This version is under R&R at *Production and Operations Management (POM)*

The Congestion Vortex: Empirical Evidence of the Adverse Effects of Delay on Workload in Court Systems

Abstract

This paper presents a methodological approach that employs empirical analysis to investigate the adverse operational effects of congestion within a court system setting. Central to this study is the premise that prolonged lead time generates additional tasks, consequently increasing congestion by consuming additional processing resources—a phenomenon we term 'the congestion vortex'. To explore this subject, we merge two distinct datasets from the Israeli court system which together provide a comprehensive map of case-related judicial tasks in all civil cases in the system for more than seven years, and the time estimated to complete each of these tasks. We also utilize two instrumental variables, exogenous shocks to the system, to mitigate potential reverse causality between judges' effort and lead time. Our results indicate that the longer a case lingers in the system, the greater the judicial workload it creates, as measured by the hours invested in producing judicial decisions. Furthermore, we demonstrate that extended lead time increases the likelihood of a case concluding with a judgment on its merits, further escalating judicial effort.

Keywords: Court Operations, Congestion, Data Driven, Instrumental Variables, Not-for-profit, Service Operations

1. Introduction

When an unfinished product in an assembly line or a patient being treated in a healthcare center must wait between stages of the process, one can easily observe the costs associated with such waiting times. However, in inherently long service-oriented contexts such as court cases, where the act of waiting is less apparent, it may seem that there are no immediate costs incurred to the system. Furthermore, a common misconception is that waiting time does not negatively impact the system, as cases in the queue are considered 'dormant' and may even drop out of the queue. In this study, we challenge this perception and posit that prolonged waiting times do have an adverse effect on the system, by altering the work content of cases in the queue.

Courts serve as the pillars of the justice system, acting as the machine that produces justice, which means that their operational performance plays a crucial role in ensuring social welfare and upholding access to justice, the latter being a fundamental civil right. Furthermore, from an economic standpoint, delays in the court system not only diminish the effectiveness of judicial decisions (Listokin, 2002), but also have negative repercussions in terms of the cost of conducting business, entrepreneurship, investments, and financial prosperity (Chemin, 2009; Chemin, 2010; Decarolis et al., 2023). Despite the paramount importance of the performance of the judicial process, court systems worldwide battle with significant congestion. This congestion arises for numerous reasons (resource constraints, to name just one) and results in prolonged waiting times, diminished service quality, and public dissatisfaction. Moreover, a continual rise in demand for judicial intervention, accompanied by escalating costs associated with court systems (CEPEJ, 2016; Decker et al., 2011), further exacerbates the problem of congestion.

This paper develops the hypothesis that delays in the court system create additional workload, thereby increasing the effective demand for judges' time, which is the constrained resource in the court system. Specifically, the longer a case remains in the system, the higher the likelihood for this case to undergo changes in its content, which, in turn, would generate additional work for the already overburdened judges. Our hypothesis suggests the existence of a "congestion vortex"; in this scenario, congestion triggers additional tasks, leading to higher congestion levels due to the increased workload, thus perpetuating a cycle.

To investigate this hypothesis, our analysis centers on the *trial phase* of the civil process, where most of the elapsed time consists of waiting time rather than processing time, as explained further in Section 3. Additionally, in the trial phase, as per civil procedure, the work content of the case has already been established during an earlier stage of the process. Consequently, waiting time during this phase does not contribute to the advancement of the process and is therefore considered a delay. According to our hypothesis, this delay has the potential to induce changes in the work content or to introduce additional tasks.

Our study utilizes two distinct datasets to explore the intricacies of civil cases within the Israeli court system. The first dataset, which was developed in partnership with Court Worm[1] (a nonprofit organization), aggregates publicly available data from the Israeli court information system. This dataset encompasses all civil cases from both magistrate and district courts, for the period from January 2000 to February 2022. The comprehensive dataset provides a rich array of case-level information, including case attributes, meetings, and decisions throughout the lifecycle of each case. The second dataset was provided by the Israeli Courts Research Division (ICRD) under the Freedom of Information Act,[2] and is centered on estimations of the effort (in work hours) required to carry out judicial tasks in the lifecycle of different case types. This dataset was created based on real-time activity logs meticulously recorded by judges and court stenographers and later validated statistically. It serves as the source of our dependent variable to proxy judicial workload – the level of judicial effort required for producing decisions in motions submitted by the parties during the trial phase, measured in work hours. By combining the above two datasets, we conduct an extensive analysis of the 48,997 cases resolved in trial between September 1st, 2012, and December 31st, 2019 for which we have verified data and can derive effort estimations.

To tackle potential reverse causality bias, we adopt the instrumental variable (IV) approach. We identify two exogenous shocks during the research period—a military conflict and a court workers' strike—that exogenously extended lead time without directly influencing judges' workload. After establishing the relevance and validity of the IVs, we propose a three-stage model: firstly, we estimate the effect of the IVs on the trial-phase lead time; second, we estimate the effect of lead time on the probability of decisions in motions during the trial phase; finally, we estimate the effect of lead time on the level of effort required for these decisions, given that there are decisions in motions.

The estimation results reveal a significant positive effect of trial-phase lead time on the level of effort exerted by judges when making decisions in motion. Specifically, the analysis suggests that a longer lead time not only increases the likelihood of changes in work content, which is proxied by a positive number of decisions in motions in the trial phase but also increases the accumulated effort required from the judge to handle the consequences of these changes. An additional analysis illustrates that extended lead time in the trial phase also increases the likelihood of a case concluding with a judgment on the merits. This task requires considerable judicial effort (rather than terminating through a settlement or other means).

[1] https://www.odata.org.il/organization/about/tolaat
[2] This dataset was established to support the study: "The Judicial Workload in the Israeli Courts System" (Aviv and Erental, 2021)

These findings further highlight the importance of considering the consequences of delay, as once a case enters the trial phase, delay impacts not only the judicial workload but also the case outcome.

The findings of this study carry significant implications for comprehending the dynamics of the judicial system and offer valuable insights for enhancing its efficiency and efficacy. Furthermore, the study contributes to the literature on congestion within service systems by using empirical data to identify and quantify adverse effects of high congestion. In addition to the relevance of this study to research in operations management (OM), it has implications for policy decisions in the judicial sphere, by showing that prolonged waiting times could substantially diminish judicial productivity — a concern of global importance (CEPEJ, 2015; CEPEJ, 2016; Church et al., 1978; Dakolias, 1999; Voigt, 2016).

2. Related literature and hypotheses development

Our study is motivated primarily by two streams of research, and these serve as the foundations for the two main hypotheses developed in this study. Below, we outline these streams and explain how they contributed to shaping our hypotheses.

2.1. Effects of congestion in service systems

Research in the field of congested service systems has predominantly focused on the impact of workload on service durations, as reviewed by Delasay et al. (2019) under a framework known as "load effect on service times" (LEST). In early research, empirical studies unveiled the interdependence between service times and the state of the system, notably impacted by the workload (Dshalalow, 1997). With advancements in data collection techniques that provide empirical evidence, theoretical studies on queuing have integrated additional assumptions that capture the complex relationships between workload and service time (Azriel et al., 2019; Delasay et al., 2016; Gans et al., 2010; Kingman, 2009).

The trajectory of research in this realm has involved modeling the effects of workload through diverse mechanisms, elucidating distinct outcomes of workload on service time, and consequently, length of stay (LOS). An early example is the field study of Edie (1954), who investigated the impact of traffic load on toll-booth service times, and found that congestion can result in shorter average service times due to an extended preparation time available to the customers. In another setting, Hillier et al. (2009) identified significant adverse effects of workload on throughput and quality within hospital environments, ultimately leading to longer service times. Kuntz et al. (2011) investigated the negative impact of high workload on throughput in pediatric emergency departments; they reported a direct positive correlation between occupancy level and LOS. However, other studies observed a decrease in LOS with occupancy level, which occurs when servers adopt strategic behaviors to mitigate the high workload, such as early discharge and speedup (Chan et al., 2012; KC and Terwiesch, 2012). In a subsequent study, KC (2014) suggested a more refined model characterized by a U-shaped curve. According to this model, an increase in workload initially results in a reduced service time (due to the strategic behavior of servers), but as the workload continues to increase, the reduction in service time starts to level off, until eventually, the trend is reversed and a higher workload results in a longer service time. In a study of hospital emergency departments (EDs), Batt and Terwiesch (2017) also uncovered the effect of aggregated adaptive mechanisms, such as early task initiation, which can alleviate the adverse effects of high workload. They suggested an inverted-U-shaped function to explain the observed relationship between workload and service time; i.e., initially, as the workload began to increase, they observed a reduction in service time, but as the workload increased further, the servers appeared to adopt various mitigation strategies, resulting in a decrease in service time. Berry Jaeker and Tucker (2017) further refined and unveiled an N-shaped relationship between workload and LOS, suggesting the existence of a second transition point at very high workloads. Beyond this point, they observe an increase in LOS, as servers exhibit behavior described as "beyond the point of speeding up." The authors coined this a "saturation effect". They attributed it to a selection mechanism wherein servers can no longer prioritize early discharge of less severe patients during periods of high workload, as they are left with the more time-consuming patients.

The observed variation in effects across different congestion levels has led to the development of the “load effect on service times” (LEST) framework, which delineates three load-level types: changeover (from zero to positive load), load (instantaneous level of load), and extended load which is load over time (Delasay et al., 2019). Given the unique attributes of the civil judicial process and the global trend of increasing demand amid resource constraints, our study specifically focuses on an environment where workload levels have consistently ‘passed the point of speeding up’ (a term suggested by Berry Jaeker and Tucker, 2017), indicative of an extended load environment, wherein a heavy workload persists over a prolonged period. Consequently, we aim to explore the ramifications of congestion within this setting, where all unresolved cases inherently entail complexity, and servers face a 'perpetual backlog' (Gilbert, 1996).

Within the context of extended load environments, KC and Terwiesch (2009) reported instances of changes in work content. Their investigation within surgical wards unveiled a deterioration in medical quality attributable to server fatigue, which in turn led to complications (i.e., changes in work content) and subsequently, to increases in LOS. Chan et al. (2017) found that delays in the queue for the ICU, which is measured as extended emergency department LOS, might cause patient deterioration, thus changing the work content and increasing the later ICU workload and subsequently LOS. However, these authors observed delays before entering the process, while we aim to observe in-process delays (i.e., added waiting time between consecutive tasks by the server for the same customer).

The LEST framework (Delasay et al., 2019) further categorizes the effects of workload based on the system component (server, network, or customer) that influences the service time determinants. These determinants encompass changes in work content, alterations in service speed, and variations in in-process delay, all of which ultimately impact service time. However, existing literature mainly focuses on the effects instigated by servers, primarily through various behavioral mechanisms. In contrast, this study aims to observe changes in work content induced by the customer.

In the current setting of civil cases in the trial phase, we use the number of decisions in motions as a measure of change in work content induced by the customer. Motions are a formal method of communication between the parties and the judge, where the parties can request the judge's approval on different issues. These issues vary in nature, from administrative matters, such as changes in representative lawyers, to legal issues, such as requests to correct claims previously presented. These requests require a formal response from all parties involved in the case, and subsequently, a formal decision from the judge. We conjecture that since all preliminary processes are concluded before the trial phase, motions filed in this phase are indicative of changes in the content of the case. For this reason, we use the number of decisions in motions to proxy changes in work content induced by the customer, which in our setting, corresponds to the parties involved in the case.

The link we propose between decisions in motions and changes in content stems from typical scenarios where extended lead times lead to situations necessitating decisions. For instance, dates scheduled months in advance become unfeasible, damages change, attorneys change over the course of handling a case, judges are promoted to higher jurisdiction levels, and the relevance of witnesses or experts may change over time relative to when the case was initially opened. These are just a few examples of how prolonged lead times can give rise to circumstances that require judicial decisions to address evolving aspects of the case.

Building upon this foundation, our research concentrates on the additional work necessitated per task due to the evolution of the task outside of the system. In our setting, this evolution refers to changes in the circumstances of the case due to a prolonged lead time. Thus, we focus on changes in work content instigated by the customer in an extended load environment. In particular, we anticipate that in-service delays, regardless of their cause, will increase the effort needed for the server to complete the task. We therefore hypothesize as follows:

Hypothesis 1 (H_1). *In an extended load environment, in-process delay increases the server's level of effort.*

2.2. Effect of congestion on judicial outcomes

Research on court congestion has already provided evidence indicating that congestion is influenced by factors such as inefficient processes, judicial passivity, and mismanagement (Mitsopoulos and Pelagidis, 2010; Dalton and Singer, 2014; Castro and Guccio, 2015; Moffett et al., 2016; Peyrache and Zago, 2016). This congestion has been shown to result in a range of effects. Firstly, it leads to a reduction in throughput (Beenstock and Haitovsky, 2004; Mitsopoulos and Pelagidis, 2007; Dimitrova-Grajzl et al., 2012; Coviello et al., 2014). Additionally, research indicates that prolonged proceedings may hinder a judge's ability to reach a just resolution (President's Commission on Law Enforcement and Administration of Justice, 1967: p.129; Nagel and Neef, 1978; Best and Tiede, 2015).

Aware of these negative consequences of court congestion, judges often attempt to mitigate congestion by various mechanisms, such as expediting case resolution through early dismissals or settlements, thereby seeking to avoid the additional effort required for a full trial and judgment on the merits (Galanter, 2004; Helland and Klick, 2007; Narayan and Smyth, 2007; Epstein et al., 2013; Dimitrova-Grajzl et al., 2014). Engel and Weinshall (2020) found that a reduction in caseload, caused by adding more judicial resources, led to a higher proportion of cases concluding with a judgment on merits, suggesting that the added capacity allowed the judges to invest more time in the case in the form of trial and judgment. However, this stream of literature examines the consequences of changes in caseload and their impact on case outcomes, driven by judges' strategic behaviors. In contrast, we focus on the consequences of delay in case outcomes.

Building upon the claim we presented when developing hypothesis H_1, we suggest the following relationship for the effect of delays on the probability of settlements in the presence of extended load. In H_1, we predicted that, for civil cases in the trial phase, where the preliminary processes have been completed and most of the judges' efforts to induce settlements have already been exhausted, delays would cause changes in the work content. We conjecture that changes of this nature during the trial phase increase the case's complexity and thus reduce its probability of ending in ways other than a judgment on merits. Therefore, we hypothesize as follows:

Hypothesis 2 (H_2). *In an extended load environment, in-process delay increases the probability of a judgment on merits,*

3. Institutional Background

The characteristics of the operational performance of the Israeli judicial system are not unique, nor are its challenges related to congestion (CEPEJ, 2016). It shares similarities with other common-law-based systems such as those in the United Kingdom, Australia, Canada, India, Ireland, and South Africa. Consequently, these systems also encounter comparable issues related to long lead times and delays in their operations.

The Israeli court system follows a hierarchical structure consisting of three primary tiers: Magistrate Courts, District Courts, and the Supreme Court. The Magistrate Courts, spread across twenty-five locations, serve as the initial trial venues for criminal and civil cases, handling minor offenses, small claims, and family law matters. They cover a wide range of legal disputes with relatively low stakes. In contrast, the six District Courts hold broader jurisdiction, addressing more serious criminal charges, significant civil disputes, appeals from Magistrate Courts, and administrative law matters. They operate as a crucial intermediate tier in Israel's judicial hierarchy with original and appellate functions. The top tier consists of the Supreme Court, which holds appellate authority over lower court decisions and original jurisdiction in cases concerning fundamental rights and administrative law, in addition to serving as the High Court of Justice. Table 1 provides annual[3] case-flow measures across these three tiers.

Table 31: Annual Summary Statistics for the Israeli Court System

	Judges	Incoming cases	Resolved cases	Pending cases
Supreme Courts	16	8,768	9,167	3,071
District Courts	205	62,303	68,394	78,063
Magistrate Courts	451	727,680	745,903	320,557
Total	672	798,751	823,464	401,691

In this manuscript, our focus is on civil cases, both within the Magistrate Courts and the District Courts. We have chosen to concentrate on these cases not only because of the considerable delays they face but also due to their significant consumption of judicial resources (Aviv and Erental, 2021). Additionally, civil cases tend to be lengthier than other types of cases and have a substantial impact on the operational efficiency of courts (Falavigna and Ippoliti, 2023). The civil judicial process can be broadly outlined as a series of sequential events from the court's perspective, as illustrated in Figure 1. Despite the sequential nature of this process, cases may exit the process (e.g., through settlements, dismissals, withdrawals, etc.) at any point in time.

[3] Based on the 2019 Annual Report, as published by the Court Administration https://www.gov.il/BlobFolder/reports/statistics_annual_2019/he/%D7%93%D7%95%D7%97%20%D7%A9%D7%A0%D7%AA%D7%99%202019.pdf

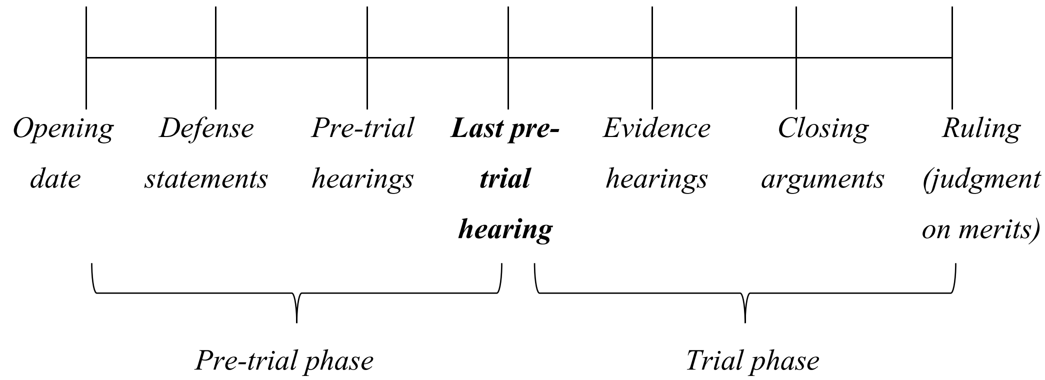

Figure 1: Sequence of Events in the Civil Judicial Process

The process commences with the filing of a civil case with the court, prompting the opposing parties to submit their written defense statements. This action usually prompts the assignment of the case to a designated judge who will oversee it henceforth. Subsequently, the evidence discovery process unfolds, wherein tasks are carried out by the involved parties under the judge's guidance. These tasks encompass, for example, obtaining expert reports from both sides, commissioning court-appointed expert reports, conducting witness depositions, and furnishing supporting documents for damage assessments.

During the pre-trial phase, judges can communicate with the involved parties through decisions and instructions, or they may summon the parties for brief court hearings. These hearings are usually scheduled in the judges' calendar between 8 am and 10 am and typically last around 30 minutes each. The frequency of these hearings can vary depending on factors such as the complexity and type of case, as well as the judge's work schedule. Between these hearings or court interactions, the parties undertake the necessary tasks as instructed. In this part of the case life cycle, judges primarily focus on familiarizing themselves with the case, assigning tasks, and actively promoting Alternative Dispute Resolutions (ADR), both within and outside the courtroom. It is worth noting that one of the hallmarks of an adversarial system is the active involvement of judges in expressing their opinions and encouraging settlements during the pre-trial phase. This proactive stance likely contributes to the high rate of cases that do not progress to the trial phase (87.87% in our validated dataset, as depicted in Figure 2, Section 4.1), noting that in the trial phase, judges are more limited in their abilities to intervene. Importantly, delays experienced during the pre-trial phase are often attributed to processes carried out by the parties outside the judicial system rather than solely a consequence of judges' workloads; thus, not all waiting periods in this phase can be categorized as in-process delays.

Considering these factors, this study concentrates on the trial phase, which forms the core part of the case. This phase encompasses the presentation of evidence, testimonies, and cross-examination of witnesses by opposing attorneys, all conducted over multiple trial hearings. Subsequently, it culminates in closing arguments, followed by the judge's ruling on the merits of the case. During this phase, judges maintain their role as impartial decision-makers and are constrained in their ability to intervene.

During the trial phase, the level of uncertainty regarding the case content tends to be relatively low because this type of uncertainty was resolved during the pre-trial phase. However, there is significant uncertainty regarding the trial phase duration, primarily due to multiple queue re-entries (each time a trial session ends, the case re-enters the judge's queue) and the judges' busy calendars. Moreover, uncertainty exists regarding the likelihood of reaching a judgment on the merits. Thus, as the work content in the trial phase is relatively well-defined, waiting time in this phase is largely attributable to the judges' busy schedules. Therefore, it can be considered a delay.

To illustrate, in our dataset, cases in magistrate courts with two trial hearings had an average *Trial-phase lead time* of 17.62 months, measured from the last pre-trial hearing to the closing date. We can estimate the required judicial effort in this phase using data from the ICRD (see Section 4.1). For example, rough calculations based on average estimations for all magistrate court case types in the dataset suggest that fewer than 12 working hours are dedicated to preparing and conducting the two trial hearing sessions and producing a judgment on the merits. As a result, the Flow Time Efficiency, or the ratio of the net time spent on the task by the server (i.e., the judge) divided by the total flow time, is approximately 0.39% (based on 22 working days per month and 8 working hours per day). This indicates that lead time in the trial phase primarily corresponds to in-process delay, which makes it a suitable proxy for this metric. Therefore, by focusing on cases during the trial phase, it is possible to use the lead time as the independent variable when testing both H_1 and H_2.

4. Empirical setting

4.1. Data

The first dataset was compiled through collaboration with an Israeli not-for-profit organization called Court Worm[4], utilizing publicly available information from the Israeli court information system. In the first step, we collected data on all civil cases from both magistrate and district courts, spanning from January 2000 to February 2022. Three separate datasets were produced from this step: cases (792,482 observations, 13 variables), decisions (1,859,993 observations, 7 variables), and meetings (2,043,195

[4] "Tola'at Hamishpat": https://xn----8hcborozt8bdd.xn--9dbq2a/%D7%97%D7%99%D7%A4%D7%95%D7%A9#gsc.tab=0

observations, 6 variables). The second step was merging these three datasets into a single case-level dataset comprising 792,482 observations. Figure 2 illustrates the data selection process from this point onwards.

After merging the datasets, we further refined our data by focusing solely on civil proceeding[5] cases resolved between September 1st, 2012, and December 31st, 2019, excluding 357,811 cases (which accounted for 45.15% of the total cases). The rationale behind this limitation was to ensure that all observations included in our analysis pertained to cases that had completed their life cycle. By specifying the closing date but not restricting the opening date, we avoided censoring cases with exceptionally long lead times. The start date of the research period was chosen to postdate the completion of the court's adoption of a new information system, which was gradually implemented during 2007-2010. The end date of the research period was set to avoid capturing the effects of COVID-19 on case management. Following further exclusions stemming from data errors and misclassifications (1,238 cases, 0.156% of overall cases), the dataset was comprised of 433,433 cases, each of which included complete case-specific information along with details about the case's meetings and decisions.

In Section 3 we described the two parts of the civil judicial process: the pre-trial phase and the trial phase, outlining the rationale behind our decision to focus our research on the trial phase. Accordingly, we excluded cases resolved during the pre-trial phase (380,845 cases, 48.06% of overall cases). Subsequently, from the remaining cases, we further excluded those that did not have at least one pre-trial hearing, as we could not verify the date marking the commencement of the trial phase for these cases (3,591 cases, 0.45% of overall cases).

To conclude, our final dataset consists of 48,997 cases with validated data and an end-date between September 1st, 2012, and December 31st, 2019, each having undergone at least one trial hearing and one pre-trial hearing. Despite the reductions applied to the dataset, we arrived at a large and comprehensive sample to test our hypotheses.

[5] Regular civil cases in the main track.

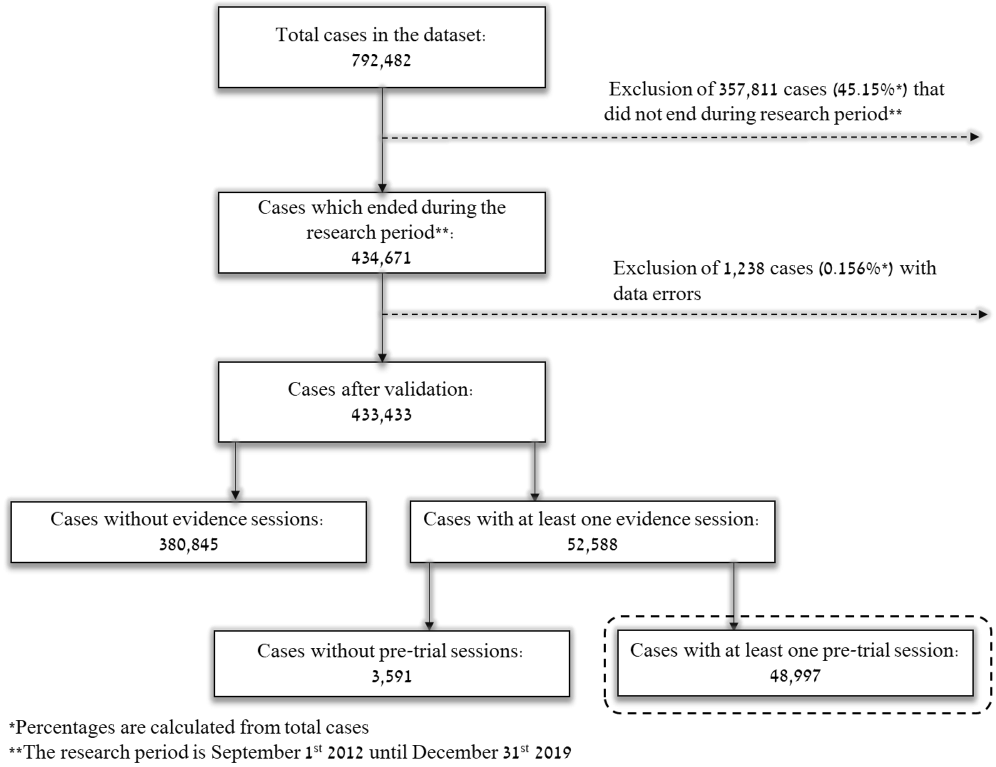

Figure 4: Data Funnel

The second dataset utilized in this research was obtained from the Israeli Courts Research Division (ICRD) under the Freedom of Information Act. This dataset includes 165 groups of cases ("case groups"), characterized by 29 variables, including classification variables (case types, case matters, court jurisdiction, and court region), statistical information regarding the mix of these cases, and the appearance rate of various judicial tasks throughout the case life cycle, and effort estimations (in judicial hours) for these tasks. These 165 case groups cover over 99% of the case flow in the system. The effort estimations were derived from activity logs completed in real-time by judges (for out-of-court activities) and court stenographers (for in-court activities). These logs were manually completed during five weeks, spanning from January 1st, 2020 to February 5th, 2020, by 88.3% of the presiding judges. Subsequently, the ICRD employed two validation methods to ensure the accuracy of the calculated average effort values. This data collection method is widely accepted in judicial workload research and is commonly utilized in most developed countries.

By combining the two comprehensive datasets described above, the case level data and the effort estimations by case groups, we gain a holistic understanding of case dynamics and judicial workload related to these cases, over more than seven years, which are suitable for testing our hypothesis. Specifically, for each case in our sample, we can use the effort estimations to calculate the time the judge spent handling specific case tasks. These case-level effort estimations can be used as a proxy for the judge's workload in testing our hypotheses. This extensive and robust dataset is unique in this context, as court systems globally encounter challenges related to data collection, coding, and validity (Shapiro, 2008). Furthermore, even when data is available within the system, access to data can still present a significant barrier (Paley et al., 2021; Dimas et al., 2023). These obstacles render empirical operations research in this domain complex and rare.

4.2. Outcome variable

To examine the impact of lead time on workload, we focus on the effort expended by judges on case-related tasks. Existing research suggests that judges are the constrained resource or bottleneck of the judicial system, at least within the Israeli court system (Bar Niv et al., 2010; Maayan et al., 2012; Azaria et al., 2023). This implies that any changes in work content requiring additional judicial effort directly influence the system's throughput. To measure the level of effort, we focus on a single case-level activity performed by judges—decisions in motions—as we explain below.

Motions are formal requests presented to the court seeking an order or judgment on various matters throughout the case life cycle. These requests cover a wide range of subjects, from issuing witness subpoenas to more intricate issues such as the dismissal of the case. Once one of the parties has submitted a motion to the court, input is required from the opposing parties, followed by the issuance of a formal decision by the judge. These decisions can take various forms, including final determinations, such as approval, rejection, or dismissal, and calls for specific actions, typically in the form of instructions. Regardless of the outcome, decisions are the formal means of communicating requests between the parties and the judge.

To estimate the level of effort invested by the judge in each case, we multiply the number of decisions made during the case from the first dataset by the effort estimation per decision (according to the case group) obtained from the second dataset as follows: $Effort_i$*=(number of decisions in motion in case i)*(effort estimation for the case group representing case i)*. This process generates an estimate of the time the judge spends on producing the described above decisions for each case in our dataset, which is measured in judicial work hours.

As outlined in Section 3, our analysis centers on the trial phase. Therefore, for each case in the dataset, our dependent variable measures the estimated effort invested in decisions that were delivered during the trial phase. This variable is used to capture the effect of trial-phase lead time on changes in work content. To provide an overview of the data, we present descriptive statistics for the judge’s level of effort throughout the case life cycle (see Table 2). This allows us to gain insights into the distribution and characteristics of decisions across different phases of the case. Note that the variable of interest (the judicial effort for decisions in motion) is a continuous variable that only takes positive values.

Table 4: Descriptive Statistics of Judicial Effort

Variable	Mean	Std. Dev.	Median
Effort for decisions in motions throughout the case life cycle (hours)	4.66	6.73	3
Effort for decisions in motions during the trial phase (hours)	1.56	3.42	0
No. of observations		48,997	

Both variables shown in Table 2 exhibit a median that is lower than the mean. This is attributed to the high incidence of an effort value of zero: 50.9% of cases have no decisions in motions during the trial phase and 14.4% experience no decisions in motions throughout the case life cycle. Figure 3 illustrates the distribution of *Effort* during the trial phase (the second variable presented in Table 2), which serves as the variable of interest when testing our hypotheses. The skewness of the graph, with the long tail, explains the findings presented in the table.

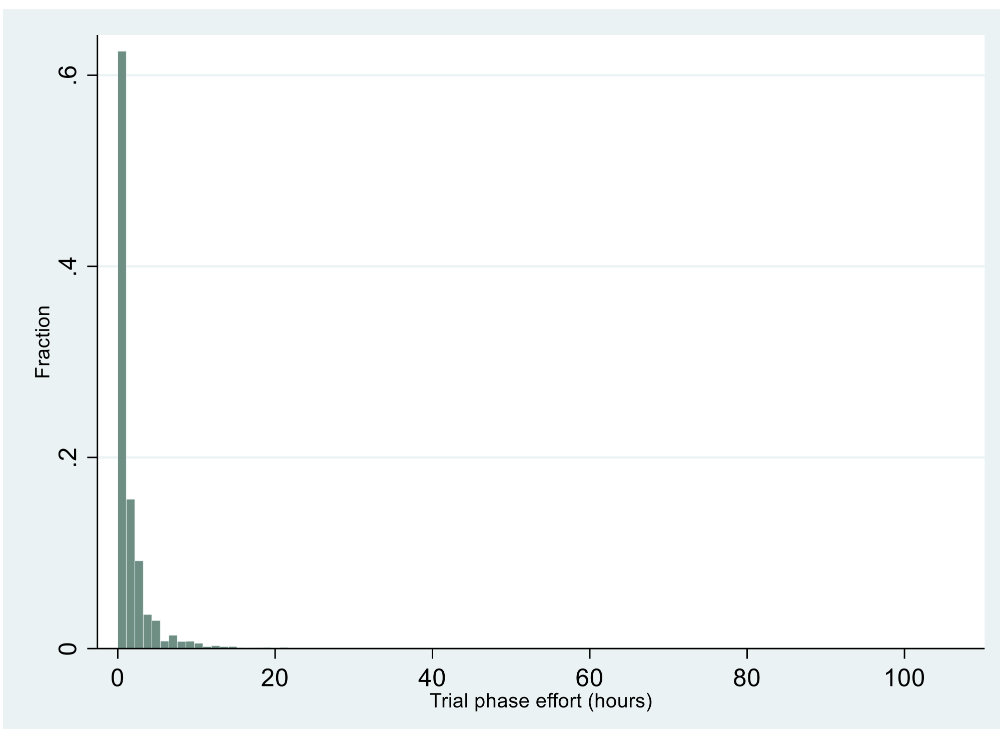

Figure 3: Distribution of Trial-phase Effort

4.3. Explanatory variables

The main explanatory variable of interest is lead time. In Table 3 we present descriptive statistics for two lead time variables – *Total lead time* and *Trial-phase lead time*. While *Total lead time* is measured from the opening of a case to its disposition, *Trial-phase lead time* is measured from the last pre-trial hearing (to denote the time point at which it was decided that the case would proceed to trial) to the case's disposition.

Table 3: Descriptive Statistics of Lead-Time Variables

Variable	Mean	Std. Dev.	Median
Total lead time (months)	39.61	17.32	36.9
Trial-phase lead time (months)	14.56	10.33	12.2
No. of observations		48,997	

From Table 3, it is evident that this environment is characterized by long lead times, with a mean *Total lead time* of 39.61 months and a mean *Trial-phase lead time* of 14.56 months. As established in Section 3, the *Trial-phase lead time*, which consists almost entirely of waiting time rather than processing time, can be considered delay. Figure 4 presents the distribution of the *Trial-phase lead time*, our variable of interest.

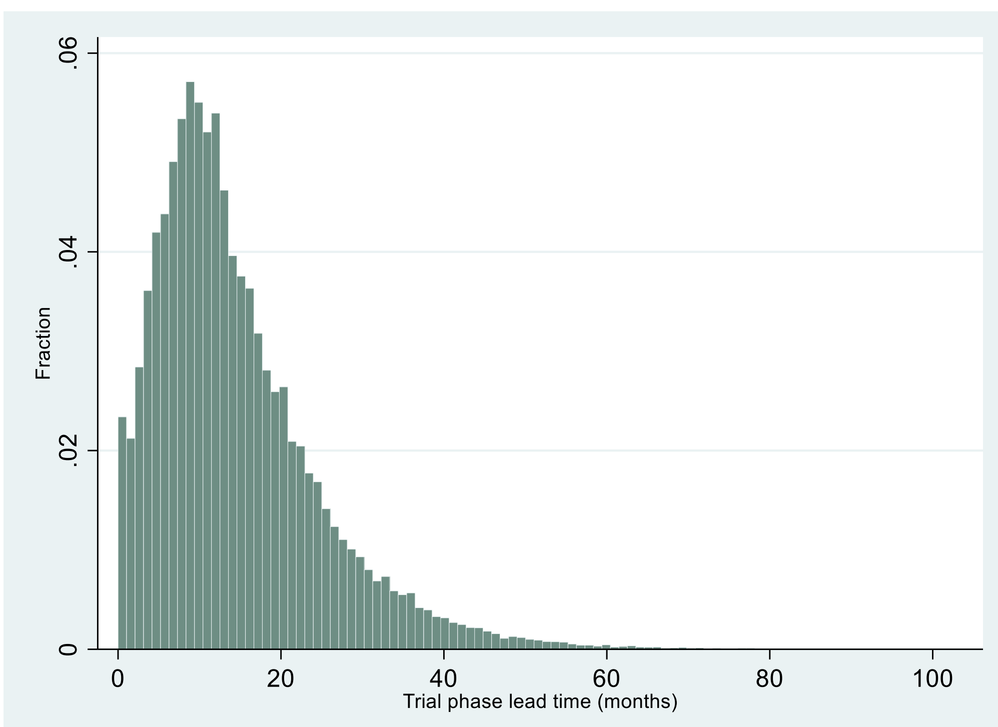

Figure 4: Distribution of Trial-phase Lead Time

Our dataset incorporates several case-level characteristics that might affect case complexity. These include *case matter*s, such as contracts, real estate, bodily harm, medical malpractice, and car accidents, and *court jurisdiction* (magistrate or district), which indicates case complexity. Table 4 presents the distribution of cases among the main categories of these two case characteristics. Additionally, we include *court region*, as practices within the same district may be more homogeneous than practices across districts.

Table 4: Case Characteristics

Variable	Most common categories	
Court jurisdiction	Magistrate court	89.99%
	District court	10. 01%
Case matter	Monetary claim	40.18%
	Bodily harm	20.33%
	Car accidents	15.09%
	Real estate	6.74%

5. Empirical model

The question we aim to answer to address H_1 is: Does delay increase the judicial effort? Ideally, to test this hypothesis, we would randomly assign cases to different levels of judicial delay, regardless of their individual characteristics (such as complexity). Unfortunately, this approach would not be feasible, as we are constrained by observational data (see Section 4.1). This setting presents a significant empirical challenge, which we address by invoking the reduced-form model for the inference, which seeks to explain the variance in effort with lead time:

$$Effort_i = \alpha + \beta \cdot LeadTime_i + \gamma \cdot \mathbf{X}_\mathrm{i} + \epsilon_i \tag{5}$$

In this equation, *Effort* is the dependent variable, representing the estimated effort exerted by a judge to produce decisions in motion during the trial phase of case *i*. The main independent variable is *Lead Time*, measured as the period between the last pre-trial hearing and the closing date of case *i*. **X** is a vector of control variables, including the case characteristics described in Section 4.3, and a time variable to account for time-dependent operational factors such as changes in demand. This time variable is an indicator of the year in which the case entered the trial phase (i.e., had its last pre-trial hearing). Lastly, ϵ captures the error term.

Estimating equation (1) using an ordinary least squares (OLS) approach is likely to yield a biased estimate of β, the effect of *Lead time* on *Effort*. This bias may arise from reverse causality, meaning that delays may arise from changes in work content. In other words, cases requiring higher effort may exhibit longer lead times. This issue would not arise if we could control lead time exogenously or if the reasons for the delay in our dataset were observable. Therefore, we need a method of capturing genuinely exogenous variation in lead time. We next describe the approach adopted to deal with this issue when testing H_1 and H_2.

5.1. Instrumental variables

To address the empirical challenge stemming from reverse causality bias, we employ the instrumental variable (IV) approach. The conditions for an acceptable IV are: (1) the IV must be relevant in explaining the variation in the potentially endogenous variable—in our case, *Lead Time*; and (2) the IV must be unrelated to unobservable factors captured in the error term ϵ_i, meaning that it must affect *Effort* only through *Lead Time* (Wooldridge, 2012).

To identify a suitable IV, we sought exogenous shocks that influence *Lead Time* without directly affecting *Effort*. Specifically, we looked for disruptive events that prolonged *Lead Time* irrespective of the work content, meaning the increase in lead time was not caused by *Effort*. We identified two disruptive events that occurred during our research period: a military conflict[6] lasting 50 days in 2014 and a court workers' strike[7] lasting two weeks in 2016. We opted to employ both IVs, as this approach improves the efficiency of the estimations and results in tighter confidence intervals (Wooldridge, 2012). We define each of the IVs as a binary indicator that captures whether the event (conflict or strike) occurred during the case's trial phase.

The military conflict, which commenced on July 8th, 2014, and concluded on August 26th, 2014, had various ramifications. These included the cancellation of all governmental programs, the closure of universities, the prohibition of gatherings of over 300 people, and other disruptions to the work routine across numerous sectors and geographical areas. To validate this IV, we investigated whether the event indeed caused a disruption in the court system's routine by affecting the ability of the system to conduct trial hearings. Specifically, we checked whether this disruption caused cancellations or postponements of trial hearings, leading to cases having to re-enter the queue, resulting in a longer *Lead Time*. We conducted a two-sample test of proportions, comparing the proportion of trial hearings held (out of the hearings

[6] https://en.wikipedia.org/wiki/2014_Gaza_War
[7] https://www.globes.co.il/news/article.aspx?did=1001143354

scheduled) for every day during the conflict period with the same period in the previous year (*difference = 0.06, p = 0.0127*) and in the subsequent year (*difference = 0.04, p = 0.0418*). In our dataset, the conflict occurred during the trial phase of 18.7% of the cases.

The second IV, the court workers' strike, commenced on July 17th, 2016 and concluded on August 1st, 2016. The strike involved the courts' administrative workers, including court stenographers, protesting their employment conditions. Initially, there were concerns that the strike might continue until September. However, after two weeks of negotiations involving all relevant stakeholders, the chair of the General Organization of Workers announced that an agreement had been reached, and the strike was terminated. Court administrative staff are a crucial resource for ensuring the proper management of the court system. Court stenographers, in particular, play a vital role in conducting hearings. While alternatives exist, such as electronic recording systems, the reports of court stenographers are the most common method of keeping records of hearings. Consequently, this strike led to hearing cancellations.

To validate the IV based on the court workers' strike, we conducted a two-sample test of proportions, comparing the proportion of trial hearings held (out of the hearings scheduled) for every day during the strike period with the same period in the previous year (*difference = 0.08, p = 0.0155*) and in the subsequent year (*difference = 0.07, p = 0.0191*). In our dataset, the strike occurred during the trial phase of 17.7% of the cases.

After establishing that these disruptions had tangible effects, we proceeded to compare cases whose trial phase occurred during at least one of the timeframes represented by the IVs, which we will refer to as the disrupted group (IV=1), with cases whose trial phase did not occur during either of these timeframes. At this stage, we opted to combine the effects of the two IVs, as only 1.88% of the cases in the dataset were affected by both. While separate estimations of the effects are still required, aggregating the IVs at the descriptive level allows for a better understanding of the differences between disrupted and undisrupted cases. First, we compared case characteristics to verify whether the groups were balanced, as presented in Table 5.

Table 5: Case Characteristics by IV Group

Variable	Most common categories	IV=0	IV=1
Court jurisdiction	Magistrate court	90.09%	89.80%
	District court	9.91%	10.20%
Case matter	Monetary claim	38.44%	43.47%
	Bodily harm	20.73%	19.56%
	Car accidents	16.11%	13.14%
	Real estate	6.79%	6.64%

Table 5 illustrates that the distribution of cases among the two types of court jurisdiction, as well as among the main categories of case matter, is very similar for the two groups. The next step was to compare descriptive statistics of the two groups (disrupted and undisrupted) to identify differences in the variables of interest, as presented in Table 6.

Table 6: Summary Statistics by IV Group

Variable	IV=0			IV=1		
	Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
Total lead time (months)	36.53	16.46	33.83	45.45	17.40	43.20
Trial-phase lead time (months)	11.20	7.52	9.87	20.92	11.83	18.57
Total effort for decisions in motions (hours)	4.35	6.32	3	5.26	7.43	3
Trial-phase effort for decisions in motions (hours)	1.28	2.84	0	2.09	4.27	1
No. of observations		32,093			16,904	

The differences between the two groups are evident. Firstly, we observe a significant difference in the mean *Total lead time*, which is 8.92 months (24.42%) higher for the disrupted group than the undisrupted group. This is primarily attributed to the difference in the mean *Trial-phase lead time*, which is 9.72 months (86.79%) higher for the disrupted group. The similarity in absolute differences (i.e., 8.92 vs. 9.72 months) stems from the definition of the IV, which is based on whether the trial phase occurred during the disruption period. The fact that the difference in lead time between the two groups primarily arises from the trial phase, suggests that the groups are balanced in terms of the pre-trial phase distribution (the Mann-Whitney test shows no significant difference between the *Pre-trial lead time* distribution of the two IV groups: $z = 0.501$, $p = 0.6164$). This difference serves as the first indication of the relevance of the IVs, as it demonstrates a correlation between the IVs and the endogenous variable *Lead Time*, fulfilling the first condition of a suitable IV. Table 6 also shows a difference in the mean *Total effort*, which is 0.91 hours (20.92%) higher for the disrupted group than the undisrupted group. This difference is primarily reflected in the mean *Trial-phase effort*, which is 0.81 hours higher for the disrupted group, corresponding to a 63.28% increase. All the differences described above were found to be statistically significant with $p < 0.001$ in mean-comparison t-tests. Additionally, we observe that while the median *Trial-phase effort* is 0 for the undisrupted group, it is 1 for the disrupted group. This indicates that a larger proportion of cases in the disrupted group had decisions in motion during the trial phase.

To further evaluate whether the two IVs meet the first condition described above, we use a simple OLS regression to test the null hypothesis that there is no relationship between each of the IVs and the endogenous variable, the *Trial-phase lead time*. The results are presented in Table 7.

Table 7: Results of the First Stage Regression

	Dependent variable: Trial-phase lead time			
	(1)	(2)	(3)	(4)
Military conflict IV	10.35***	10.18***	9.96***	11.36***
Std. E.	(0.33)	(0.33)	(0.31)	(0.28)
Workers' Strike IV	10.09***	9.90***	9.74***	13.22***
Std. E.	(0.27)	(0.27)	(0.26)	(0.29)
Intercept	10.84***	11.09***	12.92***	60.93***
Std. E.	(0.18)	(2.65)	(2.60)	(2.98)
Case matter controls	No	Yes	Yes	Yes
Court controls	No	No	Yes	Yes
Start year controls	No	No	No	Yes
R^2	0.2634	0.2770	0.2976	0.4959
No. of observations	48,997	48,997	48,997	48,997

Notes: Estimated using OLS. Robust standard errors, adjusted for 505 judges' level clusters, are reported in parentheses. *, **, *** represent statistical significance at the 10%, 5%, and 1% levels.

The direction of the effect aligns with our expectation that the disruptive events would exogenously increase the *Trial-phase lead time*. The results for all four models demonstrate a significant, positive coefficient for both IVs, with the highest estimate obtained for the workers' strike in column (4). The higher coefficient for the conflict IV than for the strike IV (13.22 months vs. 11.36 months) may seem surprising given that the conflict lasted longer. This can be explained by the fact that the conflict only caused a moderate level of interference to the day-to-day practice of holding trial hearings throughout its 7-week duration, whereas the workers' strike created a more significant shock during its 2-week period. Regarding the size of the effects (i.e., differences in lead times of many months while the disruptions only lasted for several weeks), this arises from the overburdened system; when a hearing is postponed, the case re-enters the queue, which is already months long.

We now consider the second condition for a suitable IV, known as the exclusion restriction, which requires that the two IVs be uncorrelated with unobservable factors that may affect the *Trial-phase effort*. Although it is not feasible to test this condition directly, we attempt to rule out potential mechanisms that may lead to a violation of this restriction by examining our data for case characteristics that represent unobservable factors affecting the *Trial-phase effort*, which are captured by ϵ_i. We first observe differences between the two groups (disrupted and undisrupted) in the distribution of a specific case characteristic –the number of pre-trial hearings – an unobserved factor that might affect both the *Trial-phase lead time* and the *Trial-phase effort*. Figure 5 compares the distribution of the number of pre-trial hearings (effectively a measure of case complexity) for the two IV groups.

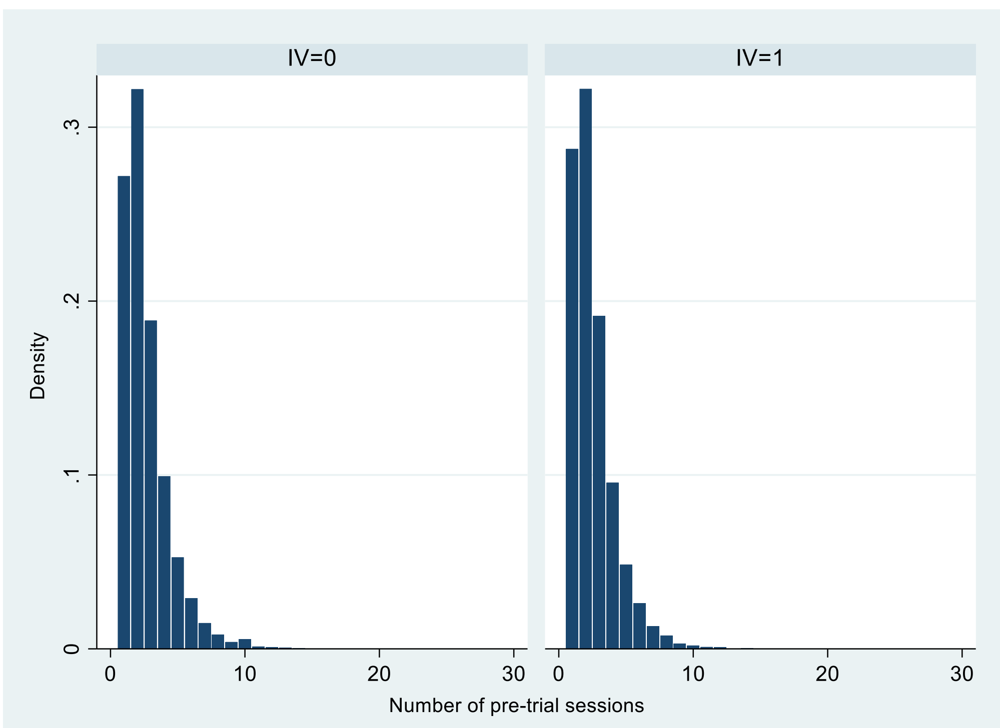

Figure 5: Distribution of Pre-Trial Hearings by IV Group

The distributions appear to be similar for the two groups, an observation that was checked using a two-sample Kolmogorov-Smirnov test. We found no significant difference (*difference = 0, p = 1*) in a one-sided test of whether cases affected by a disruption are more complex than undisrupted cases. However, we observed a small but significant difference in the two-sided test, suggesting that cases affected by the disruptions are, in fact, less complex (*difference = 0.0185, p = 0.001*) than undisrupted cases, albeit by a small margin. This implies that our estimations might be conservative, as less complex cases should result in fewer decisions in motion.

5.2. Estimation models

Having acquired evidence in support of the validity of the instrumental variables, we explore the relationship between lead time and effort. First, we plot the average *Trial-phase lead time* for every decile of the *Trial-phase effort* for each IV group (see Figure 6).

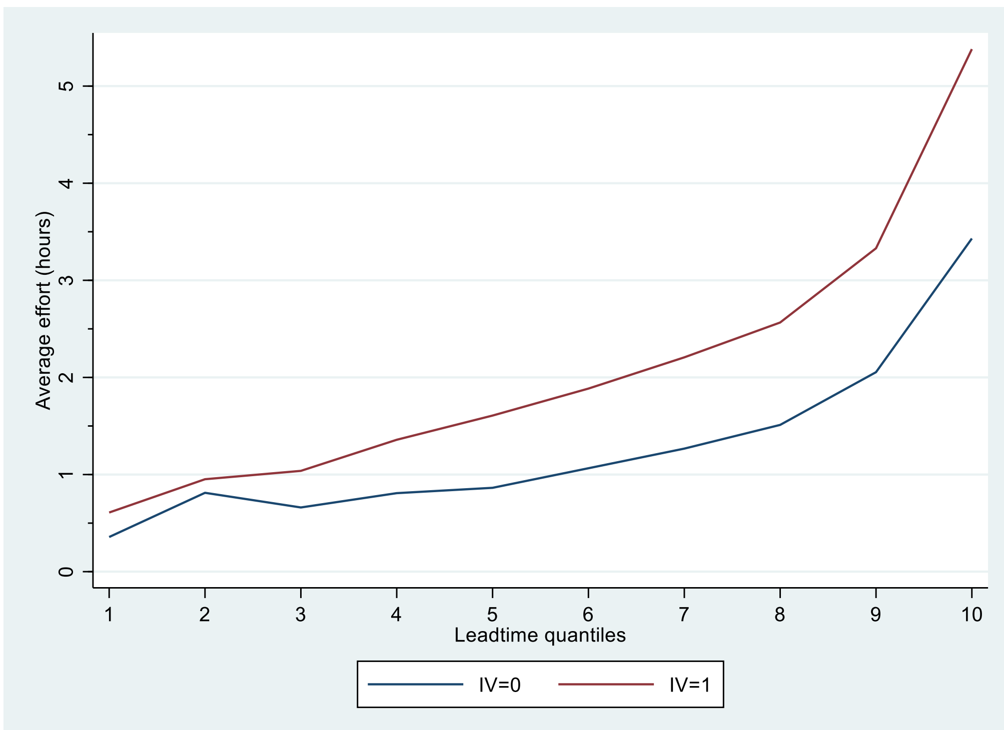

Figure 6: Average Trial-phase Effort Across Trial-phase Lead Time Deciles by IV Group

The figure illustrates an increasing trend for both groups, indicating a positive relationship between *Trial-phase lead time* and *Trial-phase effort*. However, the slope is steeper for the disrupted group, so the difference between the two groups widens as the lead time increases. For instance, the 10% of cases with the longest *Trial-phase lead time* in the disrupted group, which already experiences a longer lead time than the unaffected group (see Table 6), exhibit higher average *Trial-phase effort* than the 10% of cases with the longest *Trial-phase lead time* in the unaffected group. This supports our hypothesis that an exogenous increase in *Trial-phase lead time* will necessitate additional effort, which we attribute to changes in work content. Due to the high skewness and high proportion of censored data of the dependent variable, *Trial-phase effort*, which might indicate selection bias, we propose a three-stage model for testing H_1: (i) estimating the effect of the IVs on *Trial-phase lead time*; (ii) estimating the effect of the predicted *Trial-phase lead time* on the probability of exerting *Trial-phase effort t*; (iii) estimating the effect of the predicted *Trial-phase lead time* on the level of *Trial-phase effort* given that effort was exerted. Equation (2) accounts for the endogeneity issues by incorporating the IVs. Then, equations (3) and (4) represent the Heckman two-step correction model (Heckman, 1979), leveraging the predicted *Trial-phase lead time* from equation (2) as the independent variable. The equations for these stages are presented below:

$$LeadTime_i = \alpha + \varphi \cdot IVConflict_i + \omega \cdot IVStrike_i + \gamma \mathbf{X}_{\mathrm{i}} + \varepsilon_i \tag{1}$$

$$Prob(Effort_i > 0|\mathbf{X}_{\mathrm{i}}) = \Phi\big(\alpha + \beta_{IV} \cdot \widehat{LeadTime}_i + \gamma \mathbf{X}_{\mathrm{i}}\big) \tag{3}$$

$$\log(Effort_i|Effort_i > 0) = \alpha + \varphi\beta_{IV} \cdot \widehat{LeadTime}_i + \gamma \mathbf{X}_{\mathrm{i}} + \lambda \mathrm{IMR}_i + \epsilon_i \tag{4}$$

We remind the reader that *Lead Time* is measured as the period in months from the last pre-trial hearing to the closing date of case i, $\boldsymbol{X}$ is a vector of control variables, and *Effort* is the estimated time spent by the judge on handling decisions in the trial phase of case i. In addition, Φ is the cdf (cumulative distribution function) of the standard normal distribution, and $\mathrm{IMR_i}$ is the inverse Mills ratio derived from the first stage probit model in equation (3) to correct for selection bias.

In the next analysis, we use the exogenous variable derived from equation (2)—the predicted *Trial-phase lead time*—to test H_2 by assessing whether delay impacts the trial outcome, measured as the probability of a judgment on merits. Ending a case with a written and elaborate verdict on the matter is a task that demands significant effort from the judges, estimated to be 7.79 hours (this is the mean of the time estimations for the types of cases in our dataset). To test H_2, we estimate the following equation:

$$Prob(Judgement_i = 1|\mathbf{X}_{\mathrm{i}}) = \Phi\big(\alpha + \beta_{IV} \cdot \widehat{LeadTime}_i + \gamma \mathbf{X}_{\mathrm{i}}\big) \tag{2}$$

The next section presents and discusses the results obtained from estimating the above equations.

6. Results

Equation (2) evaluates the effect of the IVs and predicts the endogenous regressor *Trial phrase lead time* for the subsequent stages. The results, which were presented in Table 7 of Section 5.1, demonstrated a significant, positive effect of both IVs on *Trial phrase lead time*. Equation (3) was estimated using the Probit maximum likelihood model, and the results are presented in Table 8.

Table 6: Results of the Second Stage Regression for Testing H_1

	Dependent variable: Trial-phase effort (Yes=1)			
	(1)	(2)	(3)	(4)
Trial-phase lead time (predicted)	0.046***	0.044***	0.046***	0.038***
Std. E.	(0.002)	(0.002)	(0.002)	(0.002)
Intercept	-0.693***	-0.037	-0.159	-0.133
Std. E.	(0.044)	(0.235)	(0.317)	(0.371)
Case matter controls	No	Yes	Yes	Yes
Court controls	No	No	Yes	Yes
Start year controls	No	No	No	Yes
Pseudo R^2	0.0469	0.0539	0.0849	0.0884

Notes: Estimated using Probit maximum likelihood. Robust standard errors, adjusted for 505 judges' level clusters, are reported in parentheses. *, **, *** represent statistical significance at the 10%, 5%, and 1% levels.

The results in columns (1)-(4) all demonstrate a significant positive coefficient for the predicted *Trial-phase lead time*. This implies that an increase in *Trial phase lead time* elevates the probability of the judge exerting *Effort* on decisions in motions in the trial phase, which we attribute to changes in work content induced by the customers. Probit regression coefficients cannot be directly interpreted, so we derive the marginal effect of our coefficient of interest, predicted *Trial phrase lead time*, from the full regression in column (4), yielding 0.014 ($p < 0.001$). This can be interpreted as follows: an additional month of delay in the trial phase increases the probability of the judge exerting *Effort* on decisions in motions by 1.4%. (From Table 3, we observe that the mean *Trial-phase lead time* is 14.56 months.)

The final stage, described in equation (4), is a simple OLS regression of the log *Effort* on the predicted *Trial-phase lead time*, for cases where effort was exerted (*Effort > 0*). The results are presented in Table 9 below.

Table 9: Results of the Third Stage Regression for Testing H_1

	Dependent variable: Logged *Effort* (if *Effort*>0)			
	(1)	(2)	(3)	(4)
Trial-phase lead time (predicted)	0.012***	0.009***	0.018***	0.042***
Std. E.	(0.003)	(0.000)	(0.002)	(0.004)
Intercept	1.006***	0.823***	0.864***	0.161***
Std. E.	(0.035)	(0.126)	(0.154)	(0.459)
Case matter controls	No	Yes	Yes	Yes
Court controls	No	No	Yes	Yes
Start year controls	No	No	No	Yes
R^2	0.0852	0.2904	0.3333	0.3422

Notes: Estimated using OLS. Robust standard errors, adjusted for 486 judges' level clusters, are reported in parentheses. *, **, *** represent statistical significance at the 10%, 5%, and 1% levels.

The results presented in Table 9 suggest that *Trial phrase lead time* increases the logged *Trial-phase effort*, with a significant effect size of 0.042. This indicates that a unit increase in the predicted *Trial phase lead time* is associated with a statistically significant 4.23% increase in the log of effort, holding all other things equal, corrected for selection bias. These results, combined with the second stage results, support H_1 and demonstrate that delay increases the effort expended by judges, which we believe to be necessitated by a change in work content.

The results from estimating equation (5) to test H_2 are presented in Table 10.

Table 10: Results of the second stage regression for testing H_2

	Dependent variable: Judgement on merits (Yes=1)			
	(1)	(2)	(3)	(4)
Trial-phase lead time (predicted)	0.035***	0.028***	0.032***	0.032***
Std. E.	(0.002)	(0.002)	(0.002)	(0.002)
Intercept	-0.232***	0.344	0.098	-0.415
Std. E.	(0.036)	(0.256)	(0.272)	(0.69)
Case matter controls	No	Yes	Yes	Yes
Court controls	No	No	Yes	Yes
Start year controls	No	No	No	Yes
Pseudo R^2	0.0260	0.0547	0.0746	0.0752

Notes: Estimated using Probit maximum likelihood. Robust standard errors, adjusted for 505 judges' level clusters, are reported in parentheses. *, **, *** represent statistical significance at the 10%, 5%, and 1% levels.

It can be seen that a longer *Trial phrase lead time* consistently increases the probability of a case reaching a judgment on merits (Judgement = 1), which is aligned with the prediction in H_2. To interpret the results, we again derive the average marginal effect from the full model in column (4), yielding 0.011 ($p < 0.001$). This implies that for every additional month of delay, the probability of reaching a judgment on merits increases by 1.1%. It is important to note that the cases in our dataset are those that have reached the trial phase. Thus, our conjecture is that once a case enters the trial phase, the longer the delay, the greater the complexity of the case, and therefore the greater the probability that the case will be resolved by a final judgment on the merits.

7. Summary

This research investigates the influence of delay on judges' workload, as it appears in civil cases within the Israeli court system. Specifically, our goal is to address a gap in the existing literature by exploring the connection between delay, which results in changes in work content prompted by the customers, and the demand for judges' effort. Through empirical analysis, we aim to shed light on the adverse effects of prolonged lead time and propose the concept of a "congestion vortex", as it appears within the context of a heavily congested court system environment.

The study narrows its focus to the trial phase of the civil process, recognizing that waiting time during this phase adds no value and is therefore deemed delayed. Our hypothesis posits that prolonged lead times during this phase lead to alterations in the work content of cases, requiring judges to exert additional effort. Various factors, including shifting conditions over time, turnover of attorneys, and evolving relevance of witnesses and experts, contribute to these changes.

To conduct our analysis, we leverage two distinct datasets. The first dataset, developed in partnership with the organization Court Worm, includes all civil cases heard in magistrate and district courts, spanning from January 2000 to February 2022. The dataset provides a rich array of case-level information, including case attributes, meetings, and decisions throughout the lifecycle of each case. The second dataset, sourced from the Israeli Courts Research Division, consists of effort estimations for diverse case events derived from real-time activity logs completed by judges and court stenographers. The final dataset consists of 48,997 cases with validated data.

We conduct a comprehensive investigation into the dynamics of the trial phase and their influence on case outcomes. To address the inherent risk of reverse causality between lead time and workload, we adopt the instrumental variable approach. We identify two exogenous shocks—a military conflict and a court workers' strike—to construct binary instrumental variables (affected or unaffected by the disruption), on the basis that these disruptions prolonged the lead time without directly impacting the judges' workload.

The regression results indicate a notable positive impact of lead time on the level of effort exerted by judges in making decisions in motion. As lead time increases in the trial phase, there is an increased likelihood of judges exerting additional effort. Our interpretation is that longer lead times elevate the probability of work content alterations, increasing the level of effort required by judges. Furthermore, extended lead times in the trial phase amplify the likelihood of cases culminating in judgments on merits, where such judgments constitute high-effort tasks in the case life cycle.

The findings of this study carry significant implications for comprehending the inner workings of the judicial system and enhancing its operational efficiency and efficacy. By empirically demonstrating the substantial influence of delay on judges' workload and case outcomes, we debunk the misconception that delay does not detrimentally affect the court system. This research contributes to both academic scholarship and policymaking in the realm of judicial management by underscoring the criticality of tackling congestion and diminishing lead times in court systems, as exemplified in some recent works on improving court operations (Bakshi, 2023; Dimas, 2023; Azaria et al., 2023; 2024; Freund and Weng, 2024).

References

Aviv, G. & Erental, D. (2021). *The burden of the judicial work in the Israeli court system*. Jerusalem, Israel: Israeli Courts Research Division. *https://www.gov.il/he/departments/publications/reports/research_27072021_a*

Azaria S., Ronen B., & Shamir N. (2023). Justice in time: A theory of constraints approach. *Journal of Operations Management, 69*(7), 1202-1208. https://doi.org/10.1002/ joom.1234

Azaria, S., Ronen, B., & Shamir, N. (2024). Alleviating Court Congestion: The Case of the Jerusalem District Court. *INFORMS Journal on Applied Analytics, 54*(3), 205-296. https://doi.org/10.1287/inte.2023.0026

Azriel, D., Feigin, P.D. & Mandelbaum, A. (2019). Erlang-S: A data-based model of servers in queueing networks. *Management Science*, *65*(10), 4607-4635. https://doi.org/10.1287/mnsc.2018.3166

Bakshi, N., Kim, J., & Randhawa, R. S. (2023). Service Operations for Justice-On-Time: A Data-Driven Queueing Approach. *Available at SSRN 3764036*. https://dx.doi.org/10.2139/ssrn.3764036

Bar-Niv, M., Lieber, Z. & Ronen, B. (2010). Focused Management in a court system: Doing more with the existing resources. *Human Systems Management*, *29*(4), 265–277.

Batt, R.J. & Terwiesch, C. (2017). Early task initiation and other load-adaptive mechanisms in the emergency department. *Management Science*, *63*(11), 3531-3551.

Beenstock, M. & Haitovsky, Y. (2004). Does the appointment of judges increase the output of the judiciary? *International Review of Law and Economics*, *24*(3), 351–369.

Berry Jaeker, J.A. & Tucker, A.L. (2017). Past the point of speeding up: The negative effects of workload saturation on efficiency and patient severity. *Management Science*, *63*(4), 1042-1062. https://doi.org/10.1287/mnsc.2015.2387

Best, J. & Tiede, L.B. (2015). Vacancy in justice: Analyzing the impact of overburdened judges on sentencing decisions. *Available at SSRN 2417348*. https://ssrn.com/abstract=2417348

Castro, M.F. & Guccio, C. (2015). Bottlenecks or Inefficiency? An assessment of first-instance Italian courts' performance. *Review of Law & Economics*, *11*(2), 317-354. https://doi.org/10.1515/rle-2015-0030

CEPEJ (European Commission for the Efficiency of Justice) (2015). *The 2015 EU Justice Scoreboard*. https://ec.europa.eu/info/sites/default/files/justice_scoreboard_2015_en.pdf

CEPEJ (European Commission for the Efficiency of Justice) (2016). *European Judicial Systems. Efficiency and Quality of Justice, CEPEJ STUDIES No. 23, Edition 2016 (2014 data)*. https://rm.coe.int/european-judicial-systems-efficiency-and-quality-of-justice-cepej-stud/1680786b58

Chan, C. W., Farias, V. F., Bambos, N., & Escobar, G. J. (2012). Optimizing intensive care unit discharge decisions with patient readmissions. *Operations research*, *60*(6), 1323-1341.

Chan, C. W., Farias, V. F., & Escobar, G. J. (2017). The impact of delays on service times in the intensive care unit. *Management Science*, *63*(7), 2049-2072. https://doi.org/10.1287/mnsc.2016.2441

Chemin, M. (2009). The impact of the judiciary on entrepreneurship: Evaluation of Pakistan's "Access to Justice Programme." *Journal of Public Economics*, *93*(1-2), 114–125. https://doi.org/10.1016/j.jpubeco.2008.05.005

Chemin, M. (2010). Does court speed shape economic activity? Evidence from a court reform in India. *Journal of Law, Economics, and Organization*, *28*(3), 460–485. https://doi.org/10.1093/jleo/ewq014

Church, T., Carlson, A., Lee, J.L. & Tan, T. (1978). Justice delayed: The pace of litigation in urban trial courts. *State Court Journal*, *2*(4), 3–8. https://www.ojp.gov/pdffiles1/Digitization/52162NCJRS.pdf

Coviello, D., Ichino, A. & Persico, N. (2014). Time allocation and task juggling. *American Economic Review*, *104*(2), 609–623. http://dx.doi.org/10.1257/aer.104.2.609

Dakolias, M. (1999). *Court Performance Around the World: A Comparative Perspective*. Washington, DC: The World Bank.

Dalton, T. & Singer, J.M. (2014). Bigger Isn't Always Better: An Analysis of Court Efficiency Using Hierarchical Linear Modeling. *Pace Law Review*, *34*(3), 1169-1189.

Decarolis, F., Mattera, G., & Menon, C. (2023). Do local court inefficiencies delay public works?: Evidence from Italian municipalities. *OECD Regional Development Papers*, No. 43, OECD Publishing, Paris. https://doi.org/10.1787/fe4dd331-en

Decker, K., Mohlen, C. & Varela, D.F. (2011). *Improving the Performance of Justice Institutions*. Washington, D.C.: World Bank Group. http://documents.worldbank.org/curated/en/244521468230960192/Improving-the-performance-of-justice-institutions

Delasay, M., Ingolfsson, A. & Kolfal, B. (2016). Modeling load and overwork effects in queueing systems with adaptive service rates. *Operations Research*, *64*(4), 867-885. https://doi.org/10.1287/opre.2016.1499

Delasay, M., Ingolfsson, A., Kolfal, B. & Schultz, K. (2019). Load effect on service times. *European Journal of Operational Research*, *279*(3), 673-686. https://doi.org/10.1016/j.ejor.2018.12.028

Dimas, G. L., Goldkind, L., & Konrad, R. (2023). Big ideas, small data: Opportunities and challenges for data science and the social services sector. *Big Data & Society, 10*(1). https://doi.org/10.1177/20539517231171051

Dimitrova-Grajzl, V., Grajzl, P., Sustersic, J. & Zajc, K. (2012). Court output, judicial staffing, and the demand for court services: Evidence from Slovenian courts of first instance. *International review of law and economics*, *32*(1), 19-29. https://doi.org/10.1016/j.irle.2011.12.006

Dimitrova-Grajzl, V., Grajzl, P. & Zajc, K. (2014). Understanding modes of civil case disposition: Evidence from Slovenian courts. *Journal of Comparative Economics*, *42*(4), 924-939. https://doi.org/10.1016/j.jce.2014.04.006

Dshalalow, J.H. (1997). Queueing systems with state dependent parameters. In: Dshalalow, J.H., Ed., *Frontiers in queueing: models and applications in science and engineering*, 61-116. CRC Press, Boca Raton.

Edie, L.C. (1954). Traffic delays at toll booths. *Journal of the Operations Research Society of America*, *2*(2), 107-138. https://doi.org/10.1287/opre.2.2.107

Engel, C., & Weinshall, K. (2020). Manna from Heaven for Judges: Judges' Reaction to a Quasi-Random Reduction in Caseload. *Journal of Empirical Legal Studies*, *17*(4), 722-751.

Epstein, L., Landes, W.M. & Posner, R.A. (2013). *The behavior of federal judges: a theoretical and empirical study of rational choice*. Harvard University Press.

Falavigna, G. & Ippoliti, R. (2023). Data envelopment analysis to investigate the Italian legal system and its reform. *Journal of Public Affairs*, *23*(4), e2877. https://doi.org/10.1002/pa.2877

Freund, D., & Weng, W. (2024). The Dedicated Docket in US Immigration Courts: An Analysis of Fairness and Efficiency Properties. *Available at SSRN 4785713*. https://dx.doi.org/10.2139/ssrn.4785713

Galanter, M. (2004). The vanishing trial: An examination of trials and related matters in federal and state courts. *Journal of Empirical Legal Studies*, *1*(3), 459-570. https://doi.org/10.1111/j.1740-1461.2004.00014.x

Gans, N., Liu, N., Mandelbaum, A., Shen, H. & Ye, H. (2010). Service times in call centers: Agent heterogeneity and learning with some operational consequences. *Borrowing strength: theory powering applications–A Festschrift for Lawrence D. Brown*, *6*, 99-123.

Gilbert, S.M. (1996). Managing case work in professional and civil services. In: *1996 Manufacturing Service Operations Management Conference Proceedings,* Amos Tuck School of Business Administration, Dartmouth College, Hanover, NH.

Helland, E. & Klick, J. (2007). The effect of judicial expedience on attorney fees in class actions. *The Journal of Legal Studies*, *36*(1), 171–187. https://doi.org/10.1086/508266

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, *47*(1), 153-161.

Hillier, D.F., Parry, G.J., Shannon, M.W. & Stack, A.M. (2009). The effect of hospital bed occupancy on throughput in the pediatric emergency department. *Annals of Emergency Medicine*, *53*(6), 767-776. https://doi.org/10.1016/j.annemergmed.2008.11.024

KC, D.S. & Terwiesch, C. (2009). Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science*, *55*(9), 1486-1498. https://doi.org/10.1287/mnsc.1090.1037

KC, D. S., & Terwiesch, C. (2012). An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing & Service Operations Management*, *14*(1), 50-65.

KC, D.S. (2014). Does multitasking improve performance? Evidence from the emergency department. *Manufacturing & Service Operations Management*, *16*(2), 168-183. https://doi.org/10.1287/msom.2013.0464

Kingman, J.F.C. (2009). The first Erlang century—and the next. *Queueing Systems, 63*(3), 3-12. https://doi.org/10.1007/s11134-009-9147-4

Kuntz, L., Mennicken, R., & Scholtes, S. (2011). Stress on the ward–An empirical study of the nonlinear relationship between organizational workload and service quality. *Ruhr Economic Papers,* No. 277. https://hdl.handle.net/10419/61454

Listokin, Y. (2002). Efficient time bars: A new rationale for the existence of statutes of limitations in criminal law. *The Journal of Legal Studies*, *31*(1), 99-118. https://doi.org/10.1086/339290

Mitsopoulos, M. & Pelagidis, T. (2007). Does staffing affect the time to dispose cases in Greek courts? *International Review of Law and Economics*, *27*(2), 219-244. https://doi.org/10.1016/j.irle.2007.06.001

Mitsopoulos, M. & Pelagidis, T. (2010). Greek appeals courts' quality analysis and performance. *European Journal of Law and Economics*, *30*(1), 17-39. https://doi.org/10.1007/s10657-009-9128-4

Moffett, K.W., Maltzman, F., Miranda, K. & Shipan, C.R. (2016). Strategic Behavior and Variation in the Supreme Court's Caseload Over Time. *Justice System Journal*, *37*(1), 20-38. https://doi.org/10.1080/0098261X.2015.1067156

Nagel, S.S. & Neef, M. (1978). Time-oriented models and the legal process: Reducing delay and forecasting the future. *Washington University Law Quarterly*, *1978*(3), 467–527.

Narayan, P.K. & Smyth, R. (2007). What explains dissent on the High Court of Australia? an empirical assessment using a cointegration and error correction approach. *Journal of Empirical Legal Studies*, *4*(2), 401-425. https://doi.org/10.1111/j.1740-1461.2007.00093.x

Paley, A., Li Zhao, A. L., Pack, H., Servantez, S., Adler, R. F., Sterbentz, M., Pah, A., Schwartz, D., Barrie, C., Einarsson, A., & Hammond, K. (2021). From data to information: Automating data science to explore the U.S. court system. In *Proceedings of the Eighteenth International Conference for Artificial Intelligence and Law* (ICAIL'21), June 21–25, 2021, São Paulo, Brazil (pp. 1-10). ACM. https://doi.org/10.1145/3462757.3466100

Peyrache, A. & Zago, A. (2016). Large courts, small justice!: The inefficiency and the optimal structure of the Italian justice sector. *Omega*, *64*, 42–56. https://doi.org/10.1016/j.omega.2015.11.002

President's Commission on Law Enforcement and Administration of Justice (1967). *The Challenge of Crime in a Free Society*. US Government Printing Office.

Shapiro, C. (2008). Coding complexity: Bringing law to the empirical analysis of the Supreme Court. *Hastings Law Journal*, *60,* 477-539.

Voigt, S. (2016). Determinants of judicial efficiency: A survey. *European Journal of Law and Economics*, *42*(2), 183–208. https://doi.org/10.1007/s10657-016-9531-6

Wooldridge, J.M. (2012). Introductory Econometrics: A Modern Approach, 5th edition. Mason, OH: South-Western Cengage Learning.