Milestone Report

Introduction

The U.S. government incarcerates over 1,500,000 inmates in the prison system, which is the largest in the world and keeps growing. Amid concerns about the unsustainable growth of our prison system, begging in the 1980s, prison privatization became a booming industry under government programs to cut back on the federal workforce. The Justice Departments has been contracting private prison corporations for the incarceration of prisoners.

To provide clients of the private prison industry a status report of the current market, this project aims to survey the prisoner population dynamics of the U.S. justice and correction system and summarize the longitudinal trends of incarceration. It can offer important information for business decision for clients and investors in the prison industry, such as regional expansion of revenue and contract for operation of custody facilities in the future.

Data housekeeping

This project mostly utilizes data about the American prison system published by the U.S. Bureau of Justice Statistics. The data summarizes the yearly statistics of the prison population and occupancy rate from the 1970s to 2010s, as well as data of the annual admission to and release from the justice system.

Annual prisoner population, admission and release data

The U.S. Bureau of Justice Statistics surveys the inmates annually and publishes the *Prisoners In XXXX* series, reporting the demographics of the U.S. prisoner population. This project congregates the data in both federal and states' male and female prison system. **Table 1** exemplifies the structure of their data collection.

Originally, District of Columbia was treated as an independent jurisdiction in Bureau of Justice Statistics's survey. However, as of 2001, it is being considered as part of the federal jurisdiction. This created some missing values for the District of Columbia data and the federal data starting in 2001. To correct for the inconsistency, the District of Columbia data prior to 2001 is added into the federal prisoner population during data cleaning.

Jurisdiction	1978	1979	1980	…	2011	2012	2013	2014	2015	2016
Federal	29803	26371	24363	…	216362	217815	215866	210567	196455	189192
Alabama	5625	5464	6543	…	32270	32431	32381	31771	30810	28883
Alaska	712	760	822	…	5597	5633	5081	5794	5338	4434
Arizona	3456	3749	4372	…	40020	40080	41177	42259	42719	42320
Arkansas	2654	3042	2911	…	16108	14654	17235	17874	17707	17537
California	21325	22632	24569	…	149569	134534	135981	136085	129593	130390
Colorado	2486	2668	2629	…	21978	20462	20371	20646	20041	19981
Connecticut	3189	4061	4308	…	18324	17530	17563	16636	15816	14957
Delaware	1325	1419	1474	…	6739	6914	7004	6955	6654	6585
Florida	21436	19748	20735	…	103055	101930	103028	102870	101424	99974
…	…	…	…	…	…	…	…	…	…	…

Table 1. Total Population of Prisoners in the U.S.

The data table was then melted to an analysis-ready and clean format with rows representing observation while columns indicates variables (**Table 2**).

Jurisdiction	Jurisdiction Abbreviation	Year	Population
Alabama	AL	1978	5625

Alabama	AL	1979	5464
Alabama	AL	1980	6543
Alabama	AL	1981	7657
Alabama	AL	1982	9233
Alabama	AL	1983	9856
…	…	…	…

Table 2. Total Population of Prisoners in the U.S. (cleaned)

The *GeoPandas*-compatible geodata of U.S. states were merged with the population data to make possible plotting the choropleth map. The population data and geodata were then aggregated into different regions and divisions defined by the U.S. Census Bureau (**Figure 1**) to generate regional statistics and figures.

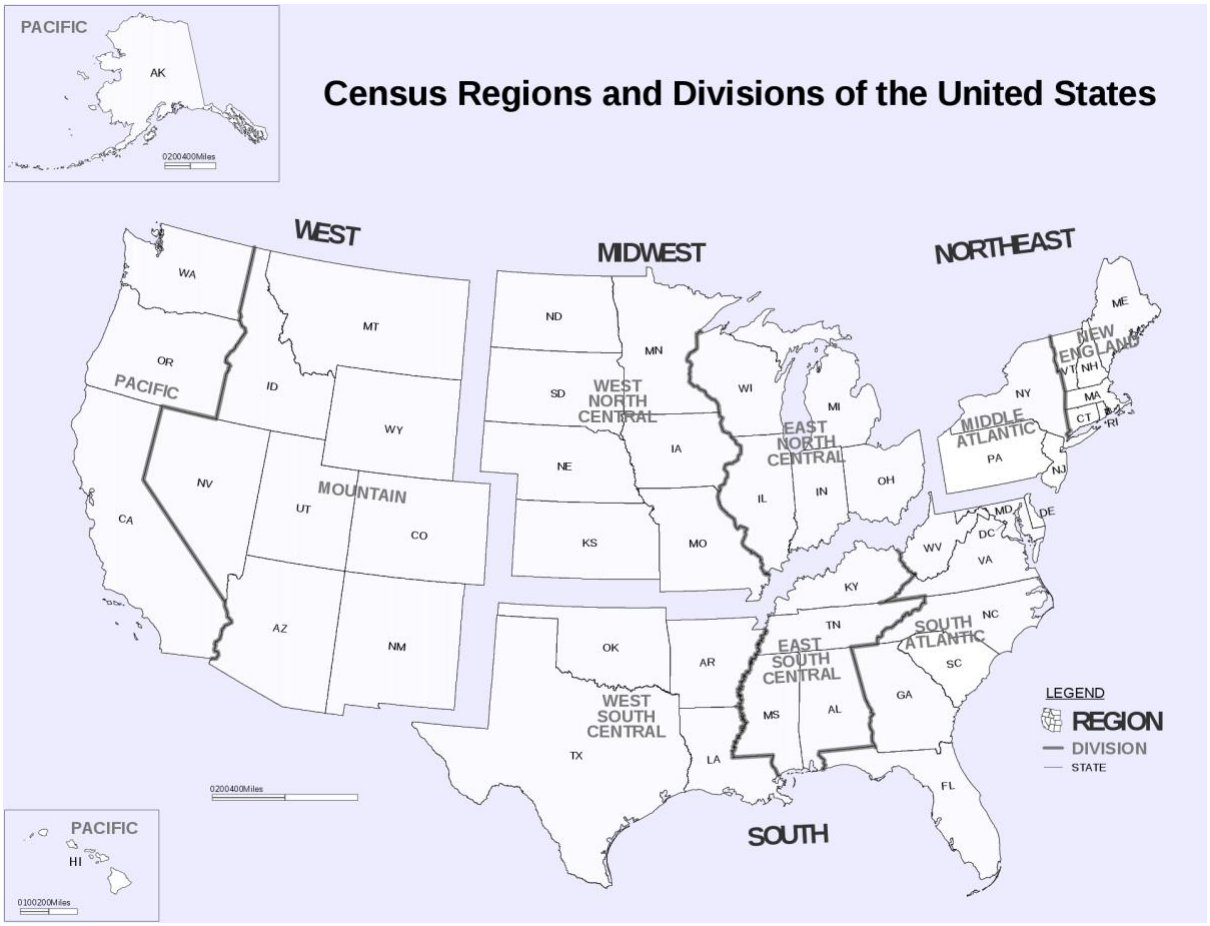

Figure 1. Census regions and divisions of the U.S.

The U.S. Bureau of Justice Statistics also surveys the statistics about annual admission and release in each jurisdiction. They were aggregated in the same fashion as the population data to generate regional statistics.

Facility occupancy rate

The Bureau of Justice Statistics also provides reports about the correction facility occupancy for the 2010s (**Table 3**). Occupancy is defined as the ratio of prisoner population to the capacity of the facilities. Some jurisdictions failed to report their data, which could be identified as *not a number* (NaN) in the table. The data table was consequentially reformatted and merged with the geodata in the same fashion as mentioned above.

Jurisdiction	2011	2012	2013	2015	2016
Federal	138	137.3	133	128	119.7
Alabama	100	99.6	100	98.2	96.3
Alaska	116	142.7	NaN	96.9	98
Arizona	79	80.4	82.4	81.9	81.7
Arkansas	100	97.1	98.7	98.2	97.2
…	…	…	…	…	…

Table 3. Prison occupancy in percentage

Exploratory analysis

Trend in prisoner population

Compared to the data in 1978, the prisoner population in most jurisdictions has increased in both the male and female systems considerably as far as the 2016 data show (**Figure 2**). The elevation is especially prominent in California, Texas and Florida.

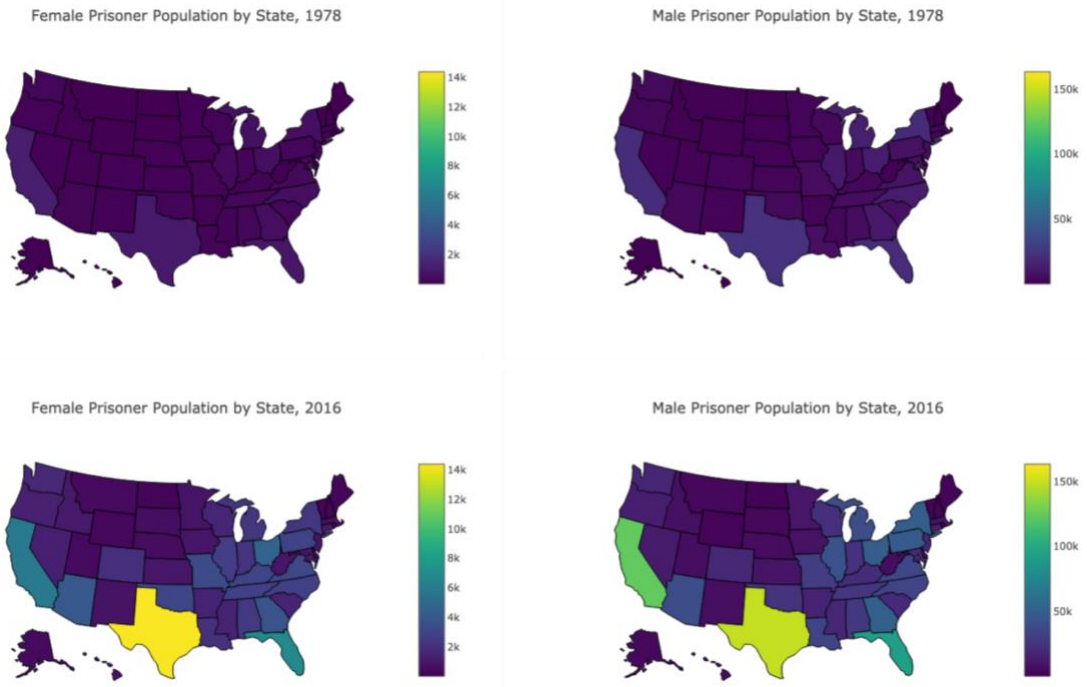

Figure 2. Prisoner population by states

The prisoner population has been growing the fastest in the South (Figure 3). It is mostly driven by the rapid increase in South Atlantic and South West Central due to growth in Florida and Texas, respectively (Figure 4).

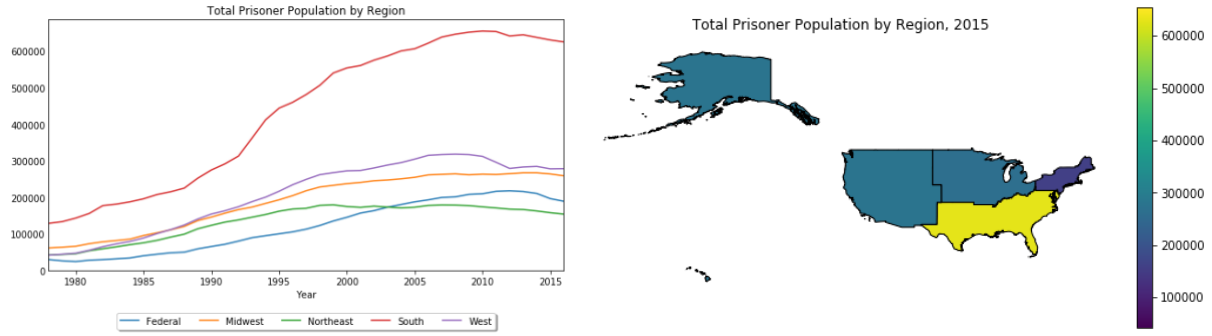

Figure 3. Prisoner population by regions

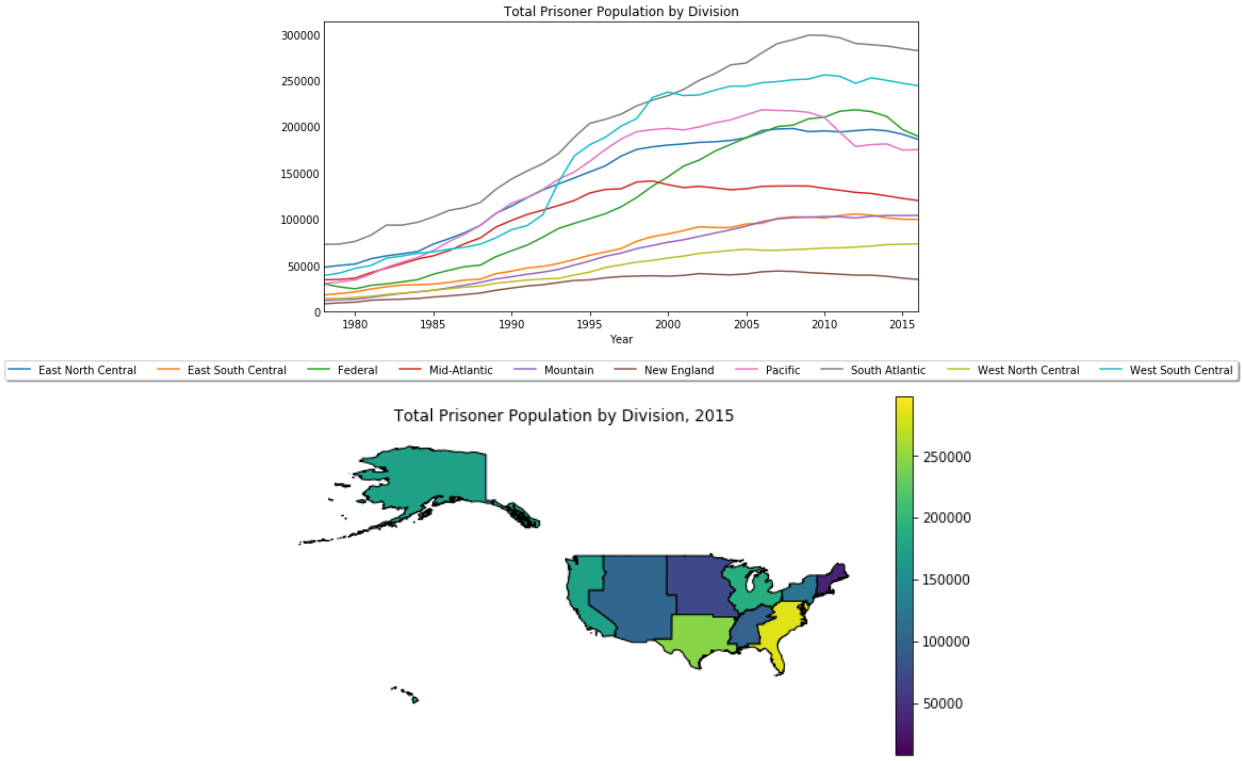

Figure 4. Prisoner population by divisions

Trend in prison occupancy

Two kinds of occupancy rates were reported by the U.S. Bureau of Justice Statistics – low estimation and high estimation (**Figure 5**). The data suggest during the second decade of the 21st century, the occupancy rates of U.S. prisons are reducing. This is especially true for some states such as California. However, for many states, their prisons are still operating over 100% of its capacity.

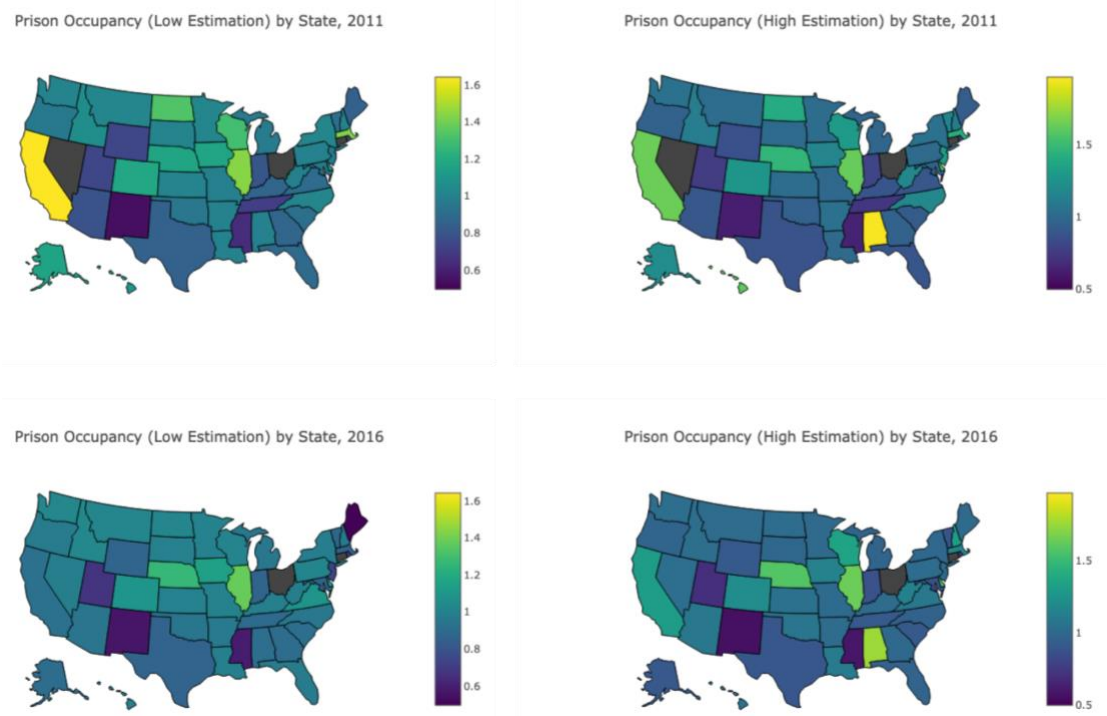

Figure 5. Prison occupancy rate by states

Model

Given data of annual population $p[t]$, admission $a[t]$ and release $r[t]$ (**Figure 6**), we can model the yearly dynamics of prisoner population with a first-degree difference stochastic process

$$p[t] - p[t-1] = a[t] - r[t] + \epsilon p[t-1], \epsilon \sim N(0, \delta^2)$$

where $\epsilon p[t-1]$ is the Gaussian error term and ϵ indicates the error is proportional to the previous year's population. ϵ should be zero-centered with a variance of δ^2. Given past annual survey of the prisoner population, admission and release, the equation above can be used to estimate the variance of ϵ.

Every year, it is assumed that any prisoner's remaining sentence to serve is also a random variable $r[t]$ which is normally distributed. The center of that distribution is year-dependent and denoted $\mu[t]$, while the variance σ^2 is assumed to be invariant. The annual release can be approximated with such distribution's cumulative distribution function

$$r[t] = F(1; \mu[t-1], \sigma^2) p[t-1]$$

$F(1)$ represents the proportion of prisoners whose remaining sentence is less than 1 year.

A second-degree difference equation is used to describe the temporal evolution of the combined remaining sentence of the whole prisoner body

$$p[t]\mu[t] = p[t-1](\mu[t-1] - 1) + 0.5r[t] + a[t]y[t] - 0.5a[t]$$

$p[t-1](\mu[t-1] - 1)$ represents the previous year's population's average number of remaining years to serve reduces by 1 year. However, because $r[t]$ prisoners are being released, 1 year of decrease is an over-estimation. On average, for each prisoner released, a half-year over-estimation is introduced. Thus, it is compensated by the $+0.5r[t]$ term given the prisoners are released evenly throughout the year. $y[t]$ is the average sentence of the admitted prisoner of that year (**Figure 6**). $a[t]y[t]$ represents the contribution of the newly admitted prisoners. Similarly, for each newly admitted prisoner, because the individual will have served some time by the end of the year, we introduced a half-year over-estimation, which is corrected by the $-0.5a[t]$ term.

Provided the annual average sentence of the newly admitted prisoners and the initial state of $\mu[t]$ ($\mu[1994]$ due to the lack of anual sentencing data prior to 1995), equation $p[t]\mu[t] = p[t-1](\mu[t-1]-1) + 0.5r[t] + a[t]y[t] - 0.5a[t]$ can be used to propagate the state of the system and resolve $\mu[t]$ (**Figure 7**).

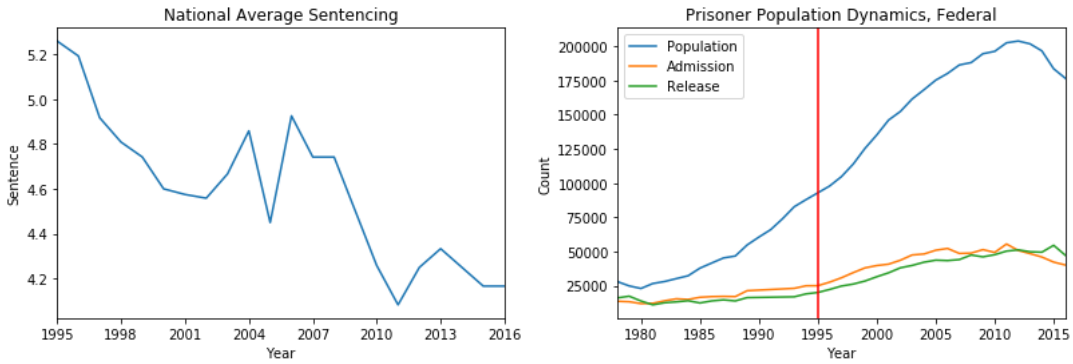

Figure 6. Annual average national sentencing and federal population, admission and release data

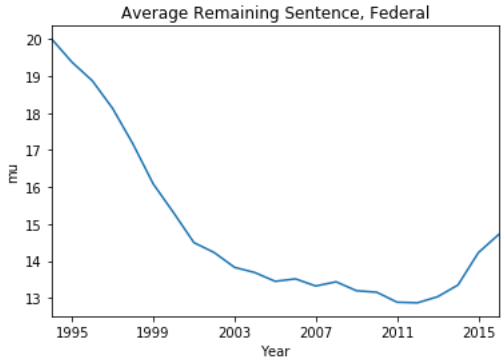

Figure 7. Resolved average remaining sentence to serve for federal prisoners

Using the resolved $\mu[t]$ and the relationship of $r[t] = F(1; \mu[t-1], \sigma^2)p[t-1]$, for each year, a σ is calculated. The average is taken for a point estimate.

Forecast

The annual admission and the average sentence for the admitted can be sampled by building a normal distribution where the variance comes from estimation via past data, while the center is the last-reported data point (2016).

Given $\mu[t-1]$ and σ, we can determine the current year's release by the former year's population using $r[t] = F(1; \mu[t-1], \sigma^2)p[t-1]$. Equation $p[t] - p[t-1] = a[t] - r[t] + \varepsilon p[t-1]$ can then be used to sample for the following year's prisoner population. Finally, equation $p[t]\mu[t] = p[t-1](\mu[t-1]-1) + 0.5r[t] + a[t]y[t] - 0.5a[t]$ can determine current year's μ. In this fashion, one full iteration of state evolution is completed.

By generating 1000 simulations, we can estimate the mean and standard deviation of the projections (**Figure 8**). Shaded area represents MEAN ± SD.

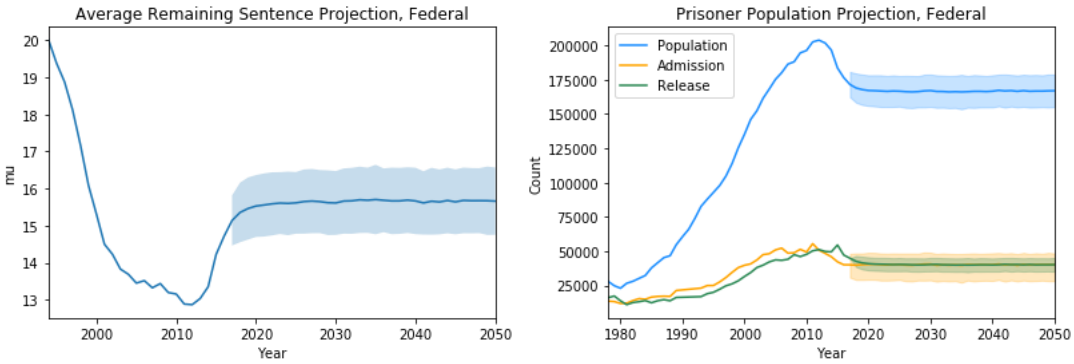

Figure 8. Projection of federal prisoner population dynamics

Conclusion

Compared to the data in 1978, the prisoner population in most jurisdictions has increased in both the male and female systems significantly. Currently, the correction facilities in most states are operating over 100% capacity.

By providing an initial state of the remaining average sentence for the prisoner population, we can extract crucial statistical parameters that make possible the simulation of future population forecast.