AN EMPIRICAL ANALYSIS OF LIFE TENURE: A RESPONSE TO PROFESSORS CALABRESI & LINDGREN

DAVID R. STRAS[*] & RYAN W. SCOTT[**]

INTRODUCTION

Opposition to life tenure has been steadily mounting in the legal academy, and Professors Steven Calabresi and James Lindgren are among those leading the charge. In an article published in the *Harvard Journal of Law & Public Policy*,[1] Calabresi

[*] Associate Professor, University of Minnesota Law School.

[**] Bristow Fellow, Office of the Solicitor General of the United States. The views expressed herein are solely those of the Authors. We are grateful to Allan Erbsen, Arthur Hellman, Toby Heytens, Kristin Hickman, Tim Johnson, and Heidi Kitrosser for their comments and suggestions on an earlier draft. Special thanks go to Tade Okediji, who read several uncirculated drafts and assisted us in formative discussions about a number of the issues raised in this Article. This Article also benefited from the excellent research assistance provided by Eric Tranby and Shaun Pettigrew.

[1] Steven G. Calabresi & James Lindgren, *Term Limits for the Supreme Court: Life Tenure Reconsidered*, 29 HARV. J.L. & PUB. POL'Y 769 (2006). The article first appeared in a collection called *Reforming the Court: Term Limits for Supreme Court Justices*, edited by Roger C. Cramton and Paul D. Carrington and published in the

and Lindgren denounce life tenure as "fundamentally flawed"[2] and "essentially a relic of pre-democratic times."[3] They deserve credit for assembling the most comprehensive critique of life tenure to date, carefully documenting what they see as its major drawbacks, and proposing a constitutional amendment providing for fixed, non-renewable eighteen-year terms for Supreme Court Justices.[4] Their article has attracted well-deserved attention among legal scholars and in the popular press. To its credit, it has found champions of all political persuasions.[5]

Calabresi and Lindgren direct three criticisms at life tenure. First, because it produces infrequent vacancies, life tenure affords the President and Senate too few opportunities to act as a democratic check on the Court through the appointment of new members.[6] Second, again because of infrequent vacancies, life tenure raises the stakes of each appointment to undesirable levels, thereby exacerbating the politicization of the appointment process.[7] Third, life tenure allows Justices to serve well into old age, increasing the risk of "mental decrepitude" on the Court.[8]

Crucial to these policy criticisms is an underlying empirical claim that "the real-world, practical meaning of life tenure" has changed dramatically since 1970.[9] Calabresi and Lindgren point to several trends as evidence of this transformation. The most frequently mentioned is the average length of tenure on the Court: Justices who left office between 1789 and 1970 served an average of 14.9 years on the bench, whereas Justices who left

spring of 2006. The revised version, published in the summer of 2006, contains updated data and additional material responding to critics. *Id.* at 769 n.*.

[2] *Id.* at 771.

[3] *Id.* at 772. Elsewhere Calabresi and Lindgren have used even more provocative language, arguing that life tenure for Supreme Court Justices has produced "a gerontocracy—like the leadership cadre of the Chinese Communist Party." Steven G. Calabresi & James Lindgren, *Supreme Gerontocracy*, WALL ST. J., Apr. 8, 2005, at A12.

[4] *See* Calabresi & Lindgren, *supra* note 1, at 809–18, 824–54.

[5] *See* Linda Greenhouse, *How Long Is Too Long for the Court's Justices?*, N.Y. TIMES, Jan. 16, 2005, at WK5; Tony Mauro, *Profs Pitch Plan for Limits on Supreme Court Service*, LEGAL TIMES., Jan. 3, 2005, at 1; Bruce Bartlett, *. . . And Tenure Traps* (July 6, 2006), http://www.ncpa.org/prs/cd/2005/20050706.htm.

[6] Calabresi & Lindgren, *supra* note 1, at 809–13.

[7] *Id.* at 813–15.

[8] *Id.* at 815–18.

[9] *Id.* at 777–78.

office between 1970 and 2006 averaged 26.1 years—an increase that Calabresi and Lindgren describe as "astonishing."[10] According to the authors, the trend marks a fundamental change in the operation of life tenure.

In the winter of 2005, we published an article taking a contrary view, defending life tenure as an institution worth preserving but proposing a package of retirement incentives—a "golden parachute" for Supreme Court Justices—to encourage mentally infirm Justices to retire in a timely manner.[11] We agreed with Calabresi and Lindgren on a number of issues, echoing their concerns about mental and physical infirmity on the Supreme Court,[12] and agreeing that statutory efforts to abolish life tenure are unconstitutional.[13]

But we also dedicated less than two pages of our seventy-page article to a rebuttal of one aspect of Calabresi and Lindgren's empirical claim, arguing that the "astonishing" increase in average tenure "depends more on the chosen period lengths than a bona fide trend."[14] Our article included two simple charts demonstrating that, by choosing a shorter period length or using non-overlapping groups of Justices rather than periods of time, the data do not reveal a dramatic recent increase.[15] The revised version of Calabresi and Lindgren's article contains fully eleven pages of new material responding to our criticisms.[16]

This Article refines and elaborates upon our critique of Calabresi and Lindgren's empirical claim, arguing that changes in average tenure are not "astonishing" and "unprecedented."

[10] *Id*. at 778–79. To a lesser extent, the authors also rely on the average number of years between vacancies on the Supreme Court, a rough corollary of length of tenure. According to Calabresi and Lindgren, "from 1881 through 1970, the average number of years between commissions stayed consistent at about 1.6 to 1.8," but since 1970 "it has nearly doubled to 3.1 years." *Id*. at 786. The authors also observe that the average age of Justices leaving the Court since 1971 has risen to 78.7 years, an increase of ten years compared with the average of 68.3 years for Justices leaving office before 1971. *Id*. at 782. We have no quarrel with Calabresi and Lindgren's analysis of the data on age at retirement, as we share their concerns about mental and physical infirmity among Supreme Court Justices. This Article therefore focuses on their principal measure of the dramatic change in life tenure: average length of tenure for Supreme Court Justices.

[11] David R. Stras & Ryan W. Scott, *Retaining Life Tenure: The Case for a "Golden Parachute,"* 83 WASH. U. L.Q. 1397 (2005).

[12] *Id*. at 1437.

[13] *Id*. at 1418–20.

[14] *Id*. at 1427.

[15] *See id*. at 1428–29 charts 1 & 2.

[16] *See* Calabresi & Lindgren, *supra* note 1, at 789–99.

Instead, we defend the conventional view that, despite short-term fluctuations, length of tenure has increased slowly and steadily over the long term, and that we can expect more slow and steady growth in the future.

The Article proceeds in two Parts. We acknowledge here, as we have from the outset,[17] that the "practical meaning" of life tenure has changed over time, and it will continue to change as Americans enjoy longer life expectancies. The question, therefore, is not whether there has been *any* change, but whether the change has been so profound and so swift to warrant the abandonment of life tenure. In Part I, we pin down the precise empirical claim advanced by Calabresi and Lindgren: that changes in the practical operation of life tenure since 1970 are both (1) "dramatic" and (2) "unprecedented." We also briefly summarize the course of the debate thus far.

In Part II, we analyze and evaluate the data on life tenure. First, we defend our original contention that period selection affects the shape and apparent magnitude of the trend in length of tenure over time. Second, we demonstrate that making a simple change to Calabresi and Lindgren's assumption that each Justice's tenure should be treated as an observation at the date of departure, rather than date of swearing-in, undermines their model by calling into question the robustness of the trend they identify. Third, we critique the cubic regression model that Calabresi and Lindgren describe as an "elegant confirmation" of their theory of a post-1970 explosion in length of tenure. Instead, applying a different nonlinear regression model, alluded to by Calabresi and Lindgren but not reported in their article, better fits the data and supports our position that changes in term length have been gradual, not "astonishing."

I. PINNING DOWN THE EMPIRICAL CLAIM

Shortly after Professors Calabresi and Lindgren published the first version of their article, we included a short critique of their empirical claim in an article proposing changes to retirement incentives for Supreme Court Justices.[18] Although we explicitly conceded that "average length [of tenure] may in-

[17] *See* Stras & Scott, *supra* note 11, at 1430.

[18] *Id.* at 1427.

crease over time due to advances in medical care,"[19] we argued that Calabresi and Lindgren's "chosen period lengths" had the effect of "exaggerat[ing] the trend."[20] To illustrate the point, we presented two simple charts. The first rendered the data in periods of ten years, rather than thirty years. It showed that average tenure during the 1830s climbed approximately as high as it has in the decades since 1971.[21] The second rendered the data in groups of five Justices by date of appointment, to ensure that each group had an equal number of observations. It revealed only a modest recent increase in average tenure.[22]

In the revised version of their article, Calabresi and Lindgren respond to our period-selection criticism in several ways. Calabresi and Lindgren criticize our two charts, finding both of them lacking for methodological reasons,[23] stating at one point that we "never bothered to check [our] factual claim" about period selection.[24] The authors include new charts showing a "lagging average," representing the average for overlapping groups of nine Justices for each of their metrics for life tenure.[25] Finally, they include the results of a cubic regression model that ostensibly "confirm[s]" their original rendering of the data.[26]

Before responding to the revised version of Calabresi and Lindgren's article, we should clearly identify the point of disagreement. Calabresi and Lindgren's empirical claim about life tenure has two features. First, they contend that changes

[19] *Id.* at 1430.

[20] *Id.* at 1427.

[21] *Id.* at 1427 chart 1.

[22] Stras & Scott, *supra* note 11, at 1428 chart 2.

[23] *See id.* at 791–94. Twice, Calabresi and Lindgren even imply that we either deliberately withheld details about our charts or affirmatively misrepresented the data. *Id.* at 791 ("They do not report the cell counts for their chart dividing the Justices by decade, perhaps for a reason."); *id.* ("Stras and Scott appear to be straining to find ways of presenting the data that lump Justices together in a way that will make the patterns in the data disappear.").

[24] *Id.* at 790.

[25] *See* Calabresi & Lindgren, *supra* note 1, at 781 chart 2, 783 chart 4, 788 chart 6.

[26] Calabresi & Lindgren, *supra* note 1, at 798–99. Professor Lindgren is a Ph.D. candidate at the University of Chicago. *See* James Lindgren, Curriculum Vitae, http://www.law.northwestern.edu/faculty/fulltime/lindgren/lindgrjacv.pdf. As a contributor to the popular law blog The Volokh Conspiracy, Professor Lindgren has said that he has "done tens of thousands of regression analyses." Posting of Jim Lindgren to The Volokh Conspiracy, http://volokh.com/posts/1162276977.shtml#155478 (Oct. 31, 2006, 10:40 EST).

in the actual operation of life tenure since 1970 have been "dramatic." Second, they argue that the changes are historically "unprecedented." Both aspects of their claim deserve fuller explanation.

First, Calabresi and Lindgren argue that the data on life tenure reveal a "dramatic" trend. They variously describe the change as "critical and significant,"[27] "remarkable,"[28] and "astounding."[29] They use the word "astonishing" twice,[30] and some form of the word "dramatic" no less than eight times.[31] The transformation in life tenure, we are assured, is nothing short of mind-boggling.

Of course, these are verbal formulations, not mathematical ones, and reasonable people might disagree about whether a particular increase is "dramatic" or "astounding." Perhaps Calabresi and Lindgren mean that the difference in data since 1971 is statistically significant. At one point, for example, they take pains to demonstrate that the difference between the length of tenure for "the last twelve retirees as a group" and the length of tenure for "the typical Justice leaving the bench through 1970" is "more than large enough to be statistically significant (p=.0002)."[32] By that standard, however, the 13.4-year increase in average tenure during the period from 1821–1850 was equally dramatic (p=.003),[33] and in hindsight, so was the period from 1789–1820 (p=.00016).[34] Moreover, as Calabresi and Lindgren acknowledge, statistical significance shows only that "[t]he data do not appear to be random,"[35] but not necessarily that the trend in the data is "astounding" or "dramatic."

Thus, when Calabresi and Lindgren say that the recent trend is "dramatic," they must mean something other than a statistically significant change. Indeed, they appear to assert that the *rate of increase* in the recent trend is especially pro-

[27] Calabresi & Lindgren, *supra* note 1, at 777.

[28] *Id*. at 786.

[29] *Id*. at 779.

[30] *Id*. at 775, 778.

[31] *Id*. at 778, 779, 780, 782, 789, 798, 807, 832.

[32] Calabresi & Lindgren, *supra* note 1, at 797.

[33] Like Calabresi and Lindgren, we used an independent samples t-test for equality of means, without assuming equality of variance. *See id*. at 797 n.79. Here, we compared data from 1789–1820 with data from 1821–1850.

[34] Here, we compared data from 1789–1820 with data from 1821–2006.

[35] Calabresi & Lindgren, *supra* note 1, at 797.

nounced. Calabresi and Lindgren write, for example, that "one of the main contributions" of their work has been to show that the increase in Supreme Court tenure is "not a gradual, steady climb."[36] Although this aspect of the empirical claim defies exact quantification, the authors plainly have in mind an exponential or accelerating increase, as opposed to a linear or decelerating increase.

Second, Calabresi and Lindgren argue that Supreme Court Justices now remain on the Court "for longer periods . . . than ever before in American history."[37] They use the term "unprecedented" at least five times.[38] This aspect of the hypothesis is more susceptible to testing: if the data reveal a historical precedent for today's average length of tenure, then we can reject it.

Both aspects of the empirical claim are crucial to Calabresi and Lindgren's policy criticisms of life tenure because they maintain that "[f]or 180 years through 1970," life tenure "worked well."[39] Only within the last few decades has the system abruptly "broken down,"[40] prompting their call for a constitutional amendment.[41] That characterization requires proof of an unprecedented change. If life tenure, historically, has produced terms as long as those that prevail today, then it would be strange to say that it "worked well" for 180 years and has since become deeply flawed. At the same time, Calabresi and Lindgren's call for reform also depends on the existence of a dramatic change. Modest, gradual, or decelerating changes in tenure on the Supreme Court hardly mark the arrival of a constitutional crisis.

We disagree with Calabresi and Lindgren's two-part empirical claim. The increase in average tenure on the Supreme Court in the last few decades has not been dramatic and unprecedented. Instead, as we explain in Part II, despite short-term fluctuations, the data are more consistent with a gradual long-term climb.

[36] *Id.*

[37] *Id.* at 771.

[38] *Id.* at 781, 783, 790, 793, 794.

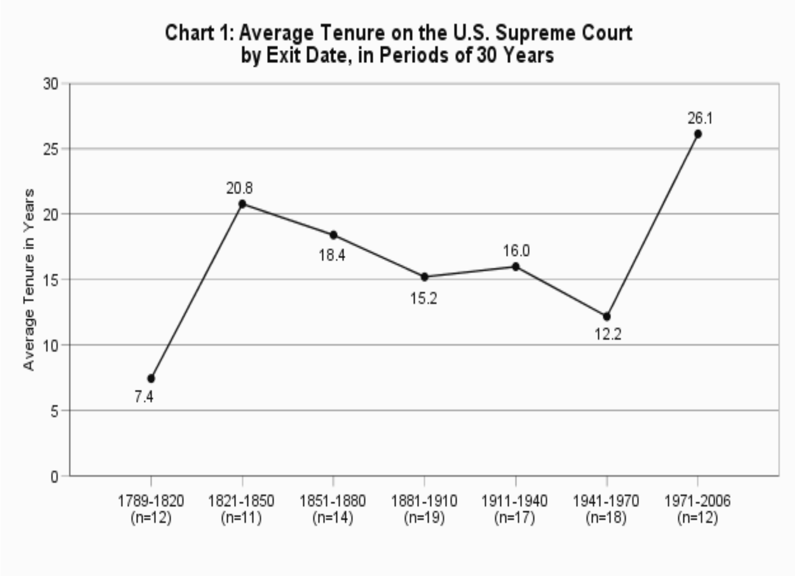
[39] Calabresi & Lindgren, *supra* note 3, at A12.

[40] *Id.*

[41] *See id.*

II. LIFE TENURE BY THE NUMBERS

More than any other aspect of Professors Calabresi and Lindgren's article, one chart has framed the debate about life tenure. It shows average tenure for Supreme Court Justices broken down into periods of approximately thirty years, with the final period covering the thirty-six years since 1970:[42]



Two aspects of the chart are striking. One is the eye-popping number for the last period: 26.1 years, as compared to the average of 14.9 years for the earlier periods combined.[43] A host of commentators have marveled at that figure, noting that it shows that average tenure today is "nearly twice as long" as it used to be.[44]

Another is the sharp jump between the penultimate period and the last period: 12.2 years for the period between 1941 and 1970 followed by 26.1 years for the post-1971 period, "an astonishing fourteen-year increase" from one period to the

[42] Calabresi & Lindgren, *supra* note 1, at 778 chart 1.

[43] *Id.*

[44] Ronald Brownstein, *Time to Bring Down the Gavel on Lifetime Tenure for Justices?*, L.A. TIMES, Oct. 17, 2005, at A10; *see also* Jeff Jacoby, *Don't Let Judges Serve for Life*, BOSTON GLOBE, May 26, 2005, at A19; Greenhouse, *supra* note 5, at WK5.

next.[45] Professor Sandy Levinson marvels at the final-period increase, calling Calabresi and Lindgren's figures "definitive."[46] Bruce Bartlett made this statistic a centerpiece of an article calling for an end to life tenure.[47] Other columnists and the blogosphere have likewise taken the figure and run with it.[48]

The uncritical acceptance of Calabresi and Lindgren's figures is unfortunate because their chart suffers from two design problems. First, the strength of the trend they identify depends on the period lengths they have selected (call this the "period-selection problem"). Second, their results depend entirely on their unexplained and unjustified decision to treat length of tenure as an observation at the date of death or retirement, rather than the date of appointment (call this the "date-of-observation problem"). We explain both problems more fully below.

A. The Period-Selection Problem

In an earlier article, we criticized the period selection in Calabresi and Lindgren's chart, arguing that the authors chose "a period length (thirty years) and a cutoff date (1971) that exaggerates the trend."[49] In the most recent version of

[45] Calabresi & Lindgren, *supra* note 1, at 778.

[46] SANFORD LEVINSON, OUR UNDEMOCRATIC CONSTITUTION 128–29 (2006). Professor Levinson recognizes the period-selection problem, however, by noting that the 1941–1970 period included several Justices "who served unusually short terms because of death or moving to other positions of service." *Id.*

[47] *See* Bartlett, *supra* note 5 (citing the increases between the last two periods in terms of average tenure, average age at retirement, and average interval between appointments).

[48] *See, e.g.*, Mark Alexander, *Supreme Consequences*, PATRIOT POST, No. 05-27, July 8, 2005, http://secure.patriotpost.us/Alexander/edition.asp?id=315; Ken Bell, *Lest Ye Be Judged*, AUSTIN REV., July 6, 2005, *available at* http://www.austinreview.com/archives/2005/07/lest_ye_be_judg.html; Posting of Publius to Publius' Forum, http://pconservablogs.com/publiusforum/2005/10/04/scotus-stats-of-interest/ (Oct. 4, 2005).

Anxious about the role that period selection might have played in generating this particular figure, Calabresi and Lindgren distance themselves from it in the most recent version of their article. "[O]ther than in Chart 1 itself and in the one paragraph discussing Chart 1," they explain, "we made only one mention of the 12.2 year tenure of Justices who left office in the 1941–1970 period." Calabresi & Lindgren, *supra* note 1, at 795. We acknowledge that Calabresi and Lindgren have consistently emphasized the 14.9-year average for Justices retiring before 1971. As the sources cited in the last three footnotes demonstrate, however, scholars and commentators relying on their research have not been so careful.

[49] Stras & Scott, *supra* note 11, at 1427.

their article, they answer that "Stras and Scott went to extraordinary lengths" in order "to suppress . . . the unprecedented length of tenure of retirees from the Court since 1971."[50] A survey of charts with alternative period lengths will illustrate our original point about the crucial role of period selection in their claim of a dramatic and unprecedented increase in length of tenure for Supreme Court Justices.[51]

Instead of periods of thirty years, we begin by rendering the data using periods of fifty years. Calabresi and Lindgren offer a powerful reason to do so: small sample sizes can cause researchers "to see patterns where there are none or to miss patterns that are present."[52] Periods of fifty years, by contrast, ensure that each period contains twenty or more observations. The results appear in Chart 2:

[50] Calabresi & Lindgren, *supra* note 1, at 793.

[51] In calculating length of service for each Justice, we subtracted (1) the date of swearing-in from (2) the date of departure from regular active service on the Supreme Court, whether by death, resignation, or the election of senior status. We used the date of swearing-in, rather than the date of confirmation or commission, because swearing-in marks the start of a Justice's active service on the Court, even for recess appointees. This methodological choice represents a slight change from our method in our last article on this subject, where we used the date of a Justice's commission. *See* Stras & Scott, *supra* note 11, at 1428 n.164. Meanwhile, we used the date of a Justice's death, resignation, or election of senior status to calculate date of departure because any one of those events triggers a new appointment opportunity for the President and marks the end of a Justice's active service on the Court. *See* 28 U.S.C. § 371(d) (2000). To convert the difference from days to years, we divided by 365.25.

Like Calabresi and Lindgren, we excluded current members of the Court from our calculations even though we know a certain *minimum* length of service for each of them. Justice Stevens, for example, has already served for more than thirty years. On the other hand, Chief Justice Roberts and Justice Alito have served for approximately four years combined. Because the current average on the Court falls well below the historical average, adding current members to the data set—as if all of them retired immediately—would distort the data. Likewise, projecting an exit date for current members, perhaps based on life expectancy or historical averages, would both distort the data and defeat the point of the analysis by assuming an answer to the question that we are analyzing. We therefore follow Calabresi and Lindgren in limiting our analysis to Justices who have already departed from the Court.

[52] Calabresi & Lindgren, *supra* note 1, at 792.

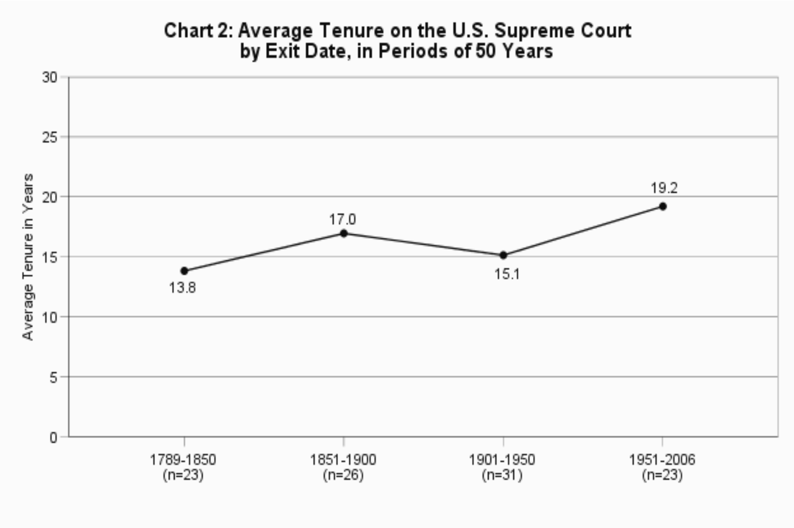


Chart 2: Average Tenure on the U.S. Supreme Court by Exit Date, in Periods of 50 Years

Although the final period reveals a 19.2-year average tenure for the period from 1951–2006, higher than for any other period, it also shows an increase of only 3.8 years over the historical average of 15.4 years from 1789–1950, and only 4.1 years over the previous-period average of 15.1 years from 1901–1950. Nothing about this chart is "dramatic" or "astonishing." Rather, the overall trend appears to be one of slow and steady growth.

Next, we render the data using periods of forty years, closer to the period length selected by Calabresi and Lindgren. Chart 3 displays the results:

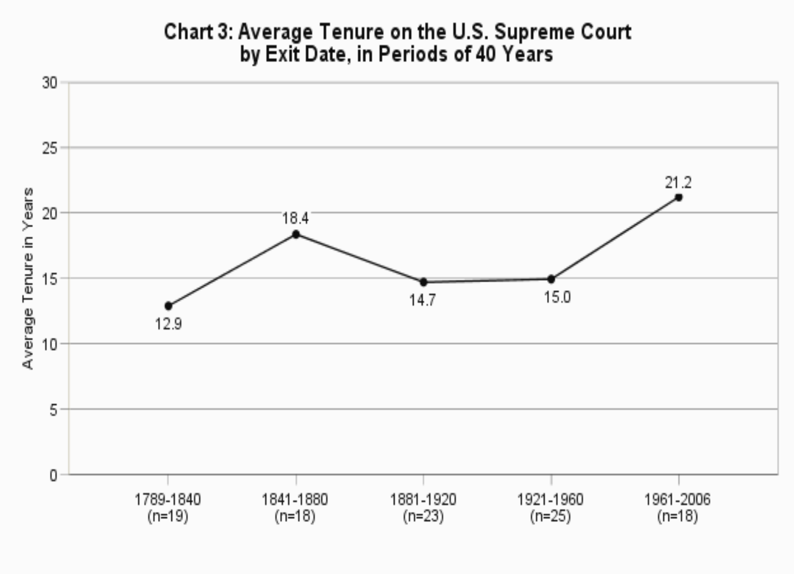
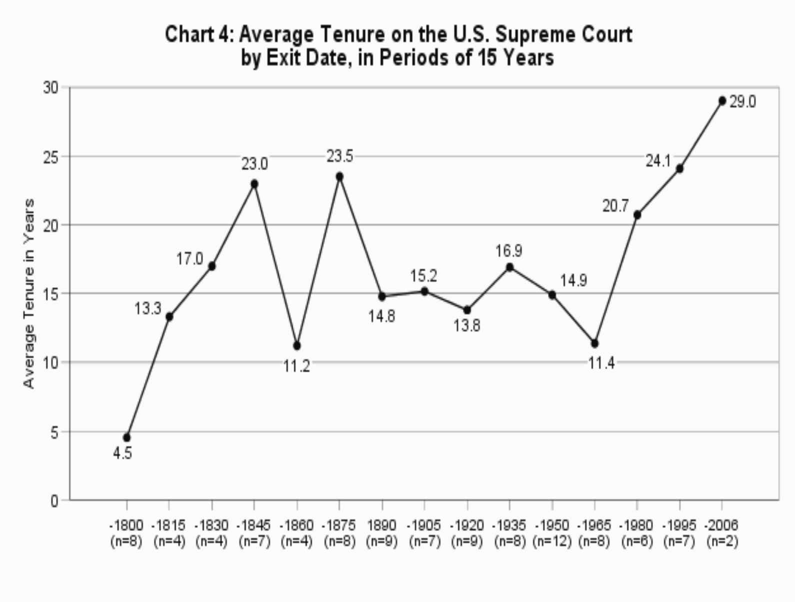


Chart 3: Average Tenure on the U.S. Supreme Court by Exit Date, in Periods of 40 Years

The final period again reveals higher-than-ever average tenure, 21.2 years for the period from 1961–2006. Like Chart 2, however, it does not show an eye-catching "dramatic" rise during the final period. The increase is an unremarkable 6.1 years from the historical average of 15.1 years from 1789–1960, and 6.2 years from the previous-period average of 15.0 years from 1921–1960. Thus, neither the forty- nor fifty-year charts support Calabresi and Lindgren's claim.[53]

If longer time periods are unpersuasive, then perhaps periods shorter than thirty years will provide some support for Calabresi and Lindgren's claim of dramatic and unprecedented growth in length of tenure since 1971. We begin with fifteen-year periods, which approximate the average length of tenure for the entire data set (16.2 years), meaning that each period roughly matches the length of time required for the Court to completely turn over its membership. The results appear in Chart 4:

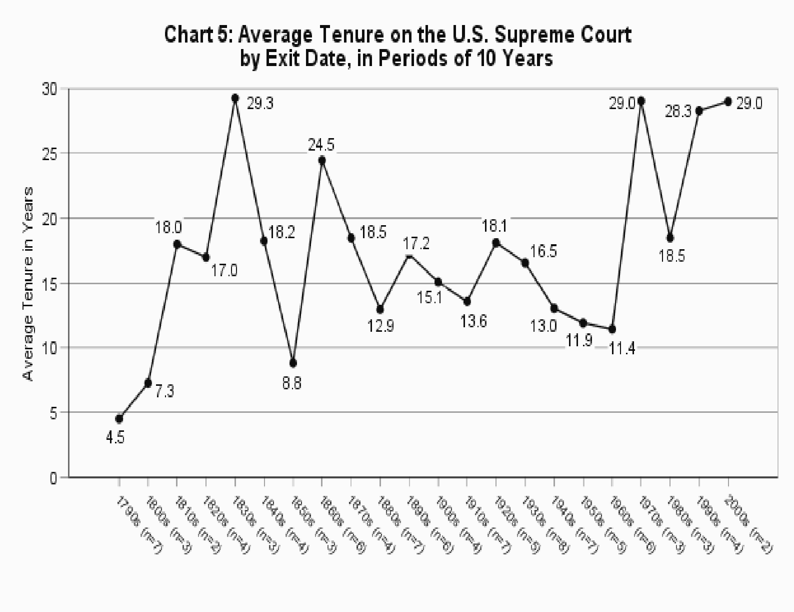


A dramatic trend is now visible in Chart 4: from 11.4 years from 1951–1965, the chart shows a huge jump to 20.7 years

[53] We are reminded of Darrell Huff's description of the "flattened" precursor to "the Gee-Whiz Graph," in *How to Lie with Statistics*: "That is very well if all you want to do is convey information. But suppose you wish to win an argument, shock a reader, move him into action, sell him something. For that, this chart lacks schmaltz." DARRELL HUFF, HOW TO LIE WITH STATISTICS 62 (2d ed. 1993).

from 1966–1980, another increase to 24.1 years from 1981–1995, and an impressive 29.0 years for the period since 1996. That's an increase of 17.6 years since 1951, nearly triple the increase shown in the last chart. Yet the trend does not look "unprecedented." Not once but twice in the nineteenth century, average tenure climbed to approximately the same levels: 23.0 years from 1831–1845, and 23.5 years from 1861–1875. Although the average of twenty-nine years from 1995–2006 represents a new high, it reflects the tenures of only two Justices (Rehnquist and O'Connor), as compared to samples of seven and eight Justices for the two highest nineteenth-century periods. The lengthy tenures of the two most recent departures from the Court do not amount to a trend.

Finally, we render the data using nice round periods of ten years, one period per decade.[54] The results are displayed in Chart 5:



In this chart, the recent increase has a clear precedent. At 29.3 years, the 1830s saw the highest average tenure of any decade in history, narrowly higher than the 29.0-year average tenure of

[54] That was, in part, our reasoning for the period selection in our earlier article, which included a close approximation of Chart 5. *See* Stras & Scott, *supra* note 11, at 1428. This time, heeding Calabresi and Lindgren's criticism, *see* Calabresi & Lindgren, *supra* note 1, at 791–92, we have disclosed the sample size for each decade.

the 1970s and 2000s. Meanwhile, the 1860s, a decade in which six Justices retired, had an average tenure of 24.5 years. As a result, the decades since 1971 do not look meaningfully different from those in the mid-nineteenth century.[55]

As Chart 1 demonstrates, only the choice of roughly thirty-year periods supports both aspects of Calabresi and Lindgren's empirical claim by showing a "dramatic" and "unprecedented" recent increase in tenure for Supreme Court Justices.[56] Thirty-year periods maximize the slope of the line between the last two periods by creating a terminal period of thirty-six years, ensuring an "astonishing" recent trend.[57] Shifting the cutoff date in either direction from 1971, even by a single Justice, would produce a chart with a less dramatic increase in average tenure.[58] Our survey of alternative period charts demonstrates that longer periods of forty or fifty years produce results that contradict the first part of Calabresi and Lindgren's hypothesis—that the increase has been "dramatic." Further, using

[55] Calabresi and Lindgren argue that even a chart using ten-year periods "tends to support, rather than reject, [their] hypothesis" because "three of the four decades with highest mean tenure are three of the last four decades." Calabresi & Lindgren, *supra* note 1, at 792. That claim, however, reflects period selection too. The decade with the fifth-highest tenure is the 1860s (24.5 years), which is noticeably higher than the 1980s (18.5 years). In any event, Calabresi and Lindgren keep moving the target. How can they seriously maintain that life tenure "worked well" before 1970, when it produced decades like the 1830s and 1860s, but "has broken down" now that it has produced decades like the 1970s and 1990s? Calabresi & Lindgren, *supra* note 3, at A12. Only period selection allows the authors to explain away the long tenures of so many mid-nineteenth-century Justices.

[56] Calabresi & Lindgren, *supra* note 1, at 778, 781. Unsurprisingly, period lengths of very close to thirty years, including thirty-five years and twenty-five years, produce charts quite similar in shape to Calabresi and Lindgren's chart. No period length, however, results in a final-period rise as stark as that created by thirty-year periods.

[57] *Id.* at 775.

[58] The revised version of Calabresi and Lindgren's article concedes that there would be a smaller increase in average tenure if they used a slightly *later* cutoff date. *See id.* at 779. They calculate that moving the cutoff to 1975 to include Justices Black, Harlan, and Douglas would increase the penultimate-period average to 15.0 years while reducing the final-period average to 25.1 years, resulting in an increase 27% smaller than the one reported in their chart. *See id.* The same is true for an *earlier* cutoff date. Moving the cutoff to 1965, for example, changes the penultimate-period average to 12.8 years while reducing the final-period average to 22.1 years, resulting in a 33% smaller increase than the one reported by Calabresi and Lindgren. Moving the cutoff date by even a single Justice in either direction also affects average tenure in the final two periods. Moving Justice Black makes the increase in the final period smaller by 13%, while shifting Chief Justice Warren renders the increase smaller by 4%.

shorter periods of ten or fifteen years produces results that contradict the second part of their hypothesis—that the increase in average tenure is "unprecedented." Thus, Calabresi and Lindgren's selection of thirty-year periods was a critical decision in presenting their claim.[59]

[59] The revised version of Calabresi and Lindgren's article includes a series of charts showing the "lagging average" for consecutive groups of nine Justices for each of their three metrics, which are meant to "smooth" the data so that it is less susceptible to significant fluctuations. *See id*. at 780–81, 783, 788 charts 2, 4, & 6. Because the new charts appear in a separate section of the article, it is not clear whether Calabresi and Lindgren intended the lagging-average analysis as a direct response to our period-selection criticism.

Nonetheless, they argue that the lagging-average charts show the data "without any period selected by the researcher." *Id*. at 780. That claim is misleading, however, because the lagging-average charts suffer from a different form of the same problem. Instead of depending on the number of years per period, the lagging-average charts depend on the number of Justices per group. *See id*. at 780. Calabresi and Lindgren have elected to display the data using a lagging average of nine Justices but they might just as easily have chosen a larger or smaller number. Thus, although they have avoided selecting a "period," it is not entirely accurate to suggest that these charts involve no selection "by the researcher." *Id*.

In a footnote, Calabresi and Lindgren anticipate this critique, and state that "[i]f [they] had included more Justices in [their] lagging averages, such as a twelve-Justice lagging average, Chart 2 would have resembled Chart 1 even more closely, and the recent rise in judicial tenure would have appeared even more dramatic." *Id*. at 781 n.39. To test their claim that "includ[ing] more Justices" only makes their argument stronger, we created lagging-average charts for groups of three, five, seven, nine, eleven, thirteen, fifteen, seventeen, and nineteen Justices. We then generated the same charts treating each Justice's tenure as an observation at the date of appointment rather than the date of departure.

We found that lagging-average charts for groups of nine, eleven, or thirteen Justices show a trend in length of tenure that is consistent with Calabresi and Lindgren's empirical claim. Because that conclusion holds true for several values and does not depend on a particular cutoff date, the lagging-average charts lend some support to Calabresi and Lindgren's decision to use thirty-year periods in Chart 1.

Like the charts displaying average tenure over various time periods, however, the lagging-average charts support Calabresi and Lindgren's empirical claim only when particular group sizes within a narrow range are selected. Lagging averages for groups of *fewer* Justices tend to refute the claim of an "unprecedented" recent increase in length of tenure. With three Justices, lagging averages in the mid-nineteenth century reached 29.7 years, longer than any of the lagging averages since 1971, and five of the eight values longer than twenty-seven years occurred between 1830 and 1880. With five Justices, the peak in the nineteenth century reached 28.1 years, higher than every value in history except 2006, and the cluster of lagging averages higher than twenty years is just as impressive in the nineteenth century as in the decades since 1971. Even with seven Justices, the lagging average in 1873 came within 0.2 years of the highest value in 1994 and was higher than the lagging averages in 2005 and 2006. Likewise, lagging averages for groups of more Justices also tend to refute the claim of an unprecedented recent increase in length of tenure. With seventeen and nineteen Justices, lagging averages in the

Originally, we offered a more abbreviated and general criticism of period selection, noting that "[a] year here or there, on one side or the other of a cutoff, and the average for the period might rise or fall considerably."[60] In the revised version of their article, Calabresi and Lindgren call that claim "simply false," and assert that even a "quick check" of the data would have revealed that a change of one year in either direction on their charts rarely would make any difference, "with the largest difference being 1.3 years."[61] "It appears," they wrote, "that Stras and Scott never bothered to check their factual claim."[62]

To the contrary, Calabresi and Lindgren have both misrepresented our claim and underestimated the impact of a one-year shift for earlier periods. First, they omit the immediately preceding sentence in our article, which makes clear that our criticism applies to period selection generally: "A fair criticism of *both* our chart *and* Calabresi and Lindgren's is that too much turns on the arbitrariness of the selected periods."[63] In Chart 5, as we claimed, "[a] year here or there" makes a huge difference.[64] Shifting each period forward by one year, for example, would noticeably reduce the values for the 1810s (by 2.3 years or 15%), the 1870s (by 3.2 years or 18%), and the 1920s (by 2.9 years or 16%); and would increase the values for the 1800s (by 5.3 years or 73%), the 1850s (by 2.4 years or 27%), the 1880s (by 2.5 years or 19%), the 1930s (by 2.9 years or 18%), and the 1980s (by 3.8 years or 21%).[65]

mid-nineteenth century were higher than any value in history. Even with fifteen Justices, the nineteenth-century lagging averages were higher than all but the 2005 and 2006 values, falling short of the 2005 and 2006 values by less than two years.

Moreover, like the period charts, the lagging-average charts depend on Calabresi and Lindgren's assumption that length of tenure should be treated as an observation on the date of death, retirement, or resignation from the Court rather than the date of appointment. *See infra* Part II.B. Changing that assumption eliminates any astonishing and unprecedented trend in the lagging-average charts, regardless of the number of Justices in each group.

[60] Stras & Scott, *supra* note 11, at 1428.

[61] Calabresi & Lindgren, *supra* note 1, at 790.

[62] *Id.*

[63] Stras & Scott, *supra* note 11, at 1428 (emphasis added).

[64] *Id.*

[65] Calabresi and Lindgren also made an arithmetic error in the most recent version of their article. They claim that the largest change resulting from a one-year shift in any of their periods is a "trivial" change for the period from 1941 to 1970. Calabresi & Lindgren, *supra* note 1, at 790. "[I]f the 1941–1970 period had started in 1942 instead of 1941," they argued, then "the mean tenure for Justices leaving the Court in that . . . period would have been 13.5 years, rather than 12.2

Second, by focusing on the periods most relevant to their theory, Calabresi and Lindgren ignore the change that would result from shifting the 1821–1850 period by one year to 1822–1851. The average tenure for the new 1822–1851 period would fall to 19.6 years, while the average for the next period would increase to 19.6 years. As a consequence, what Calabresi and Lindgren reported as a 2.4-year *decline* in average tenure during the mid-nineteenth century would have appeared as no change at all. What qualifies as a "considerable" change is of course debatable, but when a single year here or there eliminates a reported trend between periods, we think the data deserve a closer look.[66]

Regardless, our comment that "[a] year here or there"[67] can affect the shape of the resulting graph was intended as a general description of the period selection problem, and nothing more. As Charts 1 through 5 demonstrate, the choice of time periods for measuring average tenure on the Supreme Court means the difference between accepting and rejecting Calabresi and Lindgren's claim of a dramatic and unprecedented recent increase in Supreme Court tenure.[68]

years . . . ." *Id.* Ironically, they have overestimated the effect of the shift. In fact, moving the start of the 1941–1970 period to 1942 *reduces* the average tenure to 11.4 years by excluding Justices McReynolds and Hughes, and moving both the start and end dates to create a 1942–1971 period increases average tenure to only 12.9 years by including Justices Black and Harlan.

[66] *See* David H. Kaye & David A. Freedman, *Reference Guide on Statistics*, *in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 113 (Fed. Judicial Ctr. ed., 2d ed. 2000) (noting that when "the shape [of the graph] can be altered somewhat by changing the size of the bins," critical readers might find it "worth inquiring how the analyst chose the bin widths").

[67] Stras & Scott, *supra* note 11, at 1428.

[68] Justin Crowe and Christopher Karpowitz suggest yet another reason to be skeptical of Calabresi and Lindgren's empirical claim. *See* Justin Crowe & Christopher Karpowitz, *Where Have You Gone, Sherman Minton? The Decline of the Short-Term Supreme Court Justice* (Princeton Law & Public Affairs Working Paper Series, Working Paper No. 06-014, 2006), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=948813. According to the authors, any increase in long-term tenure is driven in part "by the fact that fewer justices are serving relatively short terms." *Id.* at 1. During the current period, identified by Calabresi and Lindgren as having the most dramatic increase in average tenure in history for Supreme Court Justices, Crowe and Karpowitz find that "we have witnessed the longest period without a short-term justice (37 years) in the Court's history." *Id.* at 8. Even using Calabresi and Lindgren's period lengths, removal of short-term Justices from all earlier periods shows that, "though still a historical high, [the period of 1971–2005] is not radically out of line with any earlier period in the Court's history and is only one and a half years higher than the previous peak." *Id.* at 9. In other words, any increase in average tenure may be attributed more to

B. The Date-of-Observation Problem

Equally important to Calabresi and Lindgren's analysis is their decision to treat each Justice's tenure as an observation in the year of *departure* rather than the year of *appointment*. Unlike a Justice's age at swearing-in, death, resignation, or retirement, which is measured on a single date, the length of a Justice's tenure is calculated as the difference between two dates: (1) the date of appointment (swearing-in), and (2) the date of departure (death, resignation, or retirement). When analyzing the data, researchers therefore must make a decision about when to "count" each Justice's tenure—at the beginning or at the end. In other words, does Justice Alito count as an observation in 2006, or at some date in the future when he dies, retires, or resigns?

For Calabresi and Lindgren, the choice was obvious. The revised version of their article admonishes us for reporting length of tenure as a data point at the time of appointment in one of our charts:

> Given that we are writing about a problem of delayed retirement, not appointment, it is unclear why [Stras and Scott] would seek to test our hypothesis or our groupings by using the date of appointment rather than the date of retirement. They disclose the switch in a footnote, but do not offer a theoretical reason for making it. The most likely effect of this grouping would seem to be to facilitate Type II error by suppressing any patterns in the data.[69]

All three statements are puzzling. The first, which announces that "we are writing about a problem of delayed retirement,"[70] is simply incorrect. None of us is writing *only* about a problem of delayed retirement. The length of each Justice's tenure is a product of a host of factors, including age at appointment, health, income, and pension eligibility, some of which the political branches take into account at the time of appointment.[71] Having not yet determined that there is "a problem," we are simply writing about length of tenure, and

the recent absence of short-term Justices, who arguably are not representative of ordinary life-tenured justices, *see* LEVINSON, *supra* note 46, at 128–29, than to any alarming recent change in the operation of life tenure.

[69] Calabresi & Lindgren, *supra* note 1, at 792.

[70] *Id.*

[71] *See* David R. Stras, *The Incentives Approach to Judicial Retirement*, 90 MINN. L. REV. 1417, 1439, 1444–45 (2006).

there are at least two dates at which researchers can observe that value.

The second statement, that we provided no theoretical reason for relying on the date of appointment, is a fair criticism of our last article, but it cuts both ways. Calabresi and Lindgren have offered no theoretical defense of their choice either.

The third statement, that the "most likely effect" of choosing the date of appointment would be "to facilitate Type II error,"[72] begs the question.[73] Nothing suggests that observing length of tenure at the date of appointment would have any "likely" effect one way or the other—promoting Type I or Type II error—because it simply plots the same data at a different point in each Justice's career.[74] Only by assuming that their hypothesis is correct can Calabresi and Lindgren predict that alternative methods will "likely" produce erroneous results. Also, despite using the language of formal hypothesis-testing, Calabresi and Lindgren never actually stated or tested a formal hypothesis in the initial version of their article, and neither did we. In Part I above, we have done our best to pin down their empirical claim, but without a more rigorous articulation of the hypothesis being tested, discussions of Type I and Type II error are inapposite.

Observing length of tenure at the date of appointment is an equally defensible approach. As Calabresi and Lindgren con-

[72] Calabresi & Lindgren, *supra* note 1, at 792.

[73] Type II error occurs when the null hypothesis is false, but a statistical test fails to reject it, "i.e., there is a false negative." Kaye & Freedman, *supra* note 66, at 176. Type I error, by contrast, occurs when the null hypothesis is true, but a statistical test rejects it, "i.e., there is a false positive." *Id*.

[74] Calabresi and Lindgren may have intended their critique about Type II error as a commentary solely on the second chart that we reported in our first article, which displayed average tenure in non-overlapping groups of five Justices. *See* Stras & Scott, *supra* note 11, at 1429. In that chart, we selected groups of five Justices in an effort to "flatten[] th[e] periods" used in both Calabresi and Lindgren's charts and our own. *Id*. at 1428. In a spectacular mischaracterization, Calabresi and Lindgren chastise us for deliberately "'flatten[ing]' [the] *effects*" and "'flatten[ing]'[] the *unprecedented length of tenure*" in the data. Calabresi & Lindgren, *supra* note 1, at 790, 791 n.57, 793 (emphasis added). As our text made clear, however, we sought in our second chart only to "flatten[] th[e] *periods*," i.e., to use groups that contain an equal number of Justices, rather than periods of years that contain a variable number of Justices. *See* Stras & Scott, *supra* note 11, at 1428 (emphasis added). Our goal was to eliminate a potential source of distortion caused by period selection. We did not hold, and never expressed, a desire to flatten effects in the data.

cede, "the decision to remain on the Court is one that is made continuously," and the apparent "post-1970 trend in part reflects behavior and decisions" from earlier periods.[75] Observing a Justice's entire tenure upon date of departure, however, ignores decisions to stay on the Court and backloads the data.[76] For example, Justices Black and Douglas were sworn in 1937 and 1939, respectively, and twenty-two other Justices have joined and left the Court since then. Yet, under Calabresi and Lindgren's approach, the tenures of Justices Black and Douglas count as two of the twelve most recent observations.[77]

In fact, given Calabresi and Lindgren's focus on the role of the political branches in their criticisms of life tenure,[78] date of appointment is arguably the better choice. Prospective Justices' age and anticipated length of service have become relevant considerations for the political branches when evaluating judicial appointments. By the time a Justice departs from the Court, the work of the political branches has

[75] Calabresi & Lindgren, *supra* note 1, at 779. Political scientists also recognize that the decision of whether to retire is one that is made continuously and over time. Thus, most studies that have analyzed the retirement behavior of Supreme Court Justices have used each year of a Justice's service as a separate observation. *See* Peverill Squire, *Politics and Personal Factors in Retirement from the United States Supreme Court*, 10 POL. BEHAV. 180, 184 (1988); Albert Yoon, *Pensions, Politics, and Judicial Tenure: An Empirical Study of Federal Judges, 1869–2002*, 8 AM. L. & ECON. REV. 143, 150 (2006) (using judgeship-year as the relevant unit of analysis to "account[] for secular and individual-level changes from year to year, with an eye towards examining which factors, if any, explain when judges created a judicial vacancy"); Christopher J.W. Zorn & Steven R. Van Winkle, *A Competing Risks Model of Supreme Court Vacancies, 1789–1992*, 22 POL. BEHAV. 145, 150 (2000). *But see* Timothy M. Hagle, *Strategic Retirements: A Political Model of Turnover on the United States Supreme Court*, 15 POL. BEHAV. 25, 27 (1993) (criticizing Professor Squire's approach because of the large disparity between voluntary retirements and total observations).

[76] Admittedly, observing length of tenure at the date of appointment suffers from the opposite problem by frontloading the data. Using date of appointment, however, is an equally defensible methodological choice because both approaches create similar distorting effects.

As a compromise, researchers could observe length of tenure at the exact midpoint of a Justice's career. That method also yields results that undercut Calabresi and Lindgren's hypothesis of a dramatic and unprecedented recent increase in average tenure. *See infra* note 136.

[77] Using the date of exit backloads the data for other periods as well. For example, the first Justice Harlan joined the Court in 1877, but Calabresi and Lindgren group him with the pre-World War II retirees because he remained on the Court until 1911.

[78] *See* Calabresi & Lindgren, *supra* note 1, at 809–13.

been complete for some time. Only at the time of appointment do the political branches play an active role in ensuring democratic accountability. Observing length of tenure ex ante makes sense because Calabresi and Lindgren raise ex ante concerns.

Calabresi and Lindgren criticize the decision to report tenure data by date of appointment, explaining that their "hypothesis concerns length of tenure at retirement, not appointment."[79] They repeat the criticism endlessly: "our hypothesis [is] about post-1970 retirees,"[80] "our hypothesis [is] that post-1970 retirees are different,"[81] and "[Scott and Stras] suppress[] the pattern we hypothesize."[82] Yet they offer no explanation for *why* they selected their hypothesis.[83] That is ipse dixit, not argument. Having offered a belated explanation of our decision to use date of appointment, we hope that Calabresi and Lindgren can shed some light on their own choice.

The date-of-observation problem deserves attention because it has a pronounced impact on the shape of the data. For example, whether Chief Justice Rehnquist's tenure counts as an event during the administration of Richard Nixon or George W. Bush produces vastly different conclusions about the pattern of judicial tenure. A change in the simple assumption about when to observe the data has profound consequences and leads us to reject both aspects of Calabresi and

[79] *Id.* at 793.

[80] *Id.*

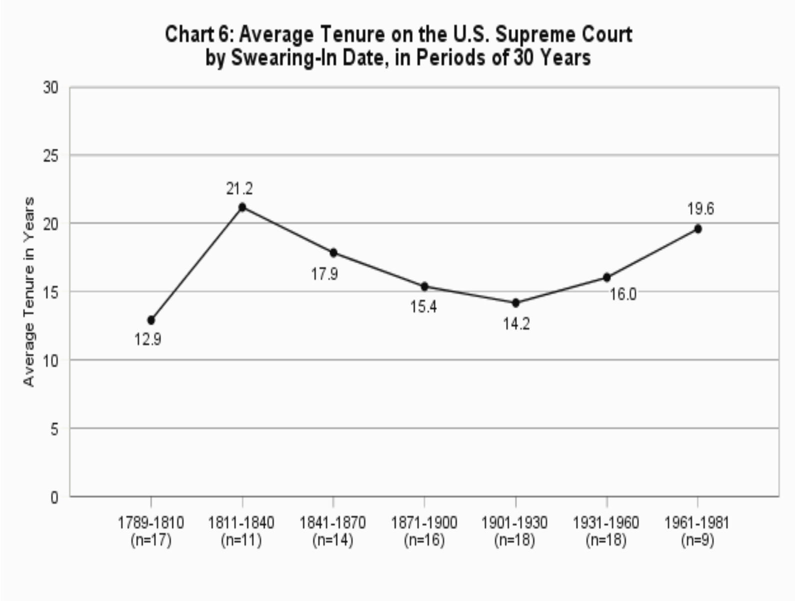
[81] *Id.* at 794.

[82] *Id.* at 792.

[83] The best reason we can imagine is that length of tenure is not known at the date of appointment, but it is known (indeed, it finally becomes known) at the date of death, retirement, or resignation. As a result, if a Justice's term in office counts at the date of appointment, the death of a long-serving Justice will retroactively affect an earlier period.

These factors, however, have never deterred statisticians from using start date (most commonly birth date) rather than end date (death date) when reporting life expectancy data, which are closely analogous to data on length of tenure. In both cases, researchers cannot be certain of the date of death (or departure) until it has occurred, and must retroactively attribute new cases to past dates (birth or appointment). Yet life expectancy is uniformly reported as a value reflecting the average lifespan (tenure) of a person *born* in a particular period rather than the average lifespan (tenure) of a person who *died* during that period. *See, e.g.*, Elizabeth Arias, *United States Life Tables, 2003*, *in* NAT'L VITAL STATS. REPS. 2006, at 34–35 tbl.12, *available at* http://www.cdc.gov/nchs/data/nvsr/nvsr54/nvsr54_14.pdf.

Lindgren's empirical claim. Measured in periods of thirty years, as they prefer, the results are reflected in Chart 6:



The most recent period, which includes Justices appointed since 1961 and no longer serving on the Court, has an average tenure of just 19.6 years. To say that the chart shows no dramatic and unprecedented rise in average tenure is an understatement. In the most recent period, 1961–1981, the data reveal an increase of just 3.6 years over the 16.0-year average for Justices appointed from 1931–1960 and an increase of just 3.7 years over the historical average of 15.9 years for Justices appointed from 1789–1930. The chart also refutes the claim that recent tenure length is unprecedented. Justices appointed from 1811–1840 served an average of 21.2 years on the Court, 1.6 years longer than the most recent period.

Sensitive to the distorting effects of period selection, we also render the length-of-tenure data by date of appointment using periods of fifteen years to approximate the average length of tenure for the entire data set. Chart 7 displays the results:

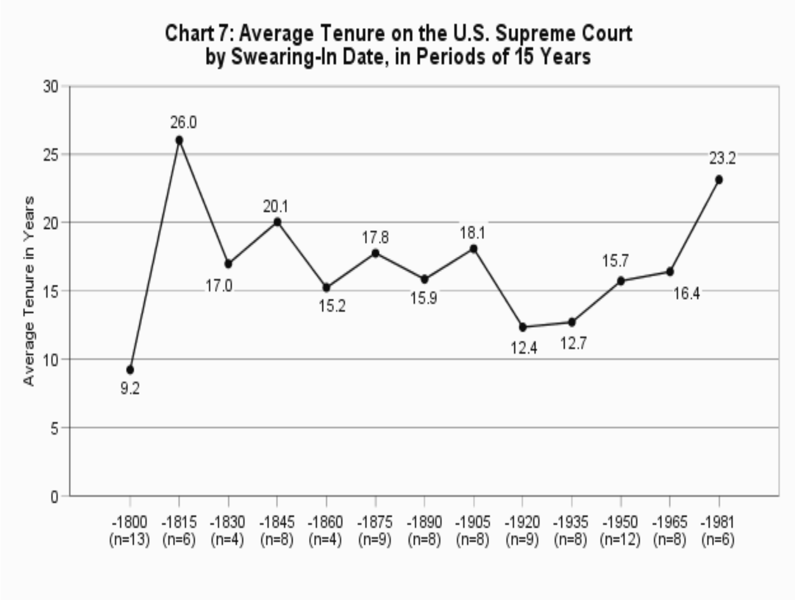


Chart 7: Average Tenure on the U.S. Supreme Court by Swearing-In Date, in Periods of 15 Years

Again the data do not reveal a dramatic and unprecedented increase in the final period. Justices appointed since 1966 have served an average of 23.2 years, an increase of 6.8 years from the 16.4-year average for Justices appointed from 1951–1965. That total still falls short, however, of the average of 26.0 years served by Justices appointed from 1801–1815.

So far we have focused on whether the date of appointment or the date of departure is the *better* choice when reporting data on length of tenure. Regardless of the outcome of that theoretical debate, however, readers should find it alarming that the choice matters at all. If the data contained a *robust* trend toward dramatically longer tenures in recent decades, then the trend should be apparent regardless of when each observation is counted.[84] If Calabresi and Lindgren have identified "a real pattern in the data," as they claim,[85] it should not crumble when such a simple and debatable premise is altered.

[84] A robust trend in length of tenure should be apparent regardless of the method of observation. For example, if average length of tenure were to remain constant for the next fifty years when measured from date of appointment, then any perceived trend during the same period, when measured from date of departure, would simply be an accident of timing in deaths, retirements, and resignations.

[85] *See* Calabresi & Lindgren, *supra* note 1, at 794.

C. Regression Models for Length of Tenure

1. Calabresi and Lindgren's Reported Cubic Model

The most impressive addition to the latest version of Calabresi and Lindgren's article is a set of regression models measuring the relationship between time (the independent variable) and length of tenure (the dependent variable). The authors "fit a linear model and ten different curvilinear ones" to the data.[86] The linear model (p=.023) as well as the cubic (p=.00001), logarithmic (p=.002), inverse (p=.002), compound (p=.003), power (p=.000002), logistic (p=.003), exponential (p=.003), S-curve (p=.00000005), and growth (p=.003) models were all statistically significant.[87] Only the quadratic model was not significant (p=.076), and the authors note that it "would have been significant if one assumed a population of 250 Justices (over perhaps the first 500 years of the United States) rather than the less realistic assumption of an infinite population of Justices . . . ."[88]

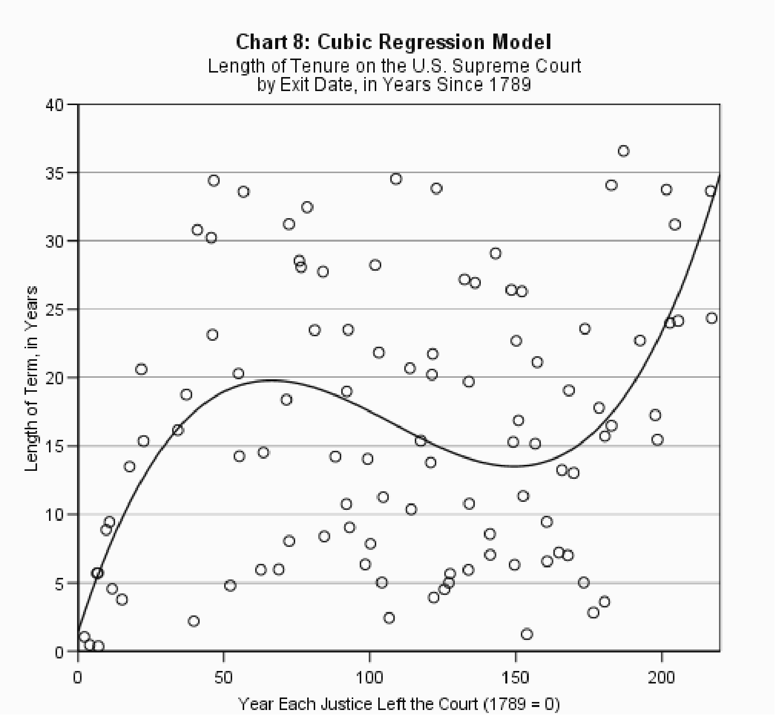
For nine of the eleven models, Calabresi and Lindgren report only the models' statistical significance (p). For the linear model, they note that the correlation coefficient (R) is .223, and we can then calculate that R^2 is .049. They report no other details. For the cubic model, however, they provide a wealth of information, including a full-page chart replicated here as Chart 8.[89]

[86] *Id.* at 797.

[87] *Id.* at 797–98.

[88] *Id.* at 798.

[89] *See id.* at 799 chart 1. We replicated Calabresi and Lindgren's model and, as far as we can tell, our results are identical. Differences in rounding explain any slight disparities in our reported values.



Model Significance=.00001

R^2=.229 Adj. R^2=.206 F=9.806

Standard Error of the Regression: 8.916

	B	Std. Error	T	Significance
	Coefficients			
Year (1789=0)	.65013	.13762	4.7240	.00001
Year Squared	-.00708	.00149	-4.75137	.00001
Year Cubed	.00002	.000005	4.79071	.00001
Constant	1.44793	3.46280	.41814	.67675

The cubic model, they explain, was "among the two best fitting models" that they analyzed.[90] The shape of the curve closely matches the trend line in their original chart using thirty-year periods. They argue that, as a result, the model

[90] Calabresi & Lindgren, *supra* note 1, at 798.

serves as "an elegant confirmation that [the] periodization in Chart 1 presented the data fairly . . . ."[91]

To many readers, the cubic model undoubtedly looks like a coup de grâce. It not only fully avoids period-selection problems, but it also lends analytical heft to what before had been glorified arithmetic. For many lawyers, law students, and journalists, a regression model with an imposing table of coefficients is at once unassailable and incomprehensible.

A brief guide to the data is therefore in order. A regression model helps to explain the relationship between two or more variables,[92] using one or more independent variables (in this case, time in years) to determine expected values for a dependent variable (in this case, length of tenure in years).[93] A linear regression model, for example, calculates the slope and intercept of the line that best "fits" the data.[94] Similarly, nonlinear regression models calculate the best-fitting curve for nonlinear functions. For example, a quadratic regression model calculates the best-fitting quadratic equation, a cubic regression model calculates the best-fitting cubic equation, and so on.

In evaluating a regression model, an important threshold attribute to consider is the *p*-value, which measures the probability of observing "data as extreme as, or more extreme than, the actual data," assuming the null hypothesis is true.[95] A variable is considered *statistically significant* when *p* falls below a pre-established significance level, such as 1% or 5%.[96] Statistical significance should not be confused with practical significance, because *p* does not tell us anything about the magnitude of the trend identified. Instead, it merely determines the level of confidence that the data observed are not

[91] *Id.*

[92] *See* Daniel L. Rubinfeld, *Reference Guide on Multiple Regression*, *in* REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, *supra* note 66, a 181.

[93] *See* Kaye & Freedman, *supra* note 66, at 171.

[94] The best-fitting line or curve is usually calculated by using the "ordinary least squares" method of estimation, which minimizes error by minimizing the sum of the squared residuals—the difference between the predicted value and actual value—for each observation. *See* MICHAEL O. FINKELSTEIN & BRUCE LEVIN, STATISTICS FOR LAWYERS 333 (1990).

[95] Kaye & Freedman, *supra* note 66, at 122.

[96] Rubinfeld, *supra* note 92, at 226.

the product of chance.[97] The significance of the variables in the cubic model, .00001, is excellent. As Calabresi and Lindgren explain, "[t]he probability that we would see the trend shown in the cubic model if the data were random is only one chance in 100,000."[98]

Conspicuously absent from their analysis, however, is any discussion of the explanatory power of their model or the closeness of fit of the model to the data. Because a regression equation almost never perfectly fits the observed values of the dependent variable, "it is important to measure how well the equation performs."[99] The most frequently-cited measure of the explanatory power of a regression model is the R^2 statistic. The R^2 statistic is a value between 0 and 1 that describes the percentage of variance in the dependent variable that is explained by the independent variable according to the regression equation.[100] R^2 is biased upward slightly because a regression equation is "optimally fitted to the data,"[101] and must be adjusted downward (yielding the "adjusted R^2") to serve as an estimate of the proportion of variability explained in the population as a whole.[102] A value of R^2 close to 1 means a strong fit, while a value of R^2 close to 0 means a weak fit.[103] Although "[m]oderate associations are the general rule in the social sciences"[104] and "there is no clear-cut answer" for the level of R^2 that indicates a satisfactory model,[105] a model with

[97] The *t*-statistic serves a similar function for each parameter. If the absolute value of *t* for a particular parameter estimate is greater than 1.96, then the estimate is statistically significant. *Id*. at 214.

[98] Calabresi & Lindgren, *supra* note 1, at 798.

[99] FINKELSTEIN & LEVIN, *supra* note 94, at 345.

[100] *Id*.

[101] *Id*. at 346.

[102] *Id*.

[103] Rubinfeld, *supra* note 92, at 215–16.

[104] Kaye & Freedman, *supra* note 66, at 136.

[105] Rubinfeld, *supra* note 92, at 216. A few statisticians have proposed rules of thumb, but such rules should be received with great caution. *See, e.g.*, R. SENTER, ANALYSIS OF DATA: INTRODUCTORY STATISTICS FOR THE BEHAVIORAL SCIENCES 433 (1969) (describing a system wherein values of *r* between 0.40 and 0.70 demonstrate a "substantial" relationship, while values of *r* less than 0.20 demonstrate a "slight, almost negligible" relationship); *see also* Stephen J. Schulhofer, *Harm and Punishment: A Critique of Emphasis on the Results of Conduct in the Criminal Law*, 122 U. PA. L. REV. 1497, 1548 (1974) ("R^2 values for these models are uniformly low, generally less than 0.30. Policy recommendations can scarcely be drawn from such inconclusive studies."). Over the years, a few expert witnesses have advanced rules of thumb as well, and the broad differences between them nicely illustrates

a higher R^2 value by definition provides a better explanation for the variability in the dependent variable.[106]

Another value reported in Chart 6, the standard of error of the regression (SER), can measure closeness of fit. The SER can be thought of as the standard deviation in the distribution of error around the regression line: "[o]ther things being equal, the larger the SER, the poorer the fit of the data to the model."[107] The SER can be used to compare regression models, but its most common application is to calculate prediction intervals around the value of the dependent variable.[108]

In Calabresi and Lindgren's cubic model, the standard error of the regression is 8.9. Although this statistic is not directly interpretable, it yields an average value for the 95% prediction interval in the cubic model of 18.02 years. In 95 cases out of 100, therefore, the cubic model's prediction of the length of tenure of a Justice will be, on average, within 18.02 years of its actual value. With an average length of tenure of only 16.2 years, such a large prediction interval indicates that the cubic model is a relatively poor fit to the data.

The cubic model's explanatory power fares no better. The R^2 is .229 and the adjusted R^2 is .206, which means that 22.9% of the observed variance in length of tenure can be explained as a (cubic) function of time. Thus, although the model is statistically significant, it identifies a weak relationship between time and tenure, failing to explain 77.1% of the variance.

the futility of the exercise. *Compare* Lucas v. Townsend, 967 F.2d 549, 552 (11th Cir. 1992) (recounting expert testimony that an R^2 of less than 0.5 "would not be strong because over 51% of the variance" in the dependent variable "would not be attributable" to the independent variable), *with* Boston Chapter, NAACP, Inc. v. Beecher, 504 F.2d 1017, 1024 n.13 (1st Cir. 1974) (recounting expert testimony that a model is "practically significant" if R^2 exceeds 0.09, thereby "explaining 9% or more of the observed variation").

[106] *See* JOHN E. FREUND, MATHEMATICAL STATISTICS 334–35 (2d ed. 1971).

[107] Rubinfeld, *supra* note 92, at 215.

[108] *See* SANFORD WEISBERG, APPLIED LINEAR REGRESSION 22 (2005). Prediction intervals are a range of values that have a specified probability (usually 95%) of containing the value of the dependent variable based on the observed value of an independent variable. *See id*. at 21. Prediction intervals are similar to confidence intervals, except that the latter estimates an unobservable population parameter, while the former predicts the distribution of individual points. *See id*. at 20.

Calabresi and Lindgren also concede that their cubic model is weak as a predictive model.[109] If their model is to be taken seriously, then average length of tenure already has increased to approximately 34 years, a total reached by only four Justices in history: Chief Justice Marshall (1835) and Justices Field (1897), Black (1971), and Douglas (1975). Their model predicts that by 2016, Justices will serve an average of 40 years on the Court. By 2036, the average will have climbed to around 60 years. Calabresi and Lindgren acknowledge that "the sharp rate of increase shown in the model for recent years is not sustainable,"[110] but they do not seem to appreciate the gravity of their concession. How can we trust the model if it is already demonstrably unreliable for predictive purposes? Evaluated on its own terms, there is less to Calabresi and Lindgren's cubic model than first appears.

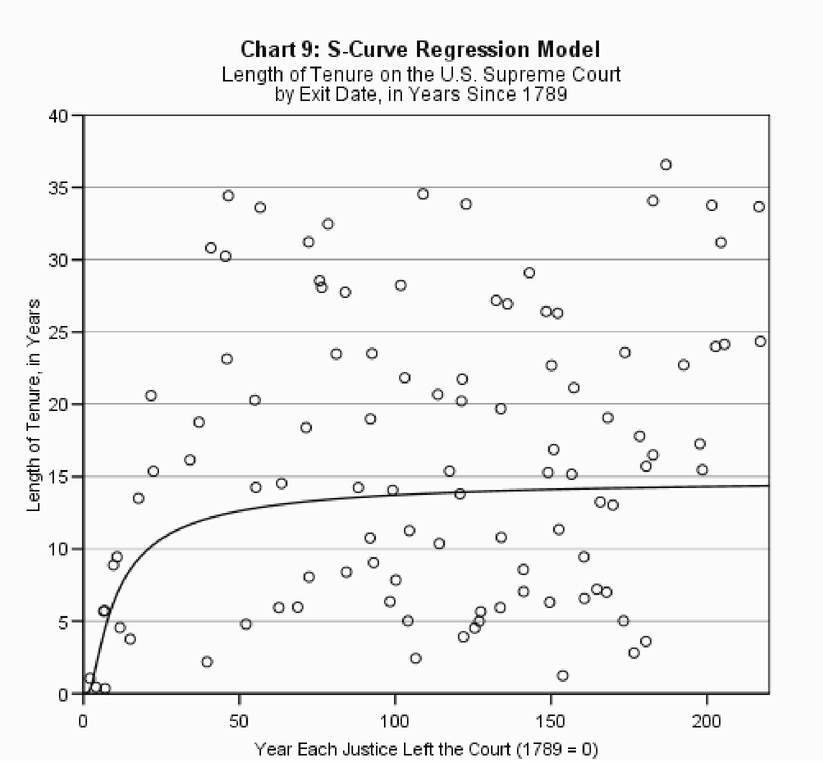
2. *Calabresi and Lindgren's Unreported S-Curve Model*

It is not at all unusual for regression models in the social sciences to have low explanatory power or a weak fit, so Calabresi and Lindgren's model would still have *some* persuasive force if it were the best-fitting model to the data. Perhaps then it would even vindicate their decision to render the tenure data in thirty-year periods, since the cubic regression has approximately the same shape as their original chart.[111] As it turns out, however, the cubic regression model is not the best-fitting curve. In our efforts to replicate Calabresi and Lindgren's findings, we discovered that an S-curve regression model—one of the ten curvilinear models they fit to the data on length of tenure—has a very different shape, greater explanatory value, and better closeness of fit.

[109] *See* Calabresi & Lindgren, *supra* note 1, at 798.

[110] *Id.*

[111] *Compare id.* at 778 chart 1 (rendering the tenure data in thirty-year periods), *with id.* at 799 chart 7 (displaying Calabresi and Lindgren's cubic regression model).



Model Significance=.00000002

R²=.266 Adj. R²=.259 F=36.582

Standard Error of the Regression: 0.788

<u>Coefficients</u>

	B	Std. Error	t	Significance
1/Year (1789=0)	-8.33931	1.37879	-6.04830	.00000002
Constant	2.70214	.08477	31.87728	<.000000001

The shape of the S-curve model bears almost no resemblance to the cubic model. Rather than an "astonishing" recent increase, it shows a sharp increase only in the nation's first fifty years, followed by slow and steady growth in tenure thereafter. It therefore contradicts both aspects of Calabresi and Lindgren's hypothesis, and calls into question their choice to render the tenure data in periods of thirty years.

The authors' unreported S-curve regression model is superior to their reported cubic regression model in several ways. First, it is statistically significant at a higher level.[112] We can be highly confident (1 in 100,000 odds) based on the cubic model that the data are not the product of chance, but we can be even more confident (greater than 1 in 1,000,000 odds) based on the S-curve model.

Second, the S-curve model is a better fit. The standard error of the regression for the S-curve model is .788, yielding an average prediction interval of 13.54 years at the 95% confidence level, as opposed to 18.02 years for Calabresi and Lindgren's cubic model. The prediction interval for the S-curve model is thus 33% smaller at the 95% level than the cubic model, indicating that the S-curve model more closely fits the data.

Third, the S-curve model has greater explanatory power. Its R^2 is .266 and its adjusted R^2 is .259, meaning that it explains 26.6% of the observed variance in length of tenure. By comparison, the cubic model explains only 22.9% of the observed variance. The S-curve model is therefore at least 16% stronger in its explanatory power than Calabresi and Lindgren's cubic model.[113]

Fourth, the S-curve model possesses superior predictive reliability. According to the model, Supreme Court Justices are currently averaging terms just under 15 years, and that average will continue to grow—slowly but steadily—over time. Those numbers seem a bit low to us; we expect that average tenure is already a bit higher, and will grow as the long-serving members of the Rehnquist Court reach the end of their terms in office. But at least the S-curve model offers a *plausible* explanation for average tenure on the Supreme Court, both at present and in the future. Calabresi and Lindgren's cubic model, by contrast, is already wildly implausible and rapidly gets worse.

[112] *Cf.* Posting of James Lindgren to The Volokh Conspiracy, http://www.volokh.com/posts/1162276977.shtml (Oct. 31, 2006, 12:42) (faulting a study using regression analyses for reporting that internet access at home reduces the incidence of rape, when in fact "the observed rape increasing effect of *computer* access is even more highly significant than the observed rape decreasing effect of *internet* access") (emphasis added).

[113] *See* Rubinfeld, *supra* note 92, at 215–16 (stating that the value of R^2 corresponds to "the percentage of variation in the dependent variable that is accounted for by all the explanatory variables"). Measured by adjusted R^2, the S-curve model is 25% stronger than the cubic model, which has an adjusted R^2 of .206.

In light of these stark differences, we are puzzled about why Calabresi and Lindgren reported only the cubic model, and not the S-curve model. Their text makes clear that they generated both, reporting the significance of the S-curve model parenthetically.[114] They also make a cryptic reference to the existence of a better curve, boasting that "the cubic model was among the two best fitting models[]."[115] Unfortunately, they never disclose which model was the "best fitting," or offer a reason for their decision to report only the model with the second-best fit. Regression models using an S-curve—that is, a sigmoidal curve—are used in human population studies where finite resources are presumed to cause rapid growth to slow as a variable approaches some ceiling.[116] An S-curve model seems especially appropriate when analyzing data on life tenure because, as Calabresi and Lindgren acknowledge, "anticipated life expectancies" may act as a ceiling and cause tenure length "to level off or grow much more slowly than the cubic model . . . would indicate."[117] Thus, we are unsure why the authors selected only the cubic model for discussion.

3. *The Null Hypothesis*

One of the most frustrating aspects of the revised version of Calabresi and Lindgren's article is that it repeatedly ascribes to us the view that there is no trend in the data on length of tenure. They repeatedly claim that we endorse that "null hypothesis,"[118] and that we have "strain[ed] to find ways of presenting the data" to reach that result.[119] They have not only badly mischaracterized our argument, but in the course of demolishing a straw man, they have inadvertently revealed weaknesses in their own claim.

Two phrases from our article apparently led Calabresi and Lindgren to believe that we endorse the null hypothesis. The first is our statement that their empirical claim "depends more

[114] *See* Calabresi & Lindgren, *supra* note 1, at 798.

[115] *Id*.

[116] *See, e.g.*, DAVID RIESMAN ET AL., THE LONELY CROWD: A STUDY OF THE CHANGING AMERICAN CHARACTER 7 (abr. and rev. ed. 2001); HENK A. DE GANS, POPULATION FORECASTS 1895–1945: THE TRANSITION TO MODERNITY 53, 99 & nn. 50–51 (1999).

[117] Calabresi & Lindgren, *supra* note 1, at 798.

[118] *Id*. at 798.

[119] *Id*. at 791.

on the chosen period lengths than a bona fide trend."[120] We did not assert that a "bona fide trend" does not exist, but only that "the chosen period lengths" were doing most of the work. That should have been clear from the first sentence of the next paragraph, omitted from Calabresi and Lindgren's discussion, in which we argued that the selection of thirty-year periods and a cutoff date of 1971 had the effect of "exaggerat[ing] the trend."[121] To use the word "exaggerate" is to concede the existence of a trend, and to dispute only the way it has been described.[122] The second is our statement that "the trend *looks* more like a random walk than a steady climb."[123] Calabresi and Lindgren interpret this language as tantamount to a claim that there exists "no time trend in these data if they were actually a random sample of Justices from a population of infinite size."[124] We intended the phrase "random walk" more in its popular sense, however, as in the central claim of *A Random Walk Down Wall Street*[125] that the long-term trend in capital markets is one of slow and steady growth, but that it is easy to be fooled by "astonishing" short-term patterns.

Even if those phrases were confusing, our explicit *refusal* to endorse the null hypothesis on the next page of our article ought to have left no doubt about our position:

> Although we believe there has been no dramatic change since 1971, and that present average tenure falls within the expectations of the founders, *we do not deny that the average length may increase over time due to advances in medical care*. The last members of the Rehnquist Court, who served together longer than any contingent since the 1820s, could nudge average term lengths upward in the coming decades. Nonetheless, given the nation's history of periods in which average tenure reached levels every bit as high as those that prevail today, we have not come close to a point of constitutional crisis.[126]

[120] Stras & Scott, *supra* note 11, at 1427.

[121] *Id.*

[122] *See* WEBSTER'S THIRD NEW INTERNATIONAL DICTIONARY 790 (1961) (defining "exaggerate" as "to misrepresent on the side of largeness (as of size, extent, or value): overstate the truth").

[123] Stras & Scott, *supra* note 11, at 1427 (emphasis added).

[124] Calabresi & Lindgren, *supra* note 1, at 797.

[125] BURTON G. MALKIEL, A RANDOM WALK DOWN WALL STREET (8th ed. 2003).

[126] Stras & Scott, *supra* note 11, at 1430 (footnotes omitted) (emphasis added).

Calabresi and Lindgren never mention this paragraph, and instead repeatedly allege that we believe the data to be entirely random. That is not now, nor has it ever been, our position.

By characterizing us as stalwart defenders of the null hypothesis, Calabresi and Lindgren were able to tilt against an easy opponent. There is an important difference, however, between *rejecting* the null hypothesis and *accepting* Calabresi and Lindgren's sea-change-since-1970 hypothesis. Indeed, in their zeal to disprove the former, they may have undermined the latter.

The battery of regression models developed by Calabresi and Lindgren were designed to roundly reject the null hypothesis. They emphasize that virtually all of the curves they fit to the data produced statistically significant results, including the linear, cubic, logarithmic, inverse, compound, logistic, exponential, S, and growth models.[127] Only the quadratic model was not significant at the 5% level (p=.076), and the authors note that it "would have been significant if one assumed a population of 250 Justices (over perhaps the first 500 years of the United States) rather than the less realistic assumption of an infinite population of Justices . . . ."[128] Based on these models, Calabresi and Lindgren argue that they can "reject the Stras and Scott null hypothesis . . . with an extremely high degree of confidence."[129]

Although the models were statistically significant, the authors provide no information about them—perhaps for a reason. The curves themselves do not appear to have been selected for particular theoretical reasons. Indeed, the choice of models seems entirely pedestrian: those ten curves are the default models that can be generated by checking all of the boxes in the "Curve Estimation" dialogue in SPSS.[130] Chart 10 summarizes the results of Calabresi and Lindgren's remaining unreported models:

[127] *See* Calabresi & Lindgren, *supra* note 1, at 797–98.

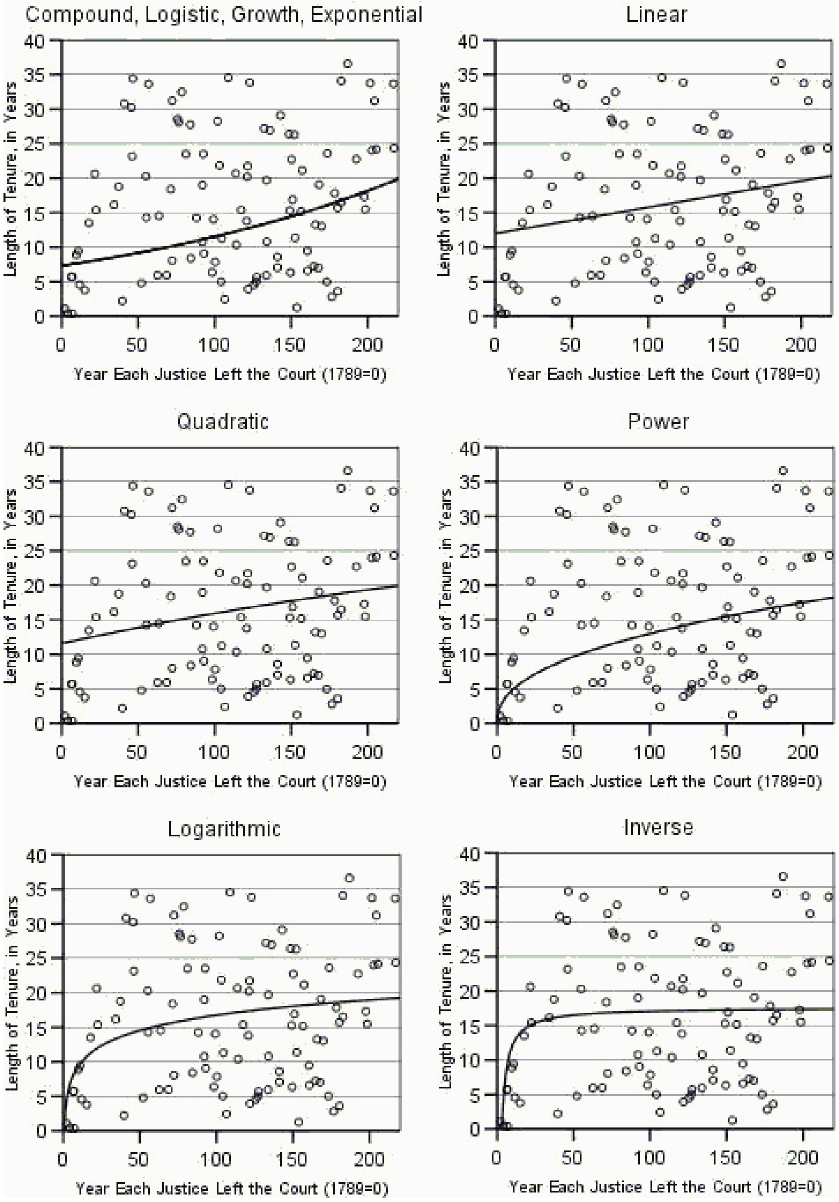
[128] *Id*. at 798.

[129] *Id*.

[130] *See* COMPREHENSIVE STATISTICAL SOFTWARE 7, *available at* http://www.spss.com/pdfs/S15SPClr.pdf (marketing brochure for SPSS, boasting that eleven trend-regression models are available: "[l]inear, logarithmic, inverse, quadratic, cubic, compound, power, S, growth, exponential, and logistic"). SPSS is a leading statistics software package for social scientists. *See id*. at 1.

Chart 10: Additional Regression Models

Length of Tenure on the U.S. Supreme Court
by Exit Date, in Years Since 1789



	Significance	R^2	Adjusted R^2	F	SER
Compound	.003	.085	.076	9.418	0.879
Growth	.003	.085	.076	9.418	0.879
Exponential	.003	.085	.076	9.418	0.879
Logistic	.003	.085	.076	9.418	0.879
Linear	.023	.050	.041	5.327	9.798
Quadratic	.075	.050	.031	2.652	9.846
Power	.000003	.196	.188	24.622	0.824
Logarithmic	.002	.090	.081	9.993	9.590
Inverse	.001	.099	.090	11.104	9.543

Four of the models (compound, logistic, growth, and exponential) are in fact identical and show only slow and steady growth, slightly accelerating over time. Three additional models (linear, quadratic, and power) likewise show no spike since 1970. Instead, they show either consistent growth over the entire data set or a slight deceleration of growth over time. The last two models (logarithmic and inverse) not only show no dramatic increase since 1970, but reveal little growth at all for the past 150 years.

Calabresi and Lindgren take pains to confirm that every one of these models is statistically significant, and by doing so they undoubtedly have disproved the null hypothesis. Yet each of these unreported models actually tends to refute their hypothesis concerning post-1970 retirees. Having developed eleven models, ten of which cut against their hypothesis, it is striking that Calabresi and Lindgren fully reported only the one favorable model, notwithstanding its inferior fit, explanatory power, and predictive reliability.

Calabresi and Lindgren's unreported regression models have the ironic effect of reinforcing our competing narrative about changes in life tenure. We maintain that, notwithstanding short-term fluctuations, the long-term trend in tenure length is slow and steady growth. Because the unreported S-curve model fits the data better than Calabresi and Lindgren's reported cubic model, our narrative, to date, offers a stronger explanation.

4. *The Date-of-Observation Problem Revisited*

The above comparison of regression models assumes, in accordance with the assumption made by Calabresi and

Lindgren, that length of tenure should be treated as an observation made at the date of departure from the Court. As we explained in our discussion of period selection, however, it is equally reasonable to treat length of tenure as an observation made at the date of appointment.[131] Further, regardless of which point of observation has greater theoretical merit, evaluating the data from both perspectives permitted us to test (and ultimately reject) Calabresi and Lindgren's conclusion that there has been dramatic and unprecedented post-1971 growth in average tenure on the Supreme Court.[132]

This technique is particularly useful in testing the robustness of Calabresi and Lindgren's regression models. In evaluating a regression model, "[t]he issue of robustness—whether regression results are sensitive to slight modifications in assumptions (e.g., that the data are measured accurately)—is of vital importance."[133] We therefore ran each of the eleven regression models treating length of tenure as an observation made at the date of appointment, rather than the date of departure.

The results are dramatically weaker. Chart 11 shows the two best-fitting models, cubic and power, side by side:

[131] *See supra* notes 78–83 and accompanying text (noting that because Calabresi and Lindgren's critique of life tenure focuses largely on the role of the political branches, which view tenure length ex ante rather than ex post, it is equally defensible to count each observation as occurring at the date of appointment rather than the date of departure).

[132] *See supra* notes 84–85 and accompanying text.

[133] Rubinfeld, *supra* note 92, at 195.

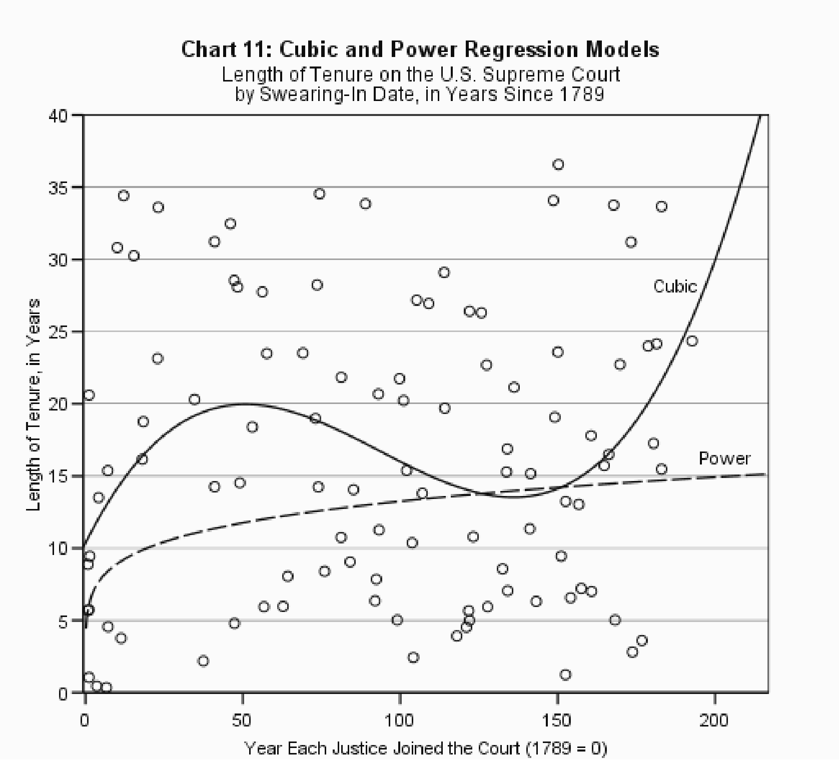


Chart 11: Cubic and Power Regression Models
Length of Tenure on the U.S. Supreme Court
by Swearing-In Date, in Years Since 1789

	Significance	R^2	Adjusted R^2	F	SER
Cubic	.030	.086	.058	3.109	9.708
Power	.009	.065	.056	7.026	0.889

The cubic model supports Calabresi and Lindgren's empirical claim, showing dramatic and unprecedented growth in length of tenure over the past several decades. The power model, by contrast, undercuts their claim by showing slow and steady growth that has decelerated in recent decades.[134]

Both models are statistically significant at the 5% confidence level, and the power model is significant even at the 1% level. Neither, however, fits the data well or has much explanatory power. The standard error of regression is high for both models, yielding an average value for the 95% prediction interval of

[134] The power model has a slope and shape similar to the S-curve model discussed in Part II.C.2.

19.63 years for the cubic model and 14.11 years for the power model. Although the cubic model (R^2=.086, adjusted R^2=.058) has slightly more explanatory power than the power model (R^2=.065, adjusted R^2=.056), the cubic model still can explain only 8.6% of the observed variation in length of tenure as a function of time. That is less than one-third of the variance explained by the S-curve model based on date of departure. By any standard, these models are simply too weak to allow us to draw any conclusions.[135]

Once again, treating length of tenure as an observation made at the date of appointment, rather than at the date of departure, dramatically changes the results.[136] Far from "elegant[ly] confirm[ing]" Calabresi and Lindgren's empirical claim,[137] these models are doubly inconclusive, pointing in opposite directions and explaining less than 9% of the variability in length of tenure. They not only call into question the robustness of the authors' cubic model, but also reduce our confidence that they have identified the correct trend in the data.

CONCLUSION

One irony of our debate with Professors Calabresi and Lindgren is that we agree on so many issues. We share their concerns about mental and physical infirmity on the Supreme Court, and have proposed a "golden parachute" to induce Justices to retire before decrepitude sets in.[138] We share their belief that statutory efforts to abolish life tenure are unconstitu-

[135] *See* SENTER, *supra* note 105, at 433; *cf.* Jeffrey S. Kinsler, *The LSAT Myth*, 20 ST. LOUIS U. PUB. L. REV. 393, 398 (2001) (describing a model in which LSAT performance "accounted for less than 4% of the variance witnessed in law school performance" as documenting "a very weak correlation by any standard").

[136] Observing length of tenure at the exact midpoint of a Justice's tenure also undercuts Calabresi and Lindgren's claim of dramatic and unprecedented recent growth in average tenure. As with the regression models examining length of tenure at date of exit, the strongest model is an S-curve model (p=.000001, R^2=.206, adjusted R^2=.199) with an almost identical shape to the curve displayed in Chart 9, not a cubic model (p=.0003, R^2=.173, adjusted R^2=.147), with a similar shape to the curve in Chart 8. Again, the S-curve model is statistically significant at a higher level of confidence, explains more of the variance in length of tenure, and has greater predictive reliability. It also fits the data better, with an average prediction interval of 13.72 years at the 95% confidence level, roughly 33% lower than the average prediction interval of 18.67 years for the cubic model.

[137] *See* Calabresi & Lindgren, *supra* note 1, at 798.

[138] *See* Stras & Scott, *supra* note 11, at 1398.

tional.[139] We also share their enthusiasm for other institutional reforms, especially concerning the Court's workload.[140] Although we remain unconvinced that Calabresi and Lindgren have adequately supported their empirical claim, we in no way disparage their valuable contributions to the field.

This Article has refined and elaborated upon our critique of Calabresi and Lindgren's empirical claim that there has been a dramatic and unprecedented increase in average tenure on the Supreme Court in recent decades. Their influential chart, documenting an "astounding" increase since 1971, has two flaws. First, it suffers from a period-selection problem. Rendering the data using longer or shorter periods blunts or eliminates the dramatic and unprecedented trend. Second, it suffers from a date-of-observation problem. Treating each Justice's term as an observation made at the date of appointment, rather than at the date of departure from the Court, also eliminates the dramatic and unprecedented trend.

The battery of regression models that Calabresi and Lindgren added to the most recent version of the article does not help their cause. To the contrary, all but one of the models—the one that they reported—refute their hypothesis by showing slow and steady growth in length of tenure over time. In particular, an S-curve model squarely refutes their claim, fits the data better, and has greater explanatory power than the cubic model that they reported. Although the data reveal a trend over time, the most accurate description of that trend is not a dramatic and unprecedented increase in average tenure on the Supreme Court since 1971.

[139] *Id*. at 1419–20.

[140] David R. Stras, *Why Supreme Court Justices Should Ride Circuit Again*, 91 MINN. L. REV. (forthcoming 2007).