

Performance Pay and Judicial Production: Evidence from Spain*

Manuel Bagues [†] Berta Esteve-Volart [‡]

June 2010

Abstract

This paper analyzes the effects of the introduction of a performance pay scheme rewarding Spanish judges. The Spanish top judicial authority established modules of production for every task judges undertake and then calculated production benchmarks. Since 2004, judges were awarded a 5% bonus if production exceeded the benchmark by 20%. We find that the introduction of this scheme increased the number of judges exceeding this threshold, and also increased average production. Nevertheless, we also observe that, consistent with a potential deterioration of intrinsic motivation, top performing judges significantly reduced their production.

Keywords: Judiciary, pay-per-performance, intrinsic motivation
JEL Classification: J33.

*We are grateful for the useful comments and suggestions provided by Mireia Artigot Golobardes, Benito Arruñada, Carlos Gómez, Fernando Gómez Pomar, Nic Morgan, Pablo Salvador Coderch, Javier Senac and by participants in seminar presentations at the Department of Law at Universitat Pompeu Fabra and at the FEDEA Workshop on Talent, Effort and Social Mobility. We are particularly grateful to the General Council of Judges (CGPJ) and to the Foundation for Applied Economic Research (FEDEA) for kindly giving us access to anonymized data on judges' performance. This paper was prepared while the second author was visiting the Bank of Spain and CEMFI, whose hospitality is gratefully acknowledged. We also acknowledge the financial support of the Social Sciences and Humanities Research Council of Canada and the Spanish Ministry of Science and Technology (research grants ECO2008-06395-C05-05 and ECO2008-01116). All remaining errors are our own.

[†]Universidad Carlos III and FEDEA, Madrid, Spain; email: mfbagues@emp.uc3m.es

[‡]York University, Toronto, Canada; email: berta@econ.yorku.ca

1 Introduction

Spain's judicial system is usually considered as being relatively inefficient. According to the World Economic Forum, Spain ranks below the world average in terms of the efficiency of its legal framework in settling disputes.¹ According to a survey conducted by the World Justice Project, one of its main problems is that cases before the Court are prone to delays.² In 2008, the number of pending cases, together with the number of new cases arising over the year, exceeded the number of resolved cases by 32% (Salvador Coderch 2010).³

In this regard, there exists a concern that judges may not be as productive as they could be, and policymakers have considered ways to increase judges' productivity. Spain's judges have traditionally been paid a fixed salary; however, in 2004 a simple performance pay scheme was introduced in order to directly link payment and performance and thus increase the productivity of judges. A production benchmark was calculated on the basis of tasks, position, and court of destination. Judges whose performance exceeded the benchmark by 20% would be rewarded with a 5% increase in their gross salary; judges whose production did not reach 80% of the benchmark might instead be penalized with a fee. The system was annulled by the Spanish Supreme Court in 2006. Upon further consultation with the judges, the system was reintroduced in 2007 with some changes. First, judges whose performance exceeded the benchmark by 20% would still be rewarded with a 5% increase in their salary, but now judges who were merely complying with the benchmark would also qualify for a bonus. In particular, judges whose production was between 100% and 120% of the benchmark were to receive a 3% bonus. In contrast, judges whose production did not reach 80% of the benchmark would not be penalized in any way.

In this paper we use information on the individual production of Spanish judges from 2003 through 2008, and study the effect of the 2004 and the 2007 performance pay schemes on the production of judges. The evidence suggests that the introduction of the first variable pay scheme resulted in an average increase in production, though the effects were mixed across the distribution. While the number of judges that exceeded the 120% threshold increased, we also find a reduction in the number of judges with very high performance. This is consistent with a deterioration of intrinsic incentives. In contrast, the second variable pay scheme resulted in increased salary costs without increasing average production. Note that the introduction of pay performance may have affected the quality of judicial decisions. Unfortunately, given that the quality of production cannot be assessed with the available data, the paper focuses on production implications, not productivity as adjusted by quality.

The rest of the paper is organized as follows. Section 2 describes the related literature. Section 3 offers background information on the Spanish Judiciary, and Section 4 describes the data. Section 5 turns to the empirical analysis. Finally, Section 6 discusses the results and concludes.

¹Spain ranks 68th out of a sample of 133 countries, just after Malawi, Mali, Zambia, Ghana and Uruguay, and immediately before Uganda. The Global Competitiveness Report 2009-2010, World Economic Forum, p. 354.

²Rule of Law Index 2009, World Justice Project, p. 142.

³This figure is calculated using data from *La Justicia Dato a Dato. Año 2008. Estadística Judicial* (Consejo General del Poder Judicial), p. 41 (Salvador Coderch 2010).

2 Related literature

There is a large literature analyzing the effects of performance pay. The rationale for introducing performance-related pay is simple: according to economic theory, paying on the basis of output should induce workers to exert more effort and therefore increase production. There is ample evidence that pay-for-performance might increase worker productivity (Lazear 2000, Oettinger 2001, Paarsch and Shearer 2000). Another advantage of performance pay might come from sorting: performance-related pay might attract the more productive employees; less productive employees may prefer to receive a salary (Lazear 1986, 2000; Oettinger 2001).

However, a number of disadvantages have been documented. For instance, performance pay can induce agents to game the system. Gaming behavior may result in larger quantity produced, at the cost of lower quality. This is a general concern associated with performance pay schemes (Paarsch and Shearer 2000). In the case of the Judiciary, where quality is difficult to observe, it may be too costly to ensure that quality does not deteriorate with performance pay (Holmstrom and Milgrom 1991). A number of authors have criticized the new compensation system in place in the Spanish Judiciary on the grounds that it may have distorted judicial practices (see Doménech Pascual 2008). Judges may have gamed the system in several ways. For instance, judges may have been taking a dissenting opinion in instances where they should not have dissented, just because dissenting opinions are well rewarded in the incentive scheme. Furthermore, judges may have split judicial processes into more than one, just to increase the number of processes, which may have been detrimental to the judicial outcome. In another example, judges may have given priority to the easier cases, in order to maximize the number of cases settled, while leaving the more difficult cases pending to future judges.

Moreover, pay performance might lead to distortions in the timing of judicial activity: agents may distribute their time between activities, or between, for example, terms in the year, in a way that maximizes individual utility, yet worsens the employer's utility. In the case of the Spanish judges, a judge may have distributed effort across semesters in order to reach the 120% production threshold, in a way that may be detrimental to the judicial outcome. Additionally, performance pay requires costly monitoring (Lazear 1986). Therefore, switching to performance pay involves costs; these will be larger the more difficult monitoring is. In the case of the Judiciary, due to the nature of the production of judges, monitoring might be particularly costly.

Finally, performance pay may reduce intrinsic motivation. There exists evidence that if an activity has an intrinsic motivation, when a monetary reward (an extrinsic motivation) is introduced, the latter displaces the former—the net effect may be a reduction in the activity (Gneezy and Rustichini 2000a,b).

3 Background

Even though performance pay was not formally introduced in the Judiciary until December 2003, the idea had precedents. In 1989, the General Council of Judges (*Consejo General del Poder Judicial*), Spain's top judicial authority, created standards establishing the expected production according to types of court. The standards were calculated simply based on one criterion: the number of sentences that were expected of every judge.

In 1998, the Council created new benchmarks, this time based on two criteria: (1) the judges' working hours, which were estimated to be 1,650 yearly, and (2) the expected time to complete every judicial task, calculated in terms of points, whereby every point was to be equivalent to one working hour. The benchmarks were to be updated frequently in light of new circumstances, in order to measure judicial activity accurately.⁴

The performance pay legislation was passed in May 2003 (BOE 126 - May 27, 2003) (henceforth, Act 15/2003), giving permission to the Council to establish production objectives or benchmarks for every court of destination in accord with either calculated modules, or some other criteria deemed appropriate (art. 8.1). The performance of every judge was then measured according to the modules, and was in turn compared to the benchmarks. Judges whose production was 20% greater than the benchmark were to be able to obtain a bonus corresponding to between 5 and 10% of their gross salary (art. 9.1). The percentage of the bonus was to be determined according to two factors: the total budget allocated to performance pay that year (never to exceed 5% of the total amount paid in salaries, art. 9.3.I), and the number of judges whose performance exceeded the objective in at least 20% (art. 9.3.II).⁵ By contrast, judges not reaching 80% of the benchmark were to experience a salary cut of 5%, provided an investigation had been launched, dealt with, and resolved by the Council for every judge, prior to the pay cut (art. 9.2).

In November 2003, Act 15/2003 refers two main points in the Act to the Council: (1) the establishment of the particular production objectives for every judicial destination, and (2) the rules to govern investigations in the cases that production was below 80% of the benchmark.

The particular production objectives were set in the Agreement of the Council plenary held on October 9, 2003; the rules to govern investigations were detailed in Rule 2/2003, as established in the Agreement of the Council plenary held on December 3, 2003.

The Rule 2/2003 and the modules set by the Council were challenged by most professional associations of judges. According to a survey of Spanish judges taken between October and November in 2005, 60% of judges opposed the performance pay system, and 36% thought that it was having a negative effect on quality.⁶ In February 2006, the Spanish Supreme Court overturned the performance pay scheme. The lack of motivation for the computation of the modules and the fact that modules were not specifically set according to court of destination were the two reasons on which the Court based its decision.

In November 2007, a different part of Act 15/2003 was invoked to create another performance pay scheme, this time in agreement with the professional associations of judges. The new system establishes a 5% bonus for judges whose production exceeds the benchmark in at least 20%, as well as a 3% bonus for judges who produce between 100 and 120% of the benchmark. It is noteworthy that the new scheme, and in line

⁴Modules were computed in 2000 considering the number of duties tackled by judges, and the time and effort expected from judges for every task. The modules were to help determine jurisdictional extensions, compatibility with other activities, nominations in discretionary positions for which dedication was important, and for the concession of honors.

⁵In practice, the total budget constraint was never binding. A decision was made to give 5%.

⁶A total of 1803 out of 4221 Spanish judges responded. "Informe sobre la encuesta a todos los jueces y magistrados en servicio activo", *Sección de Estudios Sociológicos y Estadísticos, Servicio de Planificación y Análisis de la Actividad Judicial, 2006*.

with judges' demands, does not penalize for lower production than the benchmark: judges producing less than 100% earn their entire fixed gross salary.

Figure 1 shows an example of the form (“formulario de módulo de dedicación”) that had to be filled out by judges in order to quantify the payment they were due.

4 Data

We have collected information about the personal characteristics of judges and their production in the period 2003-2008. The total number of judges is around 4,000. As shown in Table 1, about 46% of judges are female. The average judge is 43 years old, and entered the judiciary when she was 31 years old.

Using information from all public exams that have been held since 1990 we were able to identify the channel through which judges had access to the judiciary (Figure 2). Of those who entered the Judiciary after 1990, 84% did so through a public examination. Approximately 7% of judges entered through the competitions reserved for professionals with at least six years of experience (known as “third turn”), and 5% did through the competitions reserved for professionals with at least ten years of experience (known as “fourth turn”).

Figure 3 provides information about the different types of courts where judges can be assigned. The most common destinations are Civil and Criminal First Instance Courts, Regional Criminal Courts, and Superior Court of Justice.

Figures 4, 5, 6 and 7 display the available information on the production of judges in the first semester of 2003, the first semester of 2004, the second semester of 2005, and the first semester of 2008, respectively.

In 2003, production was only observable for judges producing more than 120% of the production benchmark (1275 judges out of 3890). In 2004, production was only observable for judges producing more than 80% of the production benchmark (3343 judges out of 4081). In 2005 and 2008 it is possible to observe the production achieved by all judges.

5 Empirical analysis

In this section we want to address two main questions. First, we would like to know which personal characteristics of judges are associated with higher production. Second, we would like to explore how judges reacted to the performance pay schemes. We distinguish between the pay system introduced in late 2003 and in place until 2005 (first treatment), and the system that was introduced in late 2007 (second treatment), building on the former treatment.

5.1 Determinants of production

In Table 2 we regress judicial production, as measured by the system of modules, on a number of personal characteristics of judges. In columns (1) and (2) we have the information for all years together. As seen in column (1), male judges produce significantly more—this is not surprising, because maternity leaves are not controlled for when calculating judges' production. Production is also higher among judges that entered the Judiciary when they were relatively younger. This is consistent with the

idea that individuals who need fewer years of preparation to pass the entry exam tend to be of better quality. Finally, we also observe that more experienced judges produce more. In column (2) we control for the type of court in which judges work, that is, their court of destination. Controlling for destination, we see that judges who had some professional experience before entering the Judiciary (as opposed to those who entered through a public exam, which is the omitted category here) produce more. In columns (3)-(6) we show results by year; results are roughly the same.

5.2 The effect of performance pay

5.2.1 First treatment: performance pay scheme introduced in 2004

The pay-for-performance system introduced in late 2003 was designed to award judges producing more than 20% over the benchmark, and to penalize judges whose production fell over 20% below the benchmark. In this setting, we would expect two effects: (1) that the number of judges who reach both thresholds increases, and (2) that judges cluster above these two thresholds. This sort of strategic behavior should be more likely once that judges have learned their true production in relation to the benchmarks.

Given data availability in 2003, we can only test the predictions that refer to the 120% threshold. Comparing the first and third bars in Figure 8, we observe that, consistent with the intuition above, the number of judges who were producing more than 120% increased substantially between the first semester of 2003 (before the treatment) and the second semester of 2005.⁷ Comparing the production of judges in the first semester of 2003 and the second semester of 2005 (Figure 10), we can see that the increase in the number of judges producing above 120% of the benchmark is mainly driven by the increase in the number of judges producing just above 120%.⁸ We also observe a small production peak around 100, which maybe reflects that some judges have decided to adjust their production to level of production which is expected from them by the new system.

At the same time, the figure shows that the number of judges producing more than 160% of the benchmark decreased with the first treatment. In particular, the number of judges performing above 160 fell by approximately 25%, from 329 judges to 251 judges. This is consistent with at least two hypotheses. First, it is possible that top performing judges were unaware of their relatively high performance. In that sense, giving them specific production benchmarks may have led to the realization that they were exerting more effort than most of their peers. In particular, the establishment of the benchmarks may have been seen as a signal that 100% is what is expected of every judge, but not more, and thus producing well beyond that would constitute *unnecessary* work. That may have made judges reconsider their effort levels. Second, it is also possible that the introduction of benchmarks with specific quantified modules, together with the

⁷The number of judges producing more than 120% increases relatively more at first, and then slightly decreases. That is, right after the introduction of the pay-per-performance scheme, some judges may have aimed for a certain production and, due to uncertainty about how the modules system works, may have ended up producing too much. This sort of *overshooting* behavior is also consistent with the information in Figure 9, showing the average production of judges: average production is by far the largest in the first semester of 2004; in the second semester of 2005, when treatment one is still in place, the average production decreases slightly, but is still larger than in 2003. The evolution of average production over time suggests that it may have taken some time for judges to learn how the modules system translates into production scores for them.

⁸A Kolmogorov-Smirnov test indicates that both distributions are significantly different.

monetary reward in terms of the bonus, has introduced an extrinsic motivation that has displaced the intrinsic motivation. This is consistent with the empirical evidence from other settings (Gneezy and Rustichini 2000a,b).

In summary, the introduction of the pay-per-performance scheme has two opposite effects. It has increased the number of judges producing more than 120% of the production objective, but it has reduced the number of *overproductive* judges. In order to estimate the net effect, we analyzed production over time controlling for individual fixed-effects. As shown in Figure 9, the average production of judges increased between 2003 and 2005 in around eight points (7%). In terms of cost-benefit analysis, a back of the envelope calculation suggests that the new pay-per-performance scheme was relatively successful in quantitative terms: it increased production by 7%, and it increased costs by only 2% (40% of judges obtained a 5% bonus). Naturally, in order to establish the overall effect of the pay-per-performance effect on productivity it would also be necessary to estimate the effect in terms of quality.

5.2.2 Second treatment: performance pay scheme introduced in 2008

We can now examine the effect of the second treatment: performance pay was re-introduced, with some changes, in November 2007, so we now compare the production of judges in the second semester of 2005 and in the first semester of 2008. Recall that, in the second treatment, one threshold was introduced, at 100% of the production benchmark (with 3% bonus), and one threshold was removed (at 80%). The threshold at 120% (with 5% bonus) was kept from the first treatment.

Consistent with the removal of the 80% threshold, Figure 11 shows that more judges are producing below 80% of the production benchmark. Consistent with the introduction of the 100% threshold, the figure shows an increase in the number of judges producing just above 100% and a reduction in the number of judges producing just above 120%.⁹ As shown in Figure 9, the effect in terms of the estimated average production of judges is negative: production in 2007 is one point lower than in 2005. The cost-benefit analysis of the new scheme is clearly negative: it increased costs, and it decreased production. The increase in costs is roughly equal 0.7% (while the number of judges performing above 120% is 1.5 percentage points lower, the number of judges entitled to a 3% bonus is equal to 20%). However, compared to the situation in 2003, before the first treatment was introduced, the introduction of pay-per-performance has increased production by 7% and has increased costs by only 2.6%.

6 Conclusion

This paper studies the quantitative effects of performance pay on judicial production. In 2003, the General Council of Judges, Spain's top judicial authority, established modules of production for every task judges undertake, and then calculated production benchmarks. In 2004 through 2005, judges were awarded a 5% bonus if production exceeded the benchmark by at least 20%. Since 2007, judges are awarded a 3% bonus if their production is at least 100% of the benchmark, and the bonus becomes 5% if the benchmark is exceeded by at least 20%.

We find that the production of judges is sensitive to the thresholds established. Overall, production per judge increased over the period. Nevertheless, we find that,

⁹A Kolmogorov-Smirnov test indicates that both distributions are significantly different.

while the first performance pay scheme increased production, the introduction of the 100% threshold and the elimination of the 80% threshold had a negative effect on the estimated average production.

Moreover, consistent with the deterioration of intrinsic motivation, we observe that a number of top performing judges reduce their production in the presence of performance pay. Intrinsic motivation may have deteriorated due to a number of reasons. For instance, judges may have perceived the establishment of detailed modules of production and benchmarks as a form of control, eroding their self-esteem and self-determination. Similarly, judges may have perceived the bonuses as a signal that judicial practice is not as interesting and fulfilling as they thought, or they may have interpreted the incentive scheme as distrust—judges would produce more with a bonus if they are expected to produce when a monetary reward is offered.

In further research, we would like to explore two important avenues in connection to this work. First, we would like to test for strategic timing on the part of judges (as Oyer 1998, who examined business seasonality effects on the basis of fiscal year ends). Second, we would like to test for the existence of distortions. Pay performance may have induced agents towards gaming behavior, decreasing judicial quality. In order to test this hypothesis we plan to gather data on the number of cases appealed and confirmed by a higher court.

References

- [1] Courty, Pascal and Gerard Marschke (2008), “A General Test for Distortions in Performance Measures”, *Review of Economics and Statistics*, Vol. 90(3), pp. 428-41.
- [2] Doménech Pascual, Gabriel (2008), “La Perniciosa Influencia de las Retribuciones Variables de los Jueces sobre el Sentido de sus Decisiones”, *InDret*, No. 3.
- [3] Gneezy, Uri and Aldo Rustichini (2000a), “Pay Enough or Don’t Pay at All”, *Quarterly Journal of Economics*, Vol. 115, pp. 791-810.
- [4] Gneezy, Uri and Aldo Rustichini (2000b), “A Fine Is a Price”, *Journal of Legal Studies*, Vol. 29(1), pp. 1-17.
- [5] Holmstrom, Bengt and Paul Milgrom (1991), “Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design”, *Journal of Law, Economics, & Organisation*, Vol. 7, pp. 24-52.
- [6] Lazear, Edward P. (1986), “Salaries and Piece Rates”, *Journal of Business*, Vol. 59(3), pp. 405-39.
- [7] Lazear, Edward P. (2000), “Performance Pay and Productivity”, *American Economic Review*, Vol. 90(5), pp. 1346-61.
- [8] Oettinger, G.S. (2001), “Do Piece Rates Influence Effort Choices? Evidence from Stadium Vendors”, *Economics Letters*, Vol. 73, pp. 117-23.
- [9] Oyer, Paul (1998), “Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality”, *Quarterly Journal of Economics*, Vol. 113(1), pp. 149-85.

- [10] Paarsch, Harry J. and Bruce Shearer (2000), “Piece Rates, Fixed Wages, and Incentive Effects: Statistical Evidence from Payroll Records”, *International Economic Review*, Vol. 41, pp.59-92.

Table 1: Descriptive statistics

	Mean	St. Deviation	Minimum	Maximum
Female	.46	.50	0	1
Age	43.4	9.2	26	70
Entry age	30.8	5.7	23	69
Experience	12.6	8.5	0	45

Notes: The table reports information about judges who were active in years 2003, 2004, 2005 or 2008. There are 15,961 observations, corresponding to 4,556 judges.

Table 2: Determinants of production

	All years	All years	2003	2004	2005	2008
Female	-2.85*** (0.80)	-3.22*** (1.00)	-4.80* (2.91)	-1.03 (1.62)	-3.50*** (1.34)	-4.18*** (1.45)
Age at entry in the Judiciary	-0.48*** (0.09)	-0.77*** (0.11)	-0.49 (0.31)	-0.51*** (0.17)	-0.56*** (0.16)	-0.48*** (0.18)
Experience	0.40*** (0.07)	1.28*** (0.11)	0.65** (0.27)	0.56*** (0.15)	0.72*** (0.12)	-0.06 (0.13)
Entered the Judiciary before 1990	-12.43*** (1.29)	-2.36 (1.71)	-20.67*** (4.63)	-13.27*** (2.61)	-14.12*** (2.21)	-9.41*** (2.45)
Entered the Judiciary through "fourth turn"	3.58 (2.46)	13.39*** (3.17)	-2.50 (8.91)	-4.82 (5.00)	10.55*** (4.08)	7.65* (4.45)
Entered the Judiciary through "third turn"	2.67 (1.99)	8.35*** (2.42)	1.89 (6.92)	2.41 (3.94)	2.64 (3.31)	4.01 (3.73)
Unknown channel of entry	-4.58* (2.69)	7.31** (3.34)	-7.01 (9.60)	-6.32 (5.32)	-5.49 (4.34)	-0.87 (5.16)
Constant	123.27*** (3.06)	141.02*** (5.58)	99.18*** (10.34)	128.34*** (5.79)	122.33*** (5.22)	128.29*** (6.15)
Type of court dummies	No	Yes	No	No	No	No
N	15961	9383	3890	4081	4027	3963

Notes: The table reports information from a tobit regression of judicial production. Data are left censored for judges producing less than 120 in 2003 or less than 80 in 2004. The benchmark is a male judge who entered the Judiciary through a public exam. *Entered the Judiciary through "third turn"* refers to those individuals that accessed the judiciary through a competitive process restricted to professionals with at least ten years of experience. *Entered the Judiciary through "third turn"* refers to those individuals that accessed the judiciary through a competitive process restricted to professionals with at least six years of experience. Standard errors in parentheses are clustered at the individual level. *significant at 10%, **significant at 5%; ***significant at 1%.

Figure 1: Application form for judges in a Criminal First Instance Court

Periodo de Actuación (dd/mm/aa): del _____ al _____

1er. Apellido _____ 2º Apellido _____

Nombre: _____ Con.NIF: _____

Titular del Juzgado de Instrucción núm. _____ de _____

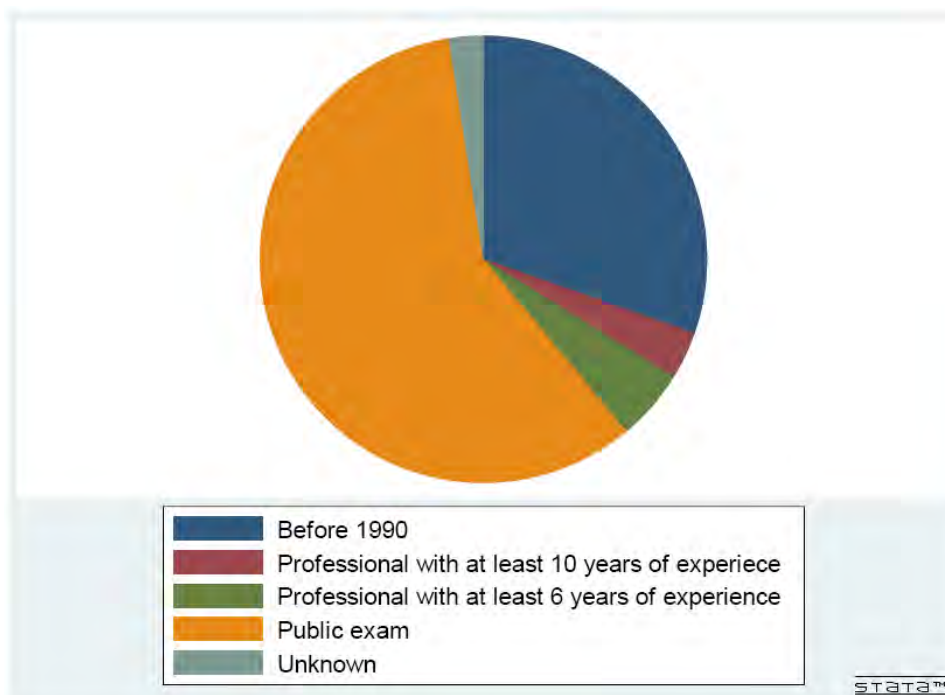
ASUNTOS TERMINADOS	Puntuación	Nº asuntos	Total H/punto
Sumarios concluidos	12		
Causas de la Ley del Jurado elevados	18		
Procedimientos Abreviados elevados (*)	3		
Juicios de faltas con sentencia (*)	1,75		
Habeas Corpus	1		
Guardias de presencia	100		
Guardias de disponibilidad	100		
Sentencias en juicios rápidos por delito con conformidad	2		
Juicio de Faltas inmediato con sentencia	2		
Resolución de solicitud de Orden de protección a víctimas de violencia doméstica	2		
Asuntos de excepcional dedicación			
TOTAL HORAS / PUNTOS			

(*) Las horas/puntos deben multiplicarse por 1,05

% sobre Módulo.

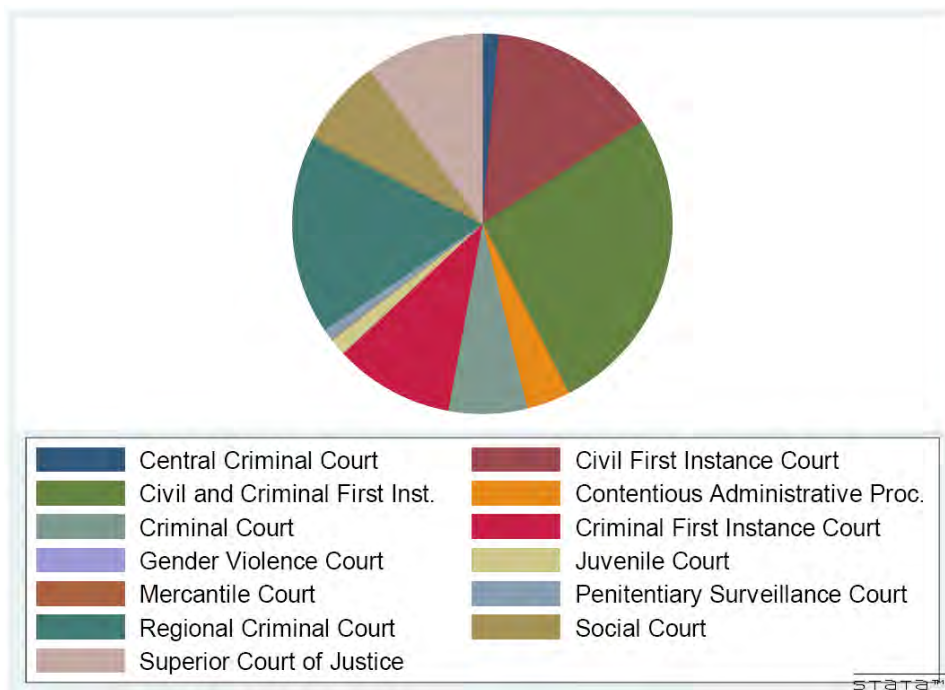
Note: The benchmark for production is 1450 points or hours. Source: Reglamento 2/2003, as published in the Official State Bulletin (BOE) on December 17th, 2003, page 44849.

Figure 2: Channel of entry in the Judiciary



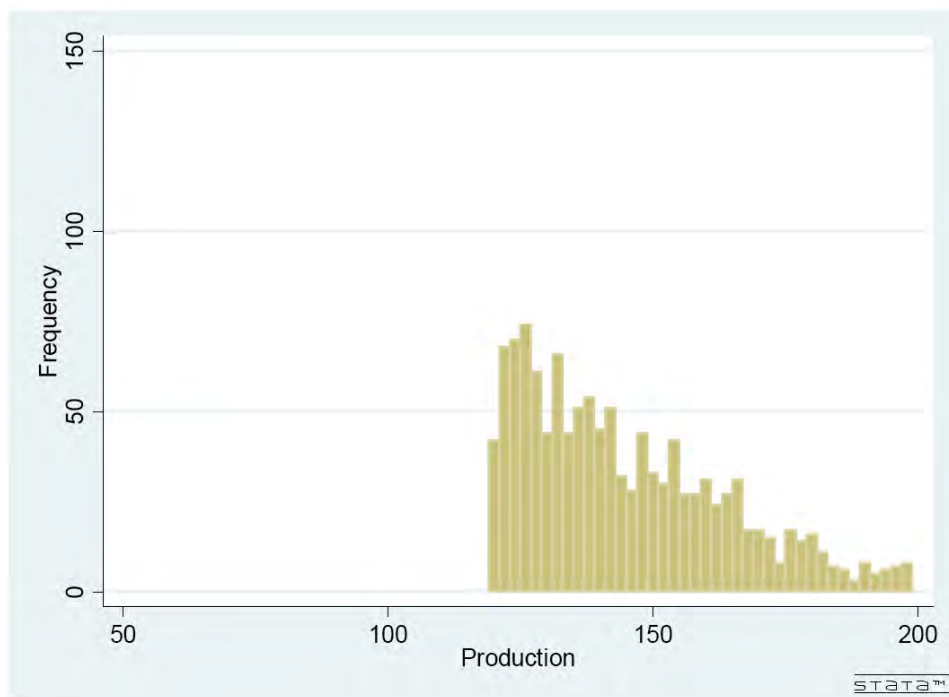
Source: Authors' calculations using information from all hirings since 1990.

Figure 3: Court of destination



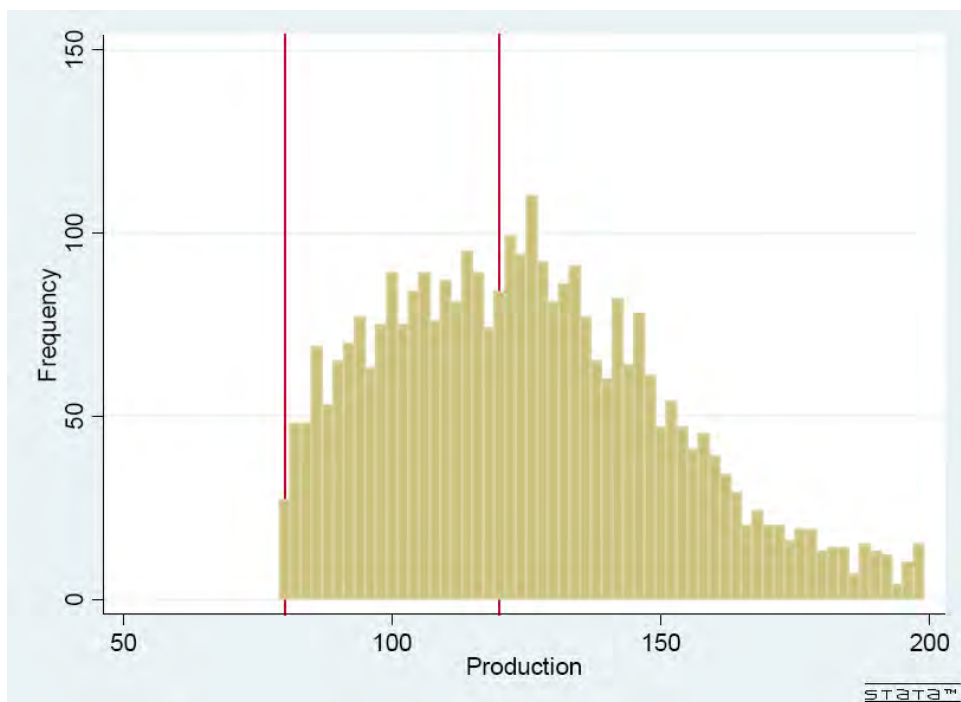
Source: Authors' calculations.

Figure 4: Production in the first semester of 2003



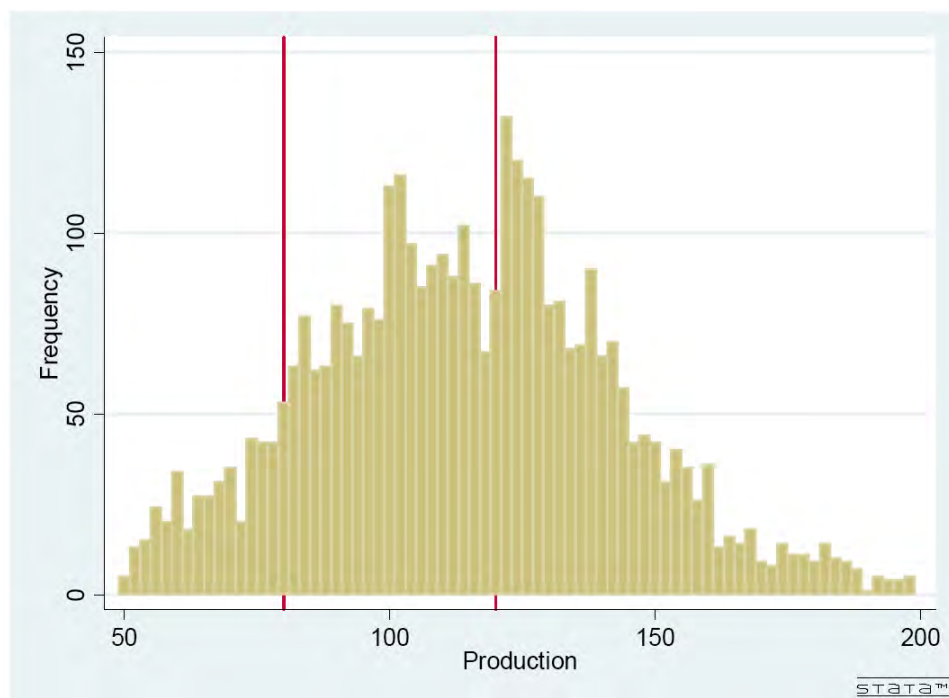
Notes: In January 2004 judges were asked to voluntarily report their production in the first and the second terms of 2003. Judges whose production was above 120% were paid a 5% bonus. Information on production that year is only available for individuals whose production was above 120% and thus obtained the bonus (1275 out of 3890 judges).

Figure 5: Production in the first semester of 2004



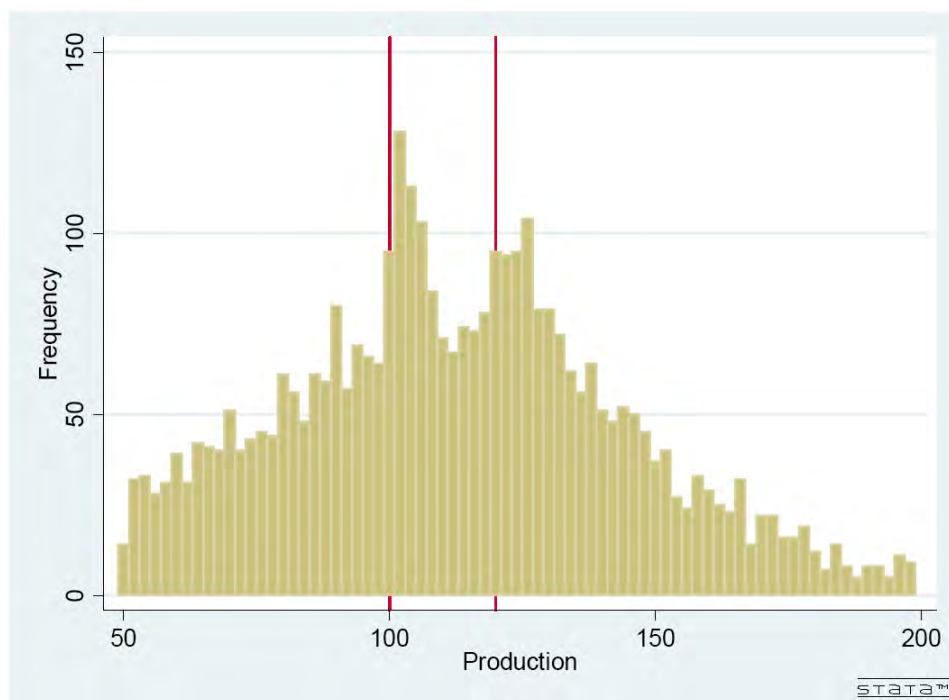
Notes: In 2004 it was mandatory for judges to report their production, but precise information about the level of production is only available for judges whose production was above 80% (3343 out of 4081 judges). The vertical bars represent the 80 and the 120 production thresholds, respectively. According to the pay-per-performance scheme at work at the time, judges whose performance exceeded the 120 benchmark were rewarded with a 5% increase in their gross salary and judges whose production did not reach 80 might instead be penalized with a 5% fee.

Figure 6: Production in the second semester of 2005



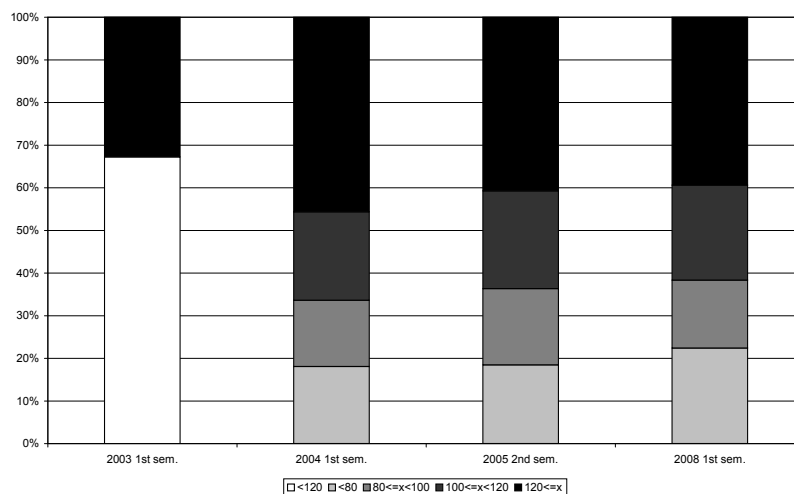
Notes: Authors' calculations. The vertical bars represent the 80 and the 120 production thresholds, respectively. According to the pay-per-performance scheme at work at the time, judges whose performance exceeded the 120 benchmark were rewarded with a 5% increase in their gross salary and judges whose production did not reach 80 might instead be penalized with a 5% fee.

Figure 7: Production in the first semester of 2008



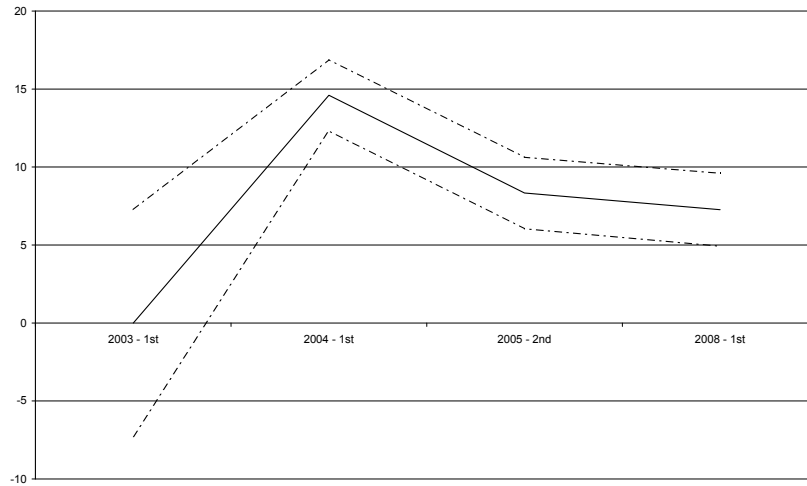
Notes: Authors' calculations. The vertical bars represent the 100 and the 120 production thresholds, respectively. According to the pay-per-performance scheme at work at the time, judges whose performance exceeded the 100 (120) benchmark were rewarded with a 3% (5%) increase in their gross salary.

Figure 8: Production in years 2003-2008



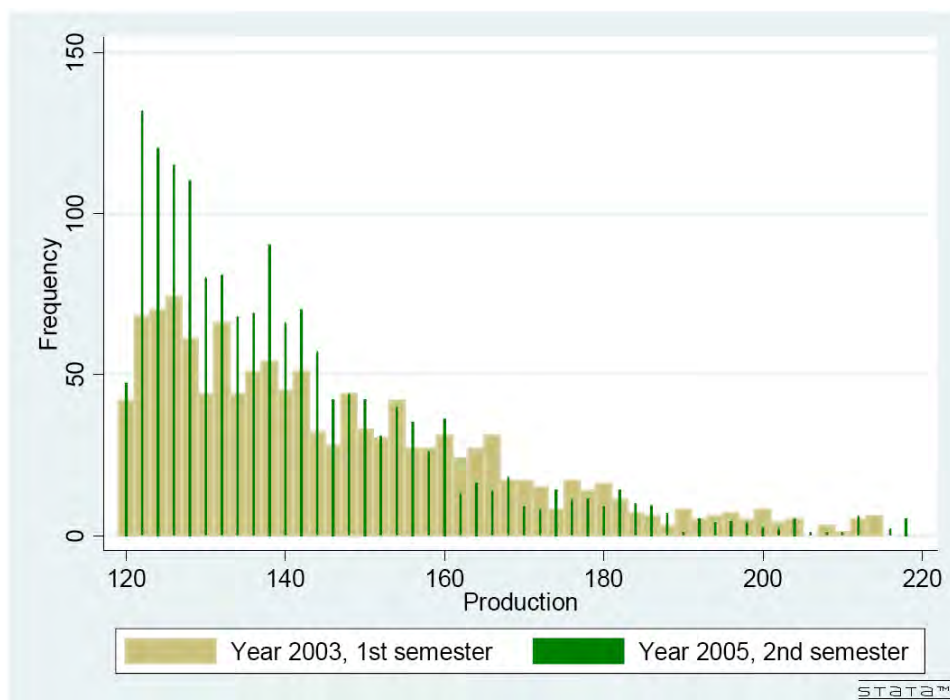
Notes: Authors' calculations. The bars indicate the share of judges whose production falls in a certain interval. For 2003 it is only possible to distinguish between judges producing more and less than 120. For years 2004, 2005 and 2008 it is possible to observe the number of judges producing less than 80, between 80 and 100, between 100 and 120, and above 120.

Figure 9: Estimated average production in years 2003-2008



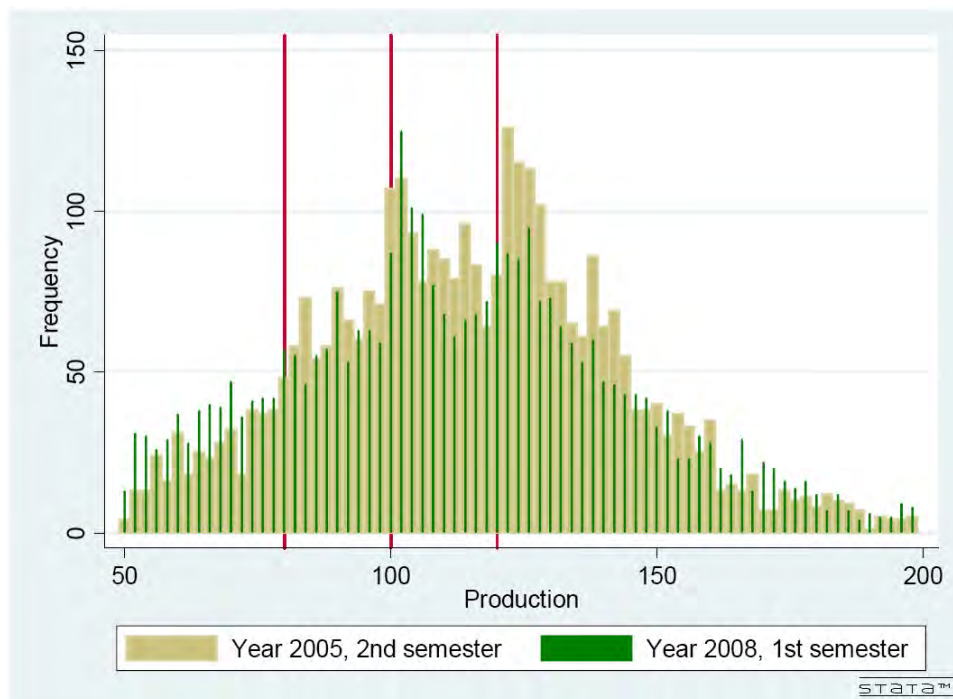
Notes: The figure reports the results of a tobit estimation where the left-hand-side is production at the individual level, and the right-hand-side includes dummies for individuals and for years. The upper and lower dashed lines indicate confidence intervals at the 95% level.

Figure 10: First treatment: comparing production in the first semester of 2003 and the second semester of 2005



Notes: The vertical lines represent the 80 and 120 production thresholds, respectively.

Figure 11: Second Treatment: comparing production in the second semester of 2005 and the first semester of 2008



Notes: The vertical lines represent the 80, 100 and 120 production thresholds, respectively.