This Morning's Breakfast, Last Night's Game:

Detecting Extraneous Influences on Judging

Daniel Li Chen[♠] & Holger Spamann[♣]

Early Draft 4/3/2014

PLEASE DO NOT CITE OR DISTRIBUTE

[♠] ETH Zurich. chendan@ethz.ch.

[♣] Harvard Law School. hspamann@law.harvard.edu.

We are grateful to Joshua Fischman, Jaya Ramji-Nogales, Andrew Schoenholtz, and Phillip Schrag for sharing data, and to Lenni Benson, Gordon Dahl, Russell Wheeler, and especially Sue Long for guiding us to and through data. We thank Erica Larson and […] for research assistance. For helpful comments, we thank seminar audiences at the law schools of Harvard University, the University of Virginia, and UCLA. Most of this work was done while Chen, Larson, and Spamann were TRAC Fellows of the Transactional Records Access Clearinghouse (TRAC) at Syracuse University, which provided the main asylum court data used in this paper.

Abstract

We detect *intra*-judge variation in judicial decisions driven by factors completely unrelated to the merits of the case, or to any case characteristics for that matter. Concretely, we show that asylum grant rates in US immigration courts differ by the success of the court city's NFL team on the night before, and by the city's weather on the day of, the decision. Our data including half a million decisions spanning two decades allows us to exclude confounding factors, such as scheduling and seasonal effects. Most importantly, our design holds the identity of the judge constant. On average, US immigration judges grant an additional 1.5% of asylum petitions on the day after their city's NFL team won, relative to days after the team lost. Bad weather on the day of the decision has approximately the opposite effect. By way of comparison, the average grant rate is 39%. We do not find comparable effects in sentencing decisions of US district courts, and speculate that this may be due to higher quality of the federal judges, more time for deliberation, or the constraining effect of the federal sentencing guidelines.

"Judge Reid is best avoided on a Monday following a weekend in which the USC football team loses."[1]

1. Introduction

What determines judicial decisions? We would like to believe it is "the law." To be sure, the law may be hard to determine or even indeterminate (Fischman, forthcoming). The identity of the judge can thus make a large difference. This simple fact was statistically established at least a century ago (Everson 1919-1920) and triggered policy responses such as the US Federal Sentencing Guidelines. In particular, judicial decisions differ by judicial ideology, as demonstrated in later studies (surveyed in Fischman and Law 2009) and recognized in the frequent judicial confirmation battles. But while these inter-judge differences show that the meaning of "the law" is not unique in practice, they are consistent with each judge consistently applying his or her version of the law. Indeed, contemporary legal theory readily admits such variation because the correct interpretation of the law is not unique, or at least not discernible for real world judges (e.g., Dworkin 1986, Kennedy 1999).

In this paper, we provide evidence from a natural experiment for a different kind of variation. We detect *intra*-judge variation driven by factors completely unrelated to the merits of the case, or to any case characteristics for that matter. Concretely, we show that asylum grant rates in US immigration courts differ by the success of the court city's NFL team on the night before, and by the city's weather on the day of, the decision. Our data including half a

[1] Kathy Morris Wolf, California Courts and Judges (1996) p. 1020 (http://bit.ly/1k7goag, visited 3/29/2014).

million decisions spanning two decades – 22,000 on Mondays after a game – allows us to exclude confounding factors, such as scheduling and seasonal effects. Most importantly, our design holds the identity of the judge constant. On average, US immigration judges grant an additional 1.5% of asylum petitions on the day after their city's NFL team won, relative to days after the team lost. Bad weather on the day of the decision has approximately the opposite effect. By way of comparison, the average grant rate is 39%.

The available data do not allow us to determine if sports outcomes and weather influence the judge directly, or indirectly through litigant behavior or other aspects of the courthouse environment. In either case, however, our results demonstrate that case outcomes depend on more than "the law," "the facts" of the case, judicial ideologies, or even constant judicial biases, for example with respect to race. We suspect that many practitioners would not be surprised by that basic claim. The contribution of our paper, however, is to provide clear causal identification of two such factors, and to measure their magnitude. To us, the measured effect of 1.5% appears large for two reasons. First, we can only measure one emotional influence out of many. Presumably, other factors such as family problems or joys, traffic jams, or health fluctuations have an even greater influence on a judge's state of mind and thus plausibly case outcomes. If we had data on these, we would expect to find much larger effects. Second, even our estimate of NFL effects is only a lower bound. For lack of better data, we crudely allocate city to one team, neglecting split allegiances within a city, let alone diverging preferences of the judges. This introduces measurement error and biases our coefficient towards zero.

In the legal literature, Jerome Frank famously claimed that "uniquely individual factors often are more important causes of judgments than anything which could be described as political, economic, or moral biases" (Frank 1930 [2009]:114). This view is often caricatured as "what the judge had for breakfast" (Schauer 2009:129n15). We do not know if the caricature targets the absurdity of the claim or its polar opposite, triviality. With respect to football and weather at least, we demonstrate that a parallel claim is not absurd but true. Whether it is trivial or not would seem to us to depend on the magnitudes. Again, we believe that they are large.

To say that there are non-trivial extraneous influences on adjudication does not mean that we impugn the judges' integrity. Our research is inspired by many field studies that have shown humans in many settings to be influenced by seemingly irrelevant factors in general and sports outcomes and weather in particular. For example, Edmans et al. (2007) show that sports results affect stock returns, Healy et al. (2010) show that sports results influence voting in political election, and Card and Dahl (2011) show that disappointing NFL football results trigger domestic violence. These studies in turn build on laboratory experiments documenting such emotional effects [to come]. In this sense, we only show that "judges are people too." That being said, we believe that this realization may have consequences for legal system design, in particular the extent to which we want to rely on a "rational" legal process. For example, there is widespread aversion to the use of explicit randomization, be it for cost-saving or experimental purposes (cf. Fischman 2013). Such aversion may diminish once it is realized that the judicial process itself involves considerable randomness.

In law, the paper closest to ours are Danziger et al. (2011a) and another paper by one of the authors (Berdejo and Chen 2013). Danziger et al. detect time-of-day patterns in Israeli judges' parole decisions. Concretely, they show that parole approval rates drop with the time from the judges' last meal. A potential problem with this research design is that the order of prisoners' appearance before the judges and the exact time judges choose to take breaks may not be random (cf. Weinshall-Margel & Shapard 2011; Danziger et al. 2011b).

Berdejo and Chen (2013) show that US federal appeals court judges become more politicized before elections. In the two quarters before elections, these judges vote more strongly along partisan lines and dissent more often from judges of the opposite political leaning in politicized cases. One concern here is that the case composition may change before elections. Berdejo and Chen show that this is unlikely to explain the result. The present paper eliminates such concerns from the outset.

Massive in*ter*-judge variation in asylum grants has been documented by Ramji-Nogales et al. (2007), who introduced the legal literature to the asylum data. They showed that grant rates for the same applicant nationality in the same city could be anywhere between, e.g., 0 and 68% depending on the judge who heard the case. Our findings complement theirs. Their findings, while shocking, would be consistent with individual judges steadily applying the same legal philosophy – their own –, but legal philosophy differing across judges. By contrast, our finding shows that there consistency is limited even within judge.

Like other researchers interested in jurisprudential questions (Fischman, forthcoming), we use asylum data primarily because they give us high statistical power: the cases are very numerous and relatively homogeneous, and the outcomes are easy to classify into grant or no grant. With hundreds or even thousands of similar cases per judge, we can thus construct fairly precise baseline approval rates for each judge from the respective judge's own decision record. And while immigration judges are not Article III judges, asylum cases involve serious, potentially life-or-death decisions (Ramji-Nogales et al. 2007), and are at least in principle adversarial proceedings. They should thus be at least comparable to other weighty judicial decisions and shed light on the operation of the legal system.

That being said, not all courts are alike, and immigration courts are perhaps on one end of the spectrum. Their case load is very high, forcing immigration judges to make important decisions with very little time (on average 7 minutes by one estimate[2]). The Board of Immigration Appeals provides little guidance on the application of the broad standard for asylum petitions, name "reasonable fear of persecution." Finally, immigration judges may be somewhat less professional than Article III judges. This lack of time for deliberation coupled with a very open-ended decision standard may amplify emotional influences.

For comparison purposes, we therefore also consider criminal sentencing by federal district judges in 1990-2011 (mostly 1998-2010), the only other large data base of comparable judicial decisions that we are aware of. In this data set, we estimate only small effects of NFL outcomes and weather, and in fact those estimates are not statistically different from zero at any conventional level of significance. We speculate that this difference is due to federal district

---

[2] […]

judges' lower workload and higher professionalism, and/or the federal sentencing guidelines, which constrained federal sentencing from 1984 until at least 2005 (cf. Yang forthcoming).

We will explain our research design in more detail in section 2. Readers familiar with natural experiments and statistics may want to skip most of this section. In section 3, we describe our data set. In section 4, we present results for asylum decisions. Section 5 discusses these results and compares them to the results for sentencing decisions.

2. Research Design

To detect the influence of an extraneous factor, we need to measure the observed outcome against a baseline, or potential outcome: What would have happened if the extraneous factor (e.g., an NFL win) had (not) been present?

While we cannot observe the counterfactual decision in any given case, our strategy is to observe decision *rates* in groups of cases that are randomly allocated to exposure or no exposure to the factor. The basic idea is thus the same as in a medical trial. We want to compare average outcomes in a group randomly exposed to the "treatment" (the treatment group) with outcomes in a group randomly not exposed to the treatment (the control group). Random assignment supports the assertion that but for the treatment, the two groups would have exhibited the same average outcomes, subject to estimable statistical fluctuation (the standard error). The difference in average outcomes is thus an estimate of the causal effect of the treatment (the treatment effect).

To implement such a design, we need a large data set of judicial decisions paired to an extraneous factor that plausibly affects the decision yet is as good as randomly assigned. We need a factor that a priori plausibly affects the decision because otherwise hopes of detecting anything are dim, and if we did, it would most likely be random noise rather than a true effect.[3] Most legally irrelevant factors that plausibly affect the decision, however, are at least suspected to be correlated with legally relevant factors. For example, race is correlated with various socio-economic variables; in asylum cases, race might even be legally relevant in as much as it pertains to the risk of racial persecution. Factors that are clearly extraneous, on the other hand, are usually only available for a small number of cases (e.g., information about a judge's own divorce). This in turn is problematic because the effect of most extraneous factors, ours included, will likely be small relative to the many legally relevant and irrelevant factors that influence the decision yet which we as researchers do not observe. Only in a large sample can we hope that those other factors cancel out and the effect of interest becomes visible.

This demanding combination of requirements led us to consider asylum decisions (and to a lesser extent, sentencing decisions), sports, and weather.[4] We now explain the requirements and our choice in more detail.

2.1.The importance of sample size and homogeneity

[3] This can be formally derived through Bayes rule. Cf. Strnad 2007.

[4] Many people asked if we had tried a multitude of "irrelevant" factors before picking football as the most "significant." The answer is no. Our original TRAC application referenced only football. We later added weather to make sense of some confusing results relating to surprise wins and losses, which we then learnedto be influenced by weather..

To appreciate the demands on sample size, consider the following numbers. Our asylum and sentencing data sets, the largest case data sets we are aware of, contain approximately 900,000 decisions each. The relevant subset of comparable decisions after a football game, however, only comprises 58,000 and 22,000 decisions, respectively (see section 3 below). To focus for now on the 22,000 asylum decisions, these divide into control and treatment groups of approximately equal size of 11,000. A 1% treatment effect is thus 110 additional grant decisions in the treatment group relative to the control group. Of course, we do not know which individual decisions these are. More to the point, we could get the same difference between the two groups by mere chance even if the true treatment effect were zero. This would happen, for example, if our control group by mere chance contained 55 additional meritless decisions and our treatment group 55 additional applications with merit. We can estimate the probability of such a chance event, which with this sample size turns out to be less than 10%. If we had only a tenth of the sample size (still 90,000 decisions!), a mere 5 such additional decisions in the control and treatment group, respectively, could create the misleading appearance of a 1% treatment effect. This is much more probable, and would prevent any reasonable inference from such smaller sample. We would not be able to claim with any certainty that the 1% estimated effect is a true effect or mere noise.

Comparability of the underlying cases greatly facilitates bounding the probability of a chance results. At the extreme, if we knew with certainty that the decisions in the control and treatment groups would be identical absent the treatment, any actual difference in decisions would have to be due to the treatment. Similarly, if we had at least a fairly good estimate of what the decisions should be absent the treatment, the actual difference would provide a fairly good estimate of the treatment effect. Inversely, if there is large variation in the underlying cases (other than what we can observe and control for), treatment and control groups may easily be very different by mere chance, and detecting a treatment effect becomes more difficult.

Based on prior research on intra-judge differences, one important factor predicting case outcomes is the identity of the judge. The first advantage of the asylum data set is therefore the relatively small number of judges. There are only 340 immigration judges in the asylum data set, compared to 1,268 district judges in the sentencing data set. Moreover, all asylum cases have the same binary potential outcome, while sentencing cases present vastly differing potential sentence ranges. One could get a good estimate of that range if one had access to the case's sentencing grid information and/or the prosecutors' recommended sentence (if there is a plea bargain, which is nearly universal), but these data are not available. By contrast, we have quite predictive covariates for the asylum decisions (see section 3 and 4 below).

2.2.(Quasi-)Random assignment

A controlled experiment actively randomizes the assignment of cases to treatment and control groups, or equivalently, of the treatment to cases. We cannot do this.[5] Instead, we piggy-back on an assignment mechanism that is as good as random with respect to the underlying population of cases. That is, the assignment mechanism may have its own rationale, but it is unrelated to case characteristics. This requires that there is neither an influence of case

[5] And arguably, we would not want to, as we are interested in detecting the effect of extraneous factors "naturally" occurring in the real world.

characteristics or outcomes on the assignment mechanism, nor a third factor influencing both. NFL outcomes and the weather satisfy both of these requirements

Obviously, the outcomes and characteristics of asylum cases do not influence NFL outcomes (neither directly nor through scheduling) or the weather. It is conceivable that case scheduling adjusts to NFL scheduling, outcomes, or the weather. For a number of reasons, however, this is extremely unlikely. First, we have learned from conversations with practitioners that the main merit hearings in asylum cases are scheduled first-in first-out, leaving no role for discretionary adjustments. Second, even if there were such room, it seems implausible that NFL and the weather would enter the picture. In fact, many cases are scheduled so far in advance that not even the NFL schedule, let alone the result or the day's weather, would be known at the time of scheduling. The NFL schedule comes out in April,[6] while asylum cases may be scheduled over a year in advance. Third, once scheduled, the main hearing date is essentially set in stone, and decisions are rendered on the spot in almost all cases (cf. section 3 below). Finally, we have verified empirically that cases heard after NFL wins are not statistically different on observable case characteristics (other than the grant decision) from cases heard after losses.

It is not easy to conceive of third factors that might influence both asylum case characteristics and NFL outcomes, let alone the weather. Perhaps cities that become wealthier attract both a better football team and a more sophisticated set of asylum petitioners. The latter would be attracted by higher wages (although one might also think that economic

[6] See, e.g., http://www.nfl.com/photoessays/0ap1000000161578.

migrants are unlikely to obtain asylum). The former would be attracted by the higher purchasing power and the concomitantly higher advertisement revenue. However fanciful, we can and will account for this possibility by controlling flexibly for city time trends.

2.3.Plausibility

[To come: Literature review of sports and weather influences.]

3. Data

3.1.Asylum data

We obtained our data on immigration judge decisions in asylum cases from 1993 to 2013 by judge and court house directly from EOIR via a FOIA request (we also obtained a nearly identical data set via TRAC, and double-checked our results on those data). This dataset includes grant-deny decisions on applications for asylum, asylum-withholding (AW), or withholding-convention against torture (WCAT), along with information on hearing dates, the completion date, whether the applicant was legally represented, whether the application was filed affirmatively or defensively (i.e., in defense of a removal proceeding), and the applicant's origin.

AW and WCAT applications are mostly subsidiary applications to asylum applications. The former two categories are almost always ancillary applications, in which case they are not independent data points. There are only 22k independent AW and WCAT applications, far fewer than the 434k asylum applications. We thus drop AW and WCAT applications to obtain a more homogenous sample while only marginally decreasing sample size.

The main problem with the asylum data is that it contains not one but two dates that could represent the main merit hearing. The main merit hearing is the hearing at which the case's substance is tried. Several practitioners have told us that the judge will almost inevitably announce the final decision at the conclusion of the hearing. This is thus the relevant date for our purposes. Unfortunately, the data do not explicitly flag the main merit hearing date. If the judge renders an oral decision at the hearing's conclusion, the main merit hearing date will be the case completion date, which is in the data. Sometimes, however, the judge reserves a written decision, generally because the government has reserved appeal. In that case, the official completion date and the main hearing date do not coincide. We could then assume that the last recorded hearing date before the completion date is the main merit hearing date. But the hearing data seem too incomplete and unreliable to say so with confidence. Consequently, we use the completion date, and drop from the data all cases for which the completion date does not coincide with a hearing date.

Another problem with the asylum data is that it does not identify family links. It does identify lead applicants and the number of subsidiary applicants (i.e., family members applying with the lead applicant). But it does not indicate who those family members are. This is a problem because the decision for the lead and subsidiary applicants is generally taken jointly, i.e., these are not independent observations. To get around this problem, we mostly work not with individual decisions, but with the average grant rate of a given judge in a given court on a given date.[7]

3.2. Federal sentencing data

We obtain data on criminal sentencing by federal district judges from TRAC.[8] The data are complete for 2004 through 2011. For earlier years, we have only a selection of sentences, and very few before 1998. In total, there are approximately 900,000 cases.

The data contain information on prison sentences, probation sentences, fines, and the death penalty. We do not consider the death penalty, as it is exceedingly rare in federal cases (71 cases). Probation sentences and monetary fines are much more frequent but still apply in only about 10% of the cases each.[9] We thus ignore them as well, and focus exclusively on prison sentences.

3.3. NFL data

We then merged the asylum and sentencing data with NFL outcome data. As almost all NFL games are played on Sundays, we dropped all other game days to keep the sample homogenous.

---

[7] For example, if judge Smith granted four applications and denied one on 4/15/2013 in Newark and granted one in New York City, we would collapse this into two data points: one data point Smith-Newark-4/15/2013 with value 0.8, and one data point Smith-NYC-4/15/2013 with value 1.

[8] Yang (forthcoming) provides an extensive description of these data.

[9] Even then, monetary fines are mostly very small relative to prison sentences. The median non-zero monetary fine is $2,000, and the 90th percentile is $15,000.

For lack of a better alternative, we matched the courthouse of the judge to the NFL team most favored by the local community in 2013 according to Facebook likes.[10] This method has three obvious shortcomings. First, the Facebook data is for 2013, while our other data is for 1993-2013. This could make a considerable difference especially for teams that moved, such as the L.A. Raiders' move to Oakland. Second, many cities are in truth divided between several teams, often bitterly so, while our measure treats (all of) any location as following only one team. For example, New York City has the Giants and the Jets. Third and perhaps most importantly, we do not know the personal preference of any given judge.[11] While it is reasonable to guess that a judge who cares about football would follow the local team, he or she may not, and in fact may not care about football at all. In sum, our assignment of teams to judges or even court houses contains considerable measurement error, and will thus bias our estimates toward zero. The lack of separate information on judges' preferences also prevents us from disentangling whether sports and weather influence decisions directly through the judges' mood, or through their environment or the other court house participants.

3.4. Weather

Finally, we merged with weather data from the National Weather Service.

4. Results

[10] Cf. http://www.facebook.com/notes/facebook-data-science/nfl-fans-on-facebook/10151298370823859.

[11] We considered surveying the immigration judges. But even a survey endorsed by the judges' union received only a 25% response rate, see […]. We figured that asking the judges about their sports preferences would generate a near zero response rate.

We now present our basic results.

4.1. Wins vs losses

We begin with the effect of NFL football results. For the reasons mentioned, we restrict the sample to asylum cases decided on Mondays after Sunday games. This leaves us with about 22,000 cases. (The results in the full sample controlling for day of the week and days after games are approximately the same.)

Preliminary bivariate tests

We begin by some simple comparisons. Table 1 compares mean grant rates on days after a win with grant rates after a loss. Looking at individual decisions (test 1), the average grant rate after a win is 3.6% higher than after a loss, or about 10% of the base grant rate. This difference is both economically large and, subject to the very important caveat in the next paragraph, highly statistically significant. Changing the unit of observation to individual judge-day grant rates (test 2) or city-day grant rates (test 3) barely changes this result. Similarly, grant rates are strongly positively correlated with wins, regardless of the level at which the data are pooled (table 2).

To be sure, these simple tests omit two important caveats. First, the simple statistical tests treat each case or ratio, as the case may be, as independent. In reality, however, observations from the same city and even more so from the same judge are subject to many of the same influences from unobserved factors. The above statistical tests are thus invalid. Second, argument that NFL wins are randomly assigned to cases becomes tenuous over long time periods. As hypothesized above, as cities get richer, they may get both better football teams and stronger asylum applicant pools.

Regressions

Consequently, in table 3, we now dig deeper into the data. We estimate a fixed effects regression for applicants *i,* judges *j*, cities *c*, and decision date *t* of the following form:

$$Grantratio_{jct} = baserate_{jc} + \delta\ T_{ct} + \beta_1 X_{jct} + \beta_2 week_t + \beta_3 season_t + \varepsilon_{jct}$$

Grantratio is the ratio of grants to the number of decisions handed down by the judge in a given court house on a given day (that is, this reduces the dataset to at most one observation per judge per day). As explained above, we use this rather than the individual grants to deal with dependence within families. $Baserate_{jc}$ is a fixed effect for judge *j* sitting in city *c*. T_{ct} indicates the treatment and δ the coefficient of interest (e.g., win or loss). X_{jct} is a vector of *average* applicant covariates for the applicants who appeared before the judge in that court on that day. In particular, it contains whether the claim was defensive or affirmative, and whether the applicant was legally represented. We also include the fraction of applicants who were Chinese; in the full sample, Chinese are over 20% of the applicants and by far the largest group. $Week_t$ is a separate dummy for each week of the year (1-52), and *season* is a separate dummy for each NFL season between 1992-2013, and ε_{ijt} is an error term.

The case covariates X_{it} are not required for identification. In fact, as already mentioned, we test that they are randomly distributed across our treatment and control groups identified by T_{ijt}. The case covariates help reduce noise, however, and thus increase our statistical power.

By contrast, it is indispensable that we account for the aforementioned dependence of observations from the same city and, a fortiori, same judge. The standard way of dealing with this problem is clustering. The problem here is that there are two levels at which cases are not independent, and they are not nested: the judge, and the city or court house (judges may sit in more than one court house). We thus cluster either by city, judge, or both (using the code of Cameron et al. 2006).

The results are in table 3. Controlling for application characteristics, the estimated effect of an NFL win is 1.3%, which is statistically significant at the 10% level. It hardly matters which way we cluster. In fact, we have found that the clustering surprisingly has hardly any effect compared to no clustering. Without application controls, the estimated effect is slightly larger, namely 1.8%. The difference between the two estimates is not statistically significant however, and could have arisen by mere chance.

Table 3 does not yet explicitly address the concern that applicant pools and NFL teams may develop in parallel. In table 4, we explicitly address this possibility in two different ways. In models 1 and 2, we include city-specific time trends, i.e., a separate time trend for each city. Here our coefficient shrinks to 1% and is no longer statistically significant. However, the city-specific linear trend is rather crude. In our opinion, the better, more flexible way to account for unobserved common trends is to match decision to their nearest neighbor. That is, rather than imposing a particular linear model, we compare each decision to the closest decision by the same judge in the same city after the opposite game result. For example, if the city's team lost on weekend 48, we compare the decision on the following Monday to decisions after the nearest win(s): weekend 46 and 48, if any; if not, weekend 44 and 49, if any; and so on. Technically, we use the matching estimator of Abadie and Imbens (2006). When we do that, we again estimate an effect of 1.5%, which is again statistically significant at the 10% level.

4.2. Weather

In table 4, we look at the effect of three types of bad weather on the day of the decision: rain, snow, and high winds. In each case, we have not only a dummy from the national weather service, but also a continuous variable measuring the intensity. We include an increasing number of covariates in models 1-4 as indicated in the bottom rows and explained in section 2 above. We again cluster the standard errors by city.

As can be seen, all three types of bad weather reduce the grant rate. For example, rain reduces grant rate by between 0.3%-0.5%. Individually, the coefficients are generally not statistically significant, perhaps in part because they are collinear. However, the F-test of joint significance strongly rejects the null hypothesis of no effect.

5. Discussion

5.1. External validity: comparison with federal sentencing

A first question is if and to what extent our results generalize to other judicial settings. As already mentioned, immigration courts are rather special. They have an extremely high workload, the judges are not Article III judges, and the applicable legal standard is rather loose.

We thus ran similar tests with the federal sentencing decisions, described in section 3.4 above. The results are in table 5. Two things are immediately apparent. First, the estimated coefficients are negative. That is, as with asylum decisions, judges appear to be, if anything, more lenient after a positive sports outcome. Second, however, the estimated effect of approximately a half month less on average after a win is not statistically different from zero.

There are two main ways of interpreting these results. One possibility is that sentencing by federal district judges is not susceptible to influence by extraneous factors like NFL games. Another possibility is that it is susceptible, but that the effect is too small to be detected. (A third possibility is that the result for immigration judges was, after all, just a statistical fluke, not repeated with other data. We leave that to the reader to decide.)

5.2.Implications

If our result stands, then we have shown intra-judicial variation caused by extraneous factors, i.e., factors unrelated to the case merits. It is not the "judge's breakfast," but close.

[to be added: implications for acceptance of experiments etc.]

References

Abadie, A., and G. W. Imbens. 2006. "Large sample properties of matching estimators for average treatment effects." Econometrica 74: 235-267

Berdejo, Carlos, and Daniel Li Chen. 2013. Priming Ideology? Electoral Cycles Without Electoral Incentives Among U.S. Judges. Working paper.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2006. "Robust Inference with multi-way clustering". NBER Technical Working Paper 0327.

Card, David, and Gordon B. Dahl. 2011. "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior." Quarterly Journal of Economics 126:103-143.

Danziger, Shai, Jonathan Levav, and Liora Avnaim-Pessoa. 2011a. "Extraneous factors in judicial decisions." Proceedings of the National Academy of Science of the United States of America 108:6889-6992.

______. 2011b. "Reply to Weinshall-Margel and Shapard: Extraneous factors in judicial decisions persist." Proceedings of the National Academy of Science of the United States of America 108:E834.

Dworkin, Ronald. 1986. Law's Empire. Cambridge, MA: Harvard University Press.

Edmans, Alex, Diego Garcia, and Oyvind Norli. 2007. Sports Sentiment and Stock Returns. Journal of Finance 62:1967-1998.

Everson, George. 1919-1920. “The Human Element in Judging.” Journal of the American Institute of Criminal Law and Criminology 10:90-99.

Fischman, Joshua. 2013. “Reuniting 'Is' and 'Ought' in Empirical Legal Scholarship.” University of Pennsylvania Law Review 162:117.

Fischman, Joshua. Forthcoming. Measuring Inconsistency, Indeterminacy, and Error in Adjudication. American Law and Economics Review.

Fischman, Joshua, and David S. Law. 2009. “What Is Judicial Ideology, and How Should We Measure It?” Washington University Journal of Law and Policy 29:133.

Frank, Jerome. 1930 [2009]. Law & The Modern Mind. New Brunswick, NJ: Transaction Publishers.

Healy, Andrew J. , Neil Malhotrab, and Cecilia Hyunjung Mo. 2010. “Irrelevant events affect voters’ evaluations of government performance.” Proceedings of the National Academy of Science of the United States of America 107:12804–12809.

Kennedy, Duncan. 1999. A Critique of Adjudication. Cambridge, MA: Harvard University Press.

Ramji-Nogales, Jaya, Andrew I. Schoenholtz & Phillip G. Schrag. 2007. “Refugee Roulette: Disparities in Asylum Adjudication.” Stanford Law Review 60:295-411.

Schauer, Frederick. 2009. Thinking Like a Lawyer. Cambridge, MA: Harvard University Press.

Jeff Strnad. 2007. “Should Legal Empiricists Go Bayesian?” American Law and Economics Review 9:195.

Weinshall-Margela, Keren, and John Shapard. 2011. “Overlooked factors in the analysis of parole decisions.” Proceedings of the National Academy of Science of the United States of America 108:E833.

Yang, Crystal S. Forthcoming. “Have Inter-Judge Sentencing Disparities Increased in an Advisory Guidelines Regime? Evidence From *Booker*.” NYU Law Review.

Table 1: Differences in mean grant rates, by NFL win/loss

Level of aggregation	After	N	Mean	*p*-value (two-sided)
(1) Case	After loss	11101	0.371	
	After win	11193	0.408	
	Difference		-0.037	0.0000
(2) Judge-day	After loss	6676	0.345	
	After win	6795	0.379	
	Difference		-0.034	0.0000
(3) City-day	After loss	2596	0.291	
	After win	2620	0.318	
	Difference		-0.027	0.0099

Table 2: Correlations between grant rates and NFL wins / win rates

Grant rates by		N	Correlation	*p*-value (two-sided)
Judge &	Day	13477	0.04	0.0000
	Year	3196	0.05	0.0060
	Total	340	0.17	0.0013
City &	Day	5216	0.04	0.0099
	Year	854	0.09	0.0107
	Total	56	0.22	0.1105

Table 3: Main NFL Regressions

Dependent variable	Judge-City-Day Ratio of Granted Asylum					
	(1)	(2)	(3)	(4)	(5)	(6)
Yesterday's NFL Win	0.018*	0.018*	0.018*	0.013*	0.013*	0.013*
	(0.007)	(0.009)	(0.009)	(0.007)	(0.007)	(0.008)
JudgeXCity Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Week Fixed Effects	-	-	-	Yes	Yes	Yes
Season Fixed Effects	-	-	-	Yes	Yes	Yes
Application controls	-	-	-	Yes	Yes	Yes
N	13,508	13,508	13,508	13,504	13,504	13,504
Clustering	No	City	+Judge	No	City	+Judge
Number of clusters	-	56	56x340	-	56	56x340

Notes: Standard errors in parentheses (* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. Applicant controls are the ratio of applications that were represented, that were defensive, and that were by Chinese applicants.

Table 4: NFL Regressions with flexible time controls

Estimation technique	OLS		Nearest-neighbor matching	
Dependent variable	Judge-City-Day Ratio of Granted Asylum			
	(2)	(3)	(5)	(6)
Yesterday's NFL Win	0.010	0.010	0.015*	0.016*
	(0.007)	(0.009)	(0.008)	(0.009)
Judge/City FE / Match	JudgeXCity		Judge	JudgeXCity
Time control	City-specific trends		Match on date	
Week Fixed Effects	Yes		No	
Application controls	Yes		No	
N	13,504	13,504	12,402	11,496
Clustering	City	+Judge	-	-
Number of clusters	56	56x340	-	-

Table 4: Judicial Decisions and Today's Weather

Dependent variable	Ratio of Asylum Decisions Granted			
	(1)	(2)	(3)	(4)
Rain (may include freezing rain) present	-0.0063*	-0.0055*	-0.0037	-0.0031
	(0.0034)	(0.0033)	(0.0028)	(0.0027)
Precipitation in mm[1]	0.0010	0.0015	-0.0006	0.0002
	(0.0019)	(0.0012)	(0.0015)	(0.0020)
Highwinds present	-0.0177	-0.0126	-0.0166	-0.0275
	(0.0134)	(0.0138)	(0.0146)	(0.0191)
Windspeed (tenths of meters per second)[1]	0.0016	-0.0028	-0.0034	-0.0041**
	(0.0031)	(0.0029)	(0.0022)	(0.0020)
Snow present	-0.0065	-0.0118	-0.0124*	-0.0120**
	(0.0067)	(0.0073)	(0.0070)	(0.0058)
Snow amount in mm[1]	-0.0038**	-0.0017	-0.00056	-0.0014
	(0.0017)	(0.0014)	(0.0013)	(0.0011)
F-Test of Joint Significance	0.021	0.061	0.032	0.007
Judge Fixed Effects	Yes	Yes	Yes	Yes
Applicant Controls	No	Yes	Yes	Yes
Time Fixed Effects	No	No	Yes	Yes
National Origin Fixed Effects	No	No	No	Yes
N	131720	127437	127437	127437
R^2	0.12	0.16	0.17	0.26

Note: Standard errors in parentheses (* p < 0.10; ** p < 0.05; *** p < 0.01). Standard errors are clustered by city. Observations are at the judge x day level. Applicant controls are the ratio of applications that have the following characteristics: defensive, primarily about torture, had representation, and family size. Time fixed effects are dummy indicators for day of week, week of year, and year. National origin fixed effects are dummy indicators for all national origins that appear at least one thousand times with an additional dummy for less frequent national origins.

[1]Coefficients are multiplied by 1000.

Table 5: Federal Sentencing

Dependent variable:	Sentence (months)			
	(1)	(2)	(3)	(4)
Yesterday's NFL Win	-0.41	-0.41	-0.56	-0.56
	(0.87)	(0.87)	(0.87)	(0.88)
JudgeXCity FE	Yes	Yes	Yes	Yes
Time FE	No	No	Week & season	Week & season
Charge controls	No	No	Yes	Yes
N	57,931	57,931	57,931	57,931
Clustering	District	+Judge	District	+Judge
Number of clusters	85	85x1085	85	85x1085

Notes: Standard errors in parentheses (* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$). Dependent variable is prison sentence in months. Charge controls are the department of the offense classification, and whether or not the case was tried. Observations are restricted to Monday decisions. All other variables are as described in Table 2.