**Does the Premium Fit the Risk? The Role of Criminal Escalation in Case Processing**

**Supplemental File:**

**Model Selection for Group-Based Trajectory Modeling and Sensitivity Tests**

**Model Selection for Group-Based Trajectory Modeling**

Group-based trajectory models (GBTM) are a set of mixture models used to divide the sample into meaningful groups. The ultimate variable of interest is group membership—an unobserved categorical variable (in the context of the current study, the type of defendants based on the patterns of their four most recent priors). We used two sets of variables to estimate these models: The x-axis is the sequence of their prior convictions, and the y-axis is the statutory exposure (see the Method section of the main manuscript). Each time, we specify the number of groups we wish the model to identify, then the statistical package estimates the best fitting model for that number of groups, as well as group characteristics. We used the package -traj- on Stata 17.0 for our analysis (Jones & Nagin, 2013).

We started by estimating a one-group GBTM model, then increased the number of groups one at a time, up to six. We selected a four-group model as our best-fitting model and presented it in detail in the Results section of the main manuscript. We made the decision based on fitting statistics, diagnostic statistics, and substantive meaning, all detailed below (for more comprehensive methodological notes, see Nagin, 2005).

First, we used the Bayesian information criterion (BIC) to evaluate model fit. BIC is a model-level indicator of model fit, and is calculated from the log likelihood (a higher value indicates better fit), the sample size, and the number of parameters (the number of groups in our context, a higher value indicates a more complicated model). BIC is positively associated with the log likelihood and negatively associated with the sample size and the number of groups. Overall, a higher BIC indicates better model fit. In most cases, BIC has a negative value, so a BIC with a smaller absolute value indicates a better model fit. As we present in Table A1, the BIC increased with the number of groups, suggesting model fit continued to improve.

Second, we used a set of diagnostic statistics, the average posterior probabilities (avePPs), to examine potential overfitting. The posterior probabilities are a set of individual-level statistics representing the chance for a given individual to belong to a given group, calculated using Bayes’ theorem (e.g., John Doe may have a 70% chance to be in Group 1, 25% chance to be in Group 2, and 5% chance to be in Group 3). Each individual is classified to the Group where they had the highest posterior probability for (in the above case, John Doe would be considered a member of Group 1). As the name suggests, the average posterior probability for each group is the average of the posterior probabilities for that group among all group members (i.e., the average of posterior probabilities for Group 1 for all individuals assigned to Group 1; the average of posterior probabilities for Group 2 for all individuals assigned to Group 2…).

Since we enter the desired number of groups each time, we may end up with too many groups, a phenomenon known as overfitting. One consequence of overfitting is the statistical package can split one group into two very similar ones. And if that happens, the groups will contain members with ambiguous group membership. In theory, when there are four groups, an individual only needs as low as a 25.1% posterior probability to become the member of a group. A high avePP for a group indicates that there is a low level of membership ambiguity for its members. If all groups in a model have high avePPs, it means the entire sample was classified into groups with a low level of ambiguity. As a rule of thumb, Nagin (2005) recommends avePP values at or above 0.7 for all groups in a given model. As we present in Table A1, this was true for all our models up to the five-group one. Our six-group model contained groups with lower avePPs, and we did not considered models with six or more groups.

Lastly, we also considered the substantive meaning of the groups. When everything else is equal, a more parsimonious model (i.e., one with fewer groups) would be more desirable (Nagin, 2005). We present our four-group and five-group models side by side in Figures A1 and A2. Three of the groups were similar between the models: a stable low-level group, a de-escalating group, and an escalating group. The four-group model identified another group with a stable and moderate level of crime seriousness. The five-group model instead identified two other stable groups, one with seriousness between 3 and 4 and the other with seriousness between 5 and 10. We conducted a set of regression analyses (not presented) and found no meaningful disparity in most of the case outcomes between that high seriousness group and the moderate seriousness groups. Moreover, the other three groups (Low Stables, De-escalators, and Escalators in the main manuscript) were highly distinctive from each other and from the Moderate Stables, but we did not feel the differentiation between Moderate and High Stables was very meaning. Although both four-group and five-group models had satisfactory overall fit (BIC) and clear-cut groups (avePPs), we selected the simpler four-group model in the end.

**Sensitivity Analyses**

In the main manuscript, we analyzed a sample of felony defendants who had four or more prior convictions, then estimated a GBTM model on their conviction crime seriousness. As a part of the analysis, we tested whether our main findings still held when we made changes to the original conditions. Our sensitivity test contains: analyses using conviction charges on samples with more extensive criminal records, analyses using arrest seriousness of prior crimes, analyses on the New York City subsample, and analyses using the average sentence length based on the charge. We present the findings of these additional analyses below.

**Conviction Trajectories for Defendants with Five Priors**

These analyses served two purposes: testing the replicability of the trajectory patterns and testing the robustness of risk convergence between De-Escalators and Low Stables. We redid the entire analyses on a new subsamples: defendants with five or more prior convictions (*n* = 44,458). We found four groups with similar patterns, presented in Figure A3. Similar to our main analysis, the crime seriousness dropped sharply for De-escalators, and nearly converged with those of Low Stables starting from the third earliest arrests respectively.

The regression analysis (Table A2, controlled for same set of variables as main analyses but only coefficients for the trajectory groups presented) found very similar results to those in the main analyses. One difference was the treatment of De-escalators. For instance, De-escalators in the 5 prior subsamples were significantly associated with Reduction to Misdemeanor, while there was non-significance in the main analyses. Additionally, Moderate Stables were not significantly associated with Dismissal decisions in the 5 prior conviction subsample, but were significant in the main analyses. Nevertheless, De-escalators and Moderate Stables were still more likely to being incarcerated, received longer incarceration lengths, and were still less likely to have the top charge reduced.

**Arrest Seriousness Trajectories**

We redid the main analyses on defendants with at least four (*n* = 56,017) and five (*n* = 44,458) prior convictions, this time using the seriousness of their prior arrests for the GBTM model. Because of plea bargaining, the seriousness of prior convictions is usually less than that of prior arrests, which may have led to different overall trajectories. We present the trajectory models in Figures A4 and A5.

For both samples of four and five prior convictions, we found the four groups similar to those in the main analysis, with some slight differences. The overall seriousness of all groups increased, and the sizes for Escalators, De-escalators, and Moderate stables slightly increased. As we discussed above, changes to the overall seriousness of the groups can likely be explained by plea bargaining. In the 4 prior and 5 prior conviction samples, De-escalators had again visible drops in the seriousness of recent crimes. Additionally, De-escalators were less likely to receive a reduction to a misdemeanor, where the main analyses found a non-significant effect. Still, regression analyses found similar patterns in the relationships between the group and case outcomes (Table A2).

**New York City Subsample**

This analysis was necessary since the majority of defendants in New York State were processed in New York City (*n* = 23,821 for four or more convictions). Model selection for the New York City (NYC) subsample occurred in a similar manner as the main analyses, with GBTM conducted with one-group to five-group models. The four-group model for the NYC subsample (Figure A6) did not produce a De-escalating group as seen in the original GBTM models. A high stable, or hyperbolic, group emerged in the four-group model, but the size of the high stable group (1.2%) is small. The other groups were similar to the Low Stable, Moderate Stable, and Escalating groups identified in the main analysis with slight differences. For instance, the fourth most recent conviction for the Moderate Stables were more serious than other GBTM models and the Escalators had a more serious second most recent conviction. Still, regression analysis (Table A2) found that Moderate Stables and Escalators are more likely to receive longer sentences compared to Low Stables. Similarly, Moderate Stables and Escalators were more likely to be incarcerated, and both were less likely to receive a charge reduction. Additionally, and unlike the main analyses, the Moderate Stable group was not associated with dismissal.

Figure A1. Four-group GBTM Model on Conviction Seriousness for Defendants with Four or More Convictions, Used as the Main Result in the Manuscript



Figure A2. Five-group GBTM Model on Conviction Seriousness for Defendants with Four or More Convictions



Figure A3. Four-group GBTM Model on Conviction Seriousness for Defendants with Five or More Convictions



Figure A4. Four-group GBTM Model on Arrest Seriousness for Defendants with Four or More Convictions



Figure A5. Four-group GBTM Model on Arrest Seriousness for Defendants with Five or More Convictions



Figure A6. Four-group GBTM Model for New York City Subsample



Table A1. BICs and Average Posterior Probabilities for Model Selection

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Total Groups | 1 | 2 | 3 | 4 | 5 |
| BIC | -281117.73 | -278202.11 | -276819.63 | -275732.16 | -275060.49 |
| AvePP |  |  |  |  |  |
| Group 1 | 1.00 | 0.78 | 0.95 | 0.82 | 0.71 |
| Group 2 | | 0.94 | 0.84 | 0.79 | 0.87 |
| Group 3 | |  | 0.77 | 0.96 | 0.80 |
| Group 4 | |  |  | 0.85 | 0.81 |
| Group 5 | |  |  |  | 0.72 |

Table A2. Sensitivity Test Regression Models

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Dismissal | Reduction to Misdemeanor | In/Out Decision to Incarcerate | Incarceration Length | |
|  |  |  |  |  | |
| *5 prior convictions (n = 44,458)* |  |  |  |  | |
| Moderate Stables | 0.05 | -0.18\*\*\* | 0.17\*\*\* | 9.15\*\*\* | |
|  | (0.03) | (0.03) | (0.05) | (1.12) | |
|  | [0.01] | [-0.07] | [0.05] |  | |
| De-escalators | 0.11 | -0.25\* | 0.26\* | 32.48\*\*\* | |
|  | (0.07) | (0.10) | (0.12) | (7.67) | |
|  | [0.03] | [-0.09] | [0.07] |  | |
| Escalators | 0.05 | -0.32\*\*\* | 0.39\*\*\* | 32.34\*\*\* | |
|  | (0.07) | (0.04) | (0.08) | (6.69) | |
|  | [0.01] | [-0.12] | [0.10] |  | |
|  |  |  |  |  | |
| *4 prior arrests (n = 56,017)* | |  |  |  | |
| Moderate Stables | 0.06\* | -0.12\*\*\* | 0.13\*\* | 11.82\*\*\* | |
|  | (0.02) | (0.03) | (0.05) | (2.10) | |
|  | [0.01] | [-0.04] | [0.01] |  | |
| De-escalators | 0.02 | -0.09\* | 0.14\*\*\* | 8.47\*\*\* | |
|  | (0.02) | (0.04) | (0.04) | (1.81) | |
|  | [0.01] | [-0.03] | [0.04] |  | |
| Escalators | 0.04 | -0.16\*\*\* | 0.25\*\*\* | 15.19\*\*\* | |
|  | (0.04) | (0.04) | (0.06) | (3.91) | |
|  | [0.01] | [-0.06] | [0.07] |  | |
|  |  |  |  |  | |
| *5 prior arrests (n = 44,458)* |  |  |  |  | |
| Moderate Stables | 0.06\*\* | -0.12\*\*\* | 0.15\*\*\* | 10.88\*\*\* | |
|  | (0.02) | (0.03) | (0.03) | (1.91) | |
|  | [0.01] | [-0.04] | [0.04] |  | |
| De-escalators | 0.05 | -0.13\*\* | 0.05 | 10.62\*\*\* | |
|  | (0.04) | (0.05) | (0.05) | (3.06) | |
|  | [0.01] | [-0.05] | [0.02] |  | |
| Escalators | -0.02 | -0.18\*\*\* | 0.22\*\*\* | 13.61\*\*\* | |
|  | (0.06) | (0.04) | (0.07) | (3.75) | |
|  | [-0.01] | [-0.07] | [0.06] |  | |
| *New York City with 4 prior convictions (n = 23,821)* | | | | | |
| Moderate Stables | 0.04 | -0.24\*\*\* | 0.25\*\*\* | 59.82\*\*\* | |
|  | (0.04) | (0.02) | (0.06) | (0.82) | |
|  | [0.01] | [-0.09] | [0.08] |  | |
| High Stable | >-0.001 | -0.42\*\*\* | 0.38\*\*\* | 43.75 | |
|  | (0.10) | (0.08) | (0.09) | | (17.65) |
|  | [>-0.001] | [-0.16] | [0.11] |  | |
| Escalators | -0.07 | -0.34\*\*\* | 0.58\*\*\* | 38.33\* | |
|  | (0.09) | (0.08) | (0.09) | (10.31) | |
|  | [-0.02] | [-0.13] | [0.15] |  | |

Note: All models also controlled for all regressors used in main analysis. For dichotomous outcomes, marginal effects when holding all other regressors at mean in brackets.

References

Jones, B. L., & Nagin, D. S. (2013). A note on a Stata plugin for estimating group-based trajectory models. *Sociological Methods & Research*, *42*(4), 608-613. <https://doi.org/10.1177/0049124113503141>

Nagin, D. S. (2005). *Group-based modeling of development*. Harvard University Press.