

# The Effect of Bail on Guilty Pleas

*MS&E 330*

Ali Alkhatib, Nikhil Garg, Martin Lindsey, Danaë Metaxa–Kakavouli

{al2, nkgarg, mlindsey, metaxa}@stanford.edu

March 17, 2016

## Abstract

Research on the subject of bail, the burden it places on defendants, and the persuasive and perhaps coercive influence it bears on the justice system has been explored to some extent ethnographically, but thus far researchers have not had the data necessary to conduct rigorous quantitative studies on bail-setting and the factors which influence pleading behavior. With a unique insight into the demographics and characteristics of defendants and their outcomes, we explore two questions: 1) Does a greater burden, as measured by bail, cause people to plead guilty? And 2) Are judges handling cases uniquely, or uniformly?

Our findings reveal that heightened burdens do seem to cause defendants to plead guilty at a higher rate, and that there is some evidence that variance in bail-setting is unique to judges.

## 1 Introduction

Maintaining an efficient justice system is in the interest of both the professionals managing and the citizens interacting with it. Computational methods including large-scale data analysis and machine learning excel at quantitatively identifying points of inefficiency. We propose to use these techniques to build models predicting points of inefficiency in the pretrial processes, in order to provide recommendations for improving efficiency to judges and other court officials.

Notably, a large majority of cases are never sent to trial; rather, charges are dropped or settled out of court. Pretrial procedures, then, such as the setting of bail or decision to pursue a case, are especially fruitful sites for consideration with regards to efficiency. A more insightful understanding of how to approach these pretrial procedures will benefit court officials in the time and money spent processing cases needlessly. For defendants, these insights could lead them to spend less time in prison awaiting trial if they are unable to pay bail, and fewer resources towards a case that will never be brought to trial.

With this potential for intervention in mind, we propose to examine the Gotham dataset, using case features available to the judge, and in particular bail, which is set by the judge, to predict whether a defendant will plead or not later in court proceedings. The goal of this procedure is to analyze one attribute of a defendant's process through the justice system from the perspective of a judge, including information the judge knows about the case and, most importantly, the decision a judge can make: setting bail.

Specifically, we conceive of bail setting as imposing a type of (intentional) burden upon a defendant. This burden can occur at several different levels, from release on a defendant's own recognizance (ROR) — a nonexistent burden, to refusal to set bail at all in order to keep a defendant in custody — a very high burden. In between these two levels are bails set which the defendant is able to pay (or find a bondsman to put up bond), and bails which the defendant is not able to pay. These two categories are irrespective of the bail amount, and instead are specific to the means of the defendant who is either able or not to pay bail. Similarly, it is worth considering whether different judges set bail in substantially different ways given similar cases, as a starting point for different analyses.

We make 2 hypotheses:

1. Effect of burden on pleading: that placing a larger bail burden on a defendant will correlate with a higher likelihood to plead later in the case. We posit a causal psychological mechanism by which defendants who have been required to put up substantial financial assets or, even further, had to spend time in jail awaiting trial are placed in a vulnerable position by the justice system and therefore are less likely to prolong their case defend themselves forcefully against charges — in other words, more likely to plead guilty.
2. Our secondary hypothesis predicts that different judges will set substantially different bail on similar cases, since we anticipate that there is significant room for human error or difference this aspect of the justice system.

The underlying goal of this work is to provide predictions which aid judges in setting effective but not exces-

sive bail by better understanding one way in which the process of setting bail influences defendant experiences with and attitudes toward the justice system.

## 2 Literature Review

Broadly, we identify 3 areas of research which provide some insight into this topic: 1) The use of *bail*, and its stated and empirical purposes, 2) the study of the effect that bail has on *burden*, and 3) the role of *plea bargaining*, both for the courts and defendants. We will discuss the intersection of these bodies of work as they relate to our research question.

### 2.1 Bail

In this section, we will explore the murky waters of the use of bail, but ultimately settle on the conclusion that the courts have made: that bail is fundamentally a necessity in a judicial system, despite the philosophical questions and issues it inherently raises. Despite the need for bail as a robust safeguard against undue imprisonment, we will illustrate that the practice of bail-setting is — perhaps decidedly — difficult to discuss in broad terms, often leaving researchers primarily with qualitative, ethnographic, and sometimes anecdotal evidence upon which to work.

Bail represents an attempt to resolve a challenging cognitive dissonance. As Foote (1959) points out, the imprisonment of a defendant who is nominally considered “innocent until found guilty” arouses challenges of principle; if the courts are to assume that a defendant is innocent (and thus treat them as such), then holding them in custody until the trial begins is patently wrong, as it deprives a *presumably* innocent person of freedom.

Nevertheless, the United States Constitution’s 8th amendment — as well as a series of laws attempting to reform bail (one in 1966, another in 1984) — ostensibly protect defendants from unreasonable and excessive bail amounts, otherwise to be set at the discretion of the presiding judge (Berg, 1985). As Berg notes, however, the most recent revisions to the Bail Reform Act — those passed in 1984 — make it explicit that judges may set or even deny bail on the basis of the judge’s assessment of the defendant’s danger to the community. In *Schall v. Martin* (1984), the court ruled that “there was nothing inherently unconstitutional in preventive detention” and that “the restraint on [the defendant]’s liberty did not amount to punishment because that was not the express purpose of the pretrial detention”.

The precedents established in the cases of *Schall v. Martin* (1984) and *United States v. Salerno* (1987) broadly supports the judge’s prerogative to set bail — and importantly, to *not* set it (in other words, to remand the defendant with no option to pay bail), if the judge determines that safety to the community (or the defendant’s appearance upon a later court date) could not be assured by any bail amount. This latitude becomes an important component that we attempt to explore in the hopes of determining whether if — at least within a single jurisdiction — judge bail-setting variance is high or not. The implication of this question may be that some judges are more merciful, and some decidedly harsher, than their peers in their assessments of defendants.

### 2.2 Burden

The 1984 Bail Reform Act thus afforded the courts a great deal of freedom in setting bail, bringing us naturally to the question of what effect this has on defendants awaiting trial. In this area, we find myriad researchers have contributed; we will attempt to highlight some of the work that is especially illuminating and representative of the body of work.

Kellough and Wortley identify qualitatively that the “moral assessments” of defendants appear to influence the burden judges assign in the form of bail-setting. This, they hypothesize, may explain (at least to some

extent) the disparity in race between those that are let free in the interim before their court date, and those that are held in custody on remand (Kellough and Wortley, 2002).

Zacharias further discusses the burden that bail may impose on defendants, in this case highlighting the leverage that prosecutors may have over defendants, for instance in the form of the unfamiliar and unwelcome environment that temporary incarceration represents, and the promise to expedite one's interaction with the justice system (Zacharias, 1997). This, finally, leads us to the outcome variable which we will ultimately investigate.

## **2.3 Plea Bargaining**

Most tangibly, the persuasive (perhaps *coercive*) force that prosecutors can exercise in tandem with judges is the pressure to encourage defendants to plead guilty in the pretrial phase. This aspect of the justice system is especially unique, both because plea-bargaining defies the adversarial model of the court and trial process and because these dynamics are so ill-understood at broad levels (Zacharias, 1997).

The research on the burden that bail represents on defendants — especially defendants with limited financial means and similarly limited access to cash — has *qualitatively* found that trial outcomes may be affected by the burden of pretrial incarceration. To illustrate this, we turn to Grossman and Katz (1983) and Alschuler (1981); Grossman and Katz estimates that as many as “90 percent of convictions ...result from a negotiated plea of guilty” (see also McCoy and Mirra, 1980; Kaplan, 1977; Newman and Remington, 1966, with similar findings).

Most of the existing research in plea bargaining, especially Grossman and Katz's work, focuses on the economic role of the plea bargain as a “screening device”; Alschuler contributes to the argument that pretrial bargaining may serve the defendant's best interests, observing that neither the courts nor the public defenders who often represent defendants have sufficient time to see all cases through open court (Alschuler, 1981). Alschuler makes the noteworthy observation that we lack sufficient information to determine whether the pressure of incarceration, the prosecutor's persuasive offer to settle, and the defending attorney's desire to expedite the closure of outstanding cases is in fact causing defendants to settle more frequently, primarily due to a dearth of broad-based data, and certainly — as Kellough and Wortley describe it — a simple “paucity of research [in plea bargaining]” (Kellough and Wortley, 2002).

## **3 Method**

In order to analyze our two research questions, we first quantify bail burden and preprocess the data, and subsequently perform statistical analyses on the data. The methods used in these steps are detailed below.

### **3.1 Bail Burden**

We conceive of the possible burden placed on a defendant by bail as fitting into one of 4 categories: 1) When a defendant is Released on his or her Own Recognizance (ROR), the burden is negligible. Defendants are required neither to put up money nor to spend time incarcerated; 2) when bail is set and a defendant is able to pay it immediately, the defendant is lightly burdened; 3) when bail is set and a defendant is able to pay it but with a delay, the defendant is significantly burdened both financially and by needing to spend time incarcerated prior to meeting bail; 4) when a defendant either has no option to be released on bail, or is unable to meet bail, he or she is *severely* burdened, having to spend the entire time until trial incarcerated.

### **3.2 Preprocessing**

Several preprocessing steps were taken to ensure the validity of our analyses. Due to some inconsistencies in volume of cases prior to 2013, only cases from after (and including) 2013 were analyzed. Additionally,

cases involving violent offenses were removed since there are extra considerations and complexities in bail-setting for violent offenders, who may be considered safety risks to the community and evaluated differently for bail. We also only consider cases that made it past arraignment, since cases where charges were dropped at arraignment are ineligible for the analyses we perform. Finally, for most analyses we consider Quality of Life (QoL) and non-Quality of Life (non-QoL) offenses separately. Quality of life offenses make up a category of illegal activity which is connected with the broken windows theory of policing — these are misdemeanor offenses such as disorderly conduct, vagrancy, or loitering. QoL and non-QoL offenses may differ substantially in the populations being charged and in their handling in the justice system, so we perform the same analyses in parallel across these two types of data.

### **3.3 Analyzing the Effect of Bail on the Probability to Plea**

We analyze the effect of bail burden on the probability of pleading through two main techniques and using both random forest models and GLM. First, we train models to predict the probability of pleading using all case characteristics through arraignment, including the burden level set. Then we create test sets composed of different burden levels (through two different techniques, described next), and analyze the test set probability of pleading conditional on each burden level. We present the results both as a CDF of the test sets' probability of pleading and as means of the distribution.

To verify the effectiveness of our models, we produce calibration plots and include them in the appendix. For each row in the original test set, we predict the probability of pleading and ROR, respectively, for our two models. We bin these probabilities and observe the true probability of observing the dependent variable for each of the bins.

#### **3.3.1 “All Else Equal”**

For our preliminary analysis, we create the test sets for each burden level through an “all else equal” assumption — that we can change the burden level of a given case while holding all other case characteristics the same. We duplicate the test set into four sets. In each data set, we set the burden level to “ROR”, “Trivial”, “Trivial with Delay”, and “Jail”, respectively. We then send these test sets to the model and view the resulting predicted probability of pleading. We note that the “all else equal” assumption is not true in general. In a murder case, for example, it would not make sense to change the burden level from “Jail” to “ROR”. However, because we analyze only non-violent cases, the assumption holds approximately.

#### **3.3.2 Propensity Score Matching**

To further support our analysis, we relax the assumption through initial propensity score matching. For this method, we split the burden into only two levels: ROR and Non-ROR. We train a model to predict the probability of ROR. Two test sets are then matched on the probability of ROR through `matchit` (Ho et al., 2011), and the burden bits are set to ROR and Non-ROR, respectively. They are then sent to the probability of pleading models. To do true propensity score matching and make sure our assumption holds, we should have filtered out cases with either very low or very high probability of ROR. Otherwise, we are still using cases where it would not make sense to change the ROR value.

### **3.4 Analyzing the Effect of Judge on Bail**

#### **3.4.1 Analysis — Random Forest**

For an initial analysis to observe judge effects, we predict the Probability of ROR being granted for each case, without using JudgeID in our prediction model. This analysis uses the same Random Forest ROR model trained in the Propensity Matching analysis. This model gives a baseline for judge independent granting of

ROR. For each judge with over 100 cases in the test set, we determine the judge’s mean predicted probability of ROR using the model. This value is the percentage of a given judge’s cases that the “average” judge would ROR. We compare this value to the true percentage of cases in which each judge actually granted ROR.

### **3.4.2 Analysis — Generalized Linear Model**

An “all else equal” approach was taken in the GLM analysis — the general strategy was to create a copy of the data with a test variable set to a given value, input these modified datasets into the model for each value the test variable can assume, and then obtain summary statistics for each possible value to then compare with the summary statistics associated with the other values; by doing this, any predictive power this model might have can be used to check for the relative influences of certain factors on the outcome that this model predicts.

The ROR model was used to check if certain arraignment judges differed substantially in how they treated ROR. The plea model was used to check if certain levels of burden had a higher predicted rate reaching a plea disposition. As a cursory additional check, matching edge cases were identified using a software package (i.e. cases that can be considered similar on model inputs but differed on observed outcomes), and in each case, the same analysis was performed on a reduced dataset of these edge cases to similar results.

Within this reduced set of edge cases, the assumption of “all else equal” has greater validity, as the matching software is trusted to compute a quantitative measure of this “equality” between cases and identify precisely those which are “sufficiently equal”. The effects of judge on ROR are summarized in figures 4b and 4a. For non-QoL cases, judges seem to have more similar ROR rates compared to QoL cases (one of the most prolific judges is also an outlier in the QoL dataset). The effects of burden on the rate of plea dispositions are summarized in figures 2b and 2a.

For both QoL and non-QoL cases, the lower levels of burden had close distributions, but the distribution for “nontrivial”, the highest level of burden had its mass shifted noticeably toward higher plea rates. Additionally, the predicted plea distributions on the QoL dataset had mass shifted more to the right compared to the non-QoL distributions. Analogous plots for the same analysis performed with the ROR model on the matched non-QoL dataset are found in figures 6a and 6b; the other combinations of QoL or non-QoL and ROR or plea model yielded matched datasets that were too small (i.e. on the order of tens of cases) to use.

## **4 Results**

### **4.1 Analyzing the Effect of Bail on the Probability to Plea**

#### **4.1.1 All Else Equal — Random Forest**

We first carry out the All Else Equal analysis through a Random Forest model. Appendix A shows the calibration information for the Random Forest models for both Non-Quality of Life and Quality of Life cases. We observe that the models are accurate and serviceable, though the Non-QoL model consistently over-predicts the probability of pleading in the unaltered test set.

Tables 1 and 2, respectively, include the mean predicted probability of pleading at different burden levels for Non-Quality of Life and Quality of Life cases, respectively. Figures 1a and 1b, show the CDFs of same values determined through the model. Applying the GLM to the dataset reveals roughly similar predictions, as illustrated by Figures 2a and 2b.

We observe that, especially for Non-Quality of Life cases, jailing someone substantively increases the predicted probability of pleading when compared other burden levels. This effect can be observed through both the higher means and the right-shift on the CDF plots. The effect is similar, though much less pronounced, for Quality of Life cases.

Burden Level	Mean Predicted Probability of Pleading
ROR	0.3631537
Trivial	0.3834242
Delay	0.3744825
Jail	698843

Table 1: Mean Probability of Pleading for Non–Quality of Life offenses through All Else Equal Analysis using Random Forest model

Burden Level	Mean Predicted Probability of Pleading
ROR	0.6072936
Trivial	0.6102411
Delay	0.5999771
Jail	0.648984

Table 2: Mean Probability of Pleading for Quality of Life offenses through All Else Equal Analysis using Random Forest model

#### 4.1.2 Propensity Score Matching — Random Forest

Finally, we carry out the Propensity Score analysis through a Random Forest model. Appendix A also includes the calibration information for the Random Forest models to predict ROR for each type of case. We observe that the ROR prediction models are accurate and unbiased.

Tables 3 and 4, respectively, include the mean predicted probability of pleading for ROR and Non–ROR, for Non–Quality of Life and Quality of Life cases, respectively. Figures 3a and 3b show the CDFs of same values determined through the model (also see findings through GLM approach in Figures 4a and 4b).

Our results match the All Else Equal analysis. Especially for Non–Quality of Life cases, a Non–ROR burden level increases the predicted probability of pleading. Future work should repeat this analysis using Jail and Non–Jail burden levels, as those levels are where the substantive difference occurs in the All Else Equal analysis.

#### 4.2 Analyzing the Effect of Judge on Bail

Figures 5a and 5b show the mean predicted probability of ROR for each judge versus the judge’s actual percentage of ROR, for Non–Quality of Life and Quality of Life cases, respectively. The plots show that some of the variation in the judges’ percentage of cases that are granted ROR is due to non–randomness in the cases that they adjudicate. However, we observe that for a given mean predicted probability of ROR (on the Y axis), there is significant judge variance on the true percentage of cases that are granted ROR. Thus, this initial analysis shows that different judges do utilize ROR differently for the same types of cases. As before, figures 6a and 6b illustrate similar patterns using a different modeling approach (GLM).

Burden Level	Mean Predicted Probability of Pleading
ROR	0.3706164
Non–ROR	0.412566

Table 3: Mean Probability of Pleading for Non–Quality of Life offenses through Random Forest Modeling

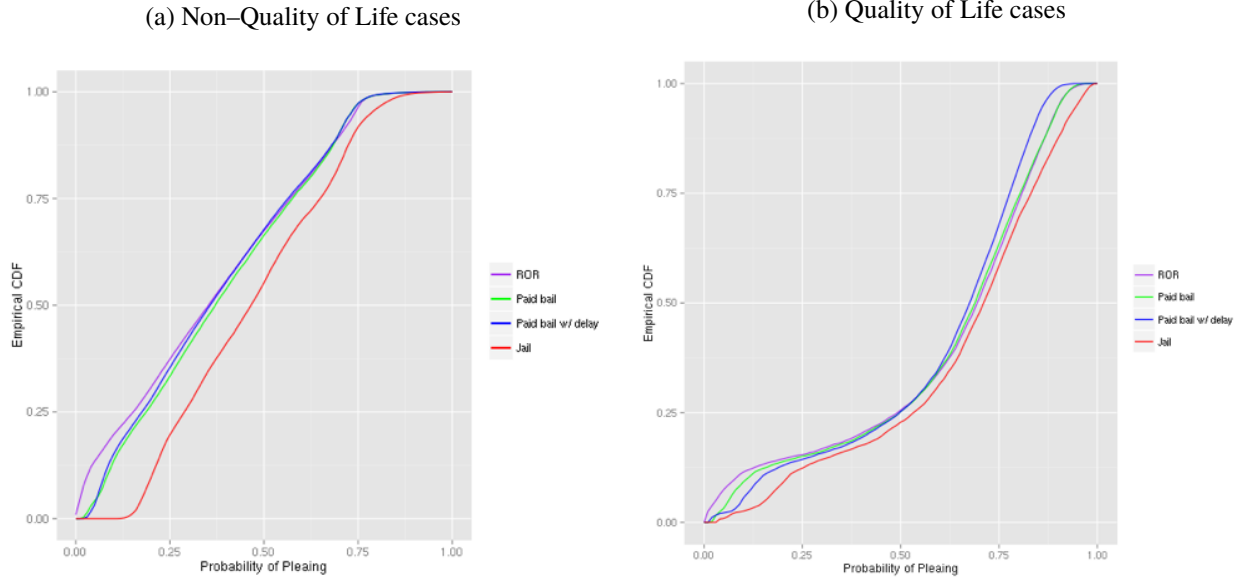


Figure 1: CDFs of predicted probability of pleading guilty at different burden levels determined through the All Else Equal Random Forest model

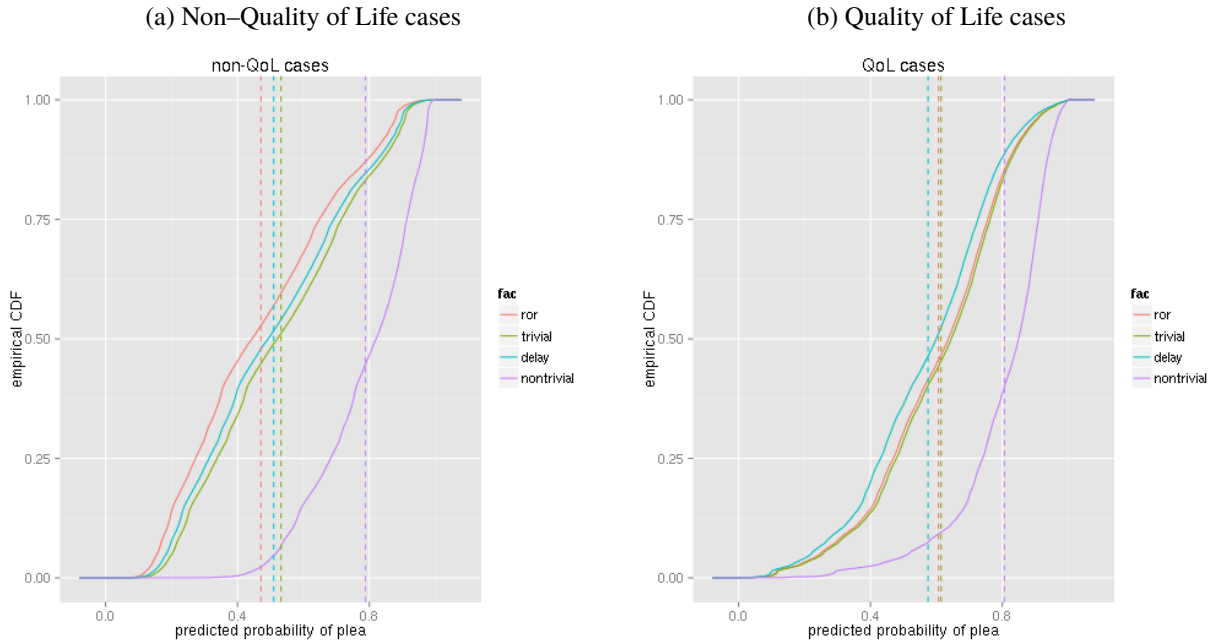


Figure 2: CDFs of predicted probability of pleading guilty at different burden levels determined through the Generalized Linear Model

## 5 Discussion

As described in the results above, we find that our first hypothesis, that a more burdensome bail will increase likelihood for a defendant to plead out, is confirmed. Our secondary hypothesis is also confirmed, though



Burden Level	Mean Predicted Probability of Pleading
ROR	0.6127946
Non-ROR	0.6363191

Table 4: Mean Probability of Pleading for Quality of Life offenses through Random Forest Modeling

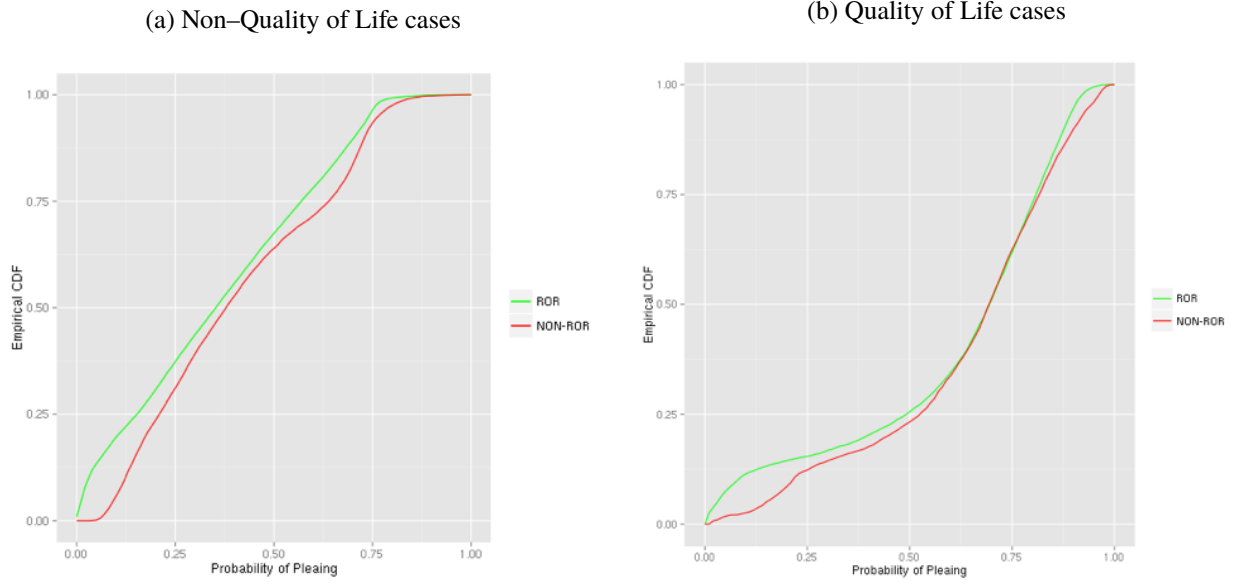


Figure 3: CDFs for the mean predicted probability of pleading for ROR and non-ROR through Random Forest Modeling

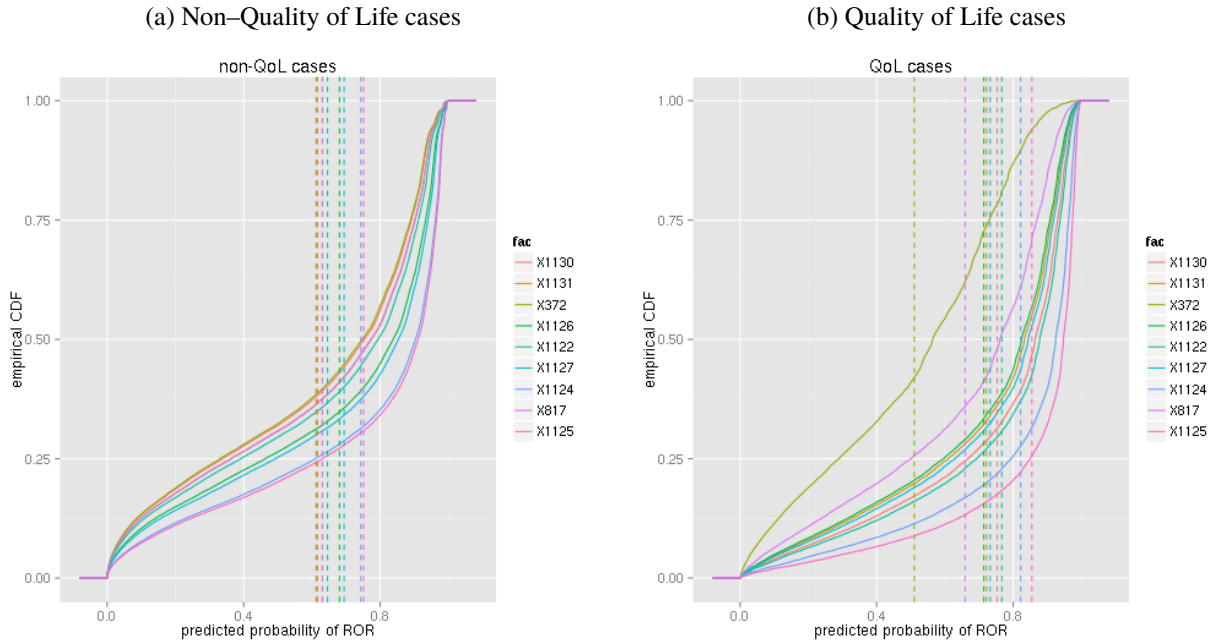


Figure 4: CDFs for the mean predicted probability of pleading for ROR and non-ROR through Generalized Linear Model

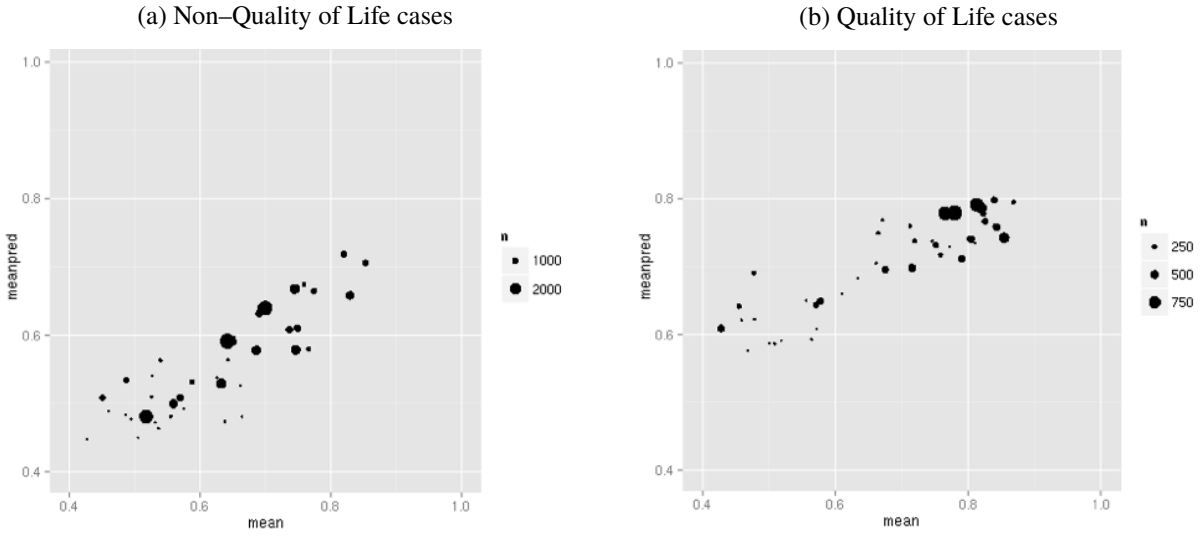


Figure 5: mean probability of ROR for each judge, vs. actual percentage of ROR modeled by Random Forest

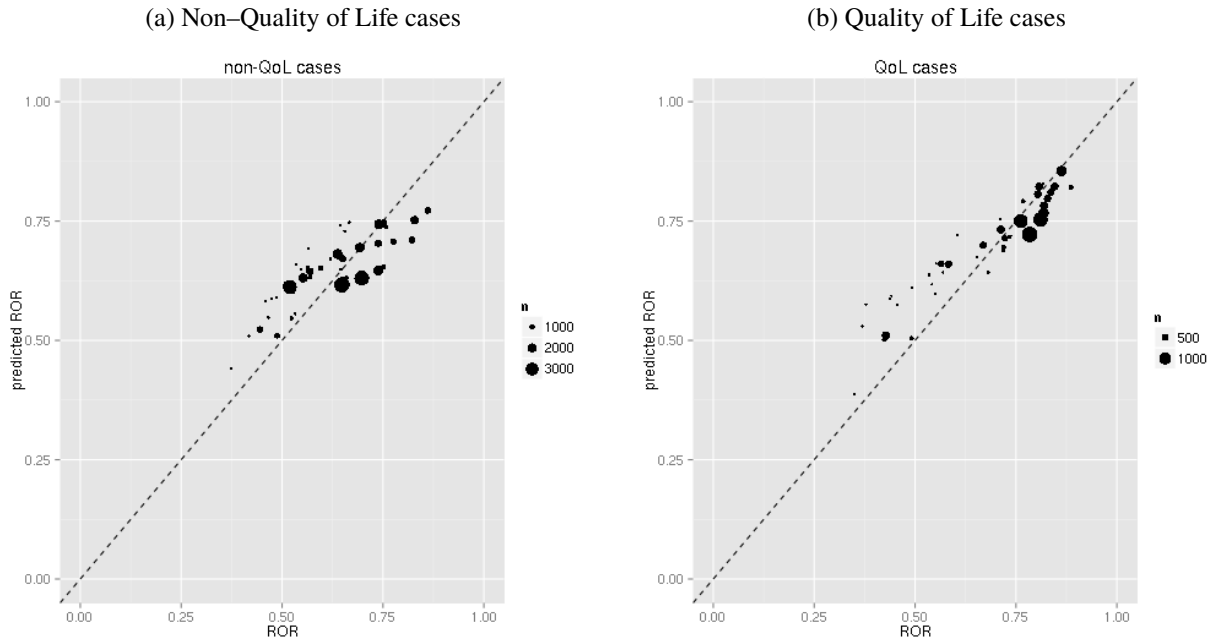


Figure 6: mean probability of ROR for each judge, vs. actual percentage of ROR modeled by GLM

less strongly.

More specifically, we observe that in non-QoL cases, there is a substantive distinction between the first two burden levels, which do not involve jail time, and the latter two burden levels, which do. Jailing a defendant increases the predicted probability of pleading compared to burdens not involving jail time. We observe a similar, though slightly smaller effect in QoL cases. With regards to our second hypothesis, we find that despite differences in the cases received by judges (which are unfortunately not randomized), some variance in the handling of cases by judges can be attributed to the specific judge handling the case.

These two findings together underscore the importance of an objective, randomized, and well-informed justice system in order to achieve fair outcomes. The degree to which a judge is informed or able to estimate a defendant's means and ability to pay bail can make a significant difference in outcomes for that defendant. This suggests that work aimed at estimating accurate and straightforward predictive heuristics for judges to use in setting bail is a promising strategy towards facilitating fairness for all defendants. Likewise, individual differences in setting bail provide the opportunity for further statistical analyses to discern the effect of different bail-setting norms on defendant outcomes, and again suggest that helping judges adjudicate consistently over time and in comparison to each other can help make the justice system more fair towards defendants.

## **6 Conclusion**

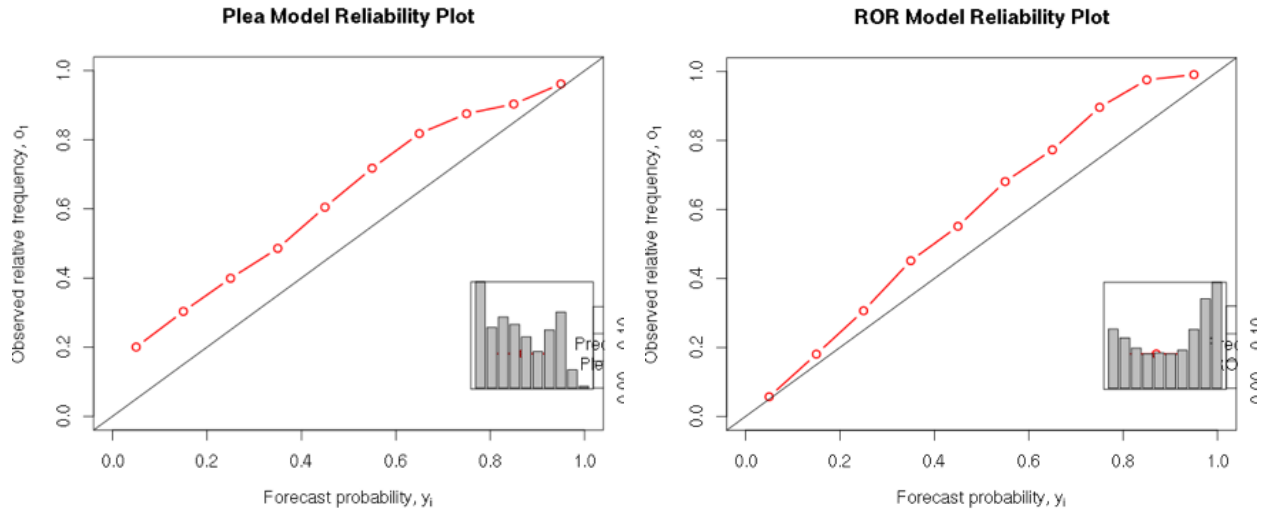
Bail-setting remains an ill-explored area of research; this is particularly the case given its critical role in the criminal justice system. Moreover, the nature of research in this space makes it challenging to run analyses on substantial quantities of data, as we were able to in this case. By adopting a number of modeling approaches and methods of analyses, we were able to verify the effect that pretrial incarceration has on a defendant's willingness to plead guilty. Perhaps unsurprisingly, we found evidence that offers credibility to the fear that bail-setting carries an undue impact on the process.

Our first attempts to quantify and measure a dimension that was previously under-explored suggests that people are experiencing worse outcomes than others on the basis of their inability to post bail. While these findings suggest that the courts are not providing the fairest possible experience to all those who pass through the justice system, we see these findings as a silver lining. Indeed, this research begins the discussion of quantifying and measuring the extent to which pretrial incarceration is affecting defendants' plea rates. This tentative step of measuring, and perhaps tracking, the effects of pretrial incarceration may yield insights to ensure that defendants are receiving just and due process.

## A Appendix

Below we provide further details including confusion matrices and reliability plots on some of the models used.

(a) Non-Quality of Life Cases



(b) Quality of Life Cases

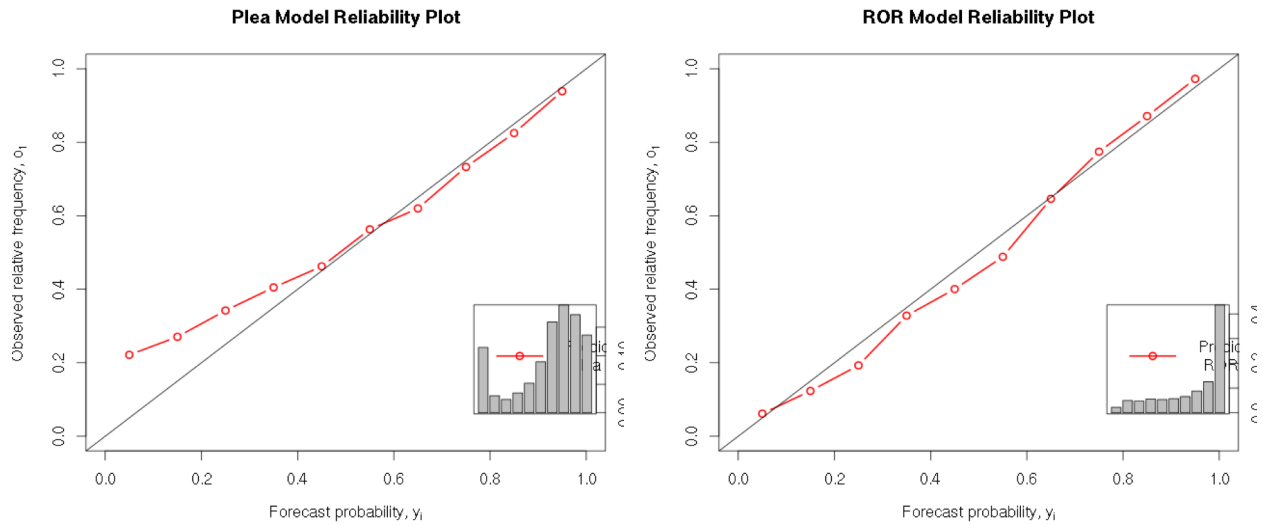
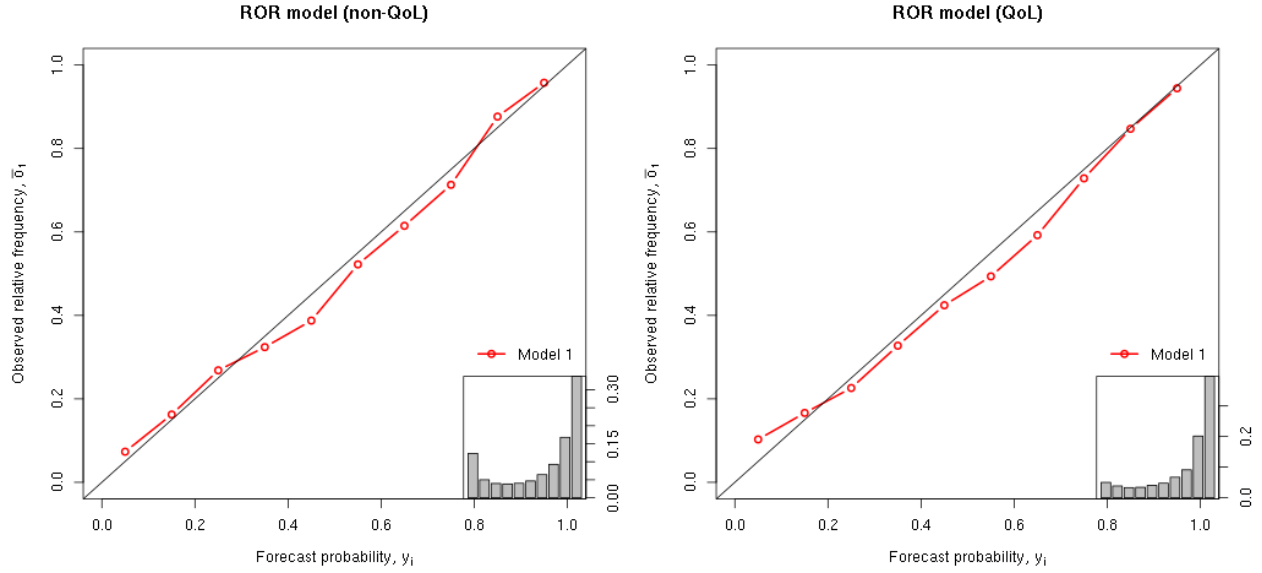
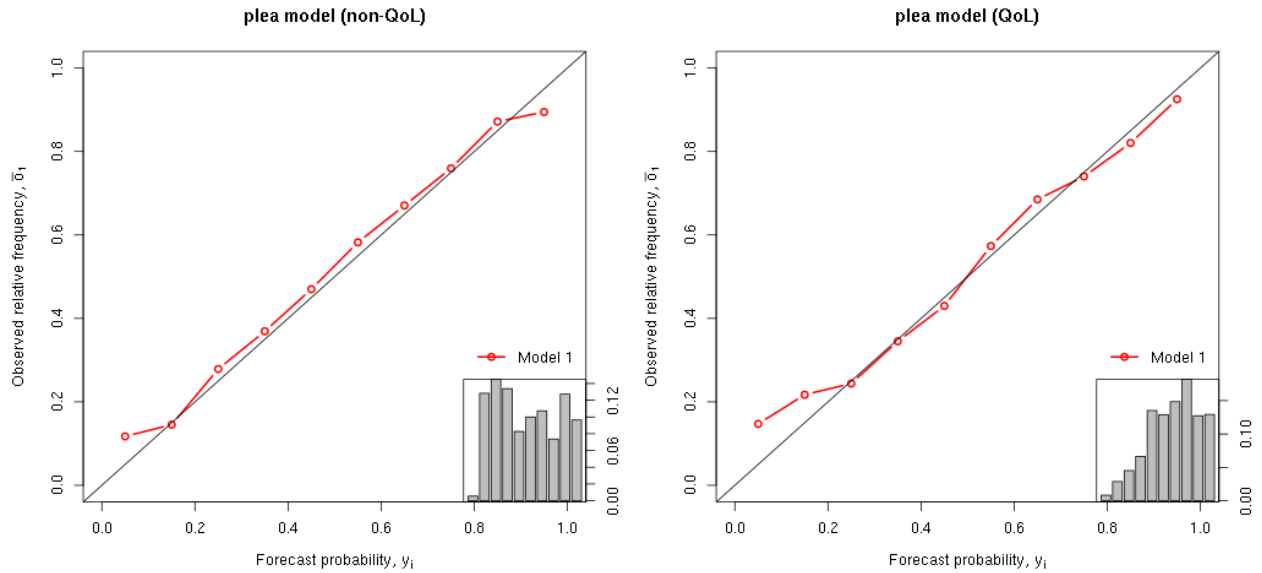


Figure 7: Random Forest Model

(a) ROR Modeling using GLM



(b) Plea Rates Modeled using GLM



## A.1 Validating GLM

With a half and half test/train split, the ROR model had acceptable performance, while the plea model's performance was rather unremarkable. Though their accuracy rates are not abysmal, it is important to keep in mind that both outcomes being predicted have relatively high frequency in the data (ROR rate is close to 65 percent and plea rate is around 50 percent).

## References

- Alschuler, A. W. (1981). The changing plea bargaining debate. *California Law Review*, 69(3):652–730.
- Berg, K. F. (1985). Bail reform act of 1984, the. *Emory LJ*, 34:685.
- Foote, C. (1959). Bail system and equal justice, the. *Fed. Probation*, 23:43.
- Grossman, G. M. and Katz, M. L. (1983). Plea bargaining and social welfare. *The American Economic Review*, 73(4):749–757.
- Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8):1–28.
- Kaplan, J. (1977). American merchandising and the guilty plea: Replacing the bazaar with the department store. *Am. J. Crim. L.*, 5:215.
- Kellough, G. and Wortley, S. (2002). Remand for plea. bail decisions and plea bargaining as commensurate decisions. *British Journal of Criminology*, 42(1):186–210.
- McCoy, T. R. and Mirra, M. J. (1980). Plea bargaining as due process in determining guilt. *Stanford Law Review*, pages 887–941.
- Newman, D. J. and Remington, F. J. (1966). *Conviction: The determination of guilt or innocence without trial*. Little, Brown Boston.
- Schall v. Martin* (1984).
- United States v. Salerno* (1987). United states v. salerno.
- Zacharias, F. C. (1997). Justice in plea bargaining. *Wm. & Mary L. Rev.*, 39:1121.