Appendix

A National Corrections Reporting Program

Data Description and Preparation

The restricted-use version of the National Corrections Reporting Program (NCRP) provides detailed information on individual admissions to prison, releases from prison, and releases from parole from 1983 until 2009. From 2005 onwards, a subset of states also reports data on the population of prisoners currently in custody. These four different record types are reported separately in four different data sets for each year. In most years, the data do not include the federal prison system.

All four types of records report basic demographic information such as the prisoner's date of birth, gender, race, Hispanic origin, and highest grade completed. The records also contain information on the prisoner's criminal history such as the most recent date of admission, the type of admission and the most serious convicted offense. In addition, the prison release records report the date and type of the most recent release from prison, and the parole release records report the date and type of both the most recent release from prison and the most recent release from parole.

We collapse the type of admission into three different categories: (1) new court commitments and probation revocations, (2) parole revocations, and (3) other admissions (such as returns from appeal or bond, returned escapees, or transfers). New court commitments and probation revocations are typically associated with criminal activity and recent arrests,[1] while parole revocations typically stem from violations of parole conditions. The number of other admissions is very small across states and years, and these admissions do not notably affect total prison populations. Further, these admissions should not be directly affected by changes in corrections policy.

To make offense codes comparable to the FBI data on crimes and arrests, we collapse offense codes into the following 14 categories: murder and manslaughter, forcible rape, robbery, aggravated assault, other assaults, burglary, larceny-theft (excluding motor vehicle theft), motor vehicle theft, other property crime, drug trafficking, drug possession or use, other sex crimes, 'white collar' crime (such as forgery and counterfeiting, fraud, embezzlement, and dealing in stolen property), and other crimes (weapon offenses, DUI and other traffic offenses, obstruction of law enforcement etc.).

For our race-specific analyses, we need a definition of race that is consistent both across the various data sources we employ and over time. As discussed in Appendix B, the FBI arrest data provide counts separately by race (White, Black, Asian, Native American) and by Hispanic origin (Yes, No). However, these data contain no cells defined by the interaction of race and Hispanic origin. Therefore, in the arrest data, it is not possible to distinguish between

[1] Some states do not distinguish court commitments and probation revocations, so a consistent distinction of these two types of admissions is not possible.

Hispanics and Non-Hispanics within each race category. In principle, the information contained in the NCRP race variable should be directly comparable to the information on race in the FBI arrest data. However, information on race is missing for a substantial share of prisoners who are reported to be of Hispanic origin.

Moreover, in 1999, the NCRP introduced an additional race category: Other. From 1999 forward, the majority of Hispanics reports race = Other, but before 1999, the majority of Hispanics reported race = White. Thus, to be consistent over time, we treat all Hispanics as Whites unless they explicitly report their race as Black, Asian, or Native American. Finally, because of their small sample sizes, we combine all race categories other than Black and White into a single Other category.

Data Quality and Selection of States with Reliable Data

A major problem with the NCRP is data availability and reliability. The series begins in 1983, and since then, only 34 states have made reports on anything approaching a regular basis. Further, the data for some of these 34 states exhibit severe internal inconsistencies or external inconsistencies with other data sets such as the NPS. To identify a set of states with reliable data, we conduct four tests. Appendix Table A5 summarizes the results of these tests.

First, we check the internal consistency of the NCRP admission and release records: By construction, the number of admissions in year x according to the admission records should weakly exceed the number of releases that correspond to admissions in year x according to the prison release records. In particular, we are concerned about data quality if subsequent releases from prisoners who report admission in year x exceed total admissions reported in year x by more than 5%. Because the FBI arrest data are not available in 1984, we never use NCRP data prior to 1985. Therefore, we do not examine the quality of admissions data prior to 1985.

The NCRP data for CA, IL, MI, MN, NE, NH, NY, ND, OH, SC, UT, VA, and WA pass this internal consistency check for every year of admission from 1985 forward. Further, data for CO, FL, GA, KY, MS, NJ, and WI fail this test in only one year. All other states show severe inconsistencies in multiple years.

Second, we check whether age-specific stocks in the post-2005 NCRP files are consistent with NCRP admission and release flows from 1987 to the 2005-2009 period.[2] In particular, if we assume that convicted offenders age 15 and under do not enter regular prisons, then, for cohorts born in 1972 and later, the cumulative difference between total admissions and total releases from 1987 through any year between 2005 and 2009 should be comparable to the stock of prisoners from that birth cohort who are recorded in the relevant 2005-2009 stock file. In particular, we question the reliability of the NCRP data if deviations exceed 10%.

We cannot perform this test on 13 of the 34 NCRP states (AL, HI, IL, KY, MS, ND, NE, NH, NV, OH, UT, VA, WV). Some never report any stock data.

[2]We choose 1987 rather than 1985 as the starting year since for some states that do not end up in our final sample, the data are incomplete prior to 1987.

Others do not report flow data in each year. Among the 21 states we can test, 12 report data that contain discrepancies that we are willing to tolerate: CA, CO, MI, MN, NJ, NY, NC, OR, SC, TX, WA, and WI.[3] Thus, the following states have data that are acceptable given our first two tests: CA, CO, MI, MN, NJ, NY, SC, WA, WI.

Finally, we check the external consistency of the NCRP admissions and release data by comparing these data to corresponding data from the NPS and by comparing annual net flows implied by NCRP data to annual changes in NPS stock data. The NCRP counts prisoners in custody, and before 1999, the NPS stock data recorded all prisoners in custody. Thus, pre-1999 changes in the NPS stock data should be comparable to net flows in the NCRP.

However, from 1999 onwards, the NPS reports the number of prisoners by jurisdiction rather than custody, and this implies that year to year changes in NPS stocks may not match net flows in the NCRP after 1998. Further, in all years, the NPS data on admissions and releases covers prisoners under jurisdiction instead of prisoners in custody, and it only covers prisoners with sentences of at least one year.

Here, we perform two sets of checks on data from the period 1983-1998. We compare (i) the NPS admission and release flows to the corresponding NCRP flows,[4] and we compare (ii) annual net NCRP flows to the corresponding annual changes in NCRP stocks. We place more weight on the comparisons in (ii) because this check involves NPS and NCRP data defined on the same universe, i.e. prisoners in custody. Starting with the set of 9 states that passed our first two tests, we eliminate only MN, leaving CA, CO, MI, NJ, NY, SC, WA, WI.[5]

A final issue with the NCRP data is that within some state-year cells, key variables such as type of admission or year of admission to prison exhibit very high missing rates or are not reported at all. Due to this missing data problem, we exclude NY from our analysis. NY data do not report the type of admission in the prison release records for 1985, 1986, and 1988-1993.

For our remaining 7 states—CA, CO, MI, NJ, SC, WA, and WI—missing rates for all key variables are typically small within state-year cells, and we impute missing information on gender, age, race, and type of offense committed using a hot-deck procedure that conditions on state, year, and the other key characteristics, treating missing values on other characteristics as a separate category. In particular, we use version 1.52 of Stata's hotdeck command, which

[3]The data from CA and WA contain the most violations of our ten percent rule. However, data from the clear majority of cohorts we test in both of these states pass our 10% deviation test.

[4]In states that do not import or export prisoners, we expect NCRP flows that are larger than the NPS flows since the NCRP flow data include prisoners sentenced to one year or less, while the NPS flow data do not.

[5]This selection procedure is more elaborate than the one in Pfaff (2011), who selects states that report for at least eleven consecutive years and for which the total number of admissions and releases does "not deviate too much (generally $< 10\%$)" (Pfaff (2011), p. 520) from the corresponding NPS data. Pfaff's (2011) rules select CA, CO, KY, IL, MI, MN, NE, NJ, SD, VA, and WA. Table A5 details the checks that KY, IL, MN, NE, SD, and VA fail in our analyses.

is downloadable from the stata homepage
(http://www.stata.com/stb/stb54/sg116_1/hotdeck.ado).

Constructing Admissions and Stock Data For 1985-2005

If the NCRP data were perfectly consistent, the following identity would hold for any cell defined by state, race, offense, and type of admission:

$$adm_y = pop_{y,Z} + \sum_{z=y}^{Z} rls_{y,z} \tag{1}$$

Here, adm_y denotes the number of admissions in year y according to the admission records, $pop_{y,Z}$ denotes the population of prisoners admitted in year y that remains in prison at the end of year $Z \geq y$, and $rls_{y,z}$ is the number of prisoners admitted in year y and released in year $z \geq y$.

However, it turns out that even in the set of states that pass all our internal consistency checks, the right hand side tends to slightly exceed the left hand side, i.e., the admission records contain too few observations compared to subsequent releases and remaining stocks. Thus, if we used the admission counts from the admission records to calculate the time served distributions in Tables 7 and 8 as well as the release probabilities that characterize the 1985 corrections policy for our simulations, we would likely overestimate the likelihood of short prison sentences and underestimate the likelihood of prison sentences that are right-censored. Moreover, for some state-race-offense cells, the number of release records that record 1985 as the admission year slightly exceeds the total admissions in the 1985 admission records. Thus, if we used both the admissions and release file to estimate release probabilities for the 1985 cohort, we would create probabilities that sum to more than 1. To avoid this problem, we construct admissions counts from the prison stock and release files using the identity in equation (1). Given our state selection rules, these time series of constructed admissions track the time series of admissions from the NCRP admission records quite closely.

After we construct these admissions data, we recursively calculate end-of year prison populations for each year in our sample. For each state, we start with the most recent annual stock file and work backwards using the fact that the prison population at the end of year $y-1$ equals the prison population at the end of year y plus releases in year y minus admissions in year y.[6]

[6]One may argue that given the slight inconsistencies across different types of records, a more precise estimate of prison populations prior to 2005 would start with the actual stock data at the end of 2005. However, Colorado does not report stock data prior to 2009, and for the remaining 6 states, the deviations between the two measures are negligible. In particular, the total prison population for these 6 states amounts to 305,200 using the actual 2005 stock data and to 304,600 using the estimated stock data—a deviation of less than .2%.

B Uniform Crime Reports

The FBI Uniform Crime Reports (UCR) provide agency-level monthly data on reported crimes and arrests for 43 offense categories. For each offense, the arrest data are reported by either (1) age and gender, (2) race, or (3) Hispanic origin. The data are available from 1980-83 and 1985-2009. We downloaded this data from the NACJD website[7] on August 3, 2012.

We use the agency-level data to create estimates for the total number of crimes and arrests for our 14 broad offense categories in each state with reliable NCRP data (see Appendix A). Since some agencies do not submit a report for each month, we use the following procedure to estimate state-wide totals of arrests.

First, we impute missing monthly information for agencies that report at least one month of data in a given year. Our default procedure is to create annual estimates by 'inflating' the number of reported arrests, i.e., we calculate average monthly reported arrests using the valid monthly reports and then multiply this number by 12. However, some agencies report so few arrests that it is more reasonable to assume that they have missing monthly reports because they only submit a report when they have actually made arrests. Other agencies tend to 'backlog file' to reduce the administrative burden of reporting to the UCR, i.e., they do not report for x months and then include the arrests for these x months in the report for month $x + 1$. Thus, we take the simple sum across reported months as the number of total arrests for an agency-year cell if one of the following conditions is satisfied:

1. The agency reports no more than 20 arrests across offense categories in a given year.

2. The agency only reports in December.

3. There is a 'spike' in reported arrests after a period of non-reporting. More specifically, the maximum monthly arrest total within a year is more than 1.5 times the average arrests in other months within this year and occurs after a period of non-reporting.

4. The agency reports for at most 6 months, the reported arrests are at least half the average across years in which the agency reports in each month, and the inflation procedure would produce an annual estimate that is more than twice as high as this average.

If none of these conditions is satisfied, we use the inflation procedure. At the end of this step, we have an estimate for the annual number of arrests by offense for any agency that reports at least one month of data in a given year.

Second, we use linear-interpolation to fill in non-reporting spells of 3 years or less.[8] After this step, we have annual estimates for a set of agencies in each state that cover about 97% of each state's population.

[7] http://www.icpsr.umich.edu/icpsrweb/content/NACJD/guides/ucr.html

[8] We do not extrapolate if the agency starts or ends with a non-reporting spell.

To create state-wide annual arrest rates, we divide the total number of arrests for each offense category in each state by the fraction of the state's population that is covered by agencies for which we have (original or imputed) arrest data.

C Simulation of Counterfactual Prison Populations

This appendix describes how we estimate counterfactual prison populations under the assumption that the sentencing and parole policies that determined time-served for the 1985 admissions cohort remained in place for future years. We proceed in three steps: First, we formalize our definition of a policy. Second, we describe the estimation of policy parameters in the base year 1985 given the data available to us. Third, we discuss how we use the estimated policy parameters for 1985 and the available data to produce counterfactual prison populations from 1986 onwards.

Our Definition of Policy

We define the policy in year y as a set of transition probabilities that apply to two at-risk populations. These populations are divided into cells defined by race, offense, and geography.

1. The policy for the population of arrested alleged offenders in year y is

$$\mathcal{P}_y^{arr} \equiv \left(\alpha_y^{cc}, \rho_{y,y}^{cc,p}, ..., \rho_{y,y+T}^{cc,p}, \rho_{y,y}^{cc,np}, ..., \rho_{y,y+T}^{cc,np}\right).$$

2. The policy for the population of parolees in year y is

$$\mathcal{P}_y^{prl} \equiv \left(\alpha_{y-P,y}^{p,pr}, .., \alpha_{y,y}^{p,pr}, \rho_{y,y}^{pr,p}, ..., \rho_{y,y+T}^{pr,p}, \rho_{y,y}^{pr,np}, ..., \rho_{y,y+T}^{pr,np}\right).$$

Here, T is the maximum possible prison spell; P is the maximum parole spell; α_y^{cc} is the probability of admission to prison in year y by a court commitment[9] given arrest in year y; $\rho_{y,z}^{cc,p}$ ($\rho_{y,z}^{cc,np}$) is the probability of release to parole (non-parole) in year z given admission to prison in year y by a court commitment; $\alpha_{x,y}^{p,pr}$ is the probability of a parole revocation in year y given release from prison to parole in year x; and $\rho_{y,z}^{pr,p}$ ($\rho_{y,z}^{pr,np}$) is the probability of release to parole (non-parole) in year z given a parole revocation in year y.

Given the available data, we must impose several strong assumptions in order to use this state-transition model. First, we assume that each convicted defendant is convicted of the offense listed in the UCR arrest records. Second, we assume that all offenders sentenced to prison actually enter prison in the

[9] As in our previous analyses, we have to pool court commitments and probation revocations because some states do not distinguish between these two types of admission.

same year they are arrested. Third, we assume that, among prisoners who last entered prison due to parole revocations, the distribution of completed spell lengths does not depend on prior time spent on parole or the number of previous parole revocations.

Further, we assume that a given individual cannot experience more than one parole revocation in a year, i.e. if an individual experiences a parole revocation in year y and is subsequently released to parole in the same year, y, the probability of another revocation in that year is zero in our simulations. While our results are not sensitive to this assumption, it greatly simplifies the task of describing how the model works.

Estimating the Policy in the Base Year

We divide the sample into cells defined by the interaction of three race groups, 14 offense categories, and two geographies. See Appendix A for definitions of the race and offense categories. We divide geography into California and our other six NCRP states. California is a large state and the history of changes to its criminal justice system is somewhat unique.

Suppressing indices for race, offense, and geography to simplify notation, let

$$
\begin{aligned}
arr_y &= \text{\# of arrests in year } y \\
adm_y^{cc} &= \text{\# of admissions in year } y \text{ due to a court commitment or probation revocation} \\
adm_y^{pr} &= \text{\# of admissions in year } y \text{ due to a parole revocation} \\
rls_{y,z}^{cc,np} &= \text{\# of people released from prison to non-parole in year } z \text{ who have been admitted in year } y \text{ due to a court commitment or probation revocation} \\
rls_{y,z}^{cc,p} &= \text{\# of people released from prison to parole in year } z \text{ who have been admitted in year } y \text{ due to a court commitment or probation revocation} \\
rls_{y,z}^{pr,np} &= \text{\# of people released from prison to non-parole in year } z \text{ who have been admitted in year } y \text{ due to a parole revocation} \\
rls_{y,z}^{pr,p} &= \text{\# of people released from prison to parole in year } z \text{ who have been admitted in year } y \text{ due to a parole revocation} \\
prl\_rev_{y,z}^{cc} &= \text{\# of people released from parole to prison in year } z \text{ who have been released from prison to parole in year } y \text{after being admitted due to a court commitement or probation revocation} \\
prl\_rev_{y,z}^{pr} &= \text{\# of people released from parole to prison in year } z \text{ who have been released from prison to parole in year } y \text{ after being admitted due to a parole revocation} \\
pop_{y,z}^{cc} &= \text{prison population at the end of year } z \text{ that has been admitted in year } y \text{ due to a court commitment or probation revocation}
\end{aligned}
$$

$pop^{pr}_{y,z}$ = prison population at the end of year z that has been admitted in year y due to a parole revocation

Also, let

$$rls^{p}_{z} = \sum_{y \leq z} \left(rls^{pr,p}_{y,z} + rls^{cc,p}_{y,z} \right)$$

be the total number of people released to parole in year z, and let $pop^{i}_{z} = \sum_{y \leq z} pop^{i}_{y,z}$ for $i \in \{cc, pr\}$ denote the total prison population at the end of year z due to a court commitment/probation revocation or a parole revocation, respectively.

Then, the policy parameters of the sentencing regime in year y can be estimated as follows:

$$\begin{aligned}
\widehat{\alpha}^{cc}_{y} &= \frac{adm^{cc}_{y}}{arr_{y}} \\
\widehat{\alpha}^{p,pr}_{x,y} &= \frac{prl_rev^{cc}_{x,y} + prl_rev^{pr}_{x,y}}{rls^{p}_{x}} \\
\widehat{\rho}^{i,j}_{y,z} &= \frac{rls^{i,j}_{y,z}}{adm^{i}_{y}} \ \forall \ i \in \{cc, pr\} \text{ and } j \in \{p, np\}
\end{aligned}$$

The UCR arrest files allow us to directly determine the number of arrests within each cell and for each year from 1985 onwards, which gives us arr_y for $y \geq 1985$. Moreover, the number of releases from prison in year z given admission in year y can be directly calculated from the NCRP prison release records. This gives us $rls^{cc,np}_{y,z}$, $rls^{cc,p}_{y,z}$, $rls^{pr,np}_{y,z}$, and $rls^{pr,p}_{y,z}$ for $y \geq 1985$ and $z \geq y$. Also, for each state in our seven state sample, we observe $pop^{cc}_{y,z}$ and $pop^{pr}_{y,z}$ for some z between 2005 and 2009 and $y \leq z$. As described in Appendix A, we do not use the admission records to calculate adm^{cc}_{y} and adm^{pr}_{y}. Instead, we use the release data and the most recent data on prison populations to construct admissions using the following identities

$$adm^{cc}_{y} = pop^{cc}_{y,Z} + \sum_{z=y}^{Z} (rls^{cc,p}_{y,z} + rls^{cc,np}_{y,z}) \tag{2}$$

and

$$adm^{pr}_{y} = pop^{pr}_{y,Z} + \sum_{z=y}^{Z} (rls^{pr,p}_{y,z} + rls^{pr,np}_{y,z}). \tag{3}$$

The data series discussed so far allow us to estimate all elements of $\mathcal{P}^{cc}_{1985}$. Estimating $\mathcal{P}^{pr}_{1985}$ requires more work. First, we need releases to parole for 1982-1984 by race and offense, which are not available in the NCRP. We estimate these release counts as follows: We take the number of total releases for each of these three years and each of our seven states from the NPS and assume that the composition of these releases with respect to release type, race, and

offense is identical to 1985. The result becomes our estimator for rls^p_x, $x \in \{1982, 1983, 1984\}$.

Next, we use the NCRP parole release records to calculate the number of parole revocations in 1985 for people released from prison to parole in or before 1985. Since CO and WA do not provide parole release records throughout the 1980s, we have to restrict our data set to five out of seven states for this calculation. Very few people get their parole revoked after more than 3 years on parole, so we assume $P = 3$ and treat parole revokees in 1985 who were released to parole before 1982 as if they were released in 1982. This yields estimates for $prl\_rev^{cc}_{x,1985} + prl\_rev^{pr}_{x,1985}$, $x \in \{1982, 1983, 1984, 1985\}$.

The resulting estimates for $\alpha^{p,pr}_{x,1985}$ can be cross-checked for each state by comparing actual parole revocations in 1985 to the number implied by the estimated policy. It turns out that our estimation procedure works well for California, but we tend to underestimate parole revocations for the other 6 states (either because the parole release records under-report releases from parole to prison or because CO and WA have higher revocation rates than MI, NJ, SC and WI). To correct for this bias, we rescale $\widehat{\alpha}^{p,pr}_{x,1985}$ by the ratio of actual parole revocations in 1985 over estimated parole revocations in 1985 that result from releases to parole 1982-1985 and $\widehat{\alpha}^{p,pr}_{x,1985}$.

Finally, estimating $\rho^{pr,p}_{y,z}$ is straightforward given constructed admissions for parole revocations in 1985 and the subsequent flow of releases 1985-2005.

Simulation of Counterfactual Prison Populations

In what follows, we let a hat over a variable denote the counterfactual value of that variable. We assume that changes in sentencing and parole policies after 1985 do not affect the time series of prison releases for those admitted to prison before the end of 1985. Thus, we let

$$\widehat{rls}^{i,j}_{y,z} = rls^{i,j}_{y,z} \ \forall \ i \in \{cc, pr\}, \ j \in \{p, np\}, \ y \leq 1985, \ z \geq 1985.$$

To begin, we ignore potential effects of changes in sentencing and parole policies on criminal activity or the probability of arrest given involvement in crime. Under this assumption, we can take the time series of arrests throughout the period as given and let

$$\widehat{arr}_y = arr_y.$$

From January 1, 1986, on, we impose P^{cc}_{1985} on the actual time series of arrests to calculate counterfactual time series of prison admissions and subsequent releases from these admissions to parole/non-parole from 1986 onwards. That is, for $y > 1985$ and $z \geq y$, we can calculate

$$\begin{aligned}
\widehat{adm}^{cc}_y &= \widehat{\alpha}^{cc}_{1985} \times arr_y \\
\widehat{rls}^{cc,p}_{y,z} &= \widehat{adm}^{cc}_y \times \widehat{\rho}^{cc,p}_{1985,1985+z-y} \\
\widehat{rls}^{cc,np}_{y,z} &= \widehat{adm}^{cc}_y \times \widehat{\rho}^{cc,np}_{1985,1985+z-y}
\end{aligned}$$

It is more difficult to calculate the corresponding counterfactual time series of parole revocations and releases of revokees to parole/non-parole from 1986 onwards. This series is based on the *counterfactual* time series of releases to parole, which is itself a result of the counterfactual policy in previous years. Thus, we apply the following procedure:

We initialize the time series of counterfactual releases to parole with

$$\widehat{rls}_x^p = rls_x^p \ \forall \ x = 1982, ..., 1985.$$

Recall, we assume that a given individual can experience no more than one parole revocation within a given year. Thus, for $y > 1985$, we have

$$\begin{aligned} \widehat{adm}_y^{pr} &= \widehat{\alpha}_{1985,1985}^{p,pr} \times \left(\widehat{rls}_{y,y}^{cc,p} + \sum_{x<y} \left(\widehat{rls}_{x,y}^{pr,p} + \widehat{rls}_{x,y}^{cc,p} \right) \right) + ... \\ &\quad ... + \widehat{\alpha}_{1984,1985}^{p,pr} \times \widehat{rls}_{y-1}^{p} + \widehat{\alpha}_{1983,1985}^{p,pr} \times \widehat{rls}_{y-2}^{p} + \widehat{\alpha}_{1982,1985}^{p,pr} \times \widehat{rls}_{y-3}^{p}. \end{aligned} \tag{4}$$

Substituting

$$\widehat{rls}_{y,y}^{cc,p} = \widehat{\rho}_{1985,1985}^{cc,p} \widehat{\alpha}_{1985}^{cc} arr_y$$

into equation (4) yields

$$\begin{aligned} \widehat{adm}_y^{pr} &= \widehat{\alpha}_{1985,1985}^{p,pr} \times \left(\widehat{\rho}_{1985,1985}^{cc,p} \widehat{\alpha}_{1985}^{cc} arr_y + \sum_{x<y} \left(\widehat{rls}_{x,y}^{pr,p} + \widehat{rls}_{x,y}^{cc,p} \right) \right) + ... \\ &\quad ... + \widehat{\alpha}_{1984,1985}^{p,pr} \times \widehat{rls}_{y-1}^{p} + \widehat{\alpha}_{1983,1985}^{p,pr} \times \widehat{rls}_{y-2}^{p} + \widehat{\alpha}_{1982,1985}^{p,pr} \times \widehat{rls}_{y-3}^{p}, \end{aligned}$$

which can be calculated from the contemporaneous data on arrests and (counterfactual) releases to parole for admissions in previous years.

Given $\widehat{adm}_y^{pr}$, we can now calculate

$$\begin{aligned} \widehat{rls}_{y,z}^{pr,p} &= \widehat{adm}_y^{pr} \times \widehat{\rho}_{1985,1985+z-y}^{pr,p} \\ \widehat{rls}_{y,z}^{pr,np} &= \widehat{adm}_y^{pr} \times \widehat{\rho}_{1985,1985+z-y}^{pr,np}. \end{aligned}$$

Thus, we can iteratively calculate $\widehat{adm}_y^{pr}$, $\widehat{rls}_{y,z}^{pr,p}$, and $\widehat{rls}_{y,z}^{pr,np}$ for $y > 1985$ and $z \geq y$ and update the time series of counterfactual releases to parole with

$$\widehat{rls}_y^p = \sum_{x \leq 1985} \left(rls_{x,y}^{pr,p} + rls_{x,y}^{cc,p} \right) + \sum_{1985 < x \leq y} \left(\widehat{rls}_{x,y}^{pr,p} + \widehat{rls}_{x,y}^{cc,p} \right).$$

The calculation of the time series of counterfactual releases to non-parole, $\widehat{rls}_y^{np}$, is analogous.

Finally, given $\left\{ pop_{1985}^i, \widehat{adm}_y^i, \widehat{rls}_{x,y}^{i,p}, \widehat{rls}_{x,y}^{i,np} \right\}_{i \in \{cc,pr\}}$, we can calculate the counterfactual time series of prison populations within each cell as

$$\widehat{pop}_y^i = \widehat{pop}_{y-1}^i + \widehat{adm}_y^i - \sum_{x \leq 1985} \left(rls_{x,y}^{i,p} + rls_{x,y}^{i,np} \right) - \sum_{1985 < x \leq y} \left(\widehat{rls}_{x,y}^{i,p} + \widehat{rls}_{x,y}^{i,np} \right).$$

Accounting for Incapacitation and Deterrence

To account for incapacitation and deterrence effects of changes in sentencing policies, we assume that there is (1) a constant elasticity between the total prison population across all offenses in year $y-1$ and criminal activity in year y and (2) a constant ratio of arrests to reported crime over time. In particular, let β_j denote the elasticity for offense category $j \in J$, $arr_{y,j}$ the number of arrests in year y for offense category j, and $pop_{y,j}$ the prison population in year y for inmates committed for offense category j. The two assumptions above imply the following relationship between the ratio of lagged counterfactual and actual total prison populations and the ratio of contemporaneous counterfactual and actual arrests:

$$\frac{\widehat{arr}_{y,j}}{arr_{y,j}} = \left[\frac{\sum_{j \in J} \left(\widehat{pop}_{y-1,j} \right)}{\sum_{j \in J} \left(pop_{y-1,j} \right)} \right]^{\beta_j}$$

or

$$\widehat{arr}_{y,j} = arr_{y,j} \left[\frac{\sum_{j \in J} \left(\widehat{pop}_{y-1,j} \right)}{\sum_{j \in J} \left(pop_{y-1,j} \right)} \right]^{\beta_j}$$

Thus, given the counterfactual end-of-year prison populations in year $y-1$ and the actual data on arrests and prison populations, we can calculate counterfactual arrests in year y. The remaining counterfactual stocks and flows for year y are then calculated just as in the simpler simulation procedure described above.

D Founding Dates for State Sentencing Commissions

This appendix describes how we found the establishment dates for sentencing commissions in various states for Table 5. We first examined the website of the National Association of Sentencing Commissions, which contains links to the websites of individual states' sentencing commissions (http://thenasc.org/aboutnasc.html). We then augmented the establishment dates collected from Frase (2005) by adding information from more recent years and from some states which Frase did not include. That information was collected from the following state websites:

Alabama:

http://sentencingcommission.alacourt.gov/about.html

Connecticut:

http://www.cga.ct.gov/asp/cgabillstatus/cgabillstatus.asp?selBillType=Bill&bill_num=5248&which_year=2010&SUBMIT1.x=8&SUBMIT1.y=13

Illinois:

http://www.icjia.state.il.us/spac/pdf/2010%20SPAC%20Annual%20Report.pdf

Louisiana:

http://www.lcle.la.gov/sentencing_commission/022410.pdf

Massachusetts:

http://www.mass.gov/courts/admin/sentcomm/background.html

New Mexico:

http://www.nmlegis.gov/lcs/lcsdocs/statutoryexecutive02.pdf

New York:

http://www.courts.state.ny.us/ip/sentencing/

Wisconsin:

http://docs.legis.wisconsin.gov/2003/statutes/statutes/973/30/1/j

Table A1

Employment and Institutionalization Rates
Black Females

Birth Year	Age					
	20-24	25-29	30-34	35-39	40-44	45-49
1910-1914						0.541
						0.013
1915-1919					0.535	
					0.008	
1920-1924				0.511		0.558
				0.008		0.007
1925-1929			0.457		0.571	
			0.008		0.005	
1930-1934		0.409		0.575		0.620
		0.008		0.007		0.004
1935-1939	0.381		0.548		0.653	
	0.006		0.005		0.003	
1940-1944		0.543		0.662		0.703
		0.006		0.004		0.005
1945-1949	0.476		0.659		0.733	
	0.006		0.005		0.005	
1950-1954		0.606		0.711		0.671
		0.005		0.008		0.007
1955-1959	0.465		0.668		0.677	
	0.004		0.009		0.009	
1960-1964		0.626		0.678		0.694
		0.011		0.011		0.006
1965-1969	0.530		0.680		0.711	
	0.008		0.011		0.007	
1970-1974		0.665		0.716		
		0.009		0.006		
1975-1979	0.572		0.691			
	0.007		0.007			
1980-1984		0.652				
		0.008				
1985-1989	0.529					
	0.005					

The top number is the employment rate and the bottom number is the institutionalization rate. This table was created using 1960-2000 census data and the 2010 ACS from IPUMS. The IPUMS website is available at http://usa.ipums.org/usa. The IPUMS variables for employment and institutionalization were EMPSTAT (=1) and GQTYPE (=1 for 1990 and later; =2, 3, 4 for 1980 and earlier). After the 1980 census, the data no longer distinguish between different types of institutionalization. Thus, to be consistent we calculate total institutionalization rates for all years. All samples are weighted by the IPUMS variable PERWT.

Table A2

Employment and Institutionalization Rates
White Females

Birth Year	Age					
	20-24	25-29	30-34	35-39	40-44	45-49
1910-1914						0.444
						0.008
1915-1919					0.414	
					0.006	
1920-1924				0.359		0.506
				0.005		0.005
1925-1929			0.312		0.489	
			0.004		0.004	
1930-1934		0.322		0.443		0.590
		0.004		0.003		0.003
1935-1939	0.439		0.399		0.615	
	0.006		0.003		0.003	
1940-1944		0.432		0.598		0.718
		0.003		0.002		0.003
1945-1949	0.529		0.589		0.745	
	0.004		0.002		0.002	
1950-1954		0.630		0.723		0.743
		0.002		0.002		0.002
1955-1959	0.627		0.697		0.731	
	0.002		0.002		0.002	
1960-1964		0.711		0.702		0.717
		0.003		0.003		0.002
1965-1969	0.682		0.687		0.708	
	0.002		0.003		0.003	
1970-1974		0.705		0.693		
		0.002		0.003		
1975-1979	0.669		0.695			
	0.002		0.003			
1980-1984		0.706				
		0.004				
1985-1989	0.639					
	0.003					

See notes for Table A1.

Table A3

Employment and Institutionalization Rates
Black Females With Less Than HS

Birth Year	Age					
	20-24	25-29	30-34	35-39	40-44	45-49
1910-1914						0.514
						0.013
1915-1919					0.500	
					0.007	
1920-1924				0.478		0.498
				0.010		0.008
1925-1929			0.409		0.510	
			0.008		0.005	
1930-1934		0.358		0.497		0.509
		0.010		0.010		0.006
1935-1939	0.315		0.459		0.523	
	0.007		0.006		0.005	
1940-1944		0.426		0.511		0.506
		0.010		0.007		0.008
1945-1949	0.340		0.460		0.498	
	0.011		0.011		0.011	
1950-1954		0.380		0.439		0.403
		0.013		0.020		0.015
1955-1959	0.283		0.377		0.402	
	0.011		0.025		0.026	
1960-1964		0.289		0.397		0.379
		0.030		0.036		0.026
1965-1969	0.231		0.381		0.392	
	0.030		0.050		0.031	
1970-1974		0.356		0.414		
		0.038		0.030		
1975-1979	0.356		0.431			
	0.025		0.031			
1980-1984		0.325				
		0.034				
1985-1989	0.263					
	0.020					

See notes for Table A1. This table was created using only data for Black females whose value for the IPUMS variable EDUCD was either less than or equal to 50 (completed less than 12 years of schooling). GED recipients are not included in this sample.

Table A4

Employment and Institutionalization Rates
White Females With Less Than HS

Birth Year	Age					
	20-24	25-29	30-34	35-39	40-44	45-49
1910-1914						0.388
						0.011
1915-1919					0.372	
					0.009	
1920-1924				0.332		0.442
				0.008		0.009
1925-1929			0.288		0.443	
			0.007		0.008	
1930-1934		0.262		0.411		0.469
		0.008		0.006		0.006
1935-1939	0.286		0.360		0.480	
	0.011		0.008		0.007	
1940-1944		0.318		0.471		0.492
		0.008		0.007		0.009
1945-1949	0.332		0.429		0.499	
	0.012		0.009		0.009	
1950-1954		0.392		0.484		0.437
		0.010		0.011		0.008
1955-1959	0.386		0.462		0.461	
	0.009		0.010		0.008	
1960-1964		0.417		0.447		0.465
		0.011		0.009		0.007
1965-1969	0.382		0.408		0.478	
	0.009		0.009		0.007	
1970-1974		0.383		0.446		
		0.007		0.010		
1975-1979	0.406		0.424			
	0.007		0.009			
1980-1984		0.373				
		0.015				
1985-1989	0.381					
	0.015					

See notes for Table A3.

Table A5 - Internal and External Consistency Checks

	Internal Consistency Checks (NCRP)		External Consistency Checks (NCRP vs. NPS)	
	Admissions vs. Releases	Flow vs. Stock	NCRP Flow vs. Change in NPS stock	NCRP vs. NPS Flow
AL	X	-	X	X
AK	X	X	X	X
CA	**√**	**(√)**	**√**	**√**
CO	**(√)**	**(√)**	**(√)**	**X**
FL	(√)	X	X	X
GA	(√)	X	√	X
HI	X	-	X	X
IL	**√**	**-**	**√**	**√**
IA	X	X	X	X
KY	(√)	-	X	X
MD	X	X	X	X
MI	**√**	**√**	**√**	**√**
MN	√	(√)	X	(√)
MS	(√)	-	X	X
MO	X	X	X	X
NE	**√**	**-**	**X**	**X**
NV	X	-	X	X
NH	√	-	X	X
NJ	**√**	**√**	**√**	**(√)**
NY	**√**	**√**	**√**	**√**
NC	X	(√)	√	X
ND	√	-	(√)	X
OH	√	-	(√)	X
OK	X	X	X	X
OR	X	√	(√)	X
PA	X	X	X	(√)
SC	**√**	**(√)**	**√**	**X**
TN	X	X	X	X
TX	X	(√)	X	X
UT	**√**	**-**	**√**	**√**
VA	√	-	(√)	X
WA	**√**	**(√)**	**√**	**X**
WV	X	-	X	X
WI	**√**	**(√)**	**(√)**	**X**

Notes:

Detailed descriptions of the external and internal checks are available in Appendix A.

"√" indicates that the state passes the test.

"(√)" indicates minor problems in only a few years.

"X" indicates major inconsistency problems in many years.

"-" indicates that the test could not be performed since the state did not report any stock data.

Table A6

Number of Persons per 1000 Arrests Who Serve Prison Terms of t

Whites

		0-1 years	1-2 years	2-3 years	3-4 years	4-5 years	5+ years
Violent Crime							
Murder & Homicide	1985	48.31	64.81	45.61	39.31	21.30	255.04
	2000	44.62	52.20	44.62	34.80	31.68	575.12
	Ratio	**0.92**	**0.81**	**0.98**	**0.89**	**1.49**	**2.26**
Forcible Rape	1985	11.52	27.76	28.11	27.76	11.35	41.20
	2000	13.73	15.12	22.68	14.32	15.91	90.11
	Ratio	**1.19**	**0.54**	**0.81**	**0.52**	**1.40**	**2.19**
Robbery	1985	30.46	43.76	26.73	17.33	9.23	18.63
	2000	39.26	41.38	26.73	17.16	13.05	57.84
	Ratio	**1.29**	**0.95**	**1.00**	**0.99**	**1.41**	**3.10**
Aggravated Assault	1985	8.17	10.24	4.91	2.28	0.73	1.69
	2000	10.31	9.28	3.89	2.62	1.87	5.76
	Ratio	**1.26**	**0.91**	**0.79**	**1.15**	**2.55**	**3.40**
Other Assault	1985	1.05	0.95	0.21	0.11	0.05	0.06
	2000	3.09	2.88	0.71	0.43	0.27	0.56
	Ratio	**2.93**	**3.04**	**3.37**	**3.82**	**5.58**	**9.84**
Property Crime							
Burglary	1985	23.63	15.43	6.62	2.96	1.24	2.38
	2000	36.29	21.41	12.01	5.39	3.07	8.24
	Ratio	**1.54**	**1.39**	**1.81**	**1.82**	**2.48**	**3.46**
Motor Vehicle Theft	1985	11.81	4.70	1.35	0.43	0.16	0.49
	2000	40.61	16.80	5.10	1.91	0.80	1.99
	Ratio	**3.44**	**3.57**	**3.78**	**4.47**	**4.84**	**4.03**
Larceny/Theft	1985	4.50	1.97	0.56	0.27	0.08	0.25
	2000	11.46	4.90	1.84	0.60	0.38	0.59
	Ratio	**2.55**	**2.50**	**3.31**	**2.22**	**4.48**	**2.38**
Other Property Crime	1985	2.00	1.51	0.79	0.46	0.18	0.23
	2000	2.79	1.95	0.80	0.53	0.31	0.72
	Ratio	**1.40**	**1.29**	**1.02**	**1.16**	**1.74**	**3.15**
Drug-Related Crime							
Drug Trafficking	1985	29.64	32.15	7.75	1.80	0.77	2.48
	2000	63.79	64.43	25.72	11.10	5.70	8.42
	Ratio	**2.15**	**2.00**	**3.32**	**6.18**	**7.44**	**3.40**
Drug Possession/Use	1985	5.28	1.59	0.22	0.07	0.03	0.36
	2000	17.42	5.46	1.64	0.53	0.31	0.56
	Ratio	**3.30**	**3.44**	**7.44**	**7.25**	**10.03**	**1.57**
Other							
Other Sex Crime	1985	9.37	16.54	13.77	11.06	5.73	17.87
	2000	19.40	22.91	24.62	11.97	16.74	61.48
	Ratio	**2.07**	**1.38**	**1.79**	**1.08**	**2.92**	**3.44**
White Collar Crime	1985	13.32	5.79	1.44	0.71	0.15	0.34
	2000	18.82	7.46	2.87	1.10	0.45	0.60
	Ratio	**1.41**	**1.29**	**1.99**	**1.54**	**3.03**	**1.77**
Other Crime	1985	1.30	0.46	0.12	0.05	0.03	0.10
	2000	2.64	1.46	0.54	0.26	0.13	0.30
	Ratio	**2.02**	**3.15**	**4.41**	**4.83**	**4.78**	**3.10**

See notes to Table 8.

Table A7

Number of Persons per 1000 Arrests Who Serve Prison Terms of t

Blacks

Violent Crime		0-1 years	1-2 years	2-3 years	3-4 years	4-5 years	5+ years
Murder & Homicide	1985	25.93	45.74	45.38	30.25	25.21	216.46
	2000	18.75	22.50	15.00	14.58	15.00	358.75
	Ratio	**0.72**	**0.49**	**0.33**	**0.48**	**0.59**	**1.66**
Forcible Rape	1985	5.73	14.42	16.54	12.30	8.48	35.41
	2000	5.77	7.42	14.84	9.48	11.13	58.13
	Ratio	**1.01**	**0.51**	**0.90**	**0.77**	**1.31**	**1.64**
Robbery	1985	24.05	33.41	20.16	13.12	8.20	21.62
	2000	29.10	32.82	22.35	17.62	14.25	81.09
	Ratio	**1.21**	**0.98**	**1.11**	**1.34**	**1.74**	**3.75**
Aggravated Assault	1985	12.88	13.34	7.06	2.98	1.86	4.81
	2000	15.37	11.05	5.84	4.74	2.41	8.93
	Ratio	**1.19**	**0.83**	**0.83**	**1.59**	**1.29**	**1.86**
Other Assault	1985	1.69	1.33	0.56	0.18	0.18	0.33
	2000	4.12	3.17	1.34	0.54	0.48	0.91
	Ratio	**2.44**	**2.38**	**2.42**	**3.04**	**2.71**	**2.71**
Property Crime							
Burglary	1985	37.81	20.79	9.46	4.05	2.30	5.49
	2000	52.76	29.58	18.53	7.78	5.99	13.82
	Ratio	**1.40**	**1.42**	**1.96**	**1.92**	**2.60**	**2.52**
Motor Vehicle Theft	1985	17.09	6.39	1.71	0.52	0.15	0.82
	2000	45.24	22.16	6.84	1.53	1.53	1.43
	Ratio	**2.65**	**3.47**	**4.00**	**2.94**	**10.30**	**1.75**
Larceny/Theft	1985	12.44	5.00	1.61	0.79	0.30	0.75
	2000	16.38	7.41	2.69	1.36	0.66	1.01
	Ratio	**1.32**	**1.48**	**1.67**	**1.72**	**2.20**	**1.34**
Other Property Crime	1985	5.15	2.47	1.76	0.99	0.42	0.70
	2000	5.69	4.18	1.80	0.72	0.58	1.73
	Ratio	**1.11**	**1.69**	**1.02**	**0.73**	**1.36**	**2.45**
Drug-Related Crime							
Drug Trafficking	1985	30.26	26.84	6.69	2.46	1.91	5.05
	2000	60.02	53.14	28.06	12.75	7.13	10.60
	Ratio	**1.98**	**1.98**	**4.20**	**5.19**	**3.73**	**2.10**
Drug Possession/Use	1985	14.02	3.68	1.10	0.55	0.19	0.85
	2000	35.51	11.70	4.64	2.01	1.25	1.86
	Ratio	**2.53**	**3.18**	**4.20**	**3.64**	**6.54**	**2.18**
Other							
Other Sex Crime	1985	11.97	20.90	14.83	10.96	7.25	27.81
	2000	31.23	26.30	23.63	15.41	19.93	66.78
	Ratio	**2.61**	**1.26**	**1.59**	**1.41**	**2.75**	**2.40**
White Collar Crime	1985	18.79	6.32	2.46	0.66	0.41	0.53
	2000	31.56	9.59	3.63	1.22	0.78	0.88
	Ratio	**1.68**	**1.52**	**1.47**	**1.86**	**1.90**	**1.65**
Other Crime	1985	3.84	0.97	0.34	0.19	0.08	0.39
	2000	4.90	2.14	0.93	0.47	0.33	0.72
	Ratio	**1.28**	**2.20**	**2.76**	**2.56**	**4.05**	**1.86**

See notes to Table 8.