Projections Methodology

Summary

The case-level population projection model is designed to simulate the state criminal justice system in order to forecast incarcerated and supervised populations. The model is an agent-based stock and flow simulation that uses patterns in the historical data and sentencing information to calculate how people flow through the criminal justice system. The case-level data enables increased granularity for the underlying populations which makes the model adaptive and accurate. It also allows enforcement of individuals' sentence lengths, such that the behaviors of all people already in the justice system at the run date of the model are projected true to reality. Furthermore, the model enables historical, current, and future state analyses of where population changes are stemming from and the cycles for distinct population groups. The model stays up to date and can be run as frequently as data is received from the DOC.

Detailed Methodology

Input Data and Data Processing

The input data is collected from a standardized direct ingest feed from the Idaho Department of Corrections (IDOC). This data represents information about persons (including PII) who have been involved in the criminal justice system at the state level and the records of their interactions with that system, e.g. charges, sentence terms, periods of incarceration, periods of supervision. The model uses this data on admissions, releases, and sentences as the primary input, such as yearly admissions to jail and prison and the accompanying sentence lengths. The model can adapt to the granularity of the data: yearly, monthly, etc. The model is designed to simulate groups that are treated differently in the system as separate sub-simulations, such as different offense types and genders. Within each sub-simulation, additional disaggregations are handled such as legal statuses (i.e. rider, termers, parole violators). The output of the model is disaggregated to the same level as the input, so more granular input data can provide a more holistic projection for different groups. For instance, if the input data is disaggregated by crime code, then the projected population output would also have an estimated population total per crime code.

Model Infrastructure

The underlying infrastructure of the model uses three core components: Cohorts, Compartments, and Compartment Transitions. These are used to simulate the movement of people through the system and count the total population at each time step.

Compartments

The Compartments component represents an area of the criminal justice system, such as incarceration (termers, riders, parole violators), supervision (parole, probation), and release/out of the system. The Compartment contains a list of all the Cohorts that have been in that area of the system at any point within the simulation.

For instance, the Compartment that represents parole supervision would maintain the total number of people on parole for each time step of the simulation. If the time step is years, then it would count the year-end total parole population.

Cohorts

The Cohort component is used to record the number of people entering part of the system at a certain time and track how many people within that group remain after each time step. For example, one Cohort could represent the population admitted to the new offense prison population compartment in 2016 for a termer incarceration sentence. This Cohort keeps a count for the number of people admitted in 2016 along with the total remaining population after 2017, 2018, and so on until the entire group has left prison. As portions of this group are released over time, additional Cohorts are created for each release period. Cohorts are not only separated by the entry year, but also by the relevant disaggregated attributes like crime type and gender.

As an example, if the yearly admissions to parole were 400 people then the Cohorts in the parole Compartment would be represented in the following table:

Parole Cohort Start Year	EOY 2017	EOY 2018	EOY 2019	EOY 2020
Pre-2017	1,000	700	500	400
2017	400	300	200	100
2018	–	400	300	200
2020	–	–	–	400
Total Population	1,400	1,400	1,400	1,000

In this example, the simulation begins in 2017 and is initialized with the population that started parole before 2017 and then estimates the new admissions to parole and parole terminations each year using the historical data. The number of people leaving the Cohort each year is defined within the Compartment Transitions explained below.

Compartment Transitions

The simulation uses sentence length distributions to calculate the number of people transitioning to a new area of the system after each time step. For example, after each year there is a proportion of the remaining parole population that will complete their sentence and will transition to the released Compartment. Likewise, there is a subset of that group that will have their parole revoked and that portion of the Cohort will transition to the state prison "parole violator" Compartment.

These transitions are defined in transition tables, which contain the probability of transitioning to a new compartment versus remaining in the existing compartment. These probabilities are specific to the time spent in the Compartment and they are computed using the historical data. For example, historical parole length data is used to initialize the parole supervision Compartment Transitions such that the transition tables match the observed sentence length distribution.

Continuing with the parole example above, if the max supervision sentence was 4 years, the parole supervision Compartment Transition table might look like:

	After 1 Year Supervision	After 2 Years Supervision	After 3 Years Supervision	After 4 Years Supervision
Remains on Supervision	75%	66%	50%	0%
Transition to Release	10%	23%	44%	99%
Transition to Incarceration	14%	10%	5%	0%
Transition to Other	1%	1%	1%	1%

Projections Methodology

This table would indicate that 50% of the group that has been on supervision for 3 years will continue on supervision, while 44% of that group will be successfully discharged from parole, 5% will be discharged for revocation and will be re-incarcerated, and 1% will be discharged for other reasons such as death. This table is generated with the available data for each simulation.

Forecasting Admissions
The model uses historical admissions to initialize historical Cohorts. For instance, the 2017 parole Cohort is created with the historical admissions to parole for 2017. However, to initialize Cohorts that have not been observed, an ARIMA forecast is used and is fit using the historical admissions trend. The model uses this to forecast admissions into the future and to backcast historical admissions.

Prediction Intervals
To communicate the confidence of the projections, the model produces prediction intervals that bound the forecast. These prediction intervals are constructed by resampling historical one-period model errors. The model estimates the distribution of model errors at time step t by summing t draws from the historical one-period model errors multiple times with replacement. From this distribution of errors, the model calculates the width of the prediction by subtracting the 2.5th percentile model error from the 97.5th percentile model error. Finally, the model centers this 95% prediction interval width on the projection, thereby bounding the forecast.

County Jails Calculation
For the historical total incarcerated population, we use the movements file to exclude people who have entered County Jails in unpaid beds. Currently, any movement period listed with the fac_cd + lu_cd code in the list 'RTSX', 'RTUT', 'CJVS', 'CJCT' is not counted until a subsequent movement is recorded without any of those codes. The county jails population puts the ingested actuals within 1% of IDOC's historical monthly average total population validated for 2019-2020. More work can be done to incorporate additional logic using the ofndr_loc_hst table in the future, but that table covers a much smaller subset of the unpaid County Jail population.

Initializing the Simulation
The model requires historical data for admissions, releases or terminations, sentence lengths, and any other relevant transitions, which could include revocation data or recidivism data. If this data is disaggregated at any level, then one simulation is initialized per disaggregated group. The model also requires data for the population that is currently in the system to initialize the Compartments, including the total population, the length of time each person has been in the compartment, and their sentence information. The current population gets separate Compartment Transition tables from the incoming population in order to account for the time that has already been served.

Model Output
The Model outputs the total population within each Compartment at each step in time, disaggregated by the same level as the input.

The current breakdowns in this version are:
- [Compartment] Incarcerated (termer, rider, parole violator)
- [Compartment] Supervision (parole, probation)
- Gender (male, female)

Model Validation
The model performance was evaluated on held out datasets from 2018 and 2019. The model was built to minimize the average absolute percent difference each year between the model output and the monthly population for each breakdown (compartment and gender). This performance metric provides a rough idea for how accurate the model projection will be 1-2 years into the future, which is the primary use case for IDOC, and the goal was less than 4% error for the first projection year. The data from 2020 was excluded from the main validation because the admission and release data was quite different due to the pandemic response. However, the projection for 2021-2025 does use the 2020 data in order to incorporate those recent trends.