The Limit of Administrative Data?

Shi Yan
Symposium at the School of Criminology and Criminal Justice
March 29th, 2019
While all research areas much concede some ground to scientific rigor in light of data constraints, it seems clear that the modal approach... yields uncertain evidence about the presence (or absence) of racial disparities because the data available for the task is likely to omit key attributes that vary in prevalence across racial groups and which are related to sentencing decisions.
Motivation: How far can we go?

• Problems: Missing information + linear assumptions
• How much can less restrictive parametric assumptions compensate the missing info?
Data: Computerized Criminal History (CCH) of New York State ($N > 13$ mil)

**CCH**
- Arrest sample
- Arrest events nested under person IDs
- No prosecutor/judge IDs
- No sentencing guidelines and broad discretion

**Conventional datasets**
- Conviction sample
- No person IDs, only arrest IDs
- May or may not have prosecutor and judge info
- Sentencing guidelines and discretion vary by state
Overview: Tasks

- Investigate criminal specialization
- Examine criminal escalation and the sentence
Study 1
How should we understand specialization?
The Assumption of Specialization
Theory & Policy: Same or Different Distributions?

**FIGURE 10.1**
Null and Alternate Hypotheses in Analysis of Variance (ANOVA)

Specialization Research as Part of the Criminal Careers Paradigm
(Piquero et al., 2003)

• Variation or stability
• Transition among offense types
The Measures
(Sullivan et al., 2009)

• Forward Specialization Coefficient
  • “Among those arrested for robbery, 12% also had robbery as their immediately subsequent arrest”

• Diversity Index
  • 0 to 1, where 0 means full specialization

• Latent class analysis
  • Groups: “Driving Specialists,” “Drug Specialists” etc.

• Plus...
  • Repeated most recent crime type?
  • Total priors for the current crime type?
However…
(Yan, 2016)

<table>
<thead>
<tr>
<th></th>
<th>Diversity Index</th>
<th>Repeated most recent</th>
<th>Priors of current crime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeated most recent</td>
<td>0.155</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Priors of current crime</td>
<td>0.023</td>
<td>0.185</td>
<td></td>
</tr>
<tr>
<td>Drug Gen.</td>
<td>-0.257</td>
<td>-0.023</td>
<td>0.259</td>
</tr>
<tr>
<td>High Invol. Gen.</td>
<td>-0.211</td>
<td>-0.031</td>
<td>0.175</td>
</tr>
<tr>
<td>Driving Spec.</td>
<td>0.228</td>
<td>0.098</td>
<td>-0.071</td>
</tr>
<tr>
<td>Property Spec.</td>
<td>0.114</td>
<td>-0.001</td>
<td>-0.038</td>
</tr>
<tr>
<td>Drug Spec.</td>
<td>0.094</td>
<td>0.054</td>
<td>-0.007</td>
</tr>
<tr>
<td>Violent Spec.</td>
<td>-0.045</td>
<td>-0.061</td>
<td>-0.119</td>
</tr>
</tbody>
</table>
To take a closer look...

<table>
<thead>
<tr>
<th>Crime</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Drug possession</td>
</tr>
<tr>
<td>2</td>
<td>Drug sale</td>
</tr>
<tr>
<td>3</td>
<td>Drug sale</td>
</tr>
<tr>
<td>4</td>
<td>Drug possession</td>
</tr>
</tbody>
</table>

• DI = 0.5  
• Class = Drug Spec.  
• Recent crime? Depends
What Does Specialization Suggest?

• Similar nature
• Common cause
• Necessity for special prevention/intervention strategy?

• Classification-prediction?
Also Similar: A Sparse Matrix Problem

\[
\begin{pmatrix}
1.0 & 0 & 5.0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 3.0 & 0 & 0 & 0 & 0 & 0 & 11.0 & 0 \\
0 & 0 & 0 & 0 & 9.0 & 0 & 0 & 0 & 0 \\
0 & 0 & 6.0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 7.0 & 0 & 0 & 0 & 0 & 0 \\
2.0 & 0 & 0 & 0 & 0 & 0 & 10.0 & 0 & 0 \\
0 & 0 & 0 & 8.0 & 0 & 0 & 0 & 0 & 0 \\
0 & 4.0 & 0 & 0 & 0 & 0 & 0 & 0 & 12.0
\end{pmatrix}
\]
Results?

• On the one hand, algorithm targets sparse matrices
• On the other hand... too sparse?
Study 2
How close can criminal records predict the sentence?

(Spoiler: Not very close)
Overview

• Criminal records are correlated with the sentence
  • But most existing research only control for number of priors
  • What happens if I put the entire rap sheet in?
We’ve seen progress in all areas

- Alternative sanctions
- Pre-conviction outcomes
- Cumulative disadvantages

- Physical appearance
- Immigration status
- Less-studied groups (Asians, Native Americans, etc.)

- Criminal career properties
- Prior delinquency and school disciplines

- Strength of evidence
- Aggravating and mitigating factors

- Sentence
- Current
- Criminal
- Extralegal

- Physical appearance
- Immigration status
- Less-studied groups (Asians, Native Americans, etc.)
However, there is one thing remaining...

• Regression is useful
  • Examines relationship between variables and the sentence
  • Predicts the sentence given observed characteristics
Can we get better predictions?
(Abrams, 2016; Piehl & Bushway, 2007)

• Regression models assume underlying functional forms
  • Human decisions can be highly non-linear

• Prediction of the sentence can be useful when...
  • Seeking to reduce extralegal disparities
  • Estimating counterfactuals
  • Just trying to decipher the sentencing process
The present study compares three modeling approaches:

- Among felony defendants, who get incarcerated?
  - In the full sample, 42.8% get incarceration.

- Key outcome: Prediction accuracy
  - Naïve guess of no incarceration for all leads to 57.2% accuracy.
  - Simple logistic regression.
  - Classification tree.
  - Random forest.
Data: New York State Computerized Criminal History

• All felony defendants between 2008 and 2012, who already had one or more prior convictions ($n = 168,811$)
  • To make sure we can connect them to their priors

• 70% random cases as training sample, 30% as testing sample
  • To prevent the model from overfitting the training data
  • “Hide” the testing sample first and train all models on training sample
  • Then examine performance on testing sample
Simple logistic model

• **DV: Incarceration (prison or jail)**
  • Severity + type of current crime
  • Number of prior felony and misdemeanor convictions
  • Race, sex, age, county & year fixed effects

• **Model findings consistent with literature**
  • Strong predictors: Crime severity + number of priors
  • Small but significant racial, gender, ethnic disparities

• **Predicts \( p(\text{inc}) \) for testing sample**
  • prediction = 1 if \( p(\text{inc}) \) >= 0.5
Accuracy = 69.01%

- Actual no incarceration: 79.2%
- Actual incarceration: 55.3%
Decision tree, CART algorithm
(Breiman et al., 1984)
What is Decision Tree?

An iris whose
- petal length is 2.8
- petal width is 1.6

is predicted as versicolor.
Accuracy = 69.70%
Random forest: Combination of trees
Random forest algorithm
(Breiman, 2001)

• Bootstrapped samples
• Random subset of predictors
• Takes minutes to resolve, not hours
Accuracy = 73.07%
Summary of findings

• Modeling non-linearity of criminal records helps somewhat...but only to a certain extent
Study 3
Is Criminal Escalation Related to the Sentence?

(Spoiler: Yes)
## Research vs. Practice

<table>
<thead>
<tr>
<th>ID</th>
<th>Prior Felonies</th>
<th>Prior Misdemeanors</th>
</tr>
</thead>
<tbody>
<tr>
<td>87465</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>98475</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>11254</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>47586</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>
Escalation *Intertwined with Specialization* (Le Blanc, 2002)
Escalation Independent from Specialization

Robbery 3\textsuperscript{rd} Degree \rightarrow \text{Robbery 2\textsuperscript{nd} Degree} \rightarrow \text{Robbery 1\textsuperscript{st} Degree}

Robbery \rightarrow \text{Robbery} \rightarrow \text{Robbery} \rightarrow \text{Robbery}
Escalation: Group-based Trajectory Models (GBTM, Nagin, 2005)

- Identifies different longitudinal patterns within sample.
- Censored normal dependent variable up to cubic term.
• Low Stable (68.0%)
• Moderate Stable (27.8%)
• High Stable (.6%)

• De-escalating (1.6%)
• Escalating (1.9%)
DV: Different Stages in Sentencing
(Cumulative disadvantage, Kutateladze et al., 2014)

• Dismissal
• Reduction to misdemeanor
• In/out decision to incarcerate
Dismissal
(1=Dismissed)
Reduction to Misdemeanor (1=Reduced)

Low Stable: 0.413
Moderate Stable: 0.351*
High Stable: 0.305*
De-escalation: 0.369*
Escalation: 0.284*
In/Out Decision
(1=Incarcerated)

<table>
<thead>
<tr>
<th>Category</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Stable</td>
<td>0.777</td>
</tr>
<tr>
<td>Moderate Stable</td>
<td>0.816*</td>
</tr>
<tr>
<td>High Stable</td>
<td>0.82*</td>
</tr>
<tr>
<td>De-escalation</td>
<td>0.791</td>
</tr>
<tr>
<td>Escalation</td>
<td>0.871*</td>
</tr>
</tbody>
</table>
Summary of Findings

• Own and unique explanatory power on and above number of priors
• Not necessarily all about escalation—less favorable outcomes as long as defendants stand somewhere high
General Discussion
Why Should Criminal Records Matter?  
(Hester et al., 2018; Roberts, 1997)

• **Not self-evident:** Levels of personal risk and diminished chances

• Considered *holistically*, especially when without guidelines—partially solved

• Cumulative process at multiple stages—not solved
  • Overall lack of exclusion restrictions
Additional Variables Are Still Necessary

• Clearly, flexibility is not everything (Study 2)
• Experimental designs and psychometric insights
• RCTs?
Thank you!

Shi Yan

shiyan@asu.edu