Simulation Study Overview

September 23, 2021

Contents

1 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 and Hartman 2017. We briefly describe both datasets below.

1.1 Sentencing Data

The sentencing data contains information about 17,671 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 (e.g. race, age, and criminal history). There are 50 judges and 46 counties in the dataset. A full list of the variables we use can be found in Table 1.

Table 1: Sentencing Data Variables

Variable	Description
date	date of sentence
county	county where the sentence was decided
circuit	circuit court where sentence was decided
judge	numerical identifier for the judge who head the case
trial	binary variable indicating whether the case went to trial
incarc	binary variable indicating whether the sentence includes incarceration
statute	the code of law that the offender broke (e.g. 56-05-0750(B)(1))
offdescr	a description of the offense (e.g. driving under the influence)
counts	numerical variable
statute_first	the code of law that the offender broke, identical to 'statute' variable
offdescr_first	a description of the first offense, similar to offdescr
offtyped	detailed categorical offense type, there are 10 values (rape, assault, theft, etc).
sgc_offcode	a numerical code that is used to determine the minimum sentence multiplier
offtypeLibHyp	categorical offense type (property, violent, drug, other)
offser	offense seriousness, ranges from 1-8
ccpnts	numeric values between 1 and 68
ccpts99	commitment score, an alternative measure of offense seriousness.
crimhist	defendant criminal history, takes 5 categorical values
ppoints	numerical values between 0-1014
male	binary variable indicating sex
age	age of defendant at sentencing, ranges 15-81
black	binary variable indicating whether defendant is black
sentence	length of sentence in months, ranges 0-11988
expmin	expected minimum sentence

1.1.1 Sample

In the raw data, there are 51 distinct judge ID's. However, according to Hester and Hartman 2017, judge 1 is a combination of several judges that had few sentencing events. As a result, we exclude judge 1 from our sample, which leaves us with 17,516 observations.

1.1.2 Imputation of Missing Dates

After removing the sentencing events heard by Judge 1, there are 1,482 sentencing events with the *date* field missing. This is the only variable with missing values in the entire dataset. We impute the dates for these events in the following way. Recall that the judge name and county are still available for sentencing events with missing dates. In general, when imputing the dates for sentencing events with missing dates, we first use the master calendar to determine a set of possible dates in which the sentencing event could have occurred, and then we assign the events with missing dates as evenly as possible across these possible dates. So, for example, if there are 5 sentencing events with missing dates and 4 potential dates, then each of the potential dates would get assigned a sentencing event, and the first of the potential dates would get the remaining one. Suppose that judge j has m sentencing events with missing dates in county c (which is in circuit k). The set of potential dates we would assign these sentencing events to would evolve in the following way. In each category, the days are ordered from earliest to latest.

1. **Matching county, GS:** Days in the master calendar in which judge j had a "GS" assignment to county c.
2. **Matching county, non-GS:** Days in the master calendar in which judge j had a non-"GS" assignment to county c.
3. **Matching circuit, GS:** Days in the master calendar in which judge j had a "GS" assignment to a county in circuit k.
4. **Matching circuit, non-GS:** Days in the master calendar in which judge j had a non-"GS" assignment to a county in circuit k.
5. **Sentencing data match:** Days in which we observe sentencing events in the sentencing data for judge j in county c.
6. **Any day, GS:** Days in the master calendar in which judge j had a "GS" assignment to any county.

So, first, we would try to assign the sentencing events to days in the first set, if that set is empty, we would move on to the next set until we found a non-empty set. Using this method, we are able to impute the missing dates for all sentencing events with missing dates. Table 2 contains the distribution of the imputation method used for pleas with missing dates.

Table 2: Distribution of missing events

Imputation Group	**Share of Pleas**
1. Matching county, GS	0.825
2. Matching county, non-GS	0.024
3. Matching circuit, GS	0.02
4. Matching circuit, non-GS	0.006
4. Sentencing data match	0.026
5. Any day, GS	0.098

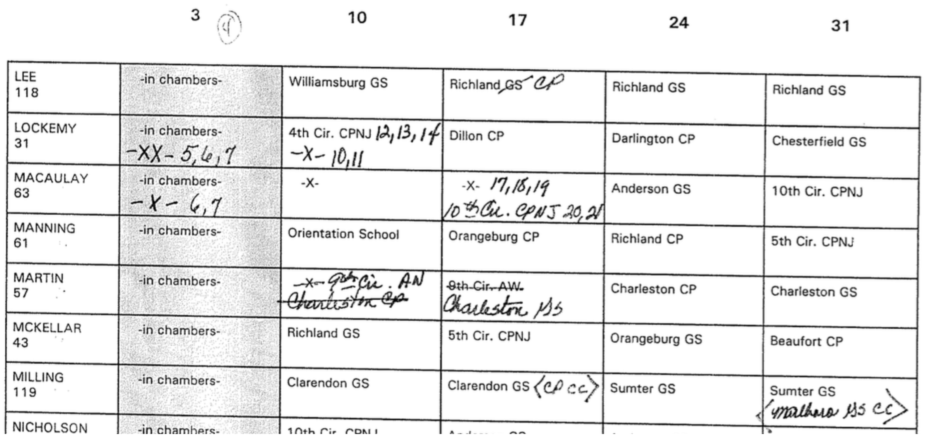
1.2 Master Calendar Data

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 1. 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).

1.3 Assignments

The judges were generally assigned to either specific counties or to circuit courts. In the master calendar data, 75% of all judge assignments were to either a specific county or to a circuit court. The remaining assignments were to one of the categories in Table 3. As can be seen, outside of "in chambers" they generally pertain to special circumstances.

Figure 1: Snapshot of Judge Calendar



	3 (4)	10	17	24	31
LEE 118	-in chambers-	Williamsburg GS	Richland ~~GS~~ CP	Richland GS	Richland GS
LOCKEMY 31	-in chambers- -XX- 5,6,7	4th Cir. CPNJ 12,13,14 -X- 10,11	Dillon CP	Darlington CP	Chesterfield GS
MACAULAY 63	-in chambers- -X- 6,7	-X-	-X- 17,18,19 10th Cir. CPNJ 20,21	Anderson GS	10th Cir. CPNJ
MANNING 61	-in chambers-	Orientation School	Orangeburg CP	Richland CP	5th Cir. CPNJ
MARTIN 57	-in chambers-	~~-X-~~ 9th Cir. AW ~~Charleston CP~~	~~8th Cir. AW~~ Charleston GS	Charleston CP	Charleston GS
MCKELLAR 43	-in chambers-	Richland GS	5th Cir. CPNJ	Orangeburg GS	Beaufort CP
MILLING 119	-in chambers-	Clarendon GS	Clarendon GS <CP CC>	Sumter GS	Sumter GS <Marlboro GS CC>
NICHOLSON	-in chambers-	10th Cir. CPNJ			

Table 3: Non county or circuit court assignments

Assignment
In chambers
Orientation School
Medical
Family Death
Sick
Military
X

1.3.1 Assignment Types

Judge calendar assignments generally have an acronym indicating the type of the assignment (e.g. Marion GS). There are eight different assignment types, each with a corresponding acronym. Assignments can have more than one type (e.g. Marion GS CC). Our study focuses on criminal cases, which are heard in general sessions (GS). Some assignment types, like CP (common pleas), are for civil cases. A full list of the assignment type acronyms and their meanings can be found in Table 4.

Table 4: Assignment Type Acronyms. The share column specifies the share of all assignments that include that assignment type.

Acronym	Meaning	Share
GS	General Session	0.36
CC	Circuit Court	0.02
SGJ	State Grand Jury	0.01
CP	Common Pleas	0.26
CPNJ	Common Pleas Non-Jury	0.13
PCR	Post Conviction Relief	0.03
Capital PCR	Capital Post Conviction Relief	< 0.01
AW	Administrative Week	0.01

1.3.2 Parsing

As can be seen from Figure 1, the assignments in the raw calendar data are on a weekly basis. However, several parts of our analysis stood to benefit from having the daily assignments for each judge, as opposed to only the weekly assignment. For example, this allowed us to more accurately count the total number of days a judge was assigned to a county, or to more accurately determine if there was discrepancy between the sentencing data and the calendar data. The goal of the parsing was to turn the weekly assignemnts in the calendar data into daily assignments. In other words, we wanted to determine where each judge *should have been* each day of the year (according to the calendar data). The end goal of the parsing is to create a dataset that contains at least one observation for each judge's location in each day of the year. This dataset will represent where the judge should have been according to the calendar data.

A judge can be assigned to one or multiple counties in a week, and we observe 5 different kinds of assignments: single assignment, single assignment with dates, multiple assignment, multiple assignment with some dates, and multiple assignment with all dates. We parse the calendar data to determine each judge's schedule for every day of the year. In the resulting data, the unit of observation is judge, day, county. Note that the only potentially ambiguous assignment types are multiple assignment and multiple assignment with some dates. By ambiguous, we mean that the calendar data does not specify a unique location for a judge. For example, in the single assignment case the judge is only supposed to be in one county (according to the calendar data). However, in the multiple assignment case, it is not clear in which county the judge should be (according to the calendar data). For these ambiguous assignment types, we set the judge to be in all counties that he is scheduled to be in. One portion of our analysis involves counting the number of days that a judge was assigned to a specific county. For days in which a judge is assigned to multiple counties, we normalize the number of days he/she is assigned to each county so that the sum of the assignments for each day is equal to one. For example, if judge 2 is scheduled for both Marion and Horry in the week of March 19-March 23, for each day of that week our data would include an observation for judge 2 in Horry and an observation for judge 2 in Marion. Furthermore, for each of these days we would say that judge 2 was assigned to 0.5 days in Horry and 0.5 days in Marion. An example of what our parsed data looks like can be seen in table 6.

Table 5: Assignment Types

Assignment	**Example**
Single	Marion
Single with dates	Marion 23, 24, 25
Multiple	Marion, Horry
Multiple some dates	Marion 23, 24, Horry
Multiple all dates	Marion 23, 24, Horry 25, 26, 27

Table 6: Example of Parsed Calendar Data, here Judge 3 is scheduled for both Horry and Greenville on March 19.

Judge	**Day**	**County**
2	2001-03-19	Marion
3	2001-03-19	Horry
3	2001-03-19	Greenville

1.4 Merging the Sentencing Data and the Calendar Data

The main challenge in using the two datasets is that they don't share a common identifier for judges. The sentencing data uses numerical IDs to identify judges, while the calendar data uses judge names to identify them. There is no mapping from judge numbers to judge names in the raw data. To map judge ID's to judge names, we assign each Judge ID to the judge name in the schedule data whose schedule overlaps the most with that Judge ID. We consider the sentencing data and schedule data to overlap for a Judge ID and a Judge name in a specific week if the Judge ID in consideration has a sentencing event in a county that that judge name is assigned to in that week. So, if in Week 12 judge 'John Doe' is assigned to 'Aitken' county, and we see a sentencing event for Judge ID n in 'Aitken' county on Week 12, we would consider their schedules to be overlapping in that week. Assigning each judge ID to the judge name with the most overlap yields a unique mapping (i.e. there are never any ties). We also tried two other methods, and they yielded the same results. The two other methods as well as the final mapping can be found in Appendix D.

2 Model

We build on the model developed by Wang 2019. 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, $c_d > 0$ (we relax this assumption later). 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. Finally, the case goes to trial if $\theta\tau + c_d < l_j(\theta\tau)$.

3 Estimation

For the simulation, we have to estimate several of the model's parameters. In this section, we provide a detailed description of how we estimate each parameter, including the sample used.

3.1 Assumptions/Restrictions

Table 7: Assumptions/Restrictions Used

Assumption/Restriction	**Section Used**
We only use the trial sentencing events to estimate θ	3.2
We only use the trial sentencing events to estimate τ	3.3
We currently only use a subset of the data to estimate c_d	3.4
We add the point $(0, 0)$ to the convex hull of every judge	3.5
We assume (TBD) when using regression model	3.7

3.2 Probability of Conviction at Trial - θ

We currently estimate this using l_2 regularized logistic regression with a penalty parameter of 1. We use the following variables to predict this quantity: Black, Offense Type, and Offense Seriousness. We represent all variables as binary variables. We split our sample into train and test sets using a 75/25 split to evaluate the performance of our classifier. Our classifier's performance can be seen in table 8. The classifier outputs prediction probabilities, which is what we use for θ. The classifier's prediction probability can be interpreted as the probability that the observation will have a value of 1 for the incarceration variable. The final classifier we use to predict θ for all observations in our dataset is trained on the full training set.

Table 8: Evaluation Metrics for Classifier

Metric	**Score**
AUC	0.93
F1	0.94
Accuracy	0.89

3.2.1 Sample

We only use the observations in the data that are trials and that have non missing values for the predictor variables. There are only 256 observations in the data meeting this criteria. Of these 256 observations, 247 resulted in a conviction.

3.2.2 Proposed Changes

We should consider jointly estimating θ and τ using a Hurdle Regression model as in Hester and Hartman 2017.

3.3 Expected Sentence Length if Convicted - τ

We currently estimate this using Negative Binomial Regression. We use the Cameron-Trivedi test for overdispersion to choose the overdispersion parameter. As in Hester and Hartman 2017, our dependent variable is the expected minimum sentence. We use the following variables to predict this quantity: Black, Age, Offense Type, and Offense Seriousness.

3.3.1 Sample

We only use the observations in the data that are trials and that have non missing values for the predictor variables. There are only 256 observations in the data meeting this criteria.

3.3.2 Proposed Changes

We should consider jointly estimating θ and τ using a Hurdle Regression model as in Hester and Hartman 2017.

3.4 Defendant Cost of Trial - c_d

We are currently estimating this using only the first method described in the Write-Up. In this method, we use the subset of cases where the sentence, s, is less than $u_j(\theta\tau)$. In these cases, our model implies that $c_d = s - \theta\tau$.

3.4.1 Proposed Changes

We should implement the maximum likelihood estimation procedure described in Nasser's document for the other cases. Here's the discussion: Recall that $i = 1, \ldots, I_j$ are the plea bargains judge j oversaw, and that their sentence is given by $s_i = \min(\theta_i\tau_i + c_d(i), u_i)$, where $u_i = u_j(\theta_i\tau_i)$ and $u_j(\cdot)$ is defined above. We define the sets $\mathcal{I}_j^1$ and $\mathcal{I}_j^2$ as follows:

$$\begin{aligned} \mathcal{I}_j^1 &= \{i = 1, \ldots, I_j : s < u_i\}, \\ \mathcal{I}_j^2 &= \{1, \ldots, I_j\} \setminus \mathcal{I}_j^1 = \{i = 1, \ldots, I_j : s_i = u_i\}. \end{aligned}$$

Then, we can infer $c_d(i) = s_i - \hat{\theta}_i\hat{\tau}_i$ for $i \in \mathcal{I}_j^1$. On the other hand, we can only infer that $c_d(i) \geq u_i - \hat{\theta}_i\hat{\tau}_i$ for $i \in \mathcal{I}_j^2$.

Next, we let K_j denote the number of trials judge j oversaw, which are indexed by $k \in \mathcal{K}_j = \{1, \ldots, K_j\}$. Recall that a case goes to trial if $\theta_i\tau_i + c_d(i) < l_i$. Thus, for $i = 1, \ldots, K_j$, we infer that $c_d(i) < l_i - \hat{\theta}_i\hat{\tau}_i$.

In summary, although we can impute $c_d(i)$ exactly for $i \in \mathcal{I}_j^1$, we can only infer that $c_d(i)$ falls into an interval for $i \in \mathcal{I}_j^2 \cup \mathcal{K}_j$. However, assuming a parametric distribution for c_d, e.g., $c_d \sim N(\mu, \sigma^2)$, we can estimate its parameters using maximum likelihood. To this end, we let F and f denote the cdf and pdf of the distribution of c_d, and define the likelihood function $\mathscr{L}_j$ as follows:

$$\mathscr{L}_j = \prod_{i \in \mathcal{I}_j^1} f(s_i - \hat{\theta}_i \hat{\tau}_i) \prod_{i \in \mathcal{I}_j^2} \bar{F}(u_i - \hat{\theta}_i \hat{\tau}_i) \prod_{i \in \mathcal{K}_j} F(l_i - \hat{\theta}_i \hat{\tau}_i).$$

Then, we let $\mathscr{L} = \prod_{j=1}^J \mathscr{L}_j$ and argmax $\mathscr{L}$ helps us choose the parameters of the distribution F.

Remark 1. Note that the approach proposed immediately above allows c_d to be negative. We can attribute this to $c_d(i)$ being the cost of trial plus an idiosyncratic shock specific to defendant i, which encompasses all factors that are not captured explicitly in the model.

Remark 2. We envision drawing defendant profiles (along with their cases) with replacement from the dataset at a particular weekly rate in our simulation study. As we do so, we can set $c_d = c_d(i)$ whenever we draw a defendant who falls into the set $\mathcal{I}_j^1$ for some judge j. Otherwise, we can consider the following two cases:

Case i) Defendant i is in the set $\mathcal{I}_j^2$ for some judge j. Then, we draw $c_d(i)$ from the conditional distribution $F(x|x \geq u_j - \hat{\theta}_i \hat{\tau}_i)$.

Case ii) Defendant i is in the set K_j for some judge j. Then, we draw $c_d(i)$ from the conditional distribution $F(x|x \leq l_j - \hat{\theta}_i \hat{\tau}_i)$.

3.5 Judge maximum and minimum plea - $l_j(\theta\tau), u_j(\theta\tau)$

Mathematical Description Recall that a key quantity for us is $\theta\tau$. Fix a judge, say judge j, and focus only on the pleas he sentenced. For each such case, we can calculate $\hat{\theta}\hat{\tau}$ using the estimates from the hurdle model described in Section (tbd). Suppose judge j handled I_j such cases, indexed by $i = 1, \ldots, I_j$. Let $\mathcal{A}_j$ denote the convex hull of the origin $(0,0)$ and the points $(\hat{\theta}_i \hat{\tau}_i, s_i)$ for $i = 1, \ldots, I_j$. Given the set $\mathcal{A}_j$, we impute $l_j(\cdot)$ and $u_j(\cdot)$ as functions of $\theta\tau$ as follows:

$$\begin{aligned} l_j(\theta\tau) &= \min\{y : (\theta\tau, y) \in \mathcal{A}_j\} \\ u_j(\theta\tau) &= \max\{y : (\theta\tau, y) \in \mathcal{A}_j\} \end{aligned}$$

Implementation Suppose we have the simplices of the convex hull for a specific judge, and that they are given as tuples of two points $((x_1, y_1), (x_2, y_2))$. Given a specific $\theta\tau$ we implement the convex hull approach by iterating over the simplices of the convex hull, and finding the two simplices on which $\theta\tau$ lies. In other words, we find the two line segments on the boundary of the convex hull whose domain includes $\theta\tau$. We then find the value of the convex hull boundary at $\theta\tau$ in each of these two line segments using linear interpolation. The largest (smallest) of these two values is $u_j(l_j)$. In cases where $\theta_1\tau_1$ is not in the domain of the convex hull because it is too large, we use the largest value in the convex hull to find $u_j(l_j)$. $\theta_1\tau_1$ will never be too small because all convex hulls include $(0,0)$ and $\theta_1\tau_1 \geq 0$. Figure 2 contains an illustration:

Note: we should think about what to do in cases where $\theta_1\tau_1$ is not in the convex hull. Figure 2 contains an illustration:

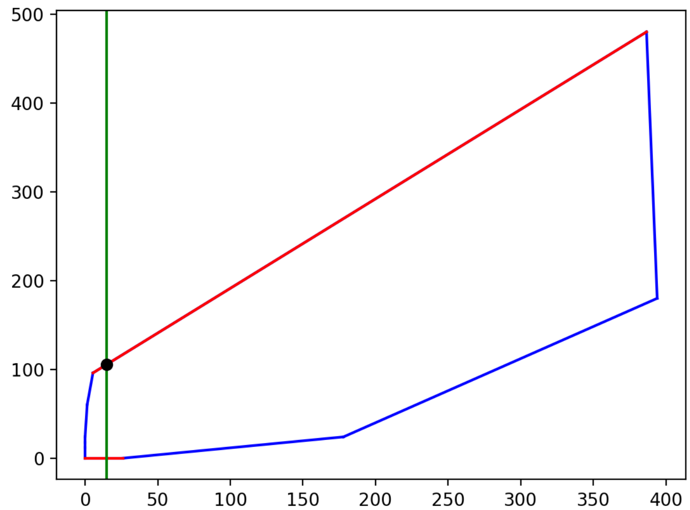


Figure 2: Example of convex hull approach. The convex hull is in blue, the two line segments containing $\theta\tau$ are in red, and the green line intercepts the x-axis at $\theta\tau$. The black dot would be the maximum plea.

3.5.1 Sample

Here, to estimate $l_j(\theta\tau), u_j(\theta\tau)$ for a specific judge, j, we use all of pleas judge j heard. In other words, we exclude all of j's trials. We also add the point $(0, 0)$ to the convex hull of every judge.

3.6 County Arrival Rates - λ_c

We set each county's arrival rate to be equal to the average number of sentencing events per week in that county, as observed in the data. Let N_{cw} denote the number of sentencing events in county c in week w, county c's arrival rate is defined as: $\lambda_c = \frac{\sum_w N_{cw}}{\sum_w 1}$. We are currently using all pleas and all weeks in our data to calculate this.

3.7 Service Rates - μ_p, μ_t

3.7.1 Samples

Plea MLE Sample This is the sample we use for the maximum likelihood estimation part of the ad-hoc algorithm. We also refer to this sample as "clean days". For this, we only consider pleas that happened on days which satisfy the following conditions:

Table 9: MLE Plea Sample Exclusion Criteria

Condition
No inconsistencies between sentencing data and calendar
Judge has at least 10 'clean' days
Judge has at least one sentencing event that day
Judge is only assigned to one county that day
Judge only sentences in one county that day
Judge never has more than 35 sentencing events in this county
Judge calendar assignment is of type "GS"

Plea Arrival Rate Sample This is the sample we use to estimate the plea arrival rate, θ in the ad-hoc algorithm. The calculation of θ involves two quantities: the total number of pleas in the data, N_p, and d, the total number of judge days. d is meant to represent the number of days in which a judge could have been working on pleas. As a result, for d we include all days of type "GS", and we include days of other work types in which we observe sentencing events. N_p currently includes all pleas in our data.

Trial Rate Sample This is the sample we use to estimate the trial service rate, μ_t. When calculating the total number of days a judge was assigned to a county, we include all days he had a "GS" assignment to that county in the master calendar, and all days of type non-GS in which we observe a sentencing event. Note, this is the same criteria as used above for the plea arrival rate sample. We include all pleas and all trials when calculating the total number of pleas and trials the judge heard in that county. We focus on the judge-county combinations that satisfy the following conditions:

Table 10: Trial Service Rate Exclusion Criteria

Condition
Judge never has more than 35 sentencing events in one day in this county Judge has at least 2 trial in this county

3.7.2 Estimation of μ_t

To estimate the trial service rate, we focus on the judge-county combinations that satisfy the conditions described in the Trial Rate Sample paragraph. Let K denote the number judge-county combinations satisfying these two conditions. We number these judge-county combinations $1, \ldots, K$ and define $\mathcal{K} = \{1, \ldots, K\}$. For judge-county $k \in \mathcal{K}$, we let $n_p(k)$ and $n_t(k)$ denote the total number of pleas and the total number of trials undertaken by this judge in this county. Similarly, for judge-county $k \in \mathcal{K}$, we let $T(k)$ denote the number of days this judge was assigned to this county.[1]

First, we assume the judges in judge-county combinations $k \in \mathcal{K}$ never idle. If this assumption was correct, the trial service rate of judge-county $k \in \mathcal{K}$ would be

$$\hat{\mu}_t(k) \ = \ \frac{n_t(k)}{T(k) - n_p(k)/\hat{\mu}_p}.$$

[1]In calculating $T(k)$ for $k \in \mathcal{K}$, we assume the judge divides his time equally among the county assignments to which he is assigned if he is assigned to multiple counties on a day.

The tuples $(n_t(k), \hat{\mu}_t(k))$ for $k \in \mathcal{K}$ are depicted in Figure 3.

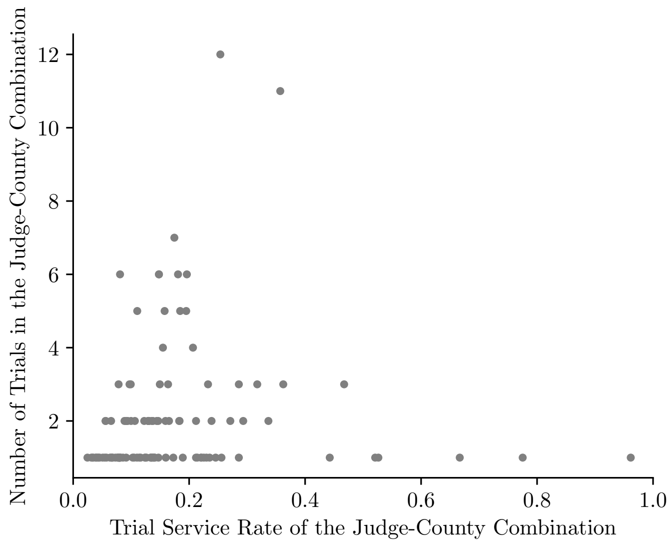


Figure 3: A graphical representation of the tuples $(n_t(k), \hat{\mu}_t(k))$ for $k \in \mathcal{K}$.

We observe that as the number of trials in a judge-county combination decreases, the variance of the $\hat{\mu}_t(k)$ estimates increase. Moreover, the $\hat{\mu}_t(k)$ estimates skew to the left as the number of trials decreases. This suggests that the judges in judge-county combinations with few trials spent some time idling. Therefore, to estimate the trail service rate, we focus on judge-county combinations for which we observe at least two trials, i.e., $k \in \tilde{\mathcal{K}} = \{k : k \in \mathcal{K}, n_t(k) \geq 2\}$. These judge-county combinations account for 72% of the trials in the dataset. The trial service rate estimate is

$$\hat{\mu}_t = \frac{\sum\limits_{k \in \tilde{\mathcal{K}}} n_t(k)}{\sum\limits_{k \in \tilde{\mathcal{K}}} T(k) - \sum\limits_{k \in \tilde{\mathcal{K}}} n_p(k)/\hat{\mu}_p}.$$

3.7.3 Ad-hoc Algorithm for Joint Estimation of μ_t, μ_p

Step 1: Let μ_p, μ_t be the current values for the plea and trial service rate. As in the estimation of μ_t, we are assuming judges only work on pleas and trials and do not idle. As a result, given the total number of trials heard, N_t, and the trial service rate, μ_t, we can calculate the expected number of days judges spent working on trials. The number of days judges spent on trials, $d_t = \frac{N_t}{\mu_t}$. We calculate the total number of days judges worked, d using the assignments from the master calendar and removing public holidays. The expected number of days judges worked on pleas is then $d_p = d - d_t$.

Step 2: Let N_p denote the total number of pleas in the data. Again, we include all pleas in our sample to calculate N_p. We set $\theta = \frac{N_p}{d_p}$. We model the plea demand for a judge as $D \sim \text{Poisson}(\theta)$, whereas the number of pleas a judge can serve in a day is denoted by X, $X \sim \text{Poisson}(\mu_p)$.

Step 3: Let $S_i = \min(D_i, X_i)$ denote the number of pleas sentenced for judge-day combination $i = 1, ..., N$. Here, we only include the judge-day combinations that satisfy our Plea MLE conditions.

We have that

$$
\begin{aligned}
P(S_i = S) &= P(X_i = S|X_i \leq D_i)P(X_i \leq D_i) + P(D_i = S|X_i > D_i)P(X_i > D_i) \\
&= \frac{\theta^s e^{-\theta}}{s!}[1 - \sum_{k=0}^{s-1} \frac{\mu_p^k e^{-\mu_p}}{k!}] + \frac{\mu_p^s e^{-\mu_p}}{s!}[1 - \sum_{k=0}^{s} \frac{\theta^k e^{-\theta}}{k!}]
\end{aligned}
$$

Let $L(\mu_p) = -\sum_{i=1}^{N} \log P(S_i = s)$. We then set $\mu_p = \text{argmin}\, L(\mu_p)$ and calculate μ_t as described in 3.7.2. Again, the judge-day combinations for which we are minimizing the negative log likelihood are those that satisfy our Plea MLE conditions. We use a gradient descent algorithm with the Adam Optimizer to find the value of μ_p that minimizes the NLL. This new value of μ_p will imply a new value of μ_t as in Section 3.7.2, and so we repeat Steps 1-3 until we converge.

4 Simulation Plan

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

State Variables For each judge, we keep track of their capacity in each period, including future periods, as well as their current and future schedules. For each county, we keep track of the judges that are scheduled to appear in every time period (including future time periods), the county's backlogged defendants, and the number of defendants that choose to go to trial.

Trial Scheduling We would schedule trials in the following way. First, for each county, we would randomly select the trial dates. Then, in each time period, we would randomly assign judges to counties as described below, if a judge is assigned to county on a day in which a trial is scheduled in that county, then the judge would stay there for 2 weeks (or however long trials take). As a result, when assigning judges and counties for the next period, the county and the judge would be removed from the list of available counties/judges. In other words, the assignment for that judge and county would already be determined for the next week as well. We could then assign the remaining judges/counties using the method described below.

Assigning judges to counties Here, we describe a method to assign a set of judges to a set of counties for T time periods. The time periods here are discrete and we think of them as weeks. We describe how the assignment would work for each individual week, but in practice all the assignments would be determined before running the rest of the simulation. Since there are 50 judges and 46 counties, in each period we would randomly draw, without replacement, 46 judges from the list of 50. Then, each of the selected judges will stay in his "home" county with probability η and with probability $1 - \eta$ he will be assigned to another county. We refer to judges that don't stay in their home county in a specific week as "rotating judges", and we refer to counties whose home judge will be rotating as "rotating counties". We assign rotating judges to rotating counties as follows: we randomly shuffle the rotating judges and the rotating counties are sorted alphabetically. So the rotating judge in the first position after shuffling would be assigned to county A, the second to County B, and so on.

Defendant Arrivals In each time period, t, we iterate over the different counties. We simulate defendant arrivals for a given county, c, as follows: first, we determine the number of arrivals, n_{ct}

by drawing from a Poisson distribution with mean λ_c. We then draw n_{ct} defendants from county c's past defendants as observed in the sentencing data. If the county has any backlogged defendants in that week, the backlogged defendants are added to that week's list of defendants, and are served first.

Defendant Choice Each defendant then chooses from the available judges that will be in county c in the next r weeks. The defendant chooses the judge, j that minimizes his expected cost, $\min(\theta\tau + c_d, u_j(\theta\tau)) + k(j)d$, where c_d is the defendant's cost of going to trial, $k(j)$ is the number of time periods until judge j will be in county c, and d is the cost of delay. Once a defendant chooses a judge, we reduce that judge's capacity for the week in which he will sentence that defendant. If there are no judges available in the next r weeks, the defendant is added to the county's list of backlogged defendants and processed again next week. If a defendant chooses to go to trial, that defendant disappears from the simulation, and all judges' capacity remains unchanged.

5 Next Steps

- Implement hurdle model for $\tau\theta$.
- Implement MLE estimation for c_d
- Implement changes to simulation

References

Hester, Rhys and Todd K Hartman (2017). "Conditional race disparities in criminal sentencing: A test of the liberation hypothesis from a non-guidelines state". In: *Journal of quantitative criminology* 33.1, pp. 77–100.

Wang, Can (2019). "Three Data-driven Model-based Policy Analyses in Criminology". PhD thesis. Stanford University.

A Sentencing Data Set

This section provides a more in-depth description of the sentencing data we used. All of the sentencing events in the data correspond to felonies or serious misdemeanors.

A.1 Data Cleaning

We discard all sentencing events in the sentencing dataset associated with Judge 1 because Judge 1 is a fictitious judge.[2]

A.2 Variables

There are two versions of the sentencing dataset: the CSV file and the STATA file. There are 17,671 sentencing events (rows) in both files. The CSV file has 33 data fields. This is the first dataset Larry obtained. The STATA file has 21 data fields. This is the second dataset Larry obtained. In what follows, we assume the sentencing dataset is cleaned as described in Appendix A.1.

[2] See the email from Rhys to Can and Larry on 05/14/2018 (Page 3 of the document Email3.pdf). In this email, Rhys states that "Judge 1 is actually a fictitious aggregate of the 155 cases that were missing a judge ID".

Table 11: Data fields of the CSV and STATA files.

Only CSV	Both	Only STATA
unnamed	datedisp/data File	expmin
statute_first (identical to statute)	circuit	jud_no
offdescr_first	county	
offtypeLibHyp	counts	
ccpnts	offser	
ccpts99	sgc_offcode	
trial	of_hom	
incarc	of_rape	
crimhist	of_rob	
ppoints	of_asslt	
male	of_burg	
age	of_dstrb	
black	of_possn	
judge	of_theft	
	of_fraud	
	of_other	
	realsent/sentence	
	statute (99.95% match)	
	offdescr (99.21% match)	

Table 11 lists the data fields that are common to the CSV file and the STATA file as well as the data fields that are unique to either the CSV file or the STATA file. As shown in Table 11, the CSV file and the STATA file have 18 data fields in common; see the middle column of Table 11. Of these 18 data fields that are in common across the two files, 16 of them are identical.[3] The remaining two common data fields are statute and offdescr. We observe that 99.95% of the statute entries and 99.21% of the offdescr entries are identical between the CSV and STATA files. Moreover, the 10 non-identical statute entries and the 141 non-identical offdescr entries seem to refer to the same statutes and offense descriptions, respectively; see Tables **??**-**??** in Appendix **??** for illustrative examples.

There are 14 data fields that appear only in CSV file; see the first column of Table 11. Similarly, there are 2 data fields that only appear in the STATA file: expmin and jud_no. The expmin data field is computed by Rhys; see item 1 in the STATA file description below. The jud_no data field of the STATA file provides a numeric identifier for the judge that sentenced each offender. Although these numeric identifiers are not identical to the numeric identifiers used in the judge data field of the CSV file, there is a one-to-one mapping between them. Consequently, we use the CSV file in our analysis (in the following sections) but supplement it with the expected minimum sentence data field from the STATA file, which is calculated using Equation (1); see Section A.2.2. Next, we describe the data fields in the CSV and STATA files in detail.

A.2.1 CSV File

The data fields of the CSV file are as follows:

[3] To be specific, after sorting the rows of the two files based on the 16 commons data fields, the entries of these 16 data fields match exactly.

Table 12: Variable descriptions for CSV file

Variable	Description
	An unnamed numeric identifier for the sentencing events.
date	Date of the sentence, ranging from 2000-07-07 to 2001-06-29.
county	County where the sentence was decided. There are 46 counties.
circuit	The circuit the county belongs to. There are 16 circuits (1-16).
judge	A numeric identifier for the judge who heard the case. There are 51 judges.
trial	Binary variable indicating whether the case went to trial.
incarc	Binary variable indicating whether the sentence included incarceration.
statute	The law the offender is accused of breaking, for example 56-05-0750(B)(1).
offdescr	A description of the offense (e.g. "Failure to Stop for a Blue Light")
counts	There are 29 distinct values, ranging between 0 and 60.
statute_first	The description of the first offense of the offender. This field is identical to 'statute'.
offdescr_first	A description of the first offense (e.g. "Pointing and Presenting Firearms at a Person")
sgc_offcode	286 distinct numeric values between 4 and 2878.
of_hom	Binary variable indicating the offense type is homicide.
of_rape	Binary variable indicating the offense type is rape
of_rob	Binary variable indicating the offense type is robbery
of_asslt	Binary variable indicating the offense type is assault
of_burg	Binary variable indicating the offense type is burglary
of_dstrb	Binary variable indicating the offense type is drug distribution
of_possn	Binary variable indicating the offense type is drug possesion
of_theft	Binary variable indicating the offense type is theft
of_fraud	Binary variable indicating the offense type is fraud
of_other	Binary variable indicating the offense type belongs to the category "other".
offtypeLibHyp	Categorical offense type. The categories: property, violent, other, and drug.
offser	Offense seriousness. There are 8 categories: 1 to 8. Categories are increasing in seriousness.
ccpnts	Numeric values between 1 and 68.
ccpts99	Commitment score. Numeric values between 1 and 12 in increasing order of seriousness.
ppoints	Numeric values between 0 and 1014.
male	Binary variable indicating whether defendant is male.
age	Age of defendant at sentencing.
black	Binary variable indicating whether defendant is black.
sentence	The length of the sentence in months. Ranges between 0 and 11988.

A.2.2 Stata File

Next, we describe the data fields of the STATA file. We only describe the two data fields that are specific to the STATA file. The remaining data fields have already been discussed above.

Table 13: Variable descriptions for STATA file

Variable	Description
expmin	The expected minimum sentence. Ranges between 0 and 470.
jud_no	A numeric identifier for the judges. There are 50 judges, numbered 1-54.[4]

A.2.3 Expected Minimum Sentence

The expected minimum sentence (reported in the STATA file) is calculated as follows:[5] The set of sgc_offcode values is partitioned into four subsets: $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$, and $\mathcal{C}_4$; see Hester and Hartman 2017 for the partitions. Let s denote the sentence length in months, c denote the sgc_offcode, and

$$m(c) = \begin{cases} 1.00 & \text{if } c \in \mathcal{C}_1, \\ 0.85 & \text{if } c \in \mathcal{C}_2, \\ 0.33 & \text{if } c \in \mathcal{C}_3, \\ 0.25 & \text{if } c \in \mathcal{C}_4, \end{cases}$$

denote the expected minimum sentence multiplier, i.e., the fraction of the sentence that has to pass before the defendant is eligible for parole. Then,

$$\text{Expected Minimum Sentence}\,(s,c) = \min\big\{\lceil m(c)\cdot s\rceil, 720\big\}. \tag{1}$$

The exceptions to this are as follows. Define the sets

$$\begin{aligned} \mathcal{A}_1 &= \{267, 383, 2362\}, \\ \mathcal{A}_2 &= \{454, 2360, 2362\}, \\ \mathcal{A}_3 &= \{79, 90, 113, 114, 139, 217, 280, 283, 312, 368, 387, 388, 389, 392, 395, 402, 457, 2356, 2359\}, \\ \mathcal{A}_4 &= \{116, 148, 281, 284, 349, 370, 452, 456, 2417\}, \end{aligned}$$

where $\mathcal{A}_i \subset \mathcal{C}_i$ for $i = 1, 2, 3, 4$. According to Hester and Hartman 2017, a mandatory minimum sentence is imposed if $c \in \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4$.

A.3 Missing Data

A.3.1 Variables with missing data

After excluding all of the sentencing events heard by Judge 1, the fictitious judge, there are 17,516 sentencing events. Amongst these sentencing events, the only data field with missing values are the *date* data field and the *expmin* data field. 1482 sentencing events are missing the *date* data field and 28 sentencing events are missing the *expmin* data field.

A.4 Events with missing dates

About 8.7% of the sentencing events (1551 sentencing events) are missing the *date* data field. This section further analyzes these sentencing events to infer the missing date from the available judge and county information.

We start by cleaning the sentencing dataset as described in Appendix A.1. Then, we fix a sentencing event that is missing the *date* data field. We record the judge number and the county associated with this sentencing event. Next, we use the master calendar (and the mapping from judge numbers to judge names described in Section **??**) to find the set of weeks on which this judge visited this county. Let us illustrate this through an example: One of the sentencing event missing the *date* date field is undertaken by Judge 4 in Newberry. According to Section **??**, Judge 4 is Judge

[5] See the code provided on `https://tkhartman.netlify.app` (accessed on October 17th, 2020) for Hester and Hartman 2017.

Baxley. Judge Baxley visited only Newberry in the week of November 6th 2000. Therefore, the set associated with this sentencing event is {November 6th 2000}. We continue this procedure for all sentencing events missing the *data* data field.

Focusing on the sentencing events who are missing the *date* data field, Figure 4 displays a histogram of the number of visits to the sentencing county by the sentencing judge using his schedule from the master calendar. For example, the first bar indicates that there are 222 sentencing events (missing the *date* data field) that occurred in a county to which the judge was never assigned (according to the master calendar). For 34 out of these 222 sentencing events, the judge visited the circuit court associated with the sentencing event in least one week. For 100 sentencing events, the judge visited the county (associated with this sentencing event) only once. We *assume* these sentencing events occurred during this week. For example, we assume the sentencing event discussed in the previous paragraph occurred in the Week of November 6th 2000.

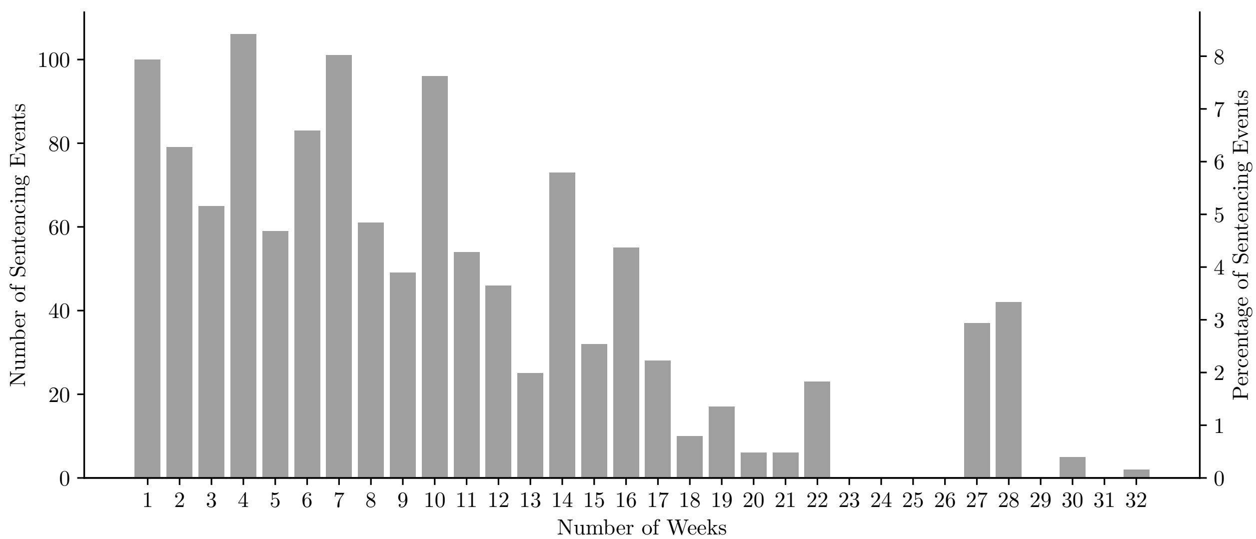


Figure 4: A histogram of the number of weeks the judge visited the county in which the sentencing events missing the *date* data field occurred.

A.5 Summary Statistics

Master

B Calendar Data

B.1 Data Cleaning

We do two things to clean the calendar data:

1. We discard all judge names in the master calendar that are not mapped to a judge number in the sentencing dataset. We discard these judge names because we do not observe any sentencing events for them. The discarded judge names are as follows: Bergdorf, Couch, Drew, Gier, Peeples, Simmons, Watts, and Young. Since we are using the same sentencing data as Hester and Hartman 2017, we can conclude that they also excluded these judges.

2. We combine the schedules of Judges "Cooper" and "Cooper, TW" in the calendar data. There are three judges in the calendar data with last name Cooper: (i) Cooper, (ii) Cooper, TW, and (iii) Cooper, GT. If we run the algorithm proposed in Section **??** using the master calendar with three distinct judges with last name Cooper, Judge 14 is mapped to either Judge Cooper (under the perfect match criterion) or Judge Cooper, TW (under the overlapping match criterion). Table 14 depicts the schedules of the three judges with last name Cooper (according to the master calendar) as well as the counties visited by Judge 14 (according to the sentencing dataset). The counties to which either Judge Cooper or Judge Cooper, TW are assigned are depicted in blue. The counties to which Judge Cooper, GT is assigned are depicted in green. The counties to which neither Judge Cooper, nor Judge Cooper, TW or Cooper, GT are assigned are depicted in red. Table 14 shows that the schedules of Judges Cooper and Cooper, TW do not overlap. Moreover, their combined schedules resemble the schedule of Judge 14 (the list of counties visited by Judge 14) the most. This suggests that Judge Cooper and Judge Cooper, TW are (most likely) the same person. Therefore, we combine the schedules of Judges "Cooper" and "Cooper, TW".[6]

[6]Incidentally, only two judges with last name Cooper are listed on the South Carolina Judicial branch's website. These judges are Thomas W. Cooper, Jr. and G. Thomas Cooper, Jr.; please see `https://www.sccourts.org/circuitCourt/displaycirjudge.cfm?judgeid=2054` and `https://www.sccourts.org/circuitCourt/displaycirjudge.cfm?judgeid=2126` (accessed on December 14th, 2020) for their bios.

Week	Judge 14	Cooper	Cooper, TW	Cooper, GT
July 10		Aiken		
July 17				Berkeley
July 24				Richland
July 31		Sumter		Greenville
August 7	Sumter	Sumter		Greenwood
August 21	Greenville	Sumter		Greenville
August 28	Williamsburg, Sumter	Williamsburg, Orangeburg		Richland, Greenville
September 4	Williamsburg, Greenville	Williamsburg		Greenville
September 11	Sumter, Clarendon, Greenville	Sumter		Greenville
September 18				Spartanburg
September 25	Greenville	Williamsburg		Greenville
October 9	Lee, Sumter	Lee		Richland
October 16	Williamsburg	Williamsburg, Aiken		Richland
October 23	Williamsburg, Horry, Georgetown	Williamsburg		
October 30	Sumter	Sumter		Richland
November 6		Williamsburg		Richland
November 13	Sumter	Sumter		
November 27		Clarendon		Newberry
December 4	Lee, Sumter	Lee		
December 11	Lee	Lee, Aiken		
January 1			Aiken	
January 8	Richland		Richland, Aiken	McCormick
January 15			Kershaw	
January 22			Aiken	
January 29	Aiken		Beaufort, Aiken	Lexington
February 5			Orangeburg	Laurens
February 12				Newberry
February 26	Richland		Richland, Orangeburg	Greenville
March 5	Richland, Edgefield			McCormick, Edgefield
March 12			Orangeburg	
March 19	Orangeburg		Orangeburg	
March 26	Richland		Richland	Edgefield
April 9	Richland		Richland	Lexington
April 16	Richland			
April 23			Richland	Lexington
April 30	Kershaw		Kershaw	Lexington
May 14				Lexington
May 21	Richland		Richland	Edgefield
May 28	Richland		Richland	
June 4	Richland		Richland	Edgefield
June 11				Edgefield
June 18	Kershaw		Kershaw	Lexington
June 25			Richland	McCormick

Table 14: The list of counties visited by Judge 14 (according to the sentencing dataset) and the counties visited by Judges Cooper, Cooper, TW, and Cooper, GT (according to the master calendar). The counties visited by Judge 14 to which either Judge Cooper or Judge Cooper, TW were assigned are written in green font. The counties visited by Judge 14 to which Judge Cooper, GT was assigned are written in blue font. The counties visited by Judge 14 to which neither Judge Cooper, nor Judge Cooper, TW or Cooper, GT were assigned are written in red font. For visual clarity, only the weeks in which at least one judge has an assignment or sentencing event are depicted.

C Discrepancies between the Sentencing Data and the Master Calendar Data

This section studies the discrepancies between the sentencing data and the calendar data. In doing so, we take advantage of the mapping from judge numbers to judge names (discussed in Section 1.4).

C.1 No Assignments Case

There are 110 judge-week combinations in this category. We observe no sentencing events (in the sentencing dataset) for 101 out of these 110 (91.8%) judge-week combinations. Recall that the judges in these combinations have no assignments on the master calendar. Therefore, there are no inconsistencies between the master calendar and the sentencing dataset for these 101 judge-week combinations. The remaining 9 judge-week combinations are listed in Table 15. We observe at least one sentencing event (in the sentencing dataset) for these judge-week combinations. Each judge visits only one county (according to the sentencing dataset) in these 9 judge-week combinations.

Table 15: Judge-week combinations in which the judge has sentencing events in a county to which he is not assigned - "No Assignment" category. The first number in the parenthesis depicts the number of pleas and the second number depicts the number of trials.

			Sentencing Data Set				
	Judge	Week	Monday	Tuesday	Wednesday	Thursday	Friday
1	7	2000-07-03					Dorchester (0,1)
2	7	2000-07-24	Orangeburg (1,0)	Orangeburg (1,0)		Orangeburg (1,0)	
3	7	2001-01-29	Dorchester (1,0)				
4	7	2001-06-04	Orangeburg (5,0)		Orangeburg (2,1)	Orangeburg (4,0)	
5	25	2000-10-30			Greenwood (1,0)		
6	25	2001-04-23	Greenwood (1,0)				
7	42	2001-03-12			Greenville (1,0)		
8	42	2001-04-16		Greenville (1,0)			
9	46	2000-10-23	Charleston (1,0)				

C.2 Single Assignment Case

C.2.1 Single assignment, no dates

We start with the single assignment, no date category. There are 2,214 judge-week combinations in this category. Although in single assignment, no date judge-week combinations (according to the master calendar) the judges have an assignment the entire week, they do not necessarily sentence every day of the week. For 1871 of the 2200 (85%) judge-week combinations, we only observe sentencing events in the county to which the judge is assigned, i.e., we do not observe sentencing events in a county other than the county to which the judge is assigned. In other words, there are no inconsistencies between the master calendar and the sentencing dataset for 1871 of these 2200 judge-week combinations. The remaining 328 judge-week combinations are listed in Table 15. We observe at least one sentencing event (in the sentencing dataset) in a county to which the judge is not assigned in these judge-week combinations.

Table 16: Judge-week combinations in which the judge has sentencing events in a county to which he is not assigned - single assignment, no date category. The counties written in green font are the counties to which the judge is assigned. The counties written in red font are the counties to which the judge is not assigned. The counties written in blue font are the counties to which the judge is not assigned, however, he is assigned to the circuit court containing these counties. So, the county assignment in the master calendar and this county belong to the same circuit court.

				Sentencing Data Set				
	Judge	Week	Master Calendar	Monday	Tuesday	Wednesday	Thursday	Friday
1	1	2000-08-07	York GS		York (2,0)	York (2,0)	York (4,0)	Kershaw (1,0), York (7,0)
2	1	2000-08-21	York GS		York (1,0)			Union (1,0), York (2,0)
3	1	2000-09-11	York GS	York (6,0)	York (2,0)		York (2,0)	Spartanburg (1,0), York (5,0)
4	1	2000-10-09	York GS			Fairfield (1,0), York (1,0)	York (5,0)	York (4,0)
5	1	2000-10-23	York GS	York (2,0)	York (3,0)	York (9,0)	York (1,0)	Kershaw (1,0), York (6,0)
6	1	2001-01-08	Chester GS	Chester (2,0)	York (1,0)			Chester (2,0)
7	1	2001-02-12	Chester CP				Lancaster (1,0)	
8	1	2001-06-04	Chester GS			Chester (0,1)	Chester (2,0), Spartanburg (1,0)	Chester (3,0)
9	2	2000-07-24	Richland GS	Richland (4,0)	Lexington (1,0), Richland (4,0)	Richland (2,0)	Richland (7,0)	
10	2	2000-09-11	Richland GS		Kershaw (1,0), Richland (6,0)	Richland (7,0)	Richland (3,0)	Richland (1,0)
.	.	.	.	.	.	.	.	.
320	49	2001-06-25	Richland GS	Lexington (1,0), Richland (1,0)				
321	50	2000-09-25	Richland GS	Richland (7,0)	Clarendon (1,0), Lexington (1,0)	Richland (2,1)	Richland (3,0)	Richland (1,0)
322	50	2000-10-02	in chambers					Richland (1,0)
323	50	2000-10-16	Orangeburg GS	Orangeburg (14,0)	Orangeburg (2,0)	Orangeburg (5,0)	Charleston (1,0), Orangeburg (3,0)	Orangeburg (1,0)
324	50	2000-12-18	in chambers	Richland (1,0)	Richland (24,0)			
325	50	2001-01-08	Orangeburg GS	Calhoun (2,0), Orangeburg (14,0)	Orangeburg (3,0)	Orangeburg (4,0)	Orangeburg (3,0)	Aiken (1,0), Orangeburg (2,0)
326	50	2001-01-22	Sumter GS	Sumter (6,0), York (1,0)	Sumter (2,0)	Sumter (3,0)	Sumter (1,0)	Sumter (3,0)
327	50	2001-02-05	Clarendon GS	Clarendon (3,0)	Clarendon (2,0)	Clarendon (1,0)		Clarendon (3,0), Sumter (1,0)
328	50	2001-04-16	Sumter GS	Clarendon (0,1), Sumter (5,0)	Sumter (3,0)	Sumter (2,0)	Sumter (3,0)	Sumter (3,0)
329	50	2001-05-28	3rd Cir. CPNJ/PCR				Sumter (1,0)	Williamsburg (1,0)

C.2.2 Single assignment with dates

There are 14 judge-week combinations in Category ii (b), i.e., a single assignment with specific dates. In 10 of the 14 judge-week combinations, we either observe no sentencing events or sentencing events only in the same county as listed in the master calendar (2 of them). In other words, there are no inconsistencies between the master calendar and the sentencing dataset for 10 of these 14 (71.4%) judge-week combinations. The remaining 4 judge-week combinations are listed in Table 15. In these 4 judge-week combinations, exactly one of the following conditions holds:

1. At least one sentencing event occurred in a county not stated on the master calendar (3 judge-week combinations); see the last three rows of Table 17.

2. At least one sentencing event occurred in the county stated on the master calendar but on a

date not stated on the master calendar (1 judge-week combination); see the first row of Table 17.

Table 17: Judge-week combinations in which the judge has sentencing events in a county to which he is not assigned - single assignment, with dates category. The counties written in green font are the counties to which the judge is assigned. The counties written in red font are the counties to which the judge is not assigned.

				Sentencing Data Set				
	Judge	Week	Master Calendar	Monday	Tuesday	Wednesday	Thursday	Friday
1	6	2001-03-19	Horry GS CP CC 21,22,23		Horry (6,0)		Horry (5,0)	
2	27	2000-09-25	Bamberg GS 25,26,27,28,29	Bamberg (10,0), Lee (1,0)	Bamberg (6,0)			
3	46	2000-10-30	Barnwell GS 1,2,3			Barnwell (1,0)	Allendale (1,0), Barnwell (2,0)	
4	48	2000-10-23	13th Cir. CPNJ 23,24,25,26					Clarendon (1,0)

C.3 Multiple Assignments Case

C.3.1 Multiple assignments, no dates

We start with the category multiple assignment, no dates. There are 19 judge-week combinations in this category. For 17 of the 19 (89.4%) judge-week combinations in this category, we only observe sentencing events in one of the two counties stated on the master calendar. In other words, there are no inconsistencies between the master calendar and the sentencing dataset for 17 of these 19 judge-week combinations. The remaining 2 judge-week combinations are listed in Table 18. The judge-week combinations in this category that satisfy (at least) one of the following conditions are potentially problematic:

1. At least one sentencing event occurred in a county not stated on the master calendar (Both rows of Table 18).
2. The judge sentenced in more than one county on a day (First row of Table 18).

Table 18: Judge-week combinations in which the judge has sentencing events in a county to which he is not assigned - multiple assignment, no dates cateogry. The county written in green font is the county to which the judge is assigned. The counties written in blue font are the counties to which the judge is not assigned, however, he is assigned to the circuit court containing this county. So, the county assignment in the master calendar and this county belong to the same circuit court.

				Sentencing Data Set				
	Judge	Week	Master Calendar	Monday	Tuesday	Wednesday	Thursday	Friday
1	2	2001-04-30	Aiken GS CC, 2nd Cir. CP/PCR			Aiken (5,0), Bamberg (1,0)		
2	9	2001-01-08	Florence GS, Lee GS			Marion (1,0)	Marion (1,0)	

C.3.2 Multiple assignments, some dates

There are 181 judge-week combinations in the multiple assignment, some dates category. For 155 of these 181 (85.6%) judge-week combinations, we only observe sentencing events in one of the two counties stated on the master calendar and only on the days stated on the master calendar. In other words, there are no inconsistencies between the master calendar and the sentencing dataset for 155 of these 181 judge-week combinations. The remaining 26 judge-week combinations are listed in Table 19. There is only one judge-week combination in the multiple assignment, some dates ambiguous category, and we do not observe any conflicts in that week.

Table 19: Judge-week combinations in which the judge has sentencing events in a county to which he is not assigned - multiple assignment, some dates category. The counties written in green font are the counties to which the judge is assigned. The counties written in red font are the counties to which the judge is not assigned. The counties written in blue font are the counties to which the judge is not assigned, however, he is assigned to the circuit court containing these counties. So, the county assignment in the master calendar and this county belong to the same circuit court.

				Sentencing Data Set				
	Judge	Week	Master Calendar	Monday	Tuesday	Wednesday	Thursday	Friday
1	6	2000-10-16	Florence CP, Florence GS CC 20				Horry (1,0)	Florence (8,0)
2	6	2000-12-18	in chambers, 12th Cir. CPNJ 18,19	Marion (1,0)				
3	9	2001-06-18	Florence GS, Florence CP 21	Florence (5,0)	Florence (3,0)	Darlington (1,0), Florence (17,0)	Florence (12,0)	Florence (5,0)
4	10	2000-10-02	in chambers, 7th Cir. CPNJ 2,3,5				Spartanburg (3,0)	
5	10	2001-06-04	8th Cir. CPNJ, Spartanburg GS CC 8	Laurens (1,0)				
6	12	2001-03-05	McCormick GS, Edgefield GS CC 8	McCormick (3,0)			Saluda (1,0)	
7	13	2000-08-28	Williamsburg GS, Orangeburg GS CC 1	Williamsburg (3,0)	Williamsburg (4,0)	Sumter (1,0), Williamsburg (2,0)	Williamsburg (6,0)	
8	13	2000-12-11	Lee GS, Aiken GS CC 15	Lee (1,0)	Lee (1,0)			Lee (0,2)
9	13	2001-01-08	Richland GS, Aiken GS CC 12	Richland (8,0)	Richland (3,0)	Richland (1,0)		Richland (0,1)
10	17	2001-03-12	Spartanburg CP, Cherokee GS CC 15			Pickens (1,0)	Cherokee (1,0)	
11	25	2001-02-26	Greenwood GS, 8th Cir. CPNJ CC 26	Greenwood (4,0)	Greenwood (10,0)	Greenwood (7,0)	Abbeville (1,0), Laurens (1,0)	Greenville (1,0), Greenwood (5,0)
12	26	2000-11-13	Greenwood GS, Medical 17	Anderson (1,0)		Greenwood (9,1)	Greenwood (5,0)	
13	28	2000-07-31	Richland GS, Lexington GS CC 3	Lexington (1,0), Richland (13,0)	Kershaw (1,0), Marlboro (1,0), Richland (10,0)	Richland (12,0)	Richland (4,0)	
14	28	2001-03-19	Darlington GS, Dillon GS CC 23	Darlington (1,0)	Darlington (8,0)	Chesterfield (1,0), Darlington (9,0)	Darlington (3,0)	
15	29	2001-06-04	16th Cir. CPNJ, Clarendon GS CC 8		Union (1,0)			Clarendon (1,0)
16	30	2000-10-02	in chambers, Greenville GS 6			Anderson (1,0)		
17	32	2001-04-09	Charleston CP, Horry GS CC 9	Charleston (1,0)				
18	32	2001-04-30	Berkeley GS, Richland GS CC 2	Berkeley (6,0)	Berkeley (1,0)	Berkeley (2,0)	Berkeley (6,0)	
19	34	2001-04-16	Florence GS, 5th Cir. CPNJ CC 20	Florence (12,0)	Florence (8,0)	Florence (16,0)	Florence (5,0)	Florence (5,0)
20	34	2001-05-14	Marion GS, 5th Cir. CPNJ CC 14	Marion (7,0)		Marion (1,0)		Marion (8,0)
21	37	2001-01-15	Georgetown CP, Georgetown GS CC 19				Marion (0,2)	
22	39	2000-08-28	Anderson CP, Oconee GS CC 1					Anderson (1,0)
23	39	2000-11-06	Anderson GS, 10th Cir. CPNJ CC 8	Anderson (3,0), Greenville (1,0)		Anderson (8,0)	Anderson (10,0)	
24	40	2001-05-07	in chambers, Greenville GS(SGJ) 7, 13th Cir. CPNJ 8	Spartanburg (1,0)				
25	43	2001-05-07	in chambers, 9th Cir. CPNJ 8		Colleton (0,1)			
26	47	2000-12-04	15th Cir. CPNJ, Horry GS CP CC 5	Horry (1,0)				

The judge-week combinations listed in Table 19 satisfy (at least) one of the following conditions, which are problematic (some satisfy multiple conditions):[7]

1. At least one sentencing event occurred in a county not stated on the master calendar, e.g., judge-week combination 2 of Table 19 (17 such judge-week combinations are listed in Table 19).

2. At least one sentencing event occurred in the county stated on the master calendar with specific dates, but on a date other than the dates specified, e.g., judge-week combination 23 of Table 19 (4 such judge-week combinations are listed in Table 19.).

3. At least one sentencing event occurred in the county stated on the master calendar without specific dates. However, this sentencing event occurred on a date specified for the other county, e.g., judge-week combination 8 of Table 19 (8 such judge-week combinations are listed in Table 19.).

4. The judge sentenced in more than one county on a day, e.g., judge-week combination 3 of Table 19 (5 such judge-week combinations are listed in Table 19.).

C.3.3 Multiple assignments, all dates

There are 57 judge-week combinations in the multiple assignment, all dates category. For 47 of these 57 (82.4%) judge-week combinations, we only observe sentencing events in one of the two counties stated on the master calendar and only on the days stated on the master calendar. In other words, there are no inconsistencies between the master calendar and the sentencing dataset for 47 of these 57 judge-week combinations. The remaining 10 judge-week combinations are listed in Table 20.

[7] For example, judge-week combination 12 satisfies all four conditions. In judge-week combination 12, the assignment of Judge 29 in the week of July 31 is Richland GS and Lexington GS CC 3. Therefore, we would expect Judge 29 to be in Lexington on Thursday, and to be in Richland on Monday, Tuesday, Wednesday, and Friday. Since we observe sentencing events in Marlboro and Kershaw (alongside Richland) on Tuesday, judge-week combination 12 satisfies Conditions 1 and 4. Since we observe sentencing events in both Lexington and Richland on Monday, judge-week combination 12 satisfies Condition 3. Since we observe sentencing events in Richland on Thursday, judge-week combination 12 satisfies Condition 2, as well.

Table 20: Judge-week combinations in which the judge has sentencing events in a county to which he is not assigned - multiple assignment, all dates category. The counties written in green font are the counties to which the judge is assigned. The counties written in red font are the counties to which the judge is not assigned. The counties written in blue font are the counties to which the judge is not assigned, however, he is assigned to the circuit court containing these counties. So, the county assignment in the master calendar and this county belong to the same circuit court.

				Sentencing Data Set				
	Judge	Week	Master Calendar	Monday	Tuesday	Wednesday	Thursday	Friday
1	5	2000-10-16	15th Cir. AW 16,19,20, Aiken 17,18		Aiken (11,0)	Aiken (11,0)	Horry (6,0), Williamsburg (1,0)	Horry (1,0)
2	9	2000-07-10	Darlington GS 11,12,13,14, XX 10		Darlington (2,0)	Darlington (5,0)	Chesterfield (1,0), Darlington (9,0)	
3	16	2000-11-27	XX 1, Aiken GS 27,28,29,30	Aiken (3,0)		Aiken (1,1)	Aiken (7,0), Spartanburg (1,0)	
4	17	2001-04-16	13th Cir. CPNJ 16, Lexington GS 17,18,19,20			Lexington (12,0)	Lexington (11,0), Richland (1,0)	
5	26	2001-05-28	10th Cir. CPNJ/PCR 28,29,30, Lexington GS 31,1			Anderson (1,0)	Lexington (15,0)	Lexington (1,0)
6	30	2001-02-26	Anderson CP 26,27,28,1, Cherokee GS 2, 10th Cir. CPNJ/GS 2					Oconee (6,0)
7	34	2001-01-08	12th Cir. CPNJ 8,9,11,12, X 10	Richland (1,0)				
8	39	2000-10-30	Anderson CP 1,2,3, Barnwell GS 30,31	Barnwell (3,0)	Aiken (1,0), Barnwell (13,0), Newberry (1,0)			
9	43	2001-05-14	Colleton GS 14,15,16, X 17,18	Colleton (4,0)	Colleton (2,0)	Colleton (3,0)		Beaufort (1,0)
10	49	2000-07-17	Lexington GS 17,18,19,20, X 21			Lexington (14,0)	Lexington (10,0)	Lexington (2,0)

The judge-week combinations listed in Table 20 satisfy (at least) one of the following conditions, which are problematic (some satisfy multiple conditions):[8]

1. At least one sentencing event occurred in a county not stated on the master calendar, e.g., judge-week combination 1 of Table 20 (8 such judge-week combinations are listed in Table 20).

2. At least one sentencing event occurred in the county stated on the master calendar with specific dates, but on a date other than the dates specified (only one such judge-week combination is listed in Table 20; see the last row of the table).

3. The judge sentenced in more than one county on a day, e.g., judge-week combination 1 of Table 20 (5 such judge-week combinations are listed in Table 20).

[8]For example, judge-week combination 1 satisfies Conditions 1 and 3. In judge-week combination 1, the assignment of Judge 6 in the week of October 16 is 15th circuit court AW 16,19,20 and Aiken 17,18. Therefore, we would expect Judge 6 to be in the 15th circuit court on Monday, Thursday, and Friday, and to be in Aiken on Tuesday and Wednesday. Since we observe sentencing events in Horry and Williamsburg on Thursday, judge-week combination 1 satisfies both Conditions 1 and 3.

D Merging the Sentencing Data and the Calendar Data

D.1 Alternative mapping methods

Recall that the master calendar lists the judge names and the counties they are assigned to each week. Also, recall that the judge names are not recorded in the sentencing dataset. Instead, the judges are numbered. This section maps the judge numbers in the sentencing dataset to the judge names in the master calendar. To do so, for each judge, we create a sequence of weekly county assignments. We do so using both the sentencing dataset (using the sentencing events and their dates) and the master calendar, resulting in two sequences of weekly county assignments for each judge. We compare these two sequences to find the mapping from the judge numbers to the judge names under the following assumptions. The justification for this assumption is discussed below after presenting the matching algorithm and the resulting mapping.

Assumption 1. We restrict attention to the county name in the assignments listed on the master calendar when constructing the mapping from the judge numbers in the sentencing dataset to the judge names in the master calendar.

For each judge, in order to measure how close the aforementioned two sequences of weekly county assignments are, we start off by comparing these sequences week by week, i.e., componentwise. To do so, fix a judge and a week, and consider the list of counties for that week in each sequence. In particular, we consider the following two notions of match:

Perfect Componentwise Match. If the set of counties visited by the judge number (obtained from the sentencing dataset) is a subset of the set of counties assigned to the judge name (obtained from the master calendar), the two sets constitute a perfect componentwise match. Otherwise, they do not constitute a perfect componentwise match. They constitute a mismatch (under the perfect componentwise matching criterion). For example, let us consider Judge 25 and Judge Hayes in the week of July 10th. In the week of July 10th, Judge 25 had sentencing events in {York} and Judge Hayes was assigned to {York}. These two sets (of counties) constitute a perfect componentwise match.

Overlapping Componentwise Match. If the intersection of the two sets is non-empty, they constitute an overlapping componentwise match. Otherwise, they constitute a mismatch (under the overlapping componentwise matching criterion). For example, let us consider Judge 25 and Judge Hayes in the week of July 17th. In the week of July 17th, Judge 25 had sentencing events in {Union,Spartanburg} and Judge Hayes was assigned to {Union}. These two sets (of counties) constitute an overlapping componentwise match. Note that this would be a mismatch under the perfect componentwise matching criterion.

Then, we use the following algorithm to map the judge numbers (in the sentencing dataset) to the judge names (in the master calendar). The algorithm seeks to map each judge number to the judge name for which the number of matching weeks is maximal.

Algorithm. Choose the matching criterion.

Step 0. Fix a judge number. Find the set of judge names that have zero weeks of mismatch with the judge number under consideration. Then, map each judge number to the judge name with which it has zero weeks of mismatch. Repeat this process for all judge numbers. Next, drop the mapped judge numbers and judge names from the consideration list and proceed to the next step.

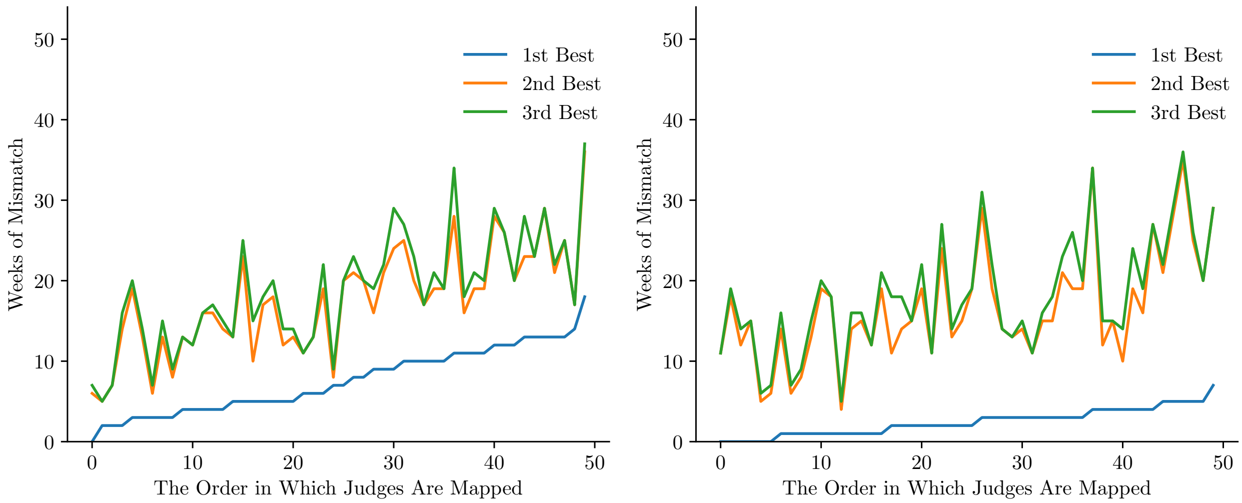
Step 1. Fix an unmapped judge number. Find the set of unmapped judge names that have 1 week of mismatch with the judge number under consideration. Then, map each unmapped judge number to the judge name with which it has 1 week of mismatch. Repeat this process for all unmapped judge numbers. Next, drop the mapped judge numbers and judge names from the consideration list. If any of the (non-fictitious) judge numbers are not yet mapped, proceed to the next step.

Step $n > 1$. Fix an unmapped judge number. Find the set of unmapped judge names that have n weeks of mismatch with the judge number under consideration. Then, map each unmapped judge number to the judge name with which it has n weeks of mismatch. Repeat this process for all unmapped judge numbers. Next, drop the mapped judge numbers and judge names from the consideration list. If any judge numbers are not yet mapped, proceed to the next step. Continue this process until all judge numbers are mapped to a judge name.

We obtain the same alphabetical mapping under both matching criterion; see Table 21. Figure 5 (in Appendix **??**) depicts the number of weeks of mismatch between the judge numbers and the judge names. In particular, it shows that for all judge numbers, the mapped judge name has considerably fewer weeks of mismatch than any other judge name. In particular, there were no ties.

Table 21: The mapping from the judge numbers to the judge names.

Judge Number	Judge Name	Judge Number	Judge Name
2	ALFORD	27	JOHNSON
3	BARBER	28	KEESLEY
4	BAXLEY	29	KINARD
5	BEATTY	30	KING
6	BREEDEN	31	KITTREDGE
7	BROGDON	32	LEE
8	BROWN	33	LOCKEMY
9	BUCKNER	34	MACAULAY
10	BURCH	35	MANNING
11	CLARY	36	MARTIN
12	COLE	37	MCKELLAR
13	COOPER, GT	38	MILLING
14	COOPER, TW	39	NEWMAN
15	COTTINGHAM	40	NICHOLSON
16	DENNIS	41	PATTERSON
17	FEW	42	PIEPER
18	FLOYD, H.	43	PYLE
19	FLOYD, S.	44	RAWL
20	GOODE	45	SAUNDERS
21	GOODSTEIN	46	SHORT
22	GREGORY	47	SMOAK
23	HALL	48	THOMAS
24	HARWELL	49	WATSON
25	HAYES	50	WESTBROOK
26	HUGHSTON	51	WILLIAMS



(a) Perfect Match. (b) Overlapping Match.

Figure 5: Weeks of mismatch for the judge name mapped to each judge number. The judge names with the second and third fewest weeks of mismatch are also depicted.

D.2 Evaluating the mapping

Next, we assess the performance of the (derived) mapping (from the judge numbers to the judge names). Under this mapping, we observe that only in 4.1% of the sentencing events with a date, the judge was sentencing in a county to which he was not assigned. These 4.1% problematic sentencing events belong to one of two categories: About 0.9% belong to judge-week combinations in which the judge did not have any county assignments on the master calendar. The remaining 3.2% belong to judge-week combinations in which the judge had county assignments on the master calendar. However, he was not assigned to the county in which the sentencing event occurred. Figure **??** depicts the number of problematic sentencing events for each judge number. Figure **??** (in Appendix **??**) depicts the number of problematic sentencing events for each week. Figures **??**-**??** show that the problematic sentencing events are not concentrated in just some judges or some weeks. These problematic sentencing events are discussed in detail in Section C.

D.3 Limitations

Assumption 1 ignores the "circuit" assignments on the master calendar. Each circuit contains multiple counties. Instead, our matching algorithm works with the most granular geographic unit of interest, the county. As mentioned above, 3.2% of the sentencing events with dates are at odds with the master calendar under the alphabetical mapping depicted in Table 21. Including "circuit" assignments and matching them on the basis of circuit courts as opposed to counties addresses only 12% of the problematic sentencing events (corresponding to 0.4% of the total number of sentencing events). As such, we stick with the choice of county for the matching purposes because it is more granular, and likely more accurate. Other assignments we ignore should be harmless because judges would not be involved in sentencing events or trials in those cases; see Figure **??**.

In addition, we focus only on the county names as opposed to the county name and the type of the assignment, which may also be listed; see Figure **??**. For example, Georgetown GS and Georgetown CP are both taken as Georgetown. There are several reasons for this: First, the type of the assignment is not always available. More importantly, all sentencing events should happen only in the general sessions (GS); and the common plea (CP) designation could be an error if a sentencing event appears in the sentencing dataset for the corresponding judge-week-county triple. Moreover, if no sentencing event occurs, then it is harmless to omit the CP designation. Consequently, focusing on the county name only makes the match easier.

Table **??** (in Appendix **??**) depicts the list of counties visited by the fictitious judge (Juge 1) alongside the list of counties to which the unmapped judge names were assigned. Table **??** shows that judge 1 is not a combination of the unmapped judge names.