Measuring the gains from labor specialization: theory and evidence*

Decio Coviello
HEC Montréal

Andrea Ichino
EUI

Nicola Persico
Northwestern

December 17, 2017

Abstract

We estimate the productivity effects of labor specialization using a judicial environment that offers a quasi-experimental setting well suited to this purpose. Judges in this environment are randomly assigned many different types of cases. This assignment generates random streaks of same-type cases which create mini-specialization events unrelated to the characteristics of judges or cases. We estimate that when judges receive more cases of a certain type they become faster, i.e., more likely to close cases of that type in any one of the corresponding hearings. Quality, as measured by probability of an appeal, is not negatively affected. We conclude that the channel through which these effects operate is learning-by-doing and that it can be generalised to other types of jobs

*This research was conducted in collaboration with the training-unit of the Court of Roma. We are grateful to Roman Acosta for outstanding research assistance; Amelia Torrice and Margherita Leone for feedbacks on early versions of the manuscript. This research was undertaken, in part, thanks to funding from the Canada Research Chairs program. The usual caveats apply.

1 Introduction

The productivity-enhancing effects of specialization have been a classic theme in economics since at least Adam Smith. While it is a truism that *some* specialization enhances productivity, it is also true that most jobs are by definition somewhat specialized, so the meaningful empirical question is whether further specialization helps *at the margin*, that is, whether there are any *unexploited* gains from specialization.

A large empirical literature estimates the gains from specialization in professions as different as surgeons, school teachers, and clerks. This literature has had to confront two key identification issues. First, *workers are in general not randomly exposed to specialization*: they choose, or are selected into their specialty. Second, the measurement of the benefits from specialization might be biased if *unobservable task characteristics influence the type and extent of specialization of the worker to which the task is assigned*. Some papers in the literature reviewed below address one source of endogeneity, but no paper that we know of addresses both. In this paper we are able to address both identification concerns due to the explicitly random process through which workers (judges, in our case) are assigned tasks.

In our setting, a computer (which, incidentally, takes no account of the judges' backlogs) randomly assigns cases to judges. This means that, occasionally, a judge will be assigned a disproportionate number of cases of a given type – Pension cases, for example. These random occurrences will periodically result in situations when a judge's docket is rich with cases of that same type, which means that *a judge is randomly exposed to specialization*. Also, the random assignment of cases ensures that unobservable task characteristics *are assigned orthogonally to the judges' specialization*. We leverage this uniquely favorable identification scenario to obtain estimates of the productivity-enhancing effects of specialization.

We estimate whether our workers get any faster and more accurate on type-A tasks when they are assigned many type-A tasks. A model is required to go from these estimates to the gains from specialization. The theory section presents such a model starting at a general level, and then specializing to the case where team production is the sum of individual workers' production functions with a convenient parametric functional form. The analysis yields mathematical conditions on the parameters of these functions such that returns from specialization are positive.

We find that judges indeed do get faster (more likely to close a case in any given hearing) during those times when their docket is rich with cases of that same type. We also find that, all else equal, having more *other case types* actually slows down the judge. As for accuracy, as best we can measure we find that more-specialized (in the above sense) judges are not differently accurate, in that we find that their decisions don't get appealed at a higher or lower rate.

After a review of the literature in Section 2, we present the theory in Section 3. Section 4 describes the data and the institutional setting, while the empirical model is presented in Section 5. Results are discussed in Section 6, where the regression estimates are translated, using the theoretical model, in an assessment of the gains from specialization. In Section 7 we show why learning-by-doing is the most likely reason of these gains. Section 8 concludes.

2 Related Literature

There is a large literature on labor specialization in many different fields. A first relevant groups contains studies of the impact of volume of surgery and specialization on patient outcomes. A meta-analysis of this literature (Chowdhury et al., 2007) finds that high-volume and specialist surgeons have significantly better outcomes (in 74 and 91 percent of the studies, respectively). However, of the 163 studies covered in this meta-study, none were randomized.[1]

KC and Staats (2012) and KC et al. (2013) study heart surgeons. After controlling for a great deal of patient characteristics, KC et al. (2013) find that experience (cumulative procedure volume) improves patient outcomes, and whereas past successes improve a surgeon's outcomes, past failures worsens them. KC and Staats (2012) partitions experience into "focal," that is, closest to the procedure at hand, and "related," more distant types of procedures, and finds that focal experience improves surgical outcomes more so than related experience. Staats and Gino (2012) use data from a home loan

[1] The authors note that "It is unlikely that randomized controlled trials will ever take place" to evaluate the effects of specialization (p. 145).

application-processing line to inquire about the effect of specialization on the productivity of data-entry clerks. They find that over the course of a single day, specialization, as compared to variety, improves worker productivity (this notion of very short-term specialization may be akin to "batching"), but when the workers' experience is examined across several days, variety appears to improve worker productivity.

Narayan et al. (2016) study the productivity of software engineers who perform maintainance tasks on different modules of a complex software; they find that experience with a given module improves productivity. Friebel and Yilmaz (2016) compare the productivity of call center agents who are "less specialized," i.e., have a greater number of certified "skills" and are more experienced, with "more specialized" agents (fewer skills, shorter tenure). Ost (2104) and Cook and Mansfield (2016) use an administrative panel of teachers rotating across subjects to parse out the relative contribution of general or subject-specific experience to productivity.

None of these papers can leverage random assignment as a source of specialization; that is, unlike our judges, these workers are not "randomly exposed to specialization." Relative to these papers, our work is unique in that it leverages an explicitly random assignment procedure *both for identifying exogenous variation in specialization and for random assignment of jobs to differently-specialized workers*. In addition, of course, the settings are different: judicial performance has great societal impact in its own right, and the findings on other occupations are not especially informative about judicial performance. Therefore our paper complements the existing literature, it does not compete with it.

We now review the literature on judicial specialization. The judicial profession is slowly specializing (see Baum 2011). But this trend is controversial because specialization is perceived to have pros and cons. Baum (2009, sec. III) discusses the pros (speed, accuracy, and uniformity) and cons (excessive assertiveness, insularity, tendency to stereotype, narrow selection into the judicial profession, vulnerability to capture by specialized interest groups) of judicial specialization. The analysis in this paper aims to quantify the first two pros: speed and accuracy.

Apart from many qualitative articles, a number of empirical analyses exist regarding the effect of specialization or experience on different measure of judicial productivity (Miller and Curry 2009; Hansford 2011; Kesan and

Ball 2011; Sustersic and Zajc 2011). These papers do not exploit exogenous variation in specialization. Our paper adds to this literature by exploiting the random assignment of cases to judges for identification.

Moving away from specialization as the explanatory variable, a number of papers study other determinants of judicial productivity. Djankov et al. (2003) argue that cross-country differences in the effectiveness of judicial systems depend primarily on the level of procedural formality of legal systems. Dimitrova-Grajzl et al. (2012) use an internal instrument to assess how judicial staffing levels impact court productivity. Bagues and Esteve-Volart (2010) study the effects of introducing incentive pay for judges, and find a complex set of effects on judicial productivity. Ash and McLeod (2014, 2016) study how the performance of US judges depends on their case load, on their tenure, and on their electoral incentives.

In previous work (Coviello et al. 2014, 2015; Bray et al. 2016) we have shown that judicial workflow management practices, and in particular multitasking, can have a significant impact on judicial productivity. This line of work is distinct from the present paper because workflow management refers to the efficient (or not) scheduling of individual hearings of different cases, whereas the present paper looks at the probability of closing a case *in a given hearing*, that is, *conditional on how the workflow has been managed.* [2]

Stepping back from judicial productivity as the outcome of interest, a number of studies have exploited the random assignment of cases to judges for identification in a variety of economic settings: see e.g. Ashenfelter et al. (1995), Kling (2006), Di Tella and Schargrodsky (2013). In addition, some recent papers explore impact of judicial reforms on a variety of economic outcomes (Lilienfeld-Toal et at. 2012, Ponticelli and Alencar 2016); this literature is only peripherally related to our work insofar as it demonstrates the judicial performance impacts economic growth.

[2]To see the difference, consider two cases A and B each of which require at most two hearings to conclude. Cases A and B are adjudicated in their first hearing with probabilities $p_{1,A},\ p_{1,B} < 0$, else a second hearing is necessary. In previous work (Coviello et al. 2014, 2015; Bray et al. 2016) we have shown that it is more efficient to wait until case A is adjudicated before starting on case B. This is workflow management. In the present paper, we ask whether $p_{1,B}$ gets larger owing to the fact that the judge has accumulated experience by working on case A.

3 Theory of labor specialization

This section presents a theory of team production, and then specializes the theory to the case where team production is the sum of individual workers' production functions. Then, a convenient parametric functional form is proposed for the individual workers' production functions, and mathematical conditions are sought on the parameters of these functions such that productivity improves if workers specialize in tasks of different types.

This setting covers many types of team productions. A classic example would be Adam Smith's pin factory, where different workers are each assigned different tasks (drawing out the wire for a single pin, straightening the wire for that pin, cutting it, etc.). In this case performance will be measured by how quickly and accurately each task is accomplished. Alternatively, a team could be a hospital surgery practice where surgeons might each specialize in different procedures (knee replacement, hip replacement, etc.), or a court where judges might each specialize in certain types of cases (labor cases, pension cases, etc). For a judge, a task might be a hearing of a given case type, and with each hearing of that case, the performance measure is the probability that the case is adjudicated in that hearing, as well as the probability that the decision is appealed.

There are J workers indexed by j. There are K task types indexed by k. Task type k has numerosity N_k. The total number of tasks is N. A worker j's total workload is fixed at N_j with the stipulation that $\sum_j N_j = N = \sum_k N_k$.

Let $n_{j,k}$ denote the number of type-k tasks allocated to worker j. We wish to allocate tasks to workers so as to maximize some objective function, for example, number of tasks accomplished in a certain time interval, or number of non-mishandled tasks (if performance quality is an issue). We denote the objective function by $f(\mathbf{n})$, where $\mathbf{n}$ is the vector with generic element $n_{j,k}$.

Our problem is:

$$\max_{\mathbf{n}} f(\mathbf{n}) \quad \text{subject to:} \tag{1}$$

$$\sum_k n_{j,k} = N_j \text{ for all } j \text{ (a judge's workload is fixed at } N_j\text{)} \tag{2}$$

$$\sum_j n_{j,k} = N_k \text{ for all } k \text{ (exactly } N_k \text{ cases are allocated)} \tag{3}$$

$$n_{j,k} \geq 0 \text{ for all } j, k \tag{4}$$

There is a natural sense in which the convexity of f captures the returns to specialization. If a strictly convex f is being maximized over some convex set X, then the maximizer(s) must be extremal, that is, they must lie at the boundaries of the set X. Extremal allocations captures "division of labor," in a sense made precise in the following proposition.

Proposition 1 ***(If f is quasi-convex it is optimal to specialize)*** *Suppose the objective function f is strictly quasi-convex. Then in the solution to problem (1) there cannot be two workers who are assigned positive amounts of the same two task types.*

Since quasi-convexity is a less restrictive condition than convexity, the following corollary holds.

Corollary 1 *If f is strictly convex it is optimal to specialize.*

In spirit, this proposition says that if f is quasi-convex then it is optimal for each worker to be fully specialized in a single case type. But this statement can't literally hold for all workers due to integer problems. So, the more nuanced statement contained in the proposition is this: if two workers are assigned a positive amount of a given (same) task type, then there can be no other task type that these two workers have in common. The following simple example illustrates the content of Proposition 1.

Example 1 ***(Illustration of Proposition 1 with two task types, two workers.)*** *There are 50 type-1 tasks, 50 type-2 tasks, and f is strictly quasi-convex (there are gains from specialization). Each worker can handle exactly*

50 tasks. Then optimality requires full specialization, that is: either worker 1 gets all the type-1 tasks and worker 2 all the type-2 tasks, or vice versa. To see this, suppose not. Then both workers must get a positive amount of both task types. But this contradicts Proposition 1.

Suppose instead that there are 60 type-1 tasks and 40 type-2 tasks. Then at the optimal allocation one of the workers must receive two types of tasks, but then by Proposition 1 the other worker must be fully specialized (in type-1 tasks, of course).

Next we provide a specific (and strictly convex, depending on parameters) functional form for the function $f(\mathbf{n})$. We want this functional form to be parsimonious, and yet to allow for learning-by-doing effects. Our basic building block is a type-specific productivity factor P^k. When this framework is applied to judges, P^k will stand for the probability with which judge j resolves a case of type k in a given hearing or, alternatively, for the probability that a case k is not appealed conditional on it being resolved. We posit that P^k depends on how many other type-k and non-type-k tasks the worker is assigned, as follows:

$$P^{j,k}(n_{j,k}, n_{j,-k}) = C_k + \gamma_j + n_{j,k}\beta_{same} + n_{j,-k}\beta_{other}, \tag{5}$$

where $n_{j,-k}$ denotes the number of non-type-k tasks assigned to the worker:

$$n_{j,-k} \stackrel{\text{def}}{=} \sum_{\kappa \neq k} n_{j,\kappa} \ .$$

If $\beta_{same} > 0$ then workers become more productive on type-k tasks by being assigned more tasks of that same type; we expect β_{same}'s estimates to be nonnegative. If $\beta_{other} > 0$ then workers get better at type-k tasks by being assigned more non-k tasks; so there is some transferability in experience across task types. If $\beta_{other} < 0$ then being assigned more non-A tasks *for given amount of A tasks* hurts a worker's productivity on type-A tasks. This might happen if the worker's memory is a finite repository that can only hold so much knowledge, and that memory is used in proportion to the type of tasks that she is assigned. We assume $C_k + \gamma_j > 0$ to ensure that even an inexperienced worker (one for whom $n_{j,k}$ and $n_{j,-k}$ equal zero) has a positive productivity.

We assume that our objective function has the following functional form:

$$f(\mathbf{n}) = A \sum_j \sum_k n_{j,k} P^{j,k}(n_{j,k}, n_{j,-k}), \tag{6}$$

where A is a positive constant. $f(\mathbf{n})$ represents the total production achieved by the entire pool of workers. Note that this function has curvature in $n_{j,k}$ even though $P^{j,k}(\cdot)$ is a linear function.

Later in the paper we will use the function $f(\mathbf{n})$ to measure two different dimensions of judicial productivity: how many cases all judges closes in a given number of hearings, and separately, how many judicial decisions are appealed. The objective function $f(\mathbf{n})$ is sufficiently flexible to capture both dimensions of productivity. If we let $P^{j,k}$ represent the "probability that a decisions is not appealed," then $f(\mathbf{n})$ represents the total number of non-appealed decisions (which it is socially desirable to maximize). Alternatively, $P^{j,k}$ may represent the "probability of closing a case in a given hearing," in which case we would like the functional form to represent the total number of decisions achieved by all judges; however, in order for this interpretation to be valid there is a gap that needs to be bridged. The gap is that our empirical counterpart for $(n_{j,k}, n_{j,-k})$ will be *number of cases*, but $P^{j,k}$ will be estimated as the probability of concluding a case within a given *hearing*. Therefore, the term $n_{j,k}$ that multiplies $P^{j,k}$ in (6) should be measured in hearings, not cases. As there are roughly 3 hearings to each case, setting $A = 3$ allows us to interpret (6) as the total amount of decisions produced by all judges within a certain number of hearings.

When objective function (6) is convex, its maximizers are extremal per Proposition 1. The next proposition spells out sufficient conditions for convexity.

Proposition 2 ***(Sufficient conditions for specialization to be optimal)*** *Suppose P^k is given by (5). The objective function f defined in (6) is strictly convex if any of the following conditions hold:*

1. $\beta_{same} > 0$ *and* $\beta_{same} \geq (K-1) \cdot \beta_{other}$
2. $\beta_{other} \geq 0$ *and* $\beta_{same} > \beta_{other}$
3. *the matrix* $\begin{bmatrix} \beta_{same} & \beta_{other} & \beta_{other} \\ \beta_{other} & \ddots & \beta_{other} \\ \beta_{other} & \beta_{other} & \beta_{same} \end{bmatrix}$ *is positive definite.*

Intuitively, this result indicates that the objective function (6) is convex if the benefits of specific learning-by-doing (measured by the coefficient β_{same}) exceed the benefits of generic learning-by-doing (measured by the coefficient β_{other}). When this is the case, it is optimal to specialize the allocation of labor.

4 Data and institutional setting

4.1 The data

Our dataset contains all the 234,050 cases filed between January 1, 2001 and December 31, 2010 in the labor court in Rome, Italy. This is the labor court of first instance in Europe's largest tribunal for number of cases.[3] The disputes occur between the firm and one or more of its workers. The nature of the dispute is coded in court filings according to the following typology: wages, promotions, working conditions, pension and sick-law rights, terminations, worker misconduct, hiring procedures, discrimination, as well as other minor issues.

We observe the entire history of each case from filing to disposition. Most dispositions take the form of a ruling (69.5%) or of a settlement between the parties (12%). The rest of the dispositions represent cases where a party withdraws its claim, or where the suit cannot be adjudicated owing to factual or procedural reasons that become known after filing, or because exceptional circumstances arise. We code all dispositions, without regard to their form, as taking effect on the date of the case's last hearing.

Cases on average last about one year, are completed in three hearings and are appealed 10% of the times. To avoid right censoring of the data, we only keep cases filed between January 1, 2001 and December 31, 2010. Allowances (22%), damages (24%), and other hypotheses (11%) represent the majority of the cases filed to this court (see Table 1 for details).

Our model is based on the idea that a judge's productivity in a given hearing is a function of her experience up to that hearing. Our main proxy for experience in a given hearing will be n, the number of cases assigned to

[3] See http://www.repubblica.it/2007/01/sezioni/cronaca/bolzoni-tribunale/bolzoni-tribunale/bolzoni-tribunale.html

Table 1: Summary statistics of the cases

	mean	sd	p25	p50	p75	n
Duration of trials	413	300	221	349	536	234050
Prob. Appeal	.095	.29	0	0	1	234050
N. hearings	3.5	2.1	2	3	4	234050
N. actors involved	2.8	3.6	2	2	3	234050
Allowances	.22	.42	0	0	0	234050
Damages	.24	.43	0	0	0	234050
Other type I	.11	.31	0	0	0	234050
Invalidity	.038	.19	0	0	0	234050
Pension	.058	.23	0	0	0	234050
Temp. Contracts	.046	.21	0	0	0	234050
Firing	.089	.28	0	0	0	234050
Qualifica	.023	.15	0	0	0	234050
Other type II	.17	.38	0	0	0	234050

Note: Statistics for all the cases filed to the Labor Court of Rome between January 1, 2001 and December 31, 2010.

the judge within the recent past. We presume that recent experience might be more relevant, but we don't want to take a stand on exactly what counts as "recent:" thus in the empirical analysis we will run three different models based on the length of the experience window: 1 year back from current hearing, 2 years back from current hearing, ever within our sample. Note that these variables are computed individually for every hearing of every case. So, for example, for a Pension-case hearing held on May 2, 2005, the variables n_{same} (n_{other}) for that hearing records how many Pension (non-Pension) cases have been assigned to the judge within 1 year, 2 years, or ever, up to May 2, 2005. Table 2 indicates that, for the average hearing, the mean number of cases of the same type assigned to the judge equals 98 in the previous year; 710 are instead the assigned cases of a different type. Similarly for other intervals.

Table 2 also reports the summary statistics on the variable h which represents the number of hearings that the judge holds (in the same intervals of 1 year before the current hearing, 2 years before, or ever within our sample.) Note that while our focus is on the outcome of cases filed in the 2001-2010 period, we compute n and h using all the data till December 31, 2014.

The cases are handled by a total, over our entire time period, of 85 full-

Table 2: Experience correlates, by hearing of each case

	mean	sd	p25	p50	p75	n
Prob. of closing the case	.29	.45	0	0	1	808583
Cases assigned (in 1,000)						
$n_{same-type}$, w/in 1yr	.098	.066	.049	.097	.13	808583
$n_{other-type}$, w/in 1yr	.71	.22	.59	.69	.78	808583
$n_{same-type}$, w/in 2yrs	.19	.11	.099	.19	.25	808583
$n_{other-type}$, w/in 2yrs	1.3	.34	1.2	1.4	1.5	808583
$n_{same-type}$, ever	.51	.37	.22	.42	.73	808583
$n_{other-type}$, ever	3.6	1.9	2.1	3.4	4.8	808583
Hearings held (in 1,000)						
$h_{same-type}$, w/in 1yr	.36	.23	.18	.32	.49	808583
$h_{other-type}$, w/in 1yr	2	.71	1.6	1.9	2.4	808583
$h_{same-type}$, w/in 2yrs	.65	.42	.31	.59	.93	808583
$h_{other-type}$, w/in 2yrs	3.7	1.5	3	3.7	4.7	808583
$h_{same-type}$, ever	1.6	1.3	.51	1.2	2.3	808583
$h_{other-type}$, ever	9.3	6.1	4.5	8.6	13	808583

Note: $n_{same-type}$, *w/in 1yr (2 yrs) [ever]* is the number of cases assigned of the same type of every case, in every hearing in the previous year (two years) [ever]. $h_{other-type}$, *w/in 1yr (2 yrs) [ever]* is the number of cases assigned of different type, in every hearing in the previous year (two years) [ever]. $h_{same-type}$, *w/in 1yr (2 yrs) [ever]* is the number of hearings held of the same type of every case, in every hearing in the previous year (two years) [ever]. $h_{other-type}$, *w/in 1yr (2 yrs) [ever]* is the number of hearings held of different type, in every hearing in the previous year (two years) [ever]. $n(h)_{same-type}$, and $n(h)_{other-type}$ in 1,000 cases.

time labor judges. We know the age and gender of these judges.

4.2 Institutional setting, including procedure for random allocation

All Italian judges hold a law degree and are selected through a public examination covering all subjects and procedural rules in law. They are paid a fixed wage that increases with seniority but is largely independent of performance. Performance matters, in addition to seniority, if and when judges request to be transferred across courts and functions.

In our court each judge is solely responsible for adjudicating the cases assigned to him or her. No jury or other judges are involved. Judges are not allowed to render themselves unavailable for assignments, unless they are sick for long periods (more than one week). In a few rare cases some judges show prolonged periods of inactivity (many months). Because their experience is atypical, we elect to drop them from our sample.

Random assignment among the "relevant" judges is required by law (Art. 25 of the Italian Constitution). The goal of this law is to ensure the absence of any relationship between the identity of judges and the characteristics of the cases assigned to them, including the identity of lawyers and the complexity of cases. In our court, random assignment is implemented by a computer that is managed by a court clerk who, in turn, is supervised by an assigned judge.

4.3 Testing random allocation of cases

Our econometric strategy relies on the random assignment of cases to judges. In this section we test for randomness in the assignment.

To provide a concrete sense of what the assignment process looks like, Table 3 reports an extract of case assignment for two consecutive weeks for six judges. These six judges receive on average 8.5 and 8.8 cases, respectively in the two weeks. In the first one, judge 38 receives seven cases; in the second week s/he receives 8 cases. Random assignment of cases across judges will occasionally generate streaks of same-type cases which create mini-specialization events that occur exogenously. Such events can be seen in Table 3: for instance, judge 38 receives no type-1 cases in the first week and s/he receives 4 type-1 cases in the following week. To test formally for random assignment during these two weeks across all judges, we report the *p-values* for Pearsons Chi-square tests computed for the 45 judges that were on duty in each of these two weeks.[4] This test checks whether judges (rows) and type of cases (columns) are independent and therefore whether cases are randomly assigned to judges. The two *p-values* are well above .10 and so the null hypothesis of random assignment cannot be rejected in the data. This test indicates that the variation in case type allocated to judges within each of these two weeks is random and not systematic.

[4] We assume that a judge is on duty if s/he receives at least a case during a particular week.

Table 3: A two-week 6-judges extract of case assignment, and p-values

Judge ID	Case type: 1	2	3	4	5	6	7	8	9	Cases assigned
					Week 18, 2006					
38	0	3	2	0	0	0	1	0	1	7
39	2	4	1	0	0	0	0	1	3	11
40	2	2	0	0	1	1	0	0	2	8
42	4	1	2	0	1	1	0	0	1	10
43	1	3	1	0	0	2	0	0	0	7
44	0	2	1	0	1	1	1	2	0	8
Random assignment (p-value)										.885
					Week 19, 2006					
38	4	2	1	0	0	1	0	0	0	8
39	2	2	1	0	2	1	0	0	0	8
40	1	4	1	0	0	1	1	1	1	10
42	1	3	0	0	1	0	1	0	2	8
43	4	1	1	0	1	0	1	0	2	10
44	4	2	1	0	0	0	0	0	2	9
Random assignment (p-value)										.994

Note: *Random assignment (p-value)* is the p-value of the Pearsons χ^2 tests computed for the judges that received at least a case in each of the weeks. These six judges are a sub-sample of the 45 judges for which we compute the tests for weekly random assigment.

Extending this logic beyond this two-week 6-judges extract, we test for random assignment by computing the Chi-square tests of independence between the judge id and several case characteristics for all weeks and all judges. These characteristics are the type of controversy in 9 categories (9 dummies); an aggregation of the type of controversy in emergency cases[5]; a dummy for the plaintiff lawyer being from Roma; the number of involved parties (capped at 10).

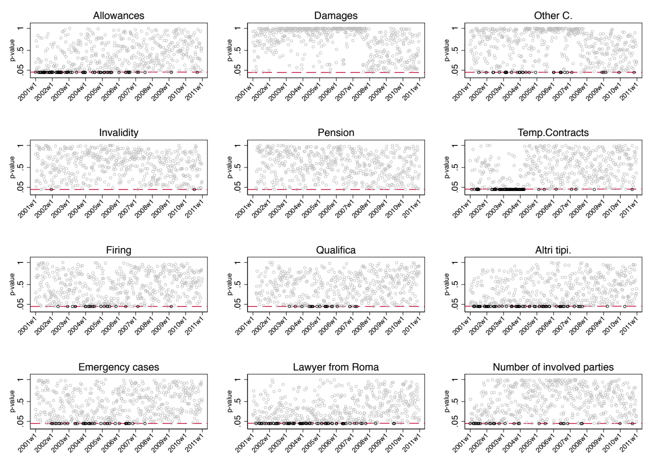
Light gray (black) circles in Figure 1 indicate the *p-values* above (below) the correct significance levels (dashed horizzontal red line) that are computed with the Benjamini and Hochberg (1995) multiple testing procedure.[6] When

[5]By analogy with what happens in a hospital emergency room, where red code cases are those that, according to judges, are urgent thus requiring immediate action and/or greater effort

[6]Summary results of the weekly tests for random assignment are presented in Table B.1. The last row presents joint results for all variables and all weeks. The first column

these correct significance levels are used, the number of rejections declines considerably as shown by the fraction of light gray circles. We can conclude that, within each week, differences in assignments are due only to small sample variability and are not systematic: in the long run, judges, receive qualitatively and quantitatively similar portfolios of controversies.

Figure 1: P-values for all weeks, all judges: evidence of random assignment



Dots are the p-values of the Chi-square tests of independence between the identity of judges and the characteristics of cases: *type of controversy* in 9 categories; a dichotomous aggregation of the types of controversy in *red code*; a dummy for *firing cases*; *zip code* of the plaintiff's lawyer (55 codes); the "number of involved parties" (capped at 10). Dashed (red) lines are correct significance levels computed with the Benjamini and Hochberg (1995) multiple testing procedure.

5 Empirical models

Our goal is to estimate the parameters β_{same} and β_{other} in the probability function (5) by exploiting random streaks of same-type cases which create mini-specialization events.

When the outcome is the probability of closing the case in a given hearing,

reports the numbers of weeks in which independence is rejected at the 5% level out of the 520 weeks on which the test is conducted. The corresponding fraction of rejections is in the second column. Since 5% is not the correct significance level in a context of multiple testing, in the third column we report the significance levels corrected with the Benjamini and Hochberg (1995) method.

the corresponding empirical model is:

$$\mathbb{I}_{i,u} = \alpha + \beta_{same} n_{j,k,t} + \beta_{other} n_{j,-k,t} + \beta_{np} np_i + \gamma_j + \delta_u + C_k + \eta_t + \mu_a + \epsilon_{i,u}. \quad (7)$$

where $\mathbb{I}_{i,u}$ is a dummy taking value one if case i is closed in its u-th hearing; j is the identifier of the judge to whom case i is assigned; k is case i's type; t is the calendar date in which the u-th hearing of case i is held. $n_{j,k,t}$ is the number of k-type cases assigned to judge j in the 365 (730, ever) days prior to the date of the u-th hearing, and $n_{j,-k,t}$ is the number of non-k-type cases assigned to judge j in the 365 (730, ever) days prior to the date t, both measured as fractions of 1,000 cases. np_i is the number of parties involved in the trial; γ_j are the judge fixed effects; δ_u are the u-th hearing fixed effects (first, second, third ...). C_k are the nine case-type fixed effects; η_t are fixed effects for the week in which the u-th hearing is held. Finally the model also includes fixed effects μ_a for the week of assignment of each case.

It should be noted that an observation is a hearing of a case. Therefore, strictly speaking, equation (7) is not correctly notated. In our database an observation is uniquely identified by the case id and the hearing counter (i, u) alone, and the indices j, k, and t in equation (7) should in fact be correctly notated as $j(i), k(i), t(i, u)$. But the correct notation is more cumbersome and, perhaps, less transparent, so we opted for the simpler notation in equation (7).

Random assignment of cases across judges guarantees that they cannot select endogenously the number of cases of each type assigned to them (which would create a problem if their selection reflected unobservables such as knowledge about a certain type of case, etc.). Random assignment also addresses also another concern: type-k cases might be more likely to be litigated during those times in which type-k jurisprudence is less settled, making type-k cases of this vintage simultaneously more numerous and more difficult to adjudicate. If this were the case then we would incorrectly attribute to specialization an effect that is in fact related to unobserved variation in the difficulty of cases. For these and similar reasons, we include the week of assignment fixed effects, μ_a so that the variation that identifies the β coefficients originates from random assignment.

We cluster standard errors at the judge and hearing week level. A possible concern with this two-way clustering strategy is that autocorrelation in backlogs might mechanically induce correlation across hearing dates, which would not be captured by the two-way clustering. Following a more conservative approach, in Appendix B we report estimates of the standard errors clustered at the judge level.

When the outcome is the probability of appeal the empirical model corresponding to (5) is:

$$Appeal_i = \alpha + \beta_{same} n_{j,k,a} + \beta_{other} n_{j,-k,a} + \beta_{np} np_i + \gamma_j + C_k + \mu_a + \epsilon_i. \quad (8)$$

where $Appeal_i$ is a dummy taking value 1 if case i is appealed and the other variables are defined as descried above. In this equation there is one observation per case, which is dated at the week of assignment a.

6 Effect of specialization on productivity

6.1 Specialization increases the probability of closing cases, has no effect on quality

Table 4 reports the estimated effects of experience on the probability of closing a case. The estimates indicate that, in all three specifications of the experience window, the estimated coefficient β_{same} is positive and greater than β_{other}. Furthermore, the difference between the two coefficients is statistically significant as indicated by the p-values.[7] Therefore, by Proposition 2 the objective function is convex and so it is optimal for judges to specialize.

Interestingly, the coefficients β_{other} are negative suggesting, according to the interpretation in Section 3, that judges get *worse* at type-k cases when they are assigned more non-k cases; apparently, there is no transferability in experience across case types.

[7] The statistical significance of these results is unchanged if we compute standard errors clustered at the judge level, see Table B.2.

Table 4: Effect of specialization on the probability of closing a case

Dep. Var.	Prob.Close	Prob.Close	Prob.Close
Method	OLS	OLS	OLS
	(1)	(2)	(3)
$n_{same-type}$, w/in 1yr	0.208***		
	(0.037)		
$n_{other-type}$, w/in 1yr	-0.060***		
	(0.012)		
$n_{same-type}$, w/in 2yrs		0.156***	
		(0.024)	
$n_{other-type}$, w/in 2yrs		-0.046***	
		(0.008)	
$n_{same-type}$, ever			0.049***
			(0.016)
$n_{other-type}$, ever			-0.019
			(0.013)
Test for $\beta_{same} \neq \beta_{other}$:	.268	.202	.068
p-value	.001	.001	.001
Judge FE	Yes	Yes	Yes
Week of hearing FE	Yes	Yes	Yes
Type of case FE	Yes	Yes	Yes
Hearing number FE	Yes	Yes	Yes
Week of assignment FE	Yes	Yes	Yes
Number of judges	85	85	85
Number of cases	234,050	234,050	234,050
Observations	808,583	808,583	808,583

Note: An observation is a hearing of a case. The dependent variable is a dummy for the closure of a case in a given hearing. For each case, $n_{same-type}$*, w/in 1yr (w/in 2yrs; ever)* is the (per 1000) number of cases of the same type assigned to the judge before the hearing within 1year (within 2years; ever). Similarly for $n_{other-type}$. All regressions control for the number of parties involved in the trial. Standard errors in parentheses are clustered at the judge and week of the hearing level (two-way clustering). *** p<0.01, ** p<0.05, * p<0.1.

The theoretical analysis in Section 3 can be applied to the probability of appeals, with the proviso that here specialization is beneficial if appeals are *reduced*, which means that the function f must now be *concave*, or equivalently, $-f$ must be convex. Condition 1 in Proposition 2, when applied to $-f$, says that specialization is beneficial in reducing appeals if $\beta_{same} < 0$ and $\beta_{same} - \beta_{other} < 0$. Table 5 hints at a possible beneficial effect of specialization on the probability of appeal, in that the estimates for $\beta_{same} - \beta_{other}$ are always negative, and statistically significant in column 2 only. This represents suggestive evidence that specialization might have a beneficial effect in terms of appeal reductions.

Table 5: Effect of specialization on the probability of appeal

Dep. Var	Prob.Appeal	Prob.Appeal	Prob.Appeal
Method	OLS	OLS	OLS
	(1)	(2)	(3)
$n_{same-type}$, w/in 1yr	-0.0419		
	(0.032)		
$n_{other-type}$, w/in 1yr	0.0172		
	(0.012)		
$n_{same-type}$, w/in 2yrs		-0.0483*	
		(0.024)	
$n_{other-type}$, w/in 2yrs		0.0105	
		(0.007)	
$n_{same-type}$, ever			-0.0059
			(0.007)
$n_{other-type}$, ever			-0.0027
			(0.004)
Test for $\beta_{same} \neq \beta_{other}$:	-0.059	-0.059	-0.003
p-value	0.145	0.041	0.632
Judge FE	Yes	Yes	Yes
Type of case FE	Yes	Yes	Yes
Week of assignment FE	Yes	Yes	Yes
Number of judges	85	85	85
Observations	234,050	234,050	234,050

Note: An observation is a case. The dependent variable is a dummy for the event that the case is appealed. For each case, $n_{same-type}$, *w/in 1yr (w/in 2yrs; ever)* is the (per 1000) number of cases of the same type assigned to the judge before the hearing within 1 year (within 2 years; ever). Similarly for $n_{other-type}$. All regressions control for the number of parties involved in the trial. Standard errors in parentheses are clustered at the judge and week of assignment level (two-way clustering). *** p<0.01, ** p<0.05, * p<0.1.

6.2 Quantitative assessment of the gains from specialization

We want to compute the effect on the amount of cases closed $f(\mathbf{n})$ of a marginal increase in specialization, namely: having judge j swapping a single case with judge j'. The switch does not affect the allocation of any judges other than j and j', hence the effect on productivity will be limited to judges j and j'. The aggregate effect of the swap on both judges' productivity is as follows.

Proposition 3 ***(productivity gains from specialization)*** *Consider two judges j, j' who are allocated $n_{j,\kappa}, n_{j',\kappa}$ type-κ and $n_{j,\kappa'}, n_{j',\kappa'}$ type-κ' cases. Suppose judge j swaps a case with judge j' so that judge j is assigned one more hearing of type κ and one fewer hearing of type κ', and vice versa for judge j'. The resulting change in the total production $f(\mathbf{n})$ is:*

$$2A\left[\left(n_{j,\kappa}-n_{j',\kappa}\right)\left(\beta_{same}-\beta_{other}\right)+\left(n_{j',\kappa'}-n_{j,\kappa'}\right)\left(\beta_{same}-\beta_{other}\right)\right].$$

The returns to specialization are increasing in the level of specialization. The latter is represented by the term $(n_{j,\kappa}-n_{j',\kappa})$ which is positive if judge j is more specialized in cases of type κ than judge j', and by the term $(n_{j',\kappa'}-n_{j,\kappa'})$ which is positive if judge j' is more specialized in cases of type κ' than judge j. Assuming that Proposition 2's sufficient conditions for convexity are met, the above expression is larger and hence total productivity is more likely to be improved by the switch, when: judge j already handles more κ-hearings than judge j', and judge j' already handles more κ'-hearings than judge j (that is, there are increasing returns from specialization); and when specific experience matters more ($\beta_{same}-\beta_{other}$ is larger). Notably, the productivity gains do not depend on the judges' ability γ_j, on the difficulty of the case types C_k, or on the judge's docket of "other" cases $n_{j,-k}$.

To get a quantitative sense of the returns to specialization, set $A=3$ (this parameter choice was discussed back on page 9) and, from Table 4 cols 1 and 2, set $\beta_{same}-\beta_{other}=0.202$ based on the estimate from the two-year specification, which is intermediate between columns 1 and 3 in Table 4. The benefits from specialization depend on the extant level of specialization, so pick a level of specialization such that judge j has 200 more type-κ cases than judge j' and judge j' has 200 more type-κ' cases than judge j. Then by

Proposition 3 the marginal return to specialization equals:

$$2 \cdot 3\left[(0.2)\,(0.202) + (0.2)\,(0.202)\right] = 0.48, \tag{9}$$

where the figure 0.2 represents the 200-case difference expressed in the unit of measure (1000 cases) which was used to estimate the β coefficients.

The above formula represents the gain from a small increase in specialization around an allocation where the difference in cases assigned is 200 per type. To get a quantitative sense of the benefits from specialization around this allocation requires comparing the benefits from *some stipulated amount* of increased specialization to the existing level of productivity. For the purpose of this calculation, the stipulated amount will be the *maximum amount of hearing switches* that are possible between two judges, that is, the amount of unexploited specialization. This amount is roughly 2000 hearings.[8] Multiplication by 0.48 yields 960, which represents the gains that the two judges can jointly achieve by going to full specialization, evaluated at the rate that prevails at the initial allocation where the difference in cases assigned is 200 per type.

We compare the gains computed above with the *total productivity* in our data, which for two judges equals: 2360 (total number of hearings held by a judge) times two (judges) times 0.29 (probability of closing a case in the average hearing, from Table 2). The resulting figure is 1369.

The ratio 960/1369 = 0.7 can be interpreted as follows. At the rate that prevails in the allocation where the difference in cases assigned is 200 per type, a marginal increase in specialization by 1% of the available amount of unexploited specialization would increase total judicial productivity by a rate of 0.7%. In other words, the elasticity of judicial productivity to specialization is 0.7.

7 Supporting the learning-by-doing channel

So far we have modeled judicial productivity as dependent on the *number of cases assigned* to a judge. It is natural to interpret our results as reflecting

[8]This figure is based on average docket of about 800 cases in total (refer back to Table 2).

learning-by-doing. But a judge learns by *holding hearings*, not merely by being assigned more cases. If hearings were randomly assigned to a judge, then we could regress productivity on "number of hearings held" and the estimates could legitimately be interpreted as measuring the effect of learning-by-doing on judicial productivity. But such randomness is unavailable because judges choose which hearings to hold (their workflow) endogenously.

A way forward is to instrument the "number of hearings held by each judge" with "the number of cases assigned to each judge," the latter being determined randomly as explained above. Thus we estimate the following model:

$$\mathbb{I}_{i,u} = \alpha + \beta_{same} h_{j,k,t} + \beta_{other} h_{j,-k,t} + \beta_{np} np_i + \gamma_j + \delta_u + C_k + \eta_t + \mu_a + \epsilon_{i,u}. \quad (10)$$

where $h_{j,k,t}$ is the (per 1000) number of hearings held for k-type cases by judge j in the 365 (730, ever) days prior to the date of the u-th hearing, and $h_{j,-k,t}$ is the (per 1000) number of hearings held for non-k-type cases assigned to judge j in the 365 (730, ever) days prior to the date t. The remaining variables are defined as in equation (7).

Table 6 provides 2SLS estimates based on this logic, in which $h_{j,k,t}$ and $h_{j,-k,t}$ are instrumented by $n_{j,k,t}$ and $n_{j,-k,t}$. The first-stage estimates (see Table B.3) have the expected sign and are strong: in Table 6, the *Cragg-Donald Wald F statistics (Joint)* are always well above 10, suggesting that "the number of assigned cases" is a significant determinant of "the number of hearings held by the judge."

Table 6: Learning by doing and the probability of closing a case

Dep. Var.	Prob.Close	Prob.Close	Prob.Close
Method	2SLS	2SLS	2SLS
	(1)	(2)	(3)
$h_{same-type}$, w/in 1yr	0.0472		
	(0.090)		
$h_{other-type}$, w/in 1yr	-0.1117*		
	(0.057)		
$h_{same-type}$, w/in 2yrs		0.0866***	
		(0.019)	
$h_{other-type}$, w/in 2yrs		-0.0310***	
		(0.006)	
$h_{same-type}$, ever			0.0134**
			(0.006)
$h_{other-type}$, ever			-0.0075
			(0.005)
Test for $\beta_{same} \neq \beta_{other}$:	.159	.118	.021
p-value	.001	.001	.001
C.-D. Wald F statistic (Joint)	4503	90563	178666
Judge FE	Yes	Yes	Yes
Week of hearing FE	Yes	Yes	Yes
Type of case FE	Yes	Yes	Yes
Hearing FE	Yes	Yes	Yes
Week of assignment FE	Yes	Yes	Yes
Number of judges	85	85	85
Number of cases	234,050	234,050	234,050
Observations	808,583	808,583	808,583

Note: An observation is a hearing of a case. The dependent variable is a dummy for the closure of a case in a given hearing. For each case, $n_{same-type}$, *w/in 1yr (w/in 2yrs; ever)* is the (per 1000) number of cases of the same type assigned to the judge before the hearing within 1yr (within 2years; ever). Similarly for $n_{other-type}$ All regressions control for the number of parties involved in the trial. $C.-D.WaldFstatistic(Joint)$ denotes the minimum eigenvalue of the joint first-stage F-statistic matrix. Standard errors in parentheses are clustered at the judge and week of the hearing level (two-way clustering). *** p<0.01, ** p<0.05, * p<0.1.

These 2SLS estimates are consistent with the ITT estimates of equation (7), in that for all specifications we find $\beta_{same} > \beta_{other}$. Therefore, they support the hypothesis that specialization increases productivity through learning by doing.

Using "number of cases assigned to a judge" to instrument for "number of hearings held by the judge," as we do in this section, requires ruling out the possibility that a judge manipulates the *difficulty of the hearings* she selects. Such a manipulation would violate the exclusion restriction. For example, upon being assigned more Pension cases the judge might react by selecting Pension hearings from easier cases. If that were the case, we would mistakenly attribute to experience what is, in fact, a selection effect. To explore this concern, we seek a measure of case difficulty. While administrative measures of case difficulty are not available, we proxy for case difficulty with the number of parties, since cases with more parties are generally viewed as more complex and indeed, in our data, they take more hearings to close. If the judge picked hearings of cases with a smaller number of parties within a case-type when confronted with a larger assignment of cases of that type, the number of parties at each hearing would not be exogenous.

To test for this type, we estimate the following model:

$$N.Parts_{j,k,t} = \alpha + \beta_{same} n_{j,k,t} + \beta_{other} n_{j,-k,t} + \gamma_j + \eta_t + C_k + \eta_t + \epsilon_{j,k,t}. \quad (11)$$

and results are presented in Table 7. The dependent variable is the average number of parts involved in cases k in a hearing held in week t. The other variables are defined as in equation 7. For this specification we cluster standard errors at the judge and week of the hearing level (two-way clustering).

Evidence for strategic selection of cases for hearings is weak at best. Only in column 2 do we find any evidence that specialisation (1,000 more cases) reduces the average number of parties in the cases heard, and the statistical significance of that coefficient is borderline. Overall, the evidence does not seem to point to systematic selection of hearing along the difficulty dimension as a major correlate of productivity.

Table 7: Assignment and selection of cases into hearings

Dep. Var	N.Parts	N.Parts	N.Parts
Method	OLS	OLS	OLS
	(1)	(2)	(3)
$n_{same-type}$, w/in 1yr	-0.290		
	(0.597)		
$n_{other-type}$, w/in 1yr	0.054		
	(0.050)		
$n_{same-type}$, w/in 2yrs		-0.522*	
		(0.267)	
$n_{other-type}$, w/in 2yrs		0.110***	
		(0.038)	
$n_{same-type}$, ever			-0.094
			(0.083)
$n_{other-type}$, ever			-0.023
			(0.050)
Judge fixed FE	Yes	Yes	Yes
Week of hearing FE	Yes	Yes	Yes
Type of case FE	Yes	Yes	Yes
Number of judges	85	85	85
Observations	226,059	226,059	226,059

Note: An observation is a hearing of a case. The dependent variable is the average number of parties involved in a case of "same-type" . For each case, $n_{same-type}$, *w/in 1yr (w/in 2yrs; ever)* is the (per 1000) number of cases of the same type assigned to the judge before the hearing within 1 year (within 2 years; ever). Similarly for $n_{other-type}$. Standard errors in parentheses are clustered at the judge and week of the hearing level (two-way clustering). *** p<0.01, ** p<0.05, * p<0.1.

8 Conclusions

The literature that estimates the gains from labor specialization has had to confront two key identification issues. First, workers are in general not randomly exposed to specialization; second, the measurement of the benefits from specialization might be biased if tasks are not randomly assigned to workers. In this paper we were able to address both identification concerns due to the explicitly random process through which our workers are assigned tasks. We have leveraged this uniquely favorable identification scenario to obtain estimates of the productivity-enhancing effects of specialization.

The estimates suggest that if judges were more specialized they would be considerably faster, i.e., more likely to close a case in any given hearing of it; quality, as measured by probability of appeal, would not be negatively affected. These results indicate large and unexploited gains from specialization for this particular group of workers, a finding that may be interpreted as a "free lunch," and thus regarded skeptically by some readers. However, when viewed from an organizational economics perspective, the judiciary is an unusual workplace: as an organization it is not exposed to competition; and its employees (judges) are, by design, insulated from authority and from monetary incentives in most work-related actions. Given high autonomy and soft incentives, it is not too surprising that large productivity gains remain unexploited.

Our analysis has policy relevance because judicial productivity matters a great deal for economic growth and development,[9] and also because the process of specialization which is taking place in the judicial profession is alive with controversy. A number of caveats must therefore be raised regarding the policy implications of this work. First, this paper is certainly not the last word; its findings need to be replicated across different courts, ideally with controlled field trials. Second, as well as benefits, judicial specialization may entail the drawbacks listed in Section 2: our estimates can hopefully provide quantitative estimates for the benefits, thus giving a sense of the magnitude of one side of the cost-benefit equation. Third, labor specialization requires scale, and accordingly, judicial specialization requires courts with

[9] According to the World Bank's "Doing Business" website, "enhancing the efficiency of the judicial system can improve the business climate, foster innovation, attract foreign direct investment and secure tax revenues."

many judges. Judicial systems that have many small courts will require mergers in order to reach the requisite scale. These mergers may be politically difficult.

References

Ash, E., and B.,W., MacLeod, (2014). Intrinsic Motivation in Public Service: Theory and Evidence from State Supreme Courts. Journal of Law and Economics, 58, 4, pp. 863-913.

Ash, E., and B.,W., MacLeod, (2016). The Performance of Elected Officials: Evidence from State Supreme Courts. *Mimeo*

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

Bagues, M., and B., Esteve-Volart, (2010). "Performance Pay and Judicial Production: Evidence from Spain." *Mimeo.*

Baum, Lawrence (2009). Probing the Effects of Judicial Specialization. Duke Law Journal, vol. 58, no. 7, 2009, pp. 16671684. JSTOR, www.jstor.org/stable/20684768.

Baum, Lawrence (2011). *Specializing the Courts.* University of Chicago Press, 2011.

Benjamini, Y., and Y., Hochberg, (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. The Journal of the Royal Statistical Society B, 57, pp. 289-300.

Bray, R., Coviello, D., A., Ichino, and N., Persico, (2016). Multitasking, Multi-Armed Bandits, and the Italian Judiciary. Manufacturing and Service Operation Management, 18(4), pp. 545-558.

Chowdhury, M. M., Dagash, H., and A. Pierro. (2007). A systematic review of the impact of volume of surgery and specialization on patient outcome. British journal of surgery, 94(2), pp. 145-161.

Coviello, D., A., Ichino, and N., Persico, (2014). Time allocation and task juggling. American Economic Review, 104(2), pp. 609623.

Coviello, D., A., Ichino, and N., Persico, (2015). The inefficiency of worker time use. Journal of the European Economic Association, 13(October), pp. 906947.

Cook, J., B., R. K. Mansfield. (2006). Task-specific experience and task-specific talent: Decomposing the productivity of high school teachers. Journal of Public Economics, Volume 140, 2016, pp. 51-72,

Di Tella, R. and E. Schargrodsky. (2012). Criminal Recidivism after Prison and Electronic Monitoring. Journal of Political Economy, 2013, 121.

Dimitrova-Grajzl, V., P., Grajzl, J., Sustersic, and K. Zajc. (2012). Court output, judicial staffing, and the demand for court services: Evidence from Slovenian courts of first instance. International Review of Law and Economics 32.1 (2012), pp. 19-29.

Djankov, S., R., LaPorta, F., Lopez-de-Silanes, and A., Shleifer, (2003). Courts. Quarterly Journal of Economics 118, 2, pp. 453-517.

Friebel, G., and L., Levent. (2017). Flexibility, specialization and individual productivity: Evidence from call center data. Manuscript.

IMF Working Paper, (2014). Judicial System Reform in ItalyA Key to Growth. by Sergi Lanau, Gianluca Esposito, and Sebastiaan Pompe. N. 14/32

KC, D. S., and B.R., Staats. (2012). Accumulating a portfolio of experience: The effect of focal and related experience on surgeon performance. Manufacturing & Service Operations Management, 14(4), pp. 618-633.

KC, D., B.R., Staats, and F. Gino. (2013). Learning from my success and from others' failure: Evidence from minimally invasive cardiac surgery. Management Science, 59(11), pp. 2435-2449.

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

LilienfeldToal, U., D., Mookherjee, and S., Visaria. (2012). The distributive impact of reforms in credit enforcement: Evidence from Indian debt recovery tribunals. Econometrica 80.2 (2012), pp. 497-558.

Narayanan, S., S., Balasubramanian, and J. M., Swaminathan. (2009). A Matter of Balance: Specialization, Task Variety, and Individual Learning in a Software Maintenance Environment. Management Science 55(11), pp. 1861-1876.

Ost, B. (2014). How do teachers improve? The relative importance of specific and general human capital. American Economic Journal: Applied Economics, pp. 127-151.

Ponticelli, J., and L. S. Alencar. (2016). Court enforcement, bank loans, and firm investment: evidence from a bankruptcy reform in Brazil. The Quarterly Journal of Economics 131, 3, pp. 1365-1413.

Shaw, K., and E.P. Lazear. (2008). Tenure and output. Labour Economics, 15(4), pp. 704-723.

Staats, Bradley R., and F. Gino (2012). Specialization and variety in repetitive tasks: Evidence from a Japanese bank. Management science 58,6, pp. 1141-1159.

World Bank (2014). Doing Business 2014: Understanding Regulations for Small and Medium-Size Enterprises.

Appendices

A Theory

A.1 Proof of Proposition 1

Proof. Let's consider the feasible set X in our problem. It is the subspace $\{n_{j,k}\} \subset \mathbb{R}^{J \times K}$ such that (2 - 4) are satisfied. Clearly, this feasible set is convex. If our objective function is convex, then the solutions must be extremal. What are the properties of extremal solutions? Consider an allocation $x = \{n_{j,k}\}$ where two judges j and j' are assigned:

$$\begin{aligned} 0 &< n_{j,k} \\ 0 &< n_{j,k'} \\ 0 &< n_{j',k} \\ 0 &< n_{j',k'} \end{aligned}$$

for some k, k'. Construct the following allocations:

Allocation y. y is equal to x in every entry except for: $y_{j,k} = n_{j,k} + \varepsilon$; $y_{j,k'} = n_{j,k'} - \varepsilon$; $y_{j',k} = n_{j,k} - \varepsilon$; $y_{j',k'} = n_{j',k'} + \varepsilon$

Allocation z. z is equal to x in every entry except for: $z_{j,k} = n_{j,k} - \varepsilon$; $z_{j,k'} = n_{j,k'} + \varepsilon$; $z_{j',k} = n_{j,k} + \varepsilon$; $z_{j',k'} = n_{j',k'} - \varepsilon$

Allocation y transfers a few type-k' cases from judge j to judge j'; and balances by transfering the same number of type-k cases from judge j' to judge j. Allocation z shifts cases in the opposite direction. These allocations are constructed so that

$$x = \frac{1}{2}y + \frac{1}{2}z.$$

Furthermore, allocations y and z are feasible because they satisfy (2 - 4):

$$\begin{aligned}\sum_k y_{j,k} &= \sum_k n_{j,k} + \varepsilon - \varepsilon = N_j \text{ for all } j \\ \sum_j y_{j,k} &= \sum_j n_{j,k} + \varepsilon - \varepsilon = N_k \text{ for all } k \\ y_{j,k} &\geq 0 \text{ for all } j,k \text{ provided } \varepsilon \text{ is sufficiently small}\end{aligned}$$

The same holds for allocation z.

Thus we have constructed two feasible allocations y, z such that $x = \alpha y + (1-\alpha) z$ for some $\alpha \in (0,1)$. It follows that $f(x) < \max[f(y), f(z)]$ for every strictly quasi-convex function f. Therefore allocation x could not be a maximizer for any strictly quasi-convex function. Thus we have shown that in the optimal allocation there cannot be two judges who are assigned a positive amount of the same two types of cases. ■

A.2 Proof of Proposition 2

We state and prove a somewhat more general version of Proposition 2. The added generality is that we allow the coefficient β_{same} to now be specific to each case type, and we denote each coefficient by β_k. In addition, we denote β_{other} by the shorter β_-. Thus, the function H^k now reads:

$$H^k(n_{j,k}, n_{j,-k}) = C_k + \gamma_j + n_{j,k}\beta_k + n_{j,-k}\beta_-, \tag{12}$$

The case dealt with in the main body of the paper is the special case where $\beta_1 = ... = \beta_K = \beta_{same.}$

Lemma 2 ***(Convexity requires specific learning-by-doing dominates generic learning-by-doing)*** *Suppose H^k is given by (12). Then objective function (6) is strictly convex if any of the following conditions hold:*

1. *$\beta_k > 0$ and $\beta_k \geq (K-1)\cdot\beta_-$ for all k*
2. *$\beta_- \geq 0$ and $\beta_k > \beta_-$ for all k*

3. the matrix $\begin{bmatrix} \beta_1 & \beta_- & \beta_- \\ \beta_- & \ddots & \beta_- \\ \beta_- & \beta_- & \beta_K \end{bmatrix}$ *is positive definite.*

Proof. The objective function can be written as:

$$\begin{aligned} & \sum_j \sum_k n_{j,k} H^{j,k} \left(n_{j,k}, n_{j,-k}\right) \\ = & \sum_j \sum_k n_{j,k} \left(C_k + \gamma_j + n_{j,k}\beta_k + n_{j,-k}\beta_-\right) \\ = & \sum_j \sum_k n_{j,k} \left(C_k + \gamma_j + n_{j,-k}\beta_-\right) + n_{j,k}^2 \beta_k \end{aligned}$$

Using the identity $n_{j,-k} = \sum_{\kappa \neq k} n_{j,k}$, the Jacobian reads:

$$\begin{aligned} \mathbf{J} &= \left[\overbrace{\left[\left(C_k + \gamma_1 + n_{1,-k}\beta_-\right) + 2n_{1,k}\beta_k + \sum_{\kappa \neq k} n_{1,\kappa}\beta_- \right]_{k=1\ldots K}}^{\text{judge } 1} \quad \cdots \quad \overbrace{\left[\left(C_k + \gamma_J + n_{J,-k}\beta_-\right) + 2n_{J,k}\beta_k + \sum_{\kappa \neq k} n_{J,\kappa}\beta_- \right]_{k=1\ldots K}}^{\text{judge } J} \right] \\ &= \left[\overbrace{\left[C_k + \gamma_1 + 2n_{1,-k}\beta_- + 2n_{1,k}\beta_k\right]_{k=1\ldots K}}^{\text{judge } 1} \quad \cdots \quad \overbrace{\left[C_k + \gamma_J + 2n_{J,-k}\beta_- + 2n_{J,k}\beta_k\right]_{k=1\ldots K}}^{\text{judge } J} \right] \end{aligned}$$

The Hessian reads:

$$\mathbf{H} = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_J \end{bmatrix}$$

where each submatrix

$$\mathbf{A}_j = 2 \cdot \begin{bmatrix} \beta_1 & \beta_- & \beta_- \\ \beta_- & \ddots & \beta_- \\ \beta_- & \beta_- & \beta_K \end{bmatrix}$$

If each block $\mathbf{A}_j$ is positive semidefinite, then $\mathbf{H}$ is also positive semidefinite (see http://math.stackexchange.com/questions/1715144/showing-that-a-partitioned-matrix-is-positive-definite).

A symmetric diagonally dominant real matrix with nonnegative diagonal entries is positive semidefinite. So $\mathbf{A}_j$ is positive definite if $\beta_k > 0$ for all k and it is diagonally dominant, that is, if $\beta_k \geq (K-1) \cdot \beta_-$.

Alternatively, note that

$$\frac{1}{2}\mathbf{A}_j = \begin{bmatrix} \beta_1 - \beta_- & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \beta_K - \beta_- \end{bmatrix} + \begin{bmatrix} \beta_- & \beta_- & \beta_- \\ \beta_- & \ddots & \beta_- \\ \beta_- & \beta_- & \beta_- \end{bmatrix},$$

so

$$\begin{aligned} \frac{1}{2}\mathbf{v}^T\mathbf{A}_j\mathbf{v} &= \mathbf{v}^T \begin{bmatrix} \beta_1 - \beta_- & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \beta_K - \beta_- \end{bmatrix} \mathbf{v} + \beta_- \mathbf{v}^T \begin{bmatrix} 1 & 1 & 1 \\ 1 & \ddots & 1 \\ 1 & 1 & 1 \end{bmatrix} \mathbf{v} \\ &= \mathbf{v}^T \begin{bmatrix} \beta_1 - \beta_- & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \beta_K - \beta_- \end{bmatrix} \mathbf{v} + \beta_- \sum_j v_j \sum_i v_i \\ &= \mathbf{v}^T \begin{bmatrix} \beta_1 - \beta_- & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \beta_K - \beta_- \end{bmatrix} \mathbf{v} + \beta_- \left(\sum_i v_i \right)^2. \end{aligned}$$

If $\beta_- > 0$ the second term is positive and a sufficient condition for positive definiteness is that the first term is positive, that is, that the matrix:

$$\begin{bmatrix} \beta_1 - \beta_- & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \beta_K - \beta_- \end{bmatrix}$$

be positive definite. ■

A.3 Proof of Proposition 3

The notation in this section follows that of Section A.2

Proof. Recall that:

$$\begin{aligned} f(\mathbf{n}) &= A \sum_j \sum_k n_{j,k} P^{j,k}(n_{j,k}, n_{j,-k}) \\ &= A \sum_j \sum_k n_{j,k} \left[C_k + \gamma_j + n_{j,k}\beta_k + n_{j,-k}\beta_- \right], \end{aligned}$$

In the algebra that follows we set the factor A to 1 for notational simplicity. We will remember to add it at the end.

The effect on productivity $f(\mathbf{n})$ of having judge j swapping a hearing with judge j' so that judge j is assigned one more hearing of type κ and one fewer hearing of type κ', and vice versa for judge j', is limited to judges j and j'. Let's first focus on the effect on judge j alone. The effect of an increase in $n_{j,\kappa}$ is:

$$\left[\frac{\partial f(\mathbf{n})}{\partial n_{j,\kappa}}\right]+\left[\frac{\partial f(\mathbf{n})}{\partial n_{j,-\kappa}}\frac{\partial n_{j,-\kappa}}{\partial n_{j,\kappa}}\right]$$
$$\left[C_\kappa+\gamma_j+2n_{j,\kappa}\beta_\kappa+n_{j,-\kappa}\beta_-\right]-\left[\sum_{k\neq\kappa}n_{j,k}\beta_-\right].$$

The effect of a decrease in $n_{j,\kappa'}$ is:

$$-\left[C_{\kappa'}+\gamma_j+2n_{j,\kappa'}\beta_{\kappa'}+n_{j,-\kappa'}\beta_-\right]+\left[\sum_{k\neq\kappa'}n_{j,k}\beta_-\right].$$

Adding the two effects together yields:

$$\begin{aligned}&\left[C_\kappa-C_{\kappa'}+2\left(n_{j,\kappa}\beta_\kappa-n_{j,\kappa'}\beta_{\kappa'}\right)+\left(n_{j,-\kappa}-n_{j,-\kappa'}\right)\beta_-\right]-\left[\left(n_{j,\kappa'}-n_{j,\kappa}\right)\beta_-\right]\\=\;&C_\kappa-C_{\kappa'}+2\left(n_{j,\kappa}\beta_\kappa-n_{j,\kappa'}\beta_{\kappa'}\right)+2\left(n_{j,-\kappa}-n_{j,-\kappa'}\right)\beta_-\ .\end{aligned}$$

The switch leaves unchanged the total number of cases N_j assigned to judge j, so substituting from the identity $n_{j,-k}=N_j-n_{j,k}$, the expression reads:

$$\begin{aligned}&C_\kappa-C_{\kappa'}+2\left(n_{j,\kappa}\beta_\kappa-n_{j,\kappa'}\beta_{\kappa'}\right)+2\left(n_{j,\kappa'}-n_{j,\kappa}\right)\beta_-\\=\;&C_\kappa-C_{\kappa'}+2n_{j,\kappa}\left(\beta_\kappa-\beta_-\right)-2n_{j,\kappa'}\left(\beta_{\kappa'}-\beta_-\right).\end{aligned}\tag{13}$$

The expression shows that judge j's productivity is more likely to increase due to the switch if, relative to type-κ' hearings, type-κ hearings are more likely to close $(C_\kappa>C_{\kappa'})$, and generate more specific learning-by-doing $(\beta_\kappa>\beta_{\kappa'})$; and, assuming that Lemma 2's sufficient conditions for convexity are met, if judge j has relatively more type-κ hearings than type-κ' hearings $(n_{j,\kappa}>n_{j,\kappa'})$.

The corresponding expression to (13) for judge j' who, recall, swaps one less κ-hearing for one more κ' hearing, is:

$$C_{\kappa'}-C_\kappa+2n_{j',\kappa'}\left(\beta_{\kappa'}-\beta_-\right)-2n_{j',\kappa}\left(\beta_\kappa-\beta_-\right).\tag{14}$$

Adding (13) and (14) yields the total effect of the swap on both judges' productivity. It is:

$$2n_{j,\kappa}\left(\beta_{\kappa}-\beta_{-}\right)-2n_{j,\kappa'}\left(\beta_{\kappa'}-\beta_{-}\right)+2n_{j',\kappa'}\left(\beta_{\kappa'}-\beta_{-}\right)-2n_{j',\kappa}\left(\beta_{\kappa}-\beta_{-}\right).$$

Now collect terms and reintroduce A back in to get:

$$2A\left[\left(n_{j,\kappa}-n_{j',\kappa}\right)\left(\beta_{\kappa}-\beta_{-}\right)+\left(n_{j',\kappa'}-n_{j,\kappa'}\right)\left(\beta_{\kappa'}-\beta_{-}\right)\right]. \tag{15}$$

■

B Additional tables and figures

Table B.1: Tests for the random assignment of cases to judges

	Rejections at 5% significance	Fraction of rejections at 5% significance	Corrected significance	Rejections at corrected significance	Fraction of rejections at corrected significance	N
	(1)	(2)	(3)	(4)	(5)	(6)
Allowances	111	.21	.0073	76	.15	520
Damages	10	.019	.000097	0	0	520
Oth.C.	61	.12	.0033	34	.065	520
Invalidity	16	.031	.0002	2	.0038	520
Pension	23	.044	.0001	0	0	520
Temp.C.	107	.21	.0078	76	.15	520
Firing	61	.12	.002	21	.04	520
Qualif.	71	.14	.0022	21	.04	520
Other.T.	125	.24	.0069	72	.14	520
Emergency	77	.15	.0037	38	.073	520
Lawyer-RM	131	.25	.007	73	.14	520
N.Parts.	70	.13	.003	31	.06	520
Overall	863	.14	.0034	412	.066	6,240

Note: The table summarizes the evidence on the weekly random assignment of cases to judges, based on Chi-square tests of independence between the identity of judges and five discrete characteristics of cases: *type of controversy* in 9 categories; a dichotomous aggregation of the types of controversy in *Emergency* cases, which are those that, according to judges, are urgent and/or complicated; a dummy for *firing cases*; *Lawyer-RM* equal one if the plaintiff's lawyer is from Rome; the "number of involved parties" (capped at 10). The last row, *Overall*, presents joint results for all variables and all weeks. Rejections at 5% significance" are the numbers of tests in which p-values are below 0.05. Correct significance levels are computed with the Benjamini and Hochberg (1995) multiple testing procedure. *Rejections at correct significance* are the numbers of tests in which p-values are below the correct significance levels.

Table B.2: Robustness: Effect of experience on the probability of closing a case, OLS with standard errors clustered at judge level

Dep. Var.	Prob.Close	Prob.Close	Prob.Close
Model	LPM	LPM	LPM
Method	OLS	OLS	OLS
	(1)	(2)	(3)
$n_{same-type}$, w/in 1yr	0.208***		
	(0.037)		
$n_{other-type}$, w/in 1yr	-0.060***		
	(0.012)		
$n_{same-type}$, w/in 2yrs		0.156***	
		(0.023)	
$n_{other-type}$, w/in 2yrs		-0.046***	
		(0.008)	
$n_{same-type}$, ever			0.049***
			(0.015)
$n_{other-type}$, ever			-0.019
			(0.012)
Diff.Coeff.	.268	.202	.068
p-value	.001	.001	.001
Judge FE	Yes	Yes	Yes
Week of hearing FE	Yes	Yes	Yes
Type of case FE	Yes	Yes	Yes
Hearing FE	Yes	Yes	Yes
Week of assignment FE	Yes	Yes	Yes
Number of judges	85	85	85
Number of cases	234,050	234,050	234,050
Observations	808,583	808,583	808,583

Note: Note: An observation is a hearing of a case. The dependent variable is a dummy for the closure of a case in a given hearing. For each case, $n_{same-type}$, *w/in 1yr (w/in 2yrs; ever)* is the (per 1000) number of cases of the same type assigned to the judge before the hearing w/in 1yr (w/in 2yrs; ever). Similarly for $n_{other-type}$. All regressions control for the number of parties involved in the trial. Standard errors in parentheses are clustered at the judge and week of the hearing level (two-way clustering). *** p<0.01, ** p<0.05, * p<0.1.

Table B.3: Assignment and learning by doing, parallel first stages

Dep. Var.	$h_{same-type}$, w/in 1yr	$h_{other-type}$, w/in 1yr	$h_{same-type}$, w/in 2yrs	$h_{other-type}$, w/in 2yrs	$h_{same-type}$, ever	$h_{other-type}$, ever
Model	LPM	LPM	LPM	LPM	LPM	LPM
Method	OLS	OLS	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
$n_{same-type}$, w/in 1yr	0.949***	-1.596***				
	(0.087)	(0.149)				
$n_{other-type}$, w/in 1yr	-0.095***	0.501***				
	(0.018)	(0.141)				
$n_{same-type}$, w/in 2yrs			1.582***	-0.837***		
			(0.098)	(0.171)		
$n_{other-type}$, w/in 2yrs			0.029	1.730***		
			(0.026)	(0.183)		
$n_{same-type}$, ever					3.103***	-0.845**
					(0.145)	(0.326)
$n_{other-type}$, ever					-0.054	2.481***
					(0.053)	(0.383)
C.-D. Wald F statistic (Joint)		4503		90563		178666
Judge FE	Yes	Yes	Yes			
Week of hearing FE	Yes	Yes	Yes	Yes	Yes	Yes
Type of case FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of judges	85	85	85	85	85	85
Observations	226,059	226,059	226,059	226,059	226,059	226,059

Note: An observation is a hearing of a judge and the type of cases. For each date of hearing-judge-type of case, $h_{same-type}$ w/in 1 yr (w/in 2 yrs; ever) is the (per 1000) number of hearings held by the judge before the hearing w/in 1yr (w/in 2yrs; ever). Similarly for $n_{other-type}$. $n_{same-type}$, *w/in 1yr (w/in 2yrs; ever)* is the (per 1000) number of cases of the same type assigned to the judge before the hearing w/in 1yr (w/in 2yrs; ever). Similarly for $n_{other-type}$. All the regressions control for the average number of parties involved in the trial. $C.-D.WaldFstatistic(Joint)$ denotes the minimum eigenvalue of the joint first-stage F-statistic matrix. Standard errors in parentheses are clustered at the judge and week of the hearing level (two-way clustering). *** p<0.01, ** p<0.05, * p<0.1.