The Effect of Age on the Performance and Retirement Decisions for Highly-Skilled Workers: The Case of State Supreme Court Judges

Elliott Ash and W Bentley MacLeod*

July 7, 2019

Abstract

This paper provides evidence on how high-skill work performance changes in response to biological aging over time using data on the judgments written by all state supreme court judges in the US for the years 1947 through 1994. Over time judge performance, as measured by influence on future decisions, rises and then falls. Within a court-year, younger judges are more productive, but this effect is mitigated in courts where there is discretion in the allocation of cases. Finally, states that introduce mandatory retirement at age 70 or 72 experience an increase in court performance.

*Elliott Ash: ashe@ethz.ch; Department of Social Sciences, ETH Zurich. W. Bentley MacLeod: bentley.macleod@columbia.edu; Department of Economics, Columbia University. We thank Yisehak Abraham, Ankeet Ball, Josh Brown, Josh Burton, Matthew Buck, Eammonn Campbell, Zoey Chopra, Daniel Deibler, Seth Fromer, Gohar Harutyunyan, Archan Hazra, Montague Hung, Dong Hyeun, Mithun Kamath, James Kim, Michael Kurish, Jennifer Kutsunai, Steven Lau, Sharon Liao, Sarah MacDougall, Justin McNamee, Sourabh Mishra, Brendan Moore, Arielle Napoli, Karen Orchansky, Bryn Paslawski, Olga Peshko, Quinton Robbins, Ricardo Rogriguez, Jerry Shi, Xiaofeng Shi, Carol Shou, Alex Swift, Holly Toczko, Tom Verderame, Sam Waters, Sophie Wilkowske, John Yang, Geoffrey Zee, Fred Zhu, and Jon Zytnick for their meticulous help in assembling data and other research assistance. Columbia University's Program for Economic Research, Columbia Law School, Princeton University's Center for Health and Wellbeing, and the National Science Foundation Grant SES-1260875 provided financial support for this research.

1 Introduction

At some point all good things must come an end, including our careers. This is sometimes reflected in mandatory retirement rules that require individuals to stop their regular employment at age 65. The benefit of such a rule is that it allows individuals to plan in advance their savings, so that they may enjoy the fruits of a long working life. The cost of such a rule is that many individuals are still productive at age 65 and would prefer to continue working in some capacity.[1] In response, the United States introduced the Age Discrimination Act of 1967 whose goal is "to promote employment of older persons based on their ability rather than age; to prohibit arbitrary age discrimination in employment; to help employers and workers find ways of meeting problems arising from the impact of age on employment."[2]

From the perspective of labor economics, mandatory retirement rules (and government's need to regulate them) are a puzzle. Physical and cognitive decline for most individuals begins in their thirties, at the beginning of most individuals' working careers. However, people continue to learn and enhance their skills, so that wages and employment continue to rise after the start of physical and cognitive decline and only begins to fall for individuals older than fifty years (Medoff and Abraham, 1980; Abraham and Farber, 1987). Such an observation is consistent with a perfectly competitive market where wages respond to a person's productivity. In a competitive market employment should continue as long as the productivity of the individual exceeds their wage.

However, measuring the productivity of a skilled professional is very difficult, and reasonable people may disagree regarding an individual's productivity. Age-related performance declines take several years, and hence it may not be clear when is the best time to step down. Moreover, if a professional is permanently employed, then dismissal must be with cause. In other words, the employer needs to provide documentary evidence that performance is not sufficient. In many cases, as in the case of the judges we consider in this study, compensation does not fall with age, and hence one cannot use wage cuts to bring compensation in line with performance. Moreover, few human resource managers are likely to enjoy telling a valued employee that their performance has declined so much that it is time for them to cease regular employment.[3]

[1] In 2010, conditional upon reaching 65, one could expect about 19 more years of life, up from about 14 years in 1960.

[2] The text of the law is included in the appendix.

[3] See MacLeod (2003) on conflict that can arise when employer and employee have different views

A mandatory retirement rule is a *decision right* that provides a potential solution to this problem. When parties have agreed in advance to a specified retirement date, then the employee's performance is not being called into question. Yet, the employer can, if she wishes, continue to employ the retiree either in a similar job, or in a job for which he is better suited. Hence, in the absence of a mandatory retirement rule the employer must provide evidence that an individual's performance is no longer acceptable. It is natural that the individual involved may not agree, or may wish to bring an age discrimination claim. Under mandatory retirement, in contrast, employment unconditionally ceases at a given date. At that point, depending upon the contract and legal jurisdiction, the employer may be free to continue employment. The key ingredient is that in the absence of an affirmative action by the employer, employment ceases without requiring an explicit evaluation of employee competence.

One reading of the Age Discrimination Act of 1967 is that the legislators recognize that a consequence of mandatory retirement is that employers avoid completely a careful evaluation of the performance of the retiree, and hence productive employment of older workers is lower than it could be. Economic theory provides little guidance since it predicts either that mandatory retirement would have no effect (due to cost-less renegotiation), or if it does have an effect then private parties should be free to choose the rule they prefer. In the end, the only way to assess the effects of these rules is with data.

In this paper we study the careers of State Supreme Court judges, an environment that is well suited for addressing these issues. These judges review appeals regarding state law, and they are among the most powerful individuals in a state. Hence, it is important to understand the best way to manage their employment. We have data from 1947 to 1994, allowing us to track the performance of judges over full careers. Meanwhile, in order to ensure that judges face no pecuniary incentive to tilt decisions, compensation does not vary between judges, is not contingent on work performance, nor does it vary substantially over time (see Landes and Posner, 2009).

A particularly appealing feature is that the job of a judge has not changed substantially over time. The job consists of hiring staff to assist in their work and hear appeals of trial court decisions that are assigned to them. They then write a decision with the help of the staff that they have hired. Much of the scientific data on the effects of age comes from time invariant physical and cognitive tests, which do not allow one to make

regarding performance.

inter-temporal comparisons. The task of judging provides a unique context to explore the effects of aging over time for this important group of professionals that would be difficult to replicate with other professions.

The challenge is that producing useful measures of worker performance, particularly for complex tasks, is very difficult. In the case of judges, the accepted measure of work quality is the number of times a decision is cited in future opinions (Choi et al., 2008). Notice that this is not a measure of whether the decision is correct or not. The job of a judge is to interpret and apply the law to new situations, and most of the time they are not setting initial legal rules. If a judge provides a new interpretation of the law, or a clarification that is particularly helpful, then it will tend to be cited in the future. Hence, citations provide a measure of performance that parties can agree upon, regardless of the normative evaluation of the rule itself. In Ash and MacLeod (2015) we provide evidence that judges do care about citations, and when given more time write longer and more detailed decisions that receive more citations.

Our first task is to show that there is indeed variation in judge quality. After a discussion of the literature and the data, Section 5 provides evidence on the variation in judge ability. Specifically, we find that there is a judge "fixed effect", and that a judge that has high work quality in one period relative to peers, is more likely to have high work quality in the next period. Some of this may be due to how cases are assigned. Fortunately, in our data most states using either random assignment or a fixed rotating assignment of cases to judges. In both cases we find that there is significant temporal auto-correlation in judge work quality as measured by citations. We also measure "work output" using the total number of words and cases as proxies, and find that there are differences between judges in total output.

Having shown that judges vary in their performance, we then ask in Section 6 if there is a systematic decline in performance over the life cycle. Empirically one faces two challenges. First, even though the job does not very much over time, there can be systematic variation in the types of cases and the number of cases over time that are not related to the age of the judge. To deal with this problem, we run regressions at the court-year level to estimate the effect of age within a court. An obvious problem with this approach is that regardless of the age at which the judge is appointed to the court, we would expect performance of new judges to increase with experience, so we add experience controls to our regressions. We find systematic evidence that performance falls with age. In addition, we can see some evidence in mitigation when we control for

the assignment rule. States with random assignment experience a larger negative age effect, but a smaller experience effect. The later result is quite interesting because it is consistent with the hypothesis that when judges are assigned a wide variety of cases it takes longer for them to improve with experience.

An additional useful feature of this context is that states may set mandatory retirement rules for judges. During the period of analysis, many states changed their mandatory retirement rule – normally, from no mandatory retirement to mandatory retirement at age 70, 72 or 75. The reason such policies are implemented is because of the perception that older judges may not be able to carry out their work effectively. The issue of an aging work force is salient in the judiciary, with recent news articles highlighting anecdotal evidence of old age interfering with judge work quality.[4] Regardless of the rule, in some states the chief justice has the right to keep on a retired judge in what the federal courts call "senior status" or "active retirement status". Therefore, we do see in our data some judges working past the official retirement age. When senior status is available, mandatory retirement does not in fact require a judge to retire from work completely. The difference is the default. In the absence of mandatory retirement, a chief judge has to convince his or her colleagues to step down, a discussion that can be difficult. In contrast, under a mandatory retirement rule the control rights switch to the court, which can decide to keep on a judge that it feels is particularly valuable.

We ask in Section 7 how introducing mandatory retirement affects court performance and find that the introduction of a retirement age of 70 or 72 leads to a younger court, and higher quality decisions. The number of cases increases slightly, but there is no effect on total amount of text produced. It is interesting to note that mandatory retirement at age 75 has no effect.

2 Background

2.1 Aging

Desjardins and Warnke (2012) review the large literature on how aging affects cognitive skills. The evidence is generally consistent with the view that while pattern recogni-

[4]See the 2011 *ProPublica* article, "Life Tenure for Federal Judges Raises Issues of Senility, Dementia," available at `https://www.propublica.org/article/life-tenure-for-federal-judges-raises-issues-of-senility-dementia`.

tion and logic skills (fluid intelligence) begin diminishing at a young age, verbal skills (i.e. writing skills) and knowledge (crystallized intelligence) improve into relatively advanced ages.

As we can see in Figure 1a, the ability to run 10 kilometers falls continuously from about age 40, and there is a very steep decrease around age 85. This figure of course has a great deal of selection because individuals who are still able to run a 10k at age 85 are a very selected group! The point is that running speed is an easily measurable activity. Hence, any employment that depends upon running speed can easily build objective performance criteria that need to be satisfied by employees. For example, professional soccer referees must meet minimum speed standards to maintain certification.

Of course, must modern jobs depend upon cognitive rather than physical skills. We have some evidence of decline based upon psychological test scores, as illustrated in Figure 1b. Here we can see a much more continuous linear decline with age starting at age 20. Since the decline is continuous from age 20, the point in time when individuals enter into the workforce, this performance measure obviously does not help to determine when one should stop work. As soon as we begin our working career this graph shows that we are in decline!

Importantly, within-person and between-person studies have found very different age-skill profiles. For example, Small et al. (2011) report a within-person study where episodic/semantic memory demonstrated no decline before the age of 75. The articles reviewed in Lindenberger (2014) suggest that an "intellectually challenging" and "socially engaged" life – such as judging – may itself mitigate cognitive decline.

A smaller literature has investigated aging effects on "wisdom" – that is, reasoning about and resolving social conflicts. Grossmann et al. (2010) show that when thinking about social dilemmas and inter-group conflict, "older people make more use of higher-order reasoning schemes that emphasize the need for multiple perspectives, allow for compromise, and recognize the limits of knowledge." These are all attractive qualities in a judge.[5]

Recent research illustrates the dangers of using cognitive tests to assess performance. Ramscar et al. (2014) find that as people age they have a larger data set in the mind. This in turn leads to slower processing speeds as they search their larger data sets. The is the classic cleverness-wisdom trade-off. Hence, the declines shown in

[5]Relatedly, Steptoe et al. (2015) review the literature on subjective well-being and aging. In most developed countries they find a U-shaped curve, where well-being decreases in early life and increases in late life.

Figure 1b are not necessarily due to a decrease in the performance, but rather a change in search time.

This highlights the importance of developing direct measures of employee performance that are relevant to the ability to carry out tasks assigned to them in the workplace. The challenge is that it is very difficult to find ways to evaluate employees over long periods of time. In particular, in the last century there have been enormous changes in the nature of work. Computers are much more important, and jobs are more complex and include "soft" factors such as the ability to manage employees. In turn, economists have argued that these changes have resulted in an increase in the return to cognitive skills (Autor et al., 2008).

The approach in labor economics is to abstract away from different types of cognitive decline and focus on the age-skill profile. The standard model features a concave relationship between age and productivity, where younger individuals invest in human capital that depreciates over the lifespan (e.g., Blundell and Macurdy, 1999). Empirical papers consistent with this pattern include Levin and Stephan (1991) (academic scientists) and Oster and Hamermesh (1998) (academic economists). In a review of age-performance trends among physicians, Choudhry et al. (2005) conclude that "older physicians possess less factual knowledge, are less likely to adhere to appropriate standards of care, and may also have poorer patient outcomes." De Bot and Makoni (2005) review the literature on variation in language skills over the lifespan.[6]

2.2 Retirement

The key policy relevance of age-performance elasticities is in the design of pension benefits and other age-related policies (Gruber and Wise, 2008). In particular, there is a large and active literature on the economics of retirement choices (Lumsdaine and Mitchell, 1999). For example, Ashenfelter and Card (2002) find that a mandatory retirement age of 70 is binding on many academic faculty, meaning that imposing this requirement significantly reduces the number of older academics.

Figure 2 shows the trend in the retirement age in the last 60 years for U.S. workers, from Munnell (2015). It has remained fairly stable between 63 and 65 for men, and has increased from 55 to 62 for women. Coile (2018) analyzes these trends and notes

[6]In the case of manual workers young than 60 Borsch-Supan and Weiss (2016) find no evidence of a performance decline. This is consistent with our results where most of the performance decline occurs after the age of 65.

the importance of increased in education and pension plan reforms.

The fruitful structural literature on retirement choice, beginning with Gustman and Steinmeier (1986) and Stock and Wise (1990), applies structural estimation methods from the industrial organization literature to predict worker responses to changes in pensions and other retirement incentives. Gustman and Steinmeier (1991) apply these methods to retirement choices for academic faculty, with comparable results to Ashenfelter and Card (2002). In political economy, Diermeier et al. (2005) and Keane and Merlo (2010) derive structural estimates of the parameters underlying retirement choices of U.S. Congressmen.

2.3 Judges

An important reason to study judge performance in this context is that the knowledge and skills relevant to good judging evolve much more slowly than those relevant to good science and good medical care. In Posner's (1995) sample of federal appellate judges, opinion quality (citations per opinion) is maintained into advanced age – into the 80s. Older judges produce fewer opinions, however. Posner argues that this is consistent with the idea that older people tend to be more reflective, less career-oriented, and less progressive. More recent studies are generally consistent with Posner's findings. These include Smyth and Bhattacharya (2003) (Australia High Court), Teitelbaum (2006) (U.S. Supreme Court), and Dimitrova-Grajzl et al. (2012) (Slovenian trial courts).

In a study of retirement among federal appellate judges, Posner (1995) notes that many judges take senior status, which allows for a reduced caseload while retaining full salary. However, only 16 percent of judges take senior status when immediately available. This suggests that there are significant non-pecuniary benefits to remaining a full-time active judge.

The political science literature has focused on how judges may strategically retire to influence the political ideology of their successor (e.g. Nixon and Haskin, 2000). Other papers have used retirement for identification, since judges planning to retire do not face the same retention-related incentives as judges who intend to stay in office (Shepherd, 2009a; Gordon and Huber, 2007; Shepherd, 2009b).

3 Institutional Context

In this project we focus upon judges because it is one of the few professions where the job description has not changed for decades. If we can measure on-the-job performance of judges, this provides a singular opportunity to measure the true productivity of a person doing a complex job over their lifespan. Our setting is state supreme courts.

While state supreme court systems vary from state to state, they also share important characteristics and structures across state lines. The fundamental role of a state judge is to rule on questions of state law (rather than federal law). These questions arise in cases appealed from lower state courts. A case begins when a plaintiff files a lawsuit or a prosecutor indicts a criminal. At trial, facts are litigated and a judge/jury gives a verdict, which the losing party can appeal. If the state has an intermediate appeals court, they will then take the case and may affirm, reverse, or modify the trial verdict. After this intermediate court's decision (or after the trial decision when the state does not have an intermediate appellate court), the ruling can be appealed to the state supreme court. In two states (Texas and Oklahoma), there are separate high courts for criminal and civil matters. In the other forty-eight states, the state supreme court is the last appeal for all matters.

If the supreme court accepts a case for review, the judges will rehear the case at oral argument and review the submitted briefs for legal error. Each judge votes whether to affirm or reverse the lower decision. One of the majority judges writes an opinion explaining the decision. In rare cases, the state supreme court ruling is appealed to the U.S. Supreme Court.

This is the institutional context in which we study judicial incentives. Importantly, the job of a supreme court judge does not change much over the course of the career. A judge in his first year of work has essentially the same task as a judge in his last. Because the nature of the work remains constant throughout a judge's career, we can analyze the effects of aging on work performance over time.

An important issue is that a "judge" is not really a single individual, but a team of individuals that includes clerks and secretarial staff. Judges select the clerks that are working for them, and hence our measures can be seen as composites that depend upon both the judge's legal skill when researching, reasoning, and writing, as well as managerial skill when selecting and directing clerks. As we know from Bloom et al. (2012), management quality varies across firms, and there are systematic relationships between management quality and firm performance.

In our data we cannot directly disentangle managerial skill from legal skill. However, we can ask if there is variation across judges in the same court, and/or across time within-judge. As shown in Ash and MacLeod (2015), there is significant variation in output and quality across judges, even after controlling for institutional and case-level characteristics. In future work, it would be interesting to collect data on clerks and other staff at state supreme courts. For now, we note that our observed effects of age and conditions could include changes in how judges hire or manage their staff.

Moreover, age-related effects may vary depending on judicial characteristics, which depend in part on how they are selected. There are three key judicial selection systems. In partisan elections, judges are selected through a partisan political process with party-specific primaries. In nonpartisan elections, party affiliations are not on the ballot and political parties are not allowed to get involved in the election process. In merit selection, judges are appointed by the governor from a list of nominees chosen by a merit commission. In the results below, we analyze the significant of these processes for the age distribution and for the age-performance profile.

4 Data

The data-set used for the empirical analysis is an extension of that used in Ash and MacLeod (2015) and Ash and MacLeod (2019). It merges information on judge biographies, state-level court institutions, and published judicial opinions. These data allow panel estimates on the effects of judge and court characteristics on performance. For this paper, we have supplemented the dataset in that paper with comprehensive data on judge birthdates and death dates, how judgeships ended, and judge retirement policies.

In particular, we have data on the characteristics of individual judges. A team of research assistants collected these data from a range of sources and built biographies for each judge in the sample. The key sources include state court web sites, judge obituaries, and Marquis Who's Who. Items that were unavailable from these sources were obtained through records requests or interviews of state court administration staff.

4.1 Aging and Retirement

The key data point for this study is the judge's birthday. For most of the judges in our data set, we were able to find their precise birthday. For almost all of the rest, we

were able to find their birth year. The handful of judges for which we could not find birth year information are not included in the analysis.

This section provides a series of descriptive statistics on the age and retirement decisions of state supreme court judges in the absence of mandatory retirement rules.

4.1.1 Summary Statistics on Age and Retirement

Figure 3 shows the age distribution for all state supreme court judges working between 1947 and 1994. Figure 4 shows the distribution of the starting age. Figure 5 shows the distribution of the ending age. Figure 6 shows the distribution of the age of death.

The figures show that there is a wide range of ages of active working state supreme court judges. Judges tend to start in their position late in life (in their 50s) and work late as well (into their 70s). These individuals are relatively healthy, many living into their 80s and 90s.

What do judges do after retirement? Figure 8 shows the trends in these career choices. At the beginning of the sample, few judges took on more work after their judgeship. That has become more common in recent years. If they do take another career, it is usually in private practice as an attorney.

4.1.2 Mandatory Retirement Rules

We are interested in the effects of mandatory retirement policies on judge performance. Tables 1 and 2 provide information on the mandatory retirement rules for state supreme courts in the United States. In 1947 (the first year in our data), 17 states had a mandatory retirement rule. By 1994 (the last year in our data), an additional 14 states had adopted mandatory retirement. We use the variation across states, and within states, to look at how mandatory retirement affects judge performance.

The introduction of mandatory retirement cannot be viewed as an exogenous event. It is a response to the perception that older judges are not performing as expected[7]. There have been some recent moves to repeal mandatory retirement rules, and hence the pressure to change operates in both directions[8]. At the very least, our results will show whether or not the rule in a state affects performance, an estimate that might

[7]See Goldstein (2011) for a discussion of the issue with respect to federal judges.

[8]See `http://ncsc.contentdm.oclc.org/cdm/ref/collection/judicial/id/440` for an update on the state situation.

be viewed as an upper bound if states adopt changes in response to currently under performing aging judges.

4.1.3 Measuring and Normalizing Judge Age

In the empirical analysis we use the age of judge j at year t as a variable in fixed-effects regressions. According to the question we are asking, we adopt a number of specifications for age. For summary statistics we use the absolute age level in years. This works for us because age is an interpretable quantity that is externally valid to other jobs.

On the other hand, the age distribution could be different across courts and (especially) across years. And age-performance gradients could also vary across courts and across time. To check sensitivity to this variation, we normalize by court-year to help compare judges to their colleagues. As a baseline, we follow the recent literature in empirical public finance on inequality and use a rank specification (Chetty et al., 2014; Asher et al., 2017), where the oldest judge is given a 1, the youngest a zero, and each judge uniformly distributed on that interval according to rank. This measure ignores any level differences across the distribution within-court-year. In case those are important, we also use an alternative measure where we standardize the age to mean zero and variance one within court-year.

An issue with court-year normalization is that the resulting measure is primarily be shifted by old judges retiring and new judges joining. Therefore results with absolute age in years are useful to include and compare. We will see that our results are robust to all specifications for age.

4.2 Judge Performance

Our performance measures are constructed from published state supreme court opinions for the years 1947 through 1994, obtained (along with some annotated meta-data) from bloomberglaw.com. The full sample includes 1,024,261 cases. We drop opinions that do not have a named author (per curium decisions), resulting in a sample of 404,928 majority opinions. This is an average 47.2 cases per judge per biennium.

4.2.1 Performance Measures

We focus on two simple metrics for judge performance, work output and work quality. The measures build off of work by previous researchers, in particular Choi et al. (2010) and Epstein et al. (2013). Data collection and processing are described in more detail by Ash and MacLeod (2015) and Ash and MacLeod (2019).

The baseline measure of *work output* is the total number of words written by a judge in opinions during a year on the job. This is a measure of the total volume of opinion-writing work that a judge is responsible for in that year. As alternatives to assess robustness, we look at number of sentences written and number of characters written.

Work quality is measured by the number of citations to a judge's opinions. Judges in a common-law system cite previous cases that are useful to their decision, and therefore citations can be seen as an expert evaluation of peer decision quality (Posner, 2008). More citations means that a case (and the authoring judge) have a stronger influence on the path of the law.

The citations measure is per case (divided by number of cases), so it is workload-adjusted. Citations are annotated as positive, negative, or distinguishing by the data provider, so for the baseline we look only at positive citations. As alternative measures, we use all cites (including negative and distinguishing), discussion cites (where the case was discussed at length by the citing court), and out-of-state cites (only citations in other jurisdictions). Because state supreme court precedents have no bindingness in other states, out-of-state citations serve as an especially strong signal of legal usefulness or influence (Choi et al., 2010). In addition, while older judges might have time to network and influence colleagues in their own court to earn cites, this concern is less pronounced for out-of-state cites.

To check for the importance of caseload changes, we report the number of opinions written as an outcome. To help assess the relative importance of output and quality, we also report a measure of *work impact* – the total number of positive citations to a judge in a year (unadjusted for number of opinions). The appendix includes a range of other outcomes, including measures of caselaw research and number of discretionary opinions written.

4.2.2 Controlling for Case Characteristics

Our measured outcomes are a product of both the type of case and the type of judge. In particular, citations are a joint measure of judge choices and case importance. For example, cases that review the constitutionality of statutes will generally get more citations than summary habeas denials. In addition, judges have some discretion over the types of cases they are chosen to author opinions for. If we want to compare the quality of judges working on the same court at the same time, we need to try to account for these non-judge factors.

Empirically, we use the full range of dummy variables for the area of law of a case, as well as the related industries of a case. These are coded for each case by Bloomberg staff attorneys, and there may be up to three legal areas and three related industrial sectors for any particular case. The case characteristics vector includes a dummy variable for each area and sector, equaling one if the case has been assigned to that area or sector. Because there are so many of these characteristics, including separate covariates for every category would almost saturate the dataset. Instead, we include the first five principal components of this matrix of controls, which explains 65% of the variance of the matrix of case controls.[9]

4.2.3 Normalizing the Measures

A challenging feature of the data is that the distributions of the outcomes are extremely variable across courts, judges, and years. This means that, when making within-court-year comparisons, for example, court-years with higher variance are upweighted in regressions using the raw data. In addition, coefficients on treatments that affect different subsets of states will not be comparable, as the different subsets will have different outcome variance.

The main goal of our analysis is to compare judges to their colleagues, working on the same court at the same time. Because the scale of our measures do not have interest in themselves, we will normalize them to minimize distributional assumptions and to make different measures comparable.

The preferred baseline measure is a within-court-year rank. The judge with the highest measure in a year is given a 1, the lowest a zero, and all other judges uniformly

[9]Using more or fewer components does not change anything. See the appendix for details. In the appendix, we report results with log transformations of the outcomes. Taking logs does not change any of the main results. It tends to strengthen significance in most cases.

distributed on that interval according to rank. Therefore the distribution does not matter, and the measure is robust to outliers. These reasons are why rank specifications have recently gained traction in the empirical work on intergenerational mobility (Chetty et al., 2014; Asher et al., 2017).

On the other hand, when outliers are due to true differences rather than noise or measurement error, the rank measure will not capture that. For example, one could imagine a year where all judges get 1 or 2 cites, except a superstar judge that gets 20 cites. To encode this differential, while still normalizing variances across courts and years, we have a second measure. Specifically, we standardize the outcomes for each judge-year by court-year, to have mean zero and variance one within each court-year subsample.

4.3 Judge assignment to cases

An important issue in our empirical study is how cases are assigned to judges for authorship. Under endogenous assignment, judges of different relative age might receive different types of cases, which in turn could drive differences in quality or other performance measures over the lifespan. On the other hand, if assignment is random, then measured differences across the life cycle are not due to judges selecting into different types of cases.

The official rules for case assignment in state supreme courts are reported in Table 3. At state supreme courts, discretionary assignment by the chief justice (the rule at the U.S. Supreme Court) is the minority rule followed in just 15 states. In 13 states, cases are randomly assigned to authoring judges. In the remaining 22 states, cases are assigned on a rotating system, with cases arbitrarily assigned to judges based on their order on the docket.

There are complex rules across states that affect the rotation. Senior judges have fewer cases. Judges can occasionally recuse themselves. On appeal after remand, the same panel normally reviews a case. There can be exceptions for specialized cases such as those involving the death penalty. (Christensen et al., 2012) show some differences in case assignment characteristics across systems. In their sample, for random assignment and rotating assignment, case characteristics and judge characteristics are only negligibly correlated. We assume these deviations from randomness are independent of our main effects.

We would like to check for conditional independence of case characteristics and

judge age. In terms of the age-quality gradient, we are most concerned with whether older judges are systematically more or less likely to author or sit on important cases. The goal is to assess randomization of state supreme court judges the same way that authors have done in the federal circuit courts. For example, Chen and Sethi (2018) use data from Boyd et al. (2010) and Sunstein et al. (2006), who code 19 case characteristics as determined by the lower court for 415 gender-discrimination Circuit Court cases, and find that case characteristics are uncorrelated with judicial panel composition. Other papers examine whether the sequence of judges assigned to cases in each Circuit Court mimics a random process. They find, for example, that the string of judges assigned to cases is statistically indistinguishable from a random string.

We test for randomness in two ways. First, we see whether the major category of case (constitutional, criminal, civil, administrative) changes over the lifespan under random assignment of judges. Second, we form predictions of case quality (citations) from detailed case type and related industry controls, and see whether that shifts by age.

The analysis is reported in Table 4. It turns out that case categories and case importance are not significantly related to judge age, either in random or in non-random assignment states. With robust (non-clustered) errors, older judges get statistically more criminal cases under discretionary assignment.

We expect judge variation to be smaller in random assignment states than in non-random assignment states, as there would be less specialization. Consistent with this idea, we found that predicting judge authorship from case text was significantly more accurate in discretionary-assignment states than in random-assignment states. Below, we compare the effects of aging and experience separately by case assignment rule.

5 Judges vary in their work performance

An initial question is whether judges vary systematically in our outcome measures. Conceptually, we ask whether there is a judge-specific effect on quality q_{jct} that persists over years of the career. Formally, we estimate

$$q_{jct} = \rho q_{jct-1} + \epsilon_{jct}, \tag{1}$$

a first-order auto-regressive model for judge j in court c at year t. The coefficient ρ summarizes the across-year within-judge correlation in performance measure q_{jct}. In

the preferred specification, q_{jct} is rank-normalized (as discussed in Subsection 4.2.3). An estimate of ρ close to one means that judges tend to keep their within-court performance ranking over time. An estimate close to zero means that the rankings are close to random.

In OLS estimates for (1), the errors are correlated and ρ will be biased. To recover a judge-specific measure, we estimate it with judge fixed effects. But the preferred specification is the Arellano and Bond (1991) GMM estimator, which instruments the lagged outcome with the twice-lagged outcome. This addresses serial correlation in performance.

Next, we would like to know whether judge persistence changes with age. Let a_{jct} be the rank-normalized age for judge j at time t. Our next estimating question is

$$q_{jct} = \rho q_{jct-1} + \beta a_{jct} + \phi a_{jct} q_{jct-q} + \epsilon_{jct}, \tag{2}$$

where we have included age as well as the interaction between age and the lagged outcome. This can again be estimated using OLS with judge fixed effects and Arellano-Bond.

Estimates for (1) and (2) are reported in Tables 5 (Work Quality) and 6 (Work Output). The Arellano-Bond estimates (Column 3) suggest that judge ranks are correlated by 13% year-to-year for quality and 27% for output. According to the age-interaction regressions (Column 6), the estimate conceals significant heterogeneity by age; older judges are more likely to change in rank over time.

6 Judge performance varies by age

6.1 Main Results

Next we provide a descriptive analysis of how differences in ages are related to differences in performance. The empirical strategy for examining the effects of aging on judicial behavior is to exploit differences in performance between judges working in the same court at the same time. The linear model for performance variable y_{jct} for judge j working in court c at year t is

$$y_{jct} = \alpha_{ct} + \gamma a_{jct} + X'_{jct}\beta + \epsilon_{jct} \tag{3}$$

where α_{ct} includes court-year fixed effects anda_{jct} is the age for judge j at t, and X_{jct} includes additional fixed effects and controls. In the preferred specification, both y_{jct} and a_{jct} are rank-normalized. We also report results with standardization to mean zero and variance one within court-year. Standard errors are clustered by state.[10]

The main source of bias we are interested in comes from the time-varying changes in the court work environment which may be correlated with age. To deal with this possibility, we include a full set of court-year fixed effects. Therefore any estimated coefficients are also relative to the court average in each year. This means the regressions effectively compare judges sitting on the same court, working at the same time, but who are of different ages.

The matrix X_{jct} includes additional fixed effects and controls for specification checks. In the most conservative specification, we include the following. First, we have cohort fixed effects, which are dummies for decade that the judge started on the court. This is meant to rule out mechanical variation due to cohort differences across the time period. In the same vein, we have court-specific linear trends in judge starting cohort: formally, judge starting-year interacted with court fixed effect. This allows for judges in different states to have a different confounding trend in starting year and performance. Next, we have fixed effects for experience rank; that is, an indicator for being the most experienced judge on the court, the second-most-experienced judge, and so on, up until the ninth-most-experienced (the maximum number of judges on a court at one time). These fixed effects flexibly control for differences in judge experience levels, which naturally covary with age. Finally, we include a set of dimension-reduced controls for case characteristics: the first five principal components of the matrix of covariates for legal topic and related industries. These controls allow us to rule out changes in performance due to changes in the types of cases that judges handle.

The main results are illustrated graphically in Figure 9 (Work Quality) and Figure 10 (Work Output). In the latter figure, we see that there is so change in output over the life cycle. In the former figure, we see a large decrease in work quality over the life cycle. We can see this in the raw data (Figure 9 top left), when standardized by court-year (top right), and using ranks (bottom left). The effect is the same when including controls for cohort, experience, and case characteristics (bottom right).

The regression estimates from Equation (3) are reported in Tables 7 and 8. There is no effect on output, but a significant negative effect on quality. A one standard

[10]Results are similar with two-way clustering by state and year.

deviation increase in age is associated with a 0.1 standard deviation decrease in positive citations per opinion. This effect holds for alternative quality measures: out-of-state cites, text-predicted cites, direction quotations by future judges, negative (rather than positive) citations, and total citations (rather than citations per opinion).

The regressions in this section include all states and years. In the appendix we include results separately by mandatory and voluntary retirement, and separate our results before/after 1970 (when WestLaw and LexisNexis were introduced). The main results on quality hold across subsamples. We also take a within-judge approach, using judge fixed effects, and find the same result. We also find the same results when dropping the first and last year, which might have partial workloads.

In the appendix we analyze a number of alternative outcomes (see Appendix Figures A.1, A.2, and A.3). There are no effect on number of majority opinions written. Judges write fewer concurrences as they age, but not fewer dissents. Judges tend to affirm (rather than reverse) more cases as they age. There is not much differences by age in the time between assignment and publication of opinions. The number of previous cases cited (table of cases length) decreases; note that this is inconsistent with greater delegation to court clerks among older judges. There is no effect on the rate the judges are overruled by the U.S. Supreme Court (although this is a very sparse outcome).

We also show significant differences in writing style by age (see Appendix Figure A.3). Older judges use shorter words and longer sentences than their colleagues. Vocabulary size (unique words used) does not change much. A text-based entropy measure increases with age. Finally, content words (nouns and adjectives) are decreasing in frequency with age, while functional words (conjunctions and prepositions) are increasing in frequency with age.

6.2 Case Allocation Rules Influence the Aging-Performance Trend

As discussed earlier, about two-thirds of our states have random or rotating assignment of judges to cases, while one-third of states have discretionary assignment. We might expect different effects of age on performance between these two groups. Under discretionary assignment, the types of cases can be calibrated by judge age to allow for specialization, or to adjust workloads according to ability. Here we analyze that possibility by estimating aging effects separately by case allocation rule.

The estimation approach is the same as that summarized in Equation (3). There are two differences. First, we zero in on how age and experience are related, by adding

an explanatory variable for years of experience on the court. We also look at the interaction between age and experience. In addition, we subset by random vs non-random states. Alternatively, we interact the age and experience terms with indicator variables for the case allocation method. As before, the results are not sensitive to using unadjusted values for age and quality, standardized by court-year, or normalizing by rank within court-year.

The first results on experience and assignment method are reported in Table 9. Columns 1 through 3 show that both age and experience have a negative effect on case quality, with a marginally positive interaction (Columns 2-3). Columns 4/5 show the results limited to discretionary assignment states, while Columns 6/7 subset to random assignment states. In both subsets, there is a negative and significant effect of age on quality. The coefficients are much larger under discretionary assignment. The experience effects are different: under discretionary assignment, there is an insignificant positive effect of experience; under random assignment, there is a significant negative effect of experience.

Are the differences between the discretionary and random states significant? We bolster this view in Table 10, which use all the data but interacts the age and experience effects on quality with an indicator variable for being in a random assignment state. Columns 1 and 2 use standardized data within court year, while Columns 3 and 4 use the rank specification. One can see that random-assignment states have a significantly *less negative* effect of aging, but a significantly *more negative* effect of experience. Equivalently, under discretionary assignment quality decreases more with age, but decreases less with experience.

To recap, case quality falls with both age and experience. The aging effect is reduced under random assignment, while the experience effect is increased under random assignment. This suggests that under discretionary assignment, cases can be matched better to more experienced judges, which mitigates the negative performance trend. However, aging effects on quality cannot be mitigated through case matching.

Consistent with this idea, in Columns 5 and 6 of Table 10, the outcome is case importance, generated from predicting case quality using case characteristics (legal topic and related industries). As already shown above, this measure does not change with age. What we add here is that there is no differential effect of age on case importance under random assignment. What we do see, though, is a statistically significant positive effect of experience on case importance under discretionary assignment. More

experienced judges get more important cases. However this effect is zero under random assignment.

7 Mandatory Retirement Rules Affect Judge Retirement Choices and Court Performance

The next analysis is to look at a set of reforms which implemented a mandatory retirement age for state supreme court judges. Identification comes from discrete changes in the rules for mandatory retirement. Sixteen states introduced a mandatory retirement age during the time period of our data (see Table 2). First, we look at how it affects the retirement decision. Second, we look at the effect on performance.

7.1 Effect of Mandatory Retirement Age on the Age Distribution

We begin with some summary statistics on these policies. First, Figure 11 illustrates the impact of mandatory retirement policies on the exit decision. This graph shows the probability of retirement at any given age, separately by the mandatory retirement rule. The blue line, with no mandatory retirement, is relatively smooth, peaking in the early 70s. The red line, with mandatory retirement at age 70, shows big increases for ages 69 and 70. We see corresponding jumps for retirement at 72 (green line) and 75 (yellow line). We see that these rules are not perfectly enforced, as some judges stay on past the mandatory retirement age due to grandfather clauses or senior judge status.

The hazard plot and age distribution in Figure 12 illustrate the same story. At any given age, the probability of exit is higher for judges under mandatory retirement. This is reflected in a shift to the left of the age distribution under mandatory retirement. Table 11 provides ordered logit estimates for how the retirement rules affect the judge retirement decision. Figure 13 provides additional graphical evidence, showing that the distribution of retirement ages changes a lot based on the retirement age.[11]

[11]Figure **??** looks at how judge retirement is related to judge longevity, separately for mandatory retirement (left panel) and voluntary retirement (right panel). The figure shows that with voluntary retirement, judges are much more likely to die within a year of leaving office. This supports the idea that mandatory retirement is an impactful policy, as judges are more likely to stay in their jobs until death otherwise. On the other hand, there is still a relatively high chance of death in the first year out

The ordered logit table and the histogram show that the 70/72 retirement age rules make more of a difference than the 75 retirement age rule. Going forward, we focus on the 70/72 reform in an event study framework. First of all, Figure 14 illustrates the short-term impact of the reform on judge age. Immediately following the reform, the average age of the judges falls.

7.2 Mandatory Retirement: Effects on Court Performance

This section looks at the diff-in-diff effect of introducing a mandatory retirement age. The regression framework is a standard differences-in-differences approach based on Bertrand et al. (2004). To control for time-invariant court characteristics that may be correlated with the retention system in various states, we include court fixed effects. To control for national trends in performance, we include year fixed effects. To control for pre-existing state trends in performance that may be confounded with the reforms, we include state-specific linear trends.

As in Ash and MacLeod (2015), we measure effects in a ten-year window around the reforms. The regressions include an indicator equaling one for the baseline time window of ten years before and ten years after a change to the retention system. The treatment variable is a dummy for the ten years after the change. Thus, with the inclusion of the court fixed effects, the estimates can be interpreted as the average difference in within-court performance for the ten years after the policy change relative to the ten years before the policy change.[12]

For additional robustness, we add state-specific treatment windows. We also add a number of additional time-varying court rules (election system, number of judges, and government expenditures on the judiciary) from Ash and MacLeod (2019). Finally, we include controls for judge experience levels.

Formally, we estimate performance y_{ist} for judge i in court s during year t as

$$y_{ist} = \alpha_{ist} + \rho R_{st} + X'_{ist}\beta + \epsilon_{ist} \tag{4}$$

where α_{ist} includes the fixed effects and treatment windows, R_{st} is a dummy variable for the ten years after the rule change (with ρ measuring the corresponding causal

of office under mandatory retirement (left panel), which may hint at a causal impact of retirement on mortality (as found in Sullivan and von Wachter, 2009). This is an important area for future work.

[12]In the appendix we include a table where we use a six-year window, fourteen-year window, or no window (all years). Our main effect on log citations is robust to all these specifications.

effect of interest), and X_{ist} includes other state and judge covariates. Standard errors are clustered by state.

In this section, we use log outcome variables without normalization. Therefore the coefficients can be interpreted as proportional changes in the outcome. Because we are interested in the performance of the court, rather than output or quality, our main outcome is log of the total positive citations to a judge during the year. This provides a measure of the influence of the court on the law.

The results for the regression are reported in Table 12. Across a range of specifications, there is a positive and significant effect of introducing a mandatory retirement age on a court's influence, as measured by citations. In the preferred specification (Column 3), we estimate a 27% increase in citations due to the reform. As seen in Column 4, about a third of the effect is explained by case type. There is a smaller, yet still marginally significant effect, with judge fixed effects (Column 5), showing that the effect is mainly driven by older judges being replaced.

We look at additional outcomes in Table 13. We report effects on log number of opinions published, work output (log number of words written, and work quality (log citations per opinion). As in Choi et al. (2010), we also report results for log out-of-state citations, which better reflect the persuasive influence of a court's precedents. We see that there are positive effects of these reforms on the number of opinions, work quality, and total out-of-state citations. In the appendix we show that there is a change in case types: after the reform, the courts take on more criminal cases and fewer administrative cases.

Figure 15 shows the main results in an event-study framework. To make these graphs, we residualized log positive citations (left panel) and log out-of-state citations (right panel) on all the fixed effects from Column 3 from Table 12. We then plotted that value by year before and after the reform. For both outcomes, we can see a clear break and increase afterward.

8 Conclusions

The goal of this paper has been to measure the effects of aging on judicial behavior. Given that judges have low-powered incentives that do not explicitly link pay to performance, these factors likely have a significant impact on judge behavior. We find that physical aging is associated with a reduction in work quality over the lifespan. Manda-

tory retirement rules increase the performance of courts as a whole. The assignment of cases to judges matters for both results.

These results will be useful to policymakers seeking to design better retirement policies for judges and other high-skill jobs. In particular, the results are useful in an era where an aging workforce is resulting in large structural changes to the economy (Acemoglu and Restrepo, 2017).

An important open question is the role of health in expert decision-making. While judges are working longer and getting more citations, it is still an open question how much this is due to differences in health. More generally, an open question is whether our changes in quality over the lifespan are due to cognitive effects of aging, changes in reputational incentives, or some other aging mechanism. In future work it would be interesting to know how aging affects other aspects of judicial decision-making.

References

Abraham, K. G. and Farber, H. S. (1987). Job duration, seniority, and earnings. *American Economic Review*, 77(3):278–297.

Acemoglu, D. and Restrepo, P. (2017). Secular stagnation? the effect of aging on economic growth in the age of automation. *NBER*.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2):277–297.

Ash, E. and MacLeod, W. B. (2015). Intrinsic motivation in public service: Theory and evidence from state supreme courts. *Journal of Law and Economics*, 58(4).

Ash, E. and MacLeod, W. B. (2019). The performance of elected officials: Evidence from state supreme courts. Working Paper 22071, National Bureau of Economic Research.

Ashenfelter, O. and Card, D. (2002). Did the elimination of mandatory retirement affect faculty retirement? *American Economic Review*, 92(4):957–980.

Asher, S., Novosad, P., and Rafkin, C. (2017). Estimating intergenerational mobility with coarse data: A nonparametric approach. Technical report, World Bank.

Autor, D. H., Katz, L. F., and Kearney, M. S. (2008). Trends in us wage inequality: Revising the revisionists. *Review Of Economics and Statistics*, 90(2):300–323.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249–275.

Bloom, N., Genakos, C., Sadun, R., and Reenen, J. V. (2012). Management practices across firms and countries. NBER Working Papers 17850, National Bureau of Economic Research, Inc.

Blundell, R. and Macurdy, T. (1999). Chapter 27 labor supply: A review of alternative approaches. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3, Part A, pages 1559–1695. Elsevier.

Borsch-Supan, A. and Weiss, M. (2016). Productivity and age: Evidence from work teams at the assembly line. *Journal of the Economics of Ageing*, 7:30–42.

Boyd, C., Epstein, L., and Martin, A. D. (2010). Untangling the causal effects of sex on judging. *American Journal of Political Science*, 54(2):389–411.

Chen, D. L. and Sethi, J. (2018). Insiders and outsiders: Does forbidding sexual harassment exacerbate gender inequality? Working Paper 16-687, Toulouse School of Econmics.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics*, 129(4):1553–1623.

Choi, S. J., Gulati, G. M., and Posner, E. A. (2008). Judicial evaluations and information forcing: Ranking state high courts and their judges. *Duke LJ*, 58:1313.

Choi, S. J., Gulati, G. M., and Posner, E. A. (2010). Professionals or politicians: The uncertain empirical case for an elected rather than appointed judiciary. *Journal of Law, Economics, and Organization*, 26(2):290.

Choudhry, N. K., Fletcher, R. H., and Soumerai, S. B. (2005). Systematic review: the relationship between clinical experience and quality of health care. *Annals of Internal medicine*, 142(4):260–273.

Christensen, R. K., Szmer, J., and Stritch, J. M. (2012). Race and gender bias in three administrative contexts: Impact on work assignments in state supreme courts. *Journal of Public Administration Research and Theory*.

Coile, C. (2018). Working longer in the us: Trends and explanations. Technical report, National Bureau of Economic Research.

De Bot, K. and Makoni, S. (2005). *Language and aging in multilingual contexts*, volume 53. Multilingual Matters.

Desjardins, R. and Warnke, A. J. (2012). Ageing and skills: A review and analysis of skill gain and skill loss over the lifespan and over time. Technical report, ZBW-German National Library of Economics.

Diermeier, D., Keane, M., and Merlo, A. (2005). A political economy model of congressional careers. *The American Economic Review*, 95(1):pp. 347–373.

Dimitrova-Grajzl, V., Grajzl, P., Zajc, K., and Sustersic, J. (2012). Judicial incentives and performance at lower courts: Evidence from slovenian judge-level data. *Review of Law & Economics*, 8(1):215–252.

Epstein, L., Landes, W. M., and Posner, R. A. (2013). *The Behavior of Federal Judges*. Harvard University Press.

Goldstein, J. (2011). Life tenure for federal judges raises issues of senility, dementia. *ProPublica*.

Gordon, S. and Huber, G. (2007). The effect of electoral competitiveness on incumbent behavior. *Quarterly Journal of Political Science*, 2(2):107–138.

Grossmann, I., Na, J., Varnum, M. E., Park, D. C., Kitayama, S., and Nisbett, R. E. (2010). Reasoning about social conflicts improves into old age. *Proceedings of the National Academy of Sciences*, 107(16):7246–7250.

Gruber, J. and Wise, D. A. (2008). *Social security and retirement around the world*. University of Chicago Press.

Gustman, A. L. and Steinmeier, T. L. (1986). A structural retirement model. *Econometrica*, 54(3):pp. 555–584.

Gustman, A. L. and Steinmeier, T. L. (1991). The Effects of Pensions and Retirement Policies on Retirement in Higher Education. *American Economic Review*, 81(2):111–15.

Keane, M. P. and Merlo, A. (2010). Money, political ambition, and the career decisions of politicians. *American Economic Journal: Microeconomics*, 2(3):pp. 186–215.

Landes, W. M. and Posner, R. A. (2009). Rational judical behavior: A statistical study. *The Journal of Legal Analysis*, 1:775–831.

Levin, S. G. and Stephan, P. E. (1991). Research productivity over the life cycle: Evidence for academic scientists. *The American Economic Review*, 81(1):pp. 114–132.

Lindenberger, U. (2014). Human cognition again: Corriger la fortune. *Science*, 346(6209):572–578.

Lumsdaine, R. L. and Mitchell, O. S. (1999). Chapter 49 new developments in the economic analysis of retirement. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3, Part C, pages 3261 – 3307. Elsevier.

MacLeod, W. B. (2003). Optimal contracting with subjective evaluation. *American Economic Review*, 93(1):216–240.

Medoff, J. L. and Abraham, K. G. (1980). Experience, performance and earnings. *Quarterly Journal of Economics*, 95:703–36.

Munnell, A. H. (2015). The average retirement age–an update. *Center for retirement research*, 1920:1960–1980.

Nixon, D. C. and Haskin, J. D. (2000). Judicial retirement strategies the judge's role in influencing party control of the appellate courts. *American Politics Research*, 28(4):458–489.

Oster, S. M. and Hamermesh, D. S. (1998). Aging and productivity among economists. *Review of Economics and Statistics*, 80(1):154–156.

Posner, R. (2008). *How Judges Think*. Harvard University Press.

Posner, R. A. (1995). *Aging and old age*. University of Chicago Press.

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., and Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science*, 6:5–42.

Shepherd, J. M. (2009a). Are appointed judges strategic too? *Duke Law Journal*, 58(7):1589–1626.

Shepherd, J. M. (2009b). The influence of retention politics on judges' voting. *The Journal of Legal Studies*, 38(1):169–206.

Small, B. J., Dixon, R. A., and McArdle, J. J. (2011). Tracking cognition–health changes from 55 to 95 years of age. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 66(suppl 1):i153–i161.

Smyth, R. and Bhattacharya, M. (2003). How fast do old judges slow down?: A life cycle study of aging and productivity in the federal court of australia. *International Review of Law and Economics*, 23(2):141–164.

Steptoe, A., Deaton, A., and Stone, A. A. (2015). Subjective wellbeing, health, and ageing. *The Lancet*, 385(9968):640–648.

Stock, J. H. and Wise, D. A. (1990). Pensions, the option value of work, and retirement. *Econometrica*, 58(5):pp. 1151–1180.

Sullivan, D. and von Wachter, T. (2009). Job displacement and mortality: An analysis using administrative data. *Quarterly Journal Of Economics*, 124(3):1265–1306.

Sunstein, C. R., Schkade, D., Ellman, L. M., and Sawicki, A. (2006). *Are Judges Political?: An Empirical Analysis of the Federal Judiciary.* Brookings Institution Press.

Teitelbaum, J. C. (2006). Age and tenure of the justices and productivity of the u.s. supreme court: Are term limits necessary? *Florida State University Law Review*, 34:161–182.

A Appendix

A.1 The Age Discrimination in Employment Act of 1967-Sec 621, section 2

The Congress hereby finds and declares that

1. in the face of rising productivity and affluence, older workers find themselves disadvantaged in their efforts to retain employment, and especially to regain employment when displaced from jobs;

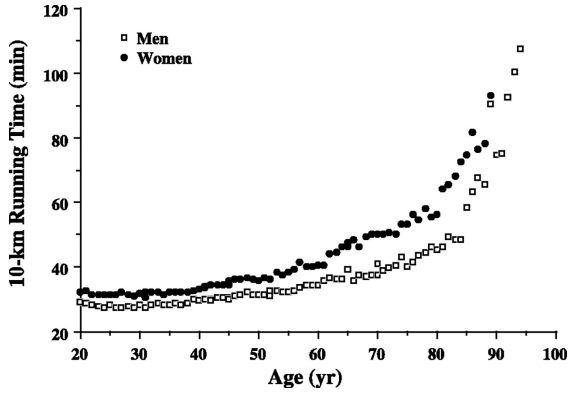
   (a) the setting of arbitrary age limits regardless of potential for job performance has become a common practice, and certain otherwise desirable practices may work to the disadvantage of older persons;

(b) the incidence of unemployment, especially long-term unemployment with resultant deterioration of skill, morale, and employer acceptability is, relative to the younger ages, high among older workers; their numbers are great and growing; and their employment problems grave;

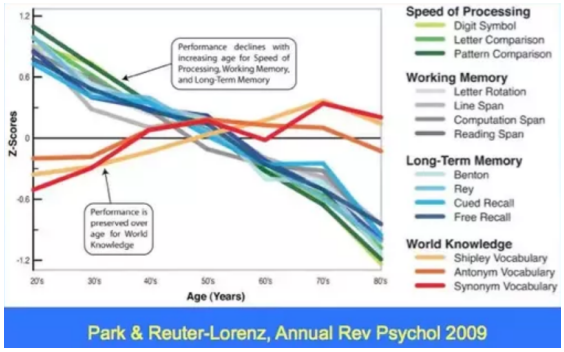
(c) the existence in industries affecting commerce, of arbitrary discrimination in employment because of age, burdens commerce and the free flow of goods in commerce.

(d) It is therefore the purpose of this chapter to promote employment of older persons based on their ability rather than age; to prohibit arbitrary age discrimination in employment; to help employers and workers find ways of meeting problems arising from the impact of age on employment.

Figure 1: Performance vs Age, Physical and Cognitive Tasks

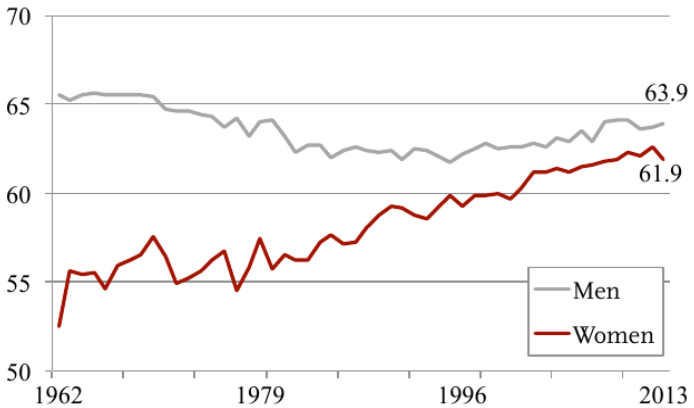


(a) 10k Running Time



(b) Cognitive Performance

Figure 2: Average Retirement Age for U.S. Workers, 1962-2013



Average retirement age by gender for U.S. workers, computed from CPS by **?**.

Figure 3: Age Distribution of Working State Supreme Court Judges

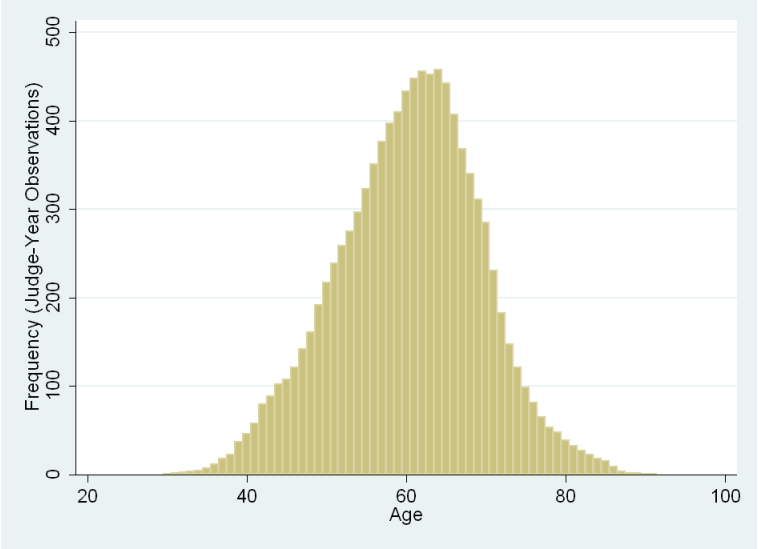


Figure 4: Starting-Age Distribution of State Supreme Court Judges

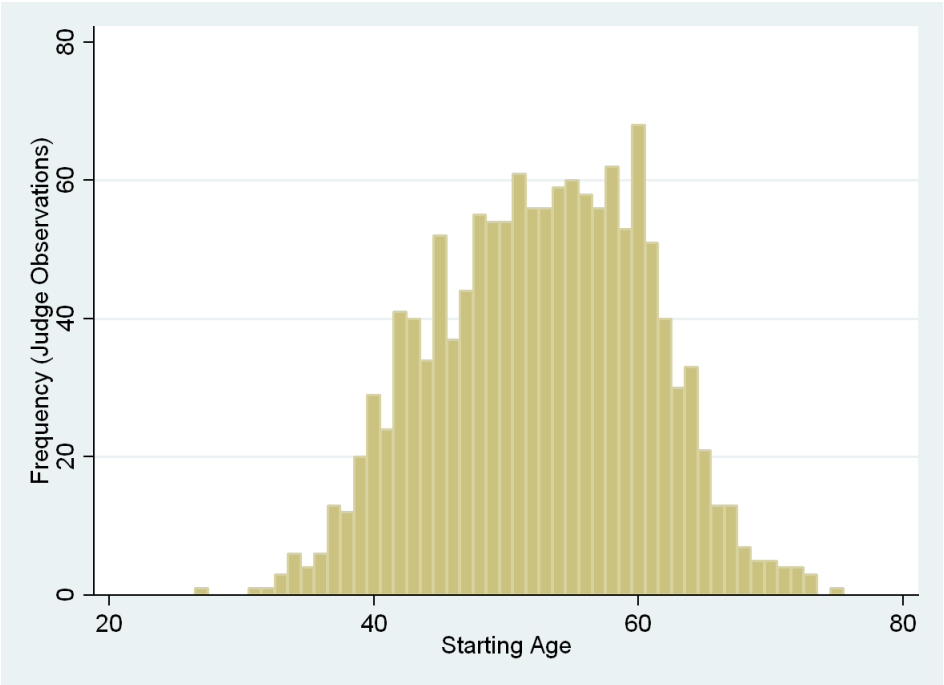


Figure 5: Ending-Age Distribution of State Supreme Court Judges

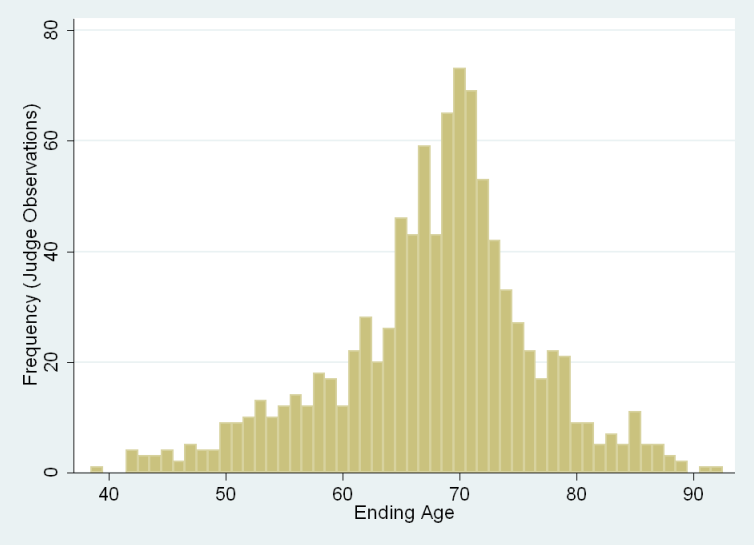


Figure 6: Age-of-Death Distribution of State Supreme Court Judges

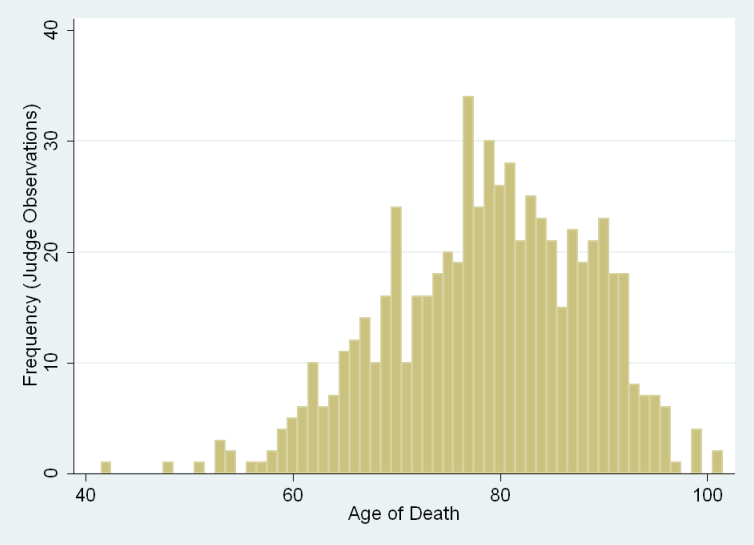
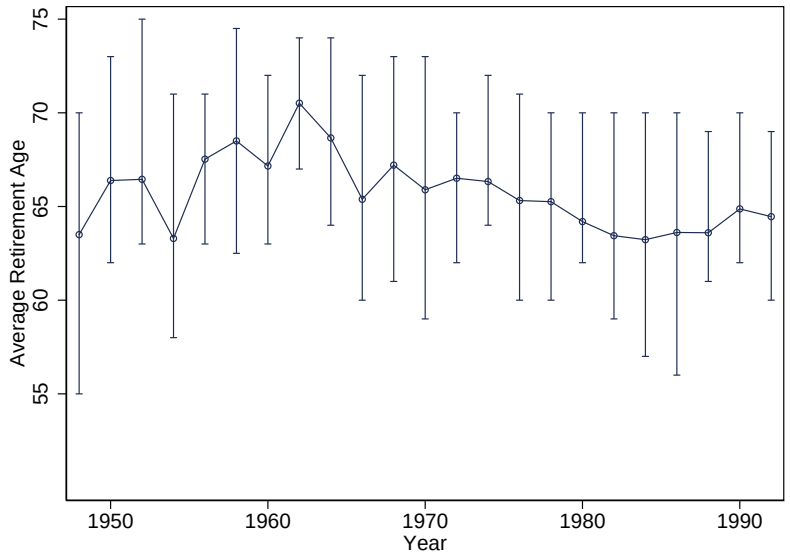
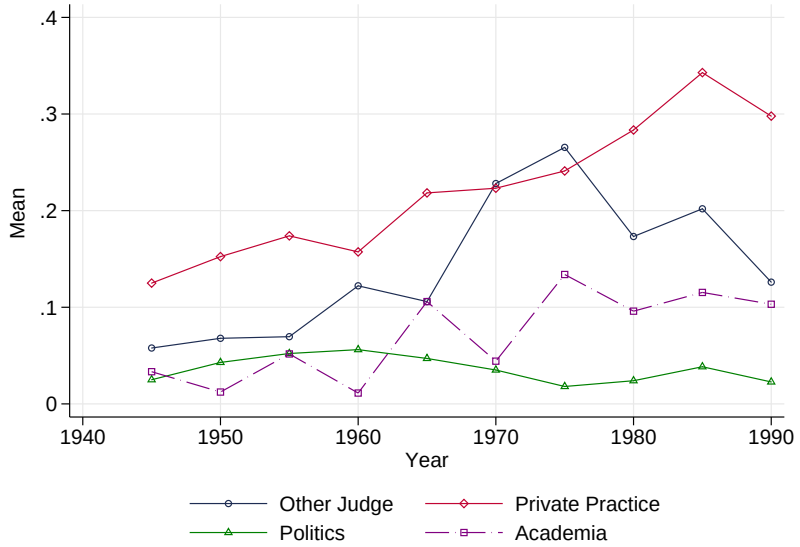


Figure 7: Average Retirement Age Over Time



Average retirement age of state supreme court judges, by year. Error spikes give 25th and 75th percentiles.

Figure 8: Post-Judgeship Careers



Proportion of judges with documented careers after their state supreme court judgeship, including other judgeship, private practice, politics, and academia. Plotted by five-year bins.

Table 1: Judge Retirement Rules By State in 1947

Retirement Rule	List of States
No Mandatory Retirement	AR, CA, DE, GA, ID, KY, ME, MS, MT, ND, NE, NM, NV, OK, RI, TN, WI, WV, VT
Retirement at Age 70	AK, HI, LA, MD, MA, MI,MO, NH, NJ, NY, OH
Retirement at Age 72	NC, SC
Retirement at Age 75	IL, IN, TX, UT

Initial retirement rules by state. Vermont (VT) has mandatory retirement at age 90; we classify it as no mandatory retirement since there are just 2 judges in our entire sample (not in Vermont) who live that long.

Table 2: Retirement Rule Changes, 1948-1993

Mandatory Retirement Age		List of States (with Year Enacted)
Before	After	
None	70	AL (1973), AZ (1992), CT (1974), FL (1972), MN (1973), PA (1968), VA (1970), WI (1955), WY (1972)
None	72	CO (1962), IA (1965), WA (1952)
None	75	KS (1993), OR (1960)
70	None	WI (1984)

Table 3: Case Assignment Rules on State Supreme Courts

Discretionary	Random	Rotating
Arizona	Idaho	Alaska
California	Louisiana	Alabama
Colorado	Mississippi	Arkansas
Connecticut	New Hampshire	Florida
Delaware	New York	Georgia
Hawaii	Ohio	Iowa
Indiana	South Dakota	Illinois
Kansas	Tennessee	Maine
Kentucky	Texas	Minnesota
Massachusetts	Virginia	Missouri
Maryland	Washington	Montana
New Jersey	Wisconsin	North Carolina
Oregon		North Dakota
Pennsylvania		Nebraska
Wyoming		New Mexico
		Nevada
		Oklahoma
		Rhode Island
		South Carolina
		Utah
		Vermont
		West Virginia

List of states by rules for case assignment in state supreme courts. Rules collected by **?**.

Table 4: Case Type and Importance by Judge Age and Case Allocation Rule

	(1) Crim Cases	(2) Civil Cases	(3) Admin Cases	(4) Con Law Cases	(5) Pred. Cites
Age × Random	0.00427	-0.00435	-0.0164	-0.0196	-0.00127
	(0.00845)	(0.00700)	(0.0115)	(0.0129)	(0.00188)
Age × Not Rand	0.0265	-0.0198	-0.00161	-0.0131	-0.0000133
	(0.0209)	(0.0230)	(0.0176)	(0.0194)	(0.00229)
Court-Year FE	X	X	X	X	X
N	13643	13643	13607	13632	13599
adj. R-sq	0.140	0.209	-0.062	-0.042	0.397

Observation is a judge working in a year. "Random" is an indicator for random-assignment states, while "Not Rand" means discretionary assignment. Age is standardized to mean zero and variance one within court-year (results are the same with unadjusted age or rank-normalized age). "Crim Cases" means proportion of cases on criminal law in a year (and respectively for civil cases, administrative cases, and constitutional law cases). "Pred. Cites" means the predicted case quality computed from an OLS regression with case characteristics (legal area and related industries). Standard errors clustered by state in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Table 5: Persistence in Judge Quality

	(1)	(2)	(3)	(4)	(5)	(6)
	Effect on Case Quality Rank					
	Panel OLS				Arellano-Bond	
Lagged Case Quality Rank	0.358	0.0272	0.361	0.0541	0.132	0.800
	(0.018)	(0.012)	(0.027)	(0.020)	(0.015)	(0.054)
Age Rank			-0.0849	-0.0623		0.534
			(0.020)	(0.030)		(0.054)
Lagged Case Quality Rank			-0.0401	-0.0605		-1.618
× Age Rank			(0.033)	(0.027)		(0.075)
N	13296	13163	12239	12062	11775	10781
Model	OLS	OLS	OLS	OLS	AB	AB
Year FE's	X	X	X	X		
Judge FE's		X		X		

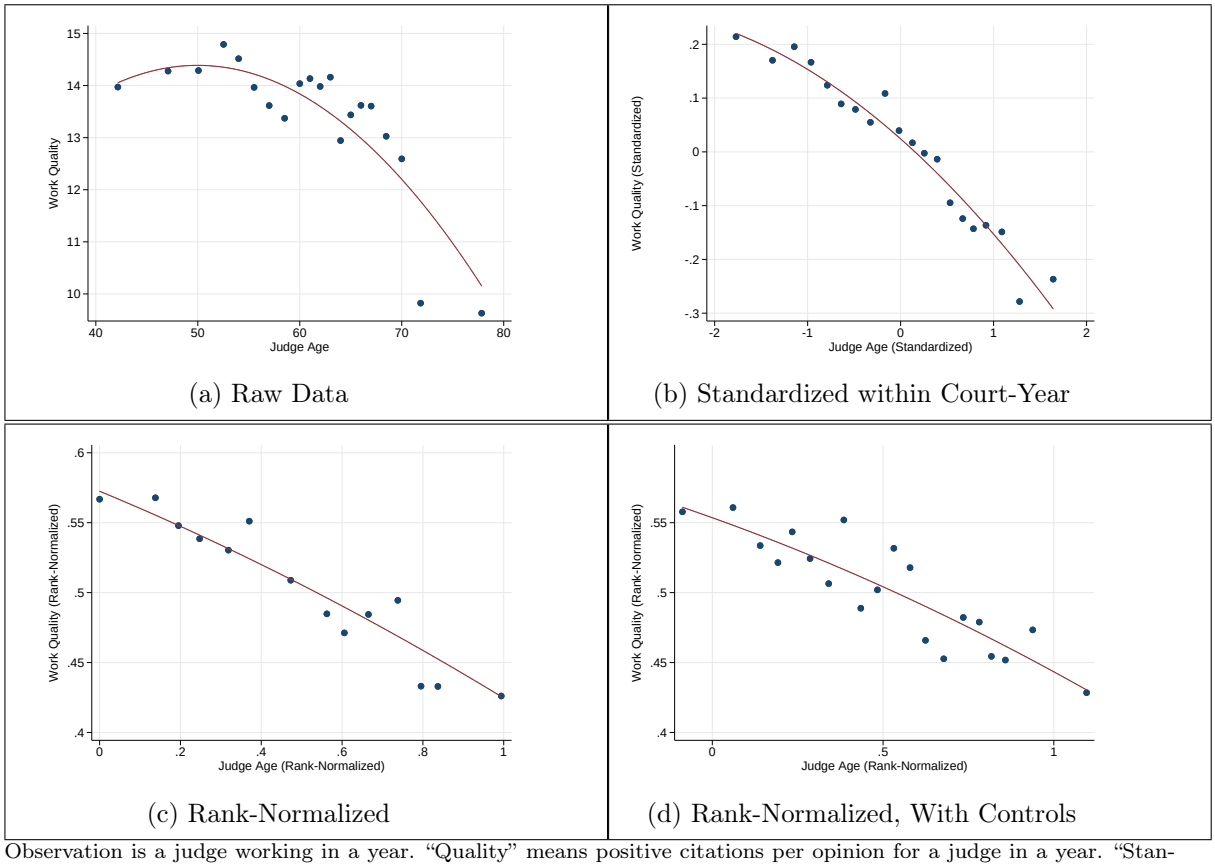
Observation is a judge working in a year. "Case Quality" means citations per opinion in a year; "rank" means judges are uniformly distributed between zero and one based on rank within court-year (0 is lowest, 1 is highest). Estimates computed with Panel OLS and Arellano-Bond, as indicated.

Table 6: Persistence in Judge Output

	(1)	(2)	(3)	(4)	(5)	(6)
	Effect on Output Rank					
	Panel OLS				Arellano-Bond	
Lagged Output Rank	0.469	0.109	0.409	0.0975	0.237	0.945
	(0.018)	(0.012)	(0.027)	(0.020)	(0.015)	(0.054)
Age Rank			-0.115	-0.0535		0.605
			(0.020)	(0.032)		(0.069)
Lagged Output Rank			0.0918	-0.00259		-1.740
× Age Rank			(0.033)	(0.034)		(0.099)
N	13296	13163	12239	12062	11775	10781
Model	OLS	OLS	OLS	OLS	AB	AB
Year FE's	X	X	X	X		
Judge FE's		X		X		

Observation is a judge working in a year. "Case Output" means total words written per year; "rank" means judges are uniformly distributed between zero and one based on rank within court-year (0 is lowest, 1 is highest). Estimates computed with Panel OLS and Arellano-Bond, as indicated.

Figure 9: Judge Age and Work Quality



(a) Raw Data

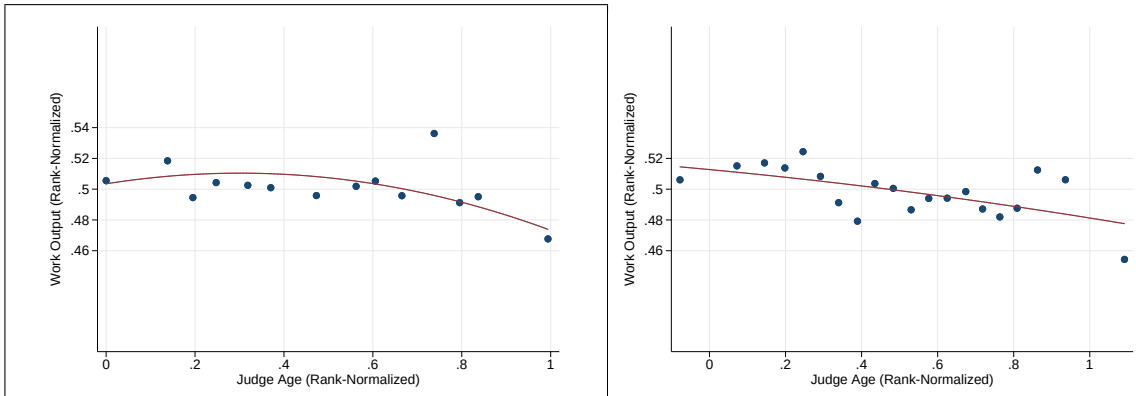
(b) Standardized within Court-Year

(c) Rank-Normalized

(d) Rank-Normalized, With Controls

Observation is a judge working in a year. "Quality" means positive citations per opinion for a judge in a year. "Standardized" means centered by court-year and divided by the court-year standard deviation. "Rank" means judges are uniformly distributed between zero and one based on rank within court-year (0 is lowest, 1 is highest). "With Controls" means court-year fixed effects, fixed effects for decade that the judge started on the court, fixed effect for the ranking (one, two, three, etc) of years of experience by judge, judge starting-year interacted with court fixed effect, and the first five principal components of the matrix of controls for legal topic and related industries. Standard errors clustered by state in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Figure 10: Judge Age and Work Output



(a) Rank-Normalized

(b) Rank-Normalized, With Controls

Observation is a judge working in a year. "Output" means number of words written by a judge in a year. "Standardized" means centered by court-year and divided by the court-year standard deviation. "Rank" means judges are uniformly distributed between zero and one based on rank within court-year (0 is lowest, 1 is highest). "With Controls" means court-year fixed effects, fixed effects for decade that the judge started on the court, fixed effect for the ranking (one, two, three, etc) of years of experience by judge, judge starting-year interacted with court fixed effect, and the first five principal components of the matrix of controls for legal topic and related industries. Standard errors clustered by state in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Table 7: Judge Age and Judge Work Quality

	(1)	(2)	(3)	(4)	(5)
	Quality (Unadjusted)	Quality (Standardized)		Quality (Rank)	
Age (Unadjusted)	-0.0605+				
	(0.0329)				
Age (Standardized)		-0.153**	-0.100**		
		(0.0177)	(0.0206)		
Age (Rank)				-0.148**	-0.102**
				(0.0187)	(0.0222)
Year FE	X	X	X	X	X
Court-Year FE		X	X	X	X
Cohort FE			X		X
Experience Rank FE			X		X
Court × Cohort Trends			X		X
Case Controls			X		X
N	13727	13637	13635	13655	13653
R-sq	0.084	0.058	0.090	0.036	0.071

Observation is a judge working in a year. "Quality" means positive citations per opinion for a judge in a year. "Standardized" means centered by court-year and divided by the court-year standard deviation. "Rank" means judges are uniformly distributed between zero and one based on rank within court-year (0 is lowest, 1 is highest). Cohort FE means fixed effect for decade that the judge started on the court. Experience Rank FE means fixed effect for the ranking (one, two, three, etc) of years of experience by judge. Court × Cohort Trends means judge starting-year interacted with court fixed effect. Case controls means the first five principal components of the matrix of controls for legal topic and related industries. Standard errors clustered by state in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Table 8: Judge Age and Judge Work Output

	(1)	(2)	(3)	(4)	(5)
	Output (Unadjusted)	Output (Standardized)		Output (Rank-Normalized)	
Age (Unadjusted)	-5.877				
	(86.09)				
Age (Standardized)		-0.0160	0.00191		
		(0.0139)	(0.0171)		
Age (Rank-Normalized)				-0.0308	-0.00952
				(0.0193)	(0.0233)
Year FE	X	X	X	X	X
Court-Year FE		X	X	X	X
Cohort FE			X		X
Experience Rank FE			X		X
Court X Cohort Trends			X		X
Case Controls			X		X
N	13727	13643	13641	13655	13653
R-sq	0.059	0.211	0.485	0.014	0.292

Observation is a judge working in a year. "Output" means number of words written by a judge in a year. "Standardized" means centered by court-year and divided by the court-year standard deviation. "Rank" means judges are uniformly distributed between zero and one based on rank within court-year (0 is lowest, 1 is highest). Cohort FE means fixed effect for decade that the judge started on the court. Experience Rank FE means fixed effect for the ranking (one, two, three, etc) of years of experience by judge. Court × Cohort Trends means judge starting-year interacted with court fixed effect. Case controls means the first five principal components of the matrix of controls for legal topic and related industries. Standard errors clustered by state in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Table 9: Age and Quality, by Case Allocation Rule

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Quality (Rank-Normalized)			
Assignment Rule		All		Discretionary		Random	
Age (Rank-Normalized)	-0.110**	-0.130**	-0.146**	-0.190**	-0.194**	-0.110**	-0.133**
	(0.0223)	(0.0305)	(0.0301)	(0.0424)	(0.0435)	(0.0382)	(0.0388)
Exper. (Rank-Normalized)	-0.0861**	-0.105**	-0.0774+	0.0145	0.0325	-0.143**	-0.116*
	(0.0200)	(0.0290)	(0.0389)	(0.0456)	(0.0455)	(0.0330)	(0.0485)
Age × Exper.		0.0379	0.0836+	0.0150	0.0450	0.0417	0.0953
		(0.0462)	(0.0486)	(0.0755)	(0.0886)	(0.0556)	(0.0579)
Court-Year FE	X	X	X	X	X	X	X
Cohort FE			X		X		X
Experience Rank FE			X		X		X
Court X Cohort Trends			X		X		X
Case Controls			X		X		X
N	13655	13655	13655	3678	3678	9977	9977
R-sq	0.042	0.042	0.068	0.045	0.071	0.046	0.073

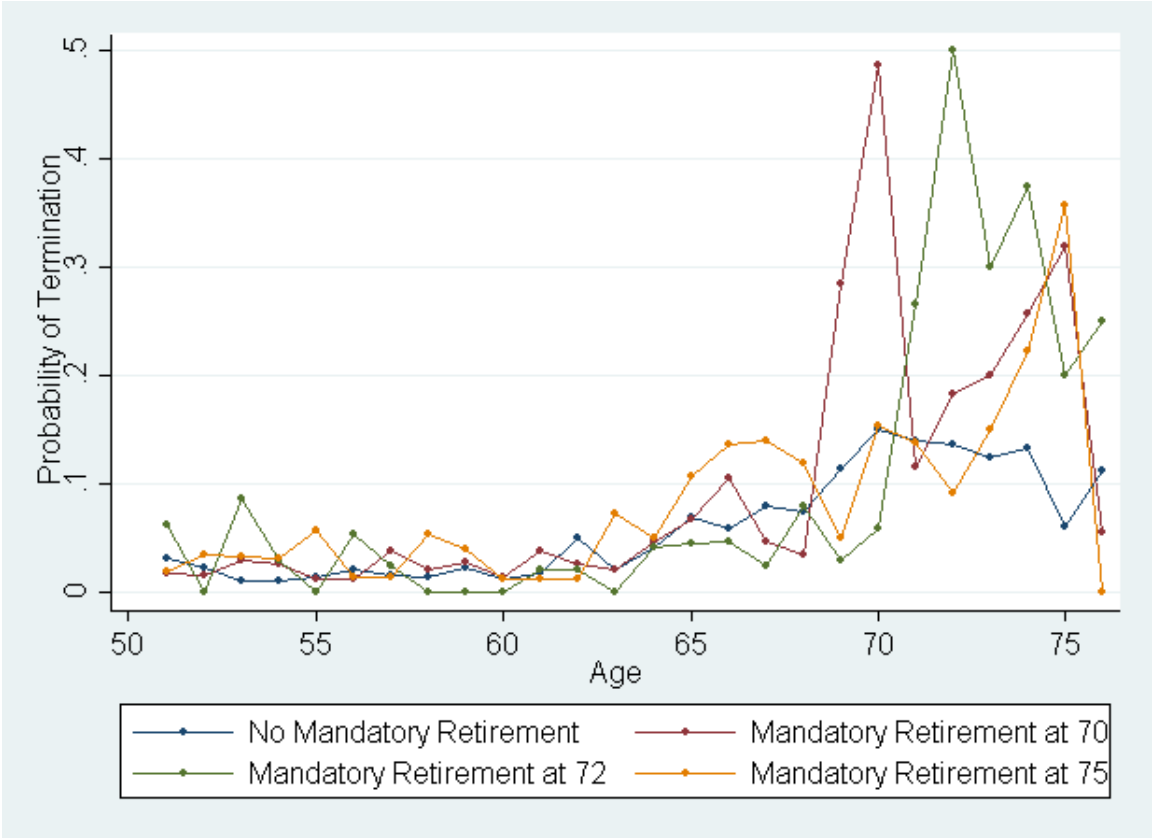
Observation is a judge working in a year. "Assignment Rule" row indicates subsetting based on rule for allocating cases to judges. "Quality" means positive citations per opinion for a judge in a year. "Standardized" means centered by court-year and divided by the court-year standard deviation. "Rank" means judges are uniformly distributed between zero and one based on rank within court-year (0 is lowest, 1 is highest). Cohort FE means fixed effect for decade that the judge started on the court. Experience Rank FE means fixed effect for the ranking (one, two, three, etc) of years of experience by judge. Court × Cohort Trends means judge starting-year interacted with court fixed effect. Case controls means the first five principal components of the matrix of controls for legal topic and related industries. Standard errors clustered by state in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Table 10: Age and Quality, by Case Allocation Rule

	(1)	(2)	(3)	(4)	(5)	(6)
	Quality (Standardized)		Quality (Rank-Normalized)		Case Importance	
Age (Standardized)	-0.179**	-0.163**			-0.00263	
	(0.0285)	(0.0307)			(0.00282)	
Age (Standardized)	0.0930*	0.0834*			0.00239	
× Random	(0.0377)	(0.0390)			(0.00341)	
Exper. (Standardized)	0.0108	0.0417			0.00497+	
	(0.0366)	(0.0655)			(0.00261)	
Exper. (Standardized)	-0.135**	-0.128+			-0.00726*	
× Random	(0.0417)	(0.0760)			(0.00328)	
Age (Rank-Normalized)			-0.182**	-0.166**		-0.0110
			(0.0330)	(0.0366)		(0.00744)
Age (Rank-Normalized)			0.0949*	0.0864+		0.00939
× Random			(0.0424)	(0.0447)		(0.00875)
Exper. (Rank-Normalized)			0.0222	0.0532		0.0152*
			(0.0342)	(0.0558)		(0.00633)
Exper. (Rank-Normalized)			-0.144**	-0.129+		-0.0207*
× Random			(0.0403)	(0.0679)		(0.00841)
Court-Year FE	X	X	X	X	X	X
Cohort FE		X		X		
Experience Rank FE		X		X		
Court X Cohort Trends		X		X		
Case Controls		X		X		
N	13429	13429	13655	13655	13393	13611
R-sq	0.065	0.086	0.045	0.069	0.669	0.672

Observation is a judge working in a year. "Random" is an indicator for random-assignment states, while "Not Rand" means discretionary assignment. "Quality" means positive citations per opinion for a judge in a year. "Standardized" means centered by court-year and divided by the court-year standard deviation. "Rank" means judges are uniformly distributed between zero and one based on rank within court-year (0 is lowest, 1 is highest). "Case Importance" means the OLS prediction of case quality based on case characteristics (legal topic and related industries). Cohort FE means fixed effect for decade that the judge started on the court. Experience Rank FE means fixed effect for the ranking (one, two, three, etc) of years of experience by judge. Court × Cohort Trends means judge starting-year interacted with court fixed effect. Case controls means the first five principal components of the matrix of controls for legal topic and related industries. Standard errors clustered by state in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Figure 11: Retirement Rates by Age, by Mandatory Retirement Age



Probability that a judge retires at a particular age, conditional on working at that age. Plotted separately by mandatory retirement rule.

Figure 12: Age Distribution and Retirement Hazards, by Mandatory Retirement

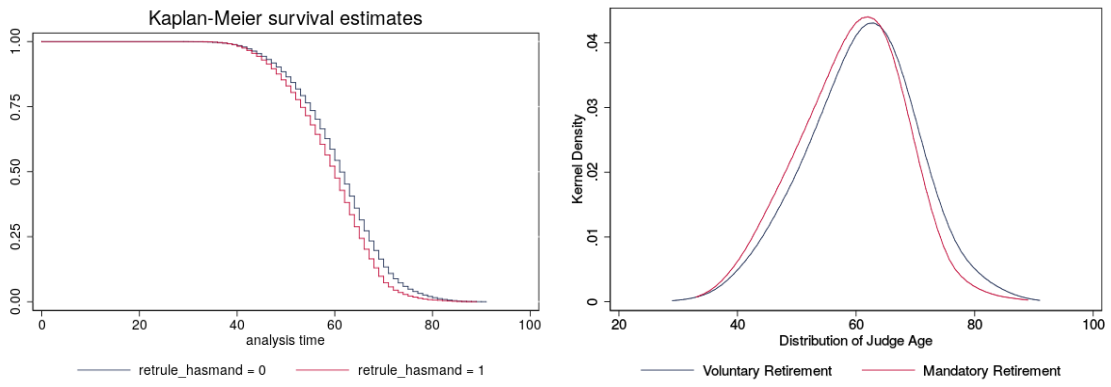


Table 11: Logit Regression of Judge Retirement

	1	2	3	4
Case Assignment	All	All	Discretionary	Random
Effect on retiring after the age of 70				
Retirement Rule = 70 or 72	-0.739**	-1.539***	-2.097***	-1.214**
	(0.243)	(0.298)	(0.477)	(0.394)
Retirement Rule = 75	-1.227***	0.150	0.805	-0.614
	(0.290)	(0.534)	(0.729)	(0.798)
Court Fixed Effects	No	Yes	Yes	Yes

Estimated from multinomial logit with Regression 1 and 2 are all states, 3 is Non-Random states, 4 is Random states

Table 12: Effect of Mandatory Retirement Reform on Log Citations

	(1)	(2)	(3)	(4)	(5)
	Log Positive Citations to Judge				
70/72 Retirement Reform	0.249*	0.283**	0.245**	0.170*	0.153+
	(0.0969)	(0.105)	(0.0890)	(0.0800)	(0.0811)
Year FE	X	X	X	X	X
Court FE	X	X	X	X	X
Court Treat Windows		X	X	X	X
Court Trends		X	X	X	X
Rule Controls			X	X	X
Experience Controls			X	X	X
Case Controls				X	X
Judge FE					X
N	14860	14860	13782	13782	13678
R-sq	0.461	0.529	0.528	0.643	0.747

Observation is a judge working in a year. "70/72 Retirement Reform" is an indicator for the ten yeras after the introduction of mandatory retirment at ages 70 or 72. Court Treat Windows means court-specific treatment windows (ten years before and after reform). Rule controls include dummies for changes to the electoral system, number of judges, and expenditures on judicial system. Experience controls include a quadratic in experience. Case controls means the first five principal components of the matrix of controls for legal topic and related industries. Standard errors clustered by state in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Figure 13: Effect of Retirement Rule on Retirement Age

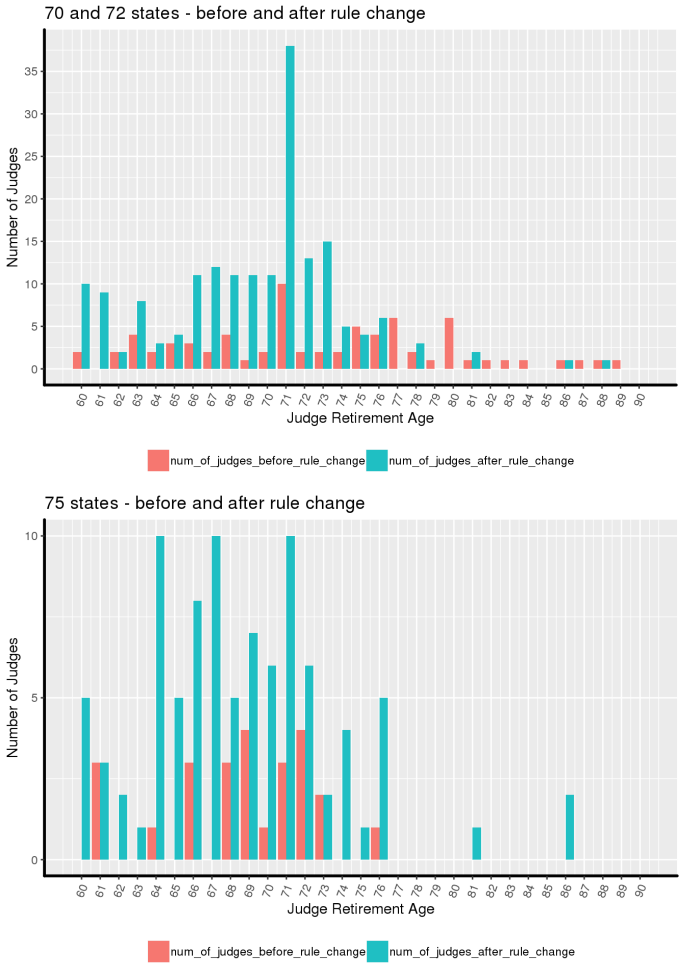
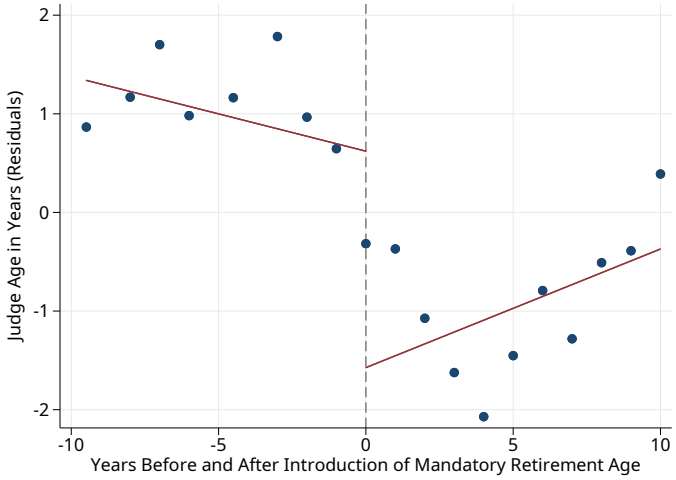


Figure 14: Event Study: Effect of 70/72 Retirement Reform on Average Judge Age



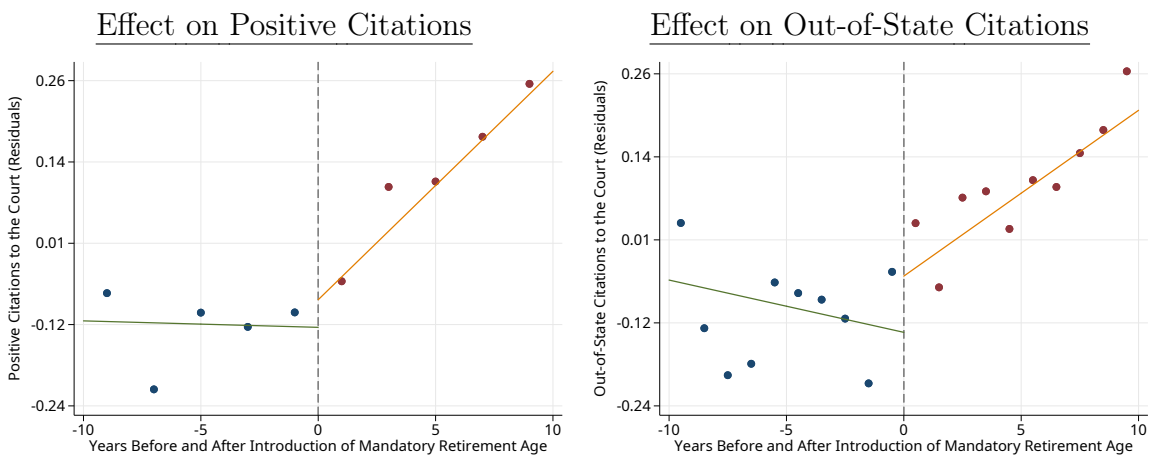
Event study effect of average judge age, before and after rules implementing mandatory retirement at age 70 or 72. Outcome is residualized on court and year fixed effects, court-specific windows and trends, court rule covariates, and experience controls. Binscatter diagram with lfit.

Table 13: Effect of Mandatory Retirement Reform, Other Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	# of Opinions		Work Output		Work Quality		Out-of-State Cites	
70/72 Retire Reform	0.136*	0.112+	0.0751	0.0775	0.0855+	0.0926*	0.173	0.191+
	(0.0538)	(0.0574)	(0.0695)	(0.0656)	(0.0484)	(0.0439)	(0.117)	(0.0978)
Year FE	X	X	X	X	X	X	X	X
Court FE	X	X	X	X	X	X	X	X
Court Treat Windows		X		X		X		X
Court Trends		X		X		X		X
Rule Controls		X		X		X		X
N	15010	13863	15010	13863	15010	13863	15010	13863
R-sq	0.325	0.512	0.266	0.386	0.649	0.718	0.471	0.521

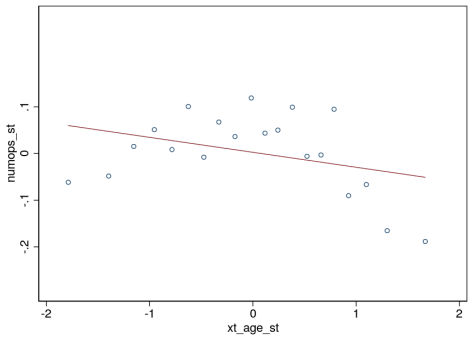
Observation is a judge working in a year. "70/72 Retirement Reform" is an indicator for the ten yeras after the introduction of mandatory retirment at ages 70 or 72. "# of Opinions" is the number of majority opinions written by a judge in a year. "Work Output" is log number of words writen in a year. "Work Quality" is number of citations per published opinion. "Total Out-of-State Cites" is Court Treat Windows means court-specific treatment windows (ten years before and after reform). Rule controls include dummies for changes to the electoral system, number of judges, and expenditures on judicial system. Experience controls include a quadratic in experience. Case controls means the first five principal components of the matrix of controls for legal topic and related industries. Standard errors clustered by state in parentheses. + p<.0.1, * p<0.05, ** p<0.01.

Figure 15: Effect of Retirement Reform on Performance, Event Study

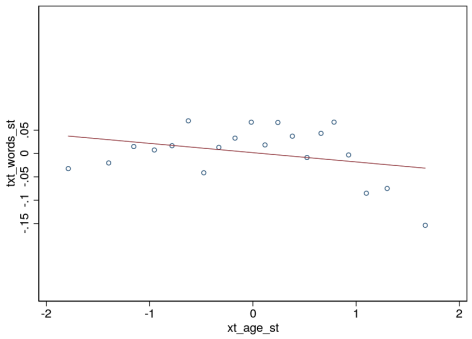


Judge performance before and after reforms implementing retirement ages of 70 or 72. Left panel outcome is log positive citations for a jugde in a year; right panel is only citations from courts in other states. Outcome is residualized on court and year fixed effects, court-specific windows and trends, court rule covariates, and experience controls. Binscatter diagram with lfit.

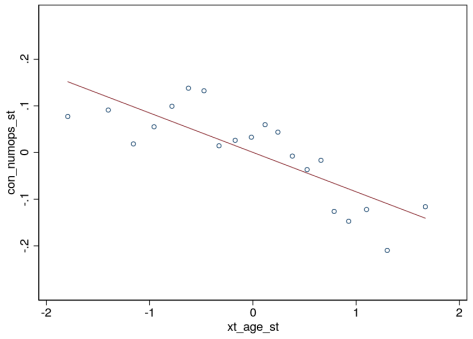
Figure A.1: Performance-Age Profile, within State-Year



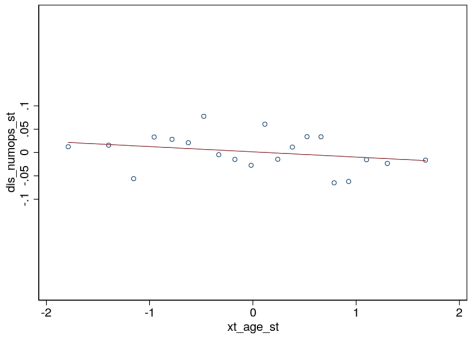
(a) Opinions Written



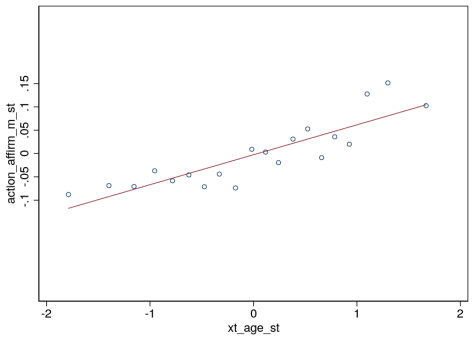
(b) Work Output



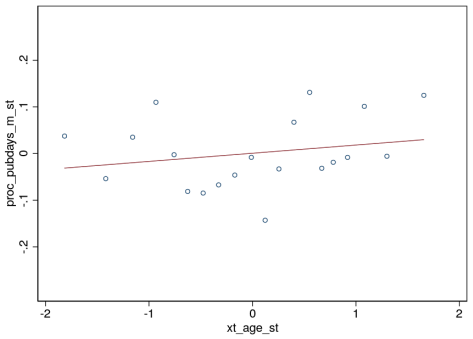
(c) Concurrences Written



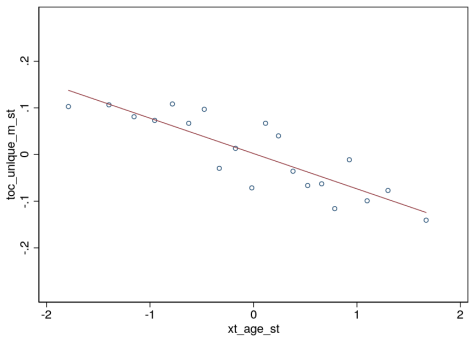
(d) Dissents Written



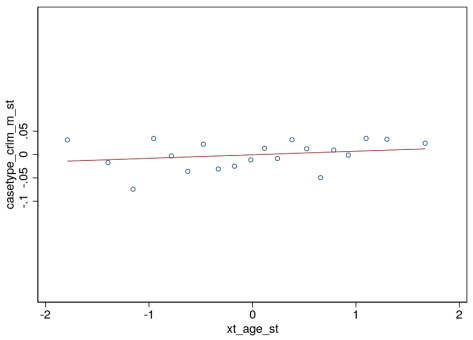
(e) Affirm Rate



(f) Publication Delay

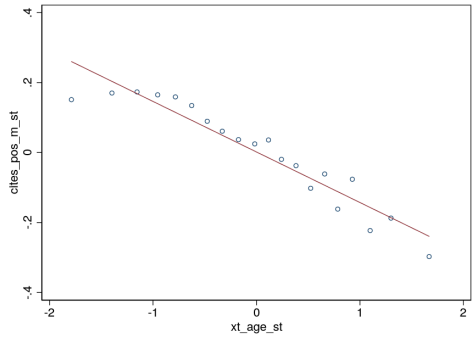


(g) Table of Cases Length

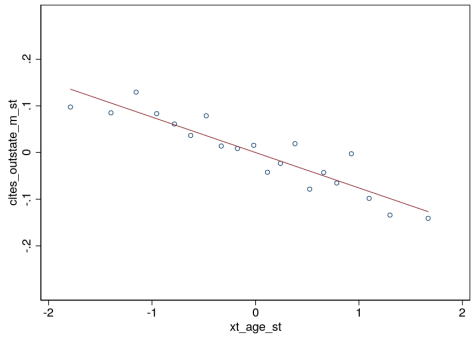


(h) Criminal Case Proportion

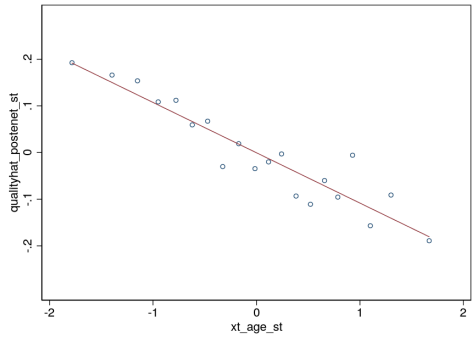
Figure A.2: Performance-Age Profile (2), within State-Year



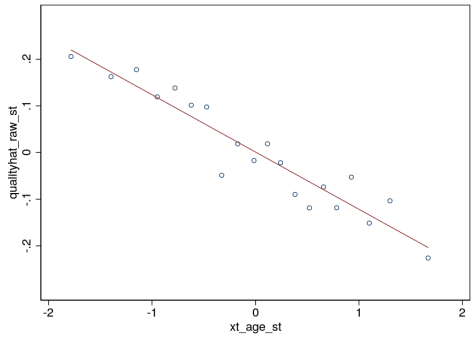
(a) Positive Citations Per Opinion



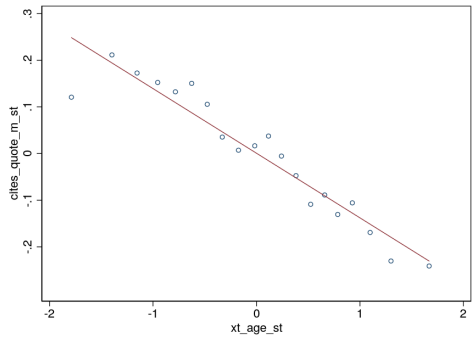
(b) Out-of-State Citations Per Opinion



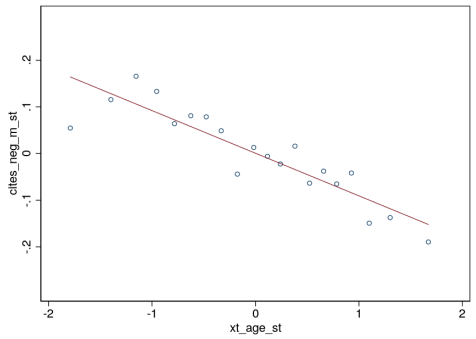
(c) Quality (Predicted from Text)



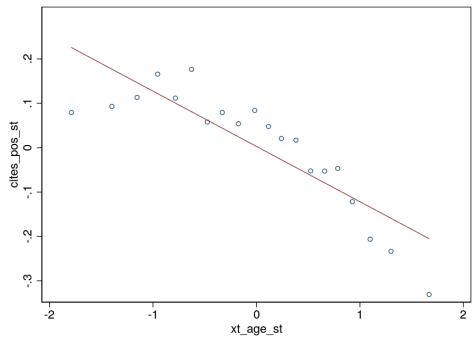
(d) Quality (Predicted from Raw Text)



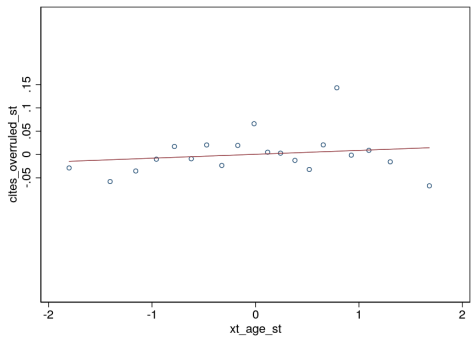
(e) Quoted-By Citations Per Opinion



(f) Negative Citations Per Opinion

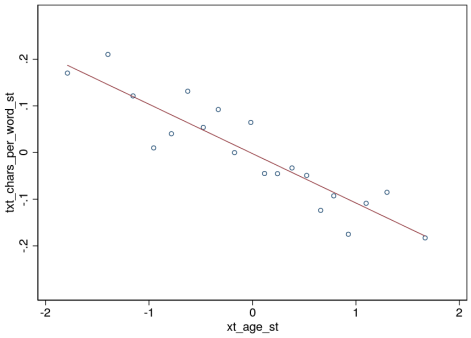


(g) Total Citations

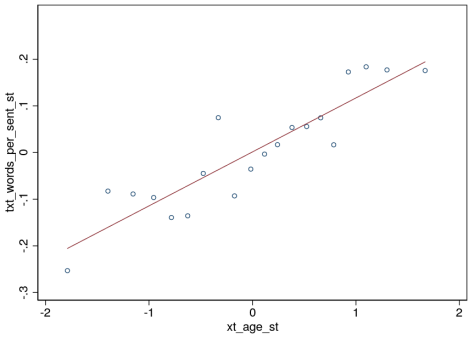


(h) Rate Overruled

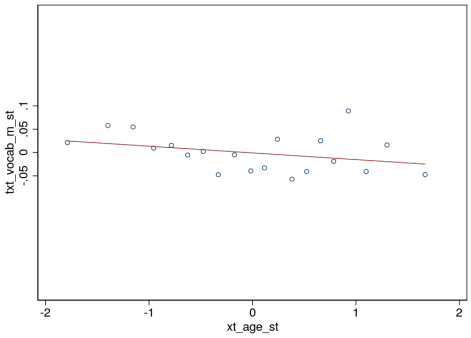
Figure A.3: Performance-Age Profile (3), within State-Year



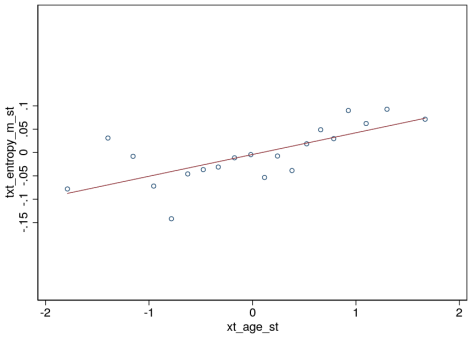
(a) Word Length



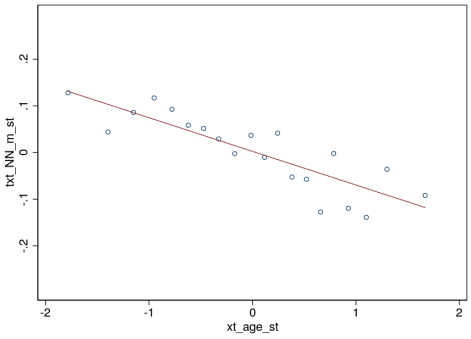
(b) Sentence Length



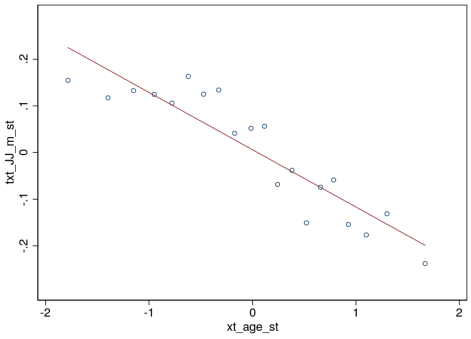
(c) Vocabulary Size



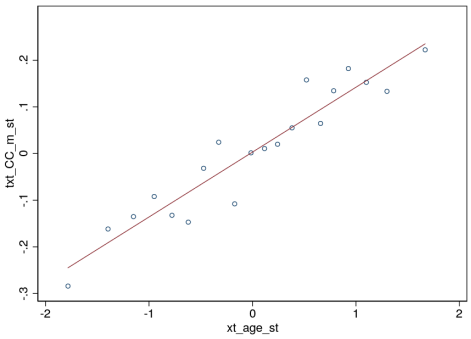
(d) Text Entropy



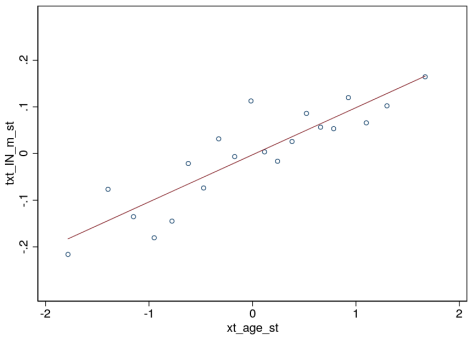
(e) Noun Frequency



(f) Adjective Frequency



(g) Conjunction Frequency



(h) Preposition Frequency

Figure A.4: Judge Output and Quality over the Life Cycle With and Without Mandatory Retirement

