ISSN 1936-5349 (print)
ISSN 1936-5357 (online)

HARVARD

JOHN M. OLIN CENTER FOR LAW, ECONOMICS, AND BUSINESS

NO, JUDGES ARE NOT INFLUENCED BY OUTDOOR TEMPERATURE (OR OTHER WEATHER): COMMENT

Holger Spamann

Discussion Paper No. 1036

07/2020

Harvard Law School
Cambridge, MA 02138

No, Judges Are Not Influenced by Outdoor Temperature (Or Other Weather): Comment

Holger Spamann [1]

Harvard Law School

July 7, 2020

[1]hspamann@law.harvard.edu. Thanks to Colby Wilkinson for downloading and cleaning the weather and pollution data, to Alex Perrault for first drawing my attention to the Philadelphia, San Antonio, and San Diego errors, and to Anthony Heyes for discussing their data with me. The discussion of coding errors, especially section 2, responds to a request by the editor and reviewers. This work was mostly done while I was a fellow at the Wissenschaftskolleg zu Berlin; I am very grateful for the hospitality.

Abstract

Heyes and Saberian (AEJ-AE 2019) estimate from 2000-2004 data that outdoor temperature reduces U.S. immigration judges' propensity to grant asylum. This estimate is the result of coding and data errors and of sample selection. Correcting the errors reduces the point estimate by two thirds, with a wide 95% confidence interval straddling zero. Enlarging the sample to 1990-2019 flips the point estimate's sign and rules out the effect size reported in Heyes and Saberian with very high confidence. An analysis of all criminal sentencing decisions by U.S. federal district judges 1992-2003 yields no evidence of temperature or other weather effects either.

1 Introduction

Heyes and Saberian 2019 (American Economic Journal: Applied Economics 11:238-265; hereinafter AHSS) estimates from immigration court cases in January 2000 through August 2004 that outdoor temperature reduces U.S. immigration judges' propensity to grant asylum. Regressing grant decisions on weather (temperature, dew point, precipitation, pressure, wind, cloud cover) and various controls including pollution (O_3, CO, $PM_{2.5}$) and spatial, temporal, and judge fixed effects, AHSS finds that every 10° F increase in temperature causes a 1.1 percentage point reduction in grants, which is 6.55% of the baseline grant rate. AHSS draws the obvious troublesome conclusion for the ideal of justice.

This comment revisits this result from three angles. First, the comment shows that AHSS's estimate is the result of coding and data entry errors, in particular the mismatching of weather measurements to location and time, the omission of 2001 data, and the use of case completions other than judicial grant/deny decisions. Correcting these errors reduces the point estimate by two thirds to 0.4 percentage points, with a standard error of 0.4.

Second, the data are actually available for 1990-2019, six times AHSS's sample period. In this larger sample, the point estimate is a 0.3 percentage point *increase* in grants for every 10° F increase in temperature. The confidence interval may or may not include zero, depending on how it is calculated, but it certainly excludes an effect of the absolute magnitude reported in AHSS, let alone with the same sign.

Third, to probe external validity, this comment analyzes all criminal sentencing decisions by U.S. federal district judges 1992-2003. (AHSS uses California parole board decisions, but those are not made by judges.) They yield no evidence of an effect of temperature or other weather either, and can again rule out an effect of the absolute size reported in AHSS with very high confidence.

All data and scripts used in this paper are available at `https://doi.org/10.7910/DVN/3LOR3R`.

2 Errors in AHSS

AHSS's data and code are publicly available on the journal's website.[1] Running AHSS's "regression.do" on their final data set "matched.dta" reproduces the point estimates and other statistics of AHSS's "preferred specification".[2] This is shown here in model 1 of table 1 only for the temperature coefficient and the joint F-statistic for weather, which will be the sole focus of discussion.[3] "Regression.do" fails to impose AHSS's (p. 247) purported restriction of weather measurements to within a 20 mile radius around the courthouse, but correcting the code on this point does not meaningfully affect the results (model 2).[4]

[1] `https://www.aeaweb.org/doi/10.1257/app.20170223.data`.

[2] AHSS's "preferred specification" is model 1 of AHSS's table 2.

[3] The F-statistic of 3.75 produced by AHSS's publicly posted data and code and reported here differs slightly from the value that AHSS reports itself (3.41).

[4] The correction assumes that "matched.dta"'s unlabeled variable "distance" records the distance to the pertinent hourly weather measurements. The issue cannot be fully resolved from AHSS's publicly posted data assembly code because it does not run and thus does not produce "matched.dta". None of this matters from model 3 onwards, which use data re-assembled with scripts written specifically for this comment.

[Table 1 about here.]

There are errors in "matched.dta" itself, however, that do have a meaningful effect on the results. The most serious issue is that half of AHSS's observations are not judicial grants or denials of asylum, and a quarter are not judicial decisions at all. As will be seen at the end of this section, this issue considerably affects the results. Since addressing the issue requires resourcing the data, however, other issues will be discussed first so as to isolate the effects of various fixes.

The second most serious issue in "matched.dta" is that weather and pollution measurements are mismatched to locations and time. Several courts have data from far-away cities (e.g., the immigration court in Arlington VA has data from Arlington TX).[5] Weather measurements are not adjusted from GMT to local time and thus record weather during the preceding night rather than AHSS's (p. 247) preferred measurement window from 6am to 4pm.[6] Fixing these errors reduces the estimated coefficient by one third (model 3).

The third serious issue in "matched.dta" is that a large amount of pollution data is missing, with dramatic consequences for sample size because AHSS drops all observations with missing data for any variable (complete case approach). Carbon monoxide (CO) is missing for all of 2001 in "matched.dta" and in AHSS's raw data, causing AHSS to lose all observations in 2001, i.e.,

[5] Other mismatches of courts to location are Oakdale (LA) to Oakdale (WI), Otay Mesa (CA) to Philadelphia, Philadelphia to Phoenix, San Antonio to San Diego, San Diego to San Francisco, and Elizabeth (NJ) to Tennessee. The issue originates with AHSS's "courtgps.dta".

[6] In "matched.dta", the temperature averaged from 6am to 4pm is on average lower than the temperature averaged over the entire day or even just from midnight to 6am.

20% of its initial sample of 269,756 observations.[7] The remaining pollution data appears to be complete in "matched.dta", but this is misleading because much of it is from remote locations much further than 20 miles from the courthosue and arguably should not be used. In particular, the median distance for measurement of particulate matter less than 2.5 microns in width ($PM_{2.5}$) is 73 miles. Complete data measured at 20 miles or less is available for only 70,843 observations. Such extreme data loss is hardly an acceptable price to pay for the inclusion of pollution controls. Wald tests yield no evidence whatsoever that the pollution controls have an effect on the outcome or, more to the point, that their exclusion leads to omitted variable bias for the weather estimates.[8] To patch this issue, model 4 and all subsequent models restrict pollution measurements to 20 miles but replace missing pollution values with zeroes while adding dummies indicating missingness, which attenuates any remaining omitted variable bias (Jones 1996). Model 4 and subsequent models also use for each combination of city, year, and variables (CO, O_3, $PM_{2.5}$) or sets of variables (weather) the one station with the most complete coverage, eliminating noise from local variation when the closest reporting measurement station changes from day to day. As expected, model 4's improvements increase the sample size and reduce the

[7]The initial sample refers to AHSS's raw "asylum.dta" file; "matched.dta" contains 269,269 observations. AHSS's title and regressions report 207,000 observations.

[8]In model 3, the joint Wald test statistic for the three pollution coefficients is $\chi^2(3) = 5.96, p = 0.11$, and the Wald test statistic for a change in the weather coefficients from omitting the pollution controls is $\chi^2(6) = 5.58, p = 0.47$. The statistics are even lower, and the p-values higher, when restricting the sample to the 70,843 complete observations with measurements taken at 20 miles or less.

standard error relative to model 3. But they reduce the point estimate even more.

This is as far as one can go with AHSS's publicly posted data. But as already mentioned, there is an even more serious issue, which is that half of AHSS's sample are not judicial decisions on the merits (i.e., to grant or deny asylum, AHSS p. 244). While AHSS's title references "207,000 Court Cases" and "matched.dta" contains 269,269 cases for January 2000 through August 2004, U.S. immigration judges only decided 149,164 asylum applications on the merits during fiscal years 2000-2004 (Executive Office for Immigration Review 2005, p. K2 (EOIR)). The official numbers (ibid p. K4) suggest, and AHSS's authors have confirmed in private communication, that AHSS's data contain all completions of asylum proceedings, including around 20% withdrawals and 7% abandonments by the *applicant*. This at least introduces noise and, worse, may make it impossible to interpret the estimates as evidence of *judicial* behavior. The issue cannot be fixed using AHSS's publicly posted data because they do not contain a variable distinguishing the different types of completion.

However, the requisite data are available directly from their ultimate source, the EOIR. AHSS's data come from the defunct website asylumlaw.org (p. 244), but asylumlaw.org in turn obtained the data from the EOIR.[9] The EOIR provided these data in response to individual Freedom of Information Act requests (e.g., Chen et al. 2016) and, since June 2018, makes them openly available online.[10] The data obtained directly from EOIR will henceforth be

[9] See data note at the bottom of `https://web.archive.org/web/20050429204003/http://www.asylumlaw.org/legal_tools/index.cfm?fuseaction=showJudges2004`.

[10] `https://fileshare.eoir.justice.gov/FOIA-TRAC-Report.zip`. The data are up-

referred to as EOIR data.

To show that the change in data source per se does not drive the subsequent results, model 5 first re-estimates model 4 with the exact same sample and specification but substituting the EOIR data for AHSS's data. The number of observations and cities is somewhat larger because model 5 also replaces AHSS's weather and pollution data with a fresh download from the NOAA ISD Lite and EPA AQS data bases, respectively. The point estimate stays virtually unchanged from model 4 while the R^2 and standard error slightly increase.

Model 6 applies the crucial fix: it limits the sample to judicial grants or denials of asylum. Now the point estimate shrinks by 18% while the standard error increases to almost the same size. The 95% confidence interval still includes AHSS's point estimate of -1.075 but it also contains a large swathe on the other side of zero. The F-statistic of the joint null hypothesis for all six weather variables, which was large in AHSS, is now very small. The bottom line is that after correcting the coding and data entry errors in AHSS, AHSS's "preferred specification" yields no evidence of an effect of outside temperature or other weather on judging in AHSS's sample period, January 2000 through August 2004.

dated monthly. This comment uses the July 2019 release, which contains data through July 2019.

3 Extended Sample Period 1990-2019

The EOIR data allow enlarging the sample period to the years 1990-2019.[11] Conceivably, the sixfold increase in sample size could reveal subtler effects than could be detected in the shorter sample. In actuality, the larger sample yields no evidence of a temperature or other weather effect either, and confidently rules out an effect of the absolute magnitude estimated in AHSS.

Models 7 and 8 are extensions of models 5 and 6, respectively, to the larger sample. When the sample is all case completions as (erroneously) in AHSS, the estimate is now a precisely estimated zero: model 7's 95% confidence interval is $[-0.2, 0.2]$ when clustering on city-month as in AHSS, i.e., model 7's upper 95% confidence bound on the absolute effect size is one fifth of AHSS's absolute point estimate. When the sample is only judicial grants and denials (model 8), the point estimate is larger in absolute value than in model 7 but still only one fifth of the size estimated in AHSS and with the opposite sign. The upper 95% confidence bound of 0.5—clustering on city-month—is half of the absolute effect size estimated in AHSS.

Model 9 makes two final improvements to the specification. First, it uses a fiscal instead of calendar year fixed effect to better account for the possibility of changing reporting conventions. Second, it uses the exogenous latest hearing date rather than the endogenous case completion date, excluding from the sample the 10% of cases with a hearing after the completion date.[12]

[11] The EOIR data begin around 1987 but coverage is incomplete until 1990.

[12] A reviewer advises that asylumlaw.org's and hence AHSS's date variable is the completion date. In half the cases decided on the merits, latest hearing and completion dates coincide because the judge decides the case orally at the end of a hearing and formally completes the case. However, in 40% of the cases, the completion date is after the latest

These changes increase the point estimate, shrink the standard error, and increase the R^2 by 10% each.

Before discussing this final result, a note on inference. Like AHSS, table 1 displays in parentheses conventional "sandwich" standard errors clustered at the city-month level. However, there are two concerns about the validity of these standard errors and resulting inference. First, the clusters are of extremely unequal size because immigration courts have vastly differing case loads: a quarter of all cases are heard in New York City, and another tenth each in Los Angeles and Miami. For example, in model 9, the 12 largest of the 661 city-month clusters—all 12 from New York City—contain one quarter of all observations, and the largest 5% of the clusters contain almost half of all observations. With such unequal cluster sizes, conventional inference can be grossly misleading (Carter et al. 2017; MacKinnon and Webb 2017). The wild bootstrap-t (Cameron et al. 2008; Roodman et al. 2019) provides a superior alternative (MacKinnon and Webb 2017). Table 1 shows wild bootstrap-t 95% confidence intervals clustered on city-month in square brackets. Second,

hearing. One reason this can happen is because, after decision on the merits, formal completion of the case is delayed by formalities, in which case the completion date is noisier than the hearing date. Alternatively, the judge can take the case into consideration and decide in writing after the hearing, in which case the judge chooses the completion date and weather on that date is no longer plausibly exogenous. In either scenario, it is preferable to use the hearing date, which is set a long time in advance (AHSS p. 245). The only problem with the hearing date is that, after completion of the asylum case, a new hearing date can be set upon a motion to reopen or reconsider (8 C.F.R. 1003.23) and in some other circumstances, in which case the new hearing date overwrites the hearing date at which the asylum claim was decided. This is the reason to exclude the 10% of cases with a latest hearing after the completion date.

however, clustering on city-month is insufficient because weather is serially correlated and hence treatment assignment in one month correlated with treatment assignment in adjoining months (cf. Abadie et al. 2017). Table 1 therefore also shows wild bootstrap-t 95% confidence intervals clustered on city in braces.[13]

Regardless of how the 95% confidence interval for temperature is constructed in model 9, it always excludes the point estimate reported in AHSS (-1.075) by a large margin and even in absolute value. The temperature effect and effect size reported in AHSS can thus be ruled out with high confidence. In fact, model 9 suggests that there is no temperature or other weather effect at all. The most credible, city-clustered bootstrap 95% confidence interval for temperature includes zero, and the joint F-statistics for weather are small. Perhaps more importantly, the positive *sign* of temperature's point estimate is the opposite of what AHSS had found and had argued one should expect if an effect existed (pp. 238, 262). Finally, the existence of weather effects in climate controlled court houses seemed rather improbable a priori, and the small point estimates are much more consistent with this skeptical view than with weather effects of a meaningful size. (For calibration, the interquartile range between individual judges is over 20 percentage points even after adjusting for all the covariates of table 1, and the interquartile range for different nationalities is even larger. See Ramji-Nogales et al. 2007; Fischman 2013.) To address remaining doubts, the next section turns

[13] "City" refers to the EOIR's BASE_CITY variable, which might be better described as "Immigration Court" because there are two in some cities like New York City, Los Angeles, or Miami.

to another data set for a fresh look at the same phenomenon.

4 External Validation: Sentencing

AHSS explicitly considers asylum adjudication merely a "test-bed" to "investigate the link from outdoor temperature to decisions made by experienced professional decision-makers." If such a link existed, one should be able to observe it also with other "experienced professional decision-makers", above all other judges. Criminal sentencing decisions by U.S. federal district judges fulfill AHSS's (pp. 239-40) criteria for an "ideal test-bed": (1) high-stakes decisions (2) made by experienced professionals (3) operating in a climate-controlled indoor environment (4) generating high-frequency data on pre-scheduled dates with a rich set of covariates (to make cases roughly comparable).[14]

The United States Sentencing Commission (USSC) makes sentencing decisions available through the present but provides decision dates only for fiscal years 1992 through 2003, comprising 610,687 cases. The data contain a rich set of case and defendant characteristics including the offense level and criminal history scores that determine the USSC's sentencing grid (see Cohen and Yang 2019 for a description of the data and their institutional background). One or more of these covariates are missing for 15% of these

[14]Sentencing decisions are partially constrained by sentencing guidelines. But judges retain plenty of discretion. The lower bound of the recommended range is at least 20% below the upper bound. Moreover, departures from the range are frequent. Economists have studied how judges use this discretion (e.g., Fischman and Schanzenbach 2012; Cohen and Yang 2019).

cases, and weather data is missing for some others, such that the useable number of observations from the USSC is 464,518.

The main outcome variable is whether the judge sentenced the defendant to prison time, and if so, for how long. To capture these two dimensions, table 2 shows regression results for the binary imprisonment decision (model 1) and, for those defendants who did receive prison time, the natural logarithm of the sentence length (model 2). Weather and pollution regressors enter, and standard errors and confidence intervals are calculated, as in models 4-9 of table 1. The other specification choices follow Yang 2015; Spamann 2018.

[Table 2 about here.]

For the probability of imprisonment, temperature's effect is fairly precisely estimated to be zero: 0.12 percentage points per 10° F, with a 95% bootstrap confidence interval of only $[-0.16, 0.35]$ percentage points even when clustering on city. For sentence length conditional on imprisonment, the point estimate is almost exactly zero (0.01 percentage points per 10° F), with a bootstrap confidence interval of $[-0.72, 0.61]$ when clustering on city. All weather variables are jointly statistically insignificant. In sum, there is even less evidence for a temperature or other weather effect in sentencing than in asylum adjudication.

5 Conclusion

This comment shows that effects of temperature and other weather on judging are at most small and probably non-existent in both asylum and sentencing. Contrary findings in AHSS resulted from errors in coding and data entry. Like other ostensible extraneous influences on judging (cf. Weinshall-Margel and Shapard 2011; Spamann 2018), the weather effect turns out to be spurious. Reports of such influences should be read with circumspection.

Judicial decision-making is not perfect. There is evidence of biases and inconsistencies in judicial decisions in general (e.g., Rachlinski and Wistrich 2017; Spamann and Klöhn 2016) and in U.S. asylum adjudication (Ramji-Nogales et al. 2007; Fischman 2013) and federal sentencing (e.g., Yang 2015; Cohen and Yang 2019) in particular. It is essential, however, to gauge the extent of the problem accurately. It is counterproductive and dangerous to paint the justice system worse than it actually is.

References

Abadie, Alberto, Susan Athey, Guido Imbens, and Jeffrey Wooldridge (2017). *When Should You Adjust Standard Errors for Clustering?*

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2008). "Bootstrap-Based Improvements for Inference with Clustered Errors". In: *The Review of Economics and Statistics* 90.3, pp. 414–427.

Carter, Andrew V., Kevin T. Schnepel, and Douglas G. Steigerwald (2017). "Asymptotic Behavior of a t-Test Robust to Cluster Heterogeneity". In: *The Review of Economics and Statistics* 99.4, pp. 698–709.

Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue (Mar. 2016). " Decision Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires". In: *The Quarterly Journal of Economics* 131.3, pp. 1181–1242.

Cohen, Alma and Crystal S. Yang (2019). "Judicial Politics and Sentencing Decisions". In: *American Economic Journal: Economic Policy* 11.1, pp. 160–91.

Executive Office for Immigration Review, U.S. Department of Justice (2005). *FY 2004 Statistical Year Book*. Falls Church, VA.

Fischman, Joshua B. (2013). "Measuring Inconsistency, Indeterminacy, and Error in Adjudication". In: *American Law and Economics Review* 16.1, pp. 40–85.

Fischman, Joshua B. and Max M. Schanzenbach (2012). "Racial Disparities Under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums". In: *Journal of Empirical Legal Studies* 9.4, pp. 729–764.

Heyes, Anthony and Soodeh Saberian (2019). “Temperature and Decisions: Evidence from 207,000 Court Cases”. In: *American Economic Journal: Applied Economics* 11.2, pp. 238–65.

Jones, Michael P. (1996). “Indicator and Stratification Methods for Missing Explanatory Variables in Multiple Linear Regression”. In: *Journal of the American Statistical Association* 91.433, pp. 222–230.

MacKinnon, James G. and Matthew D. Webb (2017). “Wild Bootstrap Inference for Wildly Different Cluster Sizes”. In: *Journal of Applied Econometrics* 32.2, pp. 233–254.

Rachlinski, Jeffrey J. and Andrew J. Wistrich (2017). “Judging the Judiciary by the Numbers: Empirical Research on Judges”. In: *Annual Review of Law and Social Science* 13.1, pp. 203–229.

Ramji-Nogales, Jaya, Andrew I. Schoenholtz, and Philip G. Schrag (2007). “Refugee Roulette: Disparities in Asylum Adjudication”. In: *Stanford Law Review* 60 (2), pp. 295–412.

Roodman, David, Morten Ørregaard Nielsen, James G. MacKinnon, and Matthew D. Webb (2019). “Fast and wild: Bootstrap inference in Stata using boottest”. In: *The Stata Journal* 19.1, pp. 4–60.

Spamann, Holger (2018). “Are Sleepy Punishers Really Harsh Punishers? Comment on Cho, Barnes, and Guanara (2017)”. In: *Psychological Science* 29.6, pp. 1006–1009.

Spamann, Holger and Lars Klöhn (2016). “Justice Is Less Blind, and Less Legalistic, than We Thought: Evidence from an Experiment with Real Judges”. In: *The Journal of Legal Studies* 45.2, pp. 255–280.

Weinshall-Margel, Keren and John Shapard (2011). “Overlooked factors in the analysis of parole decisions”. In: *Proceedings of the National Academy of Sciences* 108.42, E833–E833.

Yang, Crystal S. (2015). “Free at Last? Judicial Discretion and Racial Disparities in Federal Sentencing”. In: *The Journal of Legal Studies* 44.1, pp. 75–111.

Table 1: Asylum

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Probability of Grant (%)								
Temperature (10° F)	-1.07	-1.15	-0.67	-0.48	-0.48	-0.39	-0.02	0.27	0.30
	(0.27)	(0.29)	(0.27)	(0.24)	(0.29)	(0.36)	(0.10)	(0.14)	(0.13)
	[-1.63,-0.52]	[-1.75,-0.56]	[-1.22,-0.10]	[-0.97,-0.00]	[-1.08,0.11]	[-1.12,0.34]	[-0.23,0.19]	[-0.01,0.56]	[0.05,0.55]
	{-1.88,-0.31}	{-2.04,-0.32}	{-1.45,0.20}	{-1.14,0.04}	{-1.29,0.23}	{-1.50,0.66}	{-0.25,0.43}	{-0.24,0.79}	{-0.01,0.82}
R^2	0.17	0.17	0.17	0.17	0.20	0.25	0.23	0.33	0.37
N	206,924	195,589	202,963	248,587	267,222	138,938	1,487,308	695,084	606,754
City-Months	514	360	344	348	558	548	667	663	661
Cities	43	30	29	29	49	47	58	57	56
F (all weather)	3.75	3.72	2.32	1.82	0.89	0.31	0.11	0.86	1.41
F (temp., clouds, rain)	5.74	5.79	2.98	2.39	1.09	0.53	0.05	1.53	2.17
Data source	AHSS final	AHSS final	AHSS raw	AHSS raw	Agencies	Agencies	Agencies	Agencies	Agencies
Sample years	2000-2004	2000-2004	2000-2004	2000-2004	2000-2004	2000-2004	1990-2019	1990-2019	1990-2019
Cases (completions)	All	All	All	All	All	Merits	All	Merits	Merits
Court location	AHSS	AHSS	Actual	Actual	Actual	Actual	Actual	Actual	Actual
Weather >20 mi	Yes	No	No	No	No	No	No	No	No
Weather 6am-4pm	GMT	GMT	local	local	local	local	local	local	local
Pollution >20 mi	Yes	Yes	Yes	No	No	No	No	No	No
Pollution missingness	CC	CC	CC	MD	MD	MD	MD	MD	MD
Stable stations	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes

Linear regressions with fixed effects for judge, applicant nationality, day of the week, year, city-month, and defensive application, and controls for other weather (dew point, rain, cloud cover, pressure, wind) and pollution (O_3, CO, $PM_{2.5}$). Standard errors clustered on city-month in parentheses; wild bootstrap-t (99,999 replications) 95% confidence intervals clustered on city-month (city) in brackets (braces). Data sources: "AHSS final" refers to AHSS's "matched.dta", "AHSS raw" is AHSS's publicly posted raw data, and "Agencies" refers to fresh data downloaded from NOAA ISD Lite (weather), EPA AQS (pollution), and EOIR (other variables). The sample contains data from 1/1/2000-8/31/2004 or 1/1/1990-7/31/2019, and all completions (including abandons, withdrawals, and "other" completions) or only judicial decisions on the merits (grant/deny), as indicated. Court locations (for matching courts to weather and to pollution) are as in AHSS (including errors as discussed in the main text) or the actual locations, as indicated. Weather variables are hourly measurements averaged over 6am-4pm GMT or local time, as indicated; pollution variables are measured daily. Weather and pollution measurement stations are or are not restricted to a 20-mile radius around the court house and stable within city-year (for a given variable), as indicated. Missingness in pollution variables is handled either using CC (complete case method: dropping observations with missing values) or MD (missing dummies are added while missing values are replaced with zero). Model 1 is an exact reproduction of AHSS's "preferred specification", i.e., of model 1 of AHSS's table 2.

Table 2: Sentencing

	(1) Probability of Imprisonment (%)	(2) Ln(Length of Prison Sentence)×100
Temperature (10° F)	0.12	0.01
	(0.10)	(0.26)
	[-0.07,0.31]	[-0.51,0.53]
	{-0.16,0.35}	{-0.72,0.61}
R^2	0.49	0.75
N	471,897	391,341
City-Months	1,054	1,054
Cities	88	88
F (all weather)	0.81	0.11
F (temperature, clouds, rain)	1.35	0.16

Linear regressions with fixed effects for sentencing grid cell, type of offense, whether a statutory minimum applies, number of convictions, trial vs. plea bargain, race, Hispanic, U.S. citizen, gender, education, day of the week, fiscal year, and city-month, and controls for offender age, offender age squared, other weather (dew point, atmospheric pressure, wind speed, precipitation, sky cover), and pollution (O_3, CO, $PM_{2.5}$). Weather variables are hourly measurements averaged over 6am-4pm local time, pollution variables are measured daily, and both are restricted to 20 miles around the court house. Missing values for pollution variables have been replaced with zero, and dummies for missingness added. Data sources: NOAA ISD Lite (weather), EPA AQS (pollution), and USSC (other variables). The sample includes all federal sentencing decisions in fiscal years 1992-2003 with complete data (other than pollution). Standard errors clustered on city-month in parentheses; wild bootstrap-t (99,999 replications) 95% confidence intervals clustered on city-month (city) in brackets (braces).