Aging, Retirement, and High-Skill Work Performance: The Case of State Supreme Court Judges

Elliott Ash and Bentley MacLeod*

December 18, 2017

Preliminary Version. Comments Welcome.

Abstract

This paper provides evidence on how high-skill work performance changes in response to biological aging, and in response to mandatory retirement policies. Our data set is constructed from the work product of all state supreme court judges for the years 1947 through 1994. We use machine learning tools to construct new measures of judge work quality from a high-dimensional representation of judge writing style. We find that older judges have the same work output as younger judges but use a different writing style and get fewer citations. Mandatory retirement policies have a demotivating effect on judge performance.

*Elliott Ash: e.ash@warwick.ac.uk; Department of Economics, University of Warwick. 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, 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

The increase in both the quality and length of the human life-span poses an important public policy issue; namely: What is the appropriate employment contract as people approach retirement? It is known that employment is an important ingredient for a healthy life (Sullivan and von Wachter, 2009). But as we age, there is a decline in an individual's ability to work. One response to this has been to impose mandatory retirement for all workers at some age. But in the United States, beginning with the Age Discrimination Act of 1967, the use of mandatory retirement has been greatly reduced.[1] For most workers the current law requires employers to make employment decisions using only worker performance, and not age, as criteria.

The challenge is that producing useful measures of worker performance, particularly for complex tasks, is very difficult. Is it possible to document the changes in a worker performance over time? If so, one could use a performance measure to determine the time at which a worker may be asked to leave employment. But implementing such a rule could lead to distortions in performance as workers respond to the rule.

In this project our goal is to document the performance of a group of highly skilled workers over time - state supreme court judges. We compare the effects of the two main employment policies that are used in practice. One is the use of mandatory retirement under which an individuals is required to leave employment at a pre-specified age. The second is a policy of no fixed retirement dates, with individuals continuing to work as long as their performance is satisfactory.

We analyze a unique dataset that follows the employment patterns of state appellate court judges. We exploit the fact that over our time period a number of states implemented changes to their employment policies, and introduced mandatory retirment ages for these judges (either 70 or 75 years of age). 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.[2]

An unusual feature of judging, relative to most other jobs today, is that the nature of the work has remained essentially unchanged for decades. This allows use build

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

[2] 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`.

meaningful measures of the both the output (number of cases) and the quality (case citations) that allows us to document the variation in performance over time, and as a function of the retirement rules.

We address the following questions. First, in states with no manadatory retirement judges voluntarily choose to leave the bench. We explore two questions. First, how does the performance of a judge vary over time? Second, is there a relationship between the performance of a judge, and their voluntary decision to leave the bench?

Next, we consider the effect of introducing a manadatory retirement rule. This has two effects. First, for judges currently on the bench, the introduction of the rule changes their expectations regarding their future work career. In theory there are two countervailing effects. Judges facing retirment might work harder in order to secure a new job upon retirement. It is very common for retiring judges to enter into private work, such as judging arbitration cases or mediation work. Alternatively, finding such work may require search, in which case performance might decling with the introduction of a manadatory retirment rule.

Finally, we can ask if the performance of the court as a whole improves with the introduction of mandatory retirement. Here we exploit the fact that some states did not introduction such a change, and we can ask how the rule affect relative influence of the courts.

Appellate judges review decisions made by lower courts and then explain, through published opinions, why these decisions should be affirmed, modified, or reversed. Researching for and writing these opinions is the appellate judge's primary professional concern. As discussed at length in Choi et al. (2008) and Ash and MacLeod (2016), opinions provide relatively clean measures of quantity and quality of work output.

This paper uses the richness of our data to construct a new measure of expert work quality. The nice thing about judges is that we observe the text of written opinions, as well as the subsequent endorsing citations to those opinions by other judges. In previous work, we looked at the impacts of employment conditions on the volume of output, and the number of citations. An issue with using citations as a measure of quality is that citations are due not just to judge decisions and writing but also to many factors external to the judge, such as the legal topic and the importance of the precedent. In this work, we use methods from machine learning to address this issue. We form predictions of citations from the text features of a case portfolio, exploiting the random assignment of cases to judges to ensure that unobserved judge characteristics are orthogonal to unobserved characteristics of the assigned cases. Our preferred measure of judge quality is the prediction of citations from the text features of a judge's written opinions. Therefore we produce a measure that uses a judge's conscious choices about how to write opinions.

The first goal of this paper is to provide descriptive statistics on the performance of state appellate judges over the life cycle. We build on previous work (e.g. Posner, 1995; Smyth and Bhattacharya, 2003; Teitelbaum, 2006) by looking at a much larger sample of courts and judges, by using panel data rather than cross-sectional data, and by analyzing a broader range of behavioral outcomes. Appellate judging is an ideal profession for this inquiry because all courts operate the same way, the job has not changed much in the last 200 years, salary is not contingent on experience or performance, and there are no formal differences by age in job requirements.

The second goal is to analyze the effects on performance of judge mandatory retirement policies. Mandatory retirement could improve court performance by removing older, low-performance judges. It may also have a demotivating effect on judges who must seek alternative careers.

Our results can be summarized as follows. Relative to their younger colleagues, older judges have about the same level of work output, but write lower-quality decisions that are cited less often by future judges. Older judges use a different writing style, using shorter words and longer sentences. Mandatory retirement policies have a demotivating effect on sitting judges.

The rest of the paper unfolds as follows. Section 2 reviews the literature. Section 3 examines the mechanisms relating aging and performance. Section 4 characterizes the institutional setting, while Section 5 describes the data. Section 6 provides some descriptive evidence about aging and retirement among state supreme court judges. Section 6 estimates the impacts of mandatory retirement policies on judge performance. Section 7 concludes.

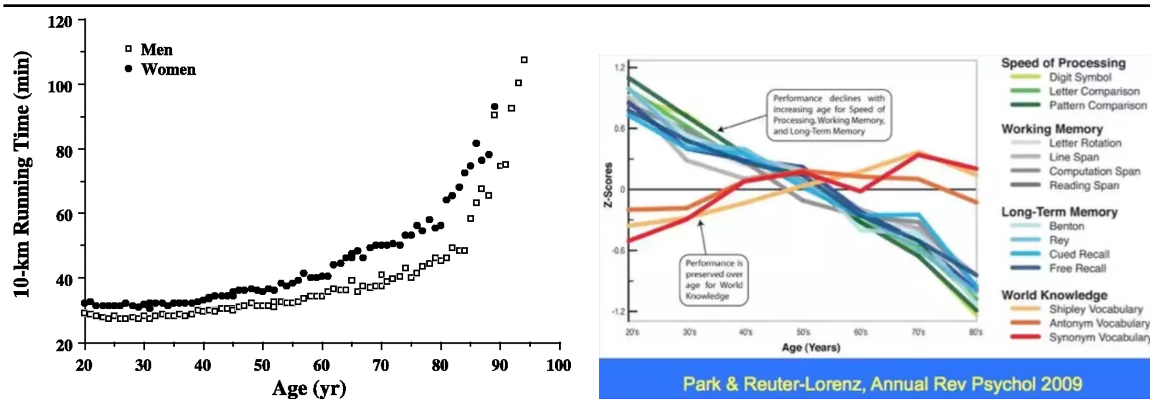
2 Background

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 recognition 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.

Figure 1: Performance vs Age, Physical and Cognitive Tasks



(a) 10k Running Time (b) Cognitive Performance

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.

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 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."

2.2 Aging in the Judiciary

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).

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.

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 (Gordon and Huber, 2007; Shepherd, 2009a,b).

The fruitful structural literature on retirement choice has not yet been applied to judges. This literature, 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.

3 Retirement Mechanisms

The purpose of this section is to discuss the incentives that are implicit in any retirement system, with a model based on Lazear (1979). We are interested in the behavior of judge j at time t. The judge j is described by a vector of characteristics at time t, X_{jt}, which include age, ability, reputation, health status, etc. Some factors, such as health, are not observed directly.

At any period t, the judge chooses an effort level $e_t \in \mathbb{R}$, quality level $q_t \in \mathbb{R}$, and whether to continue working $r_t \in \{0, 1\}$.

The judge faces the following dynamic programming problem:

$$\begin{aligned} V_{jt} = \max_{e_t, q_t, r_t} & r_t \{u(e_t, q_t; X_{jt}) + \delta V(X_{jt+1})\} \\ & + (1 - r_t) V^R(X_{jt+1}) \end{aligned}$$

where $u(\cdot)$ is the value of working as a judge and V^R is the present discounted value from retirement. If the judge does not retire, then effort and quality are choosen to solve:

$$\begin{aligned} \frac{\partial u}{\partial e_t} + \delta \frac{\partial V\left(X_{jt+1}\right)}{\partial e_t} &= 0 \\ \frac{\partial u}{\partial q_t} + \delta \frac{\partial V\left(X_{jt+1}\right)}{\partial q_t} &= 0 \end{aligned}$$

From Ash-Macleod (2015) we know that Judges have an intrinsic preference for quality, and hence even in the absence of future rewards ($\frac{\partial V(X_{jt+1})}{\partial q_t} = \frac{\partial V(X_{jt+1})}{\partial e_t} = 0$), they still choose positive effort and quality. If mandatory retirement is introduced at date $t+1$, then we have

$$V\left(X_{jt+1}\right) = V^R\left(X_{jt+1}\right),$$

and thus we can compare the returns to performance on and off the bench at date $t+1$ since we have at date t:

$$\begin{aligned} \frac{\partial u}{\partial e_t} + \delta \frac{\partial V^R\left(X_{jt+1}\right)}{\partial e_t} &= 0 \\ \frac{\partial u}{\partial q_t} + \delta \frac{\partial V^R\left(X_{jt+1}\right)}{\partial q_t} &= 0 \end{aligned}$$

.

In the absence of mandatory retirement, we observe individuals retire at all ages. This will occur at date t if and only if

$$V^R(X_{jt}) > V(X_{jt}).$$

We will explore the relationship between observable characteristics X_{jt} and their retirement decision.

4 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.

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.

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.

5 Data

The data-set used for the empirical analysis is an extension of that used in Ash and MacLeod (2015) and Ash and MacLeod (2016). 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 deathdates, how judgeships ended, and judge retirement policies.

First 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.

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.

Our institutional treatment variables are changes in state court policies affecting the judge retirement decision. These are described in more detail in Section 8.

Our performance data are constructed from the text and citations for opinions. They are described in the next section.

6 Text Features and Case Quality

This section discusses our method for using the the richness of our data to construct a new measure of expert work quality. We observe the text of written opinions, as well as the subsequent endorsing citations to those opinions by other judges.

Citations to an opinion can be viewed as a "clean" performance measure, just as future earnings can be viewed as a clean measure of a person's investment into human capital. However, this measure is of little value in assessing the performance of a currently sitting judge because it depends upon evaluations that are done years or decades into the future. In addition, citations are due not just to judge decisions and writing but also to many factors external to the judge, such as the legal topic and the importance of the precedent. In this work, we use methods from machine learning to address this issue. We form predictions of citations from the text features of a judge's opinions. Therefore we produce a measure that uses a judge's conscious choices about how to write their work product.

6.1 Predicton Model

We are interested in forming predictions of citations based on judge writing style. The outcome of interest is q_{jct}, positive citations per opinion by judge j in court c at year t. A judge's case portfolio is characterized by a set of text features X_{jct}, with element x_{jct}. These include word length, sentence length, paragraph length, frequency distributions over parts of speech, and their interactions.

We model judge quality as a function of writing style

$$q_{jct} = \alpha_{ct} + X'_{jct}\beta + I_{jct} + \epsilon_{jct} \tag{1}$$

where α_{ct} include state-year interacted fixed effects and I_{jct} includes unobserved judge characteristcs. We estimate the model in the states with random or rotating assignment of judges. With random assignment of judges at the state-year level, the inclusion of the fixed effects implies orthogonality of judge characteristics and case characteristics. Formally, we assume $\mathbb{E}\{X_{jct}I_{jct}|c,t\} = 0$.

We are not interested in the fixed effects estimates, so for computational reasons we residualize q_{jct} and X_{jct} on the fixed effects and estimate

$$\tilde{q}_{jct} = \tilde{X}'_{jct}\beta + \epsilon_{jct} \tag{2}$$

which will procure the same estimates for $\hat{\beta}$ as (1).

Using $\hat{\beta}$, we compute

$$\hat{\tilde{q}}_{jct} = \tilde{X}'_{jct}\hat{\beta}$$

and

$$\hat{q}_{jct} = X'_{jct}\hat{\beta}$$

where note that the second prediction is formed from the raw text features, rather than the residualized features. The predictions are formed for all states, not just those with random assignment of cases.

6.2 Text Features

The text features variables X is constructed from raw-text documents. We follow standard methods for document vectorization, and represent the document as frequency distributions over style features (e.g. Ash, 2016; Gentzkow et al., 2017).

Our set of text features is as follows.. We have number of paragraphs, sentences, words, and characters per document. We have sentences per paragraph, words per sentence, and characters per word. We have tfidf-weighted case length, which down-weights common and rare words (e.g. Gentzkow et al., 2016).

Next we have a text entropy measure, defined as

$$\sum_{w} -f_w \log(f_w)$$

where f_w is the frequency (probability) of word w, and the summation is over all words appearing in a case.

Finally, we have the frequency distribrution over parts of speech. This includes the set of 36 tags in the Penn TreeBank. The set includes nouns, verbs, adjectives, conjunctions, prepositions, etc.[3]

For additional predictiveness, we can also include the full set of interactions between features.

6.3 Dimension Reduction

The problem of high-dimensionality is similar to Ash (2016) and Belloni et al. (2012). We have a large number of regressors. Non-regularized least squares regression generates inconsistent estimates for parameters under these conditions. Chernozhukov et al.

[3]See `https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html` for the full list.

(2017) show how to resolve this issue with a sufficient dimension reduction and sample split.

We estimate (2) using a regularized learning model. We will start with two linear models, elastic net and partial least squares. Elastic net is OLS with L1 and L2 penalties, with penalties selected by cross-validation (Zou and Hastie, 2005). This forces all weak predictors to zero, with the remaining predictors being the optimal ones under sparsity (Belloni et al., 2012).

Similarly, we use partial least squares to simultaneously reduce the dimensionality of X and predict $\hat{q}_{jct}$, using the method from Chun and Keleş (2010). PLS provides an optimal linear combination of the features to make a prediction; a linear combination of exogenous covariates is itself exogenous. The number of PLS components is chosen by out-of-sample cross-validation.

We then have a prediction $\hat{q}_{jct}$ for each judge-year in our dataset. Changes in $\hat{q}_{jct}$ for a judge over time reflect changes in quality-relevant judge writing style.

6.4 Judge Randomization

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. 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.

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 might review a case. There can be exceptions for specialized cases such as those involving the death penalty. We assume these deviations from randomness are independent of our main effects, though we will also report omnibus checks of 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 (2011) 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.

In the current draft, our main randomization check is that the type of case (constitutional, criminal, civil, administrative) does not change significantly over the lifespan under random assignment of judges.

6.5 Discussion

One issue with this approach is that judges likely use a different writing style on important cases. So an important case, which gets high citations because it is important, also has a different set of text features because it is important. Then our coefficients are driven by a judge's response to importance, rather than a causal effect of judge's writing style on citations.

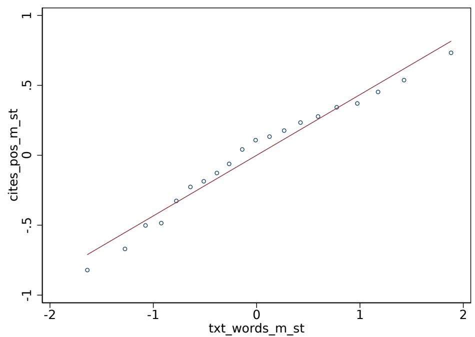
This issue also exists when using actual citation counts. In addition, this should be addressed by the random assignment of judges to cases. That said, one potential response to this issue is to instrument for X_{jct} with Z_{jct}, defined as the judge's writing style in other years.

6.6 Prediction and Validation

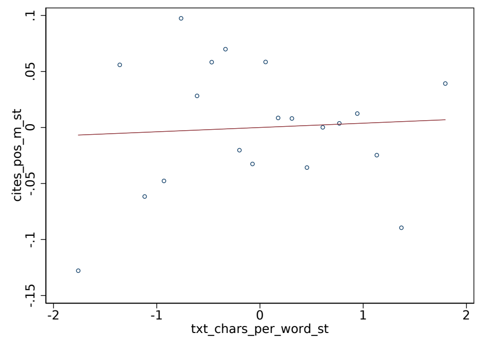
Figure 2 depicts the relation between citations and a selection of text features from judge written opinions. Perhaps unsurprisingly, judges who write longer opinions also get more citations. There is not much of a relationship between citations and the length of words or sentences. Our text entropy measure increases with citations. Noun frequency also increases, while determinant frequency decreases. In general, we found that content words (nouns, verbs, and adjectives) were increasing with citations, while function words (determinants, conjunctions, prepositions) were decreasing with citations.

We trained the predictive model as follows. First we excluded the discretionary-assignment states. Then we took half of the random-assignment states and set them as the training sample. The other half of these states were the test sample. We trained elastic net with L1 and L2 penalties, selecting penalty parameters by 10-fold cross validation (Friedman et al., 2010). We then formed predictions from the text in the whole sample. The correlation of truth and prediction in the held-out test sample is

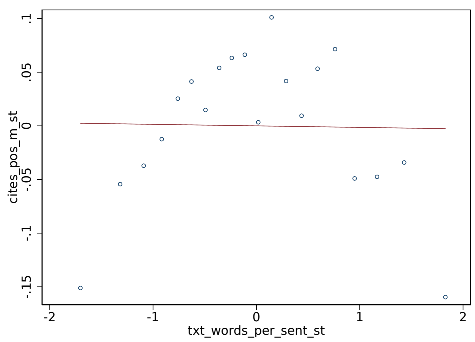
Figure 2: Quality-Text Relations, Residualized on State-Year Interacted Fixed Effects



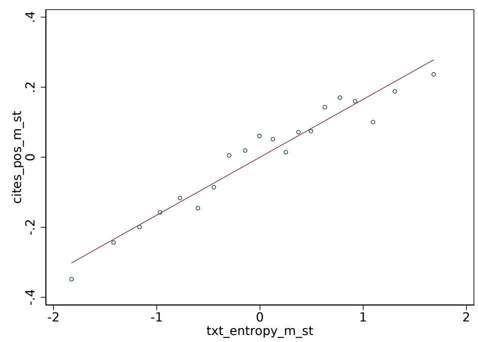
(a) Style: Words Per Opinion



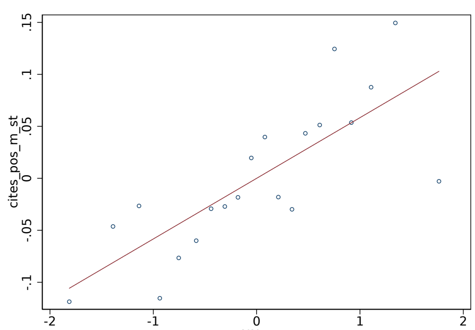
(b) Style: Word Length



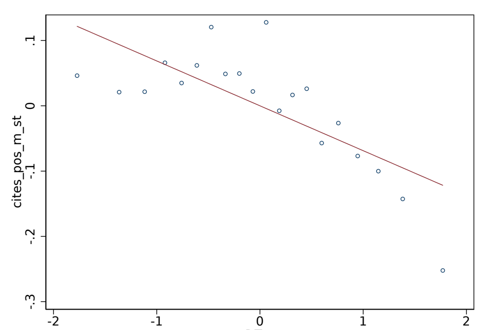
(c) Style: Sentence Length



(d) Style: Text Entropy

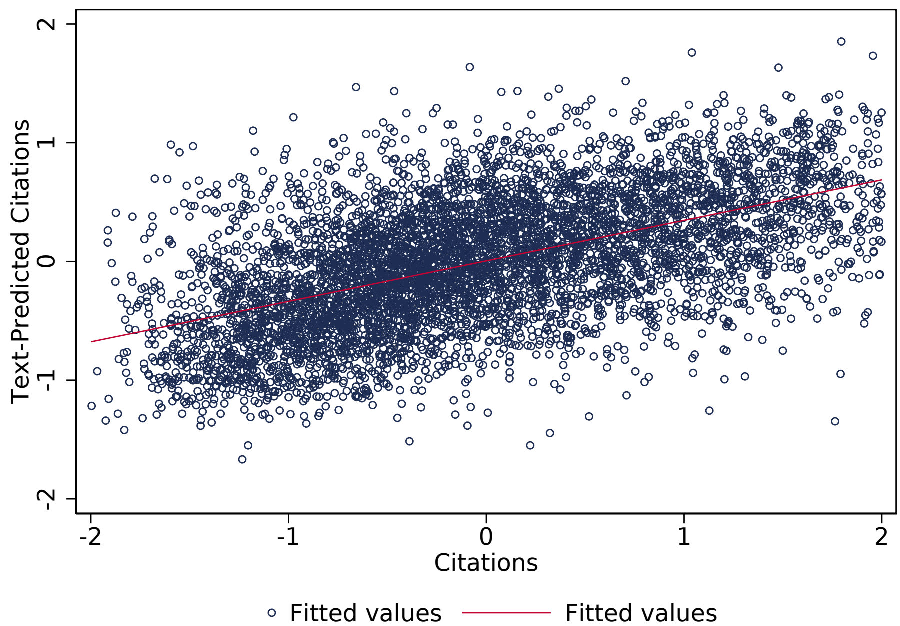


(e) Style: Noun Frequency



(f) Style: Determinant Frequency

Figure 3: Out-of-Sample Prediction of Citations Using Opinion Text Features



0.56, meaning that our set of text feature variables explains about 32% of the within-state-year variance of citations per opinion.

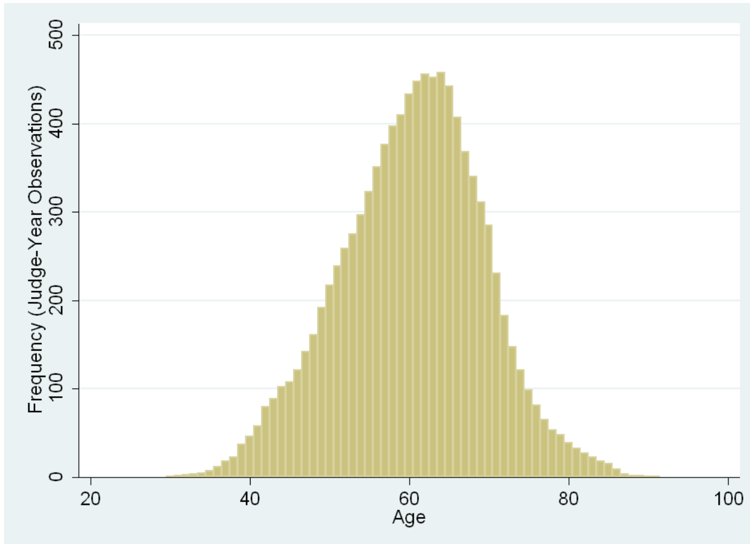
Table 2 reports the list of features selected by elastic net , along with their OLS coefficients in a post-elastic-net regression. Figure 3 includes a scatter plot of predicted citations against actual citations in a held-out test sample. There is shrinkage toward zero, reflecting that our text variables only explain 33% of the variance in citations.

One way of interpreting this is that judge choices account for at least 33% of our citation metric. There are other judge choices – such as how the legal content of the decision responds to the current legal climate – that we do not observe. Therefore our metric is a lower bound on the judge's role in citations. The other 67% that is not explained can be due to unobserbed judge factors, but also due to many other external factors outside the judge's control. Our new metric of quality serves to exclude the noise generated by these factors.

Table 1: Text Features and Quality: Post-Elastic Net Regression

Effect on Quality	Coef.	Std. Err.	t	P>t
r_txt_pars_m	0.034	0.034	1.010	0.329
r_txt_sents_m	0.153	0.055	2.760	0.013
r_txt_tfidf_m	-0.013	0.022	-0.570	0.578
r_txt_tfidf_length_m	-0.324	0.085	-3.810	0.001
r_txt_vocab_m	0.547	0.059	9.300	0.000
r_txt_vocabnorm_m	-0.076	0.023	-3.290	0.004
r_txt_words_per_sent	0.042	0.028	1.480	0.158
r_txt_chars_per_word	0.061	0.025	2.460	0.025
r_txt_sents_per_par	0.085	0.029	2.940	0.009
r_txt_CC_m	-0.033	0.020	-1.640	0.120
r_txt_CD_m	-0.028	0.017	-1.610	0.126
r_txt_DT_m	0.078	0.023	3.390	0.003
r_txt_FW_m	0.031	0.022	1.390	0.183
r_txt_IN_m	-0.074	0.027	-2.780	0.013
r_txt_JJ_m	0.082	0.024	3.360	0.004
r_txt_JJS_m	0.004	0.015	0.250	0.805
r_txt_MD_m	0.006	0.017	0.380	0.710
r_txt_NN_m	0.077	0.028	2.730	0.014
r_txt_NNS_m	-0.056	0.019	-2.930	0.009
r_txt_PDT_m	-0.017	0.018	-0.970	0.347
r_txt_PRP_m	0.041	0.027	1.520	0.147
r_txt_PRP__m	0.014	0.025	0.560	0.583
r_txt_RB_m	0.038	0.026	1.470	0.160
r_txt_RBR_m	-0.032	0.015	-2.110	0.050
r_txt_RBS_m	0.059	0.015	3.930	0.001
r_txt_RP_m	-0.041	0.021	-1.960	0.067
r_txt_TO_m	-0.020	0.023	-0.880	0.393
r_txt_VB_m	0.020	0.021	0.930	0.368
r_txt_VBD_m	-0.095	0.026	-3.630	0.002
r_txt_VBG_m	0.003	0.021	0.160	0.878
r_txt_VBN_m	-0.029	0.021	-1.380	0.184
r_txt_VBZ_m	0.055	0.024	2.250	0.038
r_txt_WDT_m	-0.016	0.025	-0.640	0.528
r_txt_WP_m	0.015	0.012	1.270	0.222
r_txt_WRB_m	0.080	0.018	4.350	0.000

Figure 4: Age Distribution of Working State Supreme Court Judges



7 Judge Age Distribution and Voluntary Retirement Decisions

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.

Figure 4 shows the age distribution for all state supreme court judges working between 1947 and 1994. Figure 5 shows the distribution of the starting age. Figure 6 shows the distribution of the ending age. Figure 7 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.

Next we look at how judge performance is related to the retirement choice. A basic question is whether judges tend to be better or worse than their colleages at the time they retire. Therefore we estimate a Cox survival regression, where the implicit outcome is retirement. We construct a dummy variable $GoodHalf_{it}$ equaling one when a judge is above the median decision quality in a court-year, and zero otherwise.

We plot survival estimates for supreme court judges, split by $GoodHalf_{it}$, in the

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

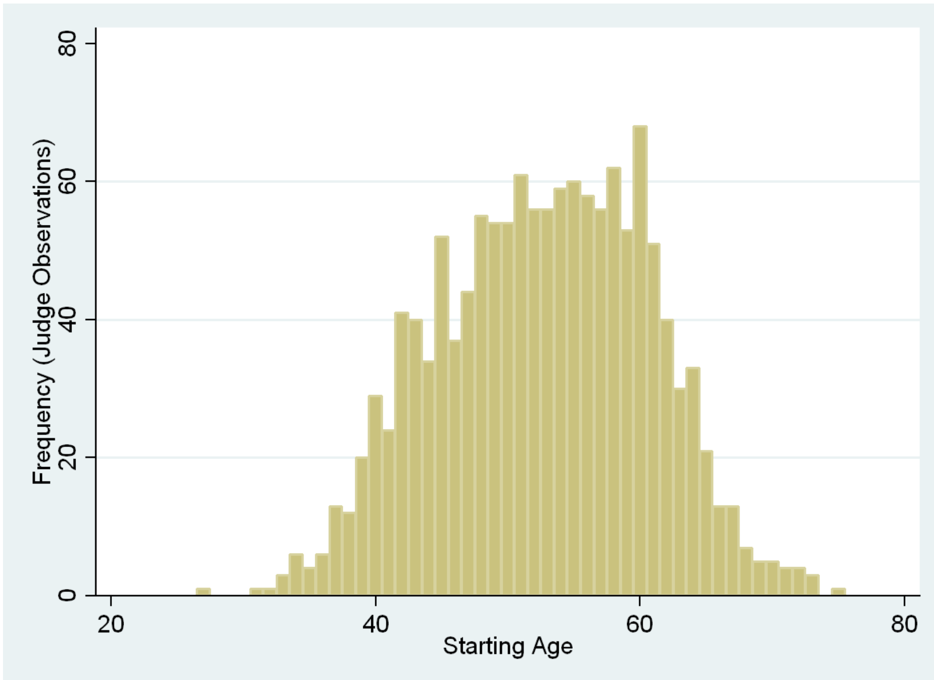


Figure 6: Ending-Age Distribution of State Supreme Court Judges

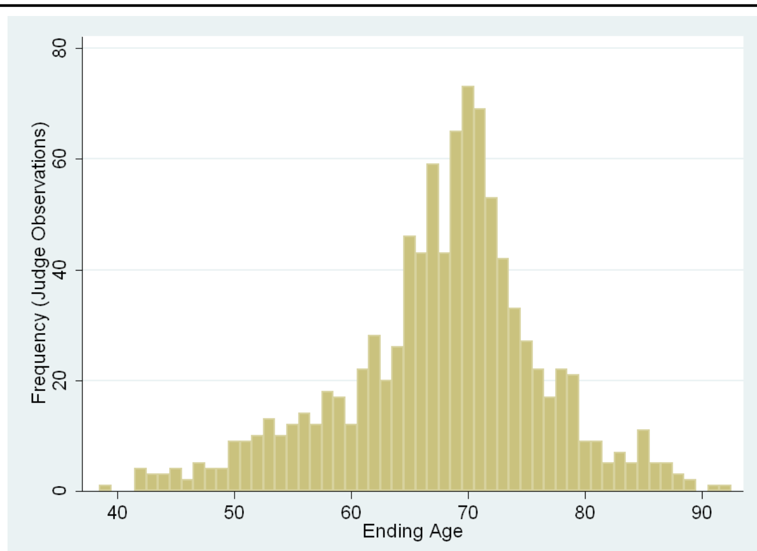
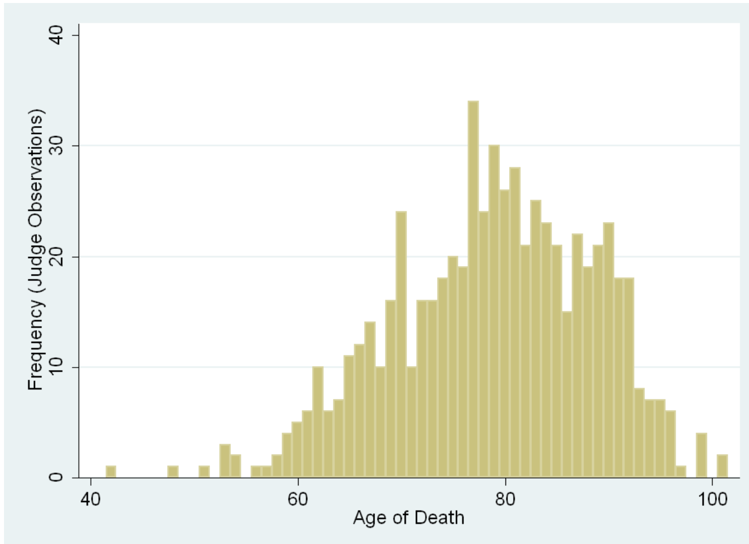


Figure 7: Age-of-Death Distribution of State Supreme Court Judges



top half of Figure 8. This graph shows that when judges retire, they tend to be better than their colleagues, conditional on age. This is consistent with better judges leaving earlier in order to pursue other career opportunities. Additional evidence consistent with this view is in the bottom panel of the figure, which shows that judges who get other jobs after judging also retire earlier.

8 Aging and Judge Performance

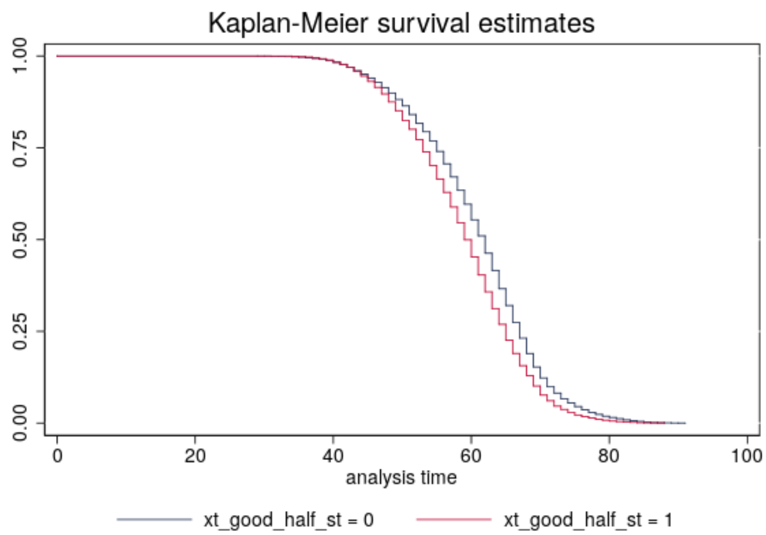
Next we provide descriptive statistics on how differences in ages affect performance. The empirical strategy for examining the effects of aging on judicial behavior is to exploit differences in performance between judges working in the same court at the same time. We look at differences in output and quality by the age of a judge controlling for other judge and court-level characteristics.

The regressions in this section include all states and years. In the appendix we include results separately by mandatory and voluntary retirement. We separate out results before/after 1970 (when WestLaw and LexisNexis were introduced). We also look at states separately based on whether they have discretionary or random assignment of cases to judges.

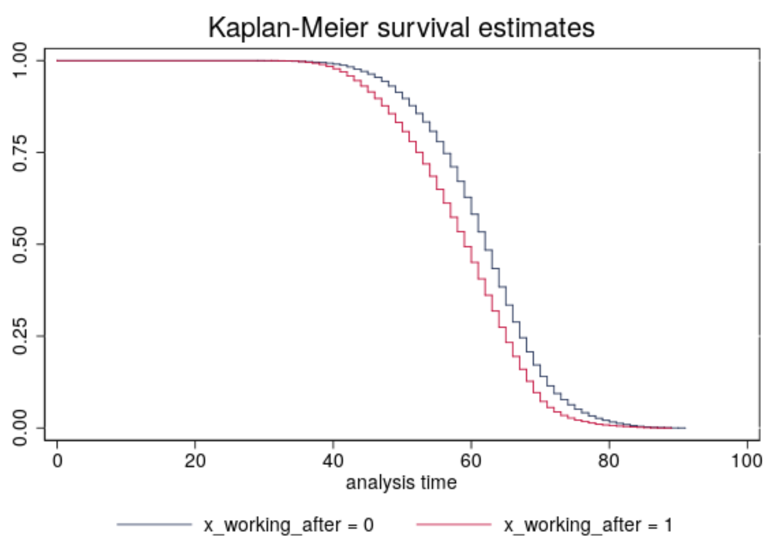
The main source of bias we are interested in comes from the time-varying changes

Figure 8: Retirement Hazards for State Supreme Court Judges

Judges Who Write Higher-Quality Decisions Tend To Retire Earlier



Judges With Outside Jobs Retire Earlier



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

The linear model for performance variable y_{jst} for judge j working in court s at year t is

$$y_{jst} = \alpha_{jst} + \rho \text{Age}_{jst} + X'_{jst}\beta + \epsilon_{jst} \tag{3}$$

where α_{jst} includes fixed effects, Age_{jst} is the age for judge j at t, and X_{jst} may include time-varying judge-level controls. We standardize the outcome and age by state-year, or by judge, depending on the relationship we are interested in estimating. Therefore our estimates can be interpreted as the average difference in standard deviations of the outcome variable, for a one-standard-deviation increase in age relative to a judge's colleagues (state-year standardization unit) or relative to the rest of a judge's career (judge level standardization unit). This is equivalent to including state-year fixed effects for α_{jst}, or including judge fixed effects for α_{jst}, respectively.

We run OLS to obtain coefficient estimates from Equation 3. These estimates are reported in Table 1. We report estimates within-state-year and within-judge, and we report estimates with and without a set of controls (and also dropping the first and last year, which might have partial workloads). There is not much difference across ages for number of opinions written or total work output. However, there is a large and significant decrease with age in the quality of decisions, as measured by citations from later judges. This effect is not significant for the within-judge regressions. However, the text-predicted quality measure is significantly negative in all specificiations.

Next, we estimate Equation 3 without the age term, and obtain the residuals. We then plot the mean residualized performance variables, binned by the residualized age variable. These are reported in Figures 9, 10, and 11. Note that the scales for the y axis are comparable across all graphs.

Figure 9 looks at a collection of variables related to workload, output, and time allocation. First we see that there are not big effects on opinions written and total words written (output), as seen from the regression table. One interesting results is that 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

Table 2: Regression Estimates, Within-State-Year Age-Performance Effect

	Standardized Within State Year							
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
	Log Opinions Written		Work Output		Quality (Citations)		Quality (Predicted)	
Age Effect	-0.0321*	-0.0431	-0.0199	-0.0183	-0.144***	-0.0999***	-0.108***	-0.0818*
	(0.0154)	(0.0223)	(0.0186)	(0.0244)	(0.0193)	(0.0249)	(0.0224)	(0.0306)
Controls		X		X		X		X
Exclude First/Last Year		X		X		X		X

	Standardized Within Judge							
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
	Log Opinions Written		Work Output		Quality (Citations)		Quality (Predicted)	
Age Effect	-0.0307	-0.0589	0.00354	-0.0182	-0.00465	-0.0478	-0.122***	-0.101**
	(0.0276)	(0.0310)	(0.0271)	(0.0276)	(0.0261)	(0.0334)	(0.0171)	(0.0297)
Controls		X		X		X		X
Exclude First/Last Year		X		X		X		X

N = 15392 judge-years, N = 13315 when dropping first and last years. In top panel, outcomes and age standardized within state-year. In bottom panel, outcomes and age standardized within judge. Controls include fixed effects for decade of birth and experiece (five-year bins). * p<0.05, ** p<0.01 *** p<0.001.

Table 3: 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

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 who live that long.

by age in the time between assignment and publication of opinions. The number of previous cases cited (table of cases length) decreases. There is not much change in the composition of their case portfolio, as seen by no change in the proportion of criminal cases (this holds for constitutional, civil, and administrative cases as well).

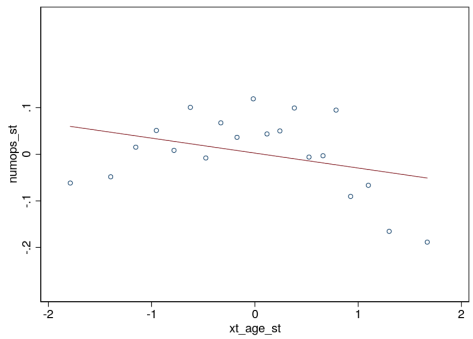
Figure 10 looks at our quality measures. As we saw in the table, there is a large decrease in opinion quality over the life cycle. This holds for out-of-state cites, text-predicted cites, direction quotations by future judges, negative (rather than positive) citations, and total citations (rather than citations per opinion). There is no effect on the rate the judges are overruled.

Figure 11 looks at differences in writing style by age. 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.

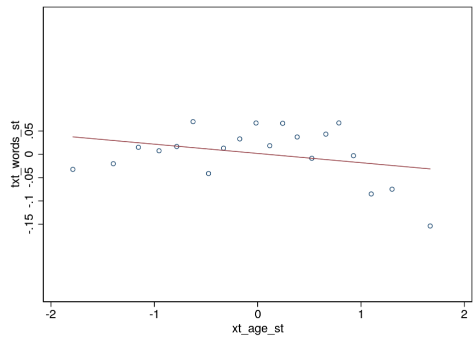
9 Effect of Mandatory Retirement Policies

This section examines the role of mandatory retirement policies in the judge retirement choice, as well as choices on the job. We are interested in the effects of mandatory retirement policies on judge output and quality.

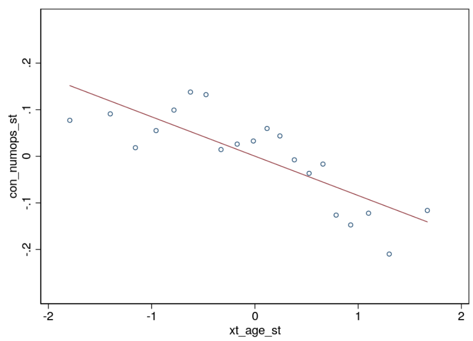
Figure 9: Performance-Age Profile, within State-Year



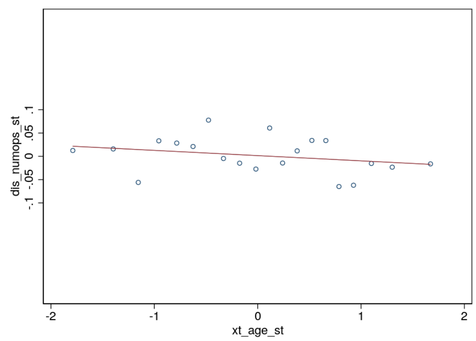
(a) Opinions Written



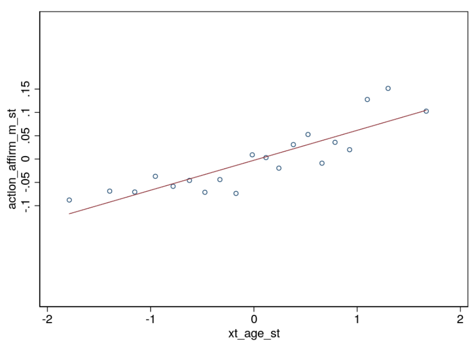
(b) Work Output



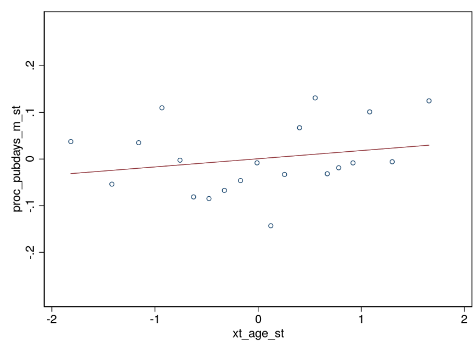
(c) Concurrences Written



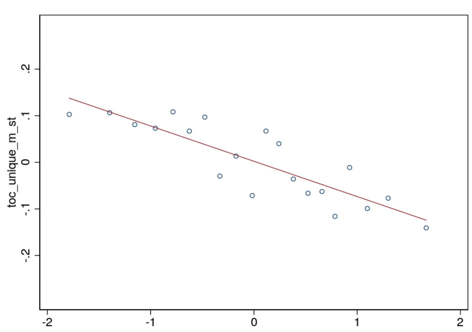
(d) Dissents Written



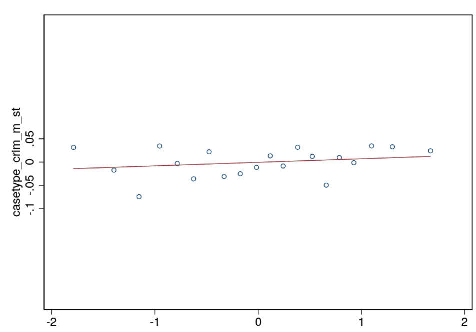
(e) Affirm Rate



(f) Publication Delay

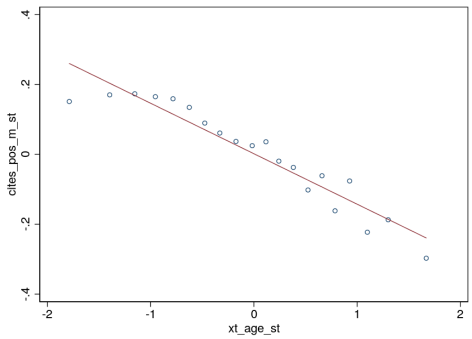


(g) Table of Cases Length

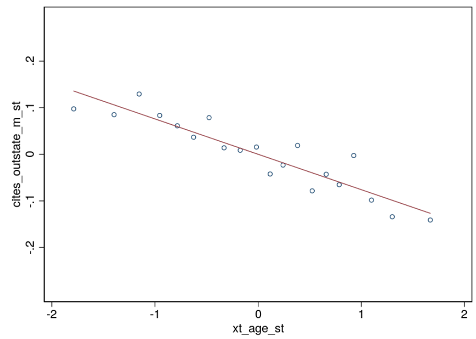


(h) Criminal Case Proportion

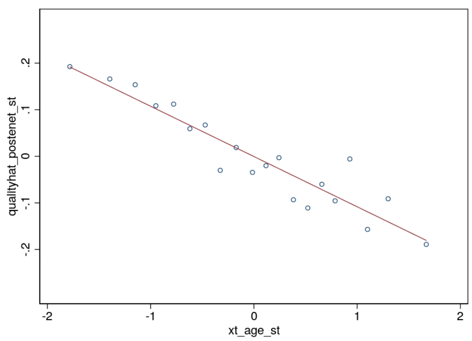
Figure 10: Performance-Age Profile (2), within State-Year



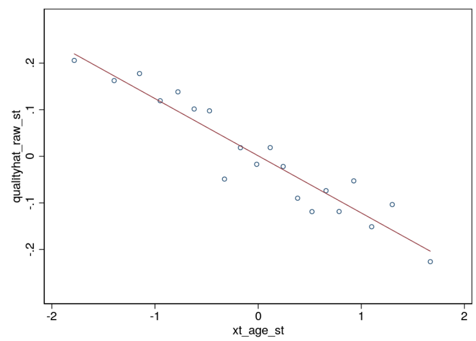
(a) Positive Citations Per Opinion



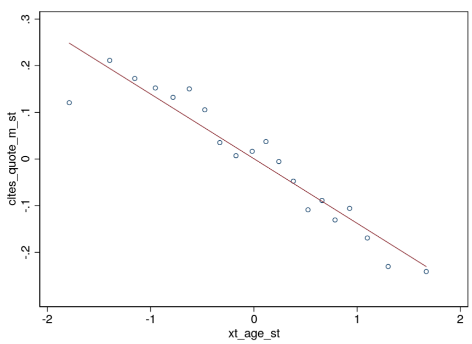
(b) Out-of-State Citations Per Opinion



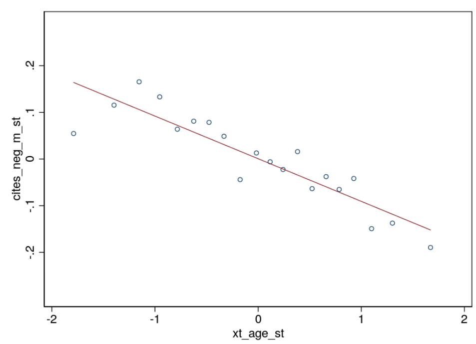
(c) Quality (Predicted from Text)



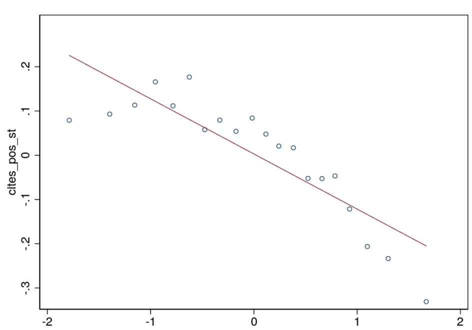
(d) Quality (Predicted from Raw Text)



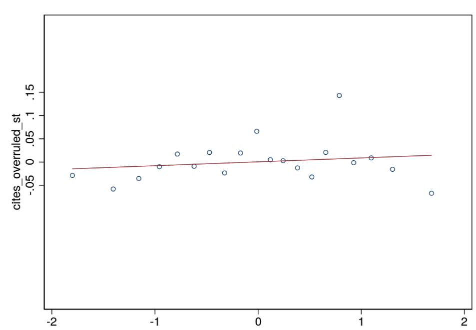
(e) Quoted-By Citations Per Opinion



(f) Negative Citations Per Opinion

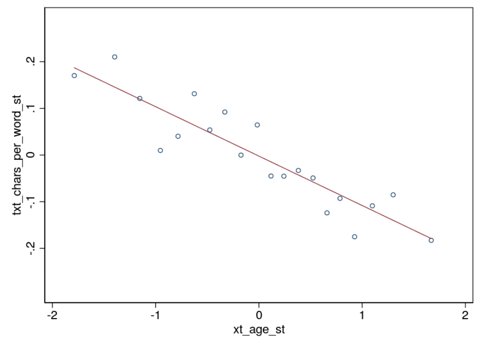


(g) Total Citations

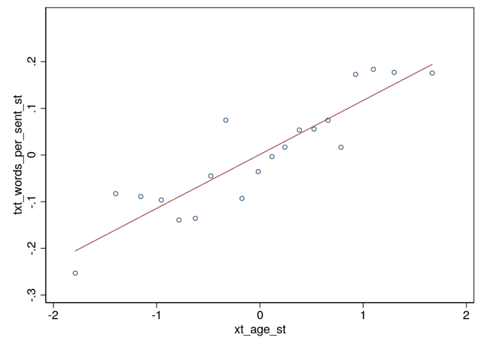


(h) Rate Overruled

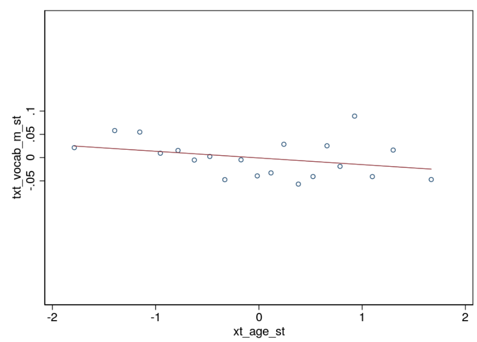
Figure 11: Performance-Age Profile (3), within State-Year



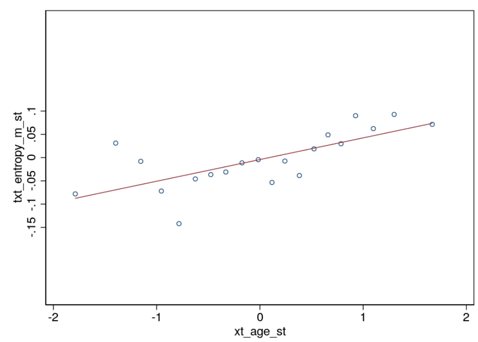
(a) Word Length



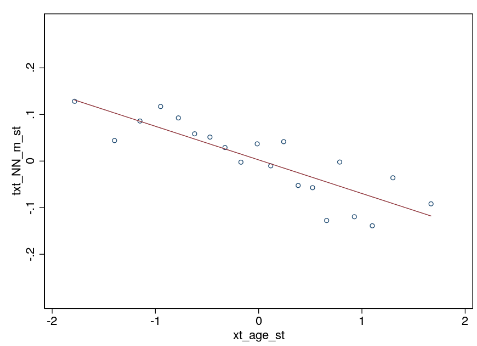
(b) Sentence Length



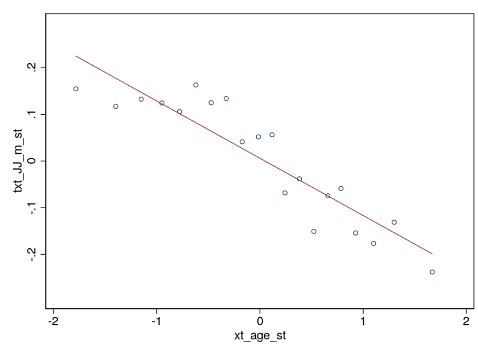
(c) Vocabulary Size



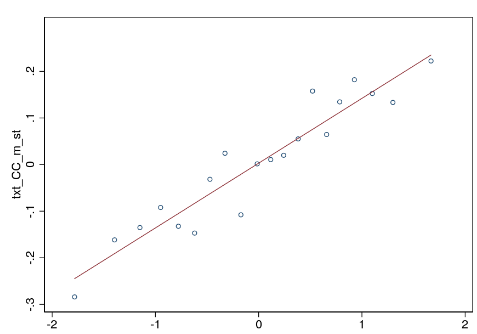
(d) Text Entropy



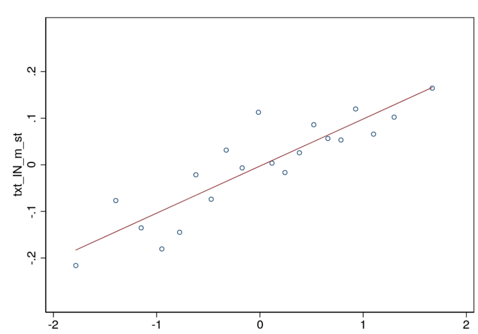
(e) Noun Frequency



(f) Adjective Frequency



(g) Conjunction Frequency



(h) Preposition Frequency

Table 4: 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)

Tables 3 and 4 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.

9.1 Mandatory Retirement and Exit Decision

Figure 12 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 13 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.

Figure 14 looks at how judge retirement is related to judge longevity, separately for mandatory retirement (left panel) and voluntary retirement (right panel). The figure

Figure 12: Retirement Rates by Age, by Mandatory Retirement Age

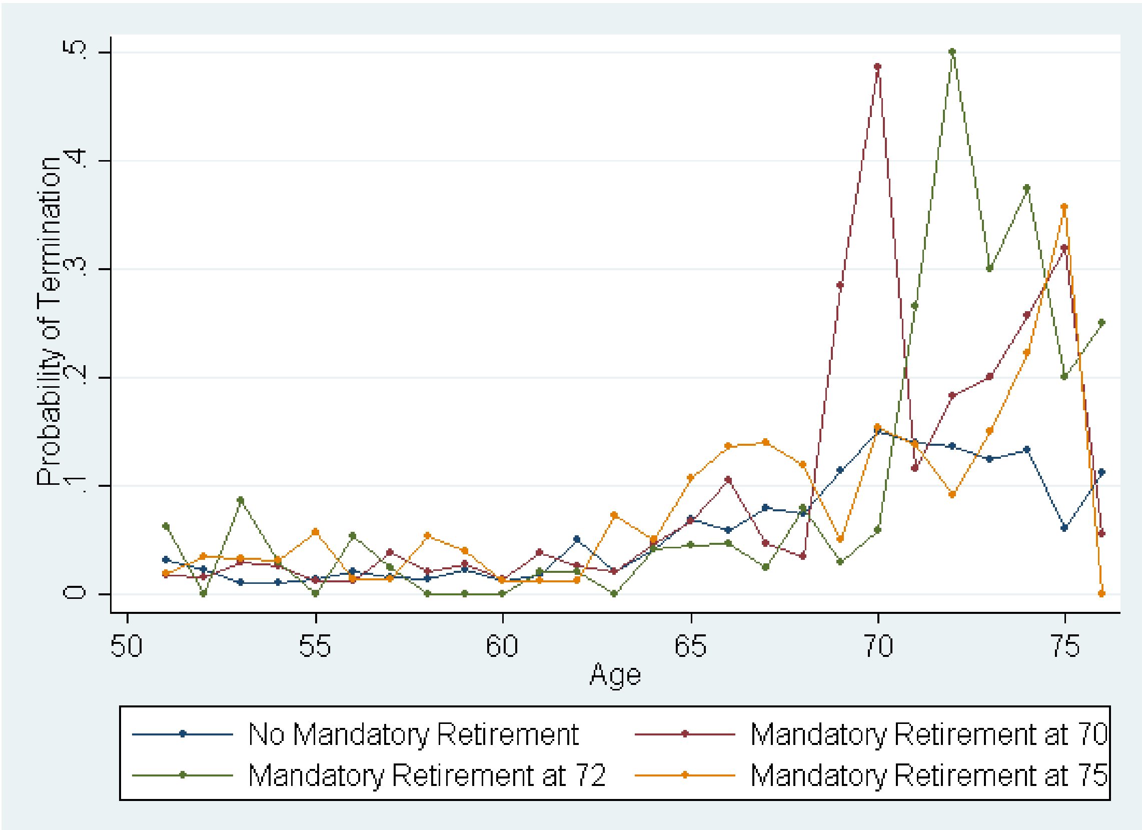


Figure 13: Age Distribution and Retirement Hazards,by Retirement Rule

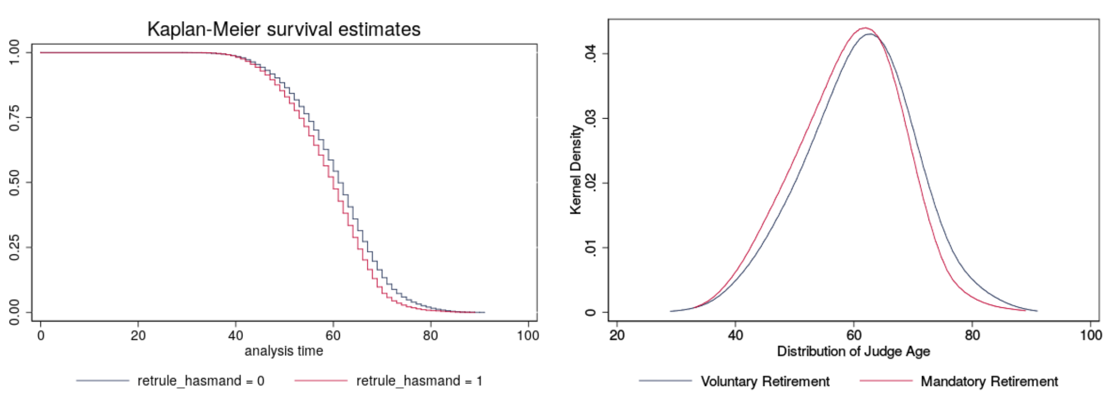
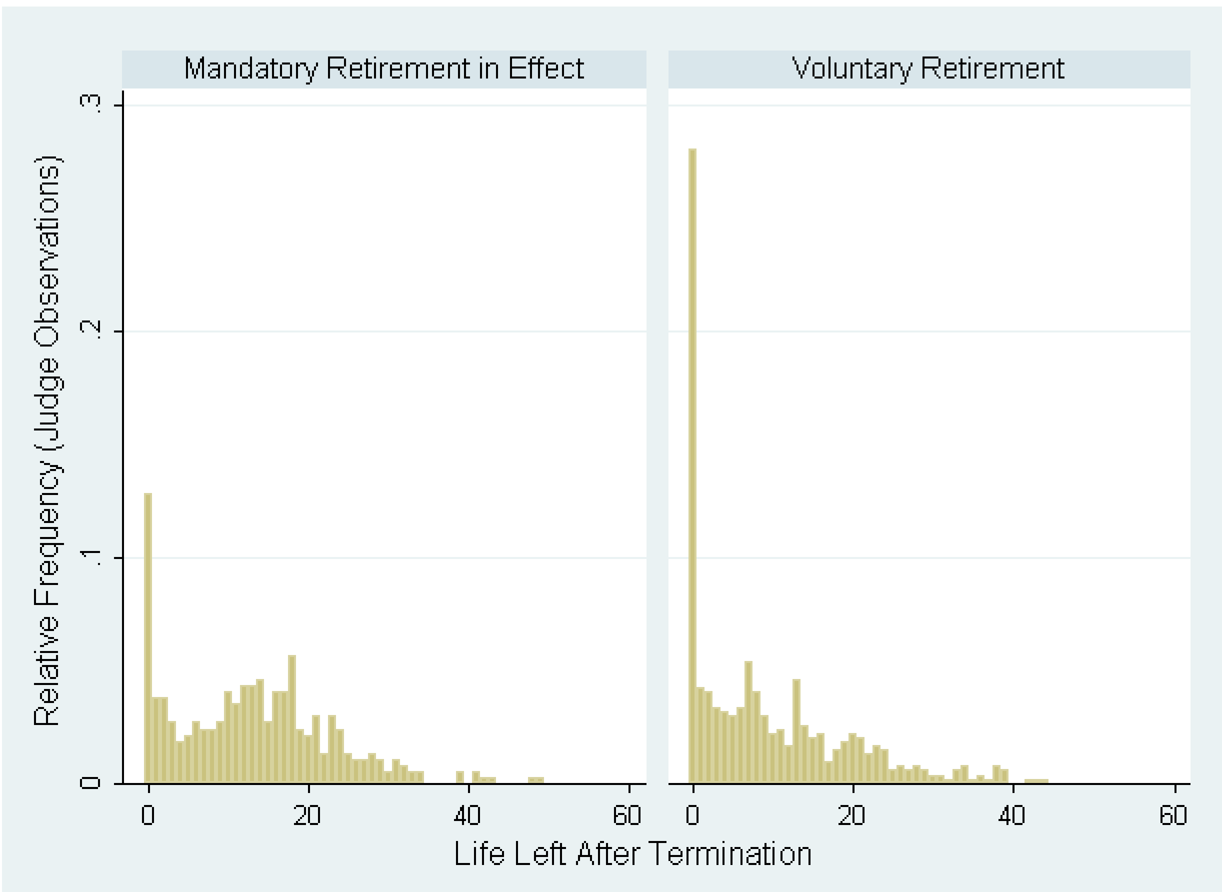


Figure 14: Distribution of Years Between Termination and Death, With and Without Mandatory Retirement



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 tend 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 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.

9.2 Mandatory Retirement Rule and Performance Trends

Now we look at how the trend in judge performance changes in the run-up to age 70, depending on whether there is mandatory retirement or not. We use the same specification as Equation 3, but we plot the residuals separately for (a) states with mandatory retirement at age 70, and (b) states without mandatory retirement.

Figure 15 shows the trends in output and quality by age for a balanced panel of judges between the ages of 50 and 70. The difference from Figure 17 is that the judges are plotted separately for states with mandatory retirement at age 70 (left panel), from states with no mandatory retirement (right panel).

In terms of case quality, there is no difference. However, for output, we see a steep decline for states with mandatory retirement. The curve is flat for states without mandatory retirement. This suggests that the mandatory retirement rule is resulting in a negative incentive effect on judge effort.

In Figure 16 we take a broader look at differences in our performance variables over the lifespan. First, we see that judges have similar trends in their workload (number of opinions) over the lifespan. In terms of output (words per year), however, there is a difference, with output increasing throughout life for mandatory retirement judges. Similarly, we see a decrease in quality for voluntary retirement judges, and increase in quality for mandatory retirement judges. This is likely due to selection, where under mandatory retirement only the best judges remain until advanced ages. This happens due to the senior judge system of active retirement, where the younger judges can choose to invite a judge back at a reduced caseload. These judges tend to have high performance. Under voluntary retirement, the lower-performance judges remain on the job longer.

We also see at the bottom of Figure 16 that language style (word length and sentence length) does not differ in its trend over the life cycle depending on the retirement rule.

Figure 15: Judge Output and Quality, Age 50-70, With and Without Mandatory Retirement

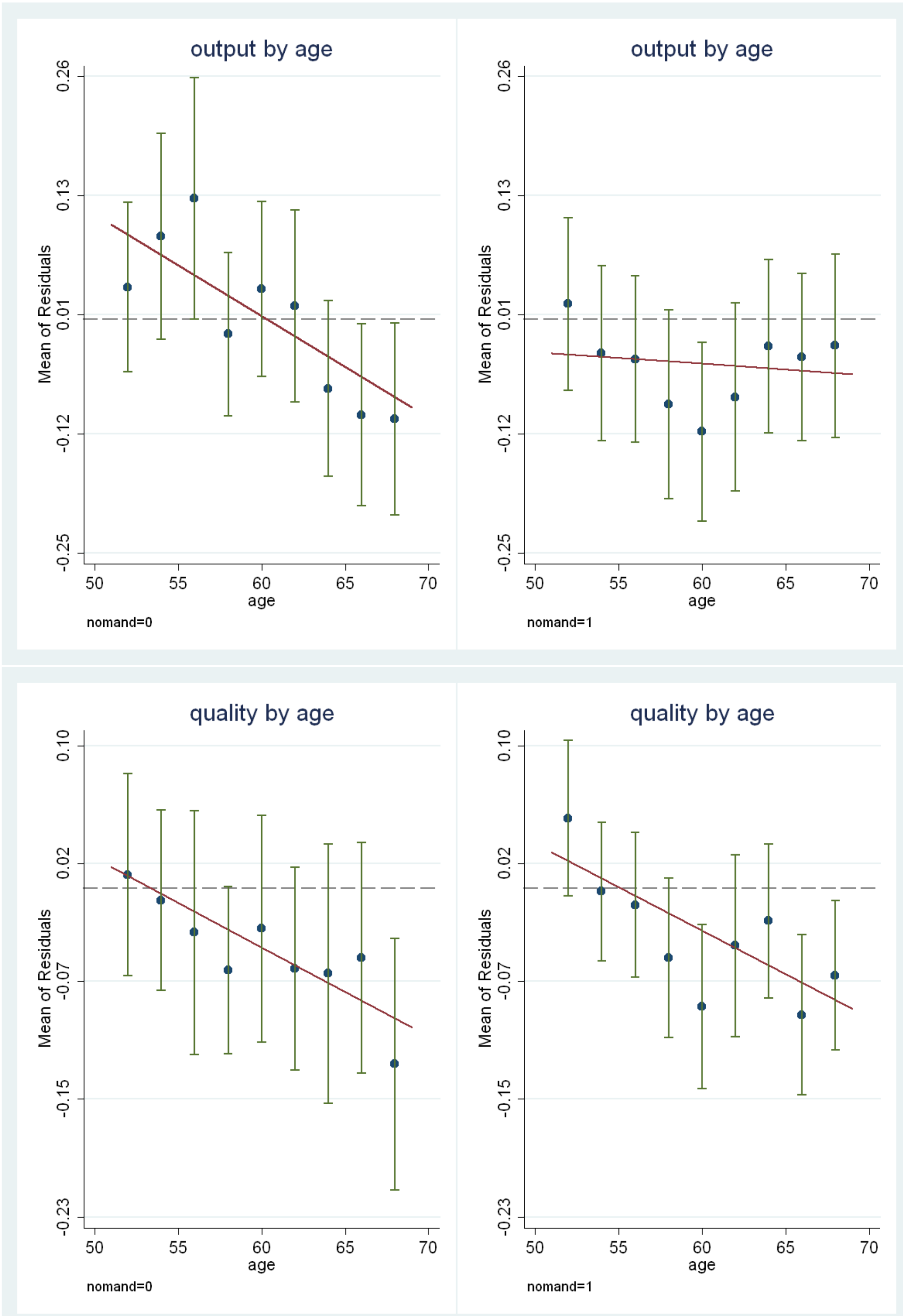
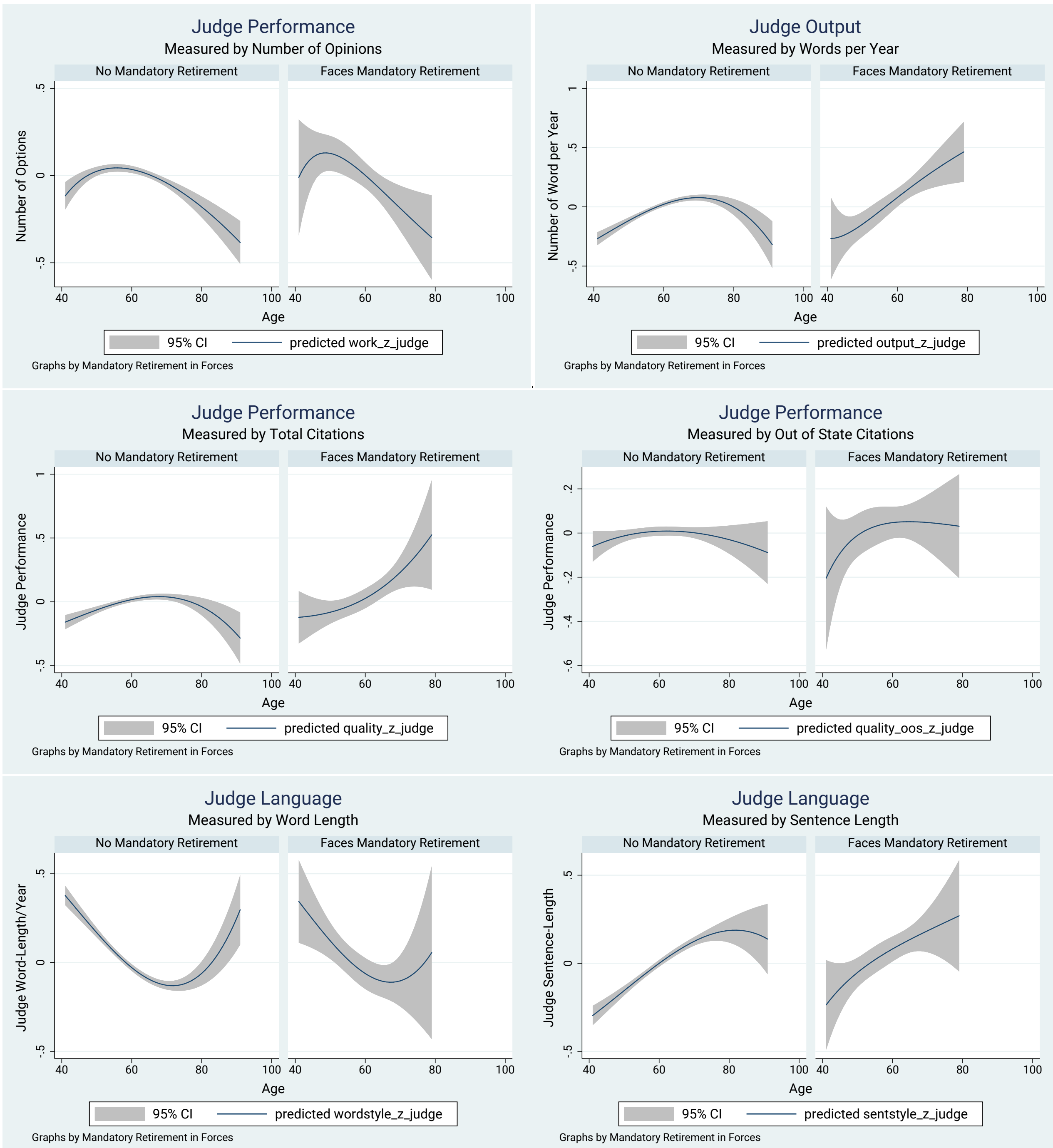


Figure 16: Judge Output and Quality over the Life Cycle With and Without Mandatory Retirement



This suggests that these language variables capture a component of physiological aging and cognitive development that does not respond to incentives.

9.3 Effect of Introducing a Mandatory Retirement Age

This section looks at the diffs-in-diffs effect of introducing a mandatory retirment age. 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 4).

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. In a handful of states, we shrank the time window if the reform occurred close to the beginning or end of the sample.[4] In the appendix we include a table using other time windows.

Formally, we estimate

$$y_{ist} = \text{YEAR}_t + \text{STATE}_s + \text{STATE}_s \times t + \bar{\rho}\bar{R}_{st} + \rho R_{st} + X'_{ist}\beta + \epsilon_{ist} \tag{4}$$

where YEAR_t is a fixed effect for the two-year period t, STATE_s is a state fixed effect, and $\text{STATE}_s \times t$ is a state-level linear time trend for state s. The term $\bar{R}_{st}$ is a dummy variable equaling one for the baseline time window of ten years before and ten years after introduction of a mandatory retirement age. R_{st} is a dummy variable for the ten

[4]These reforms are mostly enacted by voters through ballot referendums administered in November and officially going into effect the subsequent January. In these cases the dummy variable would turn on in the year following the vote. In cases where the policy is effective in the first half of the year, it is coded as turning on in that year.

Table 5: Regression Estimates, Within-State-Year Age-Performance Effect

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Opinions Written		Concurrences Written		Work Output		TOC Length	
Mandatory Retirement	0.345**	0.274*	0.0413	0.00839	-0.0794	-0.0456	-0.348**	-0.210*
	(0.112)	(0.105)	(0.0679)	(0.0692)	(0.110)	(0.0975)	(0.124)	(0.0958)
State Trends		X		X		X		X

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Citations / Op		Out-State Cites / Op		Quality (Predicted)		Total Cites	
Mandatory Retirement	-0.160*	-0.137	-0.186	-0.101	-0.237***	-0.240***	0.0601	0.0659
	(0.0774)	(0.0736)	(0.0998)	(0.0929)	(0.0594)	(0.0610)	(0.0799)	(0.0913)
State Trends		X		X		X		X

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Word Length		Sentence Length		Text Entropy		Noun Frequency	
Mandatory Retirement	0.00167	-0.0427	-0.194*	-0.247*	-0.252	-0.269*	0.215*	0.210*
	(0.0949)	(0.0836)	(0.0915)	(0.0991)	(0.134)	(0.132)	(0.0943)	(0.0970)
State Trends		X		X		X		X

N = 16032 judge-years. * p<0.05, ** p<0.01 *** p<0.001.

years after the change (with ρ measuring the corresponding causal effect of interest). X_{ist} includes other state controls when relevant. Standard errors are clustered by state.

The results for this regression are reported in Table 5. We see that after adding a mandatory retirement age, the number of majority opinions increases. There is no effect on total output, however. There is less research (Table of Cases length). There is some evidence of a decrease in work quality, especially for the text-predicted quality measure. Correspondingly, there are some changes in writing style with a decrease in sentence length, decrease in text entropy, and increase in noun frequency.

10 Conclusion

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. To this end we developed a new measure of the quality of judge writing style.

We find that physical aging is associated with a reduction in work quality over the lifespan. Mandatory retirement rules have a demotivating effect on judge performance. But this demotiving effect must be balanced against the secular decrease in decision quality due to aging.

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).

References

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

Ash, E. (2016). The political economy of tax laws in the u.s. states. Technical report.

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. (2016). 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.

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.

Belloni, A., Chen, D. L., Chernozhukov, V., and Hansen, C. (2012). Sparse models and instrument selection with an application to eminent domain. *Econometrica*, 80:2369â2429.

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.

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.

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. (2011). Insiders and outsiders: Does forbidding sexual harassment exacerbate gender inequality? Working paper, University of Chicago.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., et al. (2017). Double machine learning for treatment and causal parameters. *arXiv preprint arXiv:1608.00060*.

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.

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*.

Chun, H. and Keleş, S. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(1):3–25.

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.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1.

Gentzkow, M., Kelly, B. T., and Taddy, M. (2017). Text as data. Technical report, National Bureau of Economic Research.

Gentzkow, M., Shapiro, J. M., and Taddy, M. (2016). Measuring polarization in high-dimensional data: Method and application to congressional speech. Technical report, National Bureau of Economic Research.

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.

Lazear, E. P. (1979). Why is there mandatory retirement? *Journal of Political Economy*, 87:1261–64.

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.

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. 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.

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.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.

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;

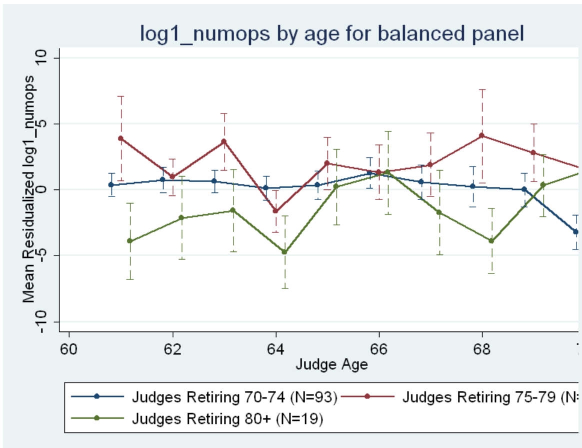
   (c) the existence in industries affecting commerce, of arbitrary discrimination in employment because of age, burdens commerce and the free flow of goods in commerce.

   (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.

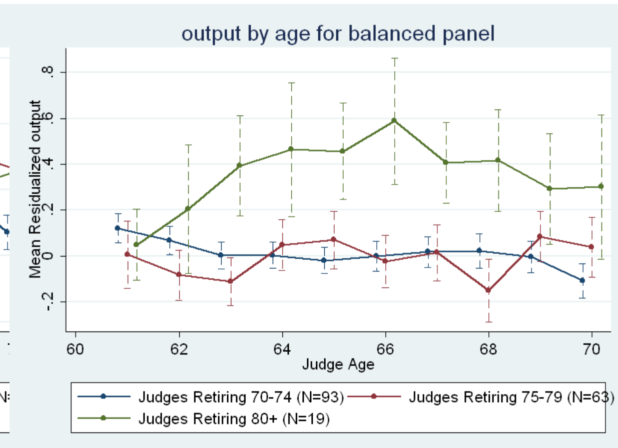
A.2 Additional Empirical Results

Figure 17 shows the major trends in output, quality, and readability by age for a balanced panel of judges. We plot the metrics separately by when the judges retired from their job. There is a clear selection effect, in the sense that the judges who last longer on the job tend to be better than judges who retire earlier.

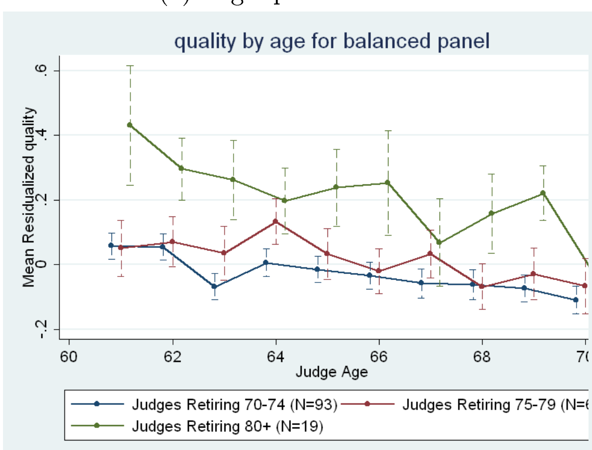
Figure 17: Judge Performance, 61-69, By Leaving Cohort



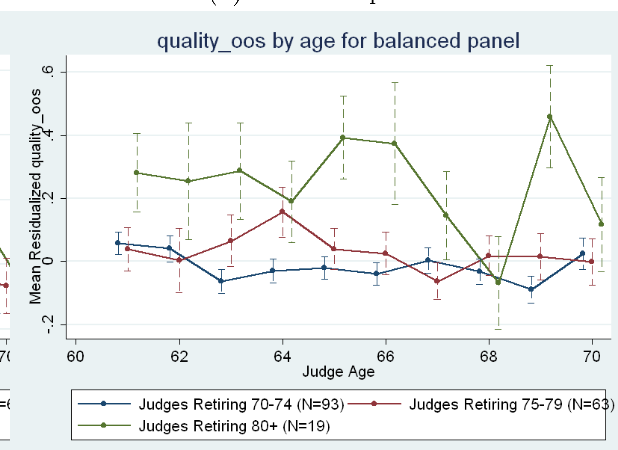
(a) Log Opinions Written



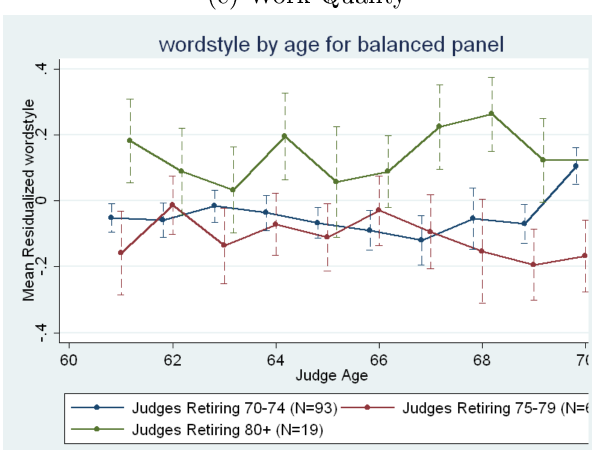
(b) Work Output



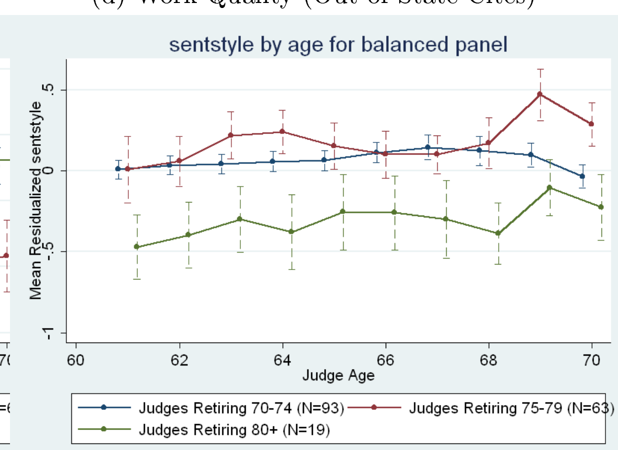
(c) Work Quality



(d) Work Quality (Out-of-State Cites)



(e) Style: Word Length



(f) Style: Sentence Length