Pre-Analysis Plan:
Event Dependence in Death Penalty Executions

August 17, 2016

Contents

1 Introduction **1**
- 1.1 Abstract . . . 1
- 1.2 Motivation . . . 1
- 1.3 Research Question . . . 1

2 Research Strategy **1**
- 2.1 Sampling . . . 1
  - 2.1.1 Sampling Frame . . . 1
  - 2.1.2 Assignment to Treatment . . . 2
- 2.2 Data Processing . . . 2

3 Empirical Analysis **4**
- 3.1 Variables . . . 4
  - 3.1.1 Event Dependence as Stratified Levels of Execution . . . 4
  - 3.1.2 Homicides . . . 4
  - 3.1.3 Racial Threat . . . 4
  - 3.1.4 Poverty . . . 4
  - 3.1.5 Population . . . 5
  - 3.1.6 Judicial Elections . . . 5
- 3.2 Model of Choice . . . 5
- 3.3 Interpreting Effects . . . 6

PRELIMINARY DRAFT: PLEASE DO NOT CITE OR DISTRIBUTE

1 Introduction

1.1 Abstract

This pre-analysis plan outlines a research strategy to test a "self-reinforcing" theory of death penalty executions, which holds that counties face decreasing marginal costs for executions (Baumgartner et al. 2016). We test this theory through examining event dependence in executions among counties that have the death penalty. To test for the presence of these self-reinforcing processes in executions, and the exogenous factors that may explain executions, we utilize an event history model that accounts for event dependence. The empirical findings of this analysis may have profound consequences for how we understand executions. Evidence of event dependence would reveal that the main determinant of whether an individual is executed is the county's previous experience with execution, which would raise many important policy, legal, and moral questions.

1.2 Motivation

Building upon previous work that has found that death penalty executions follow a power law distribution (Baumgartner et al. 2016), we seek to better understand this dynamic through specifying event history models to more precisely model event dependence. Previously, scholars have generally explained the rate of death penalty executions through the effect of exogenous actor attributes. More recently, however, researchers have examined the spatial and temporal clustering of death penalty executions (Baumgartner et al. 2016; Dieter 2013), which suggests that certain endogenous processes may drive executions. We examine this in further methodological detail while positing a novel theory to explain the event dependence in death penalty executions.

1.3 Research Question

When accounting for exogenous covariates, is there event dependence in death penalty executions? This is to say, if a county has executed an individual, is it increasingly likely that they will execute additional individuals?

2 Research Strategy

2.1 Sampling

2.1.1 Sampling Frame

Our data account for all modern executions, with a temporal domain starting in 1977 and ending in 2014. These data start with the Death Penalty Information Center database (Dieter 1995), while including data on executions (Espy 2004), homicides (Baumgartner and Gram 2015; FBI 1988), population demographics, and judicial institutions (NCSC 2016). These data include 3312 to 3637 county-year observations (contingent upon level of stratification), which reflects a census of all relevant cases. For our primary analysis, in which we plan to present the results of a model with four levels of stratification in executions, we would have 3433 observations. For a county-year to be included, the county must have had the death penalty during this period of time.

2.1.2 Assignment to Treatment

Given that we are essentially interested in the marginal change in hazard functions for each additional execution, assignment to treatment can be considered as a function of all other confounding variables that would lead to a county's decision to use the death penalty.

2.2 Data Processing

From the raw data, the data were cleaned as follows:

- Linear extrapolations were used to get annual estimates for Census data that were not available for all years. This linear extrapolations were used to generate year-level estimates. If month estimates were needed, we used the year-level estimate.
- When combining the homicides with the county-year executions data, we found a few cases where the "fips" identifier did not uniquely identify every case. One problem was two lines for Colonial Heights County, Virginia. One line had 31 (say, Line A) homicides, and the other (Line B) had 14, and this difference was solely accounted for by 1988. These were added, as in 1988 there were zero homicides in Line B, which seems unlikely.

The dataset used for our analyses will be produced using the following `R` script:

```
rm(list=ls())

options(stringsAsFactors = FALSE)

library(data.table)
library(foreign)
library(readstata13) # probably need to install this package

# Local WD
setwd("~/Dropbox/Projects/Projects_Steffensmeier/DeathPenelty/Data")

# Read in original data

cens <- read.dta13('Census_Judicial_DP_Master_782015.dta')
exec <- read.dta13('execs-by-date-by-county.dta')

setDT(cens)
setDT(exec)

cens <- cens[geo_fips > 0 & !is.na(year) & deathstate == 1,]

new <- rbindlist(lapply(1:12, function(i) {
  cens$month <- i
  cens
}))
```

```r
exec <- exec[fips > 0, .(fips, date)]
setnames(exec, 'fips', 'geo_fips')

exec[, `:=`(year = year(date), month = month(date), execInMonth = 1)]

exec <- unique(exec[, date := NULL])

exec2 <- merge(new, exec, by = c('geo_fips', 'year', 'month'), all.x = TRUE)
exec2$execInMonth[is.na(exec2$execInMonth)] <- 0

# There are 1321 execution-months in exec but only 1307 in the exec2 file...

setkeyv(exec2, c('geo_fips', 'year', 'month'))

exec2[, past_exec := shift(cumsum(execInMonth)), .(geo_fips)]
exec2$past_exec[is.na(exec2$past_exec)] <- 0


new_ex_master <- lapply(3:8, function(i) {
  # drops counties in highest strata after they execute someone (is this okay?)
  d <- exec2[past_exec < i, ]
  d[, max_strat := i]
  d[, `:=`(stratum = ifelse(past_exec == i - 1, i, past_exec + 1),
    newtime = year - 1977 + month / 12 - 1)]
  d <- d[, .(
    # can add more variables as you wish
    state_fips = unique(state_fips),
    state = unique(state),
    totalpop = mean(totalpop),
    pctwhitepop = mean(pctwhitepop),
    pctpovyes = mean(pctpovyes, na.rm = TRUE),
    judelec = mean(judelec), # this is a mean - how to aggregate judselect?
    circuit = unique(circuit),
    dp_opinion = mean(dp_opinion, na.rm = TRUE),
    homicides = mean(homicides, na.rm = TRUE),
    exec = max(execInMonth),
    time1 = min(newtime),
    time2 = max(newtime),
    max_strat = unique(max_strat)),
    .(geo_fips, stratum)]
  d[, racialthreat := abs(70 - pctwhitepop)]
  d
})
```

```
save(new_ex_master, file = 'new_ex_master.Rdata')
```

3 Empirical Analysis

3.1 Variables

3.1.1 Event Dependence as Stratified Levels of Execution

The main "variable" or "treatment" of interest is event dependence, which is to say, that we are interested in the differences of the cumulative hazard functions for counties at different levels of executions after controlling for other explanatory variables. Simply put: we are interested in comparing the cumulative hazards for counties with one execution to counties with two, three, or four or more executions. This is done through stratifying executions into four distinct categories. Data for these variables are sourced from Baumgartner et al. (2016). While we do not have a strong theoretical motivation for choosing four strata, it seems that given the typical distribution of executions, this should capture quite a bit of the distribution. As a robustness check, we will include models in the Supplementary Information (SI) Appendix that have different levels of stratification. In these models, it is important to group like-strata, especially as sample size decreases.

3.1.2 Homicides

It goes without saying that homicides in a county would be an intuitive predictor of death penalty prosecutions and executions for a county (Baumgartner and Gram 2015; FBI 1988). As such, we include the number of homicides that a county has had to control for its potential effect. The inclusion of this variable, however, may lead to violations of the proportional hazards assumption as the variable's effect may vary across strata. In order to fit a more credible model, we may transform this variable to account for proportional hazards assumption violations.

3.1.3 Racial Threat

Racial Threat Theory expects that as members of a minority group increase, the majority race will seek to impose higher levels of social control, including more punitive measures for law breaking (Eitle, D'Alessio and Stolzenberg 2002). In the case of the death penalty, as the proportion of racial minorities to whites increases, the justice system may be more likely to utilize the death penalty as a form of social control. To account for such a dynamic, we include a "Racial Threat" variable that is calculated as follows: $\mid 70 - Percentage\,White\,of\,Population \mid$.

3.1.4 Poverty

The amount of poverty in a county may drive death penalty executions and the prevalence of murder. We include a variable for the percentage of the county that lives in poverty to capture this dynamic; data for this variable is sourced through the Census and linearly extrapolated for off-years. In order to fit a more credible model, we may transform this variable to account for proportional hazards assumption violations.

3.1.5 Population

Counties with a large population may be more likely to experience large amounts of murder and just as a product of prosecutions, more executions. As such, controlling for the effect of population on executions beyond homicides is necessary. This data is sourced from the Census and linearly extrapolated to generate year-level estimates. In order to fit a more credible model, transform this variable to account for proportional hazards assumption violations.

3.1.6 Judicial Elections

In counties that elect justices, executions may be more likely as there are electoral incentives to be perceived as "hard on crime". Data for this variable is sourced from the NCSC (2016), which hosts resources on the method of justice selection for each state. Given that this variable is an indicator for whether justices are elected or not, we do not expect the inclusion of this variable to induce violations of the proportional hazards assumption.

3.2 Model of Choice

To conduct these analyses and explore the role of event dependence in predicting executions, we will specify event history models. These models allow for analysts to examine the role of observed covariates on the duration of certain phenomena. Event history models account for event dependence by allowing hazard functions to vary according to strata for different event levels. Given that we are interested in exploring the relative increase in cumulative hazard functions as the number of executions increases across strata, we believe that an event history process best describes the postulated data-generating process. Our ultimate goal is to specify the models with increasing levels of complexity, which is to say, we will move from Cox model without event dependence to the Cox model with event dependence, building up to the full conditional frailty model should sample size prove sufficient.

The conditional frailty model, introduced by Box-Steffensmeier and De Boef (2006) and Box-Steffensmeier, De Boef and Joyce (2007), is said to minimize bias and increase efficiency relative to alternative models that may be specified to examine event dependence. In particular, the conditional frailty model performs better under conditions of event dependence and heterogeneity, and as such, the conditional frailty model may be the safest choice and produce the most accurate inferences (Box-Steffensmeier and De Boef 2006; Box-Steffensmeier, De Boef and Joyce 2007).

Regardless of the model used, model specification may vary in an effort to reach a model that does not violate any of the assumptions of the Cox model. Should a model violate this proportional hazards assumption assumption, we will modify any problematic variables through interaction (typically with time), log-transformation, or stratification in an effort to fit the most credible model. We will also examine the Cox-Snell, Martingale, and Deviance residuals for each model and pay close attention to any poorly predicted observations. Such observations may be evidence for distinct data generating processes. As a result, we will estimate models excluding these observations and include them in the SI Appendix as a robustness check. Counties that have the death penalty but never execute offenders may be difficult to predict, and as such, may be produced by a distinct data generating process. This is a common problem in event history modeling. We will pay close attention to these observations and based upon our analysis, may have to consider an alternative way to model them.

Should this come to pass, we will specifying a split-population event history model. These models allow for the estimation of a splitting parameter, δ, which describes the "split" in the population of observations who survive or never survive, or in this case, execute or never execute. It would allows for the estimation of the timing of a county's execution and the probability that an execution will or will never occur. However, it is worth noting that the bias introduced through modeling never-executers using the conditional frailty model is a special case of the bias resulting from a violation of the proportional hazards assumption. As such, the results presented for a model that does not violate the proportional hazards assumption can be held as inferentially unbiased and consistent.

3.3 Interpreting Effects

Should there be increasing cumulative hazards across strata of executions, we would conclude that there is support for the hypothesis that executions reflect a learning or decreasing marginal cost process. While we are not particularly interested in the effects of other variables, we will consider and interpret them according to conventional methods of statistical significance.

References

Baumgartner, Frank R, Woody Gram, Kaneesha Johnson, Arvind Krishnamurthy and Colin Wilson. 2016. "The Geographic Distribution of US Executions.".

Baumgartner, Frank and Wallace Gram. 2015. "A Power-Law Analysis of the Uneven Geographic Distribution of Executions in the Post-Furman era of the Death Penalty.".

Box-Steffensmeier, Janet M and Suzanna De Boef. 2006. "Repeated events survival models: the conditional frailty model." *Statistics in medicine* 25(20):3518.

Box-Steffensmeier, Janet M, Suzanna De Boef and Kyle A Joyce. 2007. "Event dependence and heterogeneity in duration models: The conditional frailty model." *Political Analysis* 15(3):237–256.

Dieter, Richard C. 1995. "Death Penalty Information Center." *Retrieved January* 29:2007.

Dieter, Richard C. 2013. "The 2% Death Penalty: How a Minority of Counties Produce Most Death Cases at Enormous Costs to All." *Washington, DC: The Death Penalty Information Center. Retrieved on December* 13:2013.

Eitle, David, Stewart J D'Alessio and Lisa Stolzenberg. 2002. "Racial threat and social control: A test of the political, economic, and threat of black crime hypotheses." *Social Forces* 81(2):557–576.

Espy, M Watt. 2004. "Executions in the United States, 1608-2002: The Espy File [Computer file]. 4th ICPSR ed. Compiled by M. Watt Espy and John Ortiz Smykla, University of Alabama." *Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor]* .

FBI. 1988. Uniform Crime Reports: County Level Arrest and Offenses Data, 1977-1983. Technical report Federal Bureau of Investigation and US Dept of Justice and United States of America.

NCSC. 2016.
URL: *http://judicialselection.us/*