REFINING THE JUDICIAL SALARY/JUDICIAL PERFORMANCE DEBATE: A RESPONSE TO PROFESSORS CROSS, CZARNEZKI, HENDERSON, MARKS, AND ZORN

SCOTT BAKER[*]

Three years ago, I began collecting data for the article "Should We Pay Federal Circuit Judges More?"[1] At the outset, I had a hunch: Low judicial pay was affecting judicial performance. Specifically, low pay resulted in federal circuit judges that were more partisan, more prone to leisure, and more motivated by the prospect of their own influence. I suspected to discover a statistically significant and economically meaningful link between judicial pay and judicial performance. Scholars, after all, always treat the converse – statistically insignificant results – with skepticism. The failure to reject a null hypothesis of no association does not prove that the variables, in fact, lack association.

After conducting the analysis, the data did not support my hunch. For most of my measures of judicial performance, I did not find a statistically significant effect. These "non-results" were fairly precise, however. The confidence intervals of the estimates were tight around zero, enabling me to reject large effects of salary on the performance measures.[2] Given this, the results stood in stark contrast to Chief Justice John Roberts's hypothesis that low judicial pay was causing a constitutional crisis. So, I decided to publish the article. At the same time, I placed my data and the statistical programs underlying the analysis in the public domain. That way, other researchers could replicate, critique, and improve on the project. In a welcome development, that is exactly what has happened.

The three replies in this issue represent generous and illuminating responses to the work. They offer valid criticisms and important refinements to the claims made in the paper. Indeed, I agree with most of what these scholars say. But after all this discussion about statistics, economic theory, and data, a question remains unanswered: What, if any, impact of judicial pay on judicial performance justifies a pay raise? Framed this way, notice how the data have

[1] Scott Baker, *Should We Pay Federal Circuit Judges More?*, 88 B.U. L. REV. 63 (2008).

[2] In the original paper, all the significance results are from two-tailed tests. The results are much the same if a one-tailed significance test is employed instead. *See* Christopher Zorn, William D. Henderson, & Jason J. Czarnezki, *Working Class Judges*, 88 B.U. L. REV. 829, 834 tbl.1 (2008).

shifted the debate from the assertion of a "constitutional crisis" toward a deeper investigation about the size and kind of concrete results we hope to achieve with higher pay. My article just starts that discussion. Coupled with these replies, my hope is that this work will spur on further efforts to uncover links between judicial pay and judicial performance.[3]

My response to the replies comes in three parts. Part I responds to Frank Cross's concerns about the statistical analysis itself and the inferences drawn from that analysis. Part II considers the effects of salary increases on judges coming from "top-five" markets as identified by Jason Czarnezki, Bill Henderson, and Chris Zorn (collectively "CHZ"). Part III comments on Stephen Marks's two objections to my measure of a judge's opportunity cost.

I.   WHAT'S THE NULL?

Professor Cross makes four points in his reply. First, he suggests several reasons why my estimate of opportunity cost (NETCOST) does not capture the real opportunity cost for a specific judge. Most salient is the crudeness of the law firm salary data – it reflects average partnership income by region. There is no reason to suspect that judges from a specific city in a region would make the average partner salary for that region overall. Second, he questions whether any of the judicial performance measures used truly capture "judicial quality." If not, the failure to find a statistical correlation between a judge's financial sacrifice and those measures means little. Third, he argues that my voting pattern results are limited because I fail to control for possible panel effects. The omission of this relevant independent variable will tend to bias the results. Depending on the direction of the bias, the regressions will overestimate or underestimate the true effect of financial sacrifice on judicial performance. Finally, he asserts that I put too much faith in the failure to reject the null hypothesis. In conventional statistical significance testing, the null hypothesis is that no relationship exists between the independent and dependent variables. Strictly speaking, the study cannot confidently rule out that the estimated coefficients are the result of chance, rather than reflecting an

---

[3] The ball has started rolling on this topic. *See* Stephen Choi, Mitu Gulati, & Eric Posner, *Are Judges Overpaid?: A Skeptical Response to the Judicial Salary Debate* (Univ. of Chicago Law and Economics, Olin Working Paper 376), *available at* papers.ssrn.com/sol3/papers.cfm?abstract_id=1077295 (finding that salary does not have much of an impact on the behavior of state court judges); Reed Watson & Matthew Wolfe, *Comparing Judicial Compensation: Apples, Oranges, and Cherry Picking* (unpublished manuscript on file with author) (finding that, when making international comparisons of judicial salaries, the justices "cherry pick the highest paid judiciaries, but not necessarily the best performing ones"). The debate about judicial salaries continues to rage in the public domain. *See* George F. Will, *Bargain Basement Judiciary*, WASH. POST, Mar. 23, 2008, at B07. A New York State Supreme Court Justice has even filed a lawsuit trying to force the state comptroller to increase pay. *See* Anemona Hartocollis, *New York's Top Judge Sues Over Judicial Pay*, N.Y. TIMES, Apr. 11, 2008, at A4.

underlying association between the variables of interest.[4] In short, Professor Cross argues that my data and analysis are just too limited to do the job asked, i.e., assessing whether there is empirical support for the proposition that judicial pay impacts judicial performance. Given the failure of the statistical project, he advocates relying on anecdotal evidence and economic theory, including consideration of behavioral economics.

I agree that the opportunity cost measure is imprecise. The article corrects the imprecision a few different ways, but none of the corrections are completely satisfying.[5] The best data would come from the judges themselves – self-reports of the income they gave up for the bench. Absent that, I don't know of a better way to estimate opportunity costs or a better data source to use. For the reasons discussed in the article, the estimate likely correlates with a judge's true opportunity cost.[6] Even if all the judicial candidates would have been "above average" partners, rather than "average" partners, the analysis still holds if average and above-average partnership incomes move together.[7] True, if some candidates are better private sector lawyers than others, the assumption that all nominees forgo an average or above average partnership salary weakens the analysis. Nonetheless, the "private practice" dummy variable should pick up part of any differential effects. Suppose, as is likely, that nominees coming directly from private practice are the better, more successful private sector lawyers – their relative success made them more likely to remain in private practice before their appointment. The "private practice" dummy variable should then capture differences in opportunity cost attributable to differences in the success at practicing law. In the end, unfortunately, the precise degree of correlation between NETCOST and the judges' true opportunity cost is hard to know. As such, all the results must be taken with this measurement error in mind.[8]

---

[4] *See* DAVID FREEDMAN ET AL., STATISTICS 478 (3d ed. 1998).

[5] Corrections include: (1) adding a dummy variable, TOPFIVE, for whether the judge came from a city in a top five legal market and (2) adding a variable interacting TOPFIVE with the NETCOST measure. For a fuller discussion of this interaction dummy, see Zorn, Henderson & Czarneski, *supra* note 2 and *infra* Part II.

[6] Most of the judges in the sample (239 out of 259) remained in the same region for the ten years prior to taking the bench. Part III.A, *infra*, discusses the consequences of relaxing the assumption that no judges would have left their region for a law firm job in a higher paying region.

[7] *See* JEFFREY M. WOOLDRIDGE, INTRODUCTORY ECONOMETRICS 40 (explaining how the variance in – not the absolute value of – independent variables determines the predictive power of a regression analysis).

[8] On the significant consequences of mis-measuring independent variables, see PETER KENNEDY, A GUIDE TO ECONOMETRICS 137 (3d ed. 1992).

Professor Cross is also correct that my judicial performance measures don't perfectly capture judicial quality.[9] But if not these measures, what measures exist to assess the quality of federal circuit court judges? The short answer – none – is unsatisfying. The "I-know-a-good-federal-circuit-judge-when-I-see-one" angle is hard to test. The article employs every metric I could think of, including most of the metrics used by scholars studying the circuit courts.[10] At the start of the project, my sense was that Congress would care about these measures when considering a judicial pay raise. Suppose that the study had found statistically significant and economically meaningful correlations between financial sacrifice and voting patterns in controversial cases, dissent rates, the time it takes to render decisions, citation practices in opinion writing, and the number of outside circuit citations opinions tend to garner. In that case, I suspect Chief Justice John Roberts himself would have pointed to the study to "prove" to Congress the need for higher salaries.

---

[9] Professor Marks raises this same concern in his reply. *See* Steven Marks, *A Comment on the Relationship Between Judicial Salary and Judicial Quality*, 88 B.U. L. REV. 843 (2008).

[10] For scholars studying voting patterns in the circuit courts, see, for example, VIRGINIA A. HETTINGER, STEFANIE A. LINDQUIST & WENDY L. MARTINEK, JUDGING ON A COLLEGIAL COURT: INFLUENCES ON FEDERAL APPELLATE DECISION MAKING (2006); Frank B. Cross, *Decisionmaking in the U.S. Circuit Courts of Appeals*, 91 CAL. L. REV. 1457 (2003); Orin S. Kerr, *Shedding Light on* Chevron*: An Empirical Study of the* Chevron *Doctrine in the U.S. Courts of Appeals*, 15 YALE J. ON REG. 35-37 (1998); Richard L. Revesz, *Environmental Regulation, Ideology, and the D.C. Circuit*, 83 VA. L. REV. 1717 (1997); Cass R. Sunstein, David Schkade & Lisa Michelle Ellman, *Ideological Voting on Federal Courts of Appeals: A Preliminary Investigation*, 90 VA. L. REV. 301 (2004); Donald R. Songer & Sue Davis, *The Impact of Party and Region on Voting Decisions in the United States Courts of Appeals, 1955-1986*, 43 W. POL. Q. 317 (1990). For scholars studying dissenting behavior in the circuit courts, see Stephen J. Choi & G. Mitu Gulati, *Mr. Justice Posner? Unpacking the Statistics*, 61 N.Y.U. ANN. SURV. AM. L. 19 (2005); Jeffrey A. Lefstin, *The Measure of the Doubt: Dissent, Indeterminacy, and Interpretation at the Federal Circuit*, 58 HASTINGS L.J. 1025 (2007); Sunstein et al., *supra*. For scholars studying the time it takes for decisions, see Stefanie A. Lindquist, *Bureaucratization and Balkanization: The Origins and Effects of Decision-Making Norms in the Federal Appellate Courts*, 41 U. RICH. L. REV. 659 (2007). For scholars using the impact of outside circuit citations as a measure of opinion quality, see Stephen J. Choi & G. Mitu Gulati, *Choosing the Next Supreme Court Justice: An Empirical Ranking of Judge Performance*, 78 S. CAL. L. REV. 23 (2004); William M. Landes et al., *Judicial Influence: A Citation Analysis of Federal Courts of Appeals Judges*, 27 J. LEGAL STUD. 271 (1998).

Professor Cross's suggestion to control for panel effects also resonated.[11] So, I did just that. Panel effects arise when a circuit judge's vote is influenced by the political proclivities of the other judges on the panel deciding a particular case. The new voting pattern regressions are reported in Table 1. The panel effects have the expected sign and significance level. Democratic-appointees voting with two other democratic-appointees were eleven percent more likely to cast a liberal vote. Republican-appointees voting with two other republican-appointees were three percent more likely to case a conservative vote. Inclusion of panel effects did not alter the results on the opportunity cost variable.

---

[11] Professor Cross was the first legal scholar to consider panel effects. *See* Frank B. Cross & Emerson H. Tiller, *Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeals*, 107 YALE L.J. 2155 (1998). Political scientists had looked at such effects earlier. *See, e.g.*, Burton M. Atkins, *Judicial Behavior and Tendencies Toward Conformity in a Three-Person Small Group: A Case Study of Dissent Behavior on the U.S. Court of Appeals*, 54 SOCIAL SCI. Q. 41 (1973). The panel effects literature has now blossomed. *See generally* Sean Farhang & Gregory Wawro, *Institutional Dynamics on the U.S. Court of Appeals: Minority Representation Under Panel Decision Making*, 20 J.L. ECON. & ORG. 299 (2004); Pauline T. Kim, *Deliberation and Strategy on the United States Courts of Appeals: An Empirical Exploration of Panel Effects* (unpublished manuscript *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1115357); Thomas J. Miles & Cass R. Sunstein, *The Real World of Arbitrariness Review*, 75 U. CHI. L. REV. (forthcoming 2008); Sunstein et al., *supra* note 10.

Table 1
Relationship Between Financial Sacrifice and Voting Patterns Controlling for
Panel Effects
Probit Model

Regressors	Model(1) dem. judges		Model (2) dem. judges networth sample		Model (3) rep. judges		Model (4) rep. judges networth sample	
NETCOST	0.01	(0.89)	0.01	(0.65)	0.007	(0.77)	0.006	(0.49)
selpref	0.107	(1.10)	0.253	(1.71)	0.014	(0.28)	-0.079	(0.82)
Age	0.002	(0.92)	0	(0.11)	0.003	(1.14)	0.004	(1.05)
Sex	-0.007	(0.27)	0.009	(0.24)	0.025	(0.78)	0.079	(1.91)
Top Five	-0.07	(1.11)	-0.21	(1.74)	0.088	(1.35)	0.072	(0.75)
PrivatePractice	-0.04	(0.81)	-0.132	(1.53)	0.004	(0.11)	-0.017	(0.33)
Professor	-0.008	(0.14)	-0.081	(0.75)	0.006	(0.15)	0.052	(0.60)
Judge	-0.044	(0.88)	-0.119	(1.47)	0.024	(0.67)	-0.053	(0.95)
TOPFIVE NETCOST	0.025	(0.97)	0.121	(1.50)	-0.024	(1.37)	-0.04	(1.51)
demjudge/ dempanel	0.117	(3.92)**	0.145	(3.55)**	N/A		N/A	
demjudge/ repubpanel	-0.01	(0.41)	-0.009	(0.27)	N/A		N/A	
repjudge/ dempanel	N/A		N/A		0.027	(1.09)	0.023	(0.68)
repjudge/ reppanel	N/A		N/A		-0.037	(2.13)*	-0.056	(2.26)*
NETWORTH	N/A		0.001	(0.53)	N/A		-0.002	(0.49)
NETCOST (topfive)	0.03	(1.37)	0.13	(1.62)	-0.16	(0.89)	-0.03	(1.26)
circuit dummies	Yes		Yes		Yes		Yes	
Observations	2338		1166		3934		1957	
Pseudo R-squared	0.04		0.05		0.02		0.02	

Robust z statistics in parentheses * significant at 5%; ** significant at 1%
Estimated coefficients reflect marginal effects when all independent variables are measured at their mean. The base category for the panel effects is a judge voting with a split panel: one democratic-appointee, one republican-appointee. My dataset did not include judges appointed before 1974, after 2004, and district court judges sitting by designation. Since I constructed panel effects for those cases where three judges in my dataset participated in the decision,the number of observations differs from those reported in the original article. In light of CHZ's reply, I also report NETCOST (Topfive) as the estimate for judges from top-five markets.

Finally, Professor Cross correctly points out that researchers rarely rely on statistically insignificant results. The lack of significance could mean a bunch of things. It could be the result of mis-measured data, not enough data, too much correlation between the independent variables, or it could mean no

association between the variables of interest.[12] A small number of studies do, however, rely on and report statistically insignificant results.[13] And when they do, even with all the limitations noted above, it is because our intuition, economic theory, or the previous literature tells us that there should be a correlation.

The link between judicial salaries and judicial performance fits that bill. The reason is the nature of the claims advanced by the advocates of higher judicial pay, especially the Chief Justice. Conceding all the problems identified by Professor Cross, my data and analysis tell another side to the "constitutional crisis" story bandied about in the public domain and before Congress.[14]

---

[12] *See* WOOLDRIDGE, *supra* note 7, at 135 (explaining the consequence of small sample sizes); *see also* KENNEDY, *supra* note 8, at 179-99 (explaining the consequences of multicollinearity); *id.* at 137 (explaining the consequences of mismeasured data).

[13] Such studies appear, on rare occasion, in the leading peer-reviewed economics journals. *See, e.g.*, Koleman Strumpf & Felix Oberholzer, *The Effect of File Sharing on Record Sales: An Empirical Analysis*, 115 J. POL. ECON. 1, 1 (2007) (finding that downloads had "an effect on [music] sales which is statistically indistinguishable from zero"). On rare occasions, they appear in the leading peer-reviewed sociology journals. *See, e.g.*, Alexandra Kalev et al., *Best Practices or Best Guesses? Diversity Management and the Remediation of Inequality*, 71 AM. SOC. REV. 589, 610 (based on statistically insignificant results, concluding that some popular diversity programs don't help women or African-Americans reach management positions). Occasionally, they appear in leading peer-reviewed law and economics journals. *See, e.g.*, Orley Ashenfelter et al., *Politics and the Judiciary: The Influence of Judicial Background on Case Outcomes*, 24 J. LEGAL STUD. 257, 281 (1995) (stating that "we cannot find that Republican judges differ from Democratic judges in their treatment of civil rights cases"). And they sometimes appear in the leading law reviews. *See, e.g.*, Thomas J. Miles & Cass R. Sunstein, *Do Judges Make Regulatory Policy? An Empirical Investigation of* Chevron, 73 U. CHI. L. REV. 823, 858-59 (2006) (finding that "[f]or politically mixed panels, the [agency] validation rates of Democratic and Republican judges are very similar to each other; all but one of the differences are 10 percentage points or less and are statistically insignificant" and, concluding from this, "the influence of panel composition on judicial decisionmaking appears largely cabined to politically unified panels").

[14] *See* Chief Justice John G. Roberts, *2006 Year-End Report on the Federal Judiciary*, 39 THE THIRD BRANCH: NEWSLETTER OF THE FEDERAL COURTS (Admin. Office of the U.S. Courts, Wash. D.C.), Jan. 2007, at 1, *available at* http://www.uscourts.gov/ttb/ jan06ttb/yearend/index.html; *see also Fed. Judicial Compensation: Oversight Hearing Before the Subcomm. on the Courts, the Internet, and Intellectual Property of the H. Comm. on the Judiciary*, 110th Cong. 4 (2007) (statement of Justice Samuel Alito) ("Without serious salary reform, the country faces a very real threat to its judiciary."); *Fed. Judicial Compensation: Oversight Hearing Before the Subcomm. on the Courts, the Internet, and Intellectual Property of the H. Comm. on the Judiciary*, 110th Cong. 1 (2007) (statement of Justice Stephen Breyer) ("I believe that something has gone seriously wrong with the judicial compensation system."); *Judicial Security and Independence: Hearing Before the S. Comm. on the Judiciary*, 110th Cong. 7 (2007) (statement of Justice Anthony M. Kennedy) ("The current [judicial salary] situation . . . is a matter of grave systemic concern."); Chief

To see this, rather than consider standard statistical significance, slice the data another way. Look at the confidence intervals reported for NETCOST and each judicial performance measure. Table 2 reports these results.[15]

Table 2

Confidence Intervals for Impact of $400,000 Salary Increase on NETCOST

Performance Models	Confidence Interval
Voting – Democratic Appointees*	
Model 1 (Full Sample)	[-.01, .03]
Model 2 (Subsample w/NETWORTH)	[-.01, .03]
Voting – Republican Appointees*	
Model 1 (Full Sample)	[-.01, .02]
Model 2 (Subsample w/NETWORTH)	[-.01, .03]
Citation Bias Analysis	[-.01,.009]
Dissents Analysis	
Model 1 (Full Sample)	[-.01, -.002]
Model 2 (Sample w/NETWORTH)	[-.01, -.007]
Speed of Disposition	
Model 1 (Full Sample)	[-5.2, 6.6]
Model 2 (Sample w/NETWORTH)	[-1.4, 14.8]
Extra-Circuit Citations: Total Influence	[-.03, .13]
Extra-Circuit Citations: Avg. Influence	[-.004, .08]

 * Voting pattern regressions include panel effects.

Justice William H. Rehnquist, *2002 Year-End Report on the Federal Judiciary*, 35 THE THIRD BRANCH: NEWSLETTER OF THE FEDERAL COURTS (Admin. Office of the U.S. Courts, Wash. D.C.), Jan. 2003, at 2 ("[T]he need to increase judicial salaries . . . remains the most pressing issue [facing the judiciary].").

[15] Confidence intervals for the regressions considering strength of the nominee pool can be found here: http://www.law.unc.edu/faculty/directory/details.aspx?cid=3.

All the confidence intervals involve two-tailed tests. Using the expected sign from the theory, I also conducted a one-tailed test to find the threshold value the data rejects. This test yielded similar results and is not reported here.

These intervals mean that I can reject, at a 95-percent confidence level, any null hypothesis outside the interval.[16] Now let the Chief Justice set the null: Low pay is creating a constitutional crisis. What counts as a crisis is tough to quantify. Any number would be contestable, so I won't even try. Suppose that a constitutional crisis means increasing the chance that democratic appointees cast a liberal vote by more than two percent. I can reject that "crisis null" at 95-percent confidence. Suppose a constitutional crisis means increasing the chance that republican appointees will cast a conservative vote by more than one percent. I can reject that null at 95-percent confidence. Suppose a constitutional crisis means that the expected days between oral argument and a final decision decrease by more than 5 days. I can reject that null at 95 percent. And so on. In short, even with the imprecise judicial performance measures, the limited proxy for financial sacrifice, and the multicollinearity, the confidence intervals for most performance measures are tight around zero.[17] This means that, for almost all my measures, the data rejects a large effect from a salary change.

Yet this analysis leaves an issue open: What is a large effect? Maybe improving the total number of outside circuit citations for each opinion by more than 12 percent or reducing partisan voting by more than two percent are worth the cost of the judicial pay raise. Who knows? This is ultimately a political, not a statistical question, which requires some estimate of the social return from having a "better" judiciary as measured along these lines.

II. ARE JUDGES FROM TOP-FIVE MARKETS DIFFERENT?

In their reply, CHZ point out that a judicial pay raise is likely to impact judges from the top-five markets differently than judges in other markets. All the regressions in my study included a term interacting NETCOST with whether the judge came from a top-five market. This interaction alleviated some of the measurement error created by using regional partnership data as a judge's opportunity cost.

Private practitioners in top-five markets make more than the average partner in their respective region. The use of regional partnership data thus likely underestimated the opportunity cost for judges in the mega-markets. The interaction term mitigated this concern because it allows a one-unit increase in NETCOST to have different and presumably greater effect on judges in top-five markets. My article, however, reports the estimate on NETCOST as the overall effect for all judges, not distinguishing between top-five markets and other markets. CHZ correctly point out that the impact of a change in

---

[16] *See* FUMIO HAYSASHI, ECONOMETRICS 38 (2000).

[17] For judges from top-five markets, the results are different when it comes to voting patterns for republican-appointed judges and the speed of disposition. For all the other regressions, the results reported in Table 2 are a good estimate of the effect of a salary change on the behavior of all judges. For a fuller discussion of why this is so, see *infra* Part II.

NETCOST might differ between judges from top five markets and judges from other markets (which is why I used the interaction term in the first place). CHZ show how those differences play out. Further, CHZ use new data on the lateral market for government attorneys moving to law firms in top-five markets to show exactly how much I might have underestimated the opportunity cost for judges in these markets.

For three of the eleven judicial performance models, CHZ find a change in NETCOST has significant effects on the performance of judges from top-five markets.[18] These results stand in contrast to the insignificant effect for judges from other markets. In addition, CHZ do not find a significant effect on dissent patterns for these judges – a result in contrast to the statistically significant and negative dissent results for judges in other markets from my original article. Interestingly, CHZ interpret their findings as evidence that judges in top-five markets are more willing to trade off salary for voting power and influence, whereas judges in other markets are more willing to trade off salary for leisure.

To see more clearly what is going on with the interaction term, Table 4 reports NETCOST, the coefficient estimate for judges in non-top-five markets, and TOPFIVENETCOST, the estimate on the interaction term.

---

[18] Those regressions were: (1) democratic-appointee voting patterns (subsample with netowrth data); (2) republican-appointee voting pattern (full sample); (3) extra-circuit citations: average influence and (4) extra-circuit citations: total influence. Christopher Zorn, William D. Henderson, & Jason J. Czarnezki, *supra* note 2, at 834.

Table 3

Interaction Between NETCOST and Top-Five Markets

Performance Models	NETCOST		TOPFIVENETCOST	
Voting - Democratic Appointees (Probit)				
Model 1 (Full Sample)	0.001	(0.15)	0.02	(0.97)
Model 2 (Subsample w/NETWORTH)	0.004	(0.34)	0.12	(1.70)
Voting - Republican Appointees (Probit)				
Model 1 (Full Sample)	0.003	(0.47)	-0.031	(2.08)**
Model 2 (Subsample w/NETWORTH)	0.01	(0.98)	-0.04	(2.17)**
Citation Bias Analysis (OLS)	-0.001	(0.14)	-0.01	(1.39)
Dissents Analysis (Probit)				
Model 1 (Full Sample)	-0.006	(3.29)**	0.005	(1.42)
Model 2 (Subsample w/NETWORTH)	-0.01	(4.13)**	0.009	(1.81)
Speed of Disposition (OLS)				
Model 1 (Full Sample)	0.699	(0.23)	-12.8	(2.42)**
Model 2 (Subsample w/NETWORTH)	6.67	(1.61)	-19.42	(2.25)**
Extra-Circuit Citations: Total Influence (OLS)	0.05	(1.25)	0.1	(1.62)
Extra-Circuit Citations: Avg. Influence (OLS)	0.039	(1.77)	0.025	(0.72)

In four of the eleven regressions, the interaction term is statistically significant.[19] For these regressions, CHZ are right. Their results should be taken as an important qualification to the results reported in the original article. For the remaining seven regressions, the interaction term is insignificant. It is these regressions I want to focus on now.

Insignificance of the interaction term means that I can't reject the hypothesis that judges in top-five markets react the same to changes in opportunity cost as judges in other markets. Yet, in these regressions, CHZ find different effects depending on whether the judge comes from a major market. If we can't reject the hypothesis that top-five market judges respond similarly to changes in NETCOST as do judges in other markets, why do CHZ find that the effect depends on the judge's home market in these regressions? More importantly, which effect – the one for judges from a top five market or the one for judges from the other markets – best represents the "true" effect of a change in NETCOST on judicial performance for *all* judges.[20]

This puzzle and an ambiguity in interpreting the effect of changes in NETCOST on judicial performance can be seen more clearly with a little math.

Adopting CHZ's notation, my typical regression took the following form:

(1)

$$f^{-1}(Performance_i) =$$
$$\beta_0 + \beta_1 NETCOST_i + \beta_2 TOPPFIVE_i +$$
$$\beta_3(TOPFIVE_i \times NETCOST_i) + \mathbf{X}_i \gamma$$

As CHZ make clear, in this regression β_1 represents the effect of a change in NETCOST for judges outside the top-five markets; $\beta_1 + \beta_3$ represents the effect

---

[19] I use a two-tailed significance test here. CHZ use a one-tailed significance test in replicating the results. Under a one-tailed test, three of the eleven regressions have a significant interaction term. Zorn, Henderson & Czarnezki, *supra* note 2, at 834. Under a one-tailed test, the interaction term is significant for (1) democratic-appointee voting patterns in the networth sub-sample; (2) republican-appointee voting patterns in the full sample and (3) republican-appointee voting patterns in the networth subsample. Unlike the two-tailed test, the interaction term is insignificant for both regressions involving speed of disposition. The reason is that the coefficient doesn't have the expected sign in those regressions. The choice between a one-tailed and two-tailed test reflects how confident a researcher is that his theory gets the sign of the effect right. *See* WOOLDRIDGE, *supra* note 7, at 121-22.

[20] To avoid this ambiguity, one solution would be to drop the interaction term in all the models where it was insignificant and rerun the regressions. Then, I might have reported the NETCOST coefficient from the new regression as the overall effect. Such a move is undesirable, however, because it leads to pre-test bias of the estimates. *See* KENNEDY, *supra* note 8, at 189-91.

of a change in NETCOST on judges in top-five markets; X represents the set of controls.

I could have run the following regression instead.

(2)

$$
f^{-1}(Performance_i) =
$$
$$
\alpha_0 + \alpha_1 NETCOST_i + \alpha_2(1 - TOPPFIVE_i) +
$$
$$
\alpha_3((1 - TOPFIVE_i) \times NETCOST_i)) + \mathbf{X}_i\gamma
$$

With (2), α_1 represents the effect of a change in NETCOST for judges in top-five markets; $\alpha_1 + \alpha_3$ represents the effect of a change in NETCOST for judges in non-top-five markets; X, again, is a set of controls.

The difference between (1) and (2) is the group subject to the interaction term. In (1), NETCOST is interacted with judges from the top-five markets. In (2), NETCOST is interacted with judges from non-top five markets. Moving from (1) to (2) flips the assumption. Rather than assume regional partnership salaries under-reports the opportunity cost for judges in top-five markets, equation (2) assumes that regional partnership salaries over-reports the opportunity cost for judges outside the top-five markets. The unmeasured salary difference between the two groups remains the same. So, the assumption change, while unnatural, should be irrelevant.

The coefficient α_1 in equation (2) is the effect reported by CHZ. My article reports, β_1, the coefficient estimate from equation (1). A little algebra shows that $\alpha_1 + \alpha_3 = \beta_1$ and $\beta_3 = - \alpha_3$ no matter the size of the coefficients. For seven of the regressions, however, I can't reject that the interaction term has no effect (i.e., that $\beta_3 = - \alpha_3 = 0$). As a result, I can't reject that α_1 equals β_1. But looking at the estimates, it is clear that the coefficients aren't, in fact, equal. CHZ report different estimates than reported in the original article. In seven of those regressions, however, we can't reject that any reported differences are simply noise.

A deeper question lurks behind the results. What is the effect of a one-unit increase in NETCOST for "all" judges where the interaction term is insignificant? The answer is this: Both α_1 and β_1 are plausible candidates. Either one works and it is probably safest to report both estimates. In defense of the estimate provided in the original article as the true overall effect, that estimate has (a) the smaller standard error (it is more "accurate") and (b) the sample contains many more judges in non-top five markets, making them the more natural baseline group.

Still, CHZ advance the analysis by providing both sets of results side by side. For the regressions where the interaction term is insignificant, what happens if we accept CHZ's bigger estimate as the "true" effect of higher salaries for all judges? Not much. The economic significance of any effect is small and, in fact, the increase might even be harmful. Under CHZ's estimate,

for example, increasing salaries by $50,000 a year decreases opinion quality as measured by average outside circuit citations by six percent.

In four of the regressions, the evidence suggests that judges from top-five markets are different; they respond differently to changes in salary. CHZ show how this difference manifests itself. Most dramatically, they identify that higher salaries could diminish partisan voting among judges in top-five markets. This result is a welcome refinement to the article.

Even with this refinement, I submit, the bottom line remains the same. For judges in most places, the data allow me to exclude that a salary increase will have a large impact on the performance measured studied. Interestingly, while they don't support across the board salary increases, CHZ's results might be used to support more aggressive COLA adjustments for judges in major markets – a proposal Judge Richard Posner has been advocating for a number of years.[21]

III.   HOW DO YOU MEASURE LOST OPPORTUNITY?

In his reply, Professor Marks raises two concerns involving the appropriate measure of a judge's lost opportunity. First, he suggests the NETCOST measure is inadequate because it does not allow for the possibility that a judge in a region with low partner salaries could be giving up a position in a higher paying region when she takes the bench. Second, Marks demonstrates how measuring NETCOST in terms of judges' cumulative lost lifetime earnings may affect the results. I consider each criticism in turn.

A.   *Problems with the Mobility Assumption*

Professor Marks questions the assumption that judges won't leave their region for a higher paying law firm job elsewhere. In his well-crafted example, Professor Marks demonstrates how this simple assumption can alter the results. The judge who viewed her next best opportunity as a partnership at a law firm in the highest paid city in the country would have a higher net cost than a judge who viewed her next best financial opportunity as partnership in a law firm in her local city. Of the 259 judges in the sample, 239 hadn't moved in the ten years prior to their appointment to the bench. For these judges, it seems reasonable to suspect a hometown attachment made them unlikely to move outside the region for a law firm job.

But what about the 19 other judges? Professor Marks shows how making the wrong assumption about the mobility of these judges weakens the results. The assumption means that I consistently underestimate the opportunity cost for these judges. On this point, Professor Marks is right. In light of this critique, I investigated whether grouping the mobile judges and immobile judges together changed the analysis. To do this, I analyzed two new variables. The first variable is a dummy variable, MOBILE, for whether the

---

[21] *See* RICHARD A. POSNER, HOW JUDGES THINK 172-73 (2008).

judge moved in the ten previous years before taking the bench. The second variable is an interaction term between MOBILE and NETCOST. Similar to the interaction term between the dummy variable, TOPFIVE, and NETCOST, this term allows for a one-unit increase in NETCOST to have a greater effect on mobile judges.

Table 4 reports the results on the variables of interest. The results remain the same, except for speed of disposition and dissents. For mobile judges in markets outside the top-five, giving up lots of cash does not have a significant effect on dissenting behavior. This is in contrast to immobile judges from these markets, for whom NETCOST has a significant and negative effect. With regard to speed of disposition, the coefficient for mobile judges from non-top five markets is significant and positive. While small in magnitude (15 days), this result suggests Congress could reduce decision time for the mobile judges by increasing their salaries.

Table 4

Performance Models Controlling For Potential Mobility By Judges

Performance Models	NETCOST mobile judge non-top-five market	NETCOST immobile judge non-top-five market	NETCOST mobile judge top-five market	NETCOST immobile judge top-five market
Voting – Democratic Appointees (Probit)				
Model 1 (Full Sample)	-.06 (.78)	.01 (.94)	-.03 (.44)	.03 (1.44)
Model 2 (Subsample w/NETWORTH)	-.18 (.76)	.007 (.45)	-.06 (.24)	.13 (.1.63)
Voting – Republican Appointees (Probit)				
Model 1 (Full Sample)	-.02 (.78)	.005 (.57)	-.03 (.99)	-.002 (.15)
Model 2 (Subsample w/NETWORTH)	.06 (.45)	.003 (.29)	.04 (.28)	-.02 (.74)
Citation Bias Analysis (OLS)	.02 (1.84)	-.001 (.36)	.016 (1.15)	-.01 (1.41)
Dissents Analysis (Probit)				
Model 1 (Full Sample)	-.0009 (.15)	-.007 (3.34)**	.002 (.34)	-.003 (.97)
Model 2 (Sample w/NETWORTH)	-.01 (1.65)	-.01 (3.75)**	-.02 (1.65)	-.009 (1.63)
Speed of Disposition (OLS)				
Model 1 (Full Sample)	12.5 (1.32)	.38 (.13)	-3.12 (.33)	-15.3 (2.66)**
Model 2 (Sample w/NETWORTH)	32.42 (2.22)**	6.42 (1.54)	10.96 (.84)	-15 (10.43)
Extra-Circuit Citations: Total Influence (OLS)	-02 (.21)	.05 (1.31)	.08 (.58)	.15 (2.61)**
Extra-Circuit Citations: Avg. Influence (OLS)	.03 (.56)	.04 (1.86)	.05 (.77)	.05 (1.68)

B. *Problems with Cumulating Earnings*

Professor Marks's second concern involves my use of lost lifetime earnings to measure a judge's opportunity cost. Two examples illustrate his point.

Professor Marks's first example shows how, by looking at the lifetime stream of lost earnings, two judges that were, in fact, identical might appear different in the data. His second example demonstrates how a stream of earnings calculation might treat a judge with a weak preference for leisure as if she had a strong preference for leisure.

The first example presents a difficulty. The reason: As evidence against the theory that judicial salary matters, I take the failure to reject the hypothesis that two judges – who the data report as different, but Professor Marks shows really aren't – act the same. The second example poses a problem because the analysis relies on NETCOST being a valid proxy for the judge's taste for the judicial role, i.e., her valuation of the non-pecuniary aspects of judging. In short, Professor Marks suggests that cumulating earnings over time creates meaningless variation in the NETCOST variable. As a result, we can't be sure what is explaining the variation in the dependent judicial performance variables: the true variation in the NETCOST or the meaningless variation introduced through cumulating and then discounting net losses back to present value.

Controlling for a judge's age at appointment should mitigate some of the problem Professor Marks identifies. In both examples, meaningless variation arises because one judge serves two terms (forfeiting two years of partner income), while the other judge serves one term (forfeiting one year of partner income). The only difference between the two judges is that one judge serves longer than the other. Under the assumption that both judges serve until age sixty-five, the regression will not treat these two judges the same. The judge who took the bench at age forty-four will not be treated the same as the judge who took the bench at age forty-five. Instead the regressions, in effect, compare two judges appointed at age forty-four with different levels of opportunity cost.[22]

Even controlling for age at the time of appointment, a related concern still lingers. Take two judges appointed at age forty-five. Suppose the two judges have different opportunity costs as I measured them. The judge with the greater opportunity cost is assumed to have the more intense preference for the non-money aspects of the judicial role. NETCOST assumes each judge will serve on the bench until age sixty-five – in this example, the model would treat both judges as if they expected twenty years of judicial service. Yet, the years of expected judicial service might not be the same for the two judges. A judge with an intense preference for, say, imposing policy preferences might intend to serve longer than a judge with a weak preference for dictating policy. Despite the intense preference, this judge might have a lower NETCOST. That is to say, this judge might give up relatively little money over the twenty-year time-span, but anticipates a much longer judicial career. The same problem arises for a judge with, say, health problems. A judge appointed at age forty-five with a history of heart disease might not anticipate serving until age sixty-

---

[22] *See* WOOLDRIDGE, *supra* note 7, at 200 (providing this interpretation of a control).

five. By assuming a twenty-year judicial career, NETCOST over-estimates the intensity of this judge's preference for the judicial role.

These issues seem insurmountable. We don't have data on the likely career path for each individual judge; their health problems, if any, at the time of appointment; the likelihood they will retire at age sixty-five, remain active, or take senior status; or, if they take senior status, how long they will serve in that capacity.

Because of the difficulties in cumulating earnings over time, Professor Marks suggests a more fruitful measure of opportunity cost would examine a judge's lost earning over a single year.[23] While solving some of the problems noted above, the single period approach discards relevant data. Consider two judges, *A* and *B*. Both are appointed at the same age and forgo $50,000 in their first year on the bench. Judge *A* works in a region where law firm partnership salaries increase, on average, 25 percent a year. Judge *B* works in a region where partnership salaries increase, on average, 10 percent a year. Measuring pay as lost earnings in a single period treats these two judges as making the same financial sacrifice. Yet the truth is Judge *A* gave up more cash for the bench.

To sum up, Professor Marks is correct that cumulated earnings are an imperfect proxy for a judge's opportunity cost; yet single period earnings are also imperfect. What to do? Given these imperfections, I also considered whether the strength of the pool against which a judge competed for the nomination impacted her judicial performance. The thinking here was that higher relative judicial salaries made for a stronger pool. This alternative approach yielded similar results and should mitigate any concern over cumulating earnings for the NETCOST measure.

CONCLUSION

Let me emphasize in concluding that the study – qualified by these replies – doesn't "prove" that Congress should leave judicial salaries where they stand. It doesn't "prove" the performance measures considered reflect judicial quality. It doesn't even "prove" that higher pay wouldn't affect these measures. The basic point is that we shouldn't assume – as Chief Justice Roberts does – that pay will improve judicial performance. The article searches for a statistical significant correlation between some judicial performance measures and a crude proxy for the financial sacrifice of the judges. For most measures and most judges, it finds none. To be precise, the data rejects large effects of judicial pay on performance and fails to reject tiny or negligible effects of pay on performance, meaning that a change in salary is

---

[23] Stephen Choi, Mitu Gulati, and Eric Posner take this approach when studying the impact of pay on state court justice behavior. *See* Choi et al., *supra* note 3, at 45. Unlike the vast majority of federal judges, many state judges leave judgeships before qualifying for retirement. Hence, measuring opportunity cost as a single period loss makes more sense in the context of state court justices.

unlikely to have a meaningful (i.e., large) effect on judicial performance for most judges in most places.

   As for Professor Cross's suggested study of law professor pay, I won't do that study right now. But who knows – maybe I could be motivated to do it by a little raise.