6-18-2024

Cleaning the Dirty Pool: Testing Interaction Effects Using Different Panel Model Specifications

Rhiannon Berry

Portland State University

Cleaning The Dirty Pool: Testing Interaction Effects Using Different Panel Model Specifications

by

Rhiannon Star Berry

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science
in
Criminology and Criminal Justice

Thesis Committee:
Mark Leymon, Chair
Kris Henning
Arynn Infante

Portland State University
2024

Abstract

This exploratory study seeks to uncover the most effective approaches for constructing interaction terms within panel models. With no preconceived hypothesis in mind, the primary aim is to discern which modeling configuration yields the most robust results, laying the foundation for future research in statistical modeling. Using a large data set of sentencing reforms passed between the mid-1970s and mid-2000s, this study systematically assesses interaction terms and determines the most appropriate modeling. Different specifications of sentencing reforms at the state level within different modeling specifications will be explored to highlight which models are most appropriate in predicting imprisonment rates. By systematically examining a wide range of interaction terms and models, this research offers valuable insights into the methodology of statistical panel modeling, contributing to the advancement of statistical analysis in this field.

Table of Contents

List of Tables

List of Figures

Introduction

Criminology research often involves the analysis of panel data, sometimes referred to as cross-sectional time series data, to investigate criminal behaviors, the effectiveness of interventions, and the dynamics of the criminal justice system, among other aspects of the system. Panel data models provide a robust framework to address the complexities of criminal justice research questions (Halaby, 2004). The structural characteristics of panel data align effectively with non-randomized observational data frequently encountered in criminology (Halaby, 2004). Despite the considerable utility of panel data models in criminology research, there remains a need for more consensus regarding the optimal specifications for employing these models and how to properly specify the variables, especially related to interaction terms (Harmon, 2011). Different applications and specifications may yield very different results. This discrepancy highlights the importance of comparing different specifications, making sound statistical and theoretical decisions regarding model choices, and correctly specifying the measure.

In panel modeling, interaction terms are useful for investigating how the relationship between two variables changes or varies under different conditions or across different groups. They help researchers assess whether the effect of one variable is influenced by the characteristics of another variable. Interaction terms have considerable utility and can be employed in various statistical models (Jaccard & Turrisi, 2003). In panel models, interaction terms can be used with fixed or random effects but may produce very different results. Fixed effects models focus on capturing individual-specific characteristics, providing an in-depth understanding of how specific factors influence the studied phenomenon (Halaby, 2004).

On the other hand, random effects models consider individual characteristics as random variables and are often used when these characteristics cannot be fully accounted for (Halaby, 2004). The choice between these two approaches can significantly affect the results and conclusions drawn from the research. Additionally, some random effects models may not be appropriate due to a statistical phenomenon called omitted variable bias (OVB). Additional model specifications are necessary and may impact how interaction terms function in the models.

Panel models have been prevalent in understanding the impacts of policy changes, particularly within the realm of criminal justice. These models allow researchers to capture the dynamics and interactions of various factors over time, providing a deeper insight into complex relationships. However, the construction and inclusion of interaction terms within these models remain a point of disagreement among researchers.

The main objective of this study is to systematically contrast a range of panel models, each with distinct specifications, to identify the model that offers superior accuracy and statistical power. This is particularly crucial when dealing with multifaceted issues such as state-level sentencing reforms and their effects on imprisonment rates (Hsiao, 2003). Through this approach, I aim to discern which interaction terms and model settings configurations most effectively capture the nuances of these reform’s impacts. It provides valuable insights into how these models can be effectively applied in real-world scenarios, enhancing understanding of the methodology behind statistical analysis. The findings from this study contribute to the advancement of statistical analysis within the field of panel modeling and equip researchers and policymakers with the tools to investigate complex relationships more comprehensively. Ultimately, this study aims to expand researchers' capability to explore and interpret policies' complex interplay and outcomes by providing a clearer understanding of these methodologies and their practical applications. This is fundamental for advancing theoretical knowledge and practical applications in policy analysis.

Literature Review

The U.S. criminal justice system is a complex network, encompassing 50 states with multiple jurisdictions, ranging from federal to states to counties to municipalities, with overlap and interactions between and across levels. This complexity posed a significant challenge when researching the law-and-order sentencing reforms. Sentencing reforms have been a central focus in the criminal justice system for decades. In 1980, Minnesota became the first jurisdiction to adopt sentencing guidelines. These guidelines aimed to enhance the consistency and structure of sentencing practices, and by late 2004, about eighteen states and the District of Columbia had implemented similar systems, with other states considering similar reforms (Frase, 2005). These reforms are purported as a tool to improve the fairness and effectiveness of sentencing procedures, reduce disparities, and address the issue of mass incarceration (Brennan & Spohn, 2008). The diversity of sentencing reforms is notable, encompassing changes in sentencing guidelines and mandatory minimums (Boppre & Harmon, 2017).

There was a shift in sentencing structure from indeterminate to determinate sentencing, and from 1935 to 1975, the United States justice system operated under an indeterminate sentencing model. This approach focused on rehabilitation, providing offenders with vocational or educational programs, and was characterized by a high degree of discretion afforded to judges and parole boards (Bureau of Justice Assistance, 1996; Tonry, 1996; 2016). Sentences were tailored to the individual's rehabilitative needs, and parole release was contingent on evidence of rehabilitation. However, the model drew significant criticism for its inconsistency and perceived leniency. This dissatisfaction, driven by political debates and sensationalized media coverage, set the stage for a fundamental shift in sentencing philosophy towards more strict policies. However, it’s important to note that some individuals felt that reforms taking away judicial discretion would be more progressive due to perceived biases in sentencing decisions (Tonry, 2016).

In response, from the 1970s to the early 2000s, several key reforms were enacted across various states, marking a shift towards a more punitive model of justice (Spohn, 2000). This study examined these six key sentencing reforms: presumptive sentencing guidelines, voluntary sentencing guidelines, statutory presumptive sentencing, determinate sentencing, truth in sentencing, and three strikes laws. Presumptive sentencing guidelines provide a framework that judges are expected to follow, offering less discretion than the indeterminate model. Unlike the presumptive guidelines, voluntary sentencing guidelines are not mandatory but serve as recommendations to help standardize sentencing across courts. Statutory presumptive sentencing involves statutory mandates establishing presumed sentences for specific offenses, significantly constraining judicial discretion. Determinate sentencing eliminates parole boards and sets fixed prison terms, which offenders must serve before release. Truth in sentencing laws requires offenders to serve a substantial portion of their sentence, often 85% or more, before being eligible for parole. Three strikes laws increase the prison sentences of persons convicted of a felony previously convicted of two or more violent crimes or serious felonies (Frase, 2005a).

The first three reforms, presumptive sentencing guidelines, voluntary sentencing guidelines, and statutory presumptive sentencing, are front-end reforms. They directly influence the initial sentencing decision and are mutually exclusive; a state typically adopts one of these models or maintains an indeterminate sentencing approach. This exclusivity means that each state chooses a single framework to standardize how sentences are determined at the outset of the sentencing process. In contrast, the latter three reforms, determinate sentencing, truth in sentencing laws, and three strikes laws, are back-end reforms. These focus primarily on the execution and completion of the sentence after it has been decided. They can coexist with each other and any front-end reforms, allowing for multiple combinations and layers of sentencing rules that can apply to a single offender (Boppre & Harmon, 2017).

This variety of possible combinations and the way front-end and back-end reforms can interact illustrate the complexity of the sentencing reform landscape in the United States. Each state's choices about which reforms to implement can lead to a wide range of sentencing outcomes and administrative practices, reflecting differing priorities such as reducing discretion, increasing transparency, and ensuring the certainty and severity of sentences (Boppre & Harmon, 2017). As you can see, the interactions between these reforms can be quite complex. These variations among the reforms can make comparisons and analyses challenging. Research on the impact of sentencing reforms on imprisonment involves studying many variables, such as the timing of reforms, types of reforms, demographic characteristics, and variations across different jurisdictions.

This complexity stems from the need to understand how diverse factors contribute to changes in imprisonment rates (Boppre & Harmon, 2017). Each state in the U.S. has the authority to establish its unique sentencing policies, leading to significant variations in how they are implemented and their subsequent effects on imprisonment. Interaction terms in panel models enable researchers to explore the differences between states and assess how the relationship between sentencing reforms and imprisonment rates varies across different jurisdictions (Hsiao, 2003).

In the past, some sentencing research primarily focused on the macro-level effects of reforms within single jurisdictions (Koons-Witt, 2002; Feinman, 1994). Koons-Witt's (2002) study aimed to investigate the impact of presumptive sentencing guidelines in Minnesota on female imprisonment. Notably, Koons-Witt's study is the only one to date that has examined longitudinal differences in imprisonment across race and gender. In contrast, Feinman's (1994) research focused on the effects of mandatory imprisonment for drug offenses and second felony convictions in Florida. Koons-Witt (2002) and Feinman (1994) both found that Black women experienced a disproportionately higher increase in imprisonment compared to White women despite the implementation of sentencing guidelines and mandatory minimum sentences in Minnesota and Florida, respectively. Limiting research to a single state hinders the ability to compare data across different jurisdictions over time, reducing the opportunity for meaningful comparisons and weakening the generalizability of findings. However, more recent research has expanded its scope, examining the outcomes of multiple reforms spanning various jurisdictions over periods of time (Boppre & Harmon, 2017). This shift in research focus is evident in works by Harmon (2011; 2013), Boppre and Harmon (2017), and Stemen (2006), which get into this broader approach.

The reforms themselves exhibited significant variability in adoption and combination across states. Researchers are tasked with the challenge of effectively capturing this complexity, which may necessitate an approach that avoids oversimplification and recognizes the interdependence of these reform measures (Harmon, 2017). Some previous macro-level research failed to model the interactions of reforms and instead modeled them as independent of each other. Therefore, they have chosen more simplified ways to model interactions over complex ones. In this context, I briefly discuss recent studies that involve 40 or more states.

Testing Interaction Effects Using Different Panel Model Specifications

Stemen's (2011) study used a pooled time series design to investigate the interactions between structured sentencing, determinate sentencing, and state incarceration rates from 1978 to 2004. The results revealed that limiting release discretion through determinate sentencing is more influential than restricting sentencing discretion through structured sentencing. In line with earlier research, determinate sentencing was associated with lower incarceration rates, regardless of other policies present. In contrast to previous findings, presumptive sentencing guidelines were linked to reduced incarceration rates only when combined with determinate sentencing. This could also be due to the study's use of first-level interaction terms only, a more simplified approach to modeling variable interaction in panel models.

Boppre and Harmon's 2017 study investigated the effects of sentencing reforms on racial disparities in female imprisonment and time served across U.S. states from 1983 to 2008, with a specific emphasis on Black and White women. The final analytical model they employed features fixed effects, robust standard errors, and other crucial components, rendering it both methodologically robust and facilitating a more insightful interpretation of their findings. Other important techniques used include third-level interaction terms. Third-level interaction terms involved the simultaneous examination of three distinct reform-related variables to understand their combined influence on incarceration rates. Other studies, such as Harmon's (2013), utilized third-level interaction terms.

Testing Interaction Effects Using Different Panel Model Specifications

Harmon's (2013) study used a dependent variable of the total prison population per 100,000 people observed from 1967 to 2007 across all 50 states. The statistical modeling employed certain advanced techniques, including third-level interaction terms between reforms on imprisonment. Additionally, the study incorporated fixed-effects for states, accounting for unobserved state-level effects and enhancing the validity of the results.

The challenge of incorporating sentencing reforms into the analysis is further compounded by the necessity to consider various model specifications, which include the choice between fixed effects and random effects within regime score models or fully saturated models. Additional considerations like the impacts of trending data or stationarity can also influence how a model is constructed and how valid, consistent, and efficient the results are (Halaby, 2004). In several studies employing panel data analysis, the methods section often needs a clearer indication of whether fixed or random effects are employed. Notable examples include Anderson et al. (2011), Steffensmeier and Demuth (2006), and Steffensmeier and Haynie (2000). The selection between fixed and random effects models is critical, as these statistical approaches are instrumental in addressing unobservable factors that may influence observed relationships. Particularly, they play a crucial role in mitigating the impact of omitted variable bias. Omitted variable bias arises when a relevant variable is excluded from a regression model, leading to incorrect attributions of effects and biased estimates. In such cases, the model mistakenly assigns the influence of the omitted variable to the included variables, confounding the true relationship with the dependent variable (Halaby, 2004).

Both FE and RE models offer distinct advantages. Random effects accommodate unobserved time-invariant heterogeneity across individual units in panel data, capturing latent factors that persist consistently over time. This helps alleviate potential biases stemming from unobserved individual characteristics influencing the dependent variable. (Halaby, 2004) On the other hand, fixed effects control for time-invariant characteristics specific to each unit, enabling the management of unobserved time-invariant factors critical for addressing biases related to individual heterogeneity (Halaby, 2004; Hsiao, 2003). The choice between these approaches bears substantial implications for the validity and generalizability of study findings, emphasizing the necessity of explicitly stating the chosen method in the methodology, as emphasized by Harmon (2011).

For this analysis, the choice of FE or RE is important because it can impact how the interaction terms operate in the model and could lead to very different results. In addition to the impacts FE and RE can have on the modeling of variable interactions, in this case, interactions between sentencing reforms, the choice of how to model the reforms is an important consideration. Two ways to model interactions are fully saturated models and regime scores.

Both approaches offer unique advantages. The fully saturated models explicitly consider individual-specific characteristics, capturing heterogeneity across entities like states, regions, or countries. This approach facilitates a comprehensive examination of how the effects of sentencing reforms on imprisonment rates may vary across diverse contexts (Jaccard & Turrisi, 2003). Regime scores allow modeling shifts in the criminal justice system's policies and practices over time. This sheds light on the impact of sentencing reforms on imprisonment rates during different periods or regimes (Jaccard & Turrisi, 2003). This nuanced analysis is particularly valuable given the varied implementation times of reforms across states. Figures 1 and 2 show how the reforms are modeled distinctively between regime score and full interaction term models. Figure 1 illustrates how specific combinations of sentencing reforms are grouped into distinct regimes over time. Each row represents a unique regime, a particular set of reforms implemented together during the observed period. For instance, “VSG/DS” represents a regime where both policies are in effect simultaneously. It creates a particular legal landscape that the model treats as a single entity or regime. Figure 2 represents a full interaction model, and unlike the regime scores, this model does not group policies but allows for the analysis of every possible interaction between them. The model’s complexity allows for a nuanced understanding of how the policies might reinforce, mitigate, or independently influence the imprisonment rates. It assumes that policy impacts are cumulative and evolve over time. All these different model specification choices have both advantages and disadvantages.

Figure 1. Regime Scores for Florida

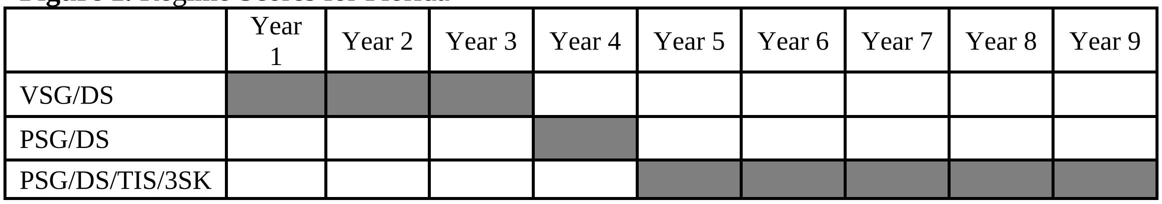

Figure 2. Full Interaction Model for Florida

Note: VSG: voluntary sentencing guidelines, PSG: presumptive sentencing guidelines, DS: Determinate sentencing, TIS: truth in sentencing, 3SK: three strikes laws

The primary aim of the current study is to analyze different ways interaction terms can be constructed and included in panel models. The best model will accurately represent the real world. This study uses state-level sentencing reforms and their relationship with imprisonment rates as a case study. It explores various modeling strategies, such as fixed effects and random effects configurations, incorporating the standard model with no interactions and other models with different configurations of the interactions. Ultimately, this contributes to the advancement of statistical analysis within panel modeling, expanding the ability for researchers to investigate complex relationships and offering a robust statistical basis to inform policy decisions. This comprehensive approach will shed light on the direct outcomes of sentencing reforms.

Data

Dependent variable

Prior to 1972, imprisonment rates remained relatively consistent over time. However, there has been a consistent and rapid increase since that year. The analysis includes a dependent variable representing the total prison population per 100,000 to estimate the potential impacts of sentencing reforms on imprisonment. This variable was assessed at the state level across all 50 states and observed from 1967 to 2007, resulting in a total of 2050 potential observations. Data for the dependent variable were obtained from the Bureau of Justice Statistics (U.S., 1965-1983, 1984-1998, 1999-2008). The measures utilized in this study assessed state-level data for general imprisonment across each year from 1978 to 2008. The start year of 1978 is significant as it marks the first instance where the racial composition of state imprisonment populations was systematically reported. The analysis concludes with data from 2008, which aligns with the last year this data was released. The Bureau of Justice Statistics (BJS) data aggregates imprisonment figures for both male and female populations into a single measure, further broken down by race. This comprehensive approach allows for an inclusive analysis of trends across different demographic groups, offering insights into racial disparities and the overall dynamics of state-level imprisonment over three decades.

The combination of time series and cross-sectional data significantly enhances the statistical power of the models (Hsiao, 2003; Wooldridge, 2002). Time series data track changes in the dependent variable over time, allowing for identifying trends and patterns.

Cross-sectional data provide information across different states, enabling comparisons and examining contextual factors. This combined approach enriches the analysis by offering a comprehensive understanding of the dynamics driving imprisonment rates and the potential effects of policy interventions. In cross-sectional analyses, it is hard to determine causality because variables are measured at one-time point. Panel analyses, however, track variables over time within the same units, allowing for a better understanding of causality and reciprocal relationships. Panel data also helps control for external variables and test for spurious associations, offering more flexibility in estimating measurement errors compared to cross-sectional data. Overall, like experimental methods, panel designs enable more rigorous assessments of causal relations (Finkel, 1995).

Independent variables (sentencing reforms)

Table 1 outlines six variables related to sentencing reform, sourced from various outlets such as the Bureau of Justice Assistance (1996), Zhang, Maxwell, and Vaughn (2009), and the Vera Institute (Stemen et al., 2006). Although there is a broad consensus on the overall aim of these reforms, substantial diversity in design and implementation has led to a need for more agreement on how they should be categorized. The study adopts the perspective of researchers who emphasize the essential legal nuances among these reforms, and these features are discussed below.

Table 1. Types of U.S. Sentencing Reforms

Reform	Description
Presumptive Sentencing Guidelines (PSG)[1]	Consists of a matrix of possible sentences with a narrower range within a sentencing category defined by an offender's criminal history (prior offenses) and offense severity, then indeterminate sentencing. Judges generally must follow the matrix as the sentence is enforced through appellate review.
Voluntary Sentencing Guidelines (VSG)[1]	Treat guidelines as formal recommendations but do not *legally mandate* the judge follow them. While judges generally follow them, an offender may not appeal deviations from the matrix.
Statutory Presumptive Sentencing (SPS)[1]	Represent an attempt to create uniformity within similarly situated crimes but acts less like a sentencing rubric. It specifies an appropriate or "normal" sentence for each offense as a baseline for a judge.
Determinate Sentencing (DS)[2]	Refers to a system *without* discretionary parole boards.
Truth in Sentencing (TIS)[2]	Requires offenders to serve a statutorily defined minimum amount of time. Only states meeting the 1994 Federal Omnibus Crime Bill minimum of 85% time-served of the original sentence are considered.
Three Strikes Laws (3STKS)[3]	A habitual offender law focused on three-time felony offenders. Generally, the law suggests a severe sentence for a third felony offense.

[1] Front-End Reforms [2] Back-End Reforms [3]Sentencing Enhancement

Scholars like Frase (2005a; 2005b), Marvell (1995), Stemen (2007), and Tonry (1995) highlight significant variations in their design and application and a need for more consensus on optimal grouping methods. This analysis, therefore, adopted a categorization system that emphasizes key legal distinctions among the reforms. It's important to note that there is disagreement within sentencing reform literature about categorization, such as placing reforms on a continuum from least restrictive to most.

The first front-end reforms analyzed were sentencing guidelines intended to standardize sentencing outcomes by reducing judicial discretion. These guidelines use a matrix system that assigns sentences based on the severity of the offense and the offender's criminal history. This system limits judges' discretion and often requires them to provide written justifications for deviations from the guidelines, ensuring a more predetermined and consistent sentencing process (Frase, 2005a).

These guidelines were categorized into two broad legal categories: presumptive sentencing guidelines and voluntary sentencing guidelines. Presumptive sentencing guidelines were strictly enforced through appellate review, meaning that a higher court can overturn a sentence that does not adhere to the guidelines, thereby strongly regulating judicial decisions. In contrast, voluntary sentencing guidelines were treated as formal recommendations without mandatory enforcement through appellate review. Judges are typically asked to justify any deviations in writing, but they keep considerable freedom in their sentencing decisions (Stemen et al., 2006). Studies, such as one by Miethe & Moore (1988), suggest that judges adhere to these guidelines approximately 85% of the time.

This distinction between presumptive and voluntary guidelines is independent of how the guidelines were initially developed. The classification of sentencing guidelines into presumptive and voluntary categories follows a methodology supported by Frase (1995) and Stemen et al. (2006). It represents a common approach within the literature despite needing to be universally accepted. This framework allows for a structured exploration of how legal nuances influence the application and effectiveness of sentencing reforms in the criminal justice system. The third front-end reform included in the analysis is statutory presumptive sentencing, which establishes baseline sentences for specific offenses. It is distinct from sentencing guidelines because they don't explicitly set up a grid defined by offense type and criminal history.

In contrast, determinate sentencing, truth in sentencing laws, and three strikes laws concentrate on the back-end mechanism of release, deviating from a rehabilitative-indeterminate model. Indeterminate sentencing is a legal approach where judges have discretion in determining the length of a criminal sentence within a range established by statute or sentencing guidelines. The actual duration of the sentence is often influenced by factors such as the offender's behavior, rehabilitation progress, and potential for reintegration into society rather than being strictly defined at the time of sentencing. Determinate sentencing refers to a system without discretionary parole boards, while truth in sentencing requires offenders to serve a minimum specified time. These reforms shift sentencing away from a more rehabilitative indeterminate model. Additionally, the three strikes laws, which impose severe sentences for repeat felony offenders, are considered additional reforms beyond the previous ones (Stemen et al. 2006).

Front-end reforms are mutually exclusive, while back-end reforms can coexist. To accommodate the lagged effect and capture an anticipated growth curve, the study adopts a logarithmic measure for each reform, departing from previous analyses heavily reliant on dummy variables. This approach provides a nuanced understanding of the sentencing reforms' impact over five years (Harmon, 2011).

Table 2 outlines the distribution of sentencing reforms across the United States as of 2008, revealing the prevalence of various reforms in different states. The counts for each reform are as follows: Presumptive Sentencing Guidelines (PSG) are present in 10 states, Voluntary Sentencing Guidelines (VSG) in 11 states, Statutory Presumptive Sentencing (SPS) in 8 states, Determinate Sentencing (DS) in 18 states, Truth in Sentencing (TIS) in 24 states, and Three Strikes Laws (3STK) in 24 states. These figures illustrate the varying adoption rates of different sentencing reforms nationwide. Transitioning between these reforms is also evident, as certain states have shifted from one sentencing approach to another over time. For example, California, Colorado, and Ohio adopted determinate sentencing in 1976, 1979, and 1996, respectively, while other states like Florida transitioned from voluntary sentencing guidelines (1983-93) to statutory presumptive sentencing (1994). These transitions reflect changes in legal frameworks and societal attitudes toward crime and punishment over the years.

Control variables

The control variables were split into three categories, and the descriptives are outlined in Table 3. The first category, crime controls, was quantified through the incidence of arrests for violent crimes and drug-related offenses. Data for this category were sourced from the FBI's Uniform Crime Report (UCR) spanning the years 1965 to

2008, subsequently transformed into state-specific rates utilizing census data from the same period (U.S. 1965-2008).

Table 2. Distribution of Sentencing Types Across the United States as of 2008

	PSG	VSG	SPS	DS	TIS	3STK
Alabama	-	2006	-	-	-	-
Alaska	-	-	1980	-	-	-
Arizona	-	-	1978	1994	1994	-
Arkansas	-	1994	-	-	-	1995
California	-	-	1976	1976	1994	1994
Colorado	-	-	1979	79-85	-	1994
Connecticut	-	-	-	81-90	1995	1994
Delaware	-	1987	-	1990	1990	-
Florida	1994	1983-93	-	1983	1995	1995
Georgia	-	-	-	-	1995	1995
Hawaii	-	-	-	-	-	-
Idaho	-	-	-	-	-	-
Illinois	-	-	-	1978	-	-
Indiana	-	-	1977	1977	-	1994
Iowa	-	-	-	-	1996	-
Kansas	1993	-	-	-	1993	1994
Kentucky	-	-	-	-	-	-
Louisiana	-	1987	-	-	-	1994
Maine	-	-	-	1976	1995	-
Maryland	-	1983	-	-	-	1994
Massachusetts	-	-	-	-	-	-
Michigan	1999	1984-98	-	-	1994	-
Minnesota	1980	-	-	1982	1993	-
Mississippi	-	-	-	1995	1995	-
Missouri	-	1997	-	-	1994	-
Montana	-	-	-	-	-	1995
Nebraska	-	-	-	-	-	-
Nevada	-	-	-	-	-	1995
New Hampshire	-	-	-	-	-	-

Table 1 continued

New Jersey	-	-	1977	-	-	1995
New Mexico	-	-	1977	1977	-	1994
New York	-	-	-	-	1995	-
North Carolina	1995	-	-	1981	1994	1994
North Dakota	-	-	-	-	1995	1995
Ohio	1996	-	-	1996	1996	-
Oklahoma	-	-	-	-	-	-
Oregon	1989	-	-	1989	1995	-
Pennsylvania	1982	-	-	-	1991	1995
Rhode Is.	-	-	1981	-	-	-
South Carolina	-	-	-	-	-	1995
South Dakota	-	-	-	-	1996	-
Tennessee	1989	-	-	-	1995	1995
Texas	-	-	-	-	-	-
Utah	-	1985	-	-	1985	1995
Vermont	-	-	-	-	-	1995
Virginia	-	1995	-	1995	1995	1994
Washington	1984	-	-	1984	1984	1993
West Virginia	-	-	-	-	-	-
Wisconsin	-	85-94 & 99	-	-	1999	1994
Wyoming	-	-	-	-	-	-
Total	10	11	8	18	24	24

This table represents the current sentencing type used by each state as of 2008. PSG stands for presumptive sentencing guidelines. VSG stands for voluntary sentencing guidelines. SPS stands for statutory presumptive sentencing. DS stands for determinate sentencing. TIS stands for truth in sentencing. 3STK refers to three strikes laws. All other states utilize indeterminate sentencing.

The crime variables were lagged to account for the time lag inherent in the impact of crime rates on imprisonment, acknowledging the inherent delay in processing arrested individuals through the legal system. These variables were further normalized into percent change scores to mitigate potential issues arising from trends in arrest rates.

The second category of controls included six demographic variables. Data for the percentage of Black and Hispanic populations, percentage of urban residents, Gini coefficient representing income inequality, and state population were drawn from the U.S. Census for 1965-2008. The percentage of urban residents was converted into a percent change score because it exhibited an upward trend over time. In contrast, the percentages of Black and Hispanic populations remained relatively stable within states, exhibiting variations across states rather than over time. The percentages of Black and Hispanic populations remained unaltered. The Gini Index is a metric used to gauge income inequality, condensing detailed income distribution data into a single statistic. It ranges from 0 to 1, with 0 representing perfect equality (where all individuals earn the same) and 1 indicating perfect inequality (where one individual or group holds all the income). It compares the actual income distribution, illustrated by the Lorenz curve, with an ideal scenario of equal income distribution (U.S. 1965-2008).

The final two demographic variables, the unemployment rate and the percentage living below the poverty line, were derived from the U.S. Bureau of Labor Statistics (1965-2008a) and the U.S. Bureau of the Census (2008), respectively. These variables, demonstrating limited temporal trends, were not subjected to transformation.

The third set of controls in the study includes a political variable denoted as FHREP. This variable represents the political party affiliation controlling the state house or assembly. A positive score indicates GOP (Republican Party) control, based on data sourced from Dubin (2007) and Hershey (2007). The variable is lagged to reflect that political control typically requires at least two years to influence state operations. Research findings indicate that all three sets of controls represent pivotal factors associated with imprisonment, justifying their incorporation into this analysis based on established literature (Barker, 2006; Beckett & Western, 2001; Blumstein & Beck, 1999; Parker & Horwitz, 1986; Raphael, 2009; Spelman, 2009).

Table 3. Descriptives of control variables.

	Mean	Std. Dev.	25th Percentile	75th Percentile	Min Value	Max Value
Crime Controls	-	-	-	-	-	-
Violent Crime	475.316	247.006	281.238	638.378	47.564	1570.000
Drug Crime	441.377	424.764	208.613	551.194	6.159	8134.423
Demographic Controls	-	-	-	-	-	-
Percent Black	10.858	8.845	3.814	15.288	0.222	37.227
Percent Hispanic	6.393	8.566	1.392	7.388	0.247	65.433
Percent Unemployment	5.753	2.001	4.400	6.700	2.200	18.000
Percent Poor	12.791	3.885	9.900	15.100	2.900	27.200
Percent Urban	70.290	14.680	59.922	82.819	30.240	100.000
Gini Coefficient	0.554	0.057	0.514	0.581	0.404	0.889
FHREP^	-0.722	2.396	-2.787	1.862	-3.862	3.615

Note: The statistics include mean, standard deviation (Std. Dev.), the 25th and 75th percentiles, and the minimum (Min Value) and maximum (Max Value) values observed for each variable. ^ FHREP represents GOP control of the state house.

Methods

In this study, six key sentencing reforms will be analyzed that were enacted by various states from the 1970s to the early 2000s: presumptive sentencing guidelines

(PSG), voluntary sentencing guidelines (VSG), statutory presumptive sentencing (SPS), determinate sentencing (DS), Truth in Sentencing (TIS), and Three Strikes Laws (3STK). These reforms, all implemented during that period, collectively represent a shift towards a more punitive justice model. Presumptive sentencing guidelines, voluntary sentencing guidelines, and statutory presumptive sentencing are mutually exclusive front-end reforms and are usually not implemented in conjunction with each other or with indeterminate sentencing. On the other hand, Truth in Sentencing and determinate sentencing, both back-end reforms, can coexist with the front-end reforms as they pertain to the release phase. The Three Strikes Law also serves as sentencing enhancements and can be implemented alongside other sentencing reforms. (Stemen et al., 2006).

Several steps are involved in arriving at an informed decision regarding the optimal statistical model. The process begins with setting up and running different model specifications. For this analysis, Stata was utilized as the statistical software program to determine the best specifications and to test the different approaches. These models include the sentencing reform variables and, in some models, the control variables outlined above. This study aimed to explore and identify the most accurate approach to modeling the interactions between the reforms, which Jaccard et al. (2003) emphasize the importance of.

To ensure the reliability and robustness of the results, the analysis also includes a Hausman test. The Hausman test is pivotal in guiding the choice between fixed effects (FE) and random effects (RE) models in panel models (Halaby, 2004). By comparing parameter estimates from these two models, the Hausman test scrutinizes the critical exogeneity assumption, evaluating whether the residuals from regressing independent variables on unobserved individual-specific effects (random effects) are correlated with the explanatory variables (Halaby, 2004). Or said another way, is it likely that there is an unobserved variable or variables correlated with the explanatory variables that should have been included but were not? A significant result assumes potential endogeneity, prompting a preference for the fixed effects model. This decision aligns with focusing on within-state variations over time, ensuring that unobservable factors are appropriately captured.

On the other hand, if no significant difference is detected, the more efficient random effects model may be favored. The Hausman test acts as a methodological safeguard, ensuring the validity and reliability of the panel model results by addressing the delicate balance between bias and efficiency associated with potential endogeneity. In doing so, it enhances the credibility of the findings, providing a robust foundation for drawing meaningful conclusions about the relationship between sentencing reforms and imprisonment rates at the state level (Wooldridge, 2002; Halaby, 2004). Although this test is being conducted, models will be run with both FE and RE.

The analysis will include both FE and RE models, regardless of the Hausman test results, to comprehensively examine various model specifications and their impacts on the study's outcome. This intentional approach allows for a thorough exploration of the sensitivity of results to different model specifications. By employing FE models, the study can capture nuanced within-state variations over time, providing insights into the specific effects of sentencing reforms within individual states (Halaby, 2004). The use of RE models facilitates the exploration of broader cross-sectional patterns, shedding light on the overall relationship between sentencing reforms and imprisonment rates across states. Through this methodological strategy, the study aims to uncover the diverse influences of different model specifications.

The primary purpose of the analysis is to explore the different ways that different combinations of reforms could be specified and which approach will produce the most accurate and robust outcomes. The analysis incorporates three reform specifications: the standard model, the fully saturated model with interaction terms, and the regime score model. Each model will encompass four variations, ensuring a comprehensive analysis. For each specification, the models will include both FE and RE configurations. Additionally, each specification will have two with and two without control variables, totaling four distinct models for each specification. This methodological approach allows for a robust exploration of panel modeling, providing comparisons between fixed and random effects models.

The standard model, which is a model without interaction terms, serves as the baseline, capturing the fundamental relationship between sentencing reforms and imprisonment rates. This model provides a straightforward examination of the overall impact of reforms on incarceration. Its simplicity offers clarity in understanding the general trend and direction of the relationship. However, the standard model's limitation lies in its oversimplification, as it may need to account for the intricate dynamics and variations associated with specific combinations of reforms. Most states have a combination of reforms enacted simultaneously and likely interact (Harmon, 2012).

Interaction terms allow for a nuanced and complex examination of how different combinations of reforms may modify the effects on imprisonment rates (Jaccard & Turrisi, 2003). A fully saturated model allows a thorough exploration of how different sentencing reforms interact by enabling them to dynamically interact with each other and for the measurement of the reforms' impacts. This approach provides a detailed understanding of the combined effects of various policy changes. However, while the fully saturated model is comprehensive, it may have some drawbacks, such as increased complexity and potential challenges in interpretation. The inclusion of interaction terms increases the complexity of the model. With each interaction term representing the combined effects of two or more independent variables, the number of parameters to estimate escalates, posing computational challenges and enhancing the intricacy of the model structure. In a small dataset, the increased loss of degrees of freedom poses the risk of diminishing statistical power (Jaccard & Turrisi, 2003).

The interpretation of results becomes more intricate with the inclusion of numerous interaction terms. Researchers must discern the main effects of individual variables and understand how these effects vary based on the levels of other interacting variables, demanding a deeper understanding of the model's intricacies. A critical concern with the fully saturated model is the risk of overfitting. As the model accommodates a multitude of interaction terms, there is a heightened susceptibility to capturing noise or random variability in the data rather than genuine relationships. While the fully saturated model thoroughly explores complex relationships, researchers must model complexity and interpretability. It is essential to ensure that the insights gained from the model are meaningful and reliable without sacrificing clarity of interpretation (Jaccard & Turrisi, 2003).

The regime-score model, also known as a "cluster model," is a statistical approach used in data analysis to categorize observations into distinct regimes or clusters based on specific criteria (Jaccard & Turrisi, 2003). In this model, the regime score variables represent each unique combination of reforms within each regime. By organizing data into different regimes, the regime-score model allows researchers to examine variations in relationships between variables across different regimes or time periods.

Compared to the fully saturated model, the regime score model can provide a more detailed explanation in certain contexts. However, its effectiveness can vary depending on how well-defined the regimes are and how accurately they reflect the underlying dynamics of the studied system. By categorizing data into different regimes based on specific criteria, this model can capture variations in the relationships between reforms across different time periods. For example, suppose there are significant policy changes or shifts in the criminal justice system over time. In that case, the regime scores model can effectively capture these changes by identifying different regimes corresponding to each period. This allows researchers to analyze how sentencing reforms impact imprisonment rates within each regime, providing insights into the varying effects of policies over time (Jaccard & Turrisi, 2003).

However, it is essential to recognize that the regime scores model simplifies the analysis by reducing the complexity of the data into discrete categories. While this approach can enhance interpretability, it may overlook subtle distinctions and interactions between variables within each regime (Jaccard & Turrisi, 2003). Additionally, defining regimes based on specific criteria can be subjective and may only sometimes capture the full complexity of the data.

The fully saturated model offers detailed insights into combined reforms but comes with added complexity. On the other hand, the regime scores model simplifies the analysis but may overlook intricate interactions. The choice between the two depends on the research objectives and the level of detail needed for a comprehensive understanding of the relationships between sentencing reforms. This methodological progression ensures a comprehensive evaluation of various model specifications.

One crucial aspect of this analysis is determining which models yield the most accurate results. I will first look at model fit tests to determine which models performed best. These tests assess how well the chosen model fits the data. A well-fitting model should provide a good representation of the underlying relationships within the data (Finkel, 1995). The analysis will also consider the R-squared statistic, which measures the model's explanatory power. A higher R-squared value indicates that the model can explain a more significant proportion of the variation in the dependent variable (Finkel, 1995). This is crucial for understanding the model's overall performance.

To aid in the assessment and to allow for comparison across different specifications, margin scores will be calculated. In this context, it refers to the differences or changes in the predicted imprisonment rate that result from alterations in the combination of sentencing reforms. These scores provide a quantitative measure of the impact of specific reforms or combinations of reforms on the outcome variable. For example, suppose the fully saturated model and the regime score model predict imprisonment rates for different combinations of sentencing reforms. In that case, the margin scores help quantify the extent of variation or difference in these predictions. They offer a numerical representation of the incremental effects of specific reforms or combinations thereof.

Calculating margin scores is valuable because it allows for a more nuanced understanding of how changes in sentencing policies influence the outcome variable. By comparing these scores between the fully saturated model and the regime score model, researchers can assess the effectiveness of different modeling approaches in capturing the complexities of sentencing reforms and their interactions (StataCorp, 2023).

The analysis systematically compares panel models with varying specifications to determine the most accurate and statistically robust model. The process begins by setting up different model specifications, incorporating fixed effects (FE) and random effects (RE) models. Additionally, the analysis includes models with and without controls to account for potential confounding variables. The examination extends to interaction terms within a fully saturated model, allowing for a nuanced exploration of the combined effects of various sentencing reforms. Margin scores are calculated to facilitate better comparisons between the fully saturated and regime score models, providing insights into the impact of different combinations of reforms. Hausman tests are employed to identify whether FE or RE models are more appropriate, ensuring the validity of the models. The assessment of statistical significance, model fit, and the R-squared statistic contribute to selecting the most appropriate model for the specific analytical context. This comprehensive approach aims to yield reliable and meaningful results, considering the intricacies of sentencing reforms and their interactions with imprisonment rates.

Results

These results depict the impact of state-level sentencing reforms on imprisonment rates, capturing the intricate dynamics of state-level sentencing reforms within the criminal justice system. The study, rooted in a comprehensive methodological framework, leverages panel models and a range of specifications to determine the optimal model. Its primary goal is to differentiate the most accurate depiction of the relationships between the different reforms and imprisonment.

Baseline Panel Models: Individual Reforms without Controls or Interaction Terms (Models 1 and 2)

Table 4 illustrates the outcomes examining individual-level reforms without control variables or interaction terms. In Model 1 with random effects, the overall regression was statistically significant (Wald chi2(6) = 1256.41, $p<0.001$), indicating that the model explained 21.7% ($R2=0.217$) of the variation in incarceration rates and that the model was a good fit. Each reform is treated as an independent effect. It is essential to interpret these results cautiously due to the likelihood of interactions between reforms in real-world applications. Among the examined reforms, voluntary sentencing guidelines, statutory presumptive sentencing, truth in sentencing, and three-strikes laws were found to be statistically significant. Notably, front-end reforms such as voluntary and statutory presumptive sentencing were associated with higher imprisonment rates than indeterminate states.

Specifically, voluntary sentencing guidelines led to a substantial percent change of 49.4% ($P < 0.000$), while statutory presumptive sentencing resulted in a percent change of 61.6% ($P < 0.000$) relative to states without reforms. Furthermore, back-end reforms also exhibited notable effects. Truth in sentencing was associated with a percent change of 70.3% ($P < 0.000$), while three-strikes laws resulted in a percent change of 53.8% ($P < 0.000$). This means states implementing these reforms tended to experience higher incarceration rates during the study period (1978–2008) than states with indeterminate sentencing. On the other hand, presumptive sentencing guidelines and the presence of determinate sentencing are statistically insignificant.

Testing Interaction Effects Using Different Panel Model Specifications

Table 4. Results of the panel analysis models with no interaction terms, no controls, random effects, and fixed effects.

	Model 1		Model 2	
Individual Reforms	Random Effects	S.E.	Fixed Effects	S.E.
Pres. Sent. Guide. (PSG)	-0.215	(13.968)	-1.654	(14.180)
Voluntary Sent Guide. (VSG)	94.839***	(11.659)	92.234***	(11.826)
Statutory Presump. Sent. (SPS)	118.349***	(24.813)	170.086***	(30.286)
Truth in Sentencing (TIS)	135.074***	(7.919)	134.330***	(7.947)
Determinate Sentencing (DS)	12.536	(10.416)	18.766	(10.684)
Three Strikes (3STK)	103.386***	(7.775)	104.483***	(7.797)
Constant	192.038***	(15.628)	182.849***	(5.770)
Observations	1550	-	1550	-
F Statistic	-	-	213.250	-
Chi2	1256.410	-	-	-
R-squared	0.217	-	0.185	-

Standard errors in parentheses.

***p < 0.001, **p < 0.01, *p < 0.05.

Model 2 introduces time and state-specific fixed effects to variables in Model 1. While still a good fit, Model 2 exhibits a slightly lower R-squared value (R2=0.185, F=213.250, p< 0.000) compared to Model 1 (0.217). There are notable similarities and distinctions in the impacts of the reforms between the two models. Presumptive sentencing guidelines and determinate sentencing maintain insignificance in both models with relatively small coefficients, signifying a consistent lack of impact on incarceration rates when compared to states with no reforms. While specific reforms, like voluntary sentencing guidelines, truth in sentencing laws, and three strikes laws, remain significant in both models, there are nuanced differences. Model 2 reports a similar percentage change for voluntary sentencing guidelines, which is 50.4% ($P<0.000$). On the other hand, Statutory had a much higher percent change for statutory sentencing (93%, $p <0.000$) than Model 1 (61.6%). It is around 1.5 times higher than in Model 1.

Baseline Panel Models: Individual Reforms with Controls and No Interaction Terms (Models 3 and 4)

Table 5 depicts the results of the expanded analysis, in which control variables were introduced to both the baseline random-effects and fixed-effects models. These controls are known factors that influence state-level incarceration rates (Barker, 2006; Beckett & Western, 2001; Blumstein & Beck, 1999; Parker & Horwitz, 1986; Raphael, 2009; Spelman, 2009). The controls include variables representing violence rates, drug rates, demographic factors (percentage of Black and Hispanic populations), economic indicators (unemployment and poverty rates), urbanization, income inequality, and GOP control of the state house (FHREP).

The random-effects model, denoted as Model 3, demonstrated a marked improvement in explanatory power, boasting an overall R-squared of 0.491 (Wald $Chi2(15) = 3010.63$, $p< 0.000$). This suggests that approximately 49.09% of the variation in state-level incarceration rates is explained by the combined influence of the reforms and the introduced control variables. Notably, this represents a substantial increase compared to the previous random-effects model without controls ($R2 = 0.217$) and the preceding fixed-effects model ($R2 = 0.185$). The Wald chi-squared test (3010.63) indicates that the model fit was nearly three times larger than model 1 (1256.410) without controls. Including control variables provide a more comprehensive understanding of the factors influencing state-level incarceration rates, likely leading to more accurate modeling of the impacts of the reforms.

Findings suggest that voluntary sentencing, truth in sentencing, and three strikes laws maintain their significance, with coefficients of 28.614 ($p<0.002$), 55.726 ($p<0.000$), and 26.885 ($p<0.000$), respectively. This signifies that those states that implement such laws experience higher incarceration rates. Interestingly, presumptive sentencing guidelines are statistically significant in this model (-49.164, $p<0.000$), unlike previous models with no controls. Unlike the other front-end reforms, the model suggests presumptive sentencing guidelines reduce imprisonment compared to non-reform states.

The control variables contribute noteworthy findings as well. The variables for the percentage of Black and Hispanic persons in the populations have coefficients of 4.456 ($p<0.000$) and 7.974 ($p<0.000$), respectively, indicating strong associations with incarceration rates. When these coefficients are standardized, they further elucidate the substantial impact of these demographic factors: the standardized beta for percent Black is 0.247, and for percent Hispanic, it is even higher at 0.427, underscoring their significant influence relative to other variables.

Table 5. Results of the panel analysis models with no interaction terms, with controls, random effects, and fixed effects.

	Model 3		Model 4	
Individual Reforms	Random Effects	S.E.	Fixed Effects	S.E.
Pres. Sent. Guide. (PSG)	-49.164***	(11.120)	-52.243***	(11.062)
Voluntary Sent Guide. (VSG)	28.614**	(9.302)	29.132**	(9.223)
Statutory Presump. Sent. (SPS)	-28.052	(19.042)	36.981	(25.567)
Truth in Sentencing (TIS)	55.726***	(6.906)	51.266***	(6.754)
Determinate Sentencing (DS)	-3.885	(8.157)	-1.004	(8.235)
Three Strikes (3STK)	26.885***	(6.579)	19.996**	(6.503)
Crime Controls	-	-	-	-
Violent Crime	0.063***	(0.013)	0.057***	(0.014)
Drug Crime	-3.01	(0.005)	0.003	(0.005)
Demographic Controls	-	-	-	-
Percent Black	4.456***	(0.646)	2.204**	(0.758)
Percent Hispanic	7.974***	(0.864)	11.255***	(1.070)
Percent Unemployment	-5.238***	(1.448)	-4.453**	(1.430)
Percent Poor	-3.901***	(1.056)	-5.077***	(1.072)
Percent Urban	-0.038	(0.126)	-0.067	(0.122)
Gini Coefficient	1379.07***	(60.832)	1295.575***	(61.447)
FHREP^	11.871***	(1.549)	14.063***	(1.593)
Constant	-530.420***	(36.416)	-470.872***	(35.124)
Observations	1359	-	1359	-
F Statistic	-	-	216.120	-
Chi2	3010.63	-	-	-
R-squared	0.491	-	0.306	-

Standard errors in parentheses. ^ FHREP represents GOP control of the state house.

***p < 0.001, **p < 0.01, *p < 0.05.

Moreover, the analysis reveals significant negative associations of unemployment and poverty rates with incarceration rates, with standardized betas of -0.066 and -0.095, respectively. These findings suggest that as unemployment and poverty decrease, incarceration rates also tend to be lower, indicating socioeconomic stability might contribute to lower crime rates. Income inequality also exhibits a profound impact, with a standardized beta of 0.495, making it one of the most influential predictors in the model. Similarly, political control, indicated by a GOP-controlled state house, shows a substantial positive association with incarceration rates, with a standardized beta of 0.178. These standardized coefficients reflect the influence of each variable on the dependent variable, imprisonment rate, adjusted for their respective scales. Variables with higher absolute values of standardized beta have a stronger impact on the dependent variable. For instance, the percentage of Hispanics and income inequality show particularly strong positive influences, whereas variables like drug rate and urban have minimal impact.

Model 4 includes fixed effects and controls and yields an overall R-squared of 0.306 ($F=216.120$, $p<0.000$), surpassing the explanatory power of the fixed-effects model without controls ($R2 = 0.185$), suggesting that adding the control variables increased the amount of variation explained in incarceration rates by 65%, underscoring the importance of control variables.

Within the individual-level reforms, Presumptive sentencing guidelines exhibit a negative coefficient of -52.243 ($p<0.000$), suggesting that imprisonment grew less when this reform was present. Voluntary sentencing guidelines maintain their statistical significance with a coefficient of 29.132 ($p<0.002$), underscoring the persistent impact of voluntary sentencing guidelines on higher imprisonment rates. Truth in sentencing and three-strike laws continue to exhibit a positive and significant association with an incarceration rate with coefficients of 51.266 ($p<0.000$) and 19.996 ($p<0.002$), again reinforcing their consistent impact across various models. Conversely, determinant sentencing remains statistically insignificant, indicating its limited influence on state-level incarceration rates.

Within the control variables, percentages of Blacks and Hispanics continue to be significant across models, with coefficients of 2.204 ($p<0.004$) and 11.255 ($p<0.000$), respectively, in Model 4. The standardized beta for Blacks is 0.122, and for Hispanics, it is notably higher at 0.603, indicating a stronger relative influence on incarceration rates. Percent Hispanic also has the highest standardized beta. Unemployment and poverty rates are both statistically significant and exhibit negative coefficients of -4.453 ($p<0.002$) and -5.077 ($p<0.000$), with their standardized betas being -0.056 and -0.123, respectively. These negative values suggest that higher unemployment and poverty rates are associated with lower incarceration rates, adjusting for other factors in the model. Income inequality had a positive coefficient of 1,295.575 ($p<0.000$) and the second-highest beta of .465. A higher absolute value of a standardized beta coefficient means that the predictor substantially impacts the dependent variable for each standard deviation change in the predictor, controlling for other variables in the model. This nuanced understanding allows for a more straightforward interpretation of how socioeconomic conditions and racial demographics impact incarceration rates across different states.

A Hausman test was conducted for Models 3 and 4 to address the choice between fixed-effects and random-effects models. The results suggest that using fixed effects is best, indicating that the individual state-specific effects are not random (Chi2(14) =184.74, $p<0.000$). While the random-effects model 3 appears to have a higher overall R-squared (49.09%), the appropriateness of each model depends on the research question and underlying assumptions.

Fully Saturated Models: No Controls and Interaction Terms (Models 5 and 6)

Table 6 showcases the results from the fully saturated models that incorporate first-level interaction terms with no control variables. Model 5 includes random effects, and the overall regression was found statistically significant (Wald Chi2(18) = 1468.57, $p < 0.000$), signifying that the model accounts for 22.14% of the variation in state-level incarceration rates. This model builds upon previous models by introducing interaction terms, providing a more nuanced understanding of the joint effects of individual-level reforms.

Table 6. Results of the panel analysis models with interaction terms, with no controls, random effects, and fixed effects.

	Model 5		Model 6	
Individual Reforms	Random Effects	S.E.	Fixed Effects	S.E.
Pres. Sent. Guide. (PSG)	94.088***	(27.232)	98.769***	(27.478)
Voluntary Sent Guide. (VSG)	82.595***	(14.922)	82.125***	(15.012)
Statutory Presump. Sent. (SPS)	129.811***	(25.367)	165.939***	(29.451)
Truth in Sentencing (TIS)	189.998***	(12.765)	189.394***	(12.799)
Determinate Sentencing (DS)	66.832***	(15.208)	74.907***	(15.477)
Three Strikes (3STK)	155.076***	(13.990)	152.991***	(14.055)
Reforms in combination	-	-	-	-
PSG + DS	-92.839**	(30.524)	-100.983***	(30.892)
PSG + TIS	-77.307	(47.810)	-79.020	(48.019)
PSG + 3STK	24.678	(41.679)	25.635	(41.948)
VSG + DS	-64.135**	(24.849)	-72.571**	(25.088)
VSG + TIS	3.1794	(20.568)	4.627	(20.644)
VSG + 3STK	21.563	(18.838)	22.701	(18.880)
SPS + DS	-105.876***	(32.983)	-93.804**	(35.691)
SPS + TIS	42.202	(30.391)	30.953	(31.474)
SPS + 3STK	-0.815	(24.183)	9.463	(24.859)
DS + TIS	-47.645**	(17.370)	-49.075**	(17.403)
DS + 3STK	-14.535	(17.703)	-15.529	(17.753)
TIS + 3STK	-123.633***	(17.782)	-120.054***	(17.934)
Constant	183.471***	(16.272)	175.515***	(5.936)
Observations	1550	-	1550	-
F Statistic	-	-	82.810	-
Chi2	1468.570	-	-	-
R-squared	0.221	-	0.196	-

Standard errors in parentheses.

***p < 0.001, **p < 0.01, *p < 0.05.

Testing Interaction Effects Using Different Panel Model Specifications

Examining the coefficients of Model 5 reveals significant positive associations for all individual-level reforms, unlike all previous models. Presumptive sentencing exhibits a robust impact with a coefficient of 94.088 (p<0.001), indicating that states adopting. These guidelines experience higher incarceration rates. In Models 3 and 4, presumptive sentencing had a negative coefficient. This discrepancy prompts further exploration into the underlying factors driving the change in direction and magnitude of the coefficient across different model iterations. Voluntary Sentencing Guidelines also suggest higher incarceration rates with a coefficient of 82.595 (p<0.000). Statutory Presumptive Sentencing also displays statistical significance with a coefficient of 129.811 (p<0.000). Truth in sentencing and three strikes have positive statistical significance as well, with the highest coefficients of 189.998 (p<0.000) and 155.076 (p<0.000), respectively. Determinate sentencing had the lowest coefficient of 66.832 (p<0.000).

The introduction of interaction terms provides additional depth to the analysis. Notably, most interaction terms involving determinate sentencing are significant with negative coefficients. Statutory and determinate sentencing has the highest negative coefficient of -105.876 (p<0.001). The interaction between presumptive and determinate sentencing reveals a negative coefficient of -92.839 (p<0.002), suggesting a potential mitigating effect. In contrast, the interaction term between presumptive sentencing and truth in sentencing exhibits a non-significant coefficient.

Model 6 includes fixed effects, first-level interaction terms, and no control variables. The model was statistically significant (F= 82.81, P<0.000). Regarding model fit, the overall R^2 of 0.196 indicates that approximately 19.6% of the variation in state-level incarceration rates is explained by the combined influence of the reforms and interaction terms, which is slightly less than Model 5. Analyzing the coefficients in Model 6 reveals significant positive associations for several reforms, and all individual-level reforms were significant, as in Model 5. Presumptive sentencing and voluntary sentencing display a noteworthy impact with coefficients of 98.769 ($p<0.000$) and 82.125 ($p<0.000$), respectively. Statutory Presumptive Sentencing substantially increases incarceration rates (coefficient = 165.939, $p<0.000$). Truth in Sentencing and Three Strikes maintain strong positive associations as well, with coefficients of 189.394 ($p<0.000$) and 152.991 ($p<0.000$), respectively. Again, determinate sentencing has the lowest coefficients of the individual-level reforms (74.907, $p<0.000$).

Most interaction terms that include determinate sentencing are significant and showcase negative coefficients except for the interaction between determinate sentencing and three strikes. This may suggest that reforms combined with determinate sentencing mitigate incarceration rates compared to states with no reforms like Model 5. The presumptive and determinate sentencing interaction term reveals a negative coefficient (-100.983, $p<0.001$). Voluntary and determinate, statutory and determinate, and truth in sentencing and determinate also reach significance with negative coefficients of -72.571 ($p<0.004$), -93.804 ($p<0.009$), and -49.075 ($p<0.005$), respectively. In combination, truth in sentencing and three-strike laws also exhibit a negative coefficient (-120.054, $p<0.000$), like Model 5, which also found a significant and negative coefficient. In contrast, the interaction terms between presumptive sentencing and truth in sentencing and presumptive sentencing and three strikes laws explore a joint influence without reaching statistical significance.

Fully Saturated Models: Controls and Interaction Terms (Models 7 and 8)

In Table 7, Models 7 and 8 incorporate first-level interaction terms and control variables. Model 7 incorporates random effects. The overall fit and significance are crucial in understanding the explanatory power of the included variables. The R-squared value (R2=.483) indicates that approximately 48.30% of the variability in incarceration rates is explained by the combination of sentencing reforms, interaction terms, and control variables. This is significantly higher than Models 5 and 6 with no controls with R-squared values at .221 and .196, respectively. Additionally, the model is a good fit, as indicated by the significant Wald Chi2(27) = 3309.96, $p<0.000$.

Truth in Sentencing, Determinate Sentencing, and Three Strikes laws exhibit statistically significant coefficients, suggesting their considerable influence on escalating imprisonment rates compared to the absence of these reforms. This is a change from Models 5 and 6, where all individual-level reforms were found significant. Truth in Sentencing emerges with the highest coefficient (91.304, $p<0.001$), followed by Three Strikes (48.377, $p<0.002$) and Determinate Sentencing (36.354, $p<0.000$), underscoring their impact on driving up incarceration rates relative to no reforms.

Testing Interaction Effects Using Different Panel Model Specifications

Table 7. Results of the panel analysis models with interaction terms, controls, random effects, and fixed effects.

	Model 7		Model 8	
Individual Reforms	Random Effects	S.E.	Fixed Effects	S.E.
Pres. Sent. Guide. (PSG)	-5.852	(20.063)	-4.782	(19.828)
Voluntary Sent Guide. (VSG)	2.503	(11.823)	4.406	(11.630)
Statutory Presump. Sent. (SPS)	-14.113	(20.011)	37.310	(25.009)
Truth in Sentencing (TIS)	91.304***	(10.793)	85.589***	(10.570)
Determinate Sentencing (DS)	36.354**	(11.699)	42.350***	(11.738)
Three Strikes (3STK)	48.377***	(11.747)	35.507**	(11.564)
Crime Controls	-	-	-	-
Violent Crime	0.067***	(0.013)	0.062***	(0.013)
Drug Crime	-0.001	(0.005)	0.001	(0.005)
Demographic Controls	-	-	-	-
Percent Black	4.246***	(0.648)	2.269**	(0.749)
Percent Hispanic	8.986***	(0.920)	12.117***	(1.119)
Percent Unemployment	-5.069***	(1.419)	-4.465***	(1.402)
Percent Poor	-2.689**	(1.046)	-3.485***	(1.062)
Percent Urban	-0.045	(0.122)	-0.081	(0.119)
Gini Coefficient	1312.402***	(60.448)	1239.524***	(60.893)
FHREP^	12.802***	(1.553)	14.856***	(1.598)
Reforms in combination	-	-	-	-
PSG + DS	-67.829**	(23.739)	-77.522***	(23.495)
PSG + TIS	42.800	(45.685)	53.848	(44.674)
PSG + 3STK	-51.796	(42.310)	-60.411	(41.385)
VSG + DS	-40.066*	(19.438)	-45.703*	(19.197)
VSG + TIS	23.661	(15.957)	23.326	(15.590)
VSG + 3STK	47.088**	(14.997)	49.626***	(14.648)
SPS + DS	-77.389**	(24.780)	-56.275*	(26.337)
SPS + TIS	29.259	(23.574)	7.206	(24.105)
SPS + 3STK	-10.494	(19.255)	3.385	(19.277)
DS + TIS	-42.369**	(13.821)	-44.232***	(13.500)
DS + 3STK	-16.505	(14.243)	-24.868	(13.920)
TIS + 3STK	-62.057***	(14.328)	-47.423***	(14.102)
Constant	-519.129***	(35.866)	-470.855***	(34.381)
Observations	1359	-	1359	-
F Statistic	-	-	131.200	-
Chi2	3309.960	-	-	-
R-squared	0.483	-	0.315	-

Standard errors in parentheses. ^ FHREP represents GOP control of the state house.

***p < 0.001, **p < 0.01, *p < 0.05.

Interestingly, the analysis of first-level interaction terms unveils intriguing dynamics. While the interaction between Voluntary Sentencing and Three Strikes laws displays a positive coefficient (47.088, p<0.002), implying a combined effect leading to higher incarceration rates than no reforms, determinate sentencing appears alongside three other reforms with significant yet negative coefficients similar to previous models (5 and 6).

The negative coefficient (-67.829, p<0.004) associated with the interaction between Presumptive and Determinate Sentencing suggests a combined effect in reducing incarceration rates compared to no reforms. This could indicate that implementing both types of sentencing reforms leads to more precise and predictable sentencing outcomes, potentially resulting in fewer individuals being incarcerated compared to the scenario without these reforms.

Similarly, the negative coefficients observed with the interactions of Statutory and Determinate Sentencing (-77.398, p<0.002) and Truth in Sentencing and Determinate Sentencing (-42.369, p<0.002) suggest a combined effect in reducing incarceration rates compared to no reforms. Additionally, the negative coefficient (-62.057, p<0.000) associated with the interaction between Truth in Sentencing and Three Strikes laws suggests a combined effect in reducing incarceration rates. This could imply that when both truth in sentencing and three strikes laws are in place, there may be a deterrent effect, leading to lower incarceration rates than the scenario without these reforms.

These findings underscore the complex interplay between different types of sentencing reforms and their combined impact on incarceration rates relative to the absence of reforms. While specific reforms individually contribute to higher incarceration rates compared to no reforms, the interactions between various reforms reveal nuanced effects that can exacerbate or mitigate the overall impact on imprisonment rates compared to states without reforms.

Among the control variables, notable findings emerged, shedding light on the influence of various demographic and socioeconomic factors on incarceration rates. Firstly, a positive and statistically significant coefficient was observed for the violent crime rate (0.067, $p < 0.001$), indicating that higher rates of violent crime were associated with increased incarceration rates. Conversely, the coefficient for drug crime rate was not statistically significant, similar to all the previous models with controls, suggesting no consistent association between drug-related crime rates and incarceration rates after controlling for other factors. Furthermore, disparities within the criminal justice system were evident in the findings related to race and ethnicity. Both the percentage of Black residents (4.246, $p < 0.001$) and the percentage of Hispanic residents (8.986, $p < 0.001$) exhibited positive and significant coefficients, indicating that higher proportions of Black and Hispanic populations within a jurisdiction were associated with higher incarceration rates.

Socioeconomic factors played a significant role in influencing incarceration rates, but the results indicate a complex story. The negative and significant coefficient for the unemployment rate (-5.069, $p < 0.001$) suggested that higher unemployment rates were associated with lower incarceration rates. Similarly, higher levels of poverty (percentage of the population living below the poverty line) exhibited a negative and statistically significant coefficient (-2.689, $p < 0.010$), indicating that higher poverty levels were associated with lower incarceration rates. However, as measured by the Gini coefficient, income inequality showed a positive and highly significant coefficient (1312.402, $p < 0.001$), suggesting that higher levels of income inequality were associated with higher incarceration rates.

Model 8 uses a fixed-effects regression with interaction terms and control variables. This model demonstrated a robust explanatory power, with an R-squared value of 0.734, indicating that approximately 73.4% of the variability in incarceration rates was explained by the included variables compared to 48.30% in model 7, which utilized random effects. It is also significantly better than Model 6, which included fixed effects and no controls, with 19.6%. Additionally, the model exhibited overall significance, as confirmed by the F-statistic ($F= 131.20$, $p < 0.000$). Models 7 and 8 had the same individual-level reforms that were found significant. Truth in Sentencing emerges again with the highest coefficient (85.589, $p<0.000$), followed by Three Strikes (48.377, $p<0.002$) and Determinate Sentencing (42.350, $p<0.000$).

The first-level interaction terms analysis unveils dynamics like model 7, which used random effects. While Voluntary Sentencing in combination with the Three Strikes laws displays a positive coefficient (49.626, $p<0.001$), implying a combined effect leading to higher incarceration rates compared to no reforms, determinate sentencing appears alongside two other reforms with significant yet negative coefficients.

The negative coefficient (-77.522, $p<0.001$) associated with the interaction between Presumptive and Determinate Sentencing suggests a combined effect in reducing incarceration rates compared to no reforms. Similarly, the negative coefficients were observed with the interactions of Truth in Sentencing and Determinate Sentencing (-44.232, $p<0.001$). There is also a negative coefficient between truth in sentencing and three strikes (-47.423, $p<0.001$).

Among the control variables, a positive and statistically significant coefficient was observed for the violent crime rate (0.062, $p < 0.000$), indicating that higher violent crime rates were associated with increased incarceration rates. Similar to model 7, both the percentage of Black residents (2.269, $p < 0.002$) and the percentage of Hispanic residents (12.117, $p < 0.000$) exhibited positive and significant coefficients, indicating that higher proportions of Black and Hispanic populations within a jurisdiction were associated with higher incarceration rates.

The socioeconomic variables had a notable influence on incarceration rates. The model revealed that an increase in unemployment rates correlated with a decrease in incarceration rates, as evidenced by a significant negative coefficient of -4.465 ($p < 0.001$) and a standardized beta of -0.0634, indicating a moderate negative influence when adjusted for scale differences among the predictors. Similarly, elevated poverty levels, defined as the percentage of the population living below the poverty line, were also linked to reduced incarceration rates, with a significant negative coefficient of -3.485 ($p < 0.001$) and a standardized beta of -0.0653, suggesting a similar moderate negative impact on incarceration rates. Conversely, income inequality, as measured by the Gini coefficient, was associated with increased incarceration rates, demonstrated by a positive and statistically significant coefficient of 1239.524 ($p < 0.000$) and a notably high standardized beta of 0.4711, which underscores its strong influence in comparison to other variables in the model.

Demographic variables also showed significant effects: the percentage of Black individuals in the population had a coefficient of 4.246 ($p < 0.000$) with a standardized beta of 0.2349, indicating a substantial positive impact on incarceration rates. Similarly, the percentage of Hispanic individuals had an even more pronounced effect, with a coefficient of 8.985 ($p < 0.000$) and the highest standardized beta of 0.4814 among the predictors, reflecting its significant influence on incarceration rates. These findings highlight the complex interplay between race, economic factors, and incarceration, emphasizing how demographic and socioeconomic conditions collectively shape social outcomes.

A Hausman test was conducted for Models 7 and 8 to address the choice between fixed-effects and random-effects models. The results suggest that using fixed effects is best, indicating that the individual state-specific effects are not random ($Chi^2(25)=102.56$, $P<0.000$).

Regime Score Models: No Controls and Regime Scores (Models 9 and 10)

In Table 8, Models 9 and 10 employ random and fixed effects using regime scores and no control variables. Using regime scores in these models represents a different approach to examining the impact of sentencing reforms on imprisonment rates than the fully saturated models that use interaction terms. Both approaches attempt to capture the interaction or co-impacts that reforms can have, but regime scores aggregate multiple reforms into a single index. These indexes allow for a comprehensive assessment of their combined effects but do not measure the specific individual impacts of the reforms or specific interactions. The lack of unique impacts can have some notable disadvantages, namely the loss of the dynamic modeling of the interactions. On the one hand, the regime score approach offers several advantages, including simplifying the analysis by reducing the number of variables and simplifying the interpretation while capturing the interactive relationships between reforms.

Model 9 incorporates random effects and reveals significant associations between regime scores representing different sentencing reforms and combinations between those reforms. The model was found statistically significant (Wald $Chi^2(27) = 1632.83$, $p<0.000$), and the R^2 value indicates that approximately 23.9% of the variability in incarceration rates is explained by the model. All individual-level reforms had significant and positive associations with incarceration rates, unlike those in Models 7 and 8. Among

Table 8. Results of the panel analysis models with regime scores, no controls, random effects, and fixed effects.

the individual reforms, truth in sentencing (179.634, $p<0.000$) and three strikes laws (162.211, $p<0.000$) exhibit the highest coefficients, indicating strong positive associations with incarceration rates compared to states with no reforms. Determinate sentencing and statutory sentencing have the following highest coefficients of 102.634 ($p<0.000$) and 145.531 ($p<0.000$), respectively. The impacts of individual-level reforms on imprisonment rates are assessed independently, without the moderating or amplifying influences of other reforms typically considered in fully saturated models.

Testing Interaction Effects Using Different Panel Model Specifications

	Model 9		Model 10	
Individual Reforms	Random Effects	S.E.	Fixed Effects	S.E.
Pres. Sent. Guide. (PSG)	95.207***	(28.664)	99.420***	(28.974)
Voluntary Sent Guide. (VSG)	99.054***	(14.458)	99.596***	(14.599)
Statutory Presump. Sent. (SPS)	145.531***	(25.721)	165.845***	(28.920)
Determinate Sentencing (DS)	102.634***	(16.265)	109.918***	(16.609)
Truth in Sentencing (TIS)	179.634***	(13.122)	179.907***	(13.196)
Three Strikes (3STK)	162.211***	(15.790)	162.033***	(15.881)
Reforms in combination	-	-	-	-
PSG + DS	95.971	(62.504)	95.740	(62.649)
VSG + DS	-241.215***	(37.796)	-243.279***	(38.161)
SPS + DS	14.855	(40.205)	40.546	(49.055)
PSG + TIS	247.805***	(28.694)	249.760***	(28.883)
VSG + TIS	(omitted)	-	(omitted)	-
SPS + TIS	(omitted)	-	(omitted)	-
PSG + 3STK	125.817	(88.210)	128.937	(88.415)
VSG + 3STK	-62.459**	(24.268)	-64.343**	(24.406)
SPS + 3STK	316.327***	(35.743)	339.350***	(39.294)
DS + TIS	241.326***	(18.958)	249.365***	(19.340)
DS + 3STK	-370.849***	(38.734)	-375.370***	(39.157)
TIS + 3STK	250.976***	(18.193)	250.552***	(18.327)
PSG + DS + TIS	125.816***	(18.534)	126.592***	(18.691)
VSG + DS +TIS	-9.670	(34.142)	-12.581	(34.513)
SPS + DS + TIS	369.115***	(38.779)	389.192***	(41.248)
PSG + DS + 3STK	(omitted)	-	(omitted)	-
VSG + DS + 3STK	38.716	(36.539)	40.365	(36.713)
SPS + DS + 3STK	165.238***	(44.805)	191.724***	(53.728)
PSG + TIS + 3STK	260.966***	(25.193)	265.211***	(25.556)
VSG + TIS + 3STK	109.567***	(24.453)	110.898***	(24.556)
SPS + TIS + 3STK	283.785***	(40.532)	307.718***	(43.144)
PSG + DS + TIS + 3STK	176.366***	(16.945)	178.009***	(17.151)
VSG + DS + TIS + 3STK	-56.688	(35.522)	-58.646	(35.774)
SPS + DS + TIS + 3STK	255.393***	(49.803)	280.479***	(58.026)
Constant	183.233***	(18.056)	178.813***	(6.036)
Observations	1550	-	1550	-
F Statistic	-	-	60.67	-
Chi2	1632.83	-	-	-
R-squared	0.239	-	0.228	-

Standard errors in parentheses.

***p < 0.001, **p < 0.01, *p < 0.05.

The results from combining multiple reforms aggregated into distinct regimes are considered next. Most of the reform indexes that were found to be significant were positive. The joint between voluntary and three-strikes and between determinate and three-strikes had negative coefficients of -62.459 ($p<0.010$) and -370.849 ($p<0.000$), respectively. These findings suggest a potential decrease in incarceration rates associated with voluntary and determinate sentencing reforms alongside three-strikes laws compared to indeterminate states. In contrast, Model 8, which does not incorporate regime scores but uses interaction terms, presents a different dynamic. Here, a positive coefficient for the interaction between voluntary sentencing and three-strikes laws suggests that in the absence of regime scores, these policies might increase incarceration rates when combined, highlighting a divergent impact from the model employing regime scores.

Voluntary sentencing and determinate sentencing reforms often prioritize judicial discretion and flexibility in sentencing, aiming to adapt penalties to the circumstances of individual cases, which may decrease incarceration rates. Conversely, three-strikes laws are characterized by stringent mandatory sentencing requirements, particularly targeting repeat offenders with escalating severity. (Stemen et al., 2006). Model 8, the fully saturated model, offers a more nuanced representation of how these policies interact to influence incarceration rates, suggesting that the interplay of reforms might be more complex than what is captured by regime scores alone.

Other impacts are observed with the conjunction between voluntary and determinate sentencing, with a negative coefficient of -241.215 ($p<0.000$). Interestingly, the combination of voluntary, three strikes, and determinate sentencing did not have statistical significance. The combination of statutory, determinate, and three strikes found a significant and positive coefficient of 165.238 (p<0.000), which may suggest that there is a combined impact with statutory sentencing that may increase incarceration rates compared to indeterminate states. This is also seen with the interaction between statutory and three strikes, with a positive coefficient of 316.327 (p<0.000).

The positive coefficient indicates an increase in imprisonment rates when presumptive sentencing guidelines and truth in sentencing are combined (247.805, p<0.000). Determinate and truth in sentencing and the combination of truth in sentencing and three strikes also saw positive coefficients of 241.326 (p<0.000) and 250.976 (p<0.000), respectively, which also suggests an increase in imprisonment rates when compared to indeterminate sentencing states. The combination of presumptive, truth in sentencing, and determinate sentencing further explains this influence with a positive coefficient of 125.816 (p<0.000). Additionally, presumptive, determinate, truth in sentencing and three strikes combined had a positive coefficient of 176.366 (p<0.000), indicating an increase in imprisonment rates when these sentencing reforms are combined compared to states with no reforms.

Model 10 utilized fixed effects and was found statistically significant (F = 60.67, p<0.000). The R-^{2} value indicates that the model explains 22.8% (R^2=.228) of the variability in incarceration rates and is very similar to the value in Model 9. Similarly to Model 9, this model exhibits significant positive coefficients for all individual-level reforms. In Model 10, a similar trend is observed concerning the combination of voluntary and determinate sentencing reforms alongside three-strikes laws. The negative coefficient obtained for the interaction between voluntary and three strikes (-64.343, p<0.008) and the joint impact between determinate and three strikes (-375.370, p<0.000) echoes the findings of Model 9. Additionally, the interaction between statutory, determinate, and three-strikes yielded a significant positive coefficient of 191.724 (p<0.000), further supporting the notion that the combined effect of these sentencing reforms increases imprisonment rates compared to no reform states. There were few differences between the two models, but Model 10 had slightly higher coefficients for many variables.

Regime Score Models: Controls and Regime Scores (Models 11 and 12)

In Table 9, Models 11 and 12 extend the analysis by incorporating regime scores alongside control variables. The integration of regime scores allows for a comprehensive assessment of the combined effects of sentencing reforms. At the same time, including control variables enables a more nuanced exploration of the factors influencing imprisonment rates, enriching the analysis with contextual insights.

Testing Interaction Effects Using Different Panel Model Specifications

Table 9. Results of the panel analysis models with regime scores, controls, random effects, and fixed effects.

	Model 11		Model 12	
Individual Reforms	Random Effects	S.E.	Fixed Effects	S.E.
Pres. Sent. Guide. (PSG)	-3.632	(20.991)	-3.370	(20.861)
Voluntary Sent Guide. (VSG)	18.880	(11.492)	20.187	(11.418)
Statutory Presump. Sent. (SPS)	-4.420	(20.638)	34.317	(24.832)
Determinate Sentencing (DS)	46.002***	(12.543)	49.932***	(12.702)
Truth in Sentencing (TIS)	94.239***	(10.873)	90.244***	(10.737)
Three Strikes (3STK)	47.788***	(13.411)	36.104**	(13.284)
Reforms in combination	-	-	-	-
PSG + DS	-18.553	(45.394)	-21.692	(44.513)
VSG + DS	-137.410***	(28.927)	-130.585***	(28.720)
SPS + DS	-86.307**	(28.487)	-22.486	(36.524)
PSG + TIS	84.705***	(23.369)	82.591***	(23.094)
VSG + TIS	(omitted)	-	(omitted)	-
SPS + TIS	(omitted)	-	(omitted)	-
PSG + 3STK	0.066	(63.767)	0.437	(62.504)
VSG + 3STK	-41.844*	(18.938)	-41.659*	(18.685)
SPS + 3STK	72.901**	(28.021)	113.330***	(31.434)
DS + TIS	80.725***	(15.627)	81.141***	(15.771)
DS + 3STK	-123.019***	(31.308)	-125.574***	(31.091)
TIS + 3STK	90.032***	(15.160)	85.580***	(14.974)
PSG + DS + TIS	4.041	(15.053)	-3.381	(14.943)
VSG + DS +TIS	-16.536	(25.897)	-18.830	(25.695)
SPS + DS + TIS	92.000**	(32.068)	119.624***	(34.597)
PSG + DS + 3STK	(omitted)	-	(omitted)	-
VSG + DS + 3STK	47.490	(32.763)	40.086	(32.214)
SPS + DS + 3STK	-70.895*	(33.496)	-18.470	(40.606)
PSG + TIS + 3STK	66.051***	(20.294)	65.709***	(20.247)
VSG + TIS + 3STK	80.389***	(19.368)	78.166***	(19.043)
SPS + TIS + 3STK	8.404	(33.512)	45.908	(36.260)

Table 9 Continued

PSG + DS + TIS + 3STK	-18.788	(15.520)	-28.956	(15.514)
VSG + DS + TIS + 3STK	-66.785*	(28.394)	-75.141**	(28.041)
SPS + DS + TIS + 3STK	-25.170	(38.494)	23.871	(44.504)
Crime Controls	-	-	-	-
Violent Crime	0.069***	(0.013)	0.065***	(0.013)
Drug Crime	-0.001	(0.005)	0.001	(0.005)
Demographic Controls	-	-	-	-
Percent Black	3.977***	(0.667)	2.258**	(0.757)
Percent Hispanic	9.157***	(0.941)	11.597***	(1.121)
Percent Unemployment	-4.521***	(1.393)	-4.019**	(1.385)
Percent Poor	-3.250**	(1.045)	-3.920***	(1.062)
Percent Urban	-0.062	(0.120)	-0.093	(0.117)
Gini Coefficient	1291.172***	(59.685)	1233.516***	(60.306)
FHREP^	12.539***	(1.555)	14.286***	(1.602)
Constant	-502.435***	(35.677)	-461.341***	(34.151)
Observations	1359	-	1359	-
F Statistic	-	-	102.31	-
Chi2	3516.43	-	-	-
R-squared	0.483	-	0.345	-

Standard errors in parentheses. ^ FHREP represents GOP control of the state house.
***p < 0.001, **p < 0.01, *p < 0.05.

Model 11 employed random effects and introduced regime scores along with control variables. The overall model was significant (Wald Chi^2(36) = 3,516.43, p<0.000), and 48.3% (R^2 =.483) of the variability in incarceration rates is explained by the model. Model 11 reveals intriguing shifts in the significance of individual-level reforms. Determinate sentencing, truth in sentencing, and three strikes emerge as the only reforms reaching statistical significance. However, their coefficients are notably less pronounced than Models 9 and 10, underscoring the nuanced interplay between reforms and control variables. For example, Determinate sentencing has a coefficient of 46.002 ($p<0.000$) in this model, but in Model 9, the coefficient was much higher (102.634). This is most likely due to the introduction of controls.

Of particular interest is the combination of three-strikes laws and other sentencing policies. In Model 11, the negative coefficient (-123.019, $p<0.00$) associated with the combined effect between determinate sentencing and three strikes suggests a potential mitigating effect, indicating that determinate sentencing may temper the severity of sentences imposed under three-strikes laws. On these same lines, voluntary sentencing and three strikes are significant, with a negative coefficient of -41.844 ($p<0.00$). This finding aligns with the notion that combining discretionary sentencing measures with mandatory sentencing requirements could lead to more nuanced and balanced outcomes in sentencing decisions.

Furthermore, the positive coefficient (72.901, $p<0.00$) observed for the index with statutory and three-strikes laws highlights a unique effect. This result implies that the simultaneous implementation of statutory sentencing guidelines alongside three-strikes laws may have an additive impact on incarceration rates, potentially reflecting a more punitive approach to sentencing. The combined interaction between presumptive sentencing, truth in sentencing, and three strikes, along with the combination of voluntary sentencing, truth in sentencing, and three strikes, both exhibit positive coefficients

(66.051, $p<0.00$ and 80.389, $p<0.00$, respectively). These findings suggest that these combined reforms lead to higher incarceration rates compared to scenarios where only one or two of these reforms are implemented.

The combined influence of statutory and determinate sentencing is significant in this model, unlike Models 9 and 10, and has a negative coefficient of -86.307 ($p<0.00$). Notably, the combination of voluntary sentencing, determinate sentencing, truth in sentencing, and three strikes emerges as significant, with a negative coefficient of -66.785, emphasizing the importance of considering the combined effects of multiple reforms in shaping incarceration rates.

Among the control variables, a positive and statistically significant relationship was observed for the violent crime rate (0.069, $p < 0.000$), indicating that higher rates of violent crime were linked to increased incarceration rates. Similarly, both the percentage of Black residents (3.977, $p < 0.000$) and the percentage of Hispanic residents (9.157, $p < 0.000$) demonstrated positive and significant coefficients, suggesting that higher proportions of Black and Hispanic populations within a jurisdiction were associated with higher rates of incarceration. The standardized beta coefficients for Black and Hispanic populations were 0.220 and 0.491, respectively, underscoring the substantial influence these demographic factors have on incarceration rates.

Socioeconomic factors played a pivotal role in influencing incarceration rates as well. The negative and statistically significant coefficient for the unemployment rate (-4.521, $p < 0.001$) indicated that higher unemployment rates corresponded to lower incarceration rates, with a standardized beta of -0.057, suggesting a moderate negative influence when scaled. Conversely, higher poverty levels, as indicated by the percentage of the population living below the poverty line, exhibited a negative and statistically significant coefficient (-3.250, $p < 0.002$), suggesting that increased poverty levels were linked to lower incarceration rates. The standardized beta for poverty was -0.079, emphasizing its comparable impact on incarceration rates. However, as measured by the Gini coefficient, income inequality displayed a positive and highly significant coefficient (1291.172, $p < 0.000$) with a standardized beta of 0.463, highlighting its strong positive effect on increasing incarceration rates, further indicating that economic disparity is a crucial determinant in the dynamics of incarceration.

Model 12 employs fixed effects, regime scores, and control variables, offering further insights into the intricate dynamics shaping imprisonment rates. The overall model was significant ($F = 102.31$, $p < 0.000$), and the model explained 34.5% ($R^2 =.345$) of the variability for incarceration rates, which was slightly lower than Model 11 at 48.3%. In terms of coefficients and significance levels, Models 11 and 12 are similar. Among the individual-level reforms, determinate sentencing, truth in sentencing, and three strikes laws were found significant with positive coefficients. The composite effects between statutory sentencing and three-strike laws had a higher coefficient (113.330, $p < 0.000$) and a lower p-value than in Model 11 (72.901, $p < 0.009$). The interaction between statutory, determinate, and truth in sentencing also had a higher coefficient and lower p-value (119.624, $p < 0.001$) than in Model 11. Notably, the higher coefficients and lower p-values for regimes in this model suggest stronger and more precise associations between specific combinations of sentencing reforms.

The control variables also followed similar patterns to Model 11. The violent crime rate had a significant and positive coefficient (0.062, $p < 0.000$), with a standardized beta of 0.100, suggesting a noticeable influence on incarceration rates. The coefficients for the percentage of Black (2.269, $p < 0.000$) and Hispanic (12.117, $p < 0.000$) residents may indicate increased incarceration rates. The standardized beta for Black residents was 0.125, and for Hispanic residents, it was a substantial 0.621. This indicates that while both are positively associated with incarceration rates, the impact of Hispanic residents is significantly stronger in this model. The coefficient for the percentage of Black residents has a lower value and higher p-value in Model 12 compared to Model 11, which suggests a potentially less significant association between the proportion of Black residents and incarceration rates in Model 12. Conversely, the higher coefficient and the larger standardized beta for Hispanic residents in Model 12 compared to Model 11 indicate a stronger association between the proportion of Hispanic residents and incarceration rates in Model 12. Other control variables, including the unemployment rate, percent of the population living below the poverty line, and income inequality, were significant with coefficients and impacts similar to those of Model 11, with standardized betas of -0.050 for unemployment, -0.095 for poverty, and a strong 0.443 for income inequality, underlining the consistent negative impact of socioeconomic disadvantages and the exacerbating effect of income inequality on incarceration rates.

Testing Interaction Effects Using Different Panel Model Specifications

A Hausman test was conducted for Models 11 and 12 to address the choice between fixed-effects and random-effects models. The results suggest that using fixed effects is best, indicating that the individual state-specific effects are not random ($Chi^2(34) = 109.52$, $p<0.000$).

Margins Scores Models 1 and 2

Table 9 presents the calculated impact of combinations of reforms of two models that incorporated the random effects and included the control variables. These models are derived from Model 7 and Model 11 from Figures 6 and 8, respectively. The calculated impacts offer valuable insights into the predicted imprisonment rates for the average state under different combinations of sentencing reforms and control variables and allow for the comparison of results when using the two different methods of modeling the interactions between reforms. Using the margin command, the expected imprisonment rates were calculated based on the coefficient estimates obtained from the regression analyses. Notably, Model 7 incorporates full interaction terms to capture the potential synergistic effects between sentencing reforms, while Model 11 utilizes regime scores to represent the combined impact of reforms. The comparison between these models sheds light on the efficacy of these two approaches in predicting incarceration rates. The margin command proves invaluable in comparing the two interaction term approaches by enabling the calculation of the expected imprisonment rate for the average state under each model's specifications. This is achieved by solving the equation derived from the regression models, substituting the respective coefficient values, and setting control variables at their mean values. Through this process, predicted imprisonment rates are derived for different combinations of sentencing reforms, allowing for a direct comparison of the two models' outcomes. By quantifying the expected effects of policy interventions in both models, the margin command facilitates a comprehensive assessment of how the inclusion of interaction terms versus regime scores influences the predicted incarceration rates for the average state. For example, suppose one model has a higher predicted imprisonment rate than another. In that case, it suggests that the higher score predicts a greater change in the incarceration rate for the same change in sentencing policies. However, it is important to note that higher margin scores do not indicate a better model. Instead, it is important to note substantial differences in the scores between these models. This is concerning, considering some indicate that a particular combination of reforms increases the imprisonment rate in one model while the same combination in another model decreases the rate.

Comparing the margin scores in Table 9 Models 1 and 2, both models have essentially the same R^2 value, with the models explaining 48.3% (R^2 =.483) of the variability in incarceration rates. The Chi2 statistic was slightly higher in the regime score model (Model 11) than the interaction term model (Model 7) by 1.1 times, suggesting the regime score model is about a 10% better fit. Indeterminate sentencing had similar margin scores in both models, indicating that it was getting around 250 inmates per 100,000 per year. Values less than 250 indicate a reduction in imprisonment compared to indeterminate sentencing and vice versa. In Models 1 and 2, the three front-end reforms, presumptive, voluntary, and statutory sentencing by themselves, were insignificant, as depicted in Models 7 and 11. The three back-end reforms, truth in sentencing, determinate sentencing, and three strikes, were significant in both these models. Among the individual-level reforms, margin scores are higher in Model 2, suggesting a potentially greater influence on incarceration rates within this model's framework. Although three strikes demonstrate a slightly higher margin score in Model 1, indicating varied sensitivity across models.

When considering combined reforms, the two models substantially differ in margin scores. For example, for the combination of voluntary sentencing and determinate, Model 1 had a margin score of 249.257, 2.2 times higher than Model 2 (112.270). Both were found significant in their respective models (7 and 11). The same goes for voluntary sentencing in combination with three strikes, where the score in Model 1 was 1.6 times that in Model 2. Determinate and three strikes together scored 2.5 times higher than the margin score in Model 2. However, the combination of reforms was only significant in the regime score model (Model 11).

Interestingly, reform combinations like statutory paired with three strikes exhibit higher margin scores in Model 2 than in Model 1 by about 1.2 times. Similarly, presumptive and determinate sentencing scored 1.1 times higher in Model 2 (231.127) than in Model 1 (213.140). The trend of higher scores between models varies for the combinations that included more than two reforms. For instance, presumptive, determinate, and truth in sentencing in Model 1 had a higher margin score by 1.2 times that of Model 2. A similar trend in sentencing is shown with voluntary, determinate, and truth in sentencing, where Model 1's score was 321.853 compared to Model 2's, 233.144, or 1.4 times higher. The higher margin score suggests that when these reforms are considered in interaction, their collective impact on incarceration rates is more pronounced in the model that accounts for full interaction terms. This could imply that these policies are not operating independently and may have a compound effect. On the other hand, Statutory, determinate, and truth in sentencing were higher in Model 2 by 1.3 times.

Table 10. Results of the Panel Analysis with margins command with random effects full interaction term model and regime score model.

	Model 1	Model 2
Individual Reforms	Random Effects with Full Interaction Terms	Random Effects with Regime Scores
Pres. Sent. Guide. (PSG)	244.614	246.049
Voluntary Sent Guide. (VSG)	252.969	268.560
Statutory Presump. Sent. (SPS)	236.353	245.260
Truth in Sentencing (TIS)	341.770^	343.920^
Determinate Sentencing (DS)	286.820^	295.683^
Three Strikes (3STK)	298.844^	297.469^
Reforms in combination	-	-
PSG + DS	213.140^	231.127
PSG + TIS	378.718	334.386^
PSG + 3STK	241.196	249.746
VSG + DS	249.257^	112.270^
VSG + TIS	367.934	(Omitted)
VSG + 3STK	348.434^	207.837^
SPS + DS	195.319^	163.374^
SPS + TIS	356.916	(Omitted)
SPS + 3STK	274.236	322.581^
DS + TIS	335.756^	330.406^
DS + 3STK	318.693	126.661^
TIS + 3STK	328.090^	339.713^

Table 10 Continued

PSG + DS + TIS	304.875^	253.721
VSG + DS +TIS	321.853^	233.144
SPS + DS + TIS	273.512^	341.680^
PSG + DS + 3STK	193.216^	(Omitted)
VSG + DS + 3STK	328.218^	297.171
SPS + DS + 3STK	216.697^	178.785^
PSG + TIS + 3STK	313.242^	315.732^
VSG + TIS + 3STK	401.342^	330.070^
SPS + TIS + 3STK	332.741^	258.085
PSG + DS + TIS + 3STK	222.894^	230.893
VSG + DS + TIS + 3STK	338.756^	182.896^
SPS + DS + TIS + 3STK	232.833^	224.511
Indeterminate Sentencing	250.466^	249.681^

^ means It was found not significant

This analysis indicates that the impact of individual reforms on incarceration rates is more pronounced in Model 2, while the interactions between different reforms generally have higher margin scores in Model 1. It is important to note that higher margin scores do not inherently indicate a 'better' model but suggest where each model may detect stronger relationships. The varying magnitudes and significance levels of these margin scores across the two models underscore the importance of considering both the analytical framework and the legal context of the reforms to draw comprehensive conclusions from the data.

Margins Scores Models 2 and 3

Table 10 compares the margin scores derived from Models 8 and 12, as depicted in Figures 7 and 9. The results in this table are from two models that utilized fixed effects and the control variables. Model 3 from Table 10 was derived from Model 8, which utilized full interaction terms, and Model 4 employed regime scores. The results present intriguing variations in the estimated effects of reforms on incarceration rates. It is important to note the differences in R^2 values between the two models. Model 12 explained 34.5% (R^2 =.345) of the variability in incarceration rates compared to Model 8, which explained a slightly lower percentage (31.5%). Overall, the differences in R^2 values between the two models suggest that the regime score model, Model 12, might be slightly more effective at explaining the variance in incarceration rates than the fully saturated model, Model 8, but this does not necessarily mean it is the superior model. Other factors, including the accurate representation of the reforms, the theoretical underpinning of the included variables, and the potential for overfitting, should also be considered when determining the best model for predicting or understanding changes in incarceration rates.

Indeterminate sentencing had similar margin scores between models 2 and 3; Indeterminate got around 243 and 244 inmates per 100,000 per year, respectively. Several observations can be made about the individual reforms and their combinations when analyzing margin scores in Models 3 and 4. Presumptive sentencing exhibits marginally higher scores in Model 4 (240.511) than in Model 3 (237.962), suggesting a slight decrease in the predicted impact on incarceration rates under the regime score framework compared to indeterminate sentencing. Both models found the impacts of Truth in Sentencing, Determinate Sentencing, and three strikes statistically significant from

Models 8 and 12. Model 4 consistently reported higher margin scores, signifying the difference between models.

Different trends are seen with the reforms in combination. For presumptive sentencing combined with determinate, the margin score is 1.1 times higher in Model 4 (222.19) than in Model 3 (202.790). In contrast, presumptive sentencing with truth in sentencing is significantly higher in Model 3, around 1.2 times than in Model 4. The voluntary and determinate sentencing combination is more than twice as high in Model 3 (243.796) as in Model 4 (113.296), both found significant in the original models. Statutory paired with three strikes is notably higher in Model 4 (357.211) than Model 3 (318.946), indicating that this combination's impact on imprisonment rates may be better captured in Model 4. For combinations involving all three reforms (PSG + DS + TIS, VSG + DS + TIS, SPS + DS +TIS), Model 3 consistently presents higher margin scores, except for SPS + DS + TIS, where Model 4 had a score 1.2 times higher than Model 3.

The statistical significance shown in the margin score models found in Models 8 and 12 indicates that certain reform combinations are robust predictors of incarceration rates. Notably, the combination of voluntary sentencing, truth in sentencing, and three strikes shows the highest margin score in Model 3 (393.775), 1.3 times higher than in Model 4 (322.048). Overall, the differences in margin scores between the two models reflect the distinct methodological approaches to capturing the relationships between sentencing reforms and incarceration rates.

Testing Interaction Effects Using Different Panel Model Specifications

Table 11. Results of the Panel Analysis with margins command with random effects full interaction term model and regime score model.

	Model 3	Model 4
Individual Reforms	Fixed Effects with Full Interaction Terms	Fixed Effects with Regime Scores
Pres. Sent. Guide. (PSG)	237.962	240.511
Voluntary Sent Guide. (VSG)	247.149	264.068
Statutory Presump. Sent. (SPS)	280.054	278.198
Truth in Sentencing (TIS)	328.332^	334.126^
Determinate Sentencing (DS)	285.094^	293.813^
Three Strikes (3STK)	278.251^	279.985^
Reforms in combination	-	-
PSG + DS	202.790^	222.19
PSG + TIS	377.399	326.473^
PSG + 3STK	213.058	244.319
VSG + DS	243.796^	113.296^
VSG + TIS	356.064	(Omitted)
VSG + 3STK	332.283^	202.222^
SPS + DS	266.129^	221.396
SPS + TIS	372.849	(Omitted)
SPS + 3STK	318.946	357.211^
DS + TIS	326.451^	325.022^
DS + 3STK	295.733	118.307^
TIS + 3STK	316.417^	329.462^
PSG + DS + TIS	297.994^	240.500
VSG + DS +TIS	308.479^	225.051
SPS + DS + TIS	314.692^	363.506^
PSG + DS + 3STK	153.018^	(Omitted)
VSG + DS + 3STK	304.062^	283.967
SPS + DS + 3STK	280.154^	225.411
PSG + TIS + 3STK	305.072^	309.591^
VSG + TIS + 3STK	393.775^	322.048^
SPS + TIS + 3STK	364.319^	289.789
PSG + DS + TIS + 3STK	200.801^	214.925
VSG + DS + TIS + 3STK	321.323^	168.740^
SPS + DS + TIS + 3STK	281.295^	267.752
Indeterminate Sentencing	242.743^	243.881^

^ means It was found to be significant

The full interaction terms in Model 8 may elucidate more complex relationships, while the regime score framework in Model 12 might offer insights into how distinct policy regimes influence incarceration rates. Considerations of the accurate representation of reforms in real-world data and the desired balance between detail and simplicity in analysis should inform the choice between models.

In sum, Model 4, which incorporates regime scores, tends to show higher margin scores for individual reforms. On the other hand, Model 3, the full interaction term model, often exhibited higher margin scores for combined reforms, suggesting that this model may be more attuned to the synergistic effects of policy interactions. The significant differences in margin scores for statistically significant reforms between the two models illustrate the importance of model selection because the results can suggest very different things.

It is important to highlight the different specifications when comparing Models 1 and 2 to Models 3 and 4. Models 1 and 2 were random effects models and assumed that the state effects were uncorrelated with the independent variables. Models 3 and 4 included fixed effects models, which consider unobserved heterogeneity by allowing each state to have its own intercept, effectively controlling all time-invariant characteristics of the states. The Hausman test was found significant in all the models associated with the margin models (7, 11, 8, and 12), which suggests that fixed effects were more appropriate. Given the Hausman test results, Models 3 and 4 findings are likely more reliable. This is because the unique characteristics of each state, which could influence the effect of sentencing policies, are controlled for, thus providing a clearer view of the causal relationship between sentencing reforms and changes in incarceration rates. This makes the fixed effects models particularly useful for policy analysis, where understanding the direct impact of policy changes is paramount. (Halaby, 2004)

Overall, these comparisons underscore the importance of considering the methodological approach when assessing the combined effects of sentencing reforms. The discrepancies observed between Models 2 and 3 emphasize the need for further exploration to interpret the factors driving these differences and to ensure the robustness of findings in empirical research on criminal justice policy.

Discussion

The results of this study shed light on the intricate dynamics of state-level sentencing reforms and their impact on imprisonment rates within the criminal justice system. The analysis systematically compares different specifications of panel models to ascertain the most accurate and statistically robust model. Various model specifications encompass FE and RE models and those with and without controls to address potential confounding variables. This study primarily aims to explore two distinct methodologies for generating interaction terms in statistical models. The approaches being compared include the fully saturated models that incorporate first-level interaction terms and the regime score model, each offering unique insights into the dynamics of policy interactions.

Summary of Key Findings

This study primarily explores the variations between different model specifications, highlighting how these differences impact understanding of the interactions among sentencing reforms and their effects on incarceration rates. This is why analyzing standard models with no interaction terms was important. A significant observation is the variability between models, especially those that exclude interaction terms, emphasizing the necessity to consider potential interactions among reforms. The analysis demonstrates that including control variables significantly affects model fit, as models that focus solely on individual-level reforms without interaction terms—both in fixed effects and random effects frameworks—exhibited the lowest R^2 values, suggesting they were less effective at explaining variations in incarceration rates.

Also, models incorporating interaction terms revealed substantial differences in the significance of individual reforms compared to standard models, underscoring the importance of these terms in capturing the complex dynamics between reforms. The advanced models, particularly the fully saturated and regime score models, identified three back-end reforms—truth in sentencing, determinate sentencing, and three strikes laws—significantly impacting incarceration rates. There was a notable variability in findings between fully saturated and regime score models. Within the fixed effects framework, the fully saturated model detected 21 reforms as statistically significant, in contrast to only 14 in the regime score model, highlighting the varying capabilities of these models in detecting significant effects. Notably, the regime model omitted three variables due to the absence of specific combinations of reforms in any state. This is a significant limitation of the fully saturated model because it includes the interaction of reforms not present in the real world.

Model Specifications

The analysis begins with the standard models with no interaction terms depicted in Models 1 through 4. Relying solely on the standard model to analyze the dynamics of sentencing reforms and their impacts on imprisonment rates has significant limitations. The standard model may not fully capture the intricate dynamics of the criminal justice system and the combination of reforms. The timing and sequencing of reforms and contextual factors such as political climate and engagement can significantly influence their effectiveness. Moreover, while certain reforms may individually contribute to higher imprisonment rates, their combined effects may vary depending on other reforms and contextual factors. Examining the effects of individual reforms in isolation may not capture their full influence, as their interactions with other reforms can modify or amplify their effects. (Harmon, 2012)

With the introduction of control variables in Models 3 and 4, the explanatory power of the models significantly improved, emphasizing the importance of considering various factors influencing imprisonment rates. While certain reforms maintained their significance, such as voluntary sentencing, truth in sentencing, and three strikes laws, the inclusion of control variables revealed additional insights. For instance, demographic factors, economic indicators, urbanization, income inequality, and GOP control of the state (fhrep) were found to have significant associations with imprisonment rates, indicating the multifaceted nature of the criminal justice system. For example, when comparing models with and without controls, observations can be made that the models that included controls generally better fit the data, as evidenced by higher R^2 values than those without controls. For instance, the R^2 value of 0.483 for Model 7, the fully saturated model with controls, demonstrates that the model accounts for approximately 48.30% of the variation in incarceration rates, attributed to the combination of sentencing reforms, interaction terms, and control variables. This represents a substantial improvement over Models 5 and 6, which did not include controls and exhibited lower R-squared values of 0.221 and 0.196, respectively. This underscores the importance of including relevant controls in policy impact analysis to account for potential confounding variables that could influence the outcomes. The investigation extends to incorporating interaction terms within the fully saturated and regime score models, facilitating a nuanced exploration of the collective impacts of diverse sentencing reforms.

The primary focus of the study is to analyze two different ways to generate interaction terms in a model. The two approaches analyzed include the fully saturated models with first-level interaction terms and the regime score model. This nuanced approach allows for a deeper understanding of how specific reforms interact with one another and the broader legal and social environment. The regime scores model and the fully saturated model (with interaction terms) provide two distinct approaches to analyzing policy impacts. The regime scores model aggregates reforms into broader categories or 'regimes,' which allows for examining their combined effects on incarceration rates. (Jaccard & Turrisi, 2003) In contrast, the fully saturated model accounts for the possibility that the impact of one reform may depend on the presence or absence of another (Harmon, 2011).

Throughout the analysis, fixed effects models consistently performed better, as indicated by the Hausman test. In the analysis, it was crucial to differentiate between Models 7 and 11, which utilized random effects, and Models 8 and 12, which employed fixed effects. The random effects models posited that state-level effects were uncorrelated with the independent variables, an assumption not made in the fixed effects models. These fixed effects models account for unobserved heterogeneity by assigning unique intercepts to each state, thereby controlling for all time-invariant characteristics that could influence outcomes. (Halaby, 2004). While random effects configurations were calculated to illustrate the range of potential model setups, this approach likely needs to be revised for the present analysis. The significance of the Hausman test across all models suggests that the unique characteristics of each state, which could influence the impact of sentencing policies, need to be controlled to provide an accurate view of the relationship between sentencing reforms and changes in incarceration rates.

The significance of the Hausman test across regime and fully saturated models with controls (7, 11, 8, and 12) underscores the appropriateness of fixed effects models for the data, suggesting they provide a more accurate analysis of the causal relationships between sentencing reforms and incarceration rates. This finding reinforces the reliability of Models 8 and 12 for policy analysis, highlighting their ability to depict the direct impacts of policy changes more accurately on incarceration rates, crucial for formulating effective criminal justice policies.

As models evolved and controls were added, R^2 values increased, reflecting an improved model fit. Including controls can significantly alter the estimated effects of sentencing reforms on incarceration rates, indicating that these factors are important in explaining the variance in the dependent variable. It is imperative to illustrate the differences between model fit with the regime score model with controls (Model 12) and the fully saturated model with controls (Model 8). Model 12 explained 34.5% of the variance in incarceration rates, slightly more than the 31.5% explained by Model 8. While the higher R^2 value suggests that Model 12, which includes regime scores, might explain the variability in incarceration rates marginally better than the fully saturated Model 8, it does not inherently signify its superiority. Choosing an optimal model should also weigh other critical factors, such as the interpretability, the theoretical foundation of the variables used, and the risk of overfitting. These considerations are vital to ensure the chosen model fits the data well, aligns with the theoretical context, and provides meaningful insights for policy implications.

The fully saturated model reveals complex synergies or conflicts between policies that are not apparent when reforms are analyzed individually or aggregated into regimes (Jaccard & Turrisi, 2003). To effectively compare the two approaches, the margins command was applied, which quantified the expected changes in imprisonment rates under different policy conditions, directly comparing the fully saturated and regime score models. This comparison is particularly vital in highlighting the diverse impacts that similar policies can have under different operational frameworks.

Margin models 3 and 4 employ fixed effects within the fully saturated and regime score frameworks, respectively. Model 3 consistently exhibited higher margin scores for combinations of reforms than Model 4. This trend was particularly pronounced when focusing only on statistically significant reforms. Conversely, the regime score models demonstrated higher margin scores for individual-level reforms. It is crucial to recognize that the higher margin scores do not denote a 'better' model but rather indicate which model may detect stronger relationships within the data.

The variations in margin scores and the difference in statistical significance across these models are the key takeaways from the analysis. For example, in Model 3, the combination of VSG, DS, TIS, and 3STK had a margin score of 321.323, indicating an increase in imprisonment rates compared to indeterminate states (242.743) and Model 4 (168.740) indicated a decrease in imprisonment comparatively. This discrepancy poses a significant challenge for policymakers, as it indicates that the models offer contrasting perspectives on the outcomes of the reforms. Such conflicting results underscore the importance of careful model selection and interpretation in policy analysis to ensure that decisions are informed by a comprehensive understanding of how different analytical approaches might influence interpretations of the data.

The fully saturated model's ability to account for the potential that the impact of one reform could be contingent on the presence or absence of others allows it to uncover complex synergies or conflicts among reforms. These dynamics are crucial for understanding the nuanced ways in which policy measures interact within the criminal justice system, potentially leading to outcomes that are not immediately apparent through simpler analytical approaches. Such interactions might reveal that certain combinations of policies amplify or mitigate the effects of one another, insights that are critical for policymakers aiming to craft effective sentencing reforms. (Harmon, 2012) The differences between modeling approaches are depicted in Figures 1 and 2. These figures offer two different perspectives on modeling policy implementation over time. While the regime score approach condenses the information into broader categories, allowing for an examination of overarching policy regimes, the full interaction model offers a more granular view, analyzing the specific ways individual policies may interact over the years.

In analyzing policy impacts, the fully saturated model approximates real-world data dynamics more closely than the regime score model. This is mainly due to its detailed approach, which does not reset or start over with each combination of reforms, unlike the regime score model, which categorizes policy combinations into discrete regimes. The interaction term model operates with the nuanced understanding that policy impacts are cumulative and evolve over time. It allows each policy's effect to be analyzed in the context of its interaction with other policies, reflecting the continuous and often overlapping nature of legislative changes. For instance, when a new sentencing reform is introduced, the interaction term model considers its influence in conjunction with existing policies, capturing the subtleties of how these reforms might amplify or negate each other's effects.

In contrast, the regime score model segments the timeline into distinct periods or 'regimes,' each characterized by a specific set of policy combinations. While this approach can simplify the analysis, it assumes that introducing a new combination of policies represents a fresh start. This segmentation can overlook the historical context in which policies are enacted and their residual effects. This leads to a less accurate portrayal of how sentencing reforms influence incarceration rates over time. Acknowledging the continuous nature of policy enactment and reform interaction, the interaction term model avoids the artificial segmentation of the policy environment, offering a more dynamic and historically grounded representation. This quality makes it particularly adept at reflecting the complexities of policy implementation and its consequences, aligning more closely with the nature of legislative processes and their outcomes. Therefore, fully saturated models are the best method to analyze reform impacts on incarceration rates.

Furthermore, regime score models do not compute effects for non-existent combinations, ensuring that only viable, observable policy interactions are considered. (Jaccard & Turrisi, 2003) Certain reform combinations are omitted in these models because they did not exist in any state during the study period. This approach avoids over-speculation about potential interactions with no empirical basis, thus maintaining the integrity of the model's predictions and focusing analysis on actionable insights. This is a weakness of interaction models because they compute effects that do not exist, which is a problem when trying to represent the data accurately. If you are using the fully saturated model, it may not be appropriate to calculate these combinations of reforms that were not present; researchers should leave these out.

Implications

The significance of this research lies in the essential role of panel models with interaction terms in examining the intricate connection between sentencing reforms and imprisonment rates within the complex framework of the criminal justice system. The study aims to distinguish a robust framework of panel modeling for examining the impact of state-level sentencing reforms on imprisonment rates. The open-ended, hypothesis-free approach ensures unbiased and comprehensive analysis, contributing to understanding statistical modeling while addressing a pertinent issue in the criminal justice field.

This study has highlighted that sentencing reforms interact in complex ways, and understanding this is crucial for policymakers and researchers alike. Consideration of reforms' cumulative and sometimes unintended effects when crafting legislation is important. The evidence that different models yield varying predictions suggests that policymakers should approach reforms with a nuanced perspective, recognizing that changes in one area of legislation can have ripple effects across the system. The results suggest a need for data-driven policy development, where decisions are informed by robust statistical analysis rather than intuition or precedent alone (Halaby, 2004; Harmon, 2012).

This study provides an in-depth understanding of different model specifications of interaction terms in panel modeling. This is advantageous to researchers for several reasons. The study enhances the researcher's ability to choose the most appropriate analytical strategy by differentiating between models that include full interaction terms and ones that aggregate reforms into regime scores. It provides an understanding of how distinct modeling approaches can capture the complex interplay of reforms over time, assisting in developing more sophisticated analyses (Wooldridge, 2002). Second, understanding that these two model approaches have contrasting results highlights the importance of tailoring their analyses for the most accurate data representation. The guidance on what model specifications are best and why is helpful for researchers to make informed decisions that enhance the validity and reliability of their findings (Finkel, 1995). This deeper understanding of differing ways to implement interactions in a model can help bridge the gap between theoretical policy analysis and empirical evidence.

The justice system can benefit from incorporating more data-driven decision-making processes that consider the complex interactions of different sentencing policies. Understanding the mitigating effects of certain policies on incarceration rates can aid in developing strategies to avoid overly punitive measures that do not necessarily enhance public safety. Sentence reforms must be continuously evaluated to understand their longterm impacts, particularly as new legislation is introduced and the policy landscape evolves.

Limitations

While this study provides a comprehensive analysis of sentencing reform's impacts on incarceration rates, it is important to acknowledge its limitations. One significant aspect is the coding of reforms; the way sentencing reforms are defined and operationalized can vary, and different coding approaches could lead to divergent results. Legal nuances and the subtleties of policy implementation are challenging to encapsulate fully within a single coding scheme, suggesting that alternative approaches may create contrasting findings (see Harmon, 2011 and Stemen & Rengifo, 2011 for examples of contrasting approaches).

Additionally, the study's focus on imprisonment rates does not consider other critical dimensions of sentencing reforms, such as their effects on recidivism rates or overall crime levels. Such outcomes are essential for a holistic evaluation of the effectiveness of criminal justice policies but are beyond the scope of this analysis.

This research is also confined to looking at two ways interaction terms can be included in a model, as well as either fixed or random effects and those with or without controls. It does not incorporate other model specifications that should have been discussed but were not a part of the scope of the study. Certain specifications, like time trend data, could be relevant given the longitudinal nature of policy evolution.

Furthermore, the study does not examine the error structure in the data or investigate the possibility that control variables might interact. These interactions could provide additional insights into the complex relationships between demographic, economic, and policy variables (Wooldridge, 2002).

The exclusion of these considerations limits the study's ability to account for all the factors that may influence the relationship between sentencing reforms and incarceration rates. As a result, the findings presented should be viewed with the understanding that they represent a snapshot of the impacts based on the specific model specifications chosen for this analysis. Any substantive interpretation should be approached with caution.

Conclusion

The results of this study underline a fundamental principle in sentencing research: the choice of model significantly influences findings. Through the systematic examination of various model specifications, it is evident that model selection alters the interpretation of the impact of sentencing reforms on incarceration rates. Interestingly, the study finds that model fit statistics are relatively similar across full models with controls. This directs attention to the importance of conceptualizing and operationalizing the reforms under investigation. It is not merely a statistical decision but a substantive one, requiring a model that mirrors the realities of policy implementation and interaction. The fully saturated model, which includes fixed effects and control variables, is optimal.

This key finding is that how interaction terms are included in a model can lead to very different conclusions. Even if the fully saturated model had demonstrated less optimal fit statistics, which it did not, the crucial determinant for model choice should be its alignment with actual policy dynamics. While useful in specific contexts, the regime score model assumes a clean slate with each new set of reforms. This assumption does not necessarily hold true in the complex landscape of criminal justice policy, where changes are often incremental and layered upon existing structures (Baker, 2006; Raphael, 2009).

Ultimately, the objective is to utilize a modeling approach that accurately mirrors policy development's detailed and progressive nature. The fully saturated model, considering interaction effects, offers a closer representation of this reality than the regime score model, which may falsely segment ongoing policy progressions. Therefore, while statistical robustness is vital, the model's fidelity to the real-world processes it aims to represent is vital (Harmon, 2012). This study highlights the need for analytical rigor, emphasizing the importance of models that accurately represent the complex interactions of reform mechanisms. This approach is essential for understanding historical and current policy frameworks and guiding future criminal justice reform. Through the lens of a policymaker, it is really important to consider the complexity and context of policies and how they might interact. Policies are often looked at in isolation within policy analysis F. It urges researchers to continuously refine their methodologies to ensure their analyses authentically reflect the nuanced realities of policy implementation.

References

Andersen, T., Morash, M., & Chesney-Lind, M. (2011). Are girls getting tougher, or are we tougher on girls? Probability of arrest and juvenile court oversight in 1980 and 2000. *Justice Quarterly*, *28*(5), 719-744.

Barker V (2006). The politics of punishing—building a state governance theory of American imprisonment variation. *Punishment & Society: The International Journal of Penology, 8*(1), 5–32.

Beckett K, Western B (2001) Governing social marginality: welfare, incarceration, and the transformation of state policy. *Punishment & Society: The International Journal of Penology, 3,* 43–59.

Blumstein, A., & Beck, A. J. (1999). Population growth in U.S. prison populations, 1980–1996. In M. H. Tonry & J. Petersilia (Eds.), *Crime and justice: A review of research* (Vol. 26, pp. 17–61). University of Chicago Press.

Boppre, B., & Harmon, M. G. (2017). The Unintended consequences of sentencing reforms: Using social chain theory to examine racial disparities in Female imprisonment. *Journal of Ethnicity in Criminal Justice*, *15*(4), 394-423.

Brennan, P. K., & Spohn, C. (2008). Empirical research on the impact of sentencing reforms: recent studies of state and federal sentencing innovations. *Journal of Contemporary Criminal Justice*, *24*(4), 340–344. https://doi.org/10.1177/1043986208323265

Bureau of Justice Assistance. (1996). National Survey of State Sentencing Structures. Washington, DC

Bureau of Justice Statistics (1978-1983). Prisoners in State and Federal Institutions on December 31, 1983. Washington, DC

Bureau of Justice Statistics (1984-1998). Correctional Populations in the United States, 1984 (through 1998). Washington, DC

Bureau of Justice Statistics (1999-2008). Prisoners, 1999 (through 2008). Washington, DC

Bureau of Labor Statistics. (1972). U.S. Department of Labor. *Geographic Profile of Employment and Unemployment, 1972 (through 2008).*

Bureau of the Census (1978-2008). *United States Census (Corresponding Year).* Washington DC

Bureau of the Census (2008). Bureau of the Census, *Historical Poverty Tables, 2004 (on line)*. Washington DC

Dubin, M. J. (2007). *Party affiliations in the state legislatures: a year by year summary, 1796-2006*. Jefferson, N.C.: McFarland.

Finkel, S. E. (1995). *Causal analysis with panel data*. Thousand Oaks, CA: Sage Publications.

Frase, R. S. (2005a). Punishment purposes. *Stanford Law Review*, *58,* 67-83. Retrieved from http://www.jstor.org/stable/40040252

Frase, R. S. (2005b). State sentencing guidelines: Diversity, consensus, and unresolved policy issues. *Columbia Law Review, 105*, 1190–1232.

Halaby, C. (2004). Panel models in sociological research: Theory into practice. *Annual Review of Sociology, 30,* 507–544.

Harmon, M. G. (2011). The imprisonment race: Unintended consequences of "fixed" sentencing on people of color over time. *Journal of Ethnicity in Criminal Justice*, *9*(2), 79-109.

Harmon, M. G. (2012). "Fixed" sentencing: The effect on imprisonment rates over time. *Journal of Quantitative Criminology*, *29*, 369-397.

Harmon, M. G., & O'Brien, R. M. (2011). Gendered arrests or gendered sentencing: Explaining the narrowing of the gender gap in imprisonment over time: 1970–2008. *Sociological perspectives*, *54*(4), 641-664.

Hershey, M. R. (2007). *Party politics in America (12th ed.).* New York: Pearson Longman. Retrieved from http://www.loc.gov/catdir/toc/ecip067/2006002846.html

Hsiao, C. (2003). *Analysis of panel data (2nd ed.).* Cambridge, UK: Cambridge University Press.

Jaccard, James., & Turrisi, Robert. (2003). *Interaction effects in multiple regression (2nd ed.)*. SAGE.

Koons-Witt, B. A., Sevigny, E. L., Burrow, J. D., & Hester, R. (2014). Gender and sentencing outcomes in South Carolina: Examining the interactions with race, age, and offense type. *Criminal Justice Policy Review*, *25*, 299–324. doi:10.1177/0887403412468884.

Miethe, T. D., & Moore, C. A. (1988). Officials' reactions to sentencing guidelines. *Journal of Research in Crime and Delinquency*, *25*, 170–187.

Parker, R. N., & Horwitz, A. V. (1986). Unemployment, crime, and imprisonment: A panel approach. *Criminology, 24*, 751–774.

Raphael, S. (2009) Explaining the rise in U.S. incarceration rates. *Criminology and Public Policy*, *8*(1),87-95

Spelman, W. (2008). Specifying the relationship between crime and prisons. *Journal of Quantitative Criminology, 24*, 149–178. https://doi.org/10.1007/s10940-008-9042-x

Spelman, W. (2009). Crime, cash, and limited options: Explaining the prison boom. *Criminology & Public Policy, 8*(1), 29–77.

Spohn, C. (2000). Thirty years of sentencing reform: The quest for a racially neutral sentencing process. *Criminal Justice, 3*, 427-501. Retrieved from http://www.justicestudies.com/pubs/livelink3-1.pdf

StataCorp. 2023. *Stata Statistical Software*: *Release 18.* College Station, TX: StataCorp LLC.

Steffensmeier, D., & Haynie, D. (2000). Gender, structural disadvantage, and urban crime: Do macrosocial variables also explain female offending rates? *Criminology*, *38*(2), 403-438.

Steffensmeier, D., & Demuth, S. (2006). Does gender modify the effects of race–ethnicity on criminal sanctioning? Sentences for male and female white, black, and Hispanic defendants. *Journal of Quantitative Criminology*, *22*, 241-261.

Stemen, D. (2007). Impact of state sentencing policies on incarceration rates in the United States, 1975-2002. *Inter-university Consortium for Political and Social Research.*

Stemen, D., Rengifo, A., & Wilson, J. (2006). Of fragmentation and ferment: Impact of state sentencing policies on incarceration rates in the United States, 1975–2002. Washington, DC: National Institute of Justice.

Stemen, D., & Rengifo, A. F. (2011). Policies and imprisonment: The impact of structured sentencing and determinate sentencing on state incarceration rates, 1978–2004. *Justice Quarterly, 28*(1), 174-201.

Tonry, M. H. (1996). *Sentencing matters*. New York, NY: Oxford University Press.

Tonry, M. (2016). *Sentencing fragments: Penal reform in America, 1975-2025*. New York: Oxford University Press.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

Zhang, Y., Maxwell, C. D., & Vaughn, M. S. (2009). The impact of state sentencing policies on the US prison population. *Journal of Criminal Justice*, *37*, 190–199. https://doi.org/10.1016/j.jcrimjus.2009.02.012