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

HARVARD

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

ARE SLEEPY PUNISHERS REALLY HARSH PUNISHERS?: COMMENT

Holger Spamann

Discussion Paper No. 898

02/2017

Harvard Law School
Cambridge, MA 02138

Are Sleepy Punishers Really Harsh Punishers?: Comment

Holger Spamann[1]

Abstract

This comment points out four severe reservations regarding Cho et al.’s (PS 2017) finding that U.S. federal judges punish more harshly on “sleepy Mondays,” the Mondays after the start of Daylights Savings Time. First, Cho et al.'s finding pertains to only one of at least two dimensions of harshness, and the opposite result obtains in the second dimension. Second, even within the first dimension, Cho et al.'s result is statistically significant only because of a variable transformation and sample restrictions that are neither transparent in the article nor theoretically sound. Third, reanalysis of the data with superior methods reveals no significant “sleepy Monday” effect in the years 1992-2003. Fourth, sentences were on average shorter on “sleepy Mondays” out of sample, namely in 2004-2016.

[1] Harvard Law School, Cambridge MA 02138; hspamann@law.harvard.edu. I thank Kyoungmin Cho for sharing data and answering questions about their analysis, and Dan Klerman, Ivan Reidel, Christopher Robertson, Jeremy Sawyer, Brooke Stanley, Tom Vogl, and Crystal Yang for very helpful feedback. I performed a part of this research as a TRAC Fellow of the Transactional Records Access Clearinghouse (TRAC) at Syracuse University, and I thank Sue Long for sharing and explaining TRAC’s data. The data and code used in this comment are available online at _____.

Cho et al. (2017) analyze criminal sentencing by U.S. federal judges in the years 1992-2003. Controlling for case covariates, Cho et al. estimate that "sentences rendered on sleepy Mondays"—Mondays following the start of daylight savings time (DST), when the night from Saturday to Sunday is one hour shorter—"were approximately 5% longer than those rendered on [the preceding and the subsequent] Mondays." Cho et al. estimate that so large a difference would arise by chance with a probability of only 0.5% if judges tended to sentence equally on the three Mondays, i.e., the p-value is .005. Cho et al. interpret this finding as evidence that sleep-deprived judges punish more harshly.

This interpretation is unwarranted for four reasons. First, Cho et al.'s finding pertains to only one of at least two dimensions of harshness, and the opposite result obtains in the second dimension. Second, even within the first dimension, Cho et al.'s result is statistically significant only because of a variable transformation and sample restrictions that are neither transparent in the article nor theoretically sound. Third, reanalysis of the data with superior methods reveals no significant "sleepy Monday" effect in the years 1992-2003. Fourth, sentences were on average *shorter* on "sleepy Mondays" out of sample, namely in 2004-2016.

Table 1 summarizes all the models discussed in this comment, showing point estimates and standard errors only for "sleepy Monday." Model 1 is an exact replication of Cho et al. (model 2 of their table 1) using their data; the estimated effect size and standard error are the same as in Cho et al.

Model 1 is restricted to cases in which the judge imposed prison time. That is, model 1 ignores the judge's initial decision to impose any prison time at all, which judges refrain from doing in 18% of cases (opting instead for alternative forms of punishment such as probation). Other articles, including the article guiding Cho et al.'s model discussion (Steffensmeier and Demuth 2000), routinely account for this equally important second dimension of harshness. Along this second dimension, however, the "sleepy Monday" estimate is *negative*: judges were *less likely* to impose prison time on "sleepy Mondays." This is shown in model 2, where the dependent variable is a dummy for whether the judge imposed any prison time.

To be sure, the negative estimate in model 2 is not statistically significant. However, the positive estimate in model 1 would not be statistically significant either but for inappropriate variable transformations and sample restrictions that are not mentioned in the article and that can be inferred only from direct examination of the data.[2] First, Cho et al.'s data contain fewer observations than are available directly from the USSC. Second, Cho et al. exclude from the sample "races" other than "black" and "white," notably the third of all cases involving "Hispanics."[3] This exclusion is unnecessary and unprincipled: racial differences are not plausibly related to the "sleepy Monday" effect, and in any event are not a priori more important than other differences that do not trigger exclusion from the sample, such as different crime types. Third, Cho et al. do not apply to the dependent variable—sentence length s—the "log transformation" ($\ln(s)$) mentioned in the article. Rather, Cho et al. use $\ln(1+s)$, as shown in model 1 of table 1. Unlike $\ln(s)$, this transformation does not support Cho et al.'s interpretation of the

[2] This paragraph's description of Cho et al.'s model and data was confirmed in personal correspondence with the authors.
[3] The article merely states that it "controlled for … race (i.e., White vs. Black)."

estimate as percentage change, nor does it generate a normally distributed variable, which is Cho et al.'s stated goal for the "log transformation."[4] There is no reason to use $\ln(1+s)$ unless the goal is to keep observations with s=0 in the sample, which Cho et al. do not. Using $\ln(s)$ and including the additional "races" and observations, the estimated "sleepy Monday" effect shrinks by 20% and the p-value rises over tenfold to .074. This is shown in model 3.

For completeness, model 4 also adds foreign defendants, whom Cho et al. exclude for unconvincing reasons. As one would expect, enlarging the sample enhances precision. But it further reduces the estimated effect size and statistical significance (p=.091).

Various other modeling choices in Cho et al. are defensible but arguably not optimal. Models 5-7 follow best practice in the analysis of sentencing data (cf., e.g., Yang 2015). As expected, the methodological improvements reduce the standard error of the "sleepy Monday" coefficients. But the more precisely estimated coefficients themselves are much smaller as well. They are statistically indistinguishable from zero at any accepted level of significance.

Finally, the "sleepy Monday" effect is rejected out of sample. Cho et al. restricted their analysis to 1992-2003 because the USSC does not publicly disclose sentencing dates for later years. The Transactional Records Access Clearinghouse (TRAC), however, has obtained such data through Freedom of Information Act requests. Controlling for all relevant covariates available from TRAC, sentences imposed by district judges on "sleepy Mondays" are on average *shorter and rarer* than on other Mondays in the years 2004-2016 (models 8-10).[5] The 95% confidence interval excludes effect sizes above 3% in the harshness dimension considered by Cho et al. (model 8).

This strongly suggests that sleepy punishers are not harsh punishers, at least not to the extent claimed by Cho et al. On a methodological level, the discussion illustrates the impact of modeling choices such as sample restrictions and variable transformations, counseling even stricter transparency about such choices.

References

Cho, K., Barnes C. M., & Guanara, C. L. (2017). Sleepy Punishers Are Harsh Punishers: Daylight Savings Time and Legal Sentences. *Psychological Science, 28,* 242-247.

Correia, S. (2014). REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects. Statistical Software Components s457874, Boston College Department of Economics, revised 6 Aug 2016.

Steffensmeier, D., & Demuth S. (2000). Ethnicity and Sentencing Outcomes in U.S. Federal Courts: Who is Punished More Harshly? *American Sociological Review, 65,* 705-729.

[4] If $s \geq 0$, as here, then $\ln(1+s) \geq 0$, whereas the support of the normal distribution is the entire real line.
[5] The estimation uses Correia (2014).

Yang, C. S. (2015). Free at Last? Judicial Discretion and Racial Disparities in Federal Sentencing. *Journal of Legal Studies, 44,* 75-111.

Table 1: Sentences (*s*) imposed on Mondays following start of Daylight Savings Time (Sleepy Monday) vs. other days

Model	(1)[a]	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent var.	Ln(1+*s*)	*s*>0	Ln(*s*)	Ln(s)	Ln(*s*)	*s*>0	*s*	Ln(*s*)	*s*>0	*s*
s=0 included?	No	Yes	No	No	No	Yes	Yes	No	Yes	Yes
Sample Races	White, black	White, black	All	All	All			All		
Citizens	U.S.	U.S.	U.S.	All	All			All		
Days	3 Mondays				All			All		
Years	1992-2003				1992-2003			2004-2016		
Controls	Cho et al.[c]	Cho et al.[c]	Cho et al.[c] Hispanic	Cho et al.[c] Hispanic citizenship	Cho et al.[c] Hispanic citizenship sentencing grid FE day-of-week, month & year FE etc.[d]			trial judge FE lead charge FE day-of-week, month & year FE		
District effect	RE slopes[b] & intercept				FE			FE		
Clustered s.e.s	No				District			District		
Model type	HLM	Logit	HLM	HLM	Linear			Linear		
Sleepy Monday	**0.061****	**-.249**	**0.048**	**0.039**	**0.024**	**-0.005**	**0.673**	**-0.028**	**-0.001**	**-0.291**
(standard error)	**(0.023)**	**(0.134)**	**(0.027)**	**(0.023)**	**(0.017)**	**(0.006)**	**(0.849)**	**(0.030)**	**(0.007)**	**(0.991)**
N	2,985	3,781	4,125	6,330	434,689	525,053	525,053	737,637	907,450	907,450
Data source	Cho et al.	Cho et al.	USSC	USSC	USSC			TRAC		

Note: In all models, *s* is right-censored at 470, as in Cho et al. RE and FE indicate random and fixed effects, and HLM hierarchical linear model.

[a] Model 1 is an exact replication of Cho et al. (model 2 of their table 1).

[b] To achieve convergence, the logit model 2 does not allow random slopes for the control variables.

[c] "Cho et al." refers to the controls used in Cho et al., which include: sentencing year (trend); criminal history FE; offense level; trial; multiple convictions indicator; defendant age, gender, race (black, white), and education (below high school, high school graduate, some college, or college graduate).

[d] Additional controls in models 5-7 include: offense type FE, age squared, number of dependents, and whether any statutory minimum applied. Cf. Yang (2015) for more details.

* *p*<.05, ***p*<.01.