

THE INEFFICIENCY OF WORKER TIME USE

Decio Coviello
HEC Montreal

Andrea Ichino
European University Institute and
University of Bologna

Nicola Persico
Kellogg School of Management,
Northwestern University

Abstract

Much work is carried out in short, interrupted segments. This phenomenon, which we label *task juggling*, has been overlooked by economists. We study the work schedules of some judges in Italy documenting that they do juggle tasks and that juggling causally lowers their productivity substantially. To measure the size of this effect, we show that although all these judges receive the same workload, those who juggle more trials at once instead of working sequentially on few of them at each unit of time, take longer to complete their portfolios of cases. Task juggling seems to have no adverse effect on the quality of the judges' decisions, as measured by the percent of decisions appealed. To identify these causal effects we estimate models with judge fixed effects and we exploit the lottery assigning cases to judges. We discuss whether task juggling can be viewed as inefficient, and provide a back-of-the-envelope calculation of the social cost of longer trials due to task juggling. (JEL: J0, K0, M5)

1. Introduction

At 8:40 a.m., while preparing documents for a 9:00 a.m. meeting about SIGMA, David notices a new email from Steven, a business analyst from CORI, the ITS client. David expected a message from Steven in regard to R6, a major software release scheduled for the next quarter, but this email is about another issue: Steven is having problems

The editor in charge of this paper was M. Daniele Paserman.

Acknowledgments: We would like to thank seminar participants at the Bank of Italy, the CEPR Public Policy meeting 2010, the NBER Summer Institute 2009, NBER Organizational Economics Workshop 2013 in Stanford, the WPEG 2009 conference, Bocconi University, European University Institute, LSE, MIT, Oxford, Royal Holloway, Timbergen Institute, Universities of Bologna, Bonn, Southampton, UPF, and Utrecht. Manuel Arellano, Daniell Li, Francesco Manaresi, Giovanni Pica, and Alfonso Rosolia gave us useful advice in several discussions. We are also grateful to the Labor Court of Milan for making the data available and we acknowledge financial support from MIUR PRIN 2009 project 2009MAATFS_001. E-mail: decio.coviello@hec.ca (Coviello); andrea.ichino@eui.eu (Ichino); nicola@nicolapersico.com (Persico)

getting reports from the Blotter system that David supervises. This issue becomes an additional unexpected working sphere that David will have to attend to this day. He calls Steven to find out more about the problem. After talking to him, he phones Phil, a developer in his team to explain the problem and explore some solutions. While talking to Phil, David is interrupted by the sudden presence of his boss Marti and Andrew who come with a question about the official holidays for the office in Munich, Germany. David was involved with Munich's operations earlier, but this working sphere is now peripheral for him as they only seek his opinion. At 9:03, David politely stops the conversation and leaves for his SIGMA meeting.

This excerpt from a time-use study¹ illustrates “just another day at the office”. While mundane and rather unremarkable, when seen from a managerial perspective it reveals something quite profound: 23 minutes of David's work time are split between three different tasks: prepare for a meeting; fix a software problem that has developed; and manage an issue concerning the office in Germany.

In Coviello, Persico, and Ichino (2014) we introduced and studied a theoretical model of a worker who, like David, juggles tasks—that is, divides his time between several projects all simultaneously requiring attention. Most people intuitively know that task juggling is prevalent in their workplace, and academic research confirms that this is so.² The purpose of this paper is to document empirically that task juggling exists; to show causally that juggling lowers productivity substantially; and that task juggling appears to have important consequences for efficiency when workers are pulled among different projects, working on all but completing few or even none.

We estimate the effect of task juggling on project duration by using an empirical specification derived from the theoretical model that we describe in Coviello et al. (2014). The data refer to the work schedules of some labor judges in Italy. We document that, just as David in his normal day at work, our judges juggle tasks and this habit causally lowers their productivity in a substantial way. In the judicial setting, of course, workers (judges) cannot be expected to work on a single case at a time because then they would be idle a lot of the time. But our theoretical model indicates that durations are minimized when the number of open cases is kept as small as possible—that is, as large as necessary to avoid idle periods but not larger. In particular, judges should take care not to open every case as soon as it is assigned, for this will steadily expand the number of cases they juggle. Instead, durations are minimized if judges “queue up” cases that are just assigned and then progressively draw from the queue as the prior backlog clears up. If judges follow this prescription the effect of a newly assigned case on the completion hazard of all other cases should be nil. If the judge does not follow this prescription and juggles, then the arrival of newly assigned cases impels the judge to open new cases (and/or change effort) and that has an externality on the completion hazard of the other cases.

To frame our empirical analysis we use a hazard model for the duration of trials as a function of the number of open cases and of effort. We remove the main source

1. Gonzalez and Mark (2005, p 148).

2. The seminal paper in the sociology literature is Perlow (1999).

of potential confounding variation including judge fixed effects in our specification. In this way we can control for any judge-specific time-invariant characteristic, concerning for example planning skills or effort costs, which might affect the judges' propensity to task juggle or their workplace habits. However, this may not be enough to identify causal effects because there can still be judge-specific but time-varying confounding factors, like health problems or good/bad mood periods. Events of this kind might generate a correlation between the two endogenous variables (task juggling, effort) and some unobservable components in the error term, even controlling for judge-specific fixed effects.

In order to address this issue we construct time-varying instruments for effort and task juggling based on the sample realization of the lottery that allocates the amount and the typology of workload to each judge. This lottery ensures that future cases of different types are randomly assigned to each judge and thus generates exogenous sources of variation which are precisely useful to study the question addressed in this paper. As for exclusion restrictions, it can be argued that they are satisfied by instruments constructed using future cases under the following identifying assumption: their arrival must affect the completion hazard of a case assigned today only if the judge begins working on them while the current case has still to be completed and/or if she changes her effort under the pressure of the new additional workload. In other words, if they affect completion hazards of previous cases only through task juggling (i.e., the opening of new cases that should be optimally queued up instead) and/or changes in the number of hearings per unit of time, which are our two endogenous variables. As in the case of any IV strategy, one could say that this assumption is not always satisfied and that also these instruments do not solve the identification problem if future workloads of different types affect the duration of a current case through other channels, like for example the time dedicated to studying cases. Given the pros and cons characterizing the two identification strategies (fixed effects and IV fixed effects), in the paper we present both sets of estimates, that define a range of plausible results obtained with different identifying assumptions.

Three features of our environment are key to our estimation strategy. First, our workers (judges) operate essentially as single units: there is no team work involved in the production of their judicial decisions.³ Secondly, we leverage the random assignment of cases to judges as a source of exogenous variation in the number and complexity of cases, the effects of which can be traced on the duration of cases. Finally, we are able to measure productivity, effort, ability, and difficulty of tasks quite accurately.

Results strongly support the hypothesis that judges respond to an increase in *future* caseload by juggling more tasks, and that in this way they exacerbate the negative effect of the caseload increase on the hazard of case completion and on the duration of all cases. At the sample mean, if a larger future caseload induces judges to increase

3. One could argue that the lawyers are part of a team with the judge. However, the reality in our empirical setting is that judges have considerable authority over lawyers in limiting their possibility to slow down the trial. The constraint on completion time is judicial time, not lawyer time. Therefore the judge is to be considered as a single worker as regards completion time.

task juggling by 1% (about two additional active cases on their desk), the completion hazard would decline approximately by a factor ranging between 1% and 2.1% which would correspond to a reduction of average duration of about 3 to 6 days (under some reasonable distributional assumptions) given an average duration of trials of 280 days. To keep completion hazards and durations constant, this increase in task juggling would need to be compensated by about 0.43 to 0.52 additional hearings per week, which would be an increase of effort ranging between 1.1% and 1.4%, given a sample average of 38 hearings per week.

We should point out that these estimates incorporate not only the mechanical consequences of task juggling which are at the core of the attention in this paper, but also the *disruption cost* of interruptions induced by task juggling, measurable in terms of additional time to reorient back to an interrupted task after the interruption is handled. A large management literature, surveyed by Mark, Gudith, and Klocke (2008) emphasizes the importance of these effects, that we cannot identify separately in our analysis, but that are likely to be potentially relevant and will be the focus of our attention in future research.

We believe these estimates are the first empirical estimates of the impact of time allocation on productivity. We also discuss whether the productivity slowdown can be thought of as inefficient behavior on the part of the judge from a private or a social viewpoint. We conclude that while there are no conclusive reasons to argue that task juggling is privately inefficient,⁴ it is undeniable that it generates a social inefficiency and we provide estimates of its size.

In addition to their general message on the effect of task juggling in any kind of work setting, these results have important implications for the Italian court system. Italy ranks 180th out of 196 countries surveyed in the 2013 “Doing Business” report of the World Bank in terms of expected duration of a commercial lawsuit (1,185 days). Not surprisingly, judicial productivity is frequently identified as a key constraint to economic growth in Italy (Esposito, Lanau, and Pompe 2014). Our estimates suggest that large productivity gains are potentially achievable simply by training judges in managing their case flow and for this reason they have been presented in several judicial conferences. As a result, some courts have attempted to rearrange their work schedules to minimize task juggling. Moreover, a software program has been created which attempts to facilitate the avoidance of task juggling and is currently being tested in three Italian tribunals. In this sense, our work has already had an impact.

A comment on our measure of productivity. We focus on completion hazards and on the duration of projects, instead of the number of completed cases, for two reasons. First, in many practical cases duration is what the worker’s principals (clients) want to minimize.⁵ Second, in our empirical application, reducing the duration of trials is a

4. This issue is reminiscent of an old debate in economics, about X-inefficiency, for which a summary can be found in Frantz (1992).

5. Many workers do not directly control the input in their productive process (such as when projects are assigned by the principal or by clients), but can control the speed at which their projects are completed. The latter tends to be especially true for workers who are not part of an assembly line. In these cases

key statutory objective.⁶ Duration of job completion is clearly not the only dimension of output: quality matters as well. We will show that a lower duration of trials is associated, if anything, with reductions in the probability that the judge's decision is appealed. Thus task juggling does not seem to generate any relevant trade-off between quantity and quality for these workers.

This paper fits broadly within the literature on the construction and estimation of production functions that can be traced back to the path-breaking paper of Cobb and Douglas (1928).⁷ Our goal is indeed to study and estimate the return to a factor of production but the focus is on individual (not firm) output. From this viewpoint, our results are more closely related to a recent literature initiated by Ichniowski, Shaw, and Prennushi (1997), suggesting that, in different areas of human behavior, individual modes of time use and activity scheduling are associated, in some cases causally, with performance for given effort.⁸ Thanks to the accurate measurement of the steps of "production", and to the access to exogenous quasi-experimental variation, in this paper we are able to identify more tightly than in this literature the causal effect on productivity of a specific and well-defined individual work practice—namely, task juggling.

What we call task juggling is an inefficiency that is also related to the concept of "bottlenecks" in the literatures on project management and project planning (see Moder, Phillips, and Davis 1983) and to the literature on network queuing, originating with Jackson (1963). We differ from the queuing literature in two ways. First, the queuing literature studies processes that are not explosive, meaning that a fraction of the time the queue is zero and the processor (the worker) is idle; this is not the case in our model, nor in our data. Second, the queuing literature is prescriptive: in our setting it would prescribe to eliminate task juggling but the literature does not offer an empirical methodology to quantitatively evaluate its impact.

it is speed, for given quality, which is the relevant performance measure. For example, an IT consultant does not control the number of customers who need her services; when there is excess demand, increased productivity can only be achieved by reducing the duration of each job, from assignment to completion. In a different setting, whenever a contractor is hired, the principal (homeowner) cares about the speed of completion, for given quality.

6. The Italian Constitution (art. 111) reads: "The law shall ensure the reasonable duration [of the trial]." And in (CSM 2010, p. 9), the Commission for the Setting of Standards in the Adjudication Process writes: "It is clear that, owing to the fundamental value attributed by the Constitution to the duration of trials, [...] a nationally-constructed index of duration must, sooner or later, become the standard measure of adjudication." At the European level there is a permanent Commission for the Efficiency of Justice (CEPEJ, see <http://www.coe.int/t/dghl/cooperation/cepej>) which is mainly focused on the duration of trials. At the global level, the "Doing Business" reports by the World Bank are concerned with the speed of dispute resolution.

7. Jorgenson (1986) surveys extensively the origins of this literature.

8. See, for example, Bertrand and Schoar (2003), Bloom and Van Reenan (2007), Bloom et al. (2009), and Bandiera et al. (2009) for CEO practices, Ameriks, Andrew, and Leahy (2003) and Lusardi and Mitchell (2008) for family financial planning and, closer to us, Aral, Brynjolfsson, and Van Alstyneet (2007) for multitasking activities and the productivity of single workers, and Garicano and Heaton (2010) for organization and productivity in the public sector. See also the recent surveys of Gibbons and Robert (2010) and DellaVigna (2009), and DellaVigna and Malmendier (2004, 2006) specifically on the issue of self-control in individual behavior.

Task juggling is also related to the sociological/management literature on time use.⁹ This literature shows how frequent are working situations in which many projects are carried along at a parallel pace. Related to it is also the literature on the *disruption cost* of interruptions, surveyed by Mark, Gudith, and Klocke (2008). These literatures do not trace empirically the effect of task juggling on output, perhaps because individual output measures are hard to obtain in many work environments and also, presumably, because establishing a causal channel is challenging outside of an experimental setting. At a more popular level, there is a large time management culture which focuses on the dynamics of distraction and on “getting things done” (see, e.g., Covey 1989 and Allen 2001).¹⁰ The success of these popular books suggests that people do indeed find it difficult to schedule tasks efficiently in the workplace.¹¹

We present the data and the institutional framework in Section 2. Section 3 provides descriptive evidence on the correlation between the productivity of judges and their effort, their ability and their propensity to juggle tasks. Analogous descriptive evidence is reported also for the quality of judges’ decisions, measured by appeal rates, and their potential determinants. In Section 4 we discuss the theoretical model that guides our econometric analysis, while Section 5 describes the econometric specification and the identification of the equations that we estimate. Results are presented in the same section, while Section 6 reports results showing that task juggling does not causally affect the quality of decisions at least as it is measured by the probability of an appeal. Section 7 estimates the size of the social consequences of task juggling. Section 8 concludes.

2. The Data and the Institutional Setting

We use data from one Italian court specialized in labor controversies for the industrial area of Milan. Our data set contains all the 50,412 cases filed between 1 January 2000 and 31 December 2005 assigned to 21 full-time judges of this court, who have been in service for at least one quarter during the period of observation. These judges are not involved in other tasks inside the tribunal and do not deal with trials of other kinds; their entire working time is dedicated to labor controversies—that is, disputes between a firm and one or more workers concerning wages, mobility inside the firm, working conditions, pension and illness rights, firing and layoff, workers’ misconducts, hiring procedures, discrimination as well as other minor issues. For this work, these judges, who must have obtained a law degree and must have passed a very selective examination on all subjects and procedural rules in law, are paid a fixed

9. See Perlow (1999) and Gonzalez and Mark (2005) for examples and a review of the literature.

10. Mullainathan and Shafir (2013, p. 38) mention task juggling in connection to the phenomenon of “tunneling”—that is, focusing on a specific aspect of a decision process and ignoring other relevant considerations.

11. For a review of the academic literature on this subject see Bellotti et al. (2004). For a speculartake on prioritization of tasks see the discussion of the “firefighting” phenomenon in Bohn (2000) and Reppenning (2001).

wage that increases with seniority. During their career they may be transferred, with or without grade promotion, from less to more important courts and functions, based on evaluations of their performance which include, among other indicators, also the average duration of the trials that they handle. As already mentioned in the Introduction, the abnormal long duration of trials is constantly at the center of the attention of the political and public opinion debate in Italy. Thus, judges do feel the pressure generated by society and by their own career concerns for a reduction of the average duration of trials, even if no explicit incentive scheme links their compensation to this objective.

We observe the complete history of all the cases from filing to disposition, which usually takes place in one of two main forms: a sentence by the judge (42%) or a settlement between parties (35%). The residual category of less relevant types of conclusion includes situations in which one party withdraws its claim, the trial cannot be decided by the judge for factual or procedural reasons that become evident after filing, or in which other kinds of exceptional circumstances emerge. Settlements, which imply durations that are on average 60% shorter than those of sentenced cases, are outcomes that judges try to induce in various ways to reduce the length of trials and to avoid having to issue a sentence. One of these ways is the calendarization of hearings, in particular of the first one. This feature of the process has implications for task juggling, to which we will return in the next section. Lawyers, particularly if defendants, can try to implement delaying practices but the judge is in control of the process and has the power to neutralize some of these actions. Given the random assignment of cases to judges that will be described and tested in what follows, there is no reason to expect that, on average, there should be a systematic correspondence between judges and the lawyers they face, particularly in a large city like Milan. Each judge is monocratically responsible for the trials assigned to him or her. No jury or other judges are involved.

Table 1 describes the unbalanced panel of judges that we study. For the judges who were already in service on 1 January 2000, we have information also on the cases that were assigned to them in the previous year and we can therefore compute a measure of their backlog at the beginning of the period under study. For the judges who took service during the period of observation (or less than one year before 1 January 2000) we analyze their productivity starting from one year after they join the court, in order to give them time to settle in. All the cases assigned to these judges during the first year of service (including those that were transferred to them from previous judges who left for another office or retired) are nevertheless counted to compute their backlog at the beginning of the second year of service in which we start to analyze their productivity. Thus all the judges that we analyze have at least one year of tenure and for each one we know the backlog of not-yet-disposed cases at the beginning of the period of observation.¹²

12. If a judge retires or is transferred to a different court (for whatever reasons) his/her cases are either all assigned to a new judge (thereby going in the initial backlog of the substitute) or they are distributed randomly to all the other judges in the court. We will later discuss the implications of these events for the econometric analysis.

TABLE 1. The panel structure.

Judge	Number of weeks of service per year						Total number of weeks of service (7)	% weeks of assignment (8)	Average number of new cases per week (9)	Standard deviation of new cases per week (10)
	2000 (1)	2001 (2)	2002 (3)	2003 (4)	2004 (5)	2005 (6)				
1	52	52	52	52	52	8	268	98	8.1	3.6
2	50	52	30	.	.	.	132	97	7.4	3.8
3	52	52	52	50	52	52	310	99	11	4.5
4	52	52	52	52	52	12	272	98	9.6	4.5
5	52	52	50	46	50	31	281	99	9.3	3.8
6	50	32	52	52	49	17	252	94	8.5	5
7	52	52	28	1	.	.	133	95	7.6	4.3
8	49	52	52	44	.	.	197	98	8.6	4.3
9	52	52	50	52	52	52	310	98	11	5.3
10	52	52	52	46	52	52	306	99	10	4.2
11	50	52	52	52	52	36	294	99	9	4
12	52	49	52	47	.	.	200	98	9	4.7
13	50	52	52	52	52	12	270	99	9.6	4
14	.	.	.	14	52	52	118	96	11	4.8
15	50	52	48	49	44	47	290	98	9.6	4.4
16	52	52	52	52	52	52	312	99	11	4.6
17	52	50	52	52	52	50	308	99	9.4	3.8
18	50	52	50	52	52	52	308	99	11	4
19	52	52	52	52	52	48	308	99	11	5
20	.	.	8	51	52	52	163	100	9.9	5.5
21	.	.	40	49	48	52	189	100	9.2	3.9
Total	921	911	928	917	867	677	5,221	98	9.7	4.5

Notes: Columns (1)–(6) report, for each judge, the number of weeks of service in each year. A week of service is a week in which the judge may receive cases. The number of yearly weeks of service is not 52 in the following situations: when we start observing the judge during the year (leaving out the first year of tenure that we do not consider, in order to let the judge settle in), when the judge leaves the court during the year or when he/she has a period of justified absence (e.g., sickness; maternity leave) of at least 1 week. Column (7) reports the total number of weeks of service for each judge over the period of observation while column (8) shows the fraction of total weeks of service in which the judge receives at least one case. Because of random assignment in some weeks of service judges may not receive cases. The last two columns report the average weekly workload (new assigned cases per week) and its standard deviation.

Columns (1)–(6) of Table 1 show, for each judge, the weeks of service in each year. Weeks of service are those in which the judge must be available to receive new cases. In Italy, as in other countries, the law (Art. 25 of the Constitution) requires that judges receive a randomly assigned portfolio of new cases.¹³ Our econometric strategy crucially relies on this random assignment, which is designed to ensure the absence of any relationship between the identity of judges and the characteristics of the cases assigned to them, including the identity of lawyers. There is a strict social control on the enforcement of random assignment in order to keep the heated industrial relations political debate out of the court as much as possible. As a result, judges are not allowed to specialize in particular types of labor controversies and cannot trade trials. Moreover, they are not allowed to render themselves unavailable for assignments, unless they are sick for long periods (more than 1 week). In the rare cases in which these events occur, assignments are typically diverted in a random way to other judges. To reduce the possibility of strategic gaming of the random assignment, cases are assigned also during periods of vacation. As a result of these rules, the yearly weeks of service in columns (2)–(7) of Table 1 are not 52 only in the following situations: when we start observing the judge during the year (leaving out the first year of tenure that we do not consider, in order to let the judge settle in), when the judge leaves the court during the year or when he/she has a period of justified absence (e.g., sickness; maternity leave) of at least 1 week.¹⁴ Column (7) reports the total number of weeks of service for each judge over the period of observation, while column (8) shows the fraction of total weeks of service in which the judge receives at least one case. There are in total 5,221 judge-weeks of service and in 81 of these some judges do not receive cases. It is evident from the table that these weeks of “no assignment while in service” are distributed randomly across judges and a formal test cannot reject this hypothesis with a p -value of 0.5375.¹⁵

Weeks of no assignment are possible, although rare, because of the way in which random assignment is specifically implemented in the court that we analyze. Every morning the judges in service are ordered alphabetically starting from a randomly extracted letter of the alphabet. The cases filed during the day are then assigned in alphabetic sequence to the judges in service. For example, if in a given day the extracted letter is B and five cases are filed, only judges with a name starting from B to F will receive an assignment on that day (assuming one judge per letter of the alphabet). Therefore, within each week judges may receive slightly different workloads in terms of size. The last two columns of the table report the number of cases assigned to each judge per week on average and the corresponding standard deviations, which are both very similar across judges and close to the overall weekly statistics of 9.7 cases

13. Other studies have exploited the random assignment of cases to judges for identification: for example Ashenfelter, Eisenberg, and Schwab (1995) and Kling (2006).

14. Evidently, the Italian legislator assumes that long sickness and maternity are events that cannot be manipulated to game the assignment of cases to judges.

15. We test this hypothesis using a linear regression for the probability that a judge receives at least a case in a week as a function of judges' fixed effects.

TABLE 2. Tests for the random assignment of cases to judges.

	Rejections at 5% significance (1)	Fraction of rejections at 5% significance (2)	Corrected significance (3)	Rejections at corrected significance (4)	Fraction of rejections at corrected significance (5)	<i>N</i> (6)
Type controversy	26	0.083	0.0008	5	0.016	312
Red code case	12	0.038	0.00016	1	0.0032	312
Firing case	18	0.058	0.00038	2	0.0064	312
Zip code of plaintiff's lawyer	26	0.083	0.00096	6	0.019	312
Number of involved parties	22	0.071	0.00081	5	0.016	312
Overall	116	0.062	0.00047	17	0.0091	1,872

Notes: The table summarizes the evidence on the weekly random assignment of cases to judges, based on Chi-square tests of independence between the identity of judges and five discrete characteristics of cases: *type of controversy* in 14 categories; a dichotomous aggregation of the types of controversy in *red code* versus *green code* cases, by analogy with what happens in a hospital emergency room, where red code cases are those that, according to judges, are urgent and/or complicated, thus requiring immediate action and/or greater effort; a dummy for *firing cases*; *zip code* of the plaintiff's lawyer (55 codes); the "number of involved parties" (capped at 10). The last row, *Overall*, presents joint results for all variables and all weeks. "Rejections at 5% significance" are the numbers of tests in which *p*-values are below 0.05. Corrected significance levels are computed with the Benjamini and Hochberg (1995) multiple testing procedure. *Rejections at corrected significance* are the numbers of tests in which *p*-values are below the corrected significance levels.

per week with a standard deviation of 4.5. Small sample variability in the number of cases given to each judge is therefore possible, but this small sample variability is not systematic and fades away over the long run.

For the same reason, also the characteristics of the assigned portfolios of cases may occasionally differ across judges within a week, but in a way that is random and not systematic. Table 2 reports the summary results of Chi-square tests of independence between the identity of judges and six characteristics of cases, conditioning on weekly assignment: type of controversy in 14 categories; an aggregation of the type of controversy in two categories ("red code" versus "green code" cases)¹⁶ by analogy with what happens in a hospital emergency room, where red code cases are those that, according to judges, are urgent and/or complicated, thus requiring immediate action and/or greater effort; a dummy for firing cases; zip code of the plaintiff's lawyer (55 codes); the number of involved parties (capped at 10). The last row, presents joint results for all variables and all weeks. The first column reports the numbers of weeks in which independence is rejected at the 5% level out of the 312 weeks on which the test is conducted. The corresponding fraction of rejections is in the second column. Since 5% is not the correct significance level in a context of multiple testing, in the third column

16. This dichotomous characterization will be used later in the econometric analysis and therefore it is separately considered here.

we report the significance levels corrected with the Benjamini and Hochberg (1995) method. When these correct significance levels are used, the number of rejections declines considerably as shown in the remaining part of the table. Therefore, we can conclude that, within each week, differences in assignments are due only to small sample variability and are not systematic: in the long run, judges, receive qualitatively and quantitatively similar portfolios of controversies.

For the purpose of identification of the causal effects of interest these are attractive and convenient features of our data that compensate for the unfortunate fact that we have no information of any kind concerning the judges under study, not even age and gender. Differently than in other data sets, which typically have some demographic characteristics but do not contain measures of ability and effort, we instead observe the entire history of all the cases assigned to each judge. With this information we can construct, as we will see in the next section, very precise time-varying measures of task juggling, productivity, work scheduling, ability, and effort for each judge.

3. Descriptive Evidence

In this section, we compare judges on the basis of average indicators of task juggling and performance per week, computed over all the weeks in which each judge is observed in service.

3.1. Task Juggling

We use the term “task juggling” to characterize any work scheduling practice as a result of which the tasks of one job are intertwined with the tasks of different jobs assigned to the same worker. Table 3 compares possible measures of task juggling for the 21 judges considered in this study, treating each trial as a job and hearings as tasks. Table 4 shows the correlations between these measures.

In Table 3, judges are ordered on the basis of their degree of task juggling as measured by the indicator reported in the first column. This indicator is the average number of *active cases* on the desk of a judge, computed over the weeks in which she is observed. Formally, here and in the rest of this paper, a case is defined as *active* at a given week if its first hearing has taken place before the end of that week but the case has not been completed yet by the same date. Of course we do not know the exact moment in which a judge starts working on cases previously assigned to her, but it seems reasonable to consider the first hearing as a good approximation of this moment. If a judge does not juggle tasks, her number of active cases must be 1. Numbers larger than 1 indicate that the judge engages in task juggling, because she has scheduled hearings of different cases between the first and the last hearings of other cases. The table shows that there is a considerable variability in the degree of task juggling across judges, even if they receive the same workload in terms of quality and quantity because of random assignment and full-time working. Judge 11 is working on average on 220

TABLE 3. Measures of task juggling.

Judge	Active cases (1)	Active cases at second hearing (2)	Other hearings between first and second hearing of a case (3)	Other non-first hearings between second and third hearings of a case (4)	Distance in weeks between first and second hearings of a case (5)	Distance in weeks between second and third hearings of a case (6)
11	220	147	384	269	11	11
17	217	143	371	284	13	16
15	217	147	405	288	12	13
13	213	145	372	279	11	12
21	203	137	371	266	13	13
10	202	135	358	263	10	11
18	197	107	429	252	14	12
1	192	120	344	252	13	15
20	187	125	407	309	11	13
3	180	116	342	247	8	8
4	169	112	308	221	8	9
19	164	109	292	215	8	8
16	159	102	295	210	8	8
5	152	110	260	206	7	8
8	134	86	261	189	7	8
14	123	58	331	206	13	13
12	122	80	232	165	6	6
7	115	70	239	159	8	8
9	89	48	195	128	6	7
2	83	50	185	135	6	7
6	70	38	163	94	7	7

Notes: The table reports, for each judge, the averages of different weekly measures of task juggling, computed over the weeks of service in which the judge is observed. *Active cases* are cases that, at the end of a week, have seen already a first hearing but are not completed yet. *Active cases at second hearing* are cases that, at the end of a week, have seen already two hearings but are not completed yet. The other measures of task juggling are described by the columns' headings.

trials in a given week, while judge 6 has only 70 cases contemporaneously open on her desk.

As mentioned in the previous section, judges may want to hold the first hearing of a case as early as possible, to explore the possibility that parties settle, and then delay to a future date the continuation of the trial if settlement does not take place. If this were the situation, judges would have many cases defined as active with respect to the first hearing, even if they refrain from task juggling when dealing with subsequent hearings. In other word, they would hold all the hearings of a trial, from the second onwards, without scheduling in the middle of them hearings of other trials, unless they are first hearings. A judge behaving in this way would have, at every moment in time, only one active case defined with respect to the second hearing, even in the presence of more

TABLE 4. Correlation between alternative measures of task juggling.

Active cases	Hearing	Active cases at second	Other hearings between first and second hearing of a case	Other non-first hearings between second and third hearings of a case	Distance in weeks between first and second hearings of a case	Distance in weeks between second and third hearings of a case
Active cases	1.00					
Active cases at second hearing	0.98 (0.001)	1.00				
Other hearings between first and second hearings of a case	0.92 (0.001)	0.84 (0.001)	1.00			
Other non-first hearings between second and third hearings of a case	0.96 (0.001)	0.93 (0.001)	0.96 (0.001)	1.00		
Distance in weeks between first and second hearings of a case	0.73 (0.001)	0.60 (0.001)	0.84 (0.001)	0.75 (0.001)	1.00	
Distance in weeks between second and third hearings of a case	0.73 (0.001)	0.63 (0.001)	0.80 (0.001)	0.77 (0.001)	0.97 (0.001)	1.00

Notes: The table reports the correlations (and, in parentheses, the *p-values* of the test that they are equal to zero) between alternative measures of task juggling. *Active cases* are cases that, at the end of a week, have seen already a first hearing but are not completed yet. *Active cases at second hearing* are cases that, at the end of a week, have seen already two hearings but are not completed yet. The other measures of task juggling are described by the column and row headings.

active cases defined with respect to the first hearing. The second column of Table 3 reports the number of *active cases at the second hearing* and it is evident that judges engage in a considerable amount of task juggling even after the second hearing and independently of the attempt to induce settlement at the first hearing. Moreover, the correlation between the two indicators of active cases is 0.98 (see Table 4), suggesting that judges who juggle first hearings, juggle also all hearings and vice versa, while settlement considerations should affect only the juggling of first hearings.

An alternative way to measure task juggling is to count how many hearings of different cases are held on average by a judge between the first and the second hearing of any given case. For a fully non-juggling judge this number should be zero, while column (3) of Table 3 shows that it actually ranges between 163 and 429. The correlation between this indicator and the number of active cases is .92 (see Table 4). A similar indicator can also be constructed for task juggling after the second hearing, by counting how many non-first hearings are held on average by the judge between the second and the third hearing of any given case. This indicator is reported in the fourth column of Table 3 and ranges between 94 and 309, confirming that task juggling extends over the entire life of cases. Also this indicator is highly correlated with all the others suggesting that juggling is independent of first hearings.

Finally, similar evidence can be derived from a third way to construct task juggling indicators, based on the average distance in weeks between subsequent hearings of cases. The last two columns of Table 3 show that the intervals between first and second hearings and between second and third hearings range on average between 6 and 14–16 weeks and the correlation between the two measures is 0.97. The correlation between these distance measures and the other indicators is lower, but still quite high (see Table 4).

To conclude, task juggling can be measured in different ways that are highly correlated across judges. More importantly, judges who juggle more first hearings seem to juggle any hearing, as if juggling were unrelated to the need of scheduling first hearings to induce a settlement between parties. The number of active cases is the measure of task juggling that we prefer to use in the analysis that follows, but, given the high correlation with other measures, results do not change qualitatively using them instead.¹⁷

3.2. Total Duration and Active Cases

The height of circles (marked by the judge id number) on the vertical axis of the top left panel of Figure 1 measures the total duration of cases assigned to each judge. Total duration is defined as the number of weeks from filing up to the week in which a sentence is deposited by the judge, or the case is settled, or other less relevant outcomes determine the end of the trial (see Section 2). On the horizontal axis judges are ordered from the slowest one on the left to the fastest one on the right. The height of the

17. We show it for active cases defined at the second hearing in the Online Appendix, downloadable from www.andreaichino.it.

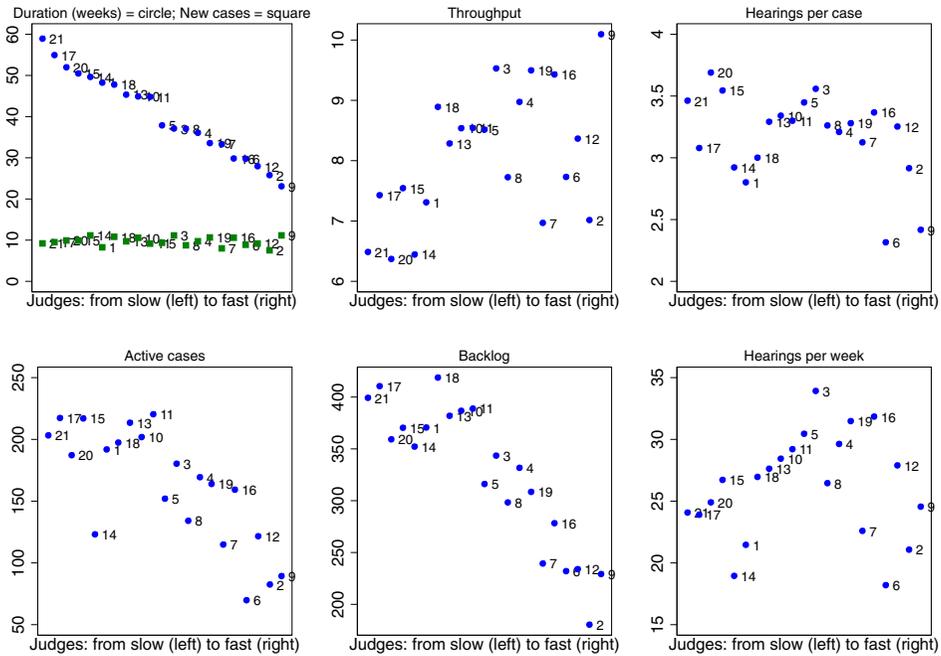


FIGURE 1. Differences of performance between judges with randomly assigned workload. For each judge, the figure displays averages of different performance indicators computed over the weeks of service. *Duration (weeks)* is computed from filing until a sentence, a settlement, or another type of case conclusion occurs. *New cases* assigned on average per week are as in column (8) of Table 1. *Active cases* are cases that have already seen a first hearing before the end of a given week but are not completed yet by that date. *Throughput* is the number of cases completed by a judge during a given week. *Backlog* is the sum of all the cases historically assigned to a judge but not yet completed by the end of a given week. *Hearings per case* is the number of hearings that on average are needed by a judge to close a case. *Hearings per week* is the number of hearings that a judge holds on average in a week.

squares in the same panel indicates the workload of new cases per week assigned to each judge on average. This graphic representation makes transparent the heterogeneity of performance, in terms of duration of trials, observed for these judges despite the fact that they receive a workload which is fairly similar in quantity (because we selected only judges who receive a full workload) and quality (because of random assignment).

The bottom left panel in the same figure plots the number of *active cases* ordering again the judges from the slowest one on the left to the fastest one on the right. The vertical comparison between the left panels of the figure highlights the high and statistically significant correlation across judges (0.77; p -value < 0.0001) between the average number of active cases and the average duration of trials. It is important to keep in mind that these differences emerge among judges of the same office, who work in exactly the same conditions, with the same secretarial assistance, and with a similar workload in terms of quantity and quality.

3.3. *Throughput and Backlog*

For the reasons explained in the Introduction, we prefer to use duration as opposed to throughput, defined as the number of completed cases per week, as a measure of productivity. The total cumulative throughput of these judges can only be equal to the input they receive, in terms of cases exogenously assigned to them. In principle, two judges may be deciding the same number of cases in a given quarter, but for one of them these cases may have been assigned just recently, while for the other they may be very old cases. What matters, really, is how long it takes to process the input. Nevertheless, in the top central panel of Figure 1 we show that if keeping too many cases opened at the same time slows down the activity of judges, also the number of cases he/she will be able to close per week is negatively affected on average. The correlation across judges between throughput and the duration of cases is negative although not strongly significant (-0.52 ; p -value = 0.015).

In line with these considerations, it is not surprising to infer, from the bottom central panel of Figure 1, that the fastest judges, with fewer active cases, have on average a lower backlog, defined as the sum of all the cases historically assigned to a judge up to the end of a given week and that are not completed yet by the same date. The correlation between the backlog and the number of active cases is high and significant (0.91; p -value < 0.0001).

3.4. *Complication of Cases, Ability and Effort of Judges*

Although suggestive, our hypothesis concerning the role of task juggling on the productivity of judges must be confronted with other more obvious and potentially relevant determinants of this performance: in particular, ability and effort.

Consider the average number of hearings that a judge needs to close a case. Without random assignment this statistic would depend on both the difficulty of the cases assigned to a judge and on her ability to handle them quickly. But given random assignment, the complication of controversies that judges face should be fairly similar, up to small random differences determined by the realization of the assignment procedure described in Section 2. Therefore, differences across judges in the average number of hearings to close a case should mostly capture the unobservable skills that determine how a judge can control the trial and the behavior of parties, lawyers, and witnesses, in order to reach quickly a decision.

This statistic is plotted in the top right panel of Figure 1, where judges are again ordered, on the horizontal axis, from the slowest one on the left to the fastest one on the right. Here we do not appreciate a correlation with duration as strong as the one observed for active cases (bottom left panel of the figure). The correlation between duration and the number of hearings per case is positive (0.41) but relatively low and less statistically significant (p -value = 0.07). Inasmuch as being able to decide a case with fewer hearings is a form of ability of a judge, this descriptive evidence does not

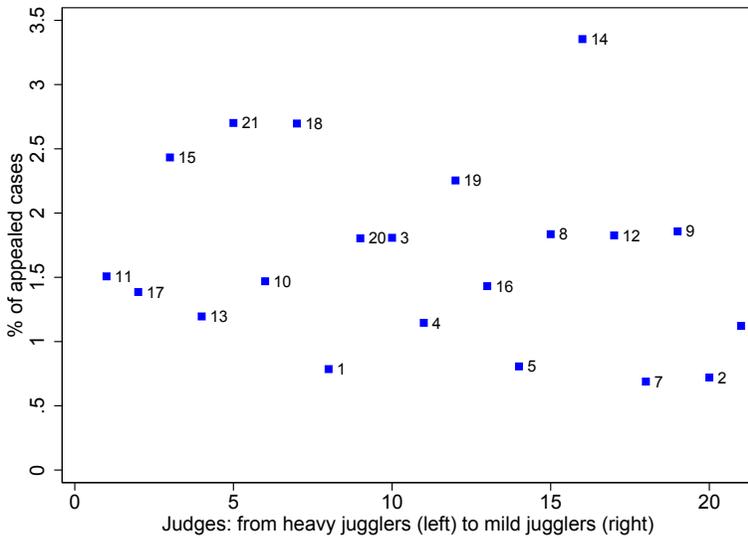


FIGURE 2. The trade-off between quantity and quality in the decision of judges with randomly assigned workload. For each judge, the figure reports the percent fraction of cases for which an appeal has been filed at a higher-level court.

suggest that such characteristics can alone explain the heterogeneity of the average duration of trials across judges.

A measure of effort is instead offered in our data by the number of hearings held by a judge in a unit of time. The idea is that, by exerting more effort, a judge can schedule more hearings per week and in this way can, *ceteris paribus*, improve her performance in terms of throughput and total duration of completed cases. Other dimensions of effort, like for example working at night on reading the documentation of cases, are not directly captured by this measure. However, note that hearings are the tasks that make a trial advance towards completion: working at night translates into *effective effort* only if it generates more hearings per week. Thus, *effective effort* is what this indicator is measuring in the bottom right panel of Figure 1. Also in this case we cannot infer an evident pattern connecting this measure of effort to performance in terms of duration: the correlation is negative (-0.12) but low and statistically insignificant (p -value = 0.60).

3.5. Is there a “Quantity versus Quality” Trade-off?

It seems important to consider the possibility of a “quantity versus quality” trade-off in the performance of judges. Could it be that judges who juggle more cases are worse judges in terms of quality of decisions? The evidence presented in Figure 2 suggests that the answer is no, as long as the percentage of appealed cases can be considered as a good measure of the quality of the judges’ decisions. We acknowledge that the appeal rate is at best an imperfect proxy for the quality of judicial decisions, since the

decision to appeal a sentence depends on a variety of factors including especially the type of trial in question. In our case the homogeneous nature of the trials (all labor litigations) and furthermore the random assignment of the cases across judges, alleviate these concerns. In using appeal rates we follow the bulk of the law and economics literature which, in the absence of more detailed information, regards appeal rates as a valid proxy for the quality of judicial decisions.¹⁸

There is no evidence that the cases assigned to heavy jugglers on the left of the figure have a lower probability of appeal than the cases assigned to mild jugglers on the right. If anything, the opposite seems to hold, given that the correlation between the average number of active cases and the percent of appealed cases across judges is positive (0.19) although statistically insignificant (p -value = 0.40). As for the effect of task juggling on completion hazards, in Section 6 we will probe further the descriptive evidence on appeal rates using an econometric model to explore the possibility that task juggling causally reduces them, thus originating a trade-off between quantity and quality.

To summarize, the descriptive evidence presented in this section suggests that task juggling, as opposed to sequential working, may reduce considerably the performance of judges in terms of total duration of the cases assigned to them, without apparent costs in terms of quality of decisions. Indicators of experience, ability, and effort are as well likely to be relevant determinants of performance, but in a possibly less significant way. To go beyond correlations we now move, in the next section, to the description of the theoretical framework that guides our multivariate econometric analysis.

4. A Theoretical Framework to Estimate the Inefficiency Caused by Task Juggling

In this section we briefly outline a theoretical model based on Coviello, Persico, and Ichino (2014) of a judge who juggles tasks. The model delivers comparative statics for the duration of each case assigned to the judge as a function of: the effort e put in by the judge (measured in number of hearings per unit of time); the complexity S of the cases that the judge deals with (measured in number of hearings necessary to adjudicate a case); the number of cases A_t the judge chooses to simultaneously work on at time t (which is a measure of the judge's juggling behavior); and, finally, the number of cases which are assigned to the judge.

A judge starts his work life with zero cases, and is thenceforth assigned cases at a constant rate α . The judge does not need to start working immediately on every case assigned to him. Rather, the judge queues up newly assigned cases and draws from that queue at rate ν . Once a case is drawn from the queue that case is processed in parallel with all other cases previously drawn from the queue and not yet completed.

18. See, for example, Appendix V in CEPEJ (2011) where the appeal rate is cited as an indicator of quality of judicial decisions for European courts. See also Mitsopoulos and Pelagidis (2010) for an in-depth analysis which takes the appeal rate as *prima facie* proxy of the quality of judicial decisions.

This modeling device allows us to model judges who choose to task juggle a lot (high ν) and those who choose to work on a few cases at a time (low ν). Each case requires an amount of effort S to complete and the judge chooses a constant effort rate e . The analysis in Coviello, Persico, and Ichino (2014) proves that, in this scenario, the judge adjudicates cases at a constant rate ω . If, moreover, e is small relative to α then Coviello, Persico, and Ichino (2014) prove that $\alpha > \omega$, namely that more cases are assigned in each time period than the judge is able to complete. This theoretical prediction is consistent with our data: the last row in column (9) of Table 1 shows that in the 5,221 weeks in which we observe our judges $\alpha = 9.7$, while from the central top panel of Figure 1 we find that $\omega = 8.2$; the number of cases assigned and not yet completed indeed grows over time. We shall therefore focus the theoretical analysis on the case $\alpha > \omega$.

Because the assignment and completion rates are constant it is straightforward to compute a case's duration. A case which is assigned to the judge at time t finds a mass of exactly

$$\alpha t - \omega t$$

cases previously assigned but not yet completed. Given an output rate of ω , these cases will take the judge exactly

$$D_t = (\alpha t - \omega t)/\omega = ((\alpha/\omega) - 1)t \quad (1)$$

units of time to work through. As soon as the time interval of length D_t has elapsed the mass of cases assigned prior to t will be completed, and in the next instant cases assigned at t will be completed too. Thus the formula for D_t yields the duration of a case assigned at time t .

We now need to substitute out the variable ω in order to express D_t in terms of variables that are chosen by the judge (effort and task juggling) in addition to the judge's workload. Proposition 1 in Coviello et al. (2014) proves that the completion rate ω is an increasing function of e/S (the more the judge chooses to work, and the fewer hearings it takes to complete a case, the more cases the judge completes per unit of time) and a decreasing function of ν , the rate at which the judge adds to the batch of cases that he simultaneously works on. Formally, we have

$$\omega = \Omega(e/S; \nu) \quad \text{with } \Omega_1 > 0 > \Omega_2.$$

The partial derivative Ω_2 is negative because a judge who opens more cases dilutes his effort among more cases and therefore ends up being slowed down.

Plugging the function Ω into equation (1) and taking partial derivatives yields the following predictions. The duration D_t (the time it takes to complete a case after it is assigned) is: increasing in α (the more cases the judge is assigned, the longer each case takes to complete on average); increasing in t (judges become overwhelmed over time—this implication follows from our parameter restriction $\alpha > \omega$); decreasing in e/S (the harder the judge works, and the fewer hearings are needed to complete a

case, the shorter the duration); and, finally, increasing in the rate ν at which the judge chooses to juggle.

In the next section the sign and magnitude of these effects will be estimated using a duration model. In the duration model the effect of the workload α will be proxied for by the total mass of cases assigned to that judge up to time t (which we call W_t and which corresponds to αt in the theoretical model). In addition, the rate ν at which the judge chooses to multitask will be proxied for by the amount of cases the judge has active at time t (corresponding to $A_t = (\nu - \omega) t$ in the theoretical model). Using A_t as a proxy for ν makes sense because if ν increases then ω decreases (recall that Ω_2 is negative) and so A_t increases.

5. The Effect of Task Juggling on the Hazard of Case Completion

The theoretical framework described so far shows that if a judge is induced by an exogenous pressure to increase task juggling, her hazard rate of case completion (i.e., the probability that her cases are completed exactly t weeks after filing given that they have survived until then) declines. In this section we present econometric evidence supporting this prediction.¹⁹

5.1. Non-Parametric Hazard Estimates

We frame the empirical analysis using “weeks from filing” as the unit of elapsed time and, in Figure 3, we begin presenting Kaplan–Meier estimates of the cumulative hazard of completing a case in week t for two group of judges that we characterize as “heavy” and “mild” jugglers in relative terms. The first group comprises those judges who keep at least 180 active cases on their desk (see Table 3), while the others are in the second group. At any week the cumulative hazard is significantly higher for the “mild” jugglers, in line with the prediction of our theoretical model.

5.2. Parametric Hazard Estimates

To probe the robustness of this evidence controlling for relevant covariates, we use a parametric hazard model with time-varying covariates. In our application we are not interested in the shape of the baseline hazard but only in how the hazard is shifted with

19. In a previous version of this paper, downloadable from www.andreaichino.it, the econometric analysis was based on the estimation of the conditional expectation function of the duration of cases using a regression analysis aimed at testing whether this conditional expectation was affected by task juggling. In that context, standard linear instrumental variable methods were used to identify and estimate causal effects, exploiting the exogenous source of variation generated by instruments constructed with the same logic that will be described here in Section 5.3. The absence of censoring in our data made that econometric approach possible and meaningful. It is reassuring that the duration analysis presented in this paper and the nonlinear control function approach that here we use to identify causal relationships, thanks to a suggestion of the editor, lead to qualitatively and quantitatively similar conclusions.

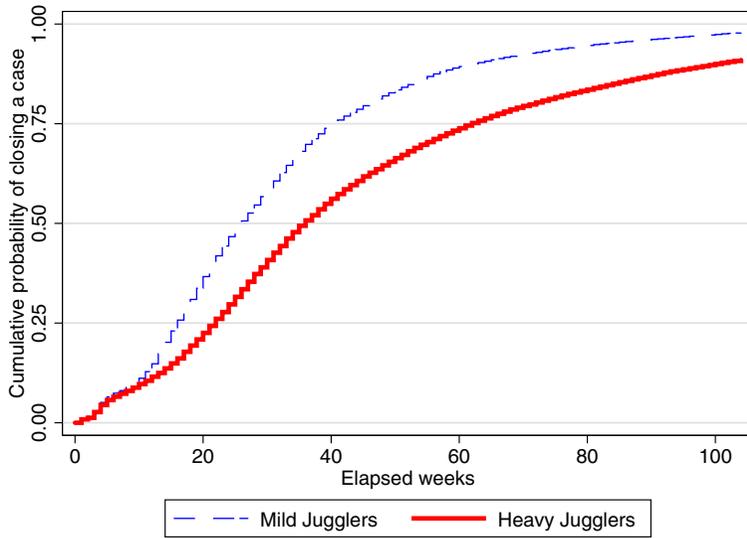


FIGURE 3. Kaplan–Meier estimate of the cumulative hazard of case completion for “heavy” and “mild” jugglers. The figure shows the cumulative distributions of the probability of closing a case in a given week conditional on survival up to that week (Kaplan–Meier failure function), splitting judges between relatively *Heavy jugglers* and *Mild jugglers*. The first group includes the 10 judges who juggle on average at least 180 active cases; the second group includes the remaining 11 judges (see Table 3).

respect to the baseline by different levels of task juggling. For this reason, we follow the partial-likelihood approach proposed by Cox (1972) and specify the hazard that case i , assigned to judge j , is completed in week t after filing as

$$\begin{aligned} h_{ijt} &= h_0(t) \exp(\beta_1 A_{ijt} + \beta_2 (e/S)_{ijt} + \beta_3 W_{ijt} + \beta_4 Z_{ijt} + \beta_5 T_t + \delta_j) \\ &= h_0(t) \lambda(X_{ijt}, \Psi), \end{aligned} \quad (2)$$

where $X_{it} = (A_{ijt}, (e/S)_{ijt}, W_{ijt}, Z_{ijt}, T_{ijt})$, $\Psi = (\beta_1, \beta_2, \beta_3; \beta_4; \beta_5; \delta_j)$, $h_0(t)$ is the baseline hazard, and $\lambda(X_{ijt}, \Psi)$ captures the deviations from the baseline hazard in which we are interested, due to the observables X_{ijt} . The A_{ijt} is the number of active cases on the desk of judge j at the end of week t of case i (i.e., task juggling).²⁰ The $(e/S)_{ijt}$ is the number of hearings held by judge j in week t of case i divided by the average number of hearings needed to close the cases for which a hearing was held in the same week (i.e., effort standardized by the difficulty of cases treated by the

20. A case is active at week t if it has seen a first hearing before the end of week t but it has not been completed yet by the same date. Evidence in the Online Appendix (downloadable from www.andreaichino.it) shows that results that use *active cases at the second hearing* as a measure of task juggling are qualitatively unchanged, which is not surprising given the high correlation with active cases discussed in Section 3.1.

judge).²¹ The W_{ijt} is the cumulated sum of cases assigned to judge j up to the end of week t of case i (i.e., total workload); T_t is a set of calendar weekly and yearly dummies to control for time effects, including seasonality, in the most flexible way. Z_{ijt} is a set of 14 dummies for the type of controversy.²² The δ_j are judges' fixed effects. It is important to keep in mind, here and in the rest of the econometric analysis, that the inclusion of judge-specific fixed effects in all our specifications should control for any time-invariant judge-specific characteristics. Descriptive statistics for these variables are displayed in Table 5.

The main focus of our analysis is on the parameter β_1 which measures the causal effect of task juggling on the completion hazard of cases assigned to judge j after t weeks from filing. Our theoretical framework states without ambiguity that this coefficient should be negative, because more task juggling should reduce the hazard of job completion. We predict instead a positive sign for β_2 , because more effort exertion for given difficulty of the cases to which that effort is applied (or the arrival of easier cases for given effort exerted by the judge) should increase (or at least not decrease) the hazard.²³ As for total workload, our model, as do many others, predicts not surprisingly that β_3 should be negative.

In Table 6, we present maximum likelihood estimates of the coefficients of the hazard specified in equation (2). In the first column, the specification does not include judge fixed effects while these effects are included in the second column. Independently of fixed effects, the estimated coefficients for the main variables of interest are statistically significant at standard significance levels and correspond to the prediction of our theoretical framework: task juggling has a negative effect on completion hazards while standardized effort has the opposite effect. This invariance with respect to the inclusion of judges' fixed effects suggests that the two main variables of interest are weakly correlated with unobservable time-invariant judge characteristics.²⁴

21. To deal with the potential confounding factor generated by the judge's choice of which hearings to hold and for which cases in the same week, we have also standardized effort using the average number of hearings needed to close all the cases held by all judges in the same week, and results are qualitatively unchanged. These results are reported in the Online Appendix (downloadable from www.andreaichino.it). An alternative way to solve the problem is the IV-fixed-effects strategy described in Section 5.3.

22. Because of random assignment of cases to judges, the inclusion or exclusion of these dummies does not affect the main coefficients of interest, as shown in the Online Appendix (downloadable from www.andreaichino.it).

23. Note that, by random assignment (see Section 2), within each unit of time all judges receive portfolios of cases that differ just because of random sampling. Therefore, if $S_{j\cdot} > S_{k\cdot}$, it must be either because judge j has randomly received a slightly more complex portfolio than judge k , or because the portfolio is effectively identical but judge j is "more able" in the sense that she can close the same portfolio of cases with fewer hearings on average than judge k . However, inasmuch as the ability of the judge is constant over time, it is captured by the fixed effects δ_j . Looking instead at the same judge across time, despite random assignment it can happen that $S_{j\cdot t} > S_{j\cdot \tau}$, with $t < \tau$, either because the ability of judge j increases over time or because the assigned cases become less difficult on average over time.

24. Only the effect of total assigned workload changes between the two columns, becoming negative, as predicted by the model, only when judge fixed effects are included. This is because, since the panel is

TABLE 5. Descriptive statistics.

	Mean	sd	p25	p50	p75	N
<i>Main outcome</i>						
Total duration in weeks (from filing to disposition)	40	34	19	31	51	50,412
<i>Regressors</i>						
Active cases	195	73	147	201	250	50,412
Standardized effort	8.2	4	5.7	8.1	10	50,412
Total workload	2133	904	1406	2052	2886	50,412
<i>Instruments</i>						
Red cases	10	7.3	5	10	15	50,412
Normal cases	34	18	26	39	46	50,412
<i>Components of standardized effort</i>						
Effort	38	17	27	39	50	50,412
Number of hearings for cases worked in a week	4.9	1.1	4.2	4.9	5.5	50,412
<i>Additional outcome</i>						
Appeal	0.017	0.13	0	0	0	50,412

Notes: Descriptive statistics for the variables used in the econometric analysis, concerning the 50,412 trials randomly assigned to 21 full-time judges of the Labor Court of Milan between 1 January 2000 and 31 December 2005. *Total duration in weeks* is computed from filing until a sentence, a settlement, or another type of case conclusion occurs. *Active cases* are cases that have already seen a first hearing before the end of a given week but are not completed yet by the same date. *Standardized effort* is the number of hearings held by a judge in a given week divided by the average number of hearings needed to close the cases for which a hearing was held in the same week; the numerator and the denominator of this ratio are the variables *effort* and *hearings needed to close cases* described in the bottom panel of the table. *Total workload* is the cumulated sum of cases assigned to a judge up to the end of a given week. As for the instruments, they are respectively defined as the numbers of *red code* and *green code* cases assigned to a judge in the five weeks that precede and include a given week (or the number of cases assigned in the first month of life of a case if the week is one of the first five of the life of the case). *Red code* cases are those that, according to a survey of judges and lawyers, are considered as urgent and/or complicated, thus requiring immediate action and/or greater effort. *Green code* cases are instead the remaining standard and simpler cases. *Appeal* is a dummy variable taking value 1 if the case is appealed to a higher court.

To get a sense of the economic significance of these results note that the estimated coefficients in columns (1), (2) and (5) of Table 6 can be interpreted as the effect of a one-unit change of each variable on the log hazard. Focusing on the fixed-effects estimates in the second column, suppose that a judge increased by 1% her level of task juggling. Since the average number of active cases at the end of a week is 195 (see Table 5), this means adding approximately two more cases to the ones already open on the judge's desk. This change of working schedule would imply a 1% decline in the completion hazard ($\approx \beta_1 \Delta A = (0.005 * 1.95) * 100$). To compensate for this hazard decline with more effort, the judge would have to hold approximately 0.43 additional hearings per week ($\approx (\beta_1 \Delta A)S/\beta_2 = (0.005 * 1.95) * 4.9/0.112$), which would represent a 1.1% increase of effort starting from the sample average of

unbalanced and random assignment is on a daily basis, judge fixed effects capture judge-specific differences in workload assignments for judges operating in non-overlapping periods.

TABLE 6. The effect of task juggling on the hazard of closing a case.

Dependent variable	$h(t X)$	$h(t X)$	Active cases	Standardized effort	$h(t X)$
Estimation method	ML	ML	OLS	OLS	ML-Control
Stage	Second	Second	First	First	Second
Variables	(1)	(2)	(3)	(4)	(5)
Active cases	-0.005 (0.0001)	-0.005 (0.0001)			-0.011 (0.004) [0.004]
Standardized effort	0.106 (0.001)	0.112 (0.001)			0.202 (0.057) [0.051]
Workload/10	0.001 (0.0001)	-0.007 (0.001)	1.322 (0.023)	0.025 (0.001)	-0.001 (0.004) [0.004]
Number of monthly red cases			0.415 (0.013)	0.043 (0.0005)	
Number of monthly normal cases			0.518 (0.011)	0.030 (0.0003)	
Cragg-Donald Wald F -statistic					425.6
Judge fixed effects	NO	YES	YES	YES	YES
Calendar fixed effects	YES	YES	YES	YES	YES
Type of case fixed effects	YES	YES	YES	YES	YES
Number of judges	21	21	21	21	21
Number of cases	50,412	50,412	50,412	50,412	50,412
Observations	2,032,449	2,032,449	2,032,449	2,032,449	2,032,449

Notes: In columns (1), (2) and (5) the dependent variable is the hazard that a case is closed in week t after filing, conditional on survival up to week t . *Active cases* are cases that have already seen a first hearing before the end of a given week but are not completed yet by the same date. *Standardized effort* is the number of hearings held by a judge in a given week divided by the average number of hearings needed to close the cases for which a hearing was held in the same week. *Total workload* is the cumulated sum of cases assigned to a judge up to the end of a given week. Columns (3) and (4) report the first-stage regressions used to construct the third-order polynomials of residuals on which the control function estimates reported in column (5) are based. The instruments are respectively defined as the number of *red code* and *green code* cases assigned to a judge in the five weeks that precede and include a given week (or the number of cases assigned in the first month of life of a case, if the week is one of the first five of the life of the case). *Red code* cases are those that, according to a survey of judges and lawyers, are considered as urgent and/or complicated, thus requiring immediate action and/or greater effort. *Green code* cases are instead the remaining standard and simpler cases. The Cragg-Donald Wald F -statistic (*Joint*) denotes the minimum eigenvalue of the joint first-stage F -statistic matrix. Standard errors in parentheses are clustered at the case level. Standard errors in square brackets are clustered at the judge level and bootstrapped with 100 repetitions.

38 hearings per week. Thus, the effect of task juggling is sizable in comparison to the effect of effort, given the observed average behavior of judges in terms of these two variables: in percentage terms, an increase in task juggling must be compensated by a similar increase in effort to keep the hazard rate constant.

To translate the effects of task juggling and effort on completion hazards into effects on the duration of cases, some distributional assumptions are needed. As suggested

by Arellano (2008), the Cox proportional hazard model can be written as a linear regression for the transformation $\Lambda(D_{ij}) = \int_0^{D_{ij}} u \, du$ of the underlying duration D_{ij} of case i assigned to judge j

$$\ln \Lambda(D_{ij}) = -X_{ij}\Psi + \eta_{ij} \quad (3)$$

if the error term η_{ij} has an extreme value distribution independent of the regressors X_{ij} . More specifically, if the baseline hazard $h_0(t) = 1$ then $\Lambda(D_{ij}) = D_{ij}$ and the regression simplifies to

$$\ln D_{ij} = -X_{ij}\Psi + \eta_{ij}, \quad (4)$$

which implies that the estimated coefficients of the Cox proportional model can be interpreted as the effect of a one-unit change of each regressor on the logarithm of duration. Under these functional assumptions, using the estimates in the second column of Table 6, a 1% increase of task juggling, corresponding to about two additional active cases, would raise the average duration of trials by approximately 1%. At the sample average of 280 days, this would be a reduction of about 3 days. A 1% increase of effort (0.38 more hearings per week) would increase *standardized* effort by almost 0.08 units reducing the duration of trials by 0.89%—that is, approximately 2.4 days at the sample average.

As explained in the next section, however, these estimates may not be consistent for the causal effects in which we are interested.

5.3. Identification of Causal Effects and Control Function Estimates

Like in the case of omitted variable bias in linear regression, Maximum Likelihood estimation of the hazard (2) does not guarantee that the causal effects in which we are interested are identified and estimated consistently because of the possibility that judge-specific time-varying unobservables are not independent of variables in X_{ijt} and at the same time affect the hazard. Such unobservables must necessarily be independent of the workload W_{ijt} which is randomly assigned (see Section 2), but may not be independent of task juggling A_{ijt} and standardized effort $(e/S)_{ijt}$. To see why, let M_{ijt} denote one such variable, for example, the bad mood or a health shock of judge j in week t of case i , which could affect at the same time task juggling, effort exertion and completion hazards in that week, generating spurious correlations between these variables with no causal interpretation. Suppose for one moment that we could observe M_{ijt} perfectly. Then the correctly specified hazard would be

$$\begin{aligned} \tilde{h}_{ijt} &= h_0(t) \exp(\beta_1 A_{ijt} + \beta_2 (e/S)_{ijt} + \beta_3 W_{ijt} + \beta_4 Z_{ijt} + \beta_5 T_t + \delta_j + \gamma M_{ijt}) \\ &= h_0(t) \lambda(X_{ijt}, \Psi) \exp(\gamma M_{ijt}) \end{aligned} \quad (5)$$

and the following condition would hold:

$$E(\omega_{ijt} - \tilde{h}_{ijt} \mid X_{ijt}, M_{ijt}) = 0, \quad (6)$$

where $\omega_{ijt} = 1$ if case i is closed exactly at t by judge j . This condition says that any shock affecting the actual completion hazard is random and independent of the determinants of the correctly specified parametric hazard \tilde{h}_{ijt} .

Now suppose that we do not observe M_{ijt} , which is the truth in our case, and that therefore we can only estimate the misspecified hazard (2). Then, using (6) the expected difference between the true and the misspecified hazard would be

$$\begin{aligned} E(\omega_{ijt} - h_{ijt} \mid X_{ijt}) &= E(\tilde{h}_{ijt} - h_{ijt} \mid X_{ijt}) \\ &= h_0(t) \left[E(\lambda(X_{ijt}, \Psi) \exp(\gamma M_{ijt}) \mid X_{ijt}) - \lambda(X_{ijt}, \Psi) \right] \\ &\neq 0 \end{aligned} \quad (7)$$

and the estimates of the causal parameters in Ψ would be inconsistent.²⁵

To address this problem we follow Palmer (2013) in using the control function approach proposed by Heckman and Robb (1985) adapted to the context of duration analysis. Consider the first-stage twin regressions

$$A_{ijt} = \alpha_0^A + \alpha_1^A V_{ijt}^1 + \alpha_2^A V_{ijt}^2 + \alpha_3^A W_{ijt} + \alpha_4^A Z_{ijt} + \alpha_5^A T_t + \delta_j^A + u_{ijt}^A \quad (8)$$

$$(e/S)_{ijt} = \alpha_0^e + \alpha_1^e V_{ijt}^1 + \alpha_2^e V_{ijt}^2 + \alpha_3^e W_{ijt} + \alpha_4^e Z_{ijt} + \alpha_5^e T_t + \delta_j^e + u_{ijt}^e, \quad (9)$$

where V_{ijt}^1 and V_{ijt}^2 are exogenous determinants of task juggling and standardized effort that are independent of M_{ijt} and do not affect the completion hazard directly. The other regressors are the exogenous determinants of the hazard which are included in X_{ijt} (note in particular, that W_{ijt} and Z_{ijt} are exogenous because of the random assignment of cases to judges). The residuals u_{ijt}^A and u_{ijt}^e in these two equations can be correlated and capture the component of A_{ijt} and $(e/S)_{ijt}$ which depend on M_{ijt} .

Conditioning also on these residuals in the hazard (2) solves the endogeneity problem and delivers consistent estimates of the causal effects of interest. To see why,

25. Note that if M_{ijt} and X_{ijt} were independent, the unobservability of M_{ijt} would be less problematic because equation (7) would become

$$E(\omega_{ijt} - h_{ijt} \mid X_{ijt}) = [E(\exp(\gamma M_{ijt})) - 1] h_0(t) \lambda(X_{ijt}, \Psi)$$

and the unobservability of M_{ijt} would affect only the estimation of the baseline hazard in which we are not interested and that would simply be scaled by $E(\exp(\gamma M_{ijt}))$. The condition (7) would be satisfied on average and the maximum likelihood estimates of the causal parameters of interest in Ψ would be unaffected and consistent, as long as M_{ijt} and X_{ijt} are independent. However, in the context of our application there is no reason to expect that this should be true.

consider the following augmented specification of the hazard:

$$\begin{aligned}\bar{h}_{ijt} &= h_0(t)\exp(\beta_1 A_{ijt} + \beta_2(e/S)_{ijt} + \beta_3 W_{ijt} + \beta_4 Z_{ijt} \\ &\quad + \beta_5 T_t + \delta_j + g(\hat{u}_{ijt}^A, \hat{u}_{ijt}^e)) \\ &= h_0(t)\lambda(X_{ijt}, \Psi)\exp(g(u_{ijt}^A, u_{ijt}^e)),\end{aligned}\tag{10}$$

where $g(u_{ijt}^A, u_{ijt}^e)$ is a polynomial in the estimated residual. Using this specification, the expected difference between the true and the augmented hazard would be

$$\begin{aligned}E(\omega_{ijt} - \bar{h}_{ijt} \mid X_{ijt}, u_{ijt}^A, u_{ijt}^e) &= E(\tilde{h}_{ijt} - \bar{h}_{ijt} \mid X_{ijt}, u_{ijt}^A, u_{ijt}^e) \\ &= h_0(t)\lambda(X_{ijt}, \Psi)\left[E(\exp(\gamma M_{ijt}) \mid X_{ijt}, u_{ijt}^A, u_{ijt}^e) \right. \\ &\quad \left. - \exp(g(u_{ijt}^A, u_{ijt}^e))\right]\end{aligned}\tag{11}$$

which is equal to zero if the control function $\exp(g(u_{ijt}^A, u_{ijt}^e))$ is equal to the conditional expectation $E(\exp(\gamma M_{ijt}) \mid X_{ijt}, u_{ijt}^A, u_{ijt}^e)$. If the control function $g(\cdot)$ is linear and the conditional distribution γM_{ijt} is normal with appropriate mean and variance, the equality holds exactly. Otherwise, identification relies on the quality of the polynomial control function in approximating the conditional expectation of $\exp(\gamma M_{ijt})$. While this quality can be assessed by showing that results are robust to different specifications of the polynomial $g(\cdot)$, which is the case in our application,²⁶ the crucial assumption on which this identification strategy stands is that the instruments V_{ijt}^1 and V_{ijt}^2 are independent of the omitted determinant M_{ijt} of the hazard. To construct instruments that satisfy this condition we exploit the lottery that assigns cases to judges in the way explained as follows.

Consider a case i , assigned to judge j , that has reached week t from filing. During the t weeks of life of this case, the judge has received (randomly) other cases that have lower seniority than case i because they were filed later. If this judge were working according to a purely sequential work schedule (no task juggling) and if her effort exertion were independent of the future workload, the number and characteristics of the cases filed after case i should be completely irrelevant for the completion hazard of case i . This is because the judge would work on the lower seniority cases only after having completed, at the usual pace, case i , whose hazard and effective duration would therefore be unaffected.

Suppose instead that the judge juggles task and/or changes effort exertion as a function of future workload. Under this assumption, the number and characteristics of the trials assigned after the filing of case i would be relevant for the completion hazard

26. In Table 6 we present results based on a third-degree polynomial, but we have experimented with polynomials of different degrees, obtaining similar results.

and the duration of case i . But this feature of future cases would solve the identification problem only under the assumption that there is no channel through which they affect the relevant outcomes different than the effort devoted by the judge to case i and/or the insertion of these cases in the batch of active trials on which she is working, instead of concentrating first on case i only.

It follows from these considerations that the cases assigned to a judge during the life of an active case could be used to construct arguably valid instruments for A_{ijt} and $(e/S)_{ijt}$:

- because they are randomly assigned (see Section 2), and thus provide an exogenous source of variation;
- if they affect the duration of case i only when the judge opens these cases before having completed case i (i.e., when the judge juggles more tasks by increasing A_{ijt}) and/or changes her standardized effort $(e/S)_{ijt}$; this can occur either because these cases, being for example more demanding, increase the propensity of the judge to hold hearings in each week (which would be a change in e_{ijt}), or because, again being more demanding, they require more hearings to be completed (which would be a change in S_{ijt}).

There is a possibility that the second of these conditions does not hold and that future workload assignments affect the duration of a current case through other channels beyond the measures of task juggling and standardized effort at our disposal, in which case the instruments described previously would violate the exclusion restrictions. For example, pressed by future cases that are either difficult or more interesting, a judge could devote less time to study a current case, or vice versa. However, it is plausible that this condition holds if what the judge can do under the pressure of future cases ultimately translates only into hearings per unit of time (effort) or into a different scheduling of hearings (task juggling) in order for the duration of cases to be affected.

To see if our conclusions are confirmed also under this alternative identification strategy we proceed in the following way. Note that, as anticipated in Sections 2 and 3, trials are not all homogeneous. On the basis of our conversations with judges and lawyers, the portfolio of cases that judges receive can be divided into two distinct categories that we label as *red code* and *green code*, by analogy with what happens in a hospital emergency room. Red code cases are those that, according to judges, are urgent and/or complicated, thus requiring immediate action and/or greater effort. Green code cases are instead the remaining standard and simpler ones. The arrival of the two types of cases is subject to potentially different stochastic processes: for example, firing and compensation litigations, which get a red code, are linked to the local business cycle, while litigations on government benefits, which fall in the green code category, are linked to changes in legislation and bureaucratic regulations whose timing follows different and more erratic rules. Moreover, as in an emergency room, while it would be perfectly reasonable that a newly arrived red code case begins to be treated before the disposition of a previously arrived green code case, the opposite

would be less reasonable. Hence we may expect that the arrival of both these types of cases follow different random processes and have different effects on the work schedules adopted by judges.

In an independent survey we therefore asked a set of judges and lawyers to assign a red code or a green code to the different possible types of cases: 22% (11,153) of the 50,412 trials considered in this study ended up being classified in the first group. On the basis of this classification, we then use as instruments V_{ijt}^1 and V_{ijt}^2 , respectively for A_{ijt} and $(e/S)_{ijt}$ in equations (8) and (9), the numbers of red code and green code cases assigned to a judge in the 5 weeks that precede and include week t .²⁷ These cases have been randomly assigned after the filing of case i and should therefore not affect the completion hazard of case i if judge j does not juggle tasks or change effort. Descriptive statistics for these instruments are reported in Table 5, while the second and third columns of Table 6 report the estimates of the first-stage regressions (8) and (9) which are strongly significant (the Cragg–Donald Wald F -statistic is equal to 425.6).

The control function estimates of the hazard (10), based on these first-stage regressions, are in the last column of Table 6. As in the case of the estimates of column (2) of the same table, which control for endogeneity just by including judge fixed effects, also these results match the predictions of our theoretical framework: task juggling and workload have a negative effect on completion hazards while standardized effort has the opposite effect. The estimated coefficients are statistically less significant, as expected, but we can still reject, at standard significance levels, that the coefficients of task juggling and standardized effort are equal to zero. Since the control function approach treats the estimated residuals from the first stages as fixed, the usual asymptotic standard errors are likely to be underestimated. For this reason the table reports also boot-strapped standard errors (with 100 replications), which are in any case very similar to the usual ones.²⁸

In terms of economic significance the causal effects of task juggling and standardized effort estimated by these coefficients are slightly larger than the effects estimated in column (2), but their relative size is comparable. Note that also the estimates of column (5) can be interpreted as the effect of a one-unit change of each variable on the log hazard. Based on these estimates and repeating the thought experiment described previously for column (1), suppose that a judge increased by 1% her level of task juggling. Since the average number of active cases at the end of a week is 195 (see Table 5), this means adding approximately two more cases to the ones already open on the judge's desk. This change of working schedule would imply a 2.1% decline in the completion hazard ($\approx \beta_1 \Delta A = (0.011 * 1.95) * 100$). To

27. In the first five weeks of life case i , the instrument is the number of cases assigned in the first month after the filing of case i . This is because if we used the general rule to construct the instruments in the first five weeks of case i , we would capture also cases assigned before the filing of case i .

28. In the Online Appendix (downloadable from www.andreaichino.it) we report results that cluster standard errors at the judge and judge-week level, and conclusions about significance are essentially unchanged.

compensate with effort this hazard decline, the judge would have to hold approximately 0.52 additional hearings per week ($\approx (\beta_1 \Delta A)S/\beta_2 = (0.011 * 1.95) * 4.9/0.202$), which would represent a 1.4% increase of effort starting from the sample average of 38 hearings per week. Thus, given our identification strategy, also in terms of causal estimates the effect of task juggling is sizable in comparison to the effect of effort and, in percentage terms, an increase in task juggling must be compensated by a significant increase in effort to keep the hazard rate constant.

Under the functional assumptions described at the end of Section 5.2 to translate the effects on completion hazards into effects on the duration of trials, using the estimates in the fifth column of Table 6, a 1% increase of task juggling, corresponding to about two additional active cases would increase the duration of cases by approximately 2.1%. At the sample average of 280 days, this would be an increase of about six days. A 1% increase of effort (0.38 more hearings per week) would increase standardized effort by almost 0.08 units reducing the duration of trials by 1.6%—that is, approximately 4.4 days at the sample average.

The reason why the IV-fixed-effects estimates of column (5) are larger in absolute size than the non-instrumented estimates of column (1) is likely to be related to the fact that different judges feel in different ways the pressure of new cases assigned after the filing of a given case, while working on it. Using the terminology of experimental econometrics, judges have different degrees of compliance with the assignment to treatment. Inasmuch as our estimates capture the causal effect of task juggling for “compliers”—that is, for judges that are more likely to feel the pressure of future work assignments, then it is not surprising that the absolute size of the estimates in column (5) is larger.

To conclude, the empirical evidence presented in Table 6 confirms the theoretical predictions concerning the effects of task juggling. Judges who are induced to juggle more tasks because they feel the pressure of future assignments of cases, require more time to complete the cases previously assigned to them. The estimated causal effect of task juggling is not only statistically significant but also quantitatively important in comparison to the causal effect of exerting more standardized effort in terms of more hearings per quarters or fewer hearings to close a case.

Before moving to a consideration of the effect of task juggling on the quality of decisions, we should mention a large management literature, surveyed by Mark, Gudith, and Klocke (2008), suggesting the existence of a *disruption cost* of interruptions, measurable in terms of additional time to reorient back to an interrupted task after the interruption is handled. In principle, also an effect in the opposite direction would be possible if it were “boring” for a judge to keep her attention continuously focused on only a few cases for a long time. In this case an increase in the number of interruptions would be good for performance because switching attention would generate a benefit from variety. Our data do not allow to separate these effects one from the other and both from the mechanical one studied in our theoretical model. However, our estimates described previously suggest that the boredom effect must be weaker and that the other two effects must prevail, so that task juggling has overall a detrimental effect on duration.

6. The Effect of Task Juggling on the Probability that a Case is Appealed

Increasing completion hazards and reducing the duration of trials are goals that could potentially come at the cost of a lower quality of decisions. Before concluding that task juggling has only detrimental effects, it is necessary to gather evidence on this possibility. For example, in the presence of positive externalities between the activities of a judge on different cases, such that she learns from some trials at the advantage of others and vice versa, task juggling could decrease completion hazards while at the same time allowing for better decisions on average. A similar effect would take place if juggling limited to first hearings (i.e., holding them as soon as possible after filing instead of waiting for the closure of cases with higher seniority) could facilitate settlements which were jointly perceived by parties as better outcomes for their cases.

In Section 3 we have anticipated descriptive evidence (see Figure 2) showing that judges who juggle more trials tend to have, on average, higher appeal rates for their cases or at least that they do not have lower rates. Unfortunately, appeal rates are not the ideal measure of the quality of decisions for a variety of reasons, as we already mentioned in Section 3.5, but it is the only measure at our disposal and, in this section, we want to see whether the correlation displayed in Figure 2 is robust in controlling for the relevant confounding factors suggested by our theoretical framework and whether it may be given a causal interpretation.

Denoting with $P_{ij} = 1$ the event that case i , assigned to judge j , is appealed, we estimate the probit model

$$\Pr(P_{ij} = 1) = \Phi\left(\beta_1 \bar{A}_{ij} + \beta_2 (\bar{e/S})_{ij} + \beta_3 \bar{W}_{ij} + \beta_4 Z_i + \beta_5 T_t + \delta_j\right), \quad (12)$$

where $\Phi(\cdot)$ is the normal cumulative distribution. The right-hand side variables are averages computed over the weeks between filing and disposition of case i . The \bar{A}_{ij} is the average number of active cases that are open on the desk of judge j . The $(\bar{e/S})_{ij}$ is the average standardized effort of judge j . The \bar{W}_{ij} is the average total workload assigned to judge j . Z_i is a set of 14 dummies for the type of controversy. T_t is a set of calendar weekly and yearly dummies for the week in which case i has been filed. δ_j are judge fixed effects.

Results are reported in Table 7, which has the same structure of Table 6. Also in this case the inclusion or exclusion of judge fixed effects (see the first two columns) leaves the estimates essentially unchanged. Task juggling has a positive effect on the probability of appeal, although marginally significant. Using the fixed-effects estimates in the second column, an increase of 100 active cases on the desk of a judge (which is approximately a 50% increase with respect to the observed sample average of 195) would increase her rate of appeal by 0.4 percentage points (a marginal effect of approximately 25% at the sample average of 1.7%). Effort has instead a negative effect which is more precisely estimated. From the viewpoint of the quantity versus quality trade-off, these results confirm the evidence of Figure 2 and suggest that, if anything, task juggling reduces the quality of decisions as measured by appeal rates.

TABLE 7. The effect of task juggling on the probability of appeal.

Dependent variable	<i>Appeal</i>	<i>Appeal</i>	Active cases	Standardized effort	<i>Appeal</i>
Estimation method	ML	ML	OLS-FE	OLS-FE	ML-Control
Stage	Second	Second	First	First	Second
Variables	(1)	(2)	(3)	(4)	(5)
Active cases/100	0.001 (0.001)	0.004 (0.002)			0.015 (0.034) [0.022]
Standardized effort	-0.001 (0.0001)	-0.001 (0.0004)			-0.008 (0.004) [0.003]
Workload/100	0.0002 (0.0002)	0.002 (0.0002)	0.046 (0.014)	-0.020 (0.021)	0.001 (0.002) [0.001]
Red code cases			0.039 (0.013)	0.120 (0.076)	
Green code cases			0.036 (0.016)	0.350 (0.050)	
Cragg–Donald Wald <i>F</i> -statistic (Joint)					77.96
Average outcome	0.017				
Judge fixed effects	NO	YES	YES	YES	YES
Calendar fixed effects	YES	YES	YES	YES	YES
Type of case fixed effects	YES	YES	YES	YES	YES
Number of judges	21	21	21	21	21
Number of cases	50,412	50,412	50,412	50,412	50,412

Notes: In columns (1), (2) and (5) the dependent variable is a dummy taking value 1 if the case has been appealed to a higher court and the estimates in each row are the *Probit* marginal effects of the correspondent variable. *Active cases/100* is the average, over the duration of each trial, of the cases that have already seen a first hearing before the end of each week but are not completed yet by the same date. *Standardized effort* is the average, over the duration of each trial, of the number of hearings held by a judge in each week divided by the average number of hearings needed to close the cases for which a hearing was held in the same week. *Workload/100* is the average, over the duration of each trial, of the cumulated sum of cases assigned to a judge up to the end of the same week. Columns (3) and (4) report the first-stage OLS regression estimates used to construct the third-order polynomials of residuals on which the control function estimates reported in column (5) are based. The instruments are respectively defined as the average number, over the duration of each trial, of “red code” and “green code” cases assigned to a judge in a given week. *Red code* cases are those that, according to a survey of judges and lawyers, are considered as urgent and/or complicated, thus requiring immediate action and/or greater effort. *Green code* cases are instead the remaining standard and simpler cases. The *Cragg–Donald Wald F-statistic (Joint)* denotes the minimum eigenvalue of the joint first-stage *F*-statistic matrix. Standard errors in parentheses are clustered at the judge level. Standard errors in square brackets are clustered at the judge level and bootstrapped with 100 repetitions.

We exploit the control function approach described in Section 5.3 to obtain causal estimates of the effect of task juggling and standardized effort on the probability of appeal based on a different identification strategy, to seek confirmation of the conclusions based on fixed effects. Results based on IV fixed effects are shown in the remaining columns of Table 7. The first stages are in columns (3) and (4). Instruments were constructed with the same logic described in Section 5.3 adapted to this context.

Specifically they are the average numbers of red and green code cases assigned to judge j during the weeks after the filing of case i and before its disposition. In the absence of juggling and if the judge does not change effort when pressed by future assignments, these cases should not affect the development and conclusion of case i , which should take place before judge j begins to look at them. These instruments are strongly significant for the measures of task juggling and standardized effort on the right-hand side of the Probit model (12). When the Probit model is augmented with third-degree polynomial control functions constructed on the basis of these instruments (see column (5) of Table 7) the effect of task juggling remains positive but is now very imprecisely estimated, and a zero causal effect cannot be excluded at standard significance levels. Similarly insignificant is the effect of workload, while standardized effort maintains a negative effect which is more precisely estimated.

In the light of these results, it seems possible to conclude that task juggling does not improve the quality of decisions, at least inasmuch as a better quality is measured by a lower rate of appeal experienced by each judge. If anything, more task juggling seems to increase the rate of appeal. So while task juggling has a clear negative effect on completion hazards and thus increases the duration of trials, the analysis of appeal rates does not suggest the existence of more desirable effects on other dimensions. For these reasons, we think it is worth asking whether the observation of task juggling among judges is evidence that judges do not use their time efficiently. We turn to this question in the next section.

7. Quantifying the Social Impact of Task Juggling

In this section we attempt to quantify the magnitude of the social gains that can be achieved by reducing task juggling.

Consider the social benefit from decreasing the duration of trials by one day ($\approx 0.4\%$ at the sample average of 280 days), which can be achieved with a task juggling reduction given by $(\Delta \ln D)/\beta_1 \approx (0.004)/\beta_1$ and thus ranging between 0.7 and 0.3 of an active case, using respectively the fixed-effects estimates in column (2) or the IV-fixed-effects estimates in column (5) of Table 6. This social benefit is equal to

$$B = \begin{array}{c} \text{Social benefit} \\ \text{from} \\ \text{one day shorter trials} \end{array} - \begin{array}{c} \text{Judge's private cost} \\ \text{from} \\ \text{decreasing task juggling} \end{array} \quad (13)$$

We examine the two components separately.

7.1. Judge's Private Cost from Decreasing Task Juggling

One might take the perspective that it costs the judge nothing to reorganize her work-scheduling practices. From this perspective the judge's private costs of decreasing task juggling should equal zero. This perspective, while legitimate, begs the question of why judges are task juggling. What could lead a judge to juggle tasks? Perhaps

the pressure of lawyers who lobby for their case to be opened sooner by the judge; or, perhaps, the judge's desire to comply with administrative rules requiring cases to be opened relatively quickly. In either case, one might say that acting against these considerations would cause the judge to incur a private cost. How large might this cost be?

We can make progress on this question if we assume that the judge chooses his amount of task juggling optimally. What would it take to reduce the durations of all the judge's cases by one day? According to our estimates in Table 6, this reduction can be achieved by an increase in effort ranging between 0.23% (in the case of IV-fixed-effects estimates) and 0.41% (in the case of fixed-effects estimates). At the margin, the judge must equate the cost of effort to the wage. Given the judge's wage, this range of estimated yearly effort increases corresponds to yearly wage increases ranging, respectively, between €274 and €493.²⁹ This is the shadow value to the judge of reducing the durations of a year's worth of cases by one day each with additional effort. The same reduction in duration can, equivalently, be achieved, at the sample average of 195 active cases, by a 0.17% decrease of task juggling in the case of the IV-fixed-effect estimates (which means reducing by 0.3 units the number of active cases on the judge's desk) or by a 0.37% decrease of task juggling in the case of the fixed-effect estimates (which means reducing by 0.7 units the number of active cases on the judge's desk). On account of the judge's optimal choice of task juggling, the shadow value of one day shorter duration must equal the cost of reducing task juggling by these amounts—that is, €274 and €493 respectively. These figures represent an (admittedly rough but reasonable) estimate for the second term in equation (13).

7.2. Social Benefit of Decreasing Task Juggling

We now turn to estimating the social benefit from shortening trial durations. We think of the trial as resulting from a worker suing the firm. The disagreement concerns the wage level (the firm, being greedy, fails to increase the workers' wages sufficiently). The model is worked out in the Appendix. The main point of the model is that the match may be socially valuable and yet the bargaining between worker and firm may fail. In this case the worker sues and, at the end of the trial, the judge will force the firm to rehire the worker. Ignoring transfers between workers and firms, which are of no consequence for welfare, the social benefit of a shorter trial is that an efficient match is recreated sooner rather than later. The magnitude of this benefit can be conceptualized as follows (on a daily basis):

$$Q - \underline{w} - \underline{\pi}, \quad (14)$$

where Q represents the social value created in the match between worker and firm; \underline{w} is the outside option earned by a worker who is waiting for the litigation to conclude; and $\underline{\pi}$ represents the return earned by the firm on the capital freed up by the missing

29. A judge's yearly gross salary is approximately €120,000 (see <http://jobspot.it/stipendio-lordomagistrato>).

worker. In the Appendix we compute these figures based on national account statistics and find that the social benefit of a *decrease by one day* of the duration of a single trial equals about €62. Given that a judge receives on average 504 new cases per year, the total yearly benefit for society of one fewer litigation day on a year's worth of trials is about $62 \times 504 = €31,248$. This number represents an estimate for the first term in equation (13).

7.3. Putting Numbers to Equation (13)

According to our back-of-the-envelope calculation the social benefits from decreasing task juggling in a year by an amount big enough to reduce trials by one day, ranges between $31,248 - 274 = €30,974$ and $31,248 - 493 = €30,755$ per judge. To grasp the size of these figures, they correspond to about 25% of a judge's annual salary. This number is large but not implausible despite the approximations needed to compute it. It should also be noted that we are not adding the direct costs of keeping a case going; if we did so the costs of litigation would increase even further relative to our current large estimate. With all the due caveats, the large size of our estimate reflects the social value of improving judicial output, which can be achieved either by reducing task juggling, or by working more hours, or by hiring more judges.

8. Conclusions

We have presented empirical evidence in favor of the theoretical hypothesis that individual work scheduling and time use have significant effects on the speed at which workers can complete assigned jobs. We test this prediction on a sample of Italian judges and show that those who are exogenously induced to juggle more trials take more time to complete similar portfolios of cases. This effect is a by-product of task juggling. We argue that task juggling is largely practiced by workers in general, and we show that it is practiced by Italian judges in particular.

In order to identify the impact of task juggling on productivity, we proceed in two ways. First, we remove the main source of endogeneity including judges fixed effects in our specifications; second, we construct time-varying instruments based on the sample realization of the lottery that allocates cases to each judge. The effects that we measure are not only statistically significant but also quantitatively important: at the sample means, an exogenously induced 1% increase of task juggling would raise the duration of trials by approximately three to six days and would need to be compensated by a 1.1% to 1.4% increase of effort in order to keep the duration of trials constant. We believe that our results provide the first empirical estimates of the impact of work scheduling on productivity.

We have considered as well the effect of task juggling on the only measure of output quality at our disposal (percentage of appeals to higher court), finding no adverse effect. For these judges, therefore, we do not detect a trade-off between quantity and quality. Obviously, this conclusion may vary when looking at different types of workers.

Finally, we discuss the extent to which our results signal that judges are inefficient in their choice about how to allocate time between tasks: specifically, whether they engage in “excessive” task juggling. Our calculations show that, in order for the current scheduling protocol to be an unconstrained private optimum for judges, the effort cost of opening *fewer* cases must equal the cost doing *more* hearings. This seems paradoxical because opening fewer cases should not represent a direct cost to the judge. Therefore we conjecture that there is some unmeasured constraint or external pressure (e.g., lawyers lobbying) which compels judges to juggle tasks. But, even if the existence of private inefficiency could be debated, we show that the *social* inefficiency of longer trials induced by excessive task juggling has strong theoretical justifications and can be empirically evaluated: if a judge could reduce task juggling enough to cut the duration of her trials by one day, this would cause a social gain of about €30,000, which corresponds approximately to one-fourth of a judge’s annual income.

We have derived our results for the specific setting of Italian judges and the policy implications for them are clear: these judges should be helped and induced to work in a more sequential way, reducing the number of active cases contemporaneously open on their desks for given assigned workload.³⁰ But the message of our paper concerning the effect of task juggling on the speed of job completion and the inefficiency of time allocation, is more general because it applies to all those situations in which more output is required, but labor or capital cannot be increased at least in the short run. A more sequential work schedule might offer a solution in these cases, because it increases output per unit of time at the cost of delaying the beginning date of some projects but not their end date.

We view the analysis in this paper and its companion (Coviello et al. 2014) as a first step into the empirical and theoretical analysis of how work scheduling affects the individual production function.

Appendix: The Inefficiency of Delay in a Firing Litigation

Consider a society with 1 unit of Capital and 1 unit of Labor. If jointly used, Capital and Labor form an employment relationship that generates an output Q , which is formally received by the firm if production is undertaken. This amount must be split between a wage w and a profit π . If used separately in alternative activities, Labor earns a wage $\underline{w} > 0$ and Capital earns a profit $\underline{\pi}$. Efficiency requires joint production if and only if $Q > \underline{w} + \underline{\pi}$.

The judge can make sure that when production is efficient, it takes place even after a breakdown in bargaining. How the judge splits the surplus $Q - \underline{w} - \underline{\pi}$ is of no consequence to efficiency, but it matters for incentives. In what follows, we assume

30. We are starting an experiment in three Italian tribunals of different size, in which we will provide judges who are willing to volunteer with an electronic agenda specifically designed to help them in reducing task juggling and in scheduling the hearings of each trial in a concentrated and sequential way. The evidence from this experiment will be a primary focus of our future research agenda.

that in the case of bargaining breakdown—that is, when production is efficient but the proposed split is not incentive compatible—the judge sets the wage at the outside option of the proposing party. This assumption allows us to simplify the proposer’s objective function, because from the proposer’s perspective there is no difference between the event in which bargaining fails and there is no room for efficiency, and the one in which bargaining fails but the failure is cured by the judge because there is room for efficient production.

A.1. Firm Makes Offer, Worker’s outside Option is Private Information

We assume that the firm can make a take it or leave it offer to the worker. Therefore, if \underline{w} were known to the firm, the firm would be able to fully expropriate the worker by setting a wage $w = \underline{w}$. In this case efficiency would prevail because the firm would only produce if $Q - \underline{w} > \underline{\pi}$, which is the efficient condition.

In our model, however, \underline{w} is not known to the firm and is considered a random variable with cumulative distribution F . If the firm offers a wage w it will be accepted only if $w > \underline{w}$, an event which has probability $F(w)$. In this case the firm makes $Q - w$ in profits. If the wage is rejected then the firm makes profits $\underline{\pi}$, whether or not the judge steps in. The firm’s expected profits as a function of the wage offer w are, therefore,

$$F(w)(Q - w) + (1 - F(w))\underline{\pi}.$$

The first-order conditions are

$$\begin{aligned} f(w)(Q - w - \underline{\pi}) - F(w) \\ = f(w)[(Q - w - \underline{\pi}) - (F(w)/f(w))]. \end{aligned}$$

The second-order conditions are satisfied if, for example, $(F(w)/f(w))$ is increasing, which is a standard assumption in the literature (uniform satisfies it, for example). In this case the first-order conditions identify a maximum of the expected profits. Let w^* be the w which solves

$$[(Q - w - \underline{\pi}) - (F(w)/f(w))] = 0,$$

that is, our equilibrium wage offer. In the case of the uniform distribution $f()$, for example, $w^* = [Q - \underline{\pi}]/2$.

So the firm will offer w^* and the offer will be rejected whenever $w^* < \underline{w}$. However, this may be inefficient. This is the case whenever $Q - \underline{\pi} > \underline{w}$. So we have an inefficiency whenever

$$w^* < \underline{w} < Q - \underline{\pi}. \quad (\text{A.1})$$

In the case of the uniform distribution $f()$, for example, we have an inefficiency whenever

$$(Q - \underline{\pi})/2 < \underline{w} < Q - \underline{\pi}. \quad (\text{A.2})$$

If the uniform distribution is between 0 and 1, this event has probability $(Q - \underline{\pi})/2$. The inefficiency in this event is that no joint production takes place due to a greedy firm which offers a wage lower than the (unknown to the firm) worker's reservation value. We can see this formally. In this case realized social value equals $\underline{w} + \underline{\pi}$ which, in light of the right-hand side of (A.2), is smaller than Q . Accepted offers are always efficient.

The judge can ascertain the true value of \underline{w} and make sure that the firm and the worker still get together, when bargaining breaks down inefficiently. In this case, the faster the better.

A.2. Worker Makes Offer; Firm's Outside Option is Private Information

We assume that the worker can make a take-it-or-leave-it offer to the firm. Therefore, if $\underline{\pi}$ were known to the firm, the worker would be able to fully expropriate the firm by setting a wage $w = Q - \underline{\pi}$. In this case efficiency would prevail because the worker would only make the offer if the wage $w = Q - \underline{\pi} > \underline{w}$, which is the efficient condition.

In our model, however, $\underline{\pi}$ is not known to the worker and is considered a random variable with cumulative distribution G . If the worker offers a wage w it will be accepted only if $Q - w > \underline{\pi}$, an event which has probability $G(Q - w)$. In this case the worker makes a surplus of w in profits, whether or not the judge steps in. If the wage is rejected then the worker makes surplus \underline{w} . The worker's expected surplus as a function of the wage offer w are, therefore,

$$wG(Q - w) + (1 - G(Q - w))\underline{w}.$$

The first-order conditions are

$$\begin{aligned} & -wg(Q - w) + G(Q - w) + g(Q - w)\underline{w} \\ & = g(Q - w)[-w + (G(Q - w)/g(Q - w)) + \underline{w}]. \end{aligned}$$

The second-order conditions are satisfied if, for example, $G(w)/g(w)$ is increasing, which is a standard assumption in the literature (uniform satisfies it, for example). In this case the first-order conditions identify a maximum of the expected profits. Let w^* be the w which solves

$$[-w + (G(Q - w)/g(Q - w)) + \underline{w}] = 0,$$

that is, our equilibrium wage offer. In the case of the uniform distribution $f()$, for example, $w^* = [Q + \underline{w}]/2$.

So the worker will offer w^* and the offer will be rejected whenever $w^* > Q - \underline{\pi}$. However, this may be inefficient. This is the case whenever $Q - \underline{\pi} > \underline{w}$. So we have an inefficiency whenever

$$w^* > Q - \underline{\pi} > \underline{w},$$

or equivalently when

$$Q - w^* < \underline{\pi} < Q - \underline{w}.$$

In the case of the uniform distribution, for example, we have an inefficiency when

$$(Q - \underline{w})/2 < \underline{\pi} < Q - \underline{w}.$$

The inefficiency in this case is that the worker asks for a high salary, which is sometimes too high.

The judge can ascertain the true value of $\underline{\pi}$ and make sure that the firm and the worker still get together, when bargaining breaks down inefficiently. In this case, the faster the better.

A.3. *The Social Benefit of One Fewer Day of Trial Duration*

Whether it is the worker or the firm that makes take-it-or-leave-it offers to the other side, in case of a litigation, society gains

$$Q_t - \underline{w}_t - \underline{\pi}_t, \tag{A.3}$$

if the judge decides on day t instead of day $t + 1$, where subscripts indicate that each variable is measured over the period of one day. This expression has empirical counterparts that allow us to approximate its monetary value in the Italian context.

We focus on year 2009, the last one for which all the necessary information is available. Using Italian National Accounting Statistics provided by ISTAT³¹ the daily value added of a full-time average worker is €171.69, which we take it as a proxy for Q_t .

Slightly more complicated is to compute \overline{W} . During a firing litigation a worker can apply for a job at a different firm and take unemployment benefits for at least 1 year after firing until she finds a new job. In principle, from the viewpoint of the worker, her outside option could be computed as the weighted mean of the daily average labor earning of an Italian employee and of the corresponding daily unemployment benefit, using the unemployment rate to construct the weights. But from the viewpoint of society, unemployment benefits are a transfer that should have no role to play in efficiency calculations. Therefore, the “productive outside option” of the worker would simply be the average daily wage multiplied by the hazard of finding a job after having been fired. We are not aware of reliable estimates of this hazard, nor do we have suitable data to generate such an estimate. To be conservative, we assume a competitive labor market, in which the fired worker can always find immediately a job at the average wage and therefore the hazard of this event is equal to 1. Using again national statistics provided by ISTAT, the average daily wage is €102.53, which is our estimate for \overline{W} . We emphasize that this is most likely an upper bound of a realistic outside option,

31. See www.istat.it. All the other statistics used in what follows come from either ISTAT or the Bank of Italy.

also because workers involved in firing litigation are probably less productive than the average worker. But we prefer to be conservative in our calculation, in the sense of avoiding overestimating the efficiency gain of a shorter trial.

Moving to the firm, we first measure the stock of physical capital per full-time employee using Bank of Italy estimates: the capital–labor ratio can be set at €222,026.78. The question is, what can the firm do with this capital per worker during the litigation? It is important to consider that the Italian labor law states that a fired worker must be rehired by the firm if the judge rules that the firing was unjustified. The risk of having to rehire the worker often induces firms to leave the capital idle during the trial. But to be conservative, we can assume that, while the judge decides, the firm can earn the riskless interest rate on the capital that would otherwise be combined with the fired worker. Since the interest rate on one year government bonds was 0.0114 during 2009, the daily alternative profit for the firm can be set as $\bar{\Pi} = \text{€}6.93$.

Putting all these figures together, a conservative estimate of the social benefit of one fewer day of trial in a firing litigation is (in euro)

$$Q_t - \bar{W}_t - \bar{\Pi}_t = 171.69 - 102.53 - 6.93 = 62.23. \quad (\text{A.4})$$

Given that a judge receives on average 504 new cases per year (9.7 on average in a week), the total benefit for society of one fewer litigation day on all the trials of the year is €31,248 (=62*504).³²

References

- Allen, David (2001). *Getting Things Done*. Penguin.
- Arellano, Manuel (2008). Duration Models. Class notes. www.cemfi.es/arellano/duration-models.pdf.
- Ameriks, John, Caplin Andrew, and John Leahy (2003). “Wealth Accumulation and the Propensity To Plan.” *Quarterly Journal of Economics*, 118, 1007–1047.
- Aral, Sinan, Erik Brynjolfsson, and Marshall Van Alstyne (2007). “Information, Technology and Information Worker Productivity.” NBER Working Paper No. 13172.
- Ashenfelter, Orly, Theodore Eisenberg, and Stewart J. Schwab (1995). “Politics and the Judiciary: The Influence of Judicial Background on Case Outcomes.” *Journal of Legal Studies*, 24, 257–281.
- Bandiera, Oriana, Luigi Guiso, Andrea Prat, and Raffaella Sadun (2008). “What CEOs do.” Working paper, London School of Economics.
- Bellotti, Victoria, Brinda Dalal, Nathaniel Good, Peter Flynn, Daniel G. Bobrow, and Nicolas Ducheneaut (2004). “What a To-Do: Studies of Task Management Towards the Design of a Personal Task List Manager.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 24–29 April 2004, Vienna, Austria, pp. 735–742.
- Benjamini, Yoav and Yosef Hochberg (1995). “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society B*, 57, 289–300.
- Bertrand, Marianne and Antoinette Schoar (2003). “Managing with Style: The Effect of Managers on Firm Policies.” *Quarterly Journal of Economics*, 118, 1169–1208.

32. This calculation rests on the simplifying assumption that the social benefit of one fewer day of firing litigation approximates the average social benefit of all other trials, for which the calculation would be more difficult because of less easily obtainable empirical counterparts.

- Bloom, Nicholas and John Van Reenen (2007). "Measuring and Explaining Management Practices Across Firms and Countries." *Quarterly Journal of Economics*, 122, 1351–1408.
- Bloom, Nicholas, Carol Propper, Stephan Seiler, and John Van Reenen, (2009). "The Impact of Competition on Management Practices in Public Hospitals." Working paper, London School of Economics.
- Bohn, Roger (2000). "Stop Fighting Fires." *Harvard Business Review*, 78, 83–91.
- CEPEJ (2011). European Commission for the Efficiency of Justice. <https://wcd.coe.int/ViewDoc.jsp?id=1872077&Site=COE>
- Cobb, W. Charles and Paul H. Douglas (1928). "A Theory of Production." *American Economic Review*, 18(2), 139–165.
- Covey, Stephen (1989). *The Seven Habits of Highly Effective Leaders*. Simon and Schuster.
- Coviello, Decio, Nicola Persico, and Andrea Ichino (2014). "Time Allocation and Task Juggling." *American Economic Review*, 104(2), 609–623.
- Cox, David R. (1972). "Regression Models and Life-Tables." *Journal of the Royal Statistical Society B*, 34, 187–220.
- CSM (2010). CONSIGLIO SUPERIORE DELLA MAGISTRATURA, Quarta Commissione. Gruppo di lavoro per la individuazione degli standard medi di definizione dei procedimenti. *Relazione 2010*.
- DellaVigna, Stefano (2009). "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature*, 47, 315–372.
- DellaVigna, Stefano and Ulrike Malmendier (2004). "Contract Design and Self-Control: Theory and Evidence." *Quarterly Journal of Economics*, 119, 353–402.
- DellaVigna, Stefano and Ulrike Malmendier (2006). "Paying Not to Go to the Gym." *American Economic Review*, 96(3), 694–719.
- Esposito, Gianluca, Sergi Lanau, and Sebastiaan Pompe (2014). "Judicial System Reform in Italy—A Key to Growth." International Monetary Fund Working Paper, 14/32.
- Frantz, Roger (1992). "X-Efficiency and Allocative Efficiency: What Have We Learned?" *American Economic Review: Papers and Proceedings*, 82(2), 434–438.
- Garicano, Luis and Paul Heaton (2010). "Information Technology, Organization, and Productivity in the Public Sector: Evidence from Police Departments." *Journal of Labor Economics*, 28, 167–201.
- Gibbons, Robert and John Roberts (2010). *Handbook of Organizational Economics*. Princeton University Press.
- González, M. Victor and Gloria Mark (2005). "Managing Currents of Work: Multi-tasking Among Multiple Collaborations." In *ECSCW 2005: Proceedings of the Ninth European Conference on Computer-Supported Cooperative Work*, 18–22 September 2005, Paris, France, edited by H. Gellersen et al., pp. 143–162.
- Heckman, J. James and Richard Robb (1985). "Alternative Methods for Evaluating the Impact of Interventions: an Overview." *Journal of Econometrics*, 30, 239–267.
- Ichniowski, Casey, Kaathryn Shaw, and Giovanna Prennushi (1997). "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines." *American Economic Review*, 87(3), 291–313.
- Jackson, R. James (1963). "Jobshop-like Queueing Systems." *Management Science*, 10, 31–142.
- Jorgenson, W. Dale (1986). "Econometric Methods for Modeling Producer Behavior." In *Handbook of Econometrics*, Vol. III, edited by Zvi Griliches and M. D. Intriligator. Elsevier, pp. 1841–1915.
- Kling, R. Jeffrey (2006). "Incarceration Length, Employment, and Earnings." *American Economic Review*, 96(3), 863–876.
- Lusardi, Annamaria and Olivia S. Mitchell (2008). "Planning and Financial Literacy." *American Economic Review: Papers and Proceedings*, 98, 413–417.
- Mark, Gloria, Daniela Gudith, and Ulrich Klocke (2008). "The Cost of Interrupted Work: More Speed and Stress." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Florence, Italy, April 2008, ACM Press, pp. 107–110.
- Mitsopoulos, Michael and Theodore Pelagidis (2010). "Greek Appeals Courts' Quality Analysis and Performance." *European Journal of Law and Economics*, 30, 17–39.

- Palmer, Christopher (2013). "Why Did So Many Subprime Borrowers Default During the Crisis." Working paper, MIT.
- Moder, J. Joseph, Cecil R. Phillips, and Edward Davis (1983). *Project Management with CPM, PERT and Project Diagramming*. Van Nostrand Reinhold.
- Mullainathan, Sendhil and Eldar Shafir (2013). *Scarcity: Why Having Too Little Means So Much*. Henry Holt.
- Perlow, Leslie (1999). "The Time Famine: Toward a Sociology of Work Time." *Administrative Science Quarterly*, 44, 57–81.
- Repenning, Nelson (2001). "Understanding Fire Fighting in New Product Development." *Journal of Product Innovation Management*, 1, 285–300.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Online Appendix for "The Inefficiency of Worker Time Use"