

Summer Paper Outline

Jose I. Velarde Morales

June 17, 2021

Contents

1	Introduction	1
2	Data	2
2.1	Sentencing Data	2
2.2	Judge Schedule Data	2
3	Model	2
4	Simulation	2
4.1	Estimated Parameters	3
5	Next Steps	3

1 Introduction

One of the chief criticisms of the US criminal justice system is that many states in the US have large regional variation in sentencing. This means that given the same crime, the probability of incarceration and the length of sentence varies considerably from county to county. To reduce variability in sentencing, many states (e.g. Florida, Minnesota, Washington) have introduced sentencing guidelines. These guidelines generally reduce the amount of discretion a judge has when deciding the length of a sentence. (cite hester) found that South Carolina has less county-level variation in sentencing than many states with sentencing guidelines.

A follow up study (cite hester 2017) suggests that this might be due to the practice of judge rotation in South Carolina. Judges in South Carolina don't sit exclusively in one county. Instead, they split their time amongst several counties in the state. For reference, the average judge hears cases in 12 counties throughout the year. Defendants are given some choice with regards to the date their case will be heard. Furthermore, defendants are made aware of which judges will be sitting in which counties in the near future. This gives them the ability to "shop" for judges, although subject to some constraints (e.g. judge availability).

The goal of this project is study how some system features affect sentencing outcomes. The system features we focus on are judge capacity and defendant choice. The outcomes we focus on are across county variation in sentencing and backlog of defendants. We study the system through simulation. We estimate relevant parameters using a rich dataset of sentencing data from South Carolina. The model we simulate builds on the work of (cite Cam).

2 Data

We have two main data sources: sentencing data and judge schedules, both of them are for the 2001 fiscal year. We obtained the data from the authors of (Hester 2017). We describe both datasets below.

2.1 Sentencing Data

The sentencing data contains information about 17,516 sentencing events in South Carolina from August 2000-July 2001. Each sentencing event contains an identifier for the judge who heard the case, the county the case was heard in, categorical variables describing the offense, and some defendant characteristics. A full list of the variables can be found in table (ref table). There are 50 judges and 46 counties in the dataset.

2.2 Judge Schedule Data

This dataset contains information about each judge’s assignment for each week of the fiscal year 2001. A snapshot of the calendar can be seen in figure (ref figure). Each assignment generally contains the county each judge was assigned to and the assignment type. The assignment type refers to the kind of cases a judge is scheduled to hear (civil or criminal).

3 Model

We use the model developed by . There are three agents: the judge, the defendant, and the prosecutor. The prosecutor proposes a plea offer, and the judge and the defendant choose whether to accept. The game evolves in the following steps:

1. The defendant chooses the judge/decides when to go to court
2. The prosecutor makes a plea offer
3. The defendant decides whether to accept the plea offer or go to trial

Each defendant is characterized by the following quantities: θ - the probability of conviction at trial, and τ - the expected sentence length if convicted in trial. Each defendant additionally has an idiosyncratic cost of going to trial, $c_d > 0$. A defendant will accept a plea offer, s , if $s \leq \theta\tau + c_d$.

Judges are modeled by their harshness, h . Given a defendant, $\theta\tau$ and a plea offer, s , the lowest plea offer the judge will accept is denoted by $l_j(\theta\tau)$ and the maximum plea offer the judge will accept is denoted by $u_j(\theta\tau)$.

As a result, given a specific judge j , the optimal sentence for the prosecutor to offer is $s^* = \min(\theta\tau + c_d, u_j(\theta\tau))$. This quantity is also the defendant’s cost.

4 Simulation

We simulate the system using the discrete choice model described in Section 3. Our simulation is as follows:

1. We set the time horizon, T , and the choice window for defendants, r .
2. At the beginning of the simulation, we randomly assign judges to counties in each period t , assigning one judge per county.
3. At each time period, for each county we simulate defendant arrivals. We draw defendants from each county's past defendants. Let S_i denote the set of judges scheduled for the next r weeks when defendant i arrives. The defendant then chooses a judge, j as follows: $j(i) = \arg \min_j \min(\theta\tau + c_d, u_j(\theta\tau)) + k(j)d$

4.1 Estimated Parameters

For our simulation, we estimate the following parameters from the data:

- θ - We estimate this using the sentencing data. We train a logistic regression model to predict this quantity for every defendant using offense type, race, and other variables.
- τ - We estimate this using a negative binomial regression model. This model is commonly used to predict this quantity in the criminology literature (see [?])
- $u_j(\theta\tau)$ and $l_j(\theta\tau)$ - we estimate these functions using a nearest neighbor approach. Given $\theta\tau$, we find the k cases most similar to that in the judge's history, and set l and u to be the minimum and maximum plea, respectively.
- μ_t and μ_p - these are judges trial and plea processing rates respectively. We jointly estimate these two parameters using an ad-hoc algorithm we developed. We find that trials have an average processing rate of about 12 days and that judges can process 13 pleas per day.
- c_d - we assume this is uniform across all defendants, and estimate it by looking at the cases where $s < u_j(\theta\tau)$. Since $s = \min(\theta\tau + c_d, u_j(\theta\tau))$, if $s < u_j(\theta\tau)$ then $c_d = s - \theta\tau$.

5 Next Steps

The next steps in our project are to refine our parameter estimates and to improve to the simulation.