Unraveling and Judge Productivity in the Market for Federal Judicial Law Clerks: Evidence and Proposal

Daniel L. Chen Yinghua He Takuro Yamashita

October 11, 2023

Federal Judicial Clerkship Market

- Every year,
  - judges compete for elite law-school graduates (as "clerks") to aid them with their workload;
  - top of those students compete for these highly prestigious positions, leading to prestigious jobs afterwards.
- The market often (mostly) unravelled.
  - Some judges move early (e.g., at the beginning of the second year), even if they agreed on a later timing of interviews.
  - Past reform agreements (1983, 86, 90, 05) all failed.
- Unraveling is a concern for many other hiring markets (and more generally, for dynamic matchings).
  - (One of) Problem: Informational inefficiency
- This paper:
  - Better understanding of this phenomenon, both theoretically and empirically;
  - Proposal for a reform, leveraging the dynamic / repeated aspect of the market explicitly.

Empirical findings

Reforms (all failed)

- Reform 1 (1983)
  - Judges not consider applications before September 15 of the students' third year.
- Reform 2 (1986)
  - Judges not consider applications before April 1.
- Reform 3 (1990)
  - No job offers allowed before May 1st of the applicant's second year.
- Reform 4 (2005)
  - Sept 15 is the date for scheduling of interviews and Sept 22 the date for offers.

Data

- 380,000 published decisions (over a million judge votes) in U.S. Circuit Courts since 1891.
- Full citation network between the cases.
- Detailed metadata for each case, including the court, publication date, and authoring judge.

Empirical result

- Our empirical result suggests:
  1. Overall, judges' production "improves" right after the reform.
  2. But there is heterogeneity:
     - High productivity judges are better off.
     - Low productivity judges are worse off.

1. Overall effect

- The output of judge i in circuit c in year t is given by:

$$q_{i,c,t} = \beta Reform_t + \alpha_i + OtherControls_{i,c,t} + \epsilon_{i,c,t},$$

where

- $Reform_t \in \{0,1\}$: 1 if year t is a reform year;
- α_i is judge fixed effects.
- *OtherControls* includes i's years of experience and its square, Circuit-specific time trends, case types...

Dependent variable (yearly)	Cases Published (1)	Citations (within) (2)	Citations (outside) (3)	Citations (total) (4)	Cases Reversed (5)
Reform	0.639***	8.397*	-3.306*	5.090	0.00109
	(0.230)	(4.588)	(1.913)	(6.084)	(0.0102)
N	7853	7853	7853	7853	7853
R-sq	0.756	0.689	0.689	0.702	0.202

Notes: Dependent variable calculated as the yearly total during a market year (September to August). Standard errors are in parentheses.

- Thus, overall, the reform increased 0.639 cases published per year = 3% of the average production (18.5 cases).

2. Heterogeneity

$$q_{i,c,t} = \beta_{low}(Reform_t \times Low_i) + \beta_{high}(Reform_t \times High_i) + \ldots,$$

- A judge is a high-prod (low-prod) judge if his productivity fixed effect if above (below) the median in non-reform years.
- (Appear to be very positively related with "prestige")

Dependent variable (yearly)	Cases Published (1)	Citations (within) (2)	Citations (outside) (3)	Citations (total) (4)	Cases Reversed (5)
Reform x Low Productivity	-1.062**	0.983	-10.81***	-9.828	-0.00290
	(0.446)	(9.461)	(4.021)	(12.68)	(0.0252)
Reform x High Productivity	3.841***	61.07***	6.418	67.49***	-0.00654
	(0.587)	(12.77)	(5.447)	(17.05)	(0.0346)
N	1939	1939	1939	1939	1939
R-sq	0.836	0.773	0.765	0.783	0.334

Notes: Dependent variable calculated as the yearly total during a market year (September to August). Standard errors are in parentheses.

- Thus, Low-prod (High-prod) judges were worse (better) off.

Takeaways

- Takeaways:
  - Reforms were not Pareto improving.
  - Low productivity judges have more incentive to deviate (Consistent with some anecdotes).
- Possible interpretation?
  - Each reform is to make all judges wait until some point, good for informational efficiency.
  - But this makes the high-prod judges take all the best students.
  - The low-prod judges could be better off by moving earlier and try cherry-picking some of them...though risky.
- We build a theoretical model based on this interpretation, in order to:
  - confirm the difficulty of achieving information efficiency, due to the low-prod judges' incentive to move early;
  - observe unravelling without any intervention; and
  - propose a possible "second-best" reform from a more dynamic/repeated-game viewpoint.

Theoretical model

Primitive

- J firms
- I workers $(I > J)$
- All workers have the same preference over the firms: $u(1) > u(2) > \ldots > u(J) > \underline{u}$, where $\underline{u}$ is the unmatched payoff.
- Worker i:
  1. receives a signal $s_i \in \{L, H\}$ at $t = 1$;
     - $p = \Pr(s_i = H)$;
     - $s_i = \Pr(\theta_i = 1 | s_i)$.
  2. realizes his ability $\theta_i \in \{0, 1\}$ at $t = 2$.
     - Hence, $q \equiv \Pr(\theta_i = 1) = pH + (1 - p)L$.
     - Assume p, L are small so that $q = O(p)$.
- All firms have the same preference over the workers, based on their abilities: i is preferred to i' iff $\theta_i = 1 > 0 = \theta_j$.

(Stage) Game

1. Each firm simultaneously decides whether to make an offer at $t = 1$, and to whom.
   - For simplicity, each firm can make only one offer. Thus, if a firm makes it at $t = 1$, he cannot make any offer at $t = 2$. "Making an offer is costly" (search, negotiation within the firm, etc.)
2. Each worker who is offered at $t = 1$ decides which one to take (or none). Any matched worker (and firm) leaves.
3. Each firm who hasn't made an offer at $t = 1$ makes an offer at $t = 2$ to one of the remaining workers.
4. Each worker who is offered at $t = 2$ decides which one to take.

- Recall: At $t = 1$, each i only knows s_i; while at $t = 2$, each i knows θ_i. Therefore, early offers tend to imply "inefficient" outcomes.

Efficiency

- Efficiency?
  - Informational efficiency:
    - It would be a waste if some firm hires i with $\theta_i = 0$ while some worker i' with $\theta_{i'} = 1$ is unmatched.
    - Measure of informational efficiency = Total θ_i among hired.
  - (Matching efficiency)
    - Maybe natural to assume supermodular preferences so that the assortative matching is most efficient? Or the opposite so that the anti-assortative one is the most efficient?
    - Here, we take an agnostic stance. Results should be robust as long as the informational efficiency is more dominant than the matching efficiency.

Result 1: Impossibility

Theorem

With sufficiently small p and $\underline{u}$, it is not an equilibrium that all firms hire at $t = 2$.

- Other parametric possibilities: I much larger than J, and/or H small.

Proof

- Let all firms hire at $t = 2$. For firm J,

$$\begin{aligned} v_J = \Pr(\#(\theta_i = 1) \geq J) &= 1 - [(1-q)^I + Iq(1-q)^{I-1} + O(p^2)] \\ &= 0 + O(p^2). \end{aligned}$$

- If J deviates and hires at $t = 1$, assuming that the offered worker would accept it, J's payoff:

$$\begin{aligned} \hat{v}_J &= (1-p)^I L + (1 - (1-p)^I) H \\ &= L + Ip(H - L) + O(p^2) \\ &> v_J. \end{aligned}$$

- Indeed, the worker would accept it given any signal at $t = 1$, as:

$$u(J) \geq Hu(1) + (1-H)\frac{\sum_{j=1}^{J} u(j) + (I-J)\underline{u}}{I}$$

if $\underline{u}$ is sufficiently small.

Result 2: Static NE

- Without any intervention, it is natural to assume that a static Nash equilibrium is to be played.
- The equilibrium must involve some early offers, as implied by Result 1.
- It may be in mixed strategies. To see this, assume $J = 2$ and small $p, \underline{u}$ (like in the first result). Payoff table (row = Firm 1; column = Firm 2):

	$t = 1$	$t = 2$
$t = 1$	$(L + Ip(H - L), L)$	$(L + Ip(H - L), (I - 1)L)$
$t = 2$	$((I - 1)L, L + Ip(H - L))$	$(Iq, 0)$

where all terms are up to $O(p^2)$.

Proposition

Assume $L + Ip(H - L) > (I - 1)L$ (e.g., L is small). Then, the unique NE is in mixed strategies.

Proof.

From any pure strategy profile, at least one firm has a strict incentive to deviate. □

- Remark: If $L + Ip(H - L) < (I - 1)L$, then $\exists$ pure NE $(t = 2, t = 1)$. However, the same condition implies a mixed NE with $J = 3$.

Repeated-game perspective?

- Firms are constantly hiring new-coming workers every year. This dynamic aspect can be exploited.
- Consider repeated games (with year $y = 1, 2, \ldots$) with
  - long-lived firms (with persistent characteristics, that is, firm 1 is the most preferred by all workers every year, etc) and
  - one-period-lived workers (they are in the hiring market once in their lives).
  - At each year y, the long-lived firms and newly arrived one-period-lived workers play the previous stage game, where each long-lived firm maximizes the discounted (by a discount factor δ) sum of his stage-game payoffs.

Result 3: Impossiblity of full efficiency again

- With large δ, the folk-theorem argument implies that many equilbria are possible, but the most efficient one is still impossible with small $p, \underline{u}$.

Proposition

For any $\delta < 1$, with sufficiently small $p, \underline{u}$, it is impossible to make all firms hire at $t = 2$ with probability one at every year y.

Proof.

If it were possible, then the discounted sum of payoffs of any firm other than 1 is $\frac{1}{1-\delta}(0 + O(p^2))$, while the deviation payoff is at least $L + lp(H - L)$. $\square$

Second-best informational efficiency

- Although the fully efficient outcome is impossible, dynamic incentives make some improvement possible.
- If only one firm hires at $t = 1$ (and the rest hire at $t = 2$), we call the outcome *second-best informationally efficient*, in the following sense.

Proposition

Consider any outcome of the stage game where $J' \subseteq J$ hire at $t = 1$ (and the rest hire at $t = 2$) with $|J'| \geq 2$. Higher informational efficiency is achieved if only $j^* = \min J'$ hires at $t = 1$

Proof.

In either case, j^* hires the same worker (say i). In case only j^* hires at $t = 1$, any worker $k \neq i$ is hired at $t = 2$ if $\theta_k = 1$. $\square$

Result 4: Random unilateral-early-mover mechanism

Definition

A random unilateral-early-mover mechanism lets one firm i move at $t = 1$ (while all the other firms wait until $t = 2$), where i may be chosen in a history-dependent and stochastic manner.

- With appropriate probabilities for each i's early moving, the dynamic incentive can be guaranteed with high δ.
- First, assume $J = 2$ with small $p, \underline{u}$, and also $L + Ip(H - L) > (I - 1)L$.
  - A static NE is in mixed strategies. The corresponding payoffs (u_1^*, u_2^*) are their minmax payoffs.
  - Hence, Nash reversion is the harshest punishment in case of any deviation.
  - What should we do "on-path"?

Rotating pattern in the optimal mechanism

- Intuition:
  - If $(t_1, t_2) = (2, 2)$ is to be played at year y, firm 2 has a static incentive to deviate. Thus, at $y + 1$, relatively high probability should be put on $(t_1, t_2) = (2, 1)$.
  - If $t = (2, 1)$ is supposed to be played at year y, firm 1 has a static incentive to deviate. Thus, at $y + 1$, relatively high probability should be put on $t = (2, 2)$.
  - Any deviation reverts to the mixed NE (with the minmax payoffs).
- Let α_{22} (α_{21}) be the probability of playing $(2, 2)$ at $y + 1$ given $(2, 2)$ ($(2, 1)$) played at y.
- Let $U_i(2, 2)$ be i's on-path continuation payoff given $(2, 2)$ is supposed to be played (and similarly $U_i(2, 1)$).

- By definition:
$$\begin{aligned} U_i(2,2) &= u_i(2,2) + \delta(\alpha_{22} U_i(2,2) + (1-\alpha_{22}) U_i(2,1)) \\ U_i(2,1) &= (I-1)L + \delta(\alpha_{21} U_i(2,2) + (1-\alpha_{21}) U_i(2,1)) \end{aligned}$$
Hence, $U_i = (I_2 - \delta\alpha)^{-1} u_i$.
- IC:
$$U_1(2,1) \geq L + Ip(H-L) + \frac{\delta}{1-\delta} u_1^*$$
$$U_1(2,2) \geq L + Ip(H-L) + \frac{\delta}{1-\delta} u_2^*.$$
- These conditions imply bounds on α_{22}, α_{21}, as a function of δ.
  - If no α can satisfy all the bounds, then the mechanism cannot assign probabilities only on $(2,2)$ and $(2,1)$.
  - If some α can satisfy all the bounds, then the optimal mechanism sets the one that maximizes the informational efficiency (higher α).

Theorem

There exists $\delta^* > 0$ s.t., for $\delta > \delta^*$, second-best information efficiency is achieved.

$J \geq 3$

- With $J \geq 3$, two changes:
  - Not only firm 2 but any firm $i \geq 3$ also has an incentive to deviate from the first best. Thus, we need to make every firm $i \neq 1$ a unilateral early mover. The corresponding probability is to make i's payoff equal to his minmax payoff.
  - As opposed to the case with $J = 2$, the NE may not attain the players minmax payoffs. More complicated off-path punishment schemes could potentially improve the informational efficiency. However, the qualitative feature would stay the same: we need to make every firm $i \neq 1$ a unilateral early mover, with a strict positive probability.

Conclusion

- Matching market between judges and clerks.
  - Prestigious and competitive
  - Unravelling: all failed past reforms
- Empirical analysis, suggesting:
  - Overall productivity improvement by reforms; but
  - Heterogeneity: Low-prod judges were hurt.
- Theoretical investigation of the optimal mechanism
  - Impossibility of the first-best information efficiency;
  - Possibility (with high δ) of the second-best information efficiency.

* Feel free to tell me if you are interested in adding a theory section (or writing a spin-off paper) to your empirical project!!