The Inefficiency of Worker Time Use*

Decio Coviello — HEC MONTRÈAL

Andrea Ichino — BOLOGNA AND EUI

Nicola Persico — NORTHWESTERN

September 18, 2013

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 are induced exogenously to 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 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-Code: J0; K0; M5.
Keywords: Individual production function, work scheduling, duration of trials.

*We would like to thank seminar participants at 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. Francesco Manaresi, Giovanni Pica and Alfonso Rosolia gave us useful advise 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.
A previous version of this paper was based on a different identification strategy. We thank Daniell Li for helpful critical comments on that version, which prompted us to improve our identification strategy in the current version. The paper refers to an Internet Appendix, where some formal result are derived, which can be downloaded at the following link: *http://nicolapersico.com/files/appendix_nopub_discrete.pdf*
Email: andrea.ichino@unibo.it; decio.coviello@gmail.com; nicola@nicolapersico.com

1 Introduction

The managerial literature on "time use" documents that workers frequently carry out a project in short incremental steps, each of which is interleaved with bits of work on other projects. For example, in a seminal study of software engineers Perlow (1999) reports that:

> "A large proportion of the time spent uninterrupted on individual activities was spent in very short blocks of time, sandwiched between interactive activities. Seventy-five percent of the blocks of time spent uninterrupted on individual activities were one hour or less in length, and, of those blocks of time, 60 percent were a half an hour or less in length."

Similarly, in their study of information consultants Gonzalez and Mark (2005, p. 151) report that:

> "The information workers that we studied engaged in an average of about 12 working spheres per day. [...] The continuous engagement with each working sphere before switching was very short, as the average working sphere segment lasted about 10.5 minutes."

The popular "self-help" literature has recognized that scheduling is a challenge for many workers. Books such as *The Myth of Multitasking: How "Doing It All" Gets Nothing Done* give workers suggestions to reduce multitasking on the job.[1]

The fact that much work is carried out in short, interrupted segments, a phenomenon which we label *task juggling*, has been overlooked by economists. This is strange because, at least in theory, task juggling directly reduces productivity. This is shown in the next example.

> **Example 1.** A worker is assigned two jobs, A and B, each requiring 2 days of undivided attention to complete. If the worker is exogenously induced to juggle both jobs, for example working on A on odd days and on B on even days, then the first task is finished after 3 days and the second after 4 days. The average duration is 3.5 days. If, instead, she is allowed to focus sequentially on each job in turn, then she completes A in 2 days and, later, B in 4 days from assignment.

[1] The first two are: Resists making active [e.g., self-initiated] switches; and Minimize all passive [e.g., other-initiated] switches.(Cited from Crenshaw 2008, p. 89).

With this sequential work schedule the average duration is 3 days only and no job is finished later. The difference between 3.5 and 3 is the mechanical effect of task juggling on average duration.

We study the work schedules of some labor judges in Italy. Our paper documents that, just as in the example, our judges juggle tasks; the paper also shows causally that juggling lowers productivity substantially. We estimate the effect of task juggling on project duration by using an empirical specification derived from a theoretical model which generalizes Example 1. In our empirical setting, of course, judges cannot be expected to work on a single case at a time. But the number of cases they should efficiently be working on should be approximately constant over time, and independent of the rate at which future cases are assigned. So if some cases are assigned earlier rather than later, the non-juggling judge should respond to this shock by keeping the "too early" cases on ice, and the effect on the duration of all other cases should be nil. But if the judge juggles, then he puts the newly assigned cases "in process" immediately and that has an externality on the duration of the other cases. In order to identify a "causal" effect 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. Cases assigned in the future should affect the duration of a case assigned today only if they are opened while the current case has still to be completed. Thus, the number of future cases satisfy the exclusion restriction because they affect the duration of a previous case only through task juggling, i.e. only through the opening of new cases that should be optimally left on ice instead. Moreover, given the lottery that assigns cases to judges, the number of future cases is randomly assigned and thus generates an exogenous source of variation which is precisely useful to study the question addressed in this paper.

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.[2] 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.

[2] 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.

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 durations of all cases. According to our estimates, at the sample mean, if an increasing future caseload induces judges to increase task juggling by 1% (approximately 5 more opened cases per year), the average duration of trials would increase by 2.4 day (0.9%) and would need to be compensated by a 0.6% increase of effort in order to keep the duration of trials constant (almost 9 more hearings per year). Our results also suggest that the negative effect of increased task juggling is partly offset by an endogenous increase in effort on the part of judges, particularly when task juggling is induced by the arrival of red code cases.

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 *et. al.* (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 both a private or a social viewpoint. We conclude that while there are plausible but not conclusive reasons to argue that task juggling is privately inefficient[3], it is undeniable that it generates a social inefficiency and we provide estimates of its size.

A comment on our measure of productivity. We focus on the duration of projects for two reasons. First, in many practical cases duration is what the worker's principals (clients) want to minimize.[4] Second, in our empirical application, reducing the duration of trials is

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

[4]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 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.

a key statutory objective.[5] Duration of job completion is clearly not the only dimension of output: quality matters as well. We will show that 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, and we can focus on trials' duration only.

Although derived within the specific setting of Italian judges, the large estimated effect of time use practices on productivity that we estimate has more general implications and raises several important issues. First, it suggests that the managerial literature may be right in focusing on task juggling and time allocation. In fact, production functions which are estimated ignoring information on time use may be substantially misspecified, and sizable inefficiencies may be overlooked. Second, it raises the issue of the social cost of rules putting pressure on workers to increase task juggling. Third, the implications of our results apply also, more generally, to 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). This delay may not be optimal for other reasons in normal times, but may be the only feasible solution during workload peaks.

This paper fits broadly within the literature on the construction and estimation of production functions that can be traced back to the path-breaking article of Cobb and Douglas (1928).[6] 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 et al. (1997), suggesting that, in different areas of human behaviour, individual modes of time use and activity scheduling are associated, in some cases causally, to performance for given effort.[7] Thanks to the accurate

[5] 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.

[6] Jorgenson (1986) surveys extensively the origins of this literature.

[7] See, for example, Bertrand and Schoar (2003), Bloom et al. (2007,2009) and Bandiera et al. (2009) for CEO practices, Ameriks et al. (2003) and Lusardi and Mitchell (2008) for family financial planning and, closer to us, Aral et al. (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

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, i.e., 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 *et al.*, 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 in 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 this literature does not offer an empirical methodology to quantitatively evaluate its impact.

Task juggling is also related to the sociological/management literature on time use.[8] 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 *et al.* 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, Allen 2001). The success of these popular books suggests that people do indeed find it difficult to schedule tasks efficiently in the workplace.[9]

In a related theoretical paper (Coviello et al. 2013), we develop a model which studies the evolution of the workload of a single worker, as a function of her work organization practices. That model is not directly applicable to the empirical analysis which constitutes the main contribution of this paper, because it does not accommodate shocks to the arrival rate of cases. Therefore, that model does not directly yield an equation which allows us to identify and estimate the causal effect of task juggling on performance.

Gibbons and Robert (2010) and Della Vigna (2009), the latter specifically on the issue of self-control in individual behaviour.

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

[9]For a review of the academic literature on this subject see Bellotti et al. (2004). For a specular take on prioritization of tasks see the discussion of the "firefighting" phenomenon in Bohn 2000 and Repenning 2001).

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. 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 discusses whether judges who engage in task juggling are behaving inefficiently, and estimates the size of the social consequences of task juggling. Section 7 concludes.

2 The data

We use data from one Italian court specialized in labor controversies for the industrial area of Milan. Our dataset contains all the 50412 cases filed between January 1, 2000 and December 31, 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. We observe the complete history of all these cases cases from filing to disposition. For the judges who were already in service on January 1, 2000, we also have information 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 January 1, 2000) we analyze their productivity starting from the fifth of their quarters of service, in order to give them time to settle in. All the cases assigned to them 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 we know the backlog of not-yet-disposed cases at the beginning of the period of observation.[10]

In the econometric analysis we will use single cases as the units of analysis, but for the purpose of describing the motivating evidence we aggregate the data at the quarterly frequency. Table 1 describes the panel of judges that we study. Six judges are observed for all the 24 quarters, while the others are observed for fewer quarters with a minimum

[10]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.

of 7 quarters. The last column of the same table reports the number of cases assigned to each judge per quarter on average. The overall average is 128 cases per quarter-judge. The characteristics of the process that assigns cases to judges are crucial for the purpose of our study and require special attention.

In Italy, as in other countries, the law (Art. 25 of the Constitution) requires that judges receive a randomly assigned portfolio of new cases.[11] This random assignment is designed to ensure the absence of any relationship between the identity of judges and the characteristics of the cases assigned to them. In the court that we consider the random assignment is implemented in the following way. 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 all judges in service. Note that this type of assignment scheme allows for small sample variability in the assignment of cases to judges, but this small sample variability is not systematic and fades away over the long run.

Table 2 shows, for example, that during the first quarter of 2000, the 18 judges in service received 129 cases on average with a standard deviation of 13 cases. The standard deviation is similar in all the other quarters. This because if, for example, in a given day the extracted letter is B and 5 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 quarter judges may receive slightly different workloads in terms of size.

For the same reason, also the characteristics of the assigned portfolios of cases may occasionaly differ across judges within a quarter. This is shown in the top part of Table 3 that reports, for each quarter, the p-value of Chi-square tests of indipendence between the identity of judges and three discrete characteristics of cases: type of controversy (14 types); zip code of the plaintif's lawyer (55 codes); the number of parties in trial (capped at 10). In the majority of quarters, independence cannot be rejected at standard significance levels, but in some quarters it is rejected at the 5% level. As shown in the second part of the table, this happens in 7 out of 24 quarters for the type of controversy, in 2 out 24 quarters for the lawyer's zip code and in 7 out of 24 quarters for the number of parties in trial. However, this occasional disomogeneity of the portfolios of cases assigned to judges fades away when the number of quarters over which judges are observed increases. This is shown in the last part of Table 3 that reports the p-values of similar Chi-square tests for all cases assigned in the period spanned by the largest balanced panel of judges identifiable in our sample. As

[11]Other studies have exploited the random assignment of cases to judges for identification: for example Ashenfelter et. al (1995) and Kling (2006).

verifiable in Table 1, this largest panel involves 14 judges observed continuously between year 2000 and year 2002. The p-values of these tests show clearly that independence cannot be rejected when we consider cases assigned over a sufficienly long number of quarters.

Therefore, we can conclude that, within a time unit, differences in assignments are due only to small sample variability and are not systematic. More specifically, they are independent of the identity of judges, who thus receive, in the long run, qualitatively and quantitatively similar portfolios of controversies. Note that, since our panel is unbalanced (see Table 1) we cannot test independence over all cases assigned to all judges in all quarters. Over the whole sample, independence is clearly rejected because judges with longer tenure receive larger numbers of cases and because different judges receive cases in different quarters and nothing guarantees the similarity of filed controversies over time. Nevertheless, the fact that independence cannot be rejected when we test over the largest balanced panel observable in our data, ensures that difference between all the judges observed in a quarter (even if they have different tenure) are not systematically connected to the identity of judges, being due only to the alphabetic process of assignment described above.

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 from other datasets, 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 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 performance per quarter, computed over all the quarters in which each judge is observed.

3.1 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 days from filing until the date in which a sentence is deposited by the judge, or the case is settled, or censoring occurs in the few cases for which we do not see the end of the trial.[12] On the horizontal axis judges are ordered from the slowest one to the left (Judge 30) to the fastest one to the right (Judge 3). The height of the squares in the same panel indicates the workload of new cases assigned to each judge on average per quarter. 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). For example, at the opposite extremes, Judges 30 and 3 receive respectively 120 and 105 cases per quarter, but the first one needs 398 days to close them while the second one need only 178 days, i.e., less than half.

The bottom left panel in the same figure plots the number cases on which each judge is contemporaneously working on average in a quarter. We call these "active" cases and they will be the focus of our analysis because they measure the extent to which judges practice task juggling. Formally, a case is defined as active at a given date if its first hearing has already taken place but the case has not been completed yet. 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. Also in this panel (as in all the others of this figure) judges are ordered 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 strikingly high and statistically significant correlation across judges (0.93; p-value < 0.0001[13]) between the average number of active cases and the average duration of trials. Comparing again extreme cases, the slowest Judge 30 keeps on average 275 files contemporaneously open on his desk while Judge 3 works on only 116 cases at the same time. In general, those who "keep more pots on the fire need more time to complete meals". 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.

3.2 Throughput and backlog

For the reasons explained in the introduction, we prefer to use duration as opposed to throughput 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.

[12] See Section 2.

[13] Here and in the rest of this Section, p-values refer to the test that reported correlations are zero.

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, Figure 1 shows that the two measures are correlated. More precisely, it shows that if keeping too many files opened at the same time slows down the activity of a judge, also the number of cases he/she will be able to close per quarter will be negatively affected on average (but not necessarily within each specific quarter as we argued in the introduction). The top central panel of Figure 1 confirms this intuition by plotting the throughput of judges ordered, as usual, from left to right according to speed of case completion. The slowest Judge 30 has almost the worst throughput (106 cases per quarter, just 8 more than the worst performer, Judge 29). The most productive in terms of throughput is Judge 11 (131 cases per quarter) who is the second best performer in terms of duration. The correlation between the number of active cases and the number of closed cases across judges per quarter is negative (-0.36). Although its statistical significance is borderline (p-value = 0.1029), the negative sign suggests that judges who work on few cases at the same time, opening new ones only when older ones are closed, tend to increase their throughput per quarter.

Consistently with this hypothesis, 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 at the beginning of each quarter. This backlog ranges from the 545 cases of Judge 18, who keeps 258 cases open at the same time and is one of the worse performers in terms of duration and throughput, to the 230 cases of the already mentioned top performer Judge 3, who has on average only 116 files on his desk at the same time. Even if all these judges receive the same number of cases per quarter their backlog is highly correlated with the number of active cases (0.94; p-value < 0.0001).

3.3 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 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 behaviour 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. In contrast with the previously examined panels of this figure, here we do not see a clear pattern jumping out of the data. Some slow judges on the left (like 30 and 18) require less than 3 hearings to close a case on average, while many faster judges need more (including in particular the top performers 3 and 14). The correlation between duration and number of hearings per case is positive (0.18) but relatively low and statistically insignificant (p-value = 0.4348). Inasmuch as being able to decide a case with fewer hearings is a form of ability of a judge, this descriptive evidence does not suggest that such characteristics has a strong effect on performance as measured by total duration of cases.

A measure of effort is instead offered in our data by the number of hearings per unit of time. The idea is that, by exerting more effort, a judge can schedule more hearings per quarter and in this way can ceteris paribus improve her performance in terms of throughput and total duration of completed cases. This statistic is plotted in the bottom right panel of Figure 1 and also in this case we cannot infer an evident pattern connecting this measure of effort to performance in terms of duration: the correlation is -0.06 and is statistically insignificant (p-value = 0.7950).

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 throughput and total duration of the cases assigned to them. Indicators of experience, ability and effort are as well likely to be relevant determinants of performance, but in a possibly less significant way. However, to properly assess the relative importance of these factors a theoretical framework and a multivariate statistical analysis are needed, to which we turn in the next Sections 4 and 5.

Before doing so, it seems important to say a word on the possibility of a "quantity versus quality" trade off in the performance of judges. Could it be that the judges with the highest throughput and the lowest total duration 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 percent of appealed cases can be considered as a good measure of the quality of the judges' decisions. There is no evidence that the cases assigned to slow judges on the left have a lower probability of appeal than the cases assigned to fast judges on the right. If anything the opposite seems to hold, given that the correlation between total duration and the percent of appealed cases is positive (0.41; p-value = 0.0648). For this reason we focus just on the effect of task juggling on duration in the rest of this paper.

4 A theoretical framework to estimate the inefficiency caused by task juggling

In this section we seek a theoretical expression for the production function of judges that we will then use as a basis for the estimation of the inefficiency caused by task juggling. The measure of output we focus on is the duration of cases.[14] We will derive an expression for the duration of a case based on the effort put in by the judge, the complexity of the case, and the way in which the judge organizes her work schedule. The latter input is the novelty of our analysis, and is measured by the number of cases the judge is working on at the same time.

We first study this production function in a steady state, where the number of cases the judge is working on at the same time is constant. Then, in Section 4.3 we study the effect of a perturbation around this steady state: i.e. an exogenously induced change in the timing at which a judge opens the cases assigned to her. The instrument that we will use in our empirical analysis mimics this exogenous perturbation.

4.1 The basic setup

The effect of effort and complexity can be appreciated even in the most stark model in which only one case is assigned to the judge. This situation is particularly simple because there is no question of how effort is distributed among different cases. The only factors that determine duration, then, are the number of hearings, or steps, that it takes to adjudicate the case (which we denote by S) and the number of hearings the judge makes per period (which we denote by e_q). Under the assumption that the judge exerts the same effort in

[14] An alternative measure of productivity could be the sum of disposed cases in a given time interval. For the reasons detailed in the introduction this is not a straightforward measure of productivity across time in this and many other contexts. Note, in particular, that judges do not control their assigned workload but only the speed at which they complete it, which translates into a cumulated sum of disposed cases at any given quarter from assignment. Therefore, the average duration of job completion and the *cumulated sum* of disposed cases from assignment are two equivalent output measure for these judges.

every period we have $e_q = e$ and thus the duration of the (single) case has a very simple expression:[15]

$$D = \frac{S}{e}. \tag{1}$$

A similar expression can be derived when e_q is not constant across period.

When the judge receives more than one case, a third factor beyond e and S affects the duration of cases, namely, how the judge allocates his time across different cases. To describe the possible ways in which this is done we need to develop some language. This language, and the associated mathematical results, are formally developed in the Internet Appendix.[16] Here we describe intuitively several possible modes of work, that is, several algorithms that the judge may use to allocate his time across different cases.

First is the sequential work schedule, in which cases are worked on sequentially: first all the steps relating to the first assigned case are accomplished, then all the steps relating to the second assigned case, etc. The polar opposite of a sequential work schedule is the full rotation one, in which, within each step, all cases that the judge has received are worked on simultaneously until each is done. A partial rotation is a generalization of the previous process: it works just as a full rotation does, except that instead of rotating on all cases the judges has received, the judge keeps some cases unopened and only gradually inserts them into the rotation. Once a case is inserted in the rotation then it receives the same amount of attention as all other open cases. We define the opening of a case as the action of inserting that case in the rotation of all the other cases that are already opened.

The full rotation and sequential work schedules are polar extremes. In the full rotation schedule cases are started as early as possible and so, at any given point in time, there is a large mass of cases being simultaneously worked on. In contrast, a sequential work schedule causes the start of a new case to be postponed as late as possible, and so in a sequential work schedule the minimum possible number of cases is simultaneously being worked on at any point in time. The partial rotation is a general family of work schedules which subsumes as special and extreme cases the full rotation and the sequential schedule. This family is parameterized by the distance from assignment at which cases are opened. That is to say, a partial rotation can take different forms depending on how early after assignment the cases are opened. If, for example, all cases are opened as early as possible (i.e. immediately after assignment) then a partial rotation becomes identical to the full rotation. If, instead, new

[15] Actually, to be precise the duration would be the smallest integer that exceeds S/e, but from now on we will ignore such integer problems.

[16] This Appendix can be downloaded at: *http://nicolapersico.com/files/appendix_nopub_discrete.pdf*

cases are opened at the slowest possible pace, then there is only one case open at any given time and so the partial rotation coincides with the sequential work schedule. This is why a partial rotation is a convenient family of work schedules to work with.

Having introduced the notion of a partial rotation, we now want to use it to describe how time allocation practices affect productivity. In other words we want to generalize equation (1) by introducing a parameter which captures how close a partial rotation is to its polar extremes. The challenge here is that the production process we study evolves over time. To meet this challenge it is easiest to start by focusing on a system that evolves in a stable way as far as the number of cases included in the rotation is concerned. To this end we now introduce the simplest possible evolution of the system over time.

Definition 1. *A judge operates according to a* ***stable rotation*** *if:*

(a) in each period the judge keeps A_0 open cases;

(b) the number α of cases assigned, their complexity S, the effort e, and the number of new opened cases ν, are all constant in each period;

(c) the work schedule is a partial rotation;

(d) the number of cases completed ω is constant across periods, and is the same as the number of new opened cases ν.

A stable rotation describes the production process of a judge who works according to a partial rotation on a number of cases which remains stable over time. Figure 3 describes a snapshot of a judge's caseload in a stable rotation. Each folder represents a case and the horizontal axis is the number of hearings (steps) of that case that have already been completed. In this example, each case requires $S = 5$ hearings to complete. At the time of the snapshot, this judge has 5 opened cases that have had one hearing, 5 opened cases that have had two, and so on. Cases which are closer to completion are colored in a lighter shade. To the left of the vertical axis are cases which have not yet been started. The white folders represent cases that are done, i.e., have received 5 hearings.

Starting from this snapshot, if we let time run forward we will see that the judge holds one hearing for every opened case; this is because the judge follows a partial rotation. Graphically, this effort moves all folders one step to the right. In addition, the judge opens the five cases to the left of the vertical axis. Let us imagine that this is all the effort the judge has time for in a period (this implies $e = 25$). In this case $A_0 = 20$, and the input rate is exactly equal to the throughput rate, as it must be in a stable rotation. The throughput in a period is exactly 5 cases, which is equal to e/S. This equality is no coincidence: in

Section 1 of the Internet Appendix we prove that the input rate and the output rate must be exactly equal to e/S for there to be a stable rotation.

This finding shows that ν, our measure of congestion and time use, is not a free parameter in a stable rotation. Therefore we will have to go beyond a stable rotation if we want to examine the effect of a change in time use. This will be done in Section 4.3.

Note that in a stable rotation the duration of cases D_q need not be constant over time. Indeed, in a stable rotation the backlog of cases will grow if the arrival rate of cases exceeds the rate at which they are opened. In Section 1 of the Internet Appendix we fully analyze how a stable rotation operates and obtain the following expression for the duration of cases.

$$D_q = \frac{S}{e}\left(A_0 + \alpha q\right) - q. \tag{2}$$

This expression solves for the duration D_q of cases assigned in period q in terms of known quantities: the exogenous assignment rate α, the measure of effort e/S, and the initial condition A_0, which is a parameter that can be specified arbitrarily. If a judge starts out with A_0 active cases in $q = 0$, and new cases are opened at the rate of e/S in periods $q = 1, 2, ...$, then cases will be solved at a rate of e/S per period and at all times there will be A_0 active cases. While the output rate of cases does not depend on A_0, the duration of each individual case does according to expression (2).

Using this expression we can illustrate some of the determinants of duration, albeit at a stable rotation. The duration of a case is increasing in α, the rate at which cases are assigned to the judge. It is decreasing in e/S, which means that judges who work hard (high e) or who have easy cases and/or are more able (low S) will have a lower duration of cases in steady state. Having a large number of active cases A_0 increases duration. Finally, the duration of cases increases with the judge's tenure ($\frac{\partial D_q}{\partial q} > 0$) if and only if $\alpha > e/S$, that is, if the arrival rate exceeds the judge's effort scaled by the perceived complexity of cases. We record these findings in a proposition.

Proposition 1. *If judges operate according to a stable rotation, the duration of a case assigned at q is increasing in α, in S/e, in A_0 and, if $\alpha > e/S$, also in q.*

Equation (2) and Proposition (1) provide a theory-based starting point for implementing an econometric analysis of the contributing factors to durations. While in standard theories of the individual production function, that ignore the scheduling of tasks, the duration of trials would depend only on the size of the workload, the difficulty of cases, the effort and the ability of a judge, our framework suggests, instead, that how time is allocated across cases for given effort and ability must be included in the specification. The natural way to test the proposition would be to estimate a linear approximation of equation (2):

$$D_{i,t} = \gamma_0 + \gamma_1 \alpha^p_{i,t} + \gamma_2 \left(\frac{e}{S}\right)^f_{i,t} + \gamma_4 t + \gamma_5 A_{i,0} + u_{i,t} \tag{3}$$

where $D_{i,t}$ is the duration (from filing to disposition) of a case assigned to judge i at date t[17], $\alpha^p_{i,t}$ is the number of cases assigned to the same judge before t (the superscript p denotes that this is the "past" workload), $\left(\frac{e}{S}\right)^f_{i,t}$ is the effort, standardized by the complexity of cases, exerted by the judge in the period after date t in which the case is handled[18] (the superscript f denotes that this is the "future" effort), $\gamma_4 t$ is a time trend, $A_{i,0}$ is the initial judge-specific condition that defines the stable number of cases on which the judge rotates tasks. The presence of the error term $u_{i,t}$ is justified because in the data the workload, effort and complexity are not constant over time, while, if they were constant, equation (2) would be an exact relationship.

However, this specification is unsatisfactory because the congestion and time misallocation induced by task juggling are constant in a stable rotation. Since we want to estimate the effects of changes in congestion, this is a problem. From an empirical viewpoint, as well, the judges are observed to depart, albeit slightly, from the model of a stable rotation. The next Section explores the extent to which a stable rotation model captures the behavior of our judges.

4.2 Are judges scheduling tasks according to a stable rotation?

To establish whether judges effectively work according to a stable rotation we have estimated a regression of the number of open cases ν on the number of closed cases ω, obtaining the following results:[19]

$$\nu = \underset{(5.55)}{5.99} + \underset{(0.04)}{1.01}\,\omega \tag{4}$$

[17]To simplify notation we omit the subscript denoting single cases.

[18]In the idealized environment of a stable rotation in which effort (number of hearings per period) is constant over time, all cases take the same number of hearings to adjudicate, the number of active cases is constant over time and standardize effort $\left(\frac{e}{S}\right)^f_{i,q}$ is exactly equal to the (constant) potential number of cases that a judge could decide in a period. But as we will see in the next section, judges do not exactly operate according to a stable rotation, and therefore the variables of interest are not constant through time. We will explain below how we effectively measure standardized effort in a more realistic non-stationary environment.

[19]The regression has been estimated on 381 quarter-judge observations and include fixed effects for the 21 judges.

where standard errors are reported in parentheses under the coefficients. According to these estimates these judges work on a schedule that is very close to a stable rotation but does not coincide exactly with it. The slope is approximately equal to 1 indicating that judges open one new case for each case that they close. But the positive intercept (even if statistically not significant) suggests that on average they also open approximately 6 new cases in every quarter on top of those that they close. As a result the number of active cases on their desk steadily increases over time albeit at a relatively low pace.

This pattern can be appreciated graphically in Figure 4. The top left panel plots the number of cases opened and closed per quarter by the seven best judges in terms of average duration. The two lines are very close one to the other, which is what should happen if these judges work according to a stable rotation, but the numbers of opened and closed cases, albeit similar, are clearly not constant overtime also because of a strong seasonal pattern (Italian judges do not work much in the summer). The top right panel repeat the exercise for the seven worst judges. For these judges it happens more frequently that the number of new opened cases is larger than the number of closed cases. It is therefore not surprising to find, in the bottom left panel, that the seven worst judges have more active cases in each quarter. This panel also shows that for both types of judges (and in particular for the worst) the number of active cases increases over time with jumps that obviously correspond closely to the quarters in which more cases are opened than closed. Finally the last panel shows that the duration of all assigned cases differs across the two groups of judges and evolves over time within each group, in line with the number of active cases, as predicted by our model.

This evidence suggests that some judges are closer than others to a stable rotation schedule. But deviations from a stable rotation exist (in both directions) and have important effects on the number of active cases and on the duration of assigned cases. Thus a stable rotation is limited in its ability to account for what we see in the data and more generally to explain what is the effect of an *increase* in congestion. Indeed, in a stable rotation the amount of congestion is constant because, by definition, cases are opened at the same rate at which they are completed. We will therefore generalize our framework in the next Section 4.3, to the more interesting and realistic case in which congestion can change.

4.3 The effect of a change in task juggling

In this section we derive and sign the effects on durations of changing the timing at which cases are opened by the judge. If a batch of cases is opened sooner rather than later, we show that the duration of all cases increases. We consider the effect of an exogenous shock that induces judges to increase by Δ the number of cases ν_q newly opened in quarter q, relative to a stable rotation where cases are opened at the constant rate ν. Figure 5 describes this event. In the following proposition we assume, for the purpose of computing the duration of a case, that cases are opened immediately upon being assigned, and so no time is spent with the case being assigned but unopened.

Proposition 2. *Suppose that there are no switching costs and that the judge operates according to a partial rotation. Suppose that a judge has so far opened ν cases per period and has exerted constant effort in each period. Suppose now that, for a given q only, the judge makes $\nu_{q-1} = \nu + \Delta$ and changes nothing else. Then the duration of cases assigned at q increases.*

Proof: See Section 2 of the Internet Appendix ■

This proposition extends the introductory example of Section 1 to the case in which the judge works on an infinite stream of cases. Like in that example we find that if more cases are opened sooner, the average duration of all cases increases. To test this proposition, that applies to judges who deviate occasionally and for exogenous reasons from a stable rotation, the econometric specification (3) must be corrected to include a variable measuring how task juggling changes with respect to the initial condition. Specifically we use the number of new cases opened by judge i in the year after the date t in which each case is filed. We denote this variable as $\nu^f_{i,t}$. Thus, the specification that we want to estimate is:

$$D_{i,t} = \beta_0 + \beta_1 \alpha^p_{i,t} + \beta_2 \left(\frac{e}{S}\right)^f_{i,t} + \beta_3 \nu^f_{i,t} + \beta_4 t + \delta_i + \epsilon_{i,t} \tag{5}$$

where δ_i is a judge specific fixed effect that absorbs the initial condition $A_{i,0}$, even if it is not observed for some judges. As in equation (3), $D_{i,t}$ is the duration (from filing to disposition) of a case assigned to judge i at date t, $\alpha^p_{i,t}$ is the number of cases assigned to the same judge in the year before t (the past workload). In this non-stationary environment, future standardized effort $\left(\frac{e}{S}\right)^f_{i,t}$ is defined as the ratio between the number of hearings held for any case treated by the judge in the year after t (independently of the date of assignment) and the number of hearings that were necessary to adjudicate the new cases assigned to the judge in the year after t. This ratio can be interpreted as the potential number of cases, among those assigned in the year after t, that a judge could have decided with the effort that she/he effectively exerted in the same period. The presence of the error term $\epsilon_{i,t}$ is justified because of the time varying unobservable components that capture judge specific preferences concerning task juggling, effort exertion and the difficulty of cases.

What signs does the theory predict for the coefficients in this relationship? The signs of β_1 and β_2 are almost predicted by Proposition (1), but not exactly since Proposition (1) deals with the case of a permanent change in $\alpha^p_{i,t}$ and $\left(\frac{e}{S}\right)^f_{i,t}$, whereas β_1 and β_2 measure the effect of a temporary increase in their respective variables. So, for example, β_1 measures the effect on duration of going from $\alpha, \alpha, \alpha, \alpha, ...$ to $\alpha, \alpha, 2\alpha, \alpha, ...$ (where subscripts and superscripts have been omitted for simplicity). To establish the signs of β_1, observe that an increase in $\alpha^p_{i,t}$ means that more cases are exogenously assigned to judge i in the year before date t. Therefore, when the time comes for the judge to work on the case assigned at date t, it will necessarily take longer to complete it whatever the scheduling of tasks chosen by the judge. Most theories of the duration of trials, would predict, like ours , that $\beta_1 > 0$.[20]

Perhaps less controversial is the prediction that $\beta_2 < 0$, because an increase in standardized effort $\left(\frac{e}{S}\right)^f_{i,t}$ means that the judge holds more hearings e in the year after date t (for whatever cases are open on her desk during the same period), or reduces the number of hearings S needed to close the new cases assigned to her in the same period. This additional effort will benefit also the case assigned at date t. Note, as discussed in Section 2, that within each unit of time, by random assignment, all judges receive portfolios of cases that differ just because of random sampling. Therefore, if $S_{i,t} > S_{j,t}$ it must be either because judge i has randomly received a slightly more complex portfolio, 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 i. Moreover, for the same judge across units of time, it could happen that $S_{i,t} > S_{i,\tau}$, with $t < \tau$, and this may happen either because the ability of judge i increases over time or because the assigned cases becomes less difficult on average over time.

The main focus of our analysis is on the parameter β_3 which measures the inefficiency of task juggling, i.e. its effect on the duration of cases assigned to judge i at any date t. Proposition (2) states without ambiguity that this coefficient should be estimated to be

[20]But in the presence of learning by doing, economies of scale or positive externalities between cases, one could imagine that a larger workload might reduce the average duration of assigned cases. We will deal with such considerations later.

positive.

Finally, Proposition (1) gives the condition for the coefficient on the time trend β_4 to be positive. We specify this trend in the most flexible way as a set of dummies for each date t, so that we can control also for seasonality

5 The effect of task juggling on the duration of trials

While $\alpha^p_{i,t}$ is randomly assigned (see Section 2), if work scheduling has a role in the determination of the duration of trials, the error term $\epsilon_{i,t}$ in equation (5) is correlated not only with standardized effort $\left(\frac{e}{S}\right)^f_{i,t}$ but also with the measure of task juggling $\nu^f_{i,t}$. This because the error term includes the unobservable components that capture judge specific preferences concerning task juggling, as well as preferences for effort exertion and ability. There is, in principle, no reason to expect that these unobservable components should be time invariant.

Therefore to estimate consistently the causal effects of standardized effort and task juggling on trials duration with equation (5), we need some exogenous source of variation of these two variables.

5.1 Identification

Consider a judge who receives a given case today and will receive a random number of additional cases of different kinds in the future (say in the following year, to focus ideas). If this judge were working according to a purely sequential work schedule (see Section 4), the number and characteristics of the future cases should be completely irrelevant for the duration of the case assigned today. This because the judge will start working on the future cases only after having completed the case assigned today, whose duration would therefore be unaffected by what the judge does afterwards.

Suppose instead that the judge works according to a partial rotation schedule (see again Section 4). Under this assumption, the number and characteristics of future trials will be relevant for the duration of the case assigned today. And, interestingly from the viewpoint of our identification problem, they will influence the duration of the previously assigned cases only inasmuch as the judge inserts the future cases in the rotation and changes her effort as a function of the new assignments.

It follows from these considerations that the future cases can be used to construct valid instruments for $\nu^f_{i,t}$ and $\left(\frac{e}{S}\right)^f_{i,t}$, because

- they are randomly assigned (see Section 2), and thus provide an exogenous source of variation;
- they can affect the duration of a previously assigned case, only if the judge opens the future cases before having completed the previous one (i.e. juggles more tasks by increasing $\nu^f_{i,t}$) and/or changes her future standardized effort $\left(\frac{e}{S}\right)^f_{i,t}$; this second effect can occur either because the future cases, being for example more demanding, increase the propensity of the judge to hold hearings in the following year (which would be a change in $e_{i,t}$), or because, again being more demanding, they require more hearings to be completed (which would be a change in $S_{i,t}$),

5.2 Estimates

To bring the model to the data we first note that while until now we have assumed that all cases are homogeneous, reality is different.

On the basis of our conversation with judges and lawyers, it is clear that the portfolio of cases that judges receive must be divided in 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 both the past and the future arrival of these two 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% (11153) of the 50412 trials considered in this study ended up being classified in the first group. On the basis of this classification, we allow the previous workload to have different effects on the duration of a case assigned at date t, depending on its red or green code; thus we estimate the following modified version of equation (5)

$$D_{i,t} = \beta_0 + \beta_1^r \rho_{i,t}^p + \beta_1^g \gamma_{i,t}^p + \beta_2 \left(\frac{e}{S}\right)_{i,t}^f + \beta_3 \nu_{i,t}^f + \beta_4 t + \delta_i + \epsilon_{i,t} \tag{6}$$

where $\rho_{i,t}^p$ and $\gamma_{i,t}^p$ are the past workloads of, respectively, red and green cases, $\alpha_{i,t}^p = \rho_{i,t}^p + \gamma_{i,t}^p$ and β_1^r and β_1^g may potentially differ.

Given the justification provided in Section 5.1, we then use as instruments for $\nu_{i,t}^f$ and $\left(\frac{e}{S}\right)_{i,t}^f$ in equation (6), the numbers $\rho_{i,t}^f$ and $\gamma_{i,t}^f$ of red code and green code cases assigned to a judge in the year *after* date t, i.e. the future assignments in the two categories. Descriptive statistics for the variables used in the estimation of equation (6) are reported in Table 4, while results are reported in Table 5.

In the first column, equation (6) is estimated with OLS, including judge, year and month-of-the-year fixed effects. The signs of the coefficients correspond to the predictions of the theoretical model. In particular, more task juggling (measured by a larger number of new opened trials in the year after the assignment of a case) increases the duration of the case, while a greater future standardized effort has the opposite effect.

The OLS estimates of column 1, however, are potentially inconsistent for the causal effects of interest. To get a sense of the economic size of these effects, we need to use the Instrumental Variable (IV) estimates reported in the second column, which are based on the instruments described above. The estimates of the coefficients of the confounded variables $\nu_{i,t}^f$ and $\left(\frac{e}{S}\right)_{i,t}^f$ are larger in size and still statistically significant. For the effect task juggling, opening one additional case in the year after the assignment of a case increases the duration of that case by 0.50 days. As for the effect of an additional unit of standardized effort, it reduces trial duration by 0.84 days.

At the mean of the distribution of new opened cases in the year after the date t of assignment of a case (486[21]), a 1% decrease of task juggling means opening approximately 5 less cases in a year. Such change of scheduling would reduce the duration of a trial assigned at date t by 2.4 days, which is a 0.9% reduction of duration.

To put the size of this effect in the right perspective we can ask how many new hearings in a year (for given difficulty of cases) the representative judge would have to hold in order to achieve the same reduction in the duration of a trial assigned at date t. Given an estimate of -0.84 for the coefficient of $\left(\frac{e}{S}\right)_{i,q}^f$, 2.9 units of standardized effort in the year after date t (a 0.6% increase at the mean of this variable, which is 518) would be needed to reduce the duration of the case assigned at date t by the same amount of 2.4 days. In other words, at the

[21]See the descriptive statistics in Table 4

mean, a 1% decrease of task juggling has the same effect as a 0.6% increase of standardized effort. Given an average number of hearings per case of $S = 3.1$, the judge would have to hold almost 9 more hearings in the year after date t to achieve the same effct of opening approximately 5 less cases.[22]

The results described so far indicate that task juggling occurs with detrimental effects on the duration of trials. When judges receive new assignments, they feel this pressure and open more new cases slowing down the processing of previously assigned cases that are still unfinished.

One could argue that this is reasonable if the new cases are of the same or perhaps higher importance than the previous ones, but the last two columns of Table 5, which report the first stage regressions[23], show that judges feel also the pressure of the less important green code cases, which should have low priority. For example, column 3 reports OLS estimates of the effects of the excluded and included instruments on the measure of task juggling $\nu^f_{i,t}$. The estimated coefficients of the number of future red code cases $\rho^f_{i,t}$ and green code cases $\gamma^f_{i,t}$ are positive and almost identical, while it would have been reasonable to expect the coefficient of red code cases to be larger

The last column of Table 5 reports the estimates of the other first stage regression in which the dependent variable is future standardized effort. Here it is interesting to note that every red code case assigned after date t increases standardized effort in the same period by 0.54 units while the effect of a green code case is just 0.12 units. This is in line with the expectation that red code cases should have a larger effect on effort

To conclude, the empirical evidence 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. This is true also for trials that should have the highest priority and whose duration should be unaffected by the subsequent arrival of less important cases. 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

[22] The 21 judges in this sample hold on average 14 hearings per day of hearing and hold approximately 2 days of hearings per week.

[23] From a strictly econometric point of view, the alphabetical procedure that assigns cases to judges, described in Section 2, ensures that our instruments are randomly assigned, but does not guarantee that the instruments are also not weak. It is however reassuring to see that the Cragg-Donald Wald F statistics, reported at the bottom of the second column of Table **??**, is largely above the critical values computed in Stock and Yogo (2005), suggesting that our instruments are sufficiently strong not to jeopardize the intepretation of the IV estimates.

of more hearings per quarters or fewer hearings to close a case.

Before moving to the efficiency implications of our results, we should mention a large management literature, surveyed by Mark et. al. (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 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 above 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 Inefficiency?

The previous sections have shown that a 1% reduction in task juggling (about 5 less new opened case per year) achieves the same effect, in terms of lower duration of cases, as a 0.6% increase in effort (about 9 more hearings per year). The question we address in this section is whether these estimates can be interpreted as evidence of inefficiency.

6.1 Private inefficiency

Let's start by considering *private* inefficiency, that is, a failure by the worker to optimally combine the available production factors. It is possible, and legitimate, to take the position that no such inefficiency can logically exist.[24] The argument here is one of revealed preferences. Under the assumption that the agent is maximizing his utility, whatever process is used by the agent to combine inputs must be interpreted as rational. If the agent appears to use the observable inputs ineffectively, the correct inference must be that the agent is economizing on an unobserved factor. Once the presence of the unobserved factor is taken into account, the agent can always be represented as operating on the production possibility frontier (PPF).

Let us follow this line of reasoning to its conclusion. Remember, our agents (judges) could produce more and simultaneously work less, simply by opening fewer cases. Let us model this scenario as a production function with two inputs, labor and task juggling. By

[24] See, for example, Stigler's (1976) argument against X-inefficiency.

introducing this additional factor, task juggling, we can now model agents as operating on the PPF. Our agents, apparently, are choosing a locus of the PPF where task juggling is not minimized. At their chosen locus, the productivity benefit of a 0.6% increase in labor equals the benefit of a 1% decrease in task juggling (we know this from our empirical estimates). If the chosen locus resulted from cost minimization for given level of productivity, the first order conditions would imply that the cost of an *additional* 0.6% in effort (about nine more hearings per year) would need to equal the cost of 1% *fewer* opened cases (about five less opened cases per year).

The difficulty with this conclusion is that it seems hard to claim that opening *less* cases has the same cost of doing *more* hearings. Remember that in order not to open a case the judge has simply not to hold its first hearing. In other words, optimality of the chosen locus would imply that not doing a (first) hearing to open a case has the same cost of doing more more (generic) hearings of other cases, which appears unlikely. Indeed, based on our conversations with judges, they do not seem to have a direct preference for opening more cases, nor do they suffer a direct cost for opening fewer cases. Moreover, once exposed to the possibility of holding fewer first hearings they recognize that this decision per se can hardly generate additional costs in terms of rescheduling or planning generic hearings for other cases.[25] Finally, they are instead unanimous in saying that holding more hearings is extremely costly because of the already tight hearings' schedule and the preparation effort that each hearing requires. If this analysis is correct, then the observed evidence is difficult to reconcile with the predictions of standard optimization.

This argument suggests that we should entertain the possibility that there is some private inefficiency. If this is the case, what accounts for the judges' failure to optimize? In our conversations with judges, we found that the great majority of them simply had not contemplated very carefully the issue of work organization. When asked to contemplate, most judges did not immediately grasp the benefits of reducing task juggling, and some judges to this day appear to believe (based on their postings on judicial mailing lists) that the benefits of reduced task juggling do not accrue when a judge is heavily overloaded (whereas, most likely, the opposite is true). Actually, two judges in a different court who recently started

[25]The reader might worry that, by opening fewer cases, a judge might risk being idle some of the time, i.e., not having enough cases on which to work. This is not a concern, however. Because of their busy calendars, judges routinely postpone hearings well beyond the time it takes for the parts to prepare for them. If judges opened fewer cases, these delays would simply get a bit shorter, but no judge would risk idleness. The data support this claim. Within our period of analysis, the number of active cases grows steadily, but the number of hearings per year is essentially stationary. This means that judges are not constrained by having too few cases on which to work; i.e, idleness is not a concern.

to follow our suggestions declared: "So far it does not seem that following this practice generates a significant increase in workload ... The benefit is not only the reduction of trial duration, but also the concentration of the trial that allows me to have well present and under control all the facts when I finally have to take a decision. This is better than when a long interval occurs between hearings." And: "In this way I have to work harder, but I am now convinced that this method will reduce the duration of my trials and such result will very well be worth the effort." These sentences suggests that excessive task juggling derives from a a failure to contemplate, rather than a conscious strategy for input deployment. So, to sum up, the evidence suggests that judges behave "as if" they failed to properly contemplate their private optimization problem.

One might point out that, while uncontemplative judges are not on "the" PPF, they are likely on *their individual* PPF, in the sense that they are on a PPF which includes a third input, their power to contemplate. Judges who use labor ineffectively are economizing on their power to contemplate. (According to this view, contemplation is comparable to technological progress which expands a judge's PPF). We are now, in our view, down to a question of terminology. We feel that in this particular case there is something a bit perverse in calling contemplation an input, akin to technological progress, and labeling the resulting misuse of labor "efficient." If the contemplation costs were large, as they might be in solving a difficult mathematical problem, then contemplation would properly be considered an input, akin to technological progress, and we would be comfortable with the notion that judges are, at any point in time, on their personal PPF. However, in our case the required contemplation is quite modest. It does not take great learning, or investment in human capital, to get the idea that doing too many things at one time slows everything down. It may take one 10 minutes to understand, and a day to convince oneself, but that's still a very small amount of effort compared with the very considerable gains. (Technical progress, in contrast, is generally defined as "new and non-obvious.") We feel, therefore, that a more proper interpretation is that the lack of contemplation reveals an imperfect optimization calculus. When optimization, in the sense of cost-benefit analysis, goes awry for lack of contemplation, there may well be the possibility of inefficiencies arising even in a single-agent setting. If something is privately inefficient when it takes exactly no resources to improve on it, then here we have a case where it takes a negligible amount of resources (contemplation) to improve significantly. We may, if we wish, call this as a "quasi-inefficiency." Language issues apart, we feel that this lack of contemplation accounts for at least part of the persistent mis-deployment of effort on the part of task juggling judges.

6.2 Social inefficiency

In the previous subsection we have argued that, whatever the judge is doing in terms of task juggling, she could be doing it better, where better is measured from her own private perspective. However, one could say that the judge would like to do it better, but is not allowed by external forces beyond her control. For example the pressure of lawyers who lobby for their case to be opened sooner by the judge, and who do not take into account the undesirable externality that this pressure generates on the duration of all other cases. Or the fear of sanctions associated with the failure to comply with rules that require judges not to wait too long before opening a case. To the extent that these external pressures are the reason why judges open too many cases, one could say that the cost of resisting these pressures is precisely the (unobserved) cost that judges take into account in planning optimally (from their private viewpoint) their work schedule. In other words, these external pressures would constitute the "unobserved factor" on which judges are economizing in an optimal way, and that, if not considered, gives an impression of private inefficiency where there is none. But even in this situation, it would still be the case that task juggling, albeit privately efficient, is socially inefficient. This because judges, in their choice of the optimal level of task juggling, would not consider the social externalities deriving from working on "too many" trials at the same time. In this section, we show that there are social consequences of task juggling which the judge may not internalize and we try to quantify them.

The social benefits from decreasing the duration of trials by one day, thanks to a 0.4% decrease of task juggling in a year[26], are equal to:

$$B = \begin{array}{c}\text{Social benefit}\\ \text{from}\\ \text{one day shorter trials}\end{array} \quad - \quad \begin{array}{c}\text{Judge's private costs}\\ \text{from}\\ \text{decreasing task juggling}\end{array} \qquad (7)$$

If we agree that the judge is carrying out task juggling to a degree which is either privately optimal or, as we argued in the previous section, privately excessive, it follows that the judge's marginal costs from decreasing task juggling must be smaller or equal than his private marginal cost of effort. Assuming that the judge's marginal cost of effort equals her wage, we can compute an upper bound for the second term in equation (7). This upper

[26]Our estimates suggest that a 1% decrease of task juggling reduces duration by 2.4 days; so a reduction of 1 day is obtained with 0.4% less task juggling.

bound equals 300 euros for a 0.25% increase in effort whose cost should be not lower than that of a 0.4% decrease of task juggling in a year, necessary to reduce trial duration by one day.[27]

We now turn to estimating the social benefit from shortening trial durations. We model a typical trial as a firing litigation. The details of this model are worked out in Appendix 8. In the model, litigation may arise when a firm, being greedy, offers a wage below the worker's (unknown) reservation value. Still, the match may be socially valuable. In this case, the worker can sue and then the judge will enjoin the firm to hire the worker. According to this model, the social benefits of a shorter trial are that an efficient match is created sooner rather than later. These benefits can be assessed as follows (on a daily basis):

$$Q - \underline{w} - \underline{\pi}, \tag{8}$$

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 worker. In Appendix 8 we compute these figures based on national accounts statistics and find that the yearly social benefit of a decrease by one day of the duration of trials equals 31675.07 euros. This number represents the first term in equation (7).

According to our back of the envelope calculation, therefore, the social benefits from decreasing task juggling in a year by an amount large enough to reduce trials by one day, are no smaller than $31675.07 - 300 = 31375.07$ euros per judge, which, to grasp its size, corresponds to about 25% of a judge's yearly salary.

7 Conclusions

We 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.

[27] A judge's yearly gross salary is approximately 120,000 euros (see http://jobspot.it/stipendio-lordo-magistrato). Therefore the cost of 0.25% of additional yearly effort equals 300 euros.

In order to identify the impact of tasks juggling on productivity, we construct time-varying instruments based on the sample realization of the lottery that allocates cases to each judge. The effect that we measure is 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 2.4 days and would need to be compensated by a 0.6% increase of effort in order to keep the duration of trials constant. We believe 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 (percent 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 *less* 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 (0.4%) to cut the duration of her trials by one day, this would cause a social gain of about thirty thousand euros, which corresponds approximately to one fourth of a judge's yearly income.

We have derived our results for the specific setting of Italian judges, 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. We view the analysis in this paper and its companion (Coviello et al. 2013) as a first step into the empirical and theoretical analysis of how work scheduling affects the individual production function.

8 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 iff $Q > \underline{w} + \underline{\pi}$.

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 that in 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.

8.1 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 ony 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

$$\begin{aligned} & f(w)(Q - w - \underline{\pi}) - F(w) \\ = \quad & f(w)\left[(Q - w - \underline{\pi}) - \frac{F(w)}{f(w)}\right]. \end{aligned}$$

The second order conditions are satisfied if, for example, $\frac{F(w)}{f(w)}$ 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

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

that is, our equilibrium wage offer. (In the case of the uniform, 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

$$w^* < \underline{w} < Q - \underline{\pi} \tag{9}$$

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

$$\frac{Q - \underline{\pi}}{2} < \underline{w} < Q - \underline{\pi}. \tag{10}$$

If the uniform is between 0 and 1, this event has probability $\frac{Q-\underline{\pi}}{2}$. 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 RHS of (10), 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.

8.2 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 ony 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

$$\begin{aligned} & -wg(Q - w) + G(Q - w) + g(Q - w)\,\underline{w} \\ = \;& g(Q - w)\left[-w + \frac{G(Q - w)}{g(Q - w)} + \underline{w}\right] \end{aligned}$$

The second order conditions are satisfied if, for example, $\frac{G(w)}{g(w)}$ 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

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

that is, our equilibrium wage offer. (In the case of the uniform, 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, for example, we have an inefficiency when

$$\frac{Q - \underline{w}}{2} < \underline{\pi} < Q - \underline{w}.$$

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

8.3 The social benefit of one less 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

$$Q_t - \underline{w}_t - \underline{\pi}_t, \tag{11}$$

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 ISTAT[28] the daily value added of a full time average worker is 171.69 euros, which we take it as a proxy for Q_t.

Slightly more complicated is to compute $\bar{W}$. During a firing litigation a worker can apply for a job at a different firm and take unemployment benefits for at least one 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

[28]See www.istat.it. All the other statistics used below come from either ISTAT or the Bank of Italy.

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 we have the 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 one. Using again national statistics provided by ISTAT, the average daily wage is 102.53 euro, which is our estimate for $\bar{W}$. We emphasize that this is most likely an upper bound of a realistic outside option, 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 to over estimate 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 euros. The question is what the firm can do with this capital per worker during the litigation. It is important to consider that the Italian labor law foresee that a fired worker must be re-hired by the firm if the judge rules that the firing was unjustified. The risk of having to re-hire 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 $\bar{\Pi} = 6.93$ euros.

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

$$Q_t - \bar{W}_t - \bar{\Pi}_t = 171.69 - 102.53 - 6.93 = 62.23 \tag{12}$$

Given that a judge receives on average 509 new cases per year, the total benefit for society of one less litigation day on all the trials of the year is 31675.07 euro (=62.23*509).[29]

[29]This calculation rests on the simplifying assumption that the social benefit of one less day of fring litigation approximates the average social beneft of all other trials, for which the calculation would be more diffcult because of less easily obtainable empirical counterparts.

References

Allen, David (2001). Getting Things Done. Viking.

Angrist D. J., and J. S. Pischke, 2008. Mostly Harmless Econometrics. An Empiricist's Companion. Princeton University Press.

Amemiya, T., 1982. Two Stage Least Absolute Deviations Estimators. Econometrica, 50, 689-711.

Ameriks, J., Caplin, A., and J. Leahy, 2003. Wealth Accumulation And The Propensity To Plan. Quarterly Journal of Economics, 118, 1007-1047.

Aral, S., Brynjolfsson, E., and M., Van Alstyne, 2007. Information, Technology and Information Worker Productivity. NBER WP., 13172.

Ashenfelter, O., T. Eisenberg, and S. J. Schwab, 1995. Politics and the Judiciary: The Influence of Judicial Background on Case Outcomes. The Journal of Legal Studies, 24, 257-281.

Bandiera, O., Guiso, L., Prat, A., and R. Sadun, 2008. What CEOs do. Manuscript, London School of Economics.

Victoria Bellotti, 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*, pp.735-742, April 24-29, 2004, Vienna, Austria.

Bertrand, M., and A., Schoar, 2003. Managing with Style: The Effect of Managers on Firm Policies. Quarterly Journal of Economics, 118, 1169-1208.

Bloom, N., and J., Van Reenen, 2007. Measuring and Explaining Management Practices Across Firms and Countries. Quarterly Journal of Economics, 122, 1351-1408.

Bloom, N., Propper, C., Seiler, S., and J., Van Reenen, 2009. The Impact of Competition on Management Practices in Public Hospitals. Manuscript, London School of Economics.

Bohn, Roger (2000) Stop Fighting Fires. Harvard Business Review, 78(4) (July-Aug 2000), pp. 83-91 .

Chernozhukov, V., and C., Hansen, 2008. Instrumental Variables Quantile Regression: A Robust Inference Approach, Journal of Econometrics, 142, 379-398.

Chesher, A., 2003. Identification in Nonseparable Models, Econometrica. 71, 5, 1405-1441

Covey, Stephen (1989). The seven habits of highly effective leaders. New York: Simon & Schuster.

Cob, C. W., and P. H., Douglas, 1928. A theory of production, American Economic Review, March 18(2), 139-165.

Coviello, D., Persico, N., and A., Ichino, 2013. Time Allocation and Task juggling. Forthcoming: American Economic Review.

CSM, 2010: CONSIGLIO SUPERIORE DELLA MAGISTRATURA, Quarta Commissione. Gruppo di lavoro per la individuazione degli standard medi di definizione dei procedimenti. Relazione 2010.

Della Vigna, S., 2009. Psychology and Economics: Evidence from the field, Journal of Economic Literature, 47, 315-372.

Dewatripont, Mathias, Ian Jewitt, and Jean Tirole, 1999. The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies. The Review of Economic Studies, 66(1) pp. 199-217.

Frantz, R., 1992. X-Efficiency and Allocative Efficiency: What Have We Learned?. The American Economic Review, Vol. 82, No. 2, Papers and Proceedings of the Hundred and Fourth Annual Meeting of the American Economic Association, pp.434-438

Garicano, L., and P., Heaton, 2010. Information Technology, Organization, and Productivity in the Public Sector: Evidence from Police Departments. Journal of Labor Economics, 28, 167-201.

Gibbons, R., and J., Roberts, 2010. Handbook of Organizational Economics. Princeton University Press.

González,Victor M. and Gloria Mark, 2005. Managing Currents of Work: Multi-tasking Among Multiple Collaborations. In H. Gellersen et al. (eds.), *ECSCW 2005: Proceedings of the Ninth European Conference on Computer-Supported Cooperative Work*, 18-22 September 2005, Paris, France, 143–162.

Griliches Zvi, and Jerry Hausman, 1986. Errors in Variables in Panel Data. Journal of Econometrics, 31,93-118.

Holmstrom, Bengt and Paul Milgrom, 1991. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. Journal of Law, Economics, and Organization, Vol. 7.

Ichniowski, C., Shaw, K., and G., Prennushi, 1997. The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines. American Economic Review, 87, 291-313.

Jackson, James R., 1963. Jobshop-like Queueing Systems. Management Science, Vol. 10, No. 1 (Oct., 1963), pp. 131-142.

Jorgenson Dale W., 1986. Econometric methods for modeling producer behavior. in "Handbook of Econometrics" Volume III, Edited by Zvi Griliches and M. D. Intriligator, Elsevier Science Publisher.

Kling, J.R., 2006. Incarceration Length, Employment, and Earnings. The American Economic Review, 96, 3, 863-876.

Lusardi, A., and O., S., Mitchell, 2008. Planning and Financial Literacy. American Economic Review: Papers and Proceedings, 98, 413-417.

Mark, G., D. Gudith, and U. Klocke, 2008, The Cost of Interrupted Work: More Speed and Stress. In *CHI '08: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 107–110, Florence, Italy, April 2008. ACM Press.

Moder,J, C Phillips, E Davis (1983). *Project management with CPM, PERT and project diagramming.* Van Nostrand Reinhold Publishers.

Perlow, Leslie (1999). The time famine: Toward a sociology of work time. Administrative Science Quarterly, 44 (1), pp. 57–81

Repenning, Nelson (2001) Understanding fire fighting in new product development. Journal of Product Innovation Management 1, pp. 85–300.

Stock, J., H., and M., Yogo, 2005. Testing for Weak Instruments in Linear IV Regression. Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. Cambridge University Press, 80-108.

Stigler, G., J., 1976. The Xistence of X-Efficiency. The American Economic Review, Vol. 66, No. 1 , pp. 213-216.

Table 1: The panel structure

Judge identifier	Number of quarters of service per year 2000	2001	2002	2003	2004	2005	Total number of quarters of service	Average number of new cases per quarter
1	4	4	4	4	4	0	20	107
3	4	4	1	0	0	0	9	105
5	4	4	4	4	4	4	24	143
6	4	4	4	4	4	0	20	129
7	4	4	4	4	4	2	22	118
8	4	1	4	4	4	0	17	119
9	4	4	1	0	0	0	9	110
10	4	4	4	2	0	0	14	118
11	4	4	4	4	4	4	24	141
12	4	4	4	2	4	4	22	138
13	4	4	4	4	4	2	22	120
14	4	4	4	2	0	0	14	125
15	4	4	4	4	4	0	20	127
18	0	0	0	0	3	4	7	152
19	2	4	4	4	2	4	20	122
20	4	4	4	4	4	4	24	137
21	4	4	4	4	4	4	24	120
22	4	4	4	4	4	4	24	138
24	4	4	4	4	4	4	24	135
29	0	0	0	2	4	4	10	150
30	0	0	0	3	4	4	11	121
Total (average in last col)	70	69	66	63	65	48	381	128

Table 2: Variability of assignments per quarter across judges

Quarter of observation	New cases per judge Average	St. Dev.	Number of judges
2000q1	129	13	18
2000q2	112	11	18
2000q3	82	7	17
2000q4	120	22	17
2001q1	137	20	17
2001q2	134	11	17
2001q3	120	14	17
2001q4	123	21	18
2002q1	134	30	18
2002q2	149	19	16
2002q3	100	11	16
2002q4	144	17	16
2003q1	147	19	16
2003q2	139	21	16
2003q3	108	12	15
2003q4	131	29	16
2004q1	139	17	15
2004q2	151	23	16
2004q3	108	23	17
2004q4	114	31	17
2005q1	123	28	13
2005q2	155	43	13
2005q3	132	18	11
2005q4	161	33	11
Average	128	28	17

Table 3: Tests for the random assignment of cases to judges

Quarter of observation	Type of controversy	Zip code of plaintif's lawyer	Number of involved parties	Number of Judges
2000q1	.089	.052	.003	18
2000q2	.003	.095	.065	18
2000q3	.230	.150	.039	17
2000q4	.045	.015	.000	17
2001q1	.430	.000	.330	17
2001q2	.000	.610	.420	17
2001q3	.760	.670	.660	17
2001q4	.770	.610	.830	18
2002q1	.032	.140	.410	18
2002q2	.130	.570	.270	16
2002q3	.048	.180	.270	16
2002q4	.008	.057	.016	16
2003q1	.720	.410	.410	16
2003q2	.620	.770	.000	16
2003q3	.350	.058	.400	15
2003q4	.120	.098	.033	16
2004q1	.850	.470	.780	15
2004q2	.950	.800	.950	16
2004q3	.190	.100	.040	17
2004q4	.140	.340	.960	17
2005q1	.580	.230	.095	13
2005q2	.004	.810	.450	13
2005q3	.660	.430	.360	11
2005q4	.160	.510	.490	11
N. of rejections	7	2	7	.
Largest balanced panel	.11	.22	.073	.

Note: The top part of this table reports, for each quarter, the p-value of a Chi-square test of indipendence between the identity of judges and three discrete characteristics of cases: type of controversy (14 types); zip code of the platintif's lawyer (55 codes); the number of parties in trial (capped at 10). The central part of the table reports the number of quarters in which independence is rejected at the 5% level. The bottom part of the table reports similar Chi-square tests as in the top part, for all cases assigned in the period spanned by the largest balanced panel of judges identifiable in our sample. As shown in Table 1 this largest panel involves 14 judges observed continuously between year 2000 and year 2002.

Figure 1: Differences of performance between judges with randomly assigned workload

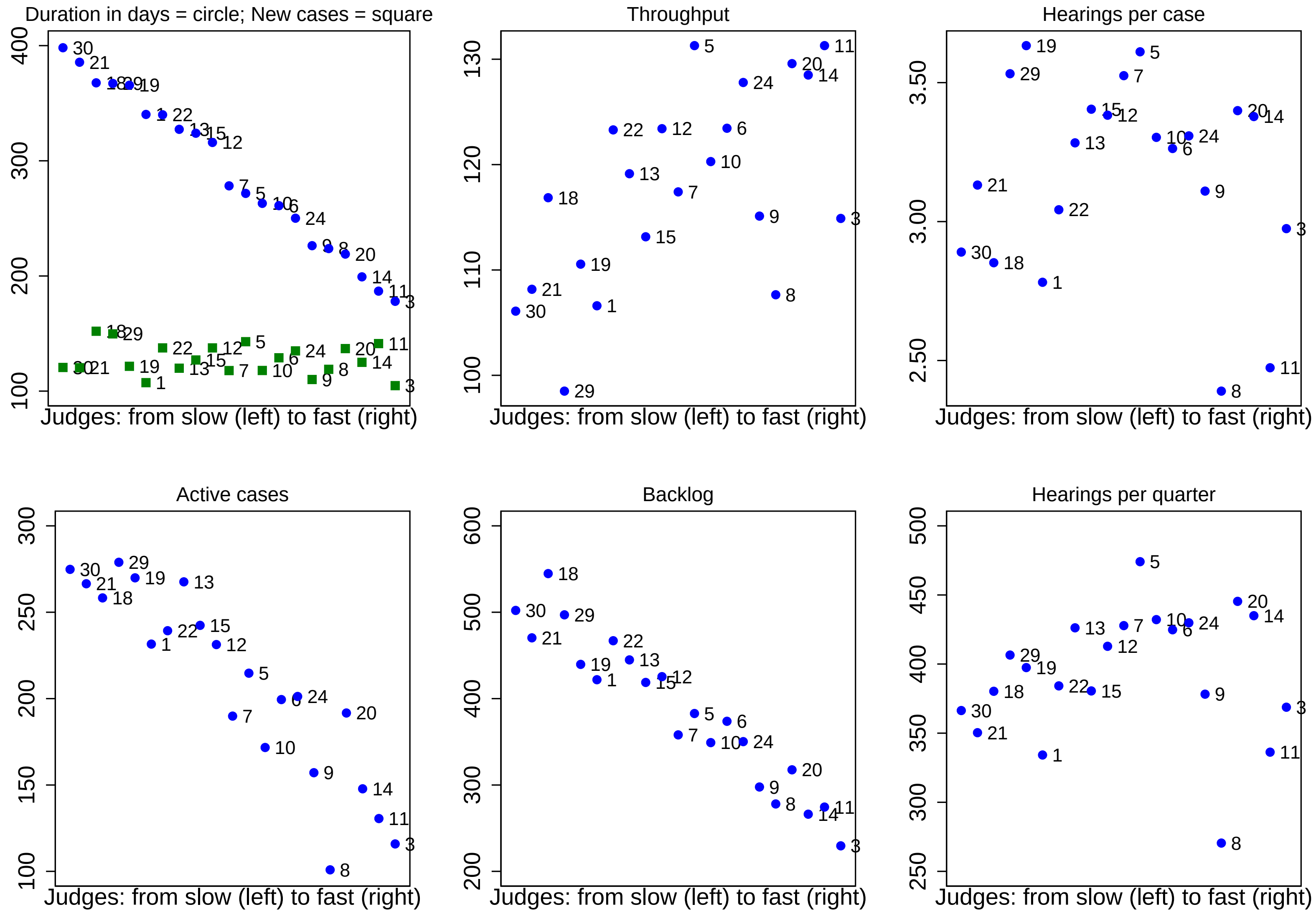

Figure 2: The trade off between quantity and quality in the decision of judges

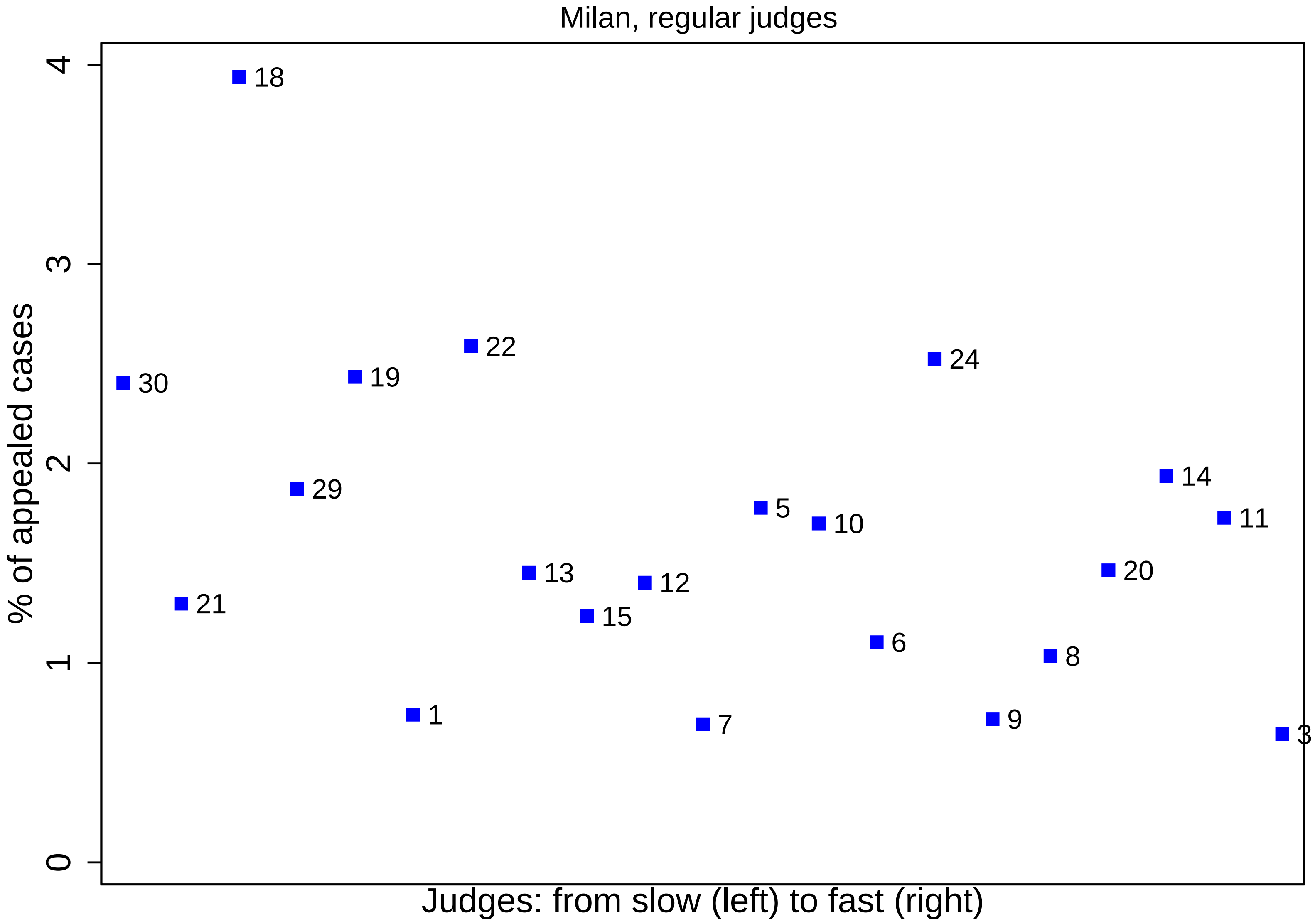

Figure 3: Work flow in a stable rotation

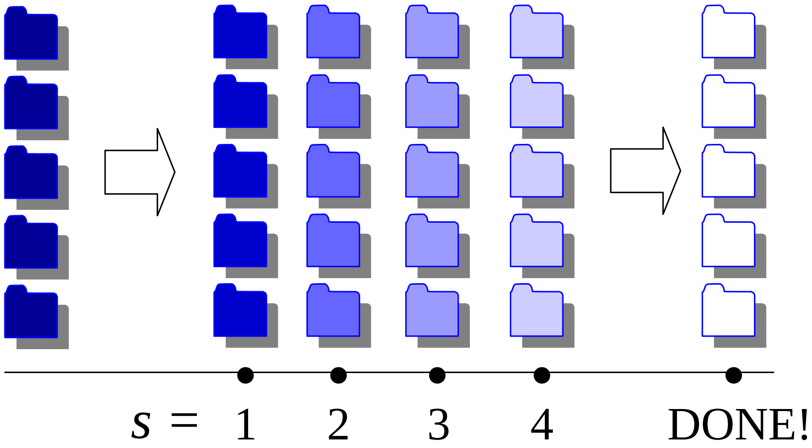

Figure 4: How far are judges from a stable rotation?

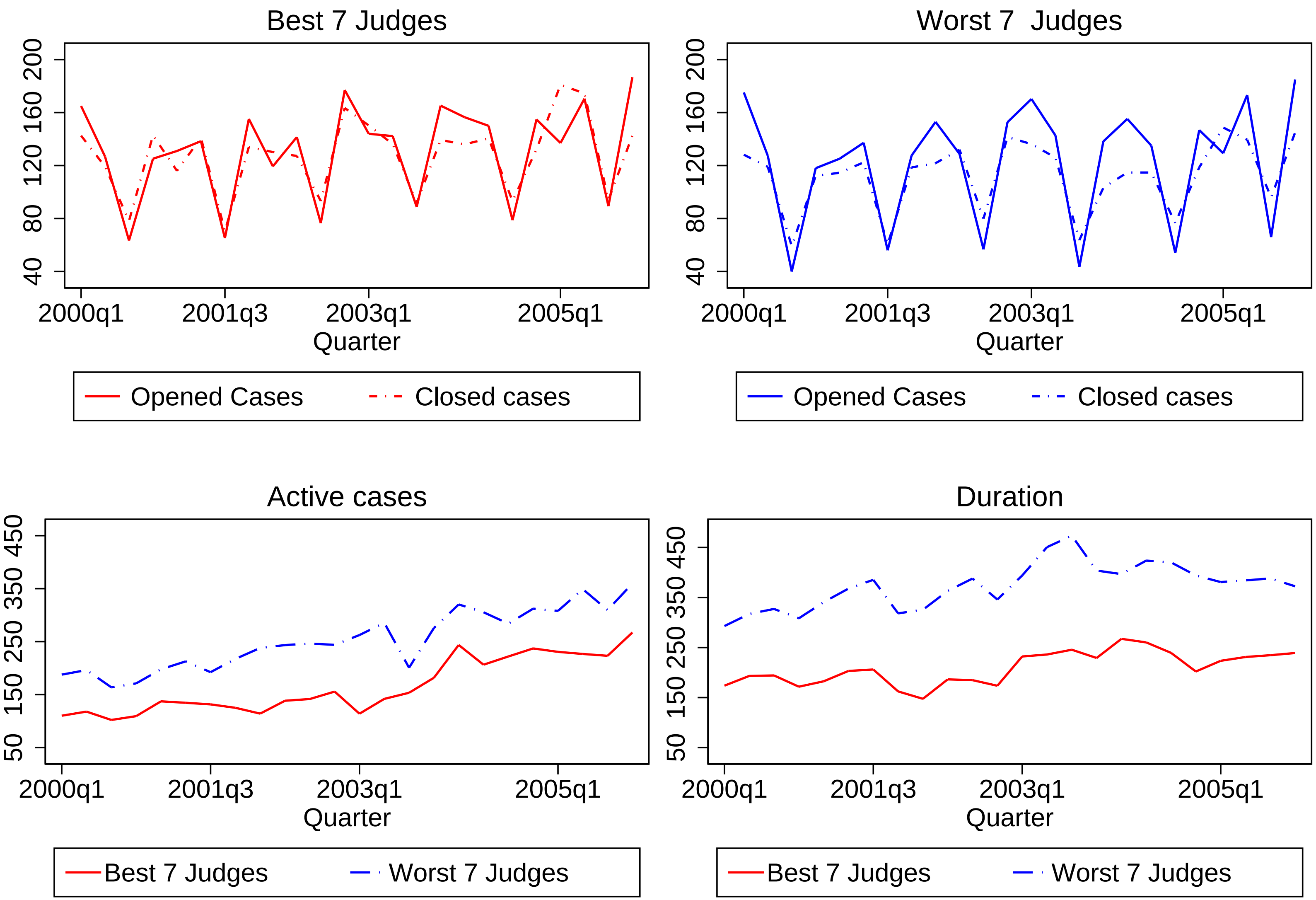

Figure 5: Deviation from a stable rotation

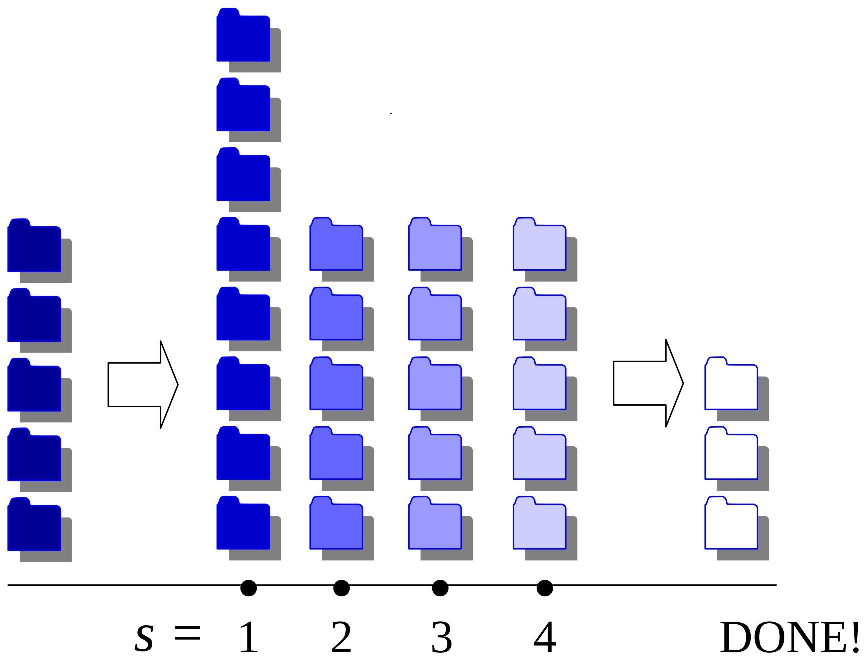

Table 4: Descriptive statistics

	Mean	sd	p25	p50	p75	n
Dependent variables for equation (6)						
$D_{i,t}$: Total duration in days (from filing to disposition)	276	237	124	210	352	50412
Regressors for equation (6)						
$\rho^p i, t$: Number of red code cases assigned in previous year	105	41	68	112	139	50412
$\gamma^p i, t$: Number of green code cases assigned in previous year	404	64	379	416	446	50412
$\nu^f_{i,t}$: Number of cases opened in future year	486	90	443	502	547	50412
$(\frac{e}{S})^f_{i,t}$: Standardized effort in future year	518	90	463	518	579	50412
Instruments for equation (6)						
$\rho^f i, t$: Number of red code cases assigned in future year	109	41	82	115	141	50412
$\gamma^f i, t$: Number of green code cases assigned in future year	357	112	329	396	433	50412
Components of standardized effort						
$e^f_{i,t}$: Number of hearings held in future year	1560	284	1399	1580	1769	50412
$S^f_{i,t}$: Number of hearings to close cases assigned in future year	3.1	.55	2.7	3.2	3.4	50412

Note: Descriptive statistics of the 50412 trials assigned to 21 full-time judges of the Labor Court of Milan between January 1, 2000 and December 31, 2005. "Previous year" is the period of 365 days preceeding the date t in which a case is assigned. "Future year" is the period of 365 days following the date t in which a case is assigned. Standardized effort in future year is defined as the ratio between the number of hearings held by the judge in the future year and the number of hearings that were necessary to decide the cases assigned to the judge in the future year. This ratio can be interpreted as the potential number of cases, among those assigned in the future year, that a judge could have completed in the same period with the effort that he/she actually exerted.

Table 5: The effect of task juggling on the duration of a trial

Estimation Method	OLS	IV	OLS	OLS
Dependent variable	$D_{i,t}$	$D_{i,t}$	$\nu_{i,t}^{f}$	$(\frac{e}{S})_{i,t}^{f}$
$\nu_{i,t}^{f}$: Number of cases opened in future year	0.30	0.50		
	(0.016)	(0.079)		
$(\frac{e}{S})_{i,t}^{f}$: Standardized effort in future year	-0.37	-0.84		
	(0.018)	(0.193)		
$\rho_{i,t}^{p}$: Number of red code cases assigned in previous year	0.06	0.26	0.09	0.64
	(0.059)	(0.104)	(0.011)	(0.017)
$\gamma_{i,t}^{p}$: Number of green code cases assigned in previous year	0.01	0.08	0.39	0.28
	(0.024)	(0.039)	(0.004)	(0.008)
$\rho_{i,t}^{f}$: Number of red code cases assigned in future year			0.60	0.54
			(0.010)	(0.017)
$\gamma_{i,t}^{f}$: Number of green code cases assigned in future year			0.56	0.12
			(0.004)	(0.006)
Judge fixed effects	YES	YES	YES	YES
Year and month-of-the-year fixed effects	YES	YES	YES	YES
Observations	50,412	50,412	50,412	50,412
Number of judges	21	21	21	21
R2	0.0958	0.0855	0.842	0.607
Cragg-Donald Wald F statistic		251.02		

Note: In columns 1 and 2 the dependent variable $D_{i,t}$ is the duration (from filing to disposition) of a case assigned at date t. "Previous year" is the period of 365 days preceeding the date t in which a case is assigned. "Future year" is the period of 365 days following the date t in which a case is assigned. Standardized effort in future year is defined as the ratio between the number of hearings held by the judge in the future year and the number of hearings that were necessary to decide the cases assigned to the judge in the future year. This ratio can be interpreted as the potential number of cases, among those assigned in the future year, that a judge could have completed in the same period with the effort that he/she actually exerted. The "Cragg-Donald Wald F statistic *(Joint)*" denotes the minimum eigenvalue of the joint first-stage F-statistic matrix. Specifically, the F statistic reported in column 2 is for the endogenous variables $\nu_{i,q}^{f}$ and $(\frac{e}{S})_{i,q}^{f}$, whose first stage estimates are reported in columns 3 and 4. The dependent variables in these columns are the instruments: respectively, the number of assigned red code cases in future year $\rho_{i,t}^{f}$ and the number of assigned green code cases in future year $\gamma_{i,t}^{f}$. Robust standard errors are in parentheses.