Supplemental Material (not copyedited or formatted) for: Sara C. Benesh, David A. Armstrong II, Zachary Wallander. 2020.
"Advisors to Elites: Untangling their Effect."
Journal of Law and Courts 8(1). DOI: 10.1086/704740.

Advisors to Elites: Untangling their Effect
Online Appendix

Sara C. Benesh, David A. Armstrong II, Zachary Wallander

Appendix A. Measure of Cert-worthiness.

One of the most flexible ways to describe the relationship between a set of observed covariates and an outcome variable is with a Bayesian Adaptive Regression Tree (BART) (see Chipman, George and McCulloch (2010) for a discussion). These models allow for arbitrary non-linearity and interactivity in the covariates. As a means for describing and discovering relationships, it is an ideal tool. Software is readily available for estimating these models in R (Chipman and McCulloch, 2014). The flexibility of these models comes at a cost of both explaining the results and understanding the functional form of the relationships between the covariates and the outcome. As is usually the case, if little or no difference exists between the predictions from this model and the predictions from a simpler parametric model, the latter are preferred.

Here, we estimate both the BART model and a logistic regression of all of the covariates on whether or not a case was granted cert. Each of the variables in our model is a binary variable. Table A1 gives the proportion of ones for each covariate given the different values of the outcome.

Since the BART model does not have coefficients, the best way to compare them is to compare the predictions. First, we can calculate $\bar{p}_{BART}$ as the posterior mean probability for each of the 726 predictions and $\bar{p}_L$ as the posterior mean probability for each of the 726 predictions. The Pearson correlation between $\bar{p}_{BART}$ and $\bar{p}_L$ is 0.962 – suggesting that one is nearly a linear transformation of the other. The plot of the two predictions against each other are in Figure A1. There are 12/726 cases that do not follow the general pattern. Their docket numbers are identified beside the points. These are cases that have only middling probabilities of being granted cert in the parametric logistic regression, but were actually granted cert by the court. The BART model is better at picking up this distinction. Overall, however, the two predictions are very closely related.

The above analysis used the posterior mean probabilities. We could also evaluate the overlap between the logistic regression posterior distribution for each prediction and the corresponding posterior from the BART model (Gimenez, Morgan and Brooks, 2009; Korner-Nievergelt and Robinson 2015). This measure evaluates the extent to which the two posterior distributions overlap. The measure is bound in the interval [0,1] with higher values meaning bigger regions shared by the both posteriors. Of the 12 unusual observations, 10 had overlap coefficients less than 0.05. The other two had overlap coefficients around 0.3. Of the remaining points, the smallest overlap coefficient is 0.41. Approximately 96% of observations have overlap coefficients greater than 50%. The average overlap coefficient is around 0.69. Thus, by and large, the posterior cover quite similar ranges. The one problem with this method is that the posterior probabilities for the unusual cases are not unimodal. This method breaks down somewhat in this case.

There is also another way of measuring overlap in posterior distributions. We simply calculate $\Delta p = p_L - p_{BART}$, which is then draws from the posterior distribution of the difference. In this case, the lowest probability of overlap is around 0.16 – indicating no interesting statistical differences between the two methods.

Supplemental Material (not copyedited or formatted) for: Sara C. Benesh, David A. Armstrong II, Zachary Wallander. 2020. "Advisors to Elites: Untangling their Effect."
Journal of Law and Courts 8(1). DOI: 10.1086/704740.

These two pieces of information lead us to conclude that the two models do not produce sufficiently different predictions to encumber our analysis with the more complicated model. The predictions from the logistic regression model are a sufficiently good measure of cert-worthiness.

Table A1: Descriptive Statistics of Covariates

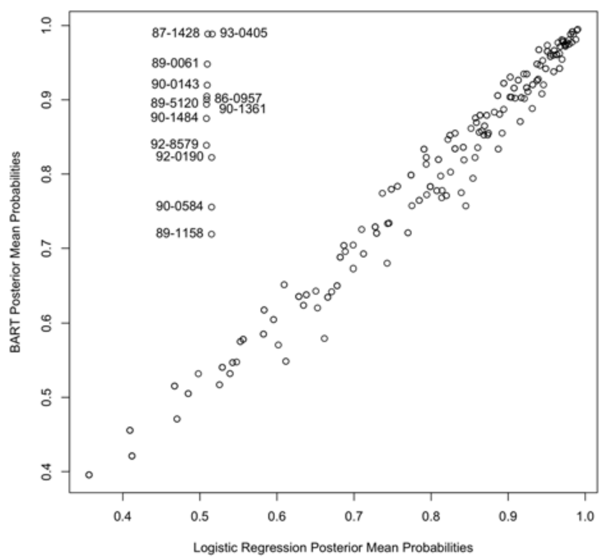
Variable	Cert Not Granted	Cert Granted
Lower court direction (0=Conservative, 1=Liberal)	0.28	0.52
Lower Court Dissent (0=No, 1=Yes)	0.27	0.32
Abortion at Issue (0=No, 1=Yes)	0.00	0.01
Death Penalty Involved (0=No, 1=Yes)	0.07	0.07
Alleged Conflict (0=No, 1=Yes)	0.63	0.79
Constitutional Claim (0=No, 1=Yes)	0.10	0.21
Lower Court Disagreement (1=Circuit Court Reversed District Court, 0=otherwise)	0.21	0.37
US is a Petitioner (0=No, 1=Yes)	0.02	0.07
True Conflict (0=No, 1=Yes)	0.17	0.54

Supplemental Material (not copyedited or formatted) for: Sara C. Benesh, David A. Armstrong II, Zachary Wallander. 2020. "Advisors to Elites: Untangling their Effect."
Journal of Law and Courts 8(1). DOI: 10.1086/704740.

Amicus Brief in Favor of Cert (0=No, 1=Yes)	0.11	0.18
Amicus Brief Against Cert (0=No, 1=Yes)	0.00	0.01

Entries are the proportion of ones of each variable given cert status

Figure A1: Posterior Mean Probabilities*



* Docket Numbers are identified for the 12 unusual cases.

By means of another robustness check, we estimate the three final models above using the BART posteriors for cert-worthiness. Table A2 shows the comparison of the two sets of coefficients. You can see that they are remarkably similar across the two methods. Thus, any misgivings regarding the discrepancies in cert-worthiness across the two methods should be allayed by the similarity evidence in Table A2 for all models.

Supplemental Material (not copyedited or formatted) for: Sara C. Benesh, David A. Armstrong II, Zachary Wallander. 2020.
"Advisors to Elites: Untangling their Effect."
Journal of Law and Courts 8(1). DOI: 10.1086/704740.

Table A2: Results of Rare Events Logistic Regression Models.

	Pool Memo Clerk		Blackmun's Clerk		Blackmun's Vote	
Variable (range/values)	Logit	BART	Logit	BART	Logit	BART
Lower-court direction {0,1}	-0.27	-0.36	**-1.21**	**-1.24**	0.07	0.02
	(0.82, 0.20)	(-0.89, 0.13)	(-1.65, -0.68)	(-1.73, -0.74)	(-0.41, 0.54)	(-0.45, 0.54)
Cert-worthiness (0,1)	**4.58**	**5.08**	**1.79**	**2.03**	**1.53**	**1.70**
	(3.19, 6.08)	(3.72, 6.67)	(0.34, 3.08)	(0.67, 3.54)	(0.25, 2.81)	(0.27, 3.14)
Abs. Distance Pool Memo Clerk to	-0.06	-0.07				
Court Median (0.01, 4.98)	(-0.27, 0.19)	(-0.29, 0.18)				
Blackmun Distance to Median			0.21	0.19	-0.08	-0.07
(1.682, 2.547)			(-0.53, 0.91)	(-0.59, 0.91)	(-0.79, 0.52)	(-0.79, 0.66)
Pool Memo Writer			**4.17**	**4.17**	0.22	0.23
Recommendation {0,1}			(2.91, 5.69)	(2.97, 5.74)	(-1.06, 1.40)	(-1.04, 1.44)
Distance of Blackmun to Pool			0.21	0.21	-0.08	-0.08
Memo Clerk (0.041, 4.560)			(-0.19 0.60)	(-0.18, 0.60)	(-0.42, 0.24)	(-0.42, 0.22)
Distance Blackmun-Pool Memo			**-0.53**	**-0.53**	-0.12	-0.13
Clerk x PMC Recommendation			(-1.02, -0.05)	(-1.00, -0.05)	(-0.57, 0.28)	(-0.52, 0.33)
Blackmun's Clerk's					**2.31**	**2.29**
Recommendation {0,1}					(1.82, 2.75)	(1.80, 2.73)
Intercept	**-4.30**	**-4.62**	**-4.60**	**-4.74**	**-3.11**	**-3.16**
	(-5.46, -3.38)	(-5.79, -3.50)	(-6.80, -2.60)	(-6.85, -2.59)	(-4.82, -1.19)	(-4.91, -1.17)

N for all models is 625.
Main entries are Rare-events Logistic Regression coefficients (i.e., posterior mean values) with 95% HPD Interval below in parentheses.
Bold coefficients are those that have at least 95% of their posterior distribution on the same side of zero. This would be similar to identifying one-tailed significance at 5% in a frequentist model.

Supplemental Material (not copyedited or formatted) for: Sara C. Benesh, David A. Armstrong II, Zachary Wallander. 2020. "Advisors to Elites: Untangling their Effect."
Journal of Law and Courts 8(1). DOI: 10.1086/704740.

Appendix B.

Recall, from fn. 23 in the main text, that we are estimating:

$$p(\beta|x,y) = \int\int p(\beta|x,y,\hat{c},\tau)p(\tau)p(\hat{c}|c,Z)d\tau d\hat{c}$$

with Monte Carlo integration to account for uncertainty in τ and cert-worthiness where ĉ is our variable of cert-worthiness (predictions from a logistic regression model where the actual cert decision c, is the outcome), Z is the design matrix for the set of variables thought to influence cert-worthiness, y is the actor's decision, and x is the set of other model regressors.

Our model specification for the Blackmun vote model includes lower court direction, the distance from Blackmun to the Court median and cert-worthiness. In addition, we want to evaluate whether the pool memo writer's recommendation matters (and if so, if it matters conditionally on the memo clerk's justice's distance to Blackmun). So, in addition to the covariates above, in model B1, we include Blackmun's clerks's recommendation (BC), the memo clerk's recommendation (MC), the distance from the memo clerk's justice to Blackmun (MD) along with all of the two way and three-way interactions ($MC \times BC, MC \times MD, BC \times MD, BC \times MC \times MD$). In Model B2, in addition to the covariates mentioned above, we include only ($BC, MC, MD, MC \times MD$). So, in total, there are four fewer terms in model B2 than model B1.

Since we are not estimating a single model, rather we are estimating a number of models to address and account for uncertainty in τ and $\hat{c}$, we cannot use the conventional likelihood ratio or Wald test to evaluate differences across nested model specifications.[1] We must, then, appeal to a different testing mechanism. First, we gather the log-likelihood from each model for each iteration of the Monte Carlo simulation. Under the null hypothesis, two times the difference between these two should follow a χ^2 distribution with 4 degrees of freedom. The observed χ^2 values range from around the 0^{th} to the 60^{th} percentiles of the theoretical distribution under the null hypothesis, so it suggests that there is no interesting difference between the two models.

Finally, we could look at the individual log-likelihoods (akin to what Clarke (2007) suggests). Here, we see that all of the simulations more of the individual log-likelihoods favored the simpler model rather than the more complex model, though in some the difference was quite small. Under the null hypothesis that the two models are equally good (i.e., that roughly half of the individual log-likelihoods are higher for each model), we would expect at least 46% of the observations to be better predicted by each model (i.e., no more than 54% of observations having higher log-likelihoods for the better model). The more complicated model

[1] Using posterior means of τ and $\hat{c}$ as plug-in estimates of the variable quantities, we could use the conventional methods of nested model comparison. Here, the likelihood ratio test has a $\chi^2 = 6.41$ on 4 degrees of freedom with a p-value of 0.093. The Wald test has $\chi^2 = 5.84$ on 4 degrees of freedom with a p-value of 0.211. Thus, the two models are not interestingly different. We get a slightly better estimate by using the posterior distribution of the coefficients from the full model and assuming the posterior distribution is multivariate normal. In this case, the Wald test χ^2 is 6.13 on 4 degrees of freedom with a p-value of 0.189. Further, the predictions from these two models correlate at 0.983.

Supplemental Material (not copyedited or formatted) for: Sara C. Benesh, David A. Armstrong II, Zachary Wallander. 2020. "Advisors to Elites: Untangling their Effect."
Journal of Law and Courts 8(1). DOI: 10.1086/704740.

has a smaller proportion than .46 of higher individual log-likelihoods around 24% of the time.

This suggests that, on average, we're better off using the simpler model.

References

Chipman, H. and McCulloch, R. (2014) BayesTree: Bayesian Adaptive Regression Trees. R package version 0.3-1.2, url: http://CRAN.R-project.org/package=BayesTree.

Chipman, H., George, E., and McCulloch R. (2010) Bayesian Additive Regression Trees. *The Annals of Applied Statistics*, 4,1, 266-298

Clarke, Kevin. 2007. "A Simple Distribution-Free Test for Nonnested Hypotheses." *Political Analysis* 15(3): 347-363.

Gimenez, O, Morgan, B.J.T. and Brooks, S.P. (2009) "Weak identiability in models for mark recapture-recovery data. Modelling Demographic Processes in Marked Populations" in Thompson, D., Coch, E. and Conroy, M. eds. Modeling Demographic Processes in Marked Populations. Springer.

Korner-Nievergelt, F. and Robinson, R. 2015. birdring: Methods to Analyse Ring Re-Encounter Data. R package version 1.3. URL: http://CRAN.R-project.org/package=birdring