES	YES	YES	YES	YES
Controls	NO	NO	NO	YES	NO	NO
Matching	NO	NO	NO	NO	YES	NO
Cons	0.518*** (0.025)	-0.000 (0.000)	-0.000 (0.838)	0.001 (0.001)	0.000 (0.012)	-0.001 (0.001)
N	1,925	1,925	1,925	1,908	1,126	722
Ibragimov Müller test						0.174
Number of Pages						
Affected	-0.196* (0.092)					
After	-0.116 (0.081)	-0.077 (0.124)	-0.077 (0.528)	-0.114 (0.121)	-0.233 (0.237)	-0.077 (0.125)
Affected × After	0.175*** (0.188)	0.398*** (0.239)	0.398*** (< 0.001)	0.333*** (0.233)	0.287*** (0.307)	0.398*** (0.241)
Judge fixed effects	NO	YES	YES	YES	YES	YES
Controls	NO	NO	NO	YES	NO	NO
Matching	NO	NO	NO	NO	YES	NO
Cons	0.615*** (0.068)	-0.001 (0.001)	-0.001 (0.518)	-0.002 (0.003)	-0.036 (0.023)	-0.004 (0.004)

Table 3: *Continued*

	Model 1	Model 2 Judge FE	Model 3 Wild Bootstrap	Model 4 Controls	Model 5 Coarsened Exact Matching	Model 6 Only Judges Active Pre and Post
<i>N</i>	1,925	1,925	1,925	1,908	1,126	722 0.022
Ibragimov Müller test						
Complete Victory						
Affected	-0.018 (0.024)	-0.071 ⁺ (0.037)	-0.071** (0.002)	-0.101* (0.047)	-0.083 ⁺ (0.044)	-0.071 ⁺ (0.037)
After	-0.013 (0.020)	0.366*** (0.064)	0.366*** (< 0.001)	0.384*** (0.064)	0.420*** (0.072)	0.366*** (0.065)
Affected × After	0.181*** (0.050)	YES	YES	YES	YES	YES
Judge fixed effects	NO	NO	NO	NO	NO	NO
Controls	NO	NO	NO	NO	YES	NO
Matching	NO	NO	NO	NO	YES	NO
Cons	0.111*** (0.015)	-0.000 (0.000)	-0.000 (0.476)	-0.001 (0.001)	-0.007 (0.007)	-0.001 (0.001)
<i>N</i>	1,925	1,925	1,925	1,908	1,126	722 0.098
Ibragimov Müller test						
Success Rate						
Affected	-0.014 (0.029)	0.004 (0.042)	0.004 (0.904)	-0.012 (0.044)	-0.037 (0.057)	0.004 (0.043)
After	0.023 (0.024)	0.241*** (0.060)	0.241*** (< 0.001)	0.245*** (0.055)	0.310*** (0.071)	0.242*** (0.061)
Affected × after	0.128** (0.046)	YES	YES	YES	YES	YES
Judge fixed effects	NO	NO	NO	YES	NO	NO
Controls	NO	NO	NO	NO	NO	NO
Matching	NO	NO	NO	NO	YES	NO
Cons	0.335*** (0.017)	-0.002 (0.001)	-0.002 (0.502)	-0.001 (0.001)	-0.003 (0.008)	-0.004* (0.002)
<i>N</i>	1,669	1,669	1,669	1,658	984	631 0.376
Ibragimov Müller test						
Cost Shifting						
Affected	-0.005 (0.038)	0.109 ⁺ (0.058)	0.109 (0.262)	0.111 ⁺ (0.063)	0.039 (0.088)	0.109 ⁺ (0.058)
After	0.007 (0.032)	0.197*** (0.082)	0.197 (0.128)	0.186* (0.083)	0.258* (0.116)	0.197* (0.083)
Affected × After	0.197*** (0.054)	YES	YES	YES	YES	YES
Judge fixed effects	NO	NO	NO	YES	NO	NO
Controls	NO	NO	NO	YES	NO	NO
Matching	NO	NO	NO	NO	YES	NO

Table 3: Continued

	Model 1	Model 2 Judge FE	Model 3 Wild Bootstrap	Model 4 Controls	Model 5 Coarsened Exact Matching	Model 6 Only Judges Active Pre and Post
Cons	0.267*** (0.025)	-0.002 (0.001)	-0.002 (0.304)	-0.005** (0.002)	-0.009 (0.011)	-0.005 (0.002)
N	1,669	1,669	1,669	1,658	984	631
Ibragimov Müller test						0.040
Cost Shifting for Plaintiff						
Affected	-0.007 (0.038)					
After	0.004 (0.030)	0.074 (0.056)	0.074 (0.384)	0.062 (0.060)	0.025 (0.087)	0.073 (0.057)
Affected × After	0.159** (0.052)	0.176* (0.080)	0.176 (0.128)	0.173* (0.081)	0.231* (0.111)	0.177* (0.081)
Judge fixed effects	NO	YES	YES	YES	YES	YES
Controls	NO	NO	NO	YES	NO	NO
Matching	NO	NO	NO	NO	YES	NO
Cons	0.233*** (0.024)	-0.002* (0.001)	-0.002 (0.334)	-0.003* (0.001)	-0.009 (0.012)	-0.004* (0.002)
N	1,669	1,669	1,669	1,658	984	631
Ibragimov Müller test						0.030

Notes: All models except Model 3 are multivariate, and hence allow for correlation among the error terms of all dependent variables. All models except Model 3 cluster standard errors at the level of judges. Model 3 is based on Cameron Gelbach wild bootstrap standard errors with clusters at the level of court district. Models 2–6 use judge fixed effects. Model 4 further controls for all case characteristics that we observe, namely, judge seniority (three bins), sum claimed (three bins), plaintiff asks for restoration, number of motions (three bins), more than one plaintiff, subject matter (four bins - 0=other, 1=client, 2=home, 3=motor), defendant type (three bins), defendant represents herself, defendant files a brief, multiple defendants; for details on controls, and the definition of bins in particular, see Table 2. Model 5 uses coarsened exact matching, and matches by the exact same list of controls. Less data in Model 3 due to missing values on single covariates. Less data in Model 5 due to coarsened exact matching. N: lower number for witness, success rate, and cost shifting variables due to missing values on these dependent variables. Model 6 only uses data from judges who have been active in the court in question both in 2011 and 2014, that is, pre and post the intervention. Standard errors in parentheses. In Model 3 p values in brackets. Ibragimov Müller tests whether clustering only at level of judges misses relevant correlation. Rejection of test indicates that standard errors in Model 2 are progressive, so that p values from Model 3 should be used. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$, + $p < 0.1$.

significant, and the interaction effect is insignificant in the bootstrap model. These two effects are therefore to be treated with a grain of doubt.

All remaining effects are very robust. Interestingly, some estimated effects are even stronger in the more conservative statistical models. With additional personnel, courts are at least 21 percent more likely to decide on the merits (Model 1 of Table 3). They are at least 8 percent less likely to only give a summary judgment (Model 1). They are at least 17 percent more likely to produce a written ruling (Model 1). They write at least 1.175 more pages (Model 1). The plaintiff is at least 18 percent more likely to obtain a full victory (Model 1), and to recover at least an additional 13 percent of the sum she claimed (Model 1).

VI. DISCUSSION

The findings provide support for the judicial approach and motivated judges theories. In the treated courts, after increasing judicial manpower, judges are significantly more likely to decide cases on the merits, less likely to use swifter case resolution like summary judgment or settlement, more likely to hear witnesses, more likely to provide a written rather than merely an oral ruling, and they write more pages.

Is this good news? From a managerial perspective, investing more resources in resolving any case, and particularly small claims, might seem inefficient. One can argue that as long as the same legal outcome is reached, writing short summary judgments, deciding cases on the spot, or pushing for settlements is more effective than hearing witnesses and writing lengthier decisions on the merits.³³ However, we show that an increase in judicial resources is reflected in substantive decisions. Hence, the judicial time devoted to resolving a case is associated with a difference in legal outcomes. In our case, plaintiffs strongly benefit from a reduction in caseload: They are more likely to win and to recover a larger portion of the claim, and courts are more likely to shift costs. Ultimately, it is for policymakers to decide whether these effects are desirable.

What are the possible causes for favoring plaintiffs under reduced caseloads? The first possible explanation is related to the cost of appeals in condensed courts. The likelihood of an appeal being filed and accepted might be higher for short and less reasoned judgments, and its cost will be higher if the appeal court decides to remand the case and return it to an already overburdened judge. Thus, for example, de Mot et al. (2015) showed that judges in courts with higher caseloads adopt legal doctrines that they believe will reduce appeal rates. It might be that judges believe that an intermediate legal outcome would yield fewer appeals from either side (as both partly win), while a full victory (or close to it) is more likely to be appealed by the losing side. This mechanism specifically explains the significant increase in full victories after treatment, yet it does not

³³Note that cost-benefit analysis will also consider aspects related to court legitimacy and procedural justice (Lind & Tyler 1988).

provide a complete explanation for the one-sided effect. Under the same reasoning, we could expect an increase in full plaintiff losses as well.

It might be argued that a pro-plaintiff decision requires more work than a pro-defendant decision. As the plaintiff in small claims carries the burden of proof and is required to support her claim by a preponderance of evidence, establishing her case might require more effort. Thus, the allocation of more time and judicial resources to hearing each case can generally work to the advantage of plaintiffs.

Likewise, the status quo bias, interacting with the effects of caseloads, can provide a possible answer: Faced with uncertainty, jurists prefer to avoid a decision and leave things as they are. In the legal civil arena, this means dismissing the claim of the plaintiff, who is perceived as trying to benefit from changing the status quo. Jurists and laypeople alike tend to view the acceptance of a civil claim as an active decision and its dismissal as a passive one (Rachlinski 1996; Zamir & Ritov 2012). Overworked judges might thus be more inclined toward the default solution of denying a claim rather than changing the status quo. This explanation is in line with studies finding that time pressure reinforces a tendency to give greater deference to primary decisionmakers, thus decreasing the likelihood of winning administrative cases and civil or criminal appeals in overworked courts (Giles et al. 2015; Guthrie & George 2004; Huang 2011).

Finally, the plaintiff's role as the initiator of a claim might drive the interaction between caseload and winning party. For example, overworked judges might be less inclined to accept plaintiffs' claims as they anticipate that high winning rates might encourage other plaintiffs to file suits, which would generate even more work. Note that this explanation does not strictly require judges to make conscious strategic decisions favoring the defendants in congested courts. Judges suffering from high caseloads might unconsciously develop an aversion to plaintiffs as those who are responsible for increasing their caseloads. Alternatively, overworked judges might suffer from higher levels of stress, bad mood, and tiredness, which are associated with a tendency to rule harshly and an inclination to deny submitted requests (Cho et al. 2016; Eren & Mocan 2018).

While further studies are needed to test the mechanisms behind the observed pro-plaintiff effect of the reduction in caseloads, and notwithstanding that the "correct" legal outcome remains unknown, the fact that increasing judicial staff affects substantive decisions suggests that swifter or more efficient case resolution can come at the expense of its legal quality.

VII. CONCLUSION

The Israeli judiciary was entertaining the idea of providing the public better service by offering late-hour adjudication. In the spirit of experimental legislation, the judiciary was trying out the idea in two of the six court districts, the Central district and Jerusalem. These districts were selected for pragmatic reasons. Courthouses should be located conveniently. There should be sufficient parking lots in the proximity. The laws' subjects

should have a chance to combine going to court with shopping or going to a restaurant. In the interest of making the test credible, the courts in the two treated districts have received additional personnel. As the intended reform was focused on the disposition of low-stakes civil cases, the additional personnel consisted of registrars. For its stated purposes, the test has been only mildly successful. Late hours have not been rolled out across the country. Even the additional personnel gradually shifted toward ordinary business hours. Yet the treated courts have kept the additional personnel.

We exploit the intervention as a quasi-experiment for the reaction of courts to the availability of additional judicial resources. Although the extreme versions of the “managerial” and “maximizing utility judges” approaches to the judiciary would not expect to see a significant effect on judicial performance and services, the “judicial” and “motivated judges” approaches predict that if capacities are freed up, judges will decide faster or invest more time and resources in their remaining cases. We have a sample of approximately 2,000 cases, half of them from 2011 when there was no sign of the intervention yet, and half of them from 2014 when the affected courts had absorbed the intervention. For each case we code a series of performance indicators. We find that treated courts are significantly more likely to decide on the merits, they are less likely to confine themselves to summary judgment or settlement, they are more likely to hear witnesses, to provide a written rather than merely an oral ruling, and they write more pages. These data provide strong support for the judicial approach and motivated judges theories. Investing more resources in resolving each case seems to work to the benefit of plaintiffs. In the treated courts they are more likely to win cases, to recover a larger fraction of their claim, and to be reimbursed for litigation costs. We discuss several mechanisms to explain the pro-plaintiff effect.

The main strength of our study is its quasi-experimental character. This provides us with fairly clean causal evidence about the effect of caseload on judicial performance. Yet every empirical study has limitations. The main limitation of our study is the identification strategy. With judge fixed effects, we can take out of the equation idiosyncrasies that characterize individual judges, or the courts in which they work. With control variables, we can condition results on all observed case characteristics. With coarsened exact matching, we can construct samples of treated and untreated courts that are balanced on all observables. Yet we cannot categorically exclude that there are further unobserved causes for the observed effects. We acknowledge this limitation, but are less worried since the reason for assigning additional judicial personnel to the Central and Jerusalem districts was completely unrelated to our research question. The court administration wanted to explore judicial services provided in the evening, not an improvement in judicial services resulting from the deployment of additional personnel. This is what has created the quasi-natural experiment that we exploit for our analysis.

Can results of this quasi-natural experiment be generalized to other cases, judges, and courts? As their name suggests, small claims tend to be easy cases that do not require a lot of judicial time and attention. The caseload impact on small claims is likely to be less pronounced than on other, more complicated, civil cases. Had we not found an effect on judicial performance regarding small claims, this would not necessarily have

meant that the intervention was pointless. It could have been that judges invest the extra time in more important cases. Yet as we show, there is an effect on small claims. It could only result from a displacement effect between case categories if, with the advent of the registrars, professional judges shift resources away from more important cases to small claims. Although we cannot exclude this with our data, we deem it highly implausible.

Will the judicial approach always prevail? Not necessarily. The relationship between the resources available for decision making and judicial services may not be linear (Lavie 2016). At some point, the marginal gain in judicial performance will likely be small. It does not seem implausible that, in that range of parameters, judges might start increasing leisure, as predicted by the managerial approach. It could therefore be that the intervention in Israel hit some sort of a sweet spot, where judges were so obviously longing for more resources that finally doing a decent job was their predominant motive. Findings thus call for further comparative investigations on caseload impact on the process and outcomes of judicial decision making.

ACKNOWLEDGEMENT

Open access funding enabled and organized by ProjektDEAL.

REFERENCES

- Adler, A. L. (2014) "Extended Vacancies, Crushing Caseloads, and Emergency Panels in the Federal Courts of Appeals," 15 *J. of Appellate Practice & Process* 163.
- Agrast, M. D., J. C. Botero, & A. Ponce (2011) *The Rule of Law Index 2011*. Washington, DC: World Justice Project.
- Aviv, G., & I. Galon (2015) *Evaluation of the Pilot Program for Evening Court Hearings in Israeli Magistrate Courts: Final Report*. Israeli Courts Research Division. Available at <https://elyon1.court.gov.il/heb/Research%20Division/doc/05072015.pdf>.
- Baker, T. E. (2006) "Applied Freakonomics: Explaining the 'Crisis of Volume'," 8 *J. of Appellate Practice & Process* 101.
- Bar-Niv Burnovski, M., Z. Lieber, & B. Ronen (2010) "Focused Management in a Court System: Doing More with the Existing Resources," 29(4) *Human Systems Management* 265.
- Beenstock, M., & Y. Haitovsky (2004) "Does the Appointment of Judges Increase the Output of the Judiciary?" 24(3) *International Rev. of Law & Economics* 351.
- Best, J., & L. Tiede (2015) *Vacancy in Justice: Analyzing the Impact of Overburdened Judges on Sentencing Decisions*. Available at SSRN 2417348.
- Blackwell, M., S. Iacus, G. King, & G. Porro (2009) "cem: Coarsened Exact Matching in Stata," 9(4) *Stata J* 524.
- Cameron, A. C., J. B. Gelbach, & D. L. Miller (2008) "Bootstrap-Based Improvements for Inference with Clustered Errors," 90(3) *Rev. of Economics & Statistics* 414.
- Carrington, P. D. (1969) "Crowded Dockets and the Courts of Appeals: The Threat to the Function of Review and the National Law," 82 *Harvard Law Rev.* 542.
- Castro, M. F., & C. Guccio (2015) "Bottlenecks or Inefficiency? An Assessment of First Instance Italian Courts' Performance," 11(2) *Rev. of Law & Economics* 317.
- Cho, K., C. M. Barnes, & C. L. Guanara (2016) "Sleepy Punishers Are Harsh Punishers: Daylight Saving Time and Legal Sentences," 28(2) *Psychological Science* 242.

- Coolsen, J. P. (2009) "Case Management Innovation in a Large, Urban Trial Court: The Critical Importance of Legal Stakeholder Attitudes," 30(1) *Justice System J.* 70.
- Cooter, R. D. (1983) "The Objectives of Private and Public Judges, 41 *Public Choice* 107.
- Dalton, T., & J. M. Singer (2014) "Bigger Isn't Always Better: An Analysis of Court Efficiency using Hierarchical Linear Modeling," 34(3) *Pace Law Rev* 1169.
- de Mot, J., M. Faure, & J. Klick (2015) "Appellate Caseload and the Switch to Comparative Negligence," 42 *International Rev. of Law & Economics* 147.
- Dimitrova-Grajzl, V., P. Grajzl, & K. Zajc (2014) "Understanding Modes of Civil Case Disposition: Evidence from Slovenian Courts," 42(4) *J. of Comparative Economics* 924.
- Eisenberg, T., T. Fisher, & I. Rosen-Zvi (2014) "Attorneys' Fees in a Loser-Pays System," 162 *Univ. of Pennsylvania Law Rev.* 1619.
- El Bialy, N (2016) The 2007 Judicial Reform and Court Performance in Egypt, 12(1) *Review of Law & Economics* 95–117.
- Engel, C., & L. Zhurakhovska (2017) "You Are in Charge: Experimentally Testing the Motivating Power of Holding a Judicial Office," 46(1) *J. of Legal Studies* 1.
- Epstein, L., W. M. Landes, & R. A. Posner (2011) "Why (and When) Judges Dissent: A Theoretical and Empirical Analysis," 3(1) *J. of Legal Analysis* 101.
- (2013) *The Behavior of Federal Judges*. Cambridge, MA: Harvard Univ. Press.
- Eren, O., & N. Mocan (2018) "Emotional Judges and Unlucky Juveniles," 10(3) *American Economic J.: Applied Economics* 171.
- Galanter, M. (1974) "Why the Haves Come Out Ahead: Speculations on the Limits of Legal Change," 9 *Law & Society Rev* 95.
- (2004) "The Vanishing Trial: An Examination of Trials and Related Matters in Federal and State Courts," 1(3) *J. of Empirical Legal Studies* 459.
- Giles, M. W., J. Strayhorn, & T. Peppers (2015) "Policymaking Under Constraint: Decision Modes in the U.S. Courts of Appeals," 36(2) *Justice System J.* 95.
- Gomes, A. O., T. Guimarães, & L. Akutsu (2016) "The Relationship Between Judicial Staff and Court Performance: Evidence from Brazilian State Courts," 8(1) *International J. for Court Administration* 12–19.
- Guthrie, C., & T. E. George (2004) "The Futility of Appeal: Disciplinary Insights into the Affirmance Effect on the United States Courts of Appeals," 32 *Florida State Univ. Law Rev.* 357.
- Guthrie, C., J. J. Rachlinski, & A. J. Wistrich (2000) "Inside the Judicial Mind," 86 *Cornell L. Rev.* 777.
- (2007) "Blinking on the Bench: How Judges Decide Cases," 93(1) *Cornell Law Rev* 1.
- Habel, P., & K. Scott (2014) "New Measures of Judges' Caseload for the Federal District Courts, 1964–2012," 2(1) *J. of Law & Courts* 153.
- Heaton, P., & E. Helland (2011) "Judicial Expenditures and Litigation Access: Evidence from Auto Injuries," 40(2) *J. of Legal Studies* 295.
- Helland, E., & J. Klick (2007) "The Effect of Judicial Expedience on Attorney Fees in Class Actions," 36(1) *J. of Legal Studies* 171.
- Huang, B. I. (2011) "Lightened Scrutiny," 124(5) *Harvard Law Rev.* 1109.
- Iacus, S., G. King, & G. Porro (2012) "Causal Inference Without Balance Checking: Coarsened Exact Matching," 20(1) *Political Analysis* 1.
- Ibragimov, R., & U. K. Müller (2016) "Inference with Few Heterogeneous Clusters," 98(1) *Rev. of Economics & Statistics* 83.
- Iverson, B. (2017) "Get in Line: Chapter 11 Restructuring in Crowded Bankruptcy Courts, 64(11) *Management Science* 5370–5394.
- Johnson, B. D. (2006) "The Multilevel Context of Criminal Sentencing: Integrating Judge and County Level Influences," 44(2) *Criminology* 259.
- Jonski, K., & D. Mankowski (2014) "Is Sky the Limit? Revisiting 'Exogenous Productivity of Judges' Argument," 6(2) *International J. for Court Administration* 53.
- Langbroeck, P. M., & M. Kleimann (2016) *Backlog Reduction Programmes and Weighted Caseload Methods for South East Europe: Two Comparative Inquiries*. Sarajevo, Bosnia and Herzegovina: Regional Cooperation Council.

- Lavie, S. (2016) "Appellate Courts and Caseload Pressure," 27 *Stanford Law & Policy Rev.* 57.
- Levy, M. K. (2013) "Judging the Flood of Litigation," 80 *Univ. of Chicago Law Rev.* 1007.
- Lienhard, A., & D. Kettiger (2011) "Research on the Caseload Management of Courts: Methodological Questions," 7(1) *Utrecht Law Rev.* 66.
- Lind, E. A., & T. R. Tyler (1988) *The Social Psychology of Procedural Justice*. New York, NY: Springer Science & Business Media.
- Maayan, E., B. Ronen, & A. Coman (2012) "Assessing the Performance of a Court System: A Performance Measures Approach," 35(11) *International J. of Public Administration* 729.
- Meador, D. J. (1974) *Appellate Courts: Staff and Process in the Crisis of Volume: An Appellate Justice Project of the National Center for State Courts*. St. Paul, MN: West Publishing Company.
- Miner, R. J. (2012–2013) "Dealing with the Appellate Caseload Crisis: The Report of the Federal Courts Study Committee Revisited," 57(3) *New York Law School Law Rev.* 517.
- Mitsopoulos, M., & T. Pelagidis (2010) "Greek Appeals Courts' Quality Analysis and Performance," 30(1) *European J. of Law & Economics* 17.
- Moffett, K. W., F. Maltzman, K. Miranda, & C. R. Shipan (2016) "Strategic Behavior and Variation in the Supreme Court's Caseload Over Time," 37(1) *Justice System J.* 20.
- Moore, P. W. H. (2015) "The Civil Caseload of the Federal District Courts," 2015(3) *Univ. of Illinois Law Rev.* 1177.
- Nagel, S., M. Neef, & N. Munshaw (1978) "Bringing Management Science to the Courts," 62 *Judicature* 128.
- Narayan, P. K., & R. Smyth (2007) "What Explains Dissent on the High Court of Australia? An Empirical Assessment Using a Cointegration and Error Correction Approach," 4(2) *J. of Empirical Legal Studies* 401.
- National Center for State Courts (2005) *CourTools Performance Measures* [measurement instrument]. Available at CourTools Website <http://www.courttools.org/>.
- Norris, S. (2018) "Judicial Errors: Evidence from Refugee Appeals, *Univ. of Chicago, Becker Friedman Institute for Economics Working Paper* 2018-75.
- Ponticelli, J., & L. S. Alencar (2016) "Court Enforcement, Bank Loans, and Firm Investment: Evidence from a Bankruptcy Reform in Brazil," 131 *Q. J. of Economics* 1365.
- Posner, R. A. (1985) *The Federal Courts: Crisis and Reform*. Cambridge, MA: Harvard Univ. Press.
- (1993) "What Do Judges Maximize? (The Same Thing Everybody Else Does)," 3 *Supreme Court Economic Rev.* 1.
- (2006) "Demand and Supply Trends in Federal and State Courts Over the Last Half Century," 8 *J. of Appellate Practice & Process* 133.
- Rachlinski, J. J. (1996) "Gains, Losses, and the Psychology of Litigation," 70 *Southern California Law Rev.* 113.
- Rachlinski, J. J., S. L. Johnson, A. J. Wistrich, & C. Guthrie (2008) "Does Unconscious Racial Bias Affect Trial Judges?" 84 *Notre Dame Law Rev.* 1195.
- Resnik, J. (1982) "Managerial Judges," 96 *Harvard Law Rev.* 376.
- Richman, W. M., & W. L. Reynolds (2012) *Injustice on Appeal: The United States Courts of Appeals in Crisis*. New York, NY: Oxford Univ. Press.
- Rosales-López, V. (2008) "Economics of Court Performance: An Empirical Analysis," 25(3) *European J. of Law & Economics* 231.
- Sisk, G. C., M. Heise, & A. P. Morriss (1998) "Charting the Influences on the Judicial Mind: An Empirical Study of Judicial Reasoning," 73 *NYU Law Rev.* 1377.
- Stras, D. R., & S. M. Pettigrew (2010) "Rising Caseload in the Fourth Circuit: A Statistical and Institutional Analysis," 61 *South Carolina Law Rev.* 421.
- Ulmer, J. T., & M. Bradley (2006) "Variation in Trial Penalties Among Serious Violent Offenses," 44 (3) *Criminology* 631.
- Ulmer, J. T., J. Eisenstein, & B. D. Johnson (2009) "Trial Penalties in Federal Sentencing: Extra-Guidelines Factors and District Variation," 27 *Justice Q.* 560.

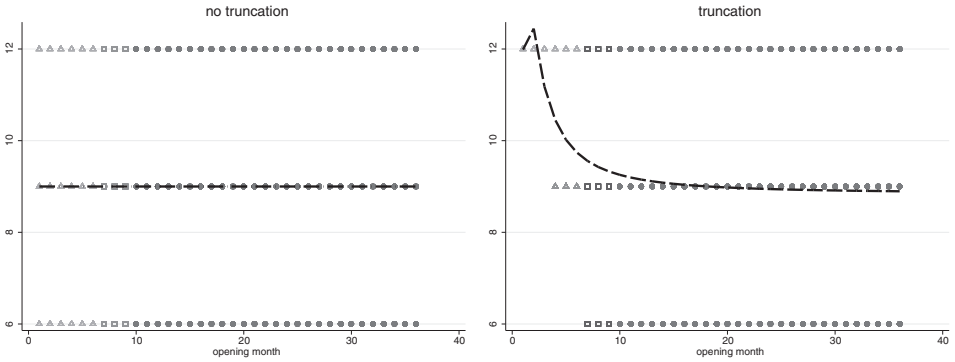
- Ulmer, J. T., & B. D. Johnson (2004) "Sentencing in Context: A Multilevel Analysis," 42(1) *Criminology* 137.
- Webber, D. (2006) *Good Budgeting, Better Justice: Modern Budget Practices for the Judicial Sector* (L&D Working Paper No. 3). Washington, DC: World Bank.
- Weber, M. ([1922](1978) *Economy and Society: An Outline of Interpretive Sociology*. Berkeley, CA: Univ. of California Press.
- Weinshall-Margel, K., & I. Galon (2013) *Interim Report for the Evaluation of the Pilot Program for Evening Court Hearings in Israeli Magistrate Courts*. Israeli Courts Research Division. Available at <https://elyon1.court.gov.il/heb/Research%20Division/doc/30072015.pdf>.
- Weinshall-Margel, K., I. Galon, & I. Taraboulos (2015) "Creating a Case Weight Index for Measuring Judicial Workload," 44 *Hebrew Univ. Law Rev.* 769.
- Yang, C. S. (2016) "Resource Constraints and the Criminal Justice System: Evidence from Judicial Vacancies," 8(4) *American Economic J.: Economic Policy* 289.
- Yeung, L. L., & P. F. Azevedo (2010) "Measuring Efficiency of Brazilian Courts with Data Envelopment Analysis (DEA)," 22(4) *IMA J. of Management Mathematics* 343.
- Zamir, E., & I. Ritov (2012) "Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation," 41(1) *J. of Legal Studies* 165.

APPENDIX A.: MECHANICAL EFFECT OF TRUNCATION

In our dataset, we have sampled the (approximately) same number of cases closed in 2011 and in 2014. Now these cases take differently long to close. If we were to try and estimate time trends during these two years of observation, they would suffer from a mechanical effect. Figure A1 demonstrates the mechanical effect with simulated data.³⁴ In the simulation, cases open on one of 36 months, with equal probability. Each month, a case opens that takes six, another that takes nine, and a third that takes 12 months to close. If we consider all cases, the estimated mean duration is, of course, nine months (the dashed line in the left panel coincides with the middle series of dots). Yet if we only consider cases that close after 12 months, we overestimate the time it takes for cases to close during the first months (dashed line in the right panel). This effect is mechanically driven by the fact that cases with a projected duration of six or nine months *that have started early* have already been closed before the beginning of the sampling period.

³⁴The code for the simulation is available from the authors upon request.

Figure A1: Simulation showing the truncation effect.



Notes: Left panel—true population effect: at each of 36 months, one new case with expected duration six months (triangles), one with expected duration nine months (squares), and one with expected duration 12 months (circles) is filed. All cases terminate as expected. Dashed line: average duration is nine months, irrespective of month in which case has been filed. Right panel—sampling by date when cases close. Cases that had been expected to close within six months are no longer in the data (lower line of dots). Only half the cases that had been expected to close within nine months are still in the data (middle line of dots). All cases that had been expected to close within 12 months are still in the data (upper line of dots). Dashed line: average duration conditional on month in which case has been filed.