Controlling Shootings by Focusing on Tail Risks

John MacDonald*
George Mohler†
P. Jeffrey Brantingham‡
Anthony A. Braga§

Abstract

Shootings are highly concentrated among a small number of people, places, and times, following a power-law distribution in which a small share of offenders and their networks accounts for most harm. Conventional violence-prevention frameworks underweight this structure, overlooking that the most consequential spikes arise from tail risk generated by the extreme end of the offender distribution. This paper develops an endogenous network model of gun violence with four components: a finite-mixture structure isolating a small latent high-risk subpopulation that drives baseline shooting intensity; a self-exciting Hawkes process in which past shootings elevate the probability of future shootings before decaying geometrically, producing temporal clustering and contagion; a deterrence-memory mechanism capturing the chilling effect of recent enforcement contact; and a gambler's-ruin process governing the structural path to incapacitation. Calibrated to Chicago shooting and arrest-network data, simulations demonstrate that modest increases in sanction certainty for high-risk individuals can substantially reduce shooting victims citywide. The model yields three policy implications. First, enforcement resources should concentrate on the high-risk subpopulation rather than be diluted across low-risk individuals. Second, speed matters: delayed enforcement allows excitation to compound before it decays, magnifying harm. Third, removing a high-risk individual from a network cools connected associates through the network's contagion lag, producing systemic relief that extends well beyond the individual case. Strategic interventions, such as focused deterrence, are therefore best understood as a form of tail-risk hedging: targeted, scalable prevention programs that stabilize public safety by neutralizing the small fraction of actors responsible for cascading violence.

Keywords: deterrence, chronic offending, Hawkes model, criminal careers

*Department of Criminology, University of Pennsylvania
†Department of Computer Science, Boston College
‡Department of Anthropology, University of California Los Angeles
§Department of Criminology, University of Pennsylvania

1 Introduction

Among the many criminal justice reforms across the U.S. over the past decade, one of the most prominent has been reducing the consequences of arrest. Diversion programs are a hallmark of criminal justice reform, based on the idea that avoiding a conviction can help mitigate the criminogenic effects of sanctions. Among felony defendants in the 75 largest U.S. counties in 2009, 6% of criminal charges for violence resulted in a formal diversion agreement (Reaves, 2013). Comparable national data for more recent years are not available, making it difficult to assess how much diversion has expanded since then. Given the wave of reform-minded policies adopted by district attorneys elected across major cities over the past decade (Mitchell & Petersen, 2025), diversion has likely grown substantially through explicit agreements as well as declinations and charge reductions that do not appear in official statistics. The criminal reform movement has embraced diversion from prosecution and reduced sentences for cases where incarceration is typical under state penal statutes (Davis, 2018), and this push has found empirical support in several studies presenting credible causal estimates that diverting first-time felony defendants charged with drug and property offenses lowers recidivism (Agan et al., 2023; Mueller-Smith & T. Schnepel, 2021).

The theoretical foundation for these reforms draws partly on the principle of marginal deterrence, which holds that the certainty of punishment, rather than its severity, is the operative margin (Nagin, 2013). Consistent with this view, the empirical literature finds little evidence that lengthier prison sentences deter criminal offending at the margin (Blumstein, 1998; Nagin, 2013). Assuming an equilibrium where everything else is held constant, it makes sense that the difference between serving 10 and 15 years in prison for an aggravated assault with a firearm may produce little benefit in general deterrence, if motivated offenders deciding to commit a shooting are generally indifferent to differences far off in the future on this scale. If chronic offenders are short-sighted, have a high discount rate, and are more risk averse, they will be more willing to tradeoff punishments that are far less certain in the distant future than those that are more certain and occur in closer proximity to the offense (Chalfin & McCrary, 2017). This theoretical logic and the growing body of evidence on the impact of police presence on crime has led some to argue that public safety benefits for hiring extra police outweigh the marginal benefits of prison sentences (Durlauf & Nagin, 2011).

Together, these strands of research suggest that reducing the punitiveness of the criminal justice system need not come at the cost of public safety, at least for the low-risk offenders most likely to be affected by diversion policies. Yet diversion and other core arguments for criminal justice reform fail to recognize the contingent relationship between the police, prosecutors, and court officials (Fryer, 2020). There is a feedback loop, such that police presence and certainty of apprehension are effective only when criminals face some nontrivial probability of sufficiently severe sanction. Social scientists who study crime control policy typically take the sanctioning environment as a given, but that assumption becomes problematic precisely when reforms alter both the certainty and severity of punishment.

Progressive reforms have been championed under the claim that reducing the consequences for the marginal offender will have little impact on public safety, and will accrue the benefits of reducing the scale of the criminal justice system (Mitchell & Petersen, 2025). Importantly, the goal is to end the era of "mass incarceration" (Travis et al., 2014). Research indicates that a primary driver of U.S. incarceration rates has been the sentencing of repeat and violent offenders to long prison terms, an effect that intensifies during periods of rising violent crime (Shen et al., 2020). Truth-in-sentencing and other reforms that limit or eliminate the use of parole as a release valve for prison populations mechanically extend prison sentences (Raphael & Stoll, 2013). It stands to reason then that progressive reforms would include directing line prosecutors to ask for the minimum necessary sentence for those convicted of crimes that face prison if their goal is to reduce the overall

incarceration rate.

However, lost in this policy discussion of progressive reform is the absence of any effort to triage where lengthy sentences are reduced while still holding the average probability of prison or jail constant. Progressive prosecutors in many locations have reduced the threshold for felony prosecution and categorically declined to prosecute offenses such as low-level drug dealing, illegal firearms carrying, and prostitution (Hogan, 2022). Some have championed a reduction in the use of arrests and across the board changes in sanctions (Harmon, 2016). This approach has normative justification when the goal is reducing punishment, particularly for first-time offenders unlikely to pose a public safety risk.

But the calculus changes for serious violence, and the central concern of this paper is how changes in sanctions regimes affect shootings specifically. In the case of deadly firearms violence policy should recognize that shootings behave like a *tail-risk* problem rather than an average-crime case.[1] In a population where offending follows a power-law or fat-tailed distribution (Sherman, 2007), most people living in disadvantaged neighborhoods suffering from high levels of gun violence are at low risk of committing a shooting, while a small subgroup of repeat offenders accounts for a disproportionate share of total firearm violence (the so-called "power few" or "fat tail") (Papachristos et al., 2012). Broad-based reductions in deterrence applied uniformly, such as lowering sanctions for the average offender for offenses like illegal carrying of firearms, may increase shootings because they do not address the tail risk of high-rate offenders that have a disproportionate effect on shootings. By contrast targeted approaches that raise the risk of sanctions for those most at-risk of shootings translates into a large reduction in the total count of shootings. As the pioneers of focused deterrence and swift-and-certain sanctions note, when it comes to "extreme public safety problems" like shootings, "a very small number of exceptional actors are at their core" (Kennedy et al., 2017, p. 161).

This logic has a close analogue in finance. Portfolio managers distinguish between strategies that protect against ordinary day-to-day fluctuations and strategies that specifically hedge against rare, extreme, capital-destroying events. The goal of tail risk hedging is to preserve capital and minimize drawdowns during market shocks. Simple hedges include holding long-term government bonds, cash, or gold reserves. A thoughtful crime-control portfolio should similarly consider sanction regimes that are predictable and scalable in protecting public safety against the ordinary case (most offenders, most of the time), while reserving a distinct strategy for the comparatively few individuals whose continued offending drives outsized social harm (Sherman, 2007). This is precisely the logic behind the most promising violence-reduction strategies, such as focused deterrence, which pairs direct messages of heightened legal risk with connections to jobs and community supports for the highest-rate offenders. Yet focused deterrence is only as effective as the accuracy with which high-rate violent offenders driving the tail are identified and prioritized for accountability (Braga et al., 2026).

This argument is not just a rhetorical framing device about policy. It has a direct implication for how a model of shooting rates in a city should be built and what a model can reveal about the design of effective mitigation strategies. This paper develops a mathematical model to formalize the tail-risk logic that the concentration of shooting risk among a small number of high-rate offenders is not an empirical curiosity but a structural feature of urban gun violence, and that crime control strategies targeting this tail yield disproportionate public safety returns. A model that structurally caps how extreme any one individual's offending can become will systematically understate both how concentrated violence already is and how large the benefit of incapacitating the tail can be. Correctly modeling shooting risk therefore requires building in the possibility of extreme individual-level concentration from the outset, and not treating it as an outlier or nuisance parameter.

[1] It is termed a "tail-risk" because of these"high-risk" individuals (residing) in the "tail" of the distribution.

In developing our modeling approach, we follow a long line of policy research on incapacitation and deterrence (Blumstein et al., 1978; Blusmtein, 1986; Nagin, 2013). As William Spelman (Spelman, 1993, p. 285) suggests, "if police and prosecutors knew how, they should focus most of their criminal investigation and prosecution resources on about 4 percent of offenders to increase the probability of incarceration given a crime by a factor of 30." Such intensive enforcement efforts could generate at least three beneficial crime control impacts: 1) the criminal incapacitation of active shooters who are also at great risk of being shot, 2) the deterrence of prospective shooters within tight-knit co-offending networks, and 3) the reduction of shooting opportunities through target removal and the disruption of ongoing conflicts. The results considerably safer neighborhoods without major increases in the incarceration rate. These arguments and an exploratory model are developed further in this paper, which is organized as follows. Section 2 establishes the empirical foundations on which the model rests. Section 3 outlines the model used to estimate shooting risk under different sanctions regimes (probabilities of deterrence). The concluding section discusses the results and explores policy implications.

2 Facts of Serious Criminal Offending and Criminal Careers

The existing literature points to six basic known facts of serious criminal offending and criminal careers that are important for understanding the dynamics of shootings that inform our model. A criminal career, or the longitudinal sequence of crimes committed by an individual, needs to be partitioned into those who participate at all in crimes and those who do so at a high rate (Blusmtein, 1986). Blumstein and colleagues note that while social services, education, and mental health interventions may reduce criminal participation, "controlling the rate of criminal activity is more central to the decisions of the criminal justice system" (Blusmtein, 1986, p. 13). Crime rates are also influenced by the supply of criminal opportunities in cities (Cook, 1986) and the effectiveness of opportunity reduction measures as offenders seek suitable targets in their immediate environments (Clarke, 1995).

First, serious violent offender and victim populations overlap extensively. Studies have long noted that homicide offenders and victims generally have the same social characteristics, share personal networks, and live in similar social settings (Wolfgang, 1958). This is particularly true for juveniles and young adults involved in violent offending and who suffer violence (Kennedy et al., 1996; Lauritsen et al., 1991). Today's shooter is often tomorrow's victim, and survivors frequently become perpetrators.

Second, there is little evidence that criminals specialize in offending (Farrington, 1986). While some evidence suggests that criminal offenders may have preferences, such that those engaged in burglaries may prefer property offenses that don't require confrontation with a victim, those that are most active in crime tend to commit a variety of offenses. The perpetrators and victims of gun violence often have extensive and varied criminal histories (Braga & Kennedy, 2021). However, despite being well known to the criminal justice system, mathematical models to predict those who only commit shootings or murders are generally exercises in futility. Predicting who will commit a shooting is difficult because violent offenders commit a wide range of offenses, and shootings for any individual will be a relatively rare event. Even machine learning models to predict rare events, like the propensity to commit a murder, are not particularly accurate (Berk et al., 2009). Predicting shooting victimization is similarly difficult, even with the number of victims exceeding the number of offenders, with the best-performing model for a study of Chicago correctly identifying only roughly one in six of the 327,181 individuals in the out-of-sample 18-month outcome period starting in 2018 as a true positive (Heller et al., 2026). Efforts that try and predict more common outcomes, like

any new offense, are much better at forecasting criminal behavior because these events are far more common (Berk & Bleich, 2013).

Third, criminal offending and in particular shooting offense is a highly skewed distribution with a small fraction of chronic offenders committing the majority of criminal offenses (Piquero et al., 2003). Depending on the source of data and sample, anywhere from 1 to 10% of offenders commit two-thirds of all serious criminal offenses. This means that within a pool of 100 criminal offenders arrested 1 to 10 of them could be responsible for 75% of all crime committed among that group of arrestees. For gun violence the numbers are even smaller, such that less than 1% offenders commit a vastly disproportionate share of shootings. This fact suggests that the biggest benefits of deterrence will come from impacting those at the highest risk for shootings and are responsible for the majority of cases (Braga et al., 2017).

Fourth, criminal offending in the population is highly age-graded, such that the majority of active criminal offending is occurring during early adulthood (Farrington, 1986; Piquero et al., 2003). But there is growing evidence that the age of the offenders arrested for shootings is less prone to decline with age, as the average age of individuals arrested for shootings in Chicago is between 2018 and 2024 was 29 years old. That means that shifts in the risk of sanctions for an arrest pool of people in their mid-20s to early-30s that are engaged in acts of violence is very consequential for public safety. Waiting for the pathway of desistance from criminal offending for this group is a threat to public safety.

Fifth, criminal offending is not generally an individual endeavor. The majority of offending among those committing burglaries, robberies, larceny, and shootings happen in small groups or networks of criminal offenders (Piquero et al., 2003). The group nature of criminal offending is particularly pronounced among younger offenders who commit violent acts like shootings (Papachristos et al., 2012). There are a few "lone wolves" who commit serious violent acts like robbery, rape, and murder, but offending in general tends to be a small group behavior. This fact suggests that contagion of behavior is also the most likely among the most active offenders whose friends and close associates are engaged in the same acts (Green et al., 2017). Contagion in the form of reductions in shooting behavior can also spread within high offending networks, a central logic of focused deterrence strategies (Braga et al., 2026).

Sixth, the most active shooters tend to be at the center of a network of friends and associates that are committing crimes (Charette & Papachristos, 2017; Nieto et al., 2022). Some of this network is transactional for property and drug offenses that require exchange of stolen goods or drugs. But other aspects of criminal networks are social and connected to local gangs, friendship networks, and families that are active in crime (Tracy et al., 2016). This fact suggests that vicarious experience is likely consequential, such that changes in sanctions regimes for shooting offenses will spread word within their networks.

Extensive research documents how these six facts of criminal offending characterize sudden outbreaks of gun violence in cities (Braga et al., 2024). For instance, during the COVID-19 pandemic in Boston, shooters and gunshot wound victims were overwhelmingly young black males involved in gangs. These gangs represented only 0.3% of Boston's residents but generated nearly half of all homicides and were involved in roughly 40% of all shootings between 2018 and 2022. Some gang homicides involved internal power struggles while others involved extended networks of gangs connected via ongoing conflicts and strategic alliances. Homicide offenders and victims had approximately 15 prior arraignments in Massachusetts courts for a variety of violent, illegal gun carrying, property, drug, and disorder offenses. The policy implication of the Boston research and similar studies conducted elsewhere in the US is clear: resources need to be focused on these high-risk people rather than diffused more broadly across city populations (Braga & Piza, 2023).

These six facts of criminal offending (population overlap; lack of specialization; chronic offenders;

semi-age-graded; group behavior; dense social network) suggest potential problems with public safety will occur if criminal justice reforms are not sufficiently focused on marginal deterrence for those at risk of committing extreme acts of violence like shootings.

2.1 The Convexity of Shootings and Enforcement Regimes

The principle of marginal deterrence, and selective incapacitation more generally, is also connected to some basic ideas about the power-law or fat tailed distribution (Taleb, 2020). On any given day, the number of active shooters in a city is small. Table 1, for example, shows for Chicago an arrest-network dataset that identifies hundreds of high-risk individuals at risk of shootings based on those arrested for a shooting and their co-offenders. Co-offending ties among these arrestees are spatially concentrated and persistent across years (Figure 1). Chicago gunshot violence follows an epidemic-like pattern transmitted through co-offending networks and social interactions (Papachristos et al., 2012). Therefore, when hundreds of high-risk individuals are densely networked around shootings, enforcement pullback affecting even a fraction of that population can allow shootings to cascade well beyond what the average deterrence effect would predict for the marginal offender.

Table 1: Individuals, Networks, Shooters, and Total Shootings by Year in Chicago

Year	Individuals in Networks	Shooter Networks	Shooters	Observed Shootings
2018	13,677	3,752	659	2,504
2019	13,975	3,779	680	2,221
2020	8,228	3,468	781	3,368
2021	6,032	3,084	645	3,637
2022	6,800	3,096	674	2,950
2023	7,402	3,371	622	2,540
2024	7,932	3,555	726	2,376

Notes: Annual counts of unique individuals belonging to social networks that include at least one identified shooter, the number of networks, the number of unique shooters, and the total observed shootings.

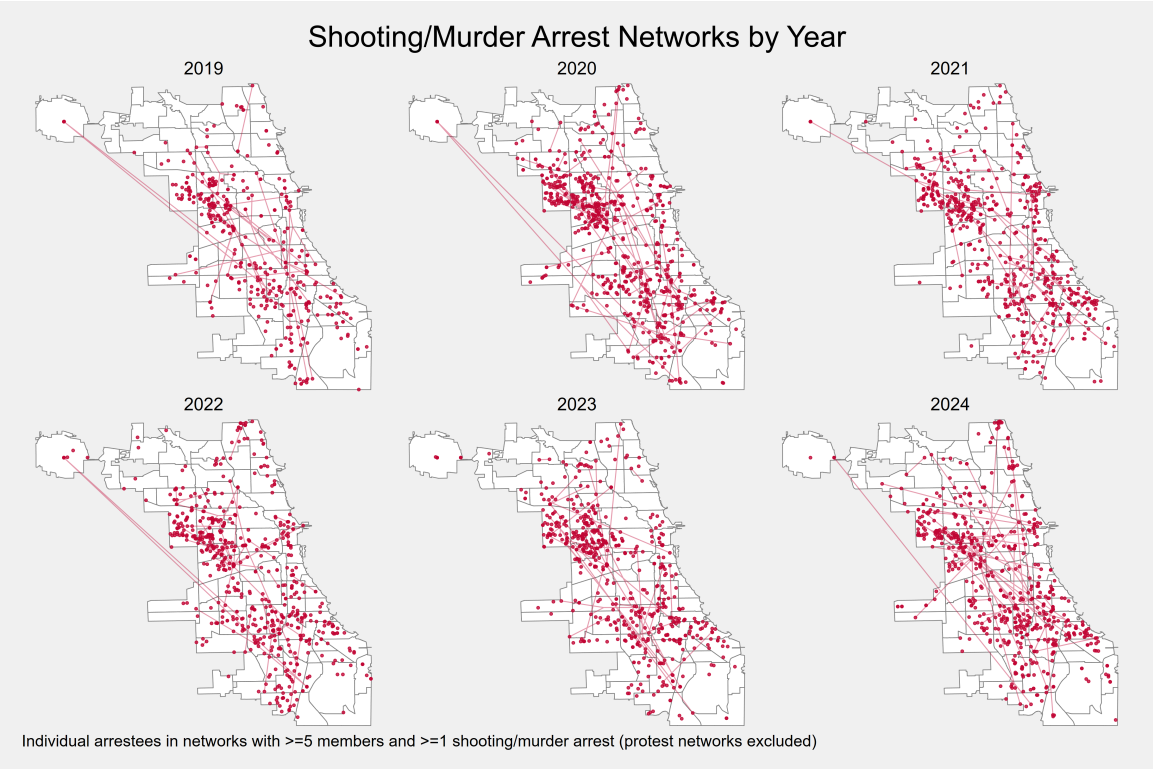


Figure 1: Shooting and Murder Arrest Co-offending Networks by Year, Chicago 2019–2024. Each point represents an individual arrestee in a network of $\geq$5 members with at least one shooting or murder arrest. Lines connect co-offenders within the same network.

One can explain shooting risk as a function of both the supply of offenders in the population at risk of committing a shooting and the "demand" for shootings under a given sanctions regime. The demand in this model is simply the prospective shooter's need to achieve some instrumental or expressive goal such as revenge, self-defense, or status. Following the standard Becker model of offending (Becker, 1968), the propensity to commit a shooting depends on the probability of punishment (p_t) and the severity of the sanction (S_t). [2] Let $\bar{\lambda}_t$ denote an individual's baseline propensity to commit a shooting absent any sanction, and let $S_t \in [0,1]$ denote the proportional reduction in the utility of offending when punishment occurs. The expected offending rate is then

$$\mathbb{E}[\lambda_t] \;=\; \bar{\lambda}_t\big[\,p_t(1-S_t)+(1-p_t)\,\big] \;=\; \bar{\lambda}_t\,(1-p_tS_t), \tag{1}$$

so that expected offending declines in both the certainty and the severity of punishment.

Rather than imposing structure through the functional form of equation (1), we state the two behavioral assumptions the carry forward in the analysis.

Assumption 1 (Convexity in enforcement)**.** *The offending rate is decreasing and convex in the certainty of punishment:* $\partial\lambda_t/\partial p_t < 0$ *and* $\partial^2\lambda_t/\partial p_t^2 > 0$.

[2] Cook and Nagin expand on this formulation and note that offenders respond not only to sanctions but also to the opportunities available to them (Cook, 1979; Nagin et al., 2015).

Convexity implies that the deterrent return to certainty is largest where enforcement is weakest: when p_t is low, small increases in the probability of apprehension yield large reductions in shootings, while at high levels of p_t the same increment buys much less. There are good reasons to expect that certainty is the margin more readily shifted in practice. Increasing p_t is fundamentally a matter of raising the likelihood of apprehension paired with some sufficiently severe consequence, which policing strategies and focused deterrence initiatives can plausibly affect, whereas severity S_t is largely determined by statutory sentencing structures that are more difficult to change and implement (Nagin et al., 2015).

The six empirical facts reviewed in Section 2 imply, however, that a single common response function is not an adequate description of the population at risk. Offending is concentrated among a small number of chronic offenders, criminal learning occurs through peer networks, and enforcement contacts accumulate over criminal careers. The probability of punishment is not exogenous and fixed, and the risk of arrest at any given criminal-justice contact depends on how actively an individual is offending.[3] To capture this heterogeneity, index individuals by an underlying risk type θ, distributed $G(\theta)$ in the population, and write the individual response function as $\lambda_t(p_t;\theta)$, increasing in θ: higher-θ individuals shoot at higher rates at any enforcement level.

Assumption 2 (Risk complementarity)**.** *The marginal deterrent effect of enforcement is increasing in offender risk:*

$$\frac{\partial^2 \lambda_t}{\partial p_t \, \partial \theta} \; < \; 0, \tag{2}$$

so that for any two risk types $\theta_H > \theta_L$ facing the same enforcement level,

$$\frac{\partial \lambda_t(p_t;\theta_H)}{\partial p_t} \; < \; \frac{\partial \lambda_t(p_t;\theta_L)}{\partial p_t} \; < \; 0. \tag{3}$$

Assumption 2 is the formal statement of the margins argument at the heart of this paper: deterrence operates most powerfully at the high-risk tail. It is this assumption that rationalizes targeted enforcement. Consider a fixed enforcement budget $\bar{p}$ allocated across individuals. By inequality (3), transferring a marginal unit of enforcement from a low-risk to a high-risk individual reduces total expected shootings by

$$\left[\frac{\partial \lambda_t(p_t;\theta_L)}{\partial p_t} - \frac{\partial \lambda_t(p_t;\theta_H)}{\partial p_t} \right] dp_t \; > \; 0, \tag{4}$$

so the shooting-minimizing allocation concentrates enforcement on the upper tail of the risk distribution. This is consistent with the empirical record on focused deterrence, in which strategies concentrating certainty of sanction on the small set of individuals and groups at highest risk of gun violence produce the largest declines in shootings.[4]

However, analysts and policy makers who evaluate a convex response at average enforcement levels confuse the effect of enforcement at the population mean with its effect at a given margin of

[3] Cook (1980) argues similarly that criminal choices may be highly responsive to even small changes in the probability of arrest when offenders closely observe one another's experiences through social networks. He further notes that the observed arrest rate is not equivalent to the probability of arrest for any single criminal opportunity. Rather, it reflects the average of accumulated arrest probabilities across opportunities selected by offenders. Consequently, improved enforcement may, under some conditions, reduce the observed arrest rate if offenders substitute away from opportunities with relatively high arrest probabilities.

[4] Under Assumption 1 alone—a common convex response function with no heterogeneity in θ—the allocation logic would run the other way: convexity in p_t implies diminishing returns to concentrating enforcement on any one individual, and a uniform allocation would minimize total offending. The case for targeting therefore rests on heterogeneity in the response function itself (Assumption 2), which the empirical facts in Section 2 support.

exposure. Let x denote an individual's exposure to enforcement intensity (e.g. $x = p_t S_t$) and let $f(x)$ denote the response in shooting intensity[5], with f decreasing and, by Assumption 1, convex:

$$f''(x) > 0. \tag{5}$$

Jensen's Inequality (Jensen, 1906) establishes that for a convex function the transformation of the mean is less than or equal to the mean of the transformation (Jensen, 1906):

$$f\big(\mathbb{E}[X]\big) \;\leq\; \mathbb{E}\big[f(X)\big]. \tag{6}$$

The left-hand side is the shooting response evaluated at average enforcement intensity—what an analyst recovers from aggregate descriptive data on average arrest clearance rates. The right-hand side is the actual population average of individual responses, integrated over the true distribution of exposure to enforcement, which varies enormously across individuals given the concentration of police contacts among active offenders. The gap between the two is the *Jensen's gap*:

$$\underbrace{\mathbb{E}\big[f(X)\big] - f\big(\mathbb{E}[X]\big)}_{\text{Jensen's gap}} \;\approx\; \tfrac{1}{2}\, f''\big(\mathbb{E}[X]\big) \cdot \mathrm{Var}(X). \tag{7}$$

This second-order Taylor approximation makes the mechanism explicit: the gap is proportional to both the *degree of convexity* $f''(\mathbb{E}[X])$ and the *variance* of enforcement exposure $\mathrm{Var}(X)$ across individuals. A more convex response curve and a more dispersed distribution of enforcement contacts both widen the gap between what average enforcement data show and what is actually happening in the population. Inference from aggregate clearance rates therefore systematically misstates the behavior of the individuals whose responses dominate the total—and by Assumption 2, those are precisely the high-risk offenders in the upper tail of $G(\theta)$, who face the steepest response functions and, under proactive enforcement regimes, the highest exposure x.

This measurement problem is not a departure from the Becker framework but a consequence of a simplification embedded in his seminal paper. Becker himself acknowledged that p_j, f_j, and u_j "are likely to differ significantly between persons" and adopted average values only "for simplicity," defining the market offense function at the population mean (Becker, 1968, p. 178). This simplification, innocuous in his theoretical exercise, has been lost in subsequent interpretation: because his own preferred constant-elasticity offense function is convex in p, evaluating deterrence at average enforcement levels systematically understates the response of the population—and equation (7) quantifies exactly the term that aggregation discards.

Figure 2 illustrates the Jensen's gap using $f(x) = e^{-\alpha x}$ with $\alpha = 2.2$, under two enforcement regimes. In the baseline regime, exposure ranges from x_{1b} (low enforcement) to x_2 (full enforcement); the chord connecting the response values at these endpoints (dashed orange) lies strictly above the curve at the mean $\mathbb{E}[X]$, so an average computed over dispersed exposures exceeds what the curve delivers at mean exposure. When enforcement pulls back, the lower endpoint shifts left from x_{1b} to x_{1p}, and the chord rotates upward (dashed red): the pullback both moves the population onto the steeper left portion of the curve and widens the dispersion of exposure, so $\mathbb{E}[f(X)]$ rises and the Jensen's gap widens. The implication for reforms that reduce the risk of sanctions for shootings is direct, and it combines both assumptions. Suppose police pull back on arrest clearances, or prosecutors decline to prosecute illegal firearms carrying among those likely to shoot someone, so that the lower bound of enforcement intensity falls from x_{1b} toward x_{1p}. Two forces compound.

[5]We note that the intensity of shootings $f(p_t, S_t)$ may be a non-linear multivariate function in both p_t and S_t, but to simplify notation and our argument we focus in this section on the one variable case with a single enforcement variable x.

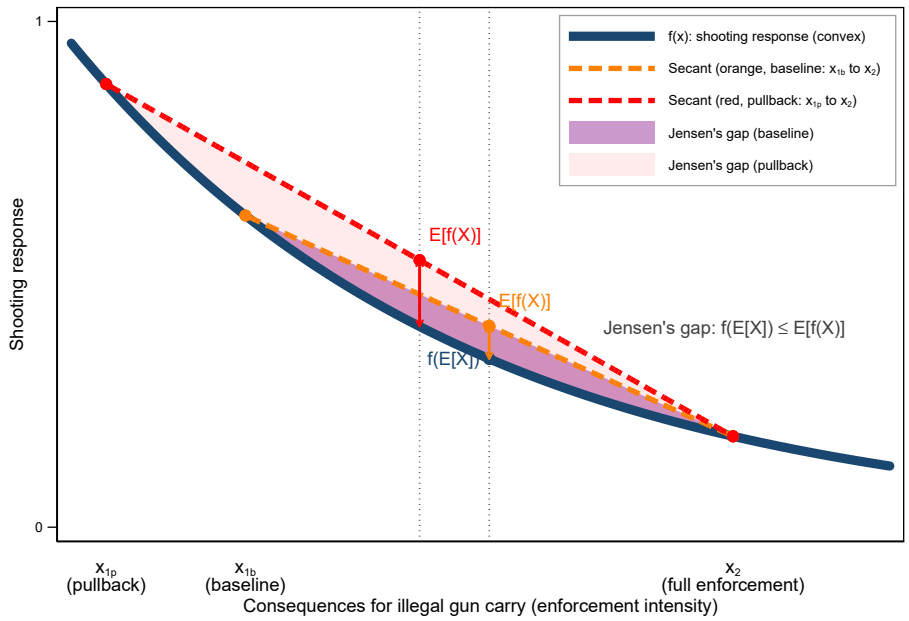


Figure 2: Jensen's Inequality applied to shooting deterrence under two enforcement regimes. The convex shooting response curve (blue) lies below the chord connecting the endpoints of each regime's exposure range: baseline (x_{1b} to x_2, dashed orange) and enforcement pullback (x_{1p} to x_2, dashed red). The vertical distance at $\mathbb{E}[X]$ between $\mathbb{E}[f(X)]$ on the chord and $f(\mathbb{E}[X])$ on the curve is the Jensen's gap.

First, by convexity (Assumption 1), the pullback moves part of the population onto the steep left portion of the response curve, where a given reduction in enforcement produces a larger increase in shootings than the same reduction would have produced at higher enforcement levels. Second, by risk complementarity (Assumption 2), the increase is largest for the high-risk offenders in the upper tail, whose response functions are steepest. The aggregate consequences of a uniform pullback are therefore concentrated in, and driven by, the tail. Meanwhile, by equation (7), deterrence estimates computed at average enforcement levels understate the population response, and the understatement grows as the pullback increases both the curvature at the mean and the dispersion of exposure—exactly the upward rotation of the chord visible in Figure 2. The result is a cascade that average-based calculations are structurally incapable of predicting: the elasticity at the mean looks small because it is dominated by individuals arrested for illegal gun carrying but at low underlying risk of committing a shooting, while shootings among high-risk offenders rise sharply. Moving from the baseline regime to the pullback regime does not merely shift the population along a flat demand curve; it moves offenders onto the steepest portion of a convex response function and amplifies the shooting impact far beyond what average deterrence estimates would suggest.

2.2 Heterogeneity in Criminal Propensity

Before building a dynamic model of shooting intensity, it is worth illustrating why the choice of how to represent individual heterogeneity in offending matters so much for the conclusions a model will

produce. To do so, we draw on a population of 781 individuals arrested for shootings in Chicago in 2020 shown in Table 1, and assume these shooters have a latent criminal propensity drawn from a Gamma distribution (γ_i), a standard and flexible way of representing unobserved heterogeneity in shooting rate among offenders. Figure 3 compares two Gamma distributions for the same population, both sharing a shape parameter of 2 (α), but differing in their scale parameter (θ), which governs both the mean and the variance of the distribution. One has a scale parameter of 1.0, implying a mean propensity of 2.0 and a variance of 2.0, while the other has a modest increase to 1.3, raising the mean propensity to 2.6 (a 30% increase) and the variance to 3.38 (a 69% increase), producing a distribution that is both shifted rightward and considerably more dispersed, with a small number of offenders reaching propensities well above the baseline case's effective range.

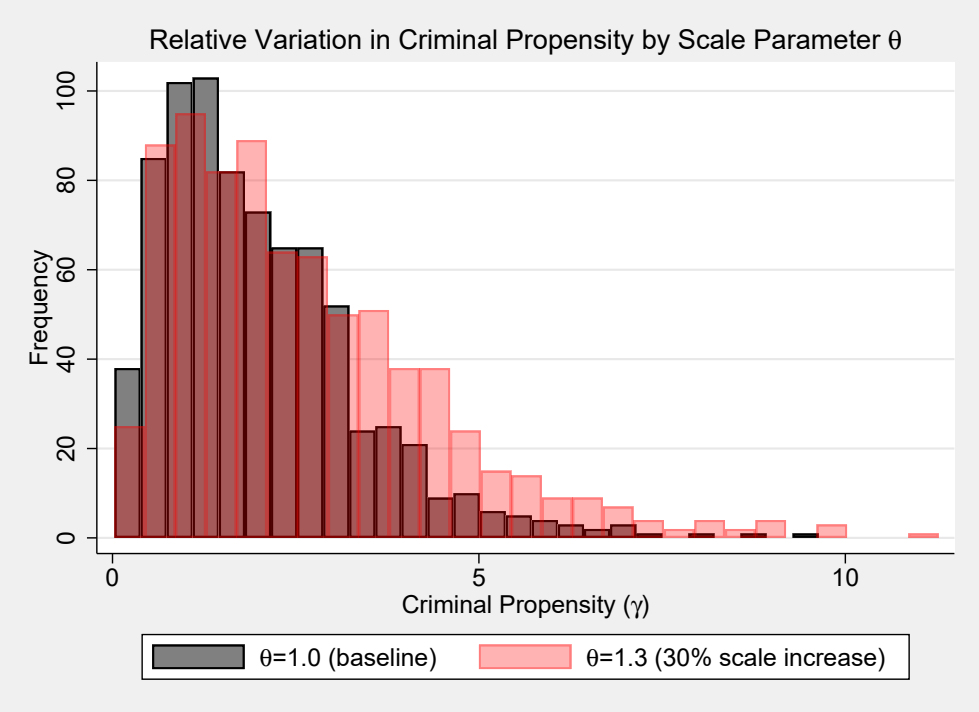


Figure 3: Criminal propensity γ for the same 781 shooters under two mixing distributions that differ only in spread (scale = 1.0 vs. scale = 1.3). Author's simulation.

The two distributions differ modestly in their means (2.07 vs. 2.61) and more sharply in their variances. Summed across all 781 shooters, the aggregate number of shootings implied by the population, representing the sum of γ_i across all offenders, rises from roughly 1,564 to roughly 2,031, an increase of about 468 shootings, or 30%. Because the expected population total is simply the number of offenders times the mean propensity, this 30% increase follows directly from the corresponding 30% increase in average propensity, and does not on its own isolate the distinct contribution of the thicker right tail, that is, the handful of individuals with γ_i above 8, 10, even 13, who do not exist in the thinner-tailed version. Even so, the exercise underscores a broader point: a model that fixes the spread of a mixing distribution too low, which is what happens implicitly whenever a Poisson process is combined with only a thin-tailed or bounded source of individual heterogeneity, risks understating the aggregate contribution of high-propensity offenders even when

it appears to get the population average roughly right.

2.3 Shootings Before Deterrence: Geometric vs. Thin-Tailed Alternatives

The same logic shows up even more sharply when we think about offending from a dynamic perspective and ask the question: how many offenses does a given shooter commit before being deterred for the first time? At each criminal-justice contact, a shooter offends with binomial probability of success (p_{success}) and is deterred with the complementary probability ($1 - p_{\text{success}}$). Given p_{success}, the number of contacts before an offender is first deterred follows a geometric distribution with mean $1/p_{\text{success}}$ and variance $(1 - p_{\text{success}})/p_{\text{success}}^2$. This geometric distribution allows for a longer right tail: a small number of shooters who are successful by avoiding sanctions can commit many shooting offenses before they are stopped. Formally, if X denotes the number of shootings committed before the first deterrence event, then X follows a geometric distribution with probability mass function

$$P(X = k) = (1 - p_{\text{success}})^{k-1} \, p_{\text{success}}, \qquad k = 1, 2, 3, \ldots, \tag{8}$$

with mean and variance

$$\mathbb{E}[X] = \frac{1}{p_{\text{success}}}, \qquad \text{Var}(X) = \frac{1 - p_{\text{success}}}{p_{\text{success}}^2}. \tag{9}$$

Figure 4 simulates this distribution for the same 781 shooters under four scenarios, labeled by the complementary deterrence value $1 - p_{\text{success}} \in \{0.25, 0.35, 0.45, 0.55\}$.

Because the mean and variance of a geometric distribution both scale with $1/p_{\text{success}}$, moving the deterrence probability across the four labeled scenarios mechanically stretches out both the center and the tail of the distribution: the average number of offenses before being deterred rises from about 1.8 (at 1-$p_{\text{success}} = 0.45$) to about 4.0 (at 1-$p_{\text{success}} = 0.75$), while the maximum value observed in 781 draws rises from 11 to 30. Summing across all 781 shooters, total simulated offenses-before-deterrence rise from about 733 to about 1,613 — an increase of roughly 880 offenses, or 120%. The point of this exercise is to show a modest 30-percentage-point shift in a single per-contact success probability compounds, through the geometric distribution's $1/p^2$ variance term, into a much larger shift in the realized tail. This is the kind of small-change but large-consequence dynamic that motivates the need for a fat-tailed distribution.

The contrast with a model that represents enforcement contact as a simple bounded Bernoulli (binomial) recursion is instructive. In a pure gambler's ruin each criminal justice contact resolves to either one shooting success or one deterrence and incapacitation (assuming the person gets incarcerated for a shooting and removed from the streets) and the compounded effect on shootings are mechanically bounded by the number of contacts observed. With only four simulated contacts and individual can only commit at most four net shooting offenses, no matter how high-rate they truly are. The geometric formulation removes that artificial ceiling because it models the *waiting time* to the first deterrence event rather than a fixed number of discrete trials, and allows the small number of truly persistent offenders to show up in the data as the extreme outliers they actually are. In the geometric formulation there is an individual with 17 offenses more than four times what a four-contact binomial recursion could ever produce for a single individual. A model that only allows the bounded, thin-tailed version will, by construction, undercount exactly the part of the distribution that crime control policy should care about most. A credible model must reproduce two features jointly: that a small number of individuals account for a grossly disproportionate share of shootings, and that each shooting raises the risk of the next until the offender is stopped by

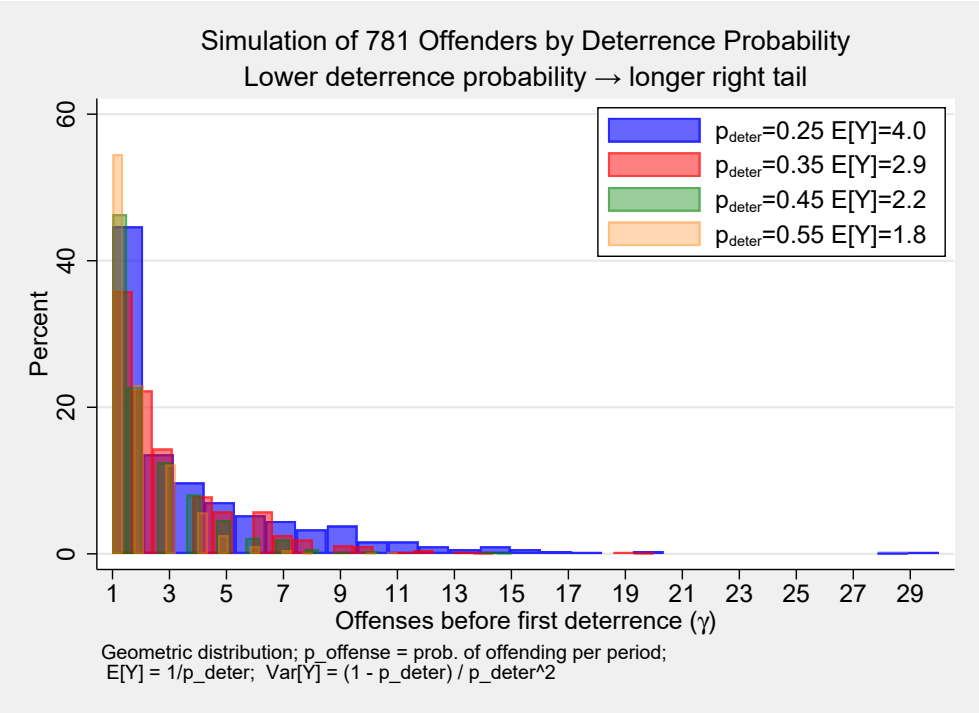


Figure 4: Simulated number of offenses before first deterrence, 781 shooters, geometric distribution parameterized by the binomial probability of deterrence $=1 - p_{\text{success}}$.

deterrence or incapacitation. A homogeneous-rate model thins the tail mechanically, treating the extremes as noise around a common mean. Where an individual's contribution to shooting rates is capped, the cap should reflect arrest and incapacitation, not analytical convenience.

3 The Dynamics of Criminal Careers

Section 2 established that a realistic model of shooting intensity needs to allow for a fat tailed distribution. This section lays out the dynamic ingredients needed to turn that static observation into an actual model of how criminal careers unfold over time, building on the same gambler's ruin logic.

3.1 Repeated Encounters as a Repeated Game

The starting intuition of gambler's ruin is that criminal-justice contact is not a single draw but a repeated game (Epstein, 2012). Just as a gambler who plays many rounds at slightly unfavorable odds faces a rapidly compounding probability of losing everything, an offender committing shootings who has repeated contact with police and the courts faces a compounding probability that an enforcement outcome (arrest, conviction, incarceration) will end their criminal career, even if the probability of that outcome at any single contact is modest. Small differences in the per-criminal justice contact probability of an unfavorable outcome therefore compound, over a career, into large differences in how many shootings accumulate before an offender is incapacitated.

One can illustrate this directly with data from Chicago shown in Table 1: the $S = 781$ individuals arrested for shootings from the 2020 calibration. At each of four simulated criminal-justice contacts, each shooter draws a Bernoulli trial that resolves either to an offense (with $1 - \phi$, where ϕ is the deterrence probability) or a deterrence event (probability ϕ); cumulative shootings are the running total of offenses across all 781 shooters over successive contacts.

Table 2 reports cumulative shootings at each of the four simulated contacts for deterrence probabilities $\phi \in \{0.25, 0.35, 0.45, 0.55\}$. The spread across deterrence levels widens steadily with each additional contact, from 256 shootings after the first contact to 1,631 after the fourth—a roughly sixfold increase. This shows that the consequences of small differences in per-contact deterrence compound rather than stay fixed as offenders accumulate contacts, motivating a focus on repeat contacts rather than first-time, low-rate offenders.

Table 2: Cumulative Shootings by Deterrence Level and Criminal-Justice Contacts, 781 Offenders

CJ Contacts	$D = .25$	$D = .35$	$D = .45$	$D = .55$	Spread
1	2,156	2,076	2,000	1,900	256
2	3,486	3,287	3,085	2,739	747
3	4,459	4,088	3,757	3,263	1,196
4	5,724	5,217	4,762	4,093	1,631

Note: D = per-contact deterrence probability. Spread = difference between $D = .25$ and $D = .55$ columns.

3.2 Model of Offending

Together, the empirical facts of criminal offending and the analytic points established above via Jensen's inequality, that variation in individual risk and in criminal-justice contact compounds into far more extreme outcomes than a single "average" offender would predict, motivate our baseline model of shooting intensity. Because heterogeneity in offending is concentrated in the tail, the model requires a fat-tailed mixing distribution, in which tail variance drives the extreme outcomes that matter most for policy. The model must therefore make arrest probability endogenous to an individual's own shooting rate, embed individuals in offending networks so that a peer's shooting raises one's own risk, and track accumulated enforcement contacts over time so that individuals can eventually be absorbed and taken out of the active risk pool through incapacitation.

3.3 From a Toy Simulation to a Model of Shooting Intensity

The simulations above are illustrative; turning them into an actual model of a city's shooting rate requires several additional structural pieces. First, we need an explicit *population structure*: a fixed number of individuals at risk of shootings, organized into peer networks, observed over a fixed number of discrete time periods. Second, we need *latent heterogeneity*: a way of distinguishing a small high-rate "shooter" subpopulation from a much larger low-rate background population. The finite-mixture structure implied in the notation of a fat tail. Third, we need a *self-exciting mechanism*: a way for past violence to elevate the probability of future violence, since shootings cluster in time rather than arriving independently. Fourth, we need a *deterrence memory*: a way for past enforcement contact to suppress future violence, decaying over time as its chilling effect fades. Fifth, we need the *gambler's ruin absorbing barrier* itself: a finite buffer of enforcement contacts that, once exhausted, removes an individual from the active risk pool entirely through incapacitation.

Each of the ingredients above captures part of the relevant dynamic of shooting propensity. A pure finite mixture tells us who is high-risk but says nothing about how risk evolves over time. A pure self-exciting (Hawkes) process captures temporal clustering but has natural endpoint. A pure deterrence-memory term captures the chilling effect of enforcement but says nothing the heterogeneity between individuals to absorb enforcement contact before exiting the population. And the gambler's ruin barrier on its own says nothing about why some individuals are more likely to face that barrier than others, or about how violence in one part of a social network spills over into another. We therefore bundle all four mechanisms (the finite mixture, self-excitation, deterrence memory, and the gambler's ruin barrier) together with a network contagion term into a single conditional-intensity model of shooting events. The model is simulated separately for each year in Table 5, using that year's population of at risk-offenders, network count, count of arrested shooters, and observed shooting total to calibrate the baseline intensities.

Table 3 summarizes these components; each is developed formally in Section 4, with the motivation for each choice connected to the tail-risk logic requiring a fat-tailed mixing distribution and a geometric, rather than bounded, representation of shootings before deterrence.

Table 3: Key Components of the Tail Risk Model of Shooting Deterrence

Symbol	Description	Notes / Values
$C_{i,t}$	Expected shooting intensity	Poisson rate for realized shootings
θ_g	Group latent trait	$\sim N(0, 0.02)$, constant within network g
μ_i	Individual latent trait	$\sim N(0, 0.01)$, time-invariant
$\varepsilon_{i,t}$	Idiosyncratic shock	$\sim N(0, 0.01)$
$f_{i,t}$	Self-excitation kernel	Fading memory of own prior shootings
$g_{i,t}$	Deterrence memory kernel	Fading memory of past enforcement contacts
λ_i^{base}	Baseline intensity	$\mu_s \approx 0.25$ (shooters); $\mu_{ns} \approx 0.01$ (others)
$S_{i,t}$	Gambler's ruin buffer	Integer $\in \{0, \ldots, M\}$; $S = 0$ is absorbing
ϕ_g	Deterrence scenario	Per-period arrest probability; $\phi_0 \in \{0.15, \ldots, 0.55\}$
ω_1	Self-excitation decay	0.30 (half-life $\approx$ 2.3 months)
ω_2	Deterrence decay	0.15 (slower: enforcement memory persists)
R_1	Excitation jump	0.10 (shooters); 0.05 (non-shooters)
R_2	Deterrence jump per arrest	0.20
β	Network contagion elasticity	0.40
M	Absorbing barrier depth	3 enforcement contacts

4 The Model

A terminological clarification is in order before presenting the formal model. We use deterrence to mean specific deterrence, or the suppression of an individual's future shooting following contact with the criminal justice system and a sanction. This is modeled jointly with incapacitation, the permanent removal of that individual from the active risk pool once their enforcement-contact buffer $S_i(t)$ is exhausted. General deterrence, in this context, is the effect of sanctions on co-offender associates who were not themselves arrested, and emerges endogenously through the network contagion mechanism $A_i(t)$. The contagion channel runs in both directions: suppressing a high-rate shooter lowers their peers' risk, while an unsuppressed one raises it through self-excitation, propagating violence outward through the network rather than containing it to the individual. General deterrence in

this model is therefore a second-order, network-transmitted consequence of specific deterrence and incapacitation, rather than a directly imposed effect.

4.1 Population and Latent Types

The population in a given year consists of N individuals indexed by i, observed over $T = 12$ monthly time periods, partitioned into G disjoint peer networks of size $|k(i)| \approx 2$–3 (assigned by modal arithmetic). Each individual is assigned a latent type $z_i \in \{0, 1\}$ drawn once, at the start of the simulation, from a Bernoulli distribution with mixing proportion $\pi = S/N$:

$$z_i \sim \text{Bernoulli}(\pi). \tag{10}$$

Individuals with $z_i = 1$ are the high-risk "shooter" type; the remaining $N - S$ individuals are the low-risk "non-shooter" type. In the 2020 calibration used for illustration throughout this paper, $N = 8{,}228$ at risk-offenders and $S = 781$ shooters, so $\pi \approx 0.095$: shooters are roughly 9.5% of the at risk offending population.

4.2 Full Shooting Intensity Equation

The shooting process for individual i in period t is a non-stationary Poisson process with conditional intensity $C_i(t)$:

$$N_i(t) \mid C_i(t) \;\sim\; \text{Poisson}\big(C_i(t)\big) \tag{11}$$

$$C_i(t) = \lambda_i^{\text{base}} \cdot \exp\Big[\underbrace{\theta_{k(i)} + \mu_i + \varepsilon_{it}}_{\text{latent heterogeneity}} + \underbrace{f_i(t)}_{\text{self-excitation}} - \underbrace{g_i(t)}_{\text{deterrence}} - \underbrace{\lambda_i^{\text{base}} \cdot \tfrac{1}{S_i(t)+1}}_{\text{gambler's ruin}} \Big] \cdot \underbrace{\big(1 + A_i(t)\big)^{\beta}}_{\text{network contagion}} \tag{12}$$

The baseline intensity λ_i^{base} depends on type and is calibrated each year from that year's observed shooting total, with shooters assumed responsible for 70% of events and non-shooters for the remaining 30%:

$$\lambda_i^{\text{base}} = \begin{cases} \mu_s & \text{if } z_i = 1 \text{ (shooter)} \\ \mu_{ns} & \text{if } z_i = 0 \text{ (non-shooter)} \end{cases}, \qquad \mu_s = \frac{0.70 \times \text{Obs.}}{S \times 12}, \qquad \mu_{ns} = \frac{0.30 \times \text{Obs.}}{(N - S) \times 12}. \tag{13}$$

In the 2020 calibration, this gives $\mu_s \approx 0.2516$ and $\mu_{ns} \approx 0.0113$, indicating that shooters have a per-period baseline rate roughly 22 times higher than non-shooters.

The three latent-heterogeneity terms in Equation (12) are drawn at different levels of aggregation and held fixed over different spans of time: $\theta_{k(i)}$ is a network effect constant across all members of i's peer network for the year, μ_i is an individual effect constant across the year for i but varying across individuals within the same network, and ε_{it} is an idiosyncratic shock that varies independently every month,

$$\theta_{k(i)} \sim \mathcal{N}(0,\, 0.02), \qquad \mu_i \sim \mathcal{N}(0,\, 0.01), \qquad \varepsilon_{it} \sim \mathcal{N}(0,\, 0.01). \tag{14}$$

The remaining four terms in Equation (12) are state variables that evolve recursively rather than being drawn once. The self-excitation kernel, deterrence kernel, and network contagion term update each period as

$$f_i(t) = (1 - \omega_1)\, f_i(t-1) \;+\; R_{1,z_i}\, N_i(t-1), \tag{15}$$

$$g_i(t) = (1 - \omega_2)\, g_i(t-1) \;+\; R_2 \cdot \mathbb{1}\{\Delta S_i(t-1) = -1\}, \tag{16}$$

$$A_i(t) = \frac{1}{|k(i)|} \sum_{j \in k(i)} N_j(t-1), \tag{17}$$

so that self-excitation occurs as a shooting raises i's own future risk before decaying away at rate ω_1. This is countered by a deterrence buffer: an enforcement contact — a decrement of the gambler's-ruin protection, $\Delta S_i(t-1) = -1$ — chills future risk before decaying away at the slower rate $\omega_2 < \omega_1$. Finally, contagion in the peer network occurs as shootings among i's peer-network members in the prior month raise i's own risk through the $(1 + A_i(t))^\beta$ term. The buffer decrement that feeds Equation (16) is itself generated by an explicit enforcement process rather than an exogenous rate. In each period, i's probability of an enforcement contact is increasing in i's own current shooting intensity, scaled by the network's deterrence baseline $\phi_{k(i)}$ and normalized by the shooter baseline rate μ_s:

$$p_i^{\downarrow}(t) = \min\Big(\phi_{k(i)} \cdot \frac{C_i(t)}{\mu_s},\ 1\Big), \qquad \Delta S_i(t) = \begin{cases} -1 & \text{w.p. } p_i^{\downarrow}(t) \quad \text{(enforcement contact)} \\ +1 & \text{w.p. } 1 - p_i^{\downarrow}(t) \quad \text{(quiet period)} \end{cases} \tag{18}$$

$$S_i(t) = \text{clip}\big(S_i(t-1) + \Delta S_i(t-1),\ 0,\ M\big), \tag{19}$$

with i permanently removed from the active risk pool — $C_i(\tau) = 0$ for all $\tau > t$ — the first period $S_i(t) = 0$. Normalizing by μ_s means a shooter operating exactly at that year's baseline intensity faces a per-period contact probability of $\phi_{k(i)}$ itself, so ϕ_0 can be read directly as an at-baseline arrest probability, as in Table 4, with the buffer decrement recalibrated automatically each year alongside μ_s.

Equations (16) and (18) are both triggered by the same enforcement contact, so the model carries two enforcement channels rather than one. The deterrence kernel $g_i(t)$ operates on the intensive margin: a temporary suppression of risk for an individual who remains in the population, decaying back to zero at rate ω_2. The gambler's-ruin buffer $S_i(t)$ operates on the extensive margin: a permanent, absorbing state governing whether an individual remains in the population at all. A single enforcement contact therefore has both a short-run behavioral effect that fades and a long-run structural effect that does not. The two channels decompose specific deterrence from incapacitation rather than duplicating one mechanism, consistent with the terminology drawn at the start of Section 4.

The simulation is initialized at $f_i(0) = g_i(0) = 0$ and $S_i(0) = M$ for every individual rather than at each variable's steady-state distribution. This reflects how the population is sampled rather than numerical convenience: the N individuals simulated in a given year are drawn from that year's co-offending arrest network and are therefore, by construction, active and non-incapacitated as the year begins. [6]

Three structural features of equation (12) carry the analytical results that follow, and we state them explicitly.

Assumption 3 (Multiplicative deterrence)**.** *Deterrence enters through the exponential term $e^{-g_i(t)}$, so enforcement removes a* proportion *of an individual's shooting intensity rather than a fixed number of events. Evaluating equation* (18) *at $C_i(t) = \mu_s$ gives a stationary per-period arrest probability $\varphi_i = \phi_{k(i)}$, under which the deterrence kernel has steady-state expectation $\bar{g}_i = \kappa\varphi_i$ with $\kappa = R_2/\omega_2$, giving the steady-state response*

$$\lambda_i(\varphi_i) \ = \ \lambda_i^{base}\, e^{-\kappa\varphi_i}. \tag{20}$$

Because suppression is proportional, the absolute number of shootings prevented by a given φ scales with λ_i^{base}: heterogeneity in baseline levels, not curvature of a common response function, is what makes targeted enforcement valuable.

[6] Initializing from a stationary distribution would presuppose that some fraction of the year's population is already near absorption or already carrying elevated excitation before month 1, which is inconsistent with a sampling frame that selects on recent activity starting at the baseline of the year rather than on steady-state tenure. A longer term model could include a discarded burn-in period if there was a clear underlying steady state.

Assumption 4 (Subcriticality of cascades)**.** *Self-excitation and network contagion generate finite cascades: each shooting produces fewer than one expected direct "offspring" event ($n_i < 1$ for all i). At the calibrated values the branching ratios are $n_s \approx 0.06$ and $n_{ns} \approx 0.001$, so the condition is amply satisfied. Subcriticality implies that the total expected events sourced by an individual, $V(\lambda) = \lambda/(1 - c\lambda)$, is increasing and strictly convex in the baseline rate—the property underlying the aggregation results of Section 2.1.*

Assumption 5 (Linear enforcement budget)**.** *Total enforcement effort is $\Phi = \sum_i \varphi_i$ with $\varphi_i \geq 0$, so allocations across individuals are comparable at a fixed aggregate budget.*

Under these assumptions, the calibrated 22:1 intensity ratio is not merely descriptive: the advantage of optimally targeted over uniform enforcement at the same budget equals the gap between the arithmetic and geometric means of λ_i^{base}, which grows with heterogeneity, while the convexity of $V(\lambda)$ implies that models or estimates calibrated to the population-average intensity systematically understate total shootings.

4.3 Parameters in Plain English

Table 4 lists every parameter in the model using the 2020 population for illustration. We now explain what each block of the model is doing in plain language.

Table 4: Model Parameters

Parameter	Symbol	Value	Role
Population (2020 illustrative calibration)			
Individuals	N	8,228	Total population
Time periods	T	12	Simulation horizon (months)
Networks	G	3,468	Peer groups (≈2–3 members each)
Shooters	S	781	High-risk mixture component
Arrest buffer max	M	3	Absorbing barrier depth
Mixture baselines			
Shooter intensity	μ_s	0.2516	Monthly shooting rate, shooter type
Non-shooter intensity	μ_{ns}	0.0113	Monthly shooting rate, non-shooter type
Self-excitation kernel			
Excitation decay	ω_1	0.30	Speed at which past shootings stop mattering
Shooter jump	$R_{1,s}$	0.10	Excitation added per shooting (shooter)
Non-shooter jump	$R_{1,ns}$	0.05	Excitation added per shooting (non-shooter)
Deterrence kernel			
Deterrence decay	ω_2	0.15	Speed at which a past arrest stops mattering
Arrest jump	R_2	0.20	Deterrence added per arrest
Network contagion			
Contagion elasticity	β	0.40	How strongly peer violence raises own risk
Gambler's ruin / deterrence scenarios			
Deterrence baseline	ϕ_0	0.15–0.55	Per-period arrest probability at baseline intensity

Population block. N, G, T, and S describes how large the total population of people tied to someone who has been arrested for a shooting, how many co-offending networks exist in the

population, how many months they are observed, and the number that are high-risk type. M is the number of enforcement contacts an individual can absorb before being permanently removed from the risk pool.

Mixture baselines. μ_s and μ_{ns} are simply each type's average monthly shooting rate before any of the dynamic adjustments are applied. The fact that μ_s is roughly 22 times μ_{ns} is exactly the tail-concentration logic discussed in the prior literature that is the basis for the model's starting point.

Self-excitation kernel. $f_i(t)$ captures the idea that a shooting makes an individual more likely to be involved in another shooting in the near future: it keeps a fading memory of i's own recent shooting involvement, which feeds back into their risk going forward. Every month it decays toward zero at rate ω_1, and every shooting adds a jump of size R_{1,z_i}. Because shooters get a bigger jump than non-shooters (0.10 vs. 0.05), a shooting event excites a shooter's future risk more than it would excite a non-shooter's risk, consistent with the retaliatory dynamics often observed in gun violence (Leovy, 2015).

Deterrence kernel. $g_i(t)$ is the fading memory of recent enforcement contact. It decays more slowly than the excitation kernel ($\omega_2 < \omega_1$), meaning the chilling effect of an arrest outlasts the excitation effect of a shooting, consistent with the idea that getting arrested has a longer-lasting behavioral effect than committing an offense (Thomas & Pogarsky, 2026).

Gambler's ruin buffer. $S_i(t)$ is the gambler's ruin variable: an integer buffer between 0 and $M = 3$ that goes down by one with each enforcement contact and can recover (up to the cap) during quiet periods. When it hits zero, the individual is permanently incapacitated. The term $-\lambda_i^{\text{base}}/(S_i(t)+1)$ in equation (12) is a penalty that grows as this buffer erodes, so risk falls mechanically as someone approaches the absorbing barrier, independent of the deterrence kernel.

Network contagion. $A_i(t)$ is the average number of shootings in an individual's own peer network last month. It enters multiplicatively, so a shooting in someone's immediate network raises their own risk of a future shooting and captures retaliation, co-offending, and the spread of violence through criminal networks.

Deterrence scenarios. ϕ_0 is the policy lever of deterrence: the network-level probability of an arrest and sanction at baseline shooting intensity. We simulate five values, $\phi_0 \in \{0.15, 0.25, 0.35, 0.45, 0.55\}$, ranging from a low-enforcement regime to a regime consistent with focused-deterrence.

Table 5 reports this structure for each year from 2018 through 2024, together with the observed annual shooting count that the model is calibrated with.

Table 5: Population Structure, Observed Shootings, and Calibrated Baseline Intensities by Year

Year	N	G	S	$N-S$	π	Observed	μ_s	μ_{ns}
2018	13,677	3,752	659	13,018	0.048	2,504	0.2216	0.0048
2019	13,975	3,779	680	13,295	0.049	2,221	0.1905	0.0042
2020	8,228	3,468	781	7,447	0.095	3,368	0.2516	0.0113
2021	6,032	3,084	645	5,387	0.107	3,637	0.3289	0.0169
2022	6,800	3,096	674	6,126	0.099	2,950	0.2553	0.0120
2023	7,402	3,371	622	6,780	0.084	2,540	0.2382	0.0094
2024	7,932	3,555	726	7,206	0.092	2,376	0.1909	0.0082

$\pi = S/N$. $\mu_s = 0.70 \times \text{Obs.}/(S \times 12)$; $\mu_{ns} = 0.30 \times \text{Obs.}/((N-S) \times 12)$.

The model is simulated separately for each year from 2018 through 2024 using parameters derived from Chicago data, using that year's N, G, S, and observed shooting total from Table 5, under each of the five deterrence scenarios. A visual depiction of the key model parameter is displayed in

Appendix Table A1.

5 Results: Model vs. Actual Shootings

Figure 5 shows for select years 2018, 2021, and 2024 the key structure of month-by-month trajectory of total simulated shootings based on for all five deterrence scenarios. Within every year, the ordering of the five deterrence curves is exactly what the model's logic predicts: higher deterrence (ϕ_0) produces a lower trajectory of total shootings, month after month, as the gambler's ruin buffer erodes faster among the shooter subpopulation and more of them are absorbed out of the active risk pool. One can see that there are materially higher yearly counts in 2021 consistent with the pandemic-era spike in observed shootings. The spread between the $\phi_0 = 0.15$ and $\phi_0 = 0.55$ curves widens after roughly month 4-5 in each year as the gambler's ruin absorption accumulates among the most active shooters under the higher enforcement scenario. Figures A1–A7 in the appendix display the Gambler's Ruin buffer, incapacitation rate, Hawkes excitation and deterrence kernels, and mean conditional intensity at the calibration anchor $\phi = 0.45$ for each year from 2018 through 2024.

Aggregating each year's monthly trajectories into an annual total and comparing them to the actual reported shooting count produces Figure 6. Two patterns stand out. First, every simulated deterrence scenario tracks the broad shape of the actual yearly shooting numbers in Chicago: the drop in 2019, the sharp rise into the 2020-2021 pandemic-era peak, and the decline through 2024. The model is not simply fitting a flat trend; it reproduces the actual year-to-year swings under every enforcement assumption. Second, and more important for the argument of this paper, the actual reported total sits *above every single simulated deterrence curve, in every year* -including the most permissive scenario, $\phi_0 = 0.15$.

Table A1, breaks the calibration down further, reporting total simulated shootings by latent type (shooter vs. non-shooter) at each of four deterrence levels by year, alongside actual shooting counts. The decomposition by latent type reveals that even under the simulation there remains a striking concentration: across all years and scenarios, shooters account for the vast majority of simulated shootings despite representing a small fraction of the population. In 2021, the peak violence year, simulated shooter counts range from 2,351 at $\phi_0 = 0.15$ to 1,683 at $\phi_0 = 0.55$—a reduction of more than 668 shootings driven entirely by deterrence acting on the shooter subpopulation. Non-shooter counts in that same year move only modestly, from 1,064 to 1,099, remaining essentially flat across all deterrence scenarios.

This asymmetry is consistent across every year in the table. In 2018, shooter counts fall from 1,615 at $\phi_0 = 0.15$ to 1,123 at $\phi_0 = 0.55$, a decline of nearly 30%, while non-shooter counts vary by fewer than 50 individuals across the same range. The same pattern holds in 2023 and 2024 as shootings declined toward pre-pandemic levels: deterrence scenarios compress shooter counts substantially, from 1,785 to 1,158 in 2023 and from 1,585 to 1,102 in 2024, while non-shooter counts remain largely unaffected. The entire deterrence lever in this model operates through the shooter subpopulation, exactly the tail-focused mechanism a credible model needs to have.

5.1 Reconciling the Model With Chicago's Clearance Rate

The gap in Figure 5 between the actual series and even the most permissive deterrence scenario of $\phi_0 = 0.15$ raises the question of why under the lowest simulated sanctions regime the model predicts fewer shootings than Chicago actually experienced in every year. The answer is that $\phi_0 = 0.15$ is not, in fact, a low estimate of real-world sanctions regime in Chicago. Chicago's arrest rate for

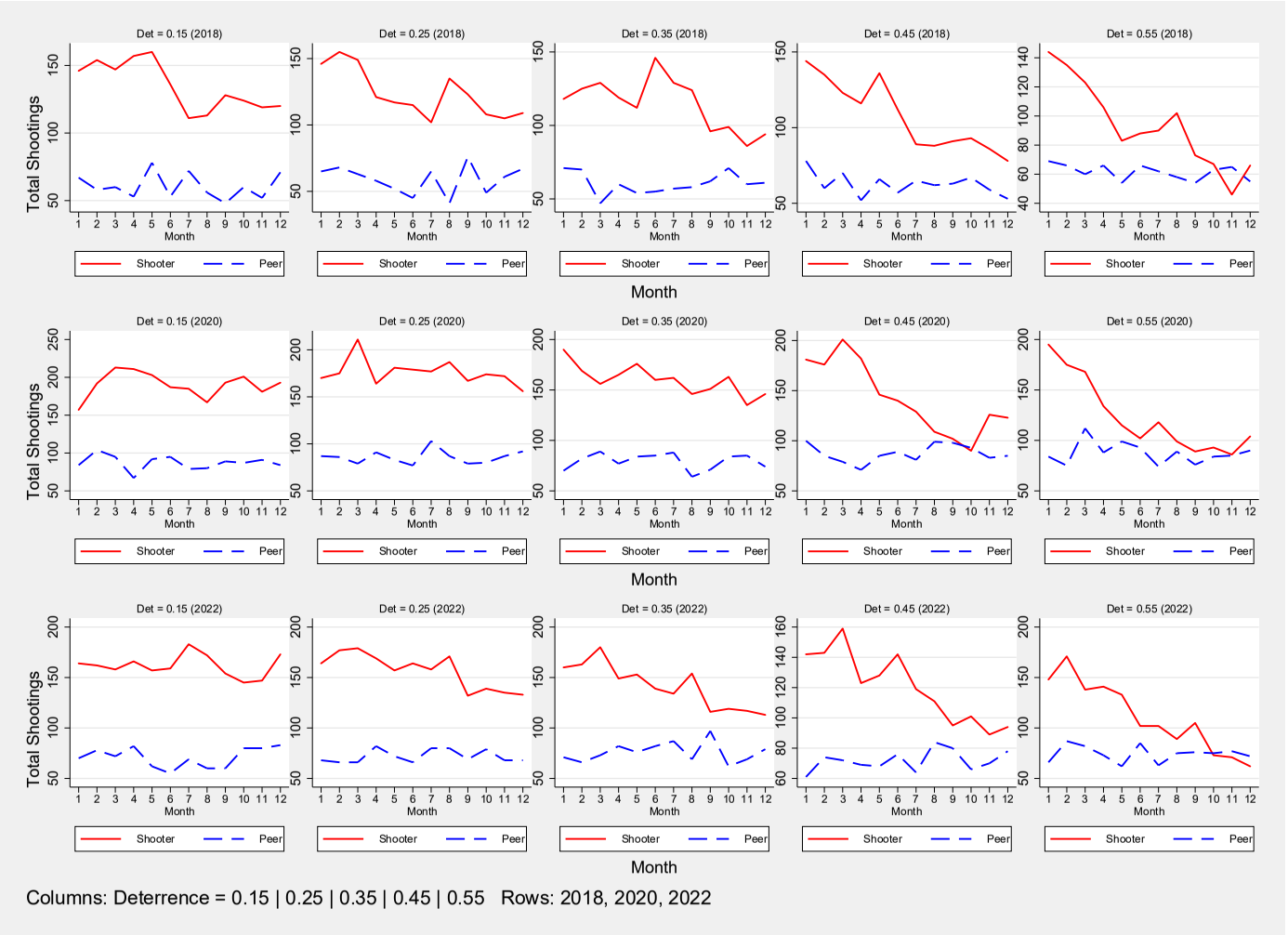


Figure 5: Total simulated shootings by month, 2018–2024, under five deterrence scenarios ($\phi_0 = 0.15$ through 0.55).

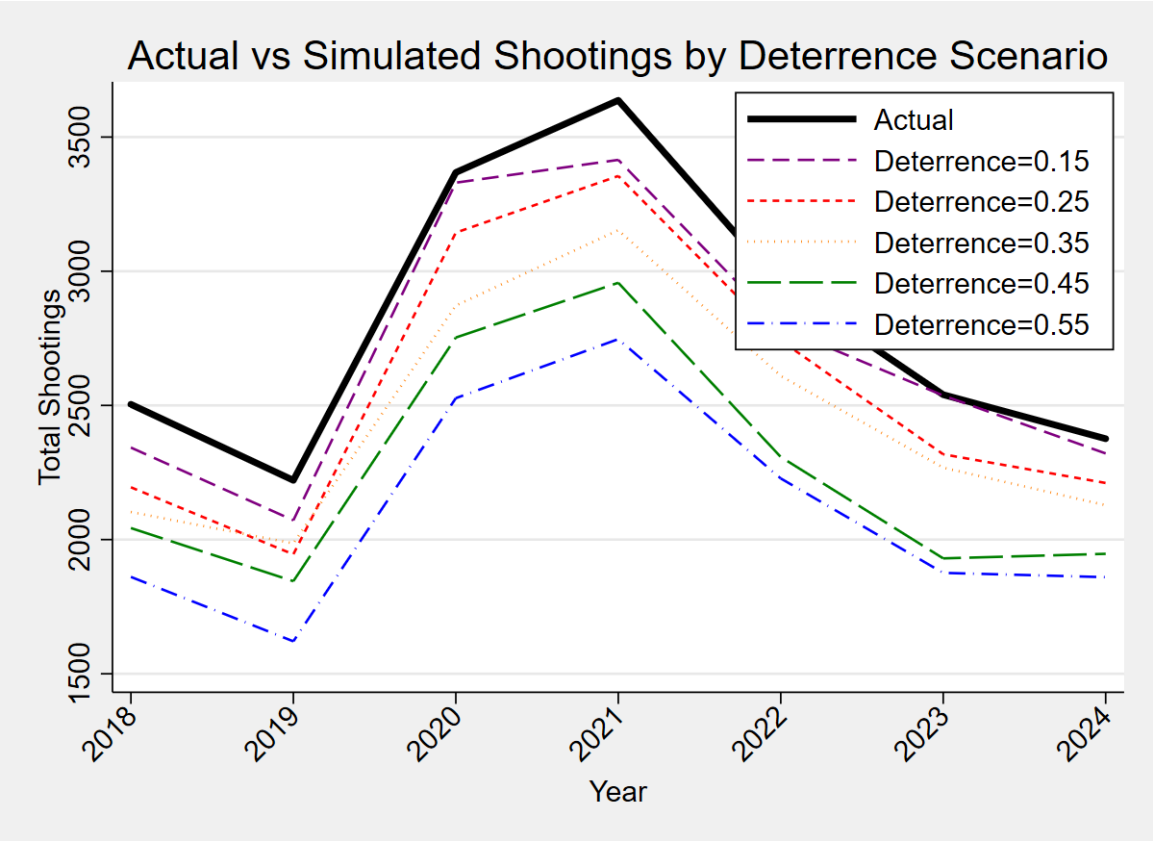


Figure 6: Actual vs. simulated total shootings by year, 2018–2024, under five deterrence scenarios.

shootings, shown in Table A2, averages 12.83%.[7] Even arrest rates for the more closely investigated subset of *fatal* shootings range between in the 23.74-32.16% for 2018-2024. Once that is understood, the pattern in Figure 6 of actual shootings exceeding every simulated curve in every year is not a sign that model is not correctly calibrated, but a sign of the direction shootings should move once we recognize that the true per-contact probability of an arrest and consequence for a shooting in Chicago is even lower than the most permissive scenario simulated.

The takeaway for policy is the mirror image of the takeaway for modeling: if even a 15% arrest probability at baseline intensity is insufficient to bring simulated violence down to the modest levels Chicago has periodically achieved, then the actual clearance rate is nowhere near the threshold needed to meaningfully constrain the upper tail of offending that drives total shootings. Figure 7 calibrates the model using the observed year-specific arrest rates for shootings as the deterrence parameter and compares simulated totals directly to actual shooting counts. The re-calibrated model tracks realized total shootings closely, falling only slightly below observed totals. Augmenting each

[7]Authors' calculations based on linking `case_number` in the Chicago Police Department arrest records to `case_number` in the Violence Reduction - Victims of Homicides and Non-Fatal Shootings file. Arrest records are available at https://data.cityofchicago.org/Public-Safety/Arrests/dpt3-jri9/about_data. Shooting incident records are available at https://data.cityofchicago.org/Public-Safety/Violence-Reduction-Victims-of-Homicides-and-Non-Fa/gumc-mgzr/about_data. The `case_number` field uniquely identifies each CPD incident report and serves as the common key across both datasets.

yearly estimate with a 95% confidence interval confirms tight convergence to the actual number of shootings, validating both the model's structure and the central role of enforcement probability in governing the upper tail of the shooting distribution.

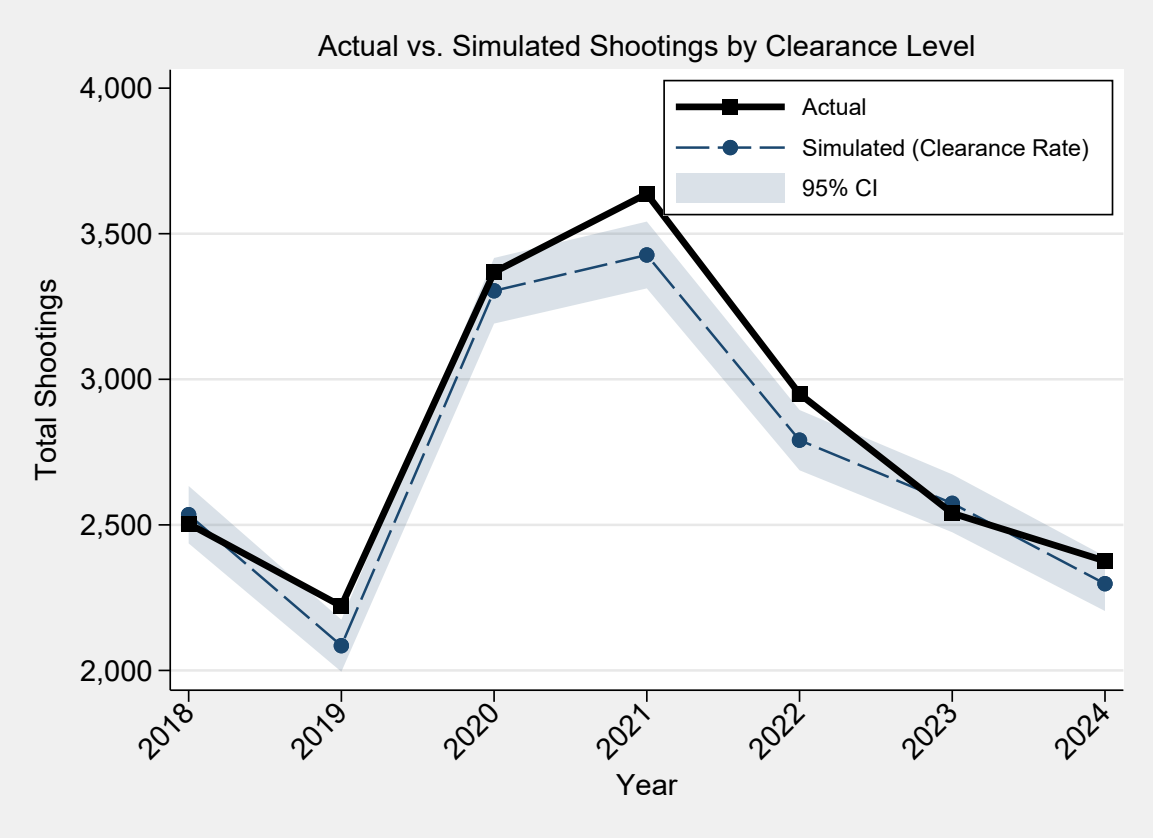


Figure 7: Actual vs. simulated total shootings by year, 2018–2024, using yearly clearance rates.

6 Discussion and Conclusion

This paper develops a shooting intensity model that captures the tail-risk structure of urban gun violence and that a small number of individuals and their immediate peers generate an outsized share of total shootings in a city. Using parameters estimated from Chicago data, we show that if shootings are concentrated among a small, identifiable subpopulation of high-rate offenders, the model closely reproduces both surges and declines in citywide shooting rates. Importantly, thin-tailed models will systematically understate the potential for shootings to rise and fall geometrically. Capturing how that concentration evolves over time requires combining a finite mixture model, a self-exciting process, a deterrence memory, and a gambler's-ruin absorbing barrier, linked across peer networks. When calibrated year by year to Chicago's actual population and shooting counts, the model reproduces the broad shape of the 2018–2024 trajectory and concentrates the effect of enforcement almost entirely within the shooter subpopulation. The fact that even the model's most permissive enforcement scenario outperforms reality is corroborated by Chicago's own clearance data

that shows the probability of an arrest for a shooting remained below 15% every year, falling below even our weakest simulated sanctions floor.

The model yields three policy implications. First, enforcement resources should concentrate on the high-risk subpopulation rather than be diluted across low-risk individuals. Second, speed matters as delayed enforcement allows excitation to compound before it decays, magnifying the number of shootings over the course of a year. Third, removing a high-risk individual from a network cools connected associates through the network's contagion lag, producing systemic relief that extends well beyond the individual case. Focused deterrence and holding offenders accountable for extreme acts of violence like shootings (Kennedy et al., 2017) is therefore best understood as a form of tail-risk hedging - a targeted, scalable intervention that stabilizes public safety by neutralizing the small fraction of actors responsible for cascading firearms violence. Public safety policy should treat tail risks of shootings as a distinct problem from average-case crime prevention, not a more extreme version of the same problem. Effective, certainty-focused policing and prosecution aimed at the highest-rate offenders is one of the few interventions with a consistent, scalable, well-replicated track record; it functions less like a bet on the average case and more like insurance against the tail. Given how far below even a modest deterrence floor Chicago's actual enforcement intensity has sat for years, the marginal payoff from closing that gap and focusing on the relatively small number of individuals responsible for most shootings is likely to be large. Progressive reforms that reduce sanctions uniformly across the offending distribution do nothing for this tail risk and, as the model shows, will mechanically increase total shootings by leaving the highest-rate sub-population and the gambler's-ruin buffer left unchecked.

A selective approach, such as that suggested by the three policy implications above, makes better use of the limited resources of the police and prosecution agencies (Spelman & Kramer, 1990). We focus on the potential for police and prosecutors serving as the front end of the criminal justice system as the key mechanism through which shootings can be reduced without generating larger correctional populations. This lofty goal seems achievable if and only if enforcement interventions can be designed to achieve higher levels of accountability the tail of the shooter distribution (as suggested by Figure 7). Efforts to increase shooting clearance rates could enhance the risk of apprehension for shooters (Cook et al., 2019) and, assuming effective prosecutions and appropriate sanctions upon conviction, generate larger deterrent and incapacitation effects in the right-tail of the gun violence risk distribution. Focused deterrence strategies also attempt to increase apprehension risks for violent gang members by using alternative enforcement efforts in the aftermath of shootings (Braga & Kennedy, 2021; Kennedy et al., 2017). While shooting investigations progress, authorities "pull every enforcement lever" available to halt violent outbreaks quickly via serving outstanding warrants, enforcing any conditions of probation and parole, increased traffic and disorder enforcement, shutting down other illicit activities such as street drug sales, and other actions. In essence, the authorities use the ongoing involvement of shooters in the criminal justice system and their varied offending activities against them. These same kinds of enforcement actions characterize gang takedowns in so-called "precision policing" schemes (Chalfin et al., 2021).

Several limitations of the present paper deserve mention. First, the model does not yield causal estimates of deterrence and incapacitation efforts. We assume a direction of effect between sanctions and offending, but the true causal structure is difficult to identify from observational data as focused enforcement efforts and offending rates are jointly determined. Second, the assumed sizes of self-excitation and network contagion are informed by the literature but remain uncertain. Small changes in these parameters can meaningfully alter model projections, and we cannot verify from arrest data alone whether the endogenous mixture model approximates the true data-generating process of shootings. Third, arrest records capture only a fraction of actual offending, meaning both the true shooting rate and the size and composition of co-offending networks are latent and only

partially observed. Networks reconstructed from co-arrest data will understate connectivity and may systematically exclude any active shooters who avoid an arrest.

Nevertheless, the model offers a practical tool for applied policy analysis and discussions of how to control shootings in a city. The underlying code is publicly available, and the framework is designed to track whether changes in policy-relevant parameters, such as shifts in sanctions regimes for shootings, produce model projections of shooting rates that are consistent with observed trends. To the extent that the model can be recalibrated as new data arrive, it provides a tractable way to evaluate whether policy levers are moving in directions consistent with reducing shooting risk.

Crime is highly concentrated in people, places, and across offending networks where a small number of individuals generate a disproportionate share of social harm (Papachristos et al., 2012; Sherman, 2007). This concentration is especially pronounced for gun violence (Braga & Cook, 2023): a tiny fraction of the population accounts for the vast majority of shootings, both as perpetrators and victims (Heller et al., 2026; Lattimore et al., 1997). Yet policy discussions on how to respond to gun violence seldom acknowledge these facts and tend to focus on aggregate trends rather than the high-risk individuals and networks driving them. The results from this model suggest that targeting the tail risk of high-risk individuals and their co-offending networks offers great potential for reducing shootings, and that criminal justice reforms should not confuse marginal deterrence for the average offender with the focused sanctions necessary to deter and incapacitate the small number of individuals driving the bulk of gun violence.

References

Agan, A., Doleac, J. L., & Harvey, A. (2023). Misdemeanor prosecution. *The Quarterly Journal of Economics*, *138*(3), 1453–1505.

Becker, G. S. (1968). Crime and punishment: An economic approach. *The Journal of Political Economy*, *169*, 176–177.

Berk, R., & Bleich, J. (2013). Statistical procedures for forecasting criminal behavior: A comparative assessment. *Criminology Public Policy*, *12*, 513.

Berk, R., Sherman, L., Barnes, G., Kurtz, E., & Ahlman, L. (2009). Forecasting murder within a population of probationers and parolees: A high stakes application of statistical learning. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*(1), 191–211.

Blumstein, A. (1998). Us criminal justice conundrum: Rising prison populations and stable crime rates. *Crime & Delinquency*, *44*(1), 127–135.

Blumstein, A., Cohen, J., & Nagin, D. (1978). *Deterrence and incapacitation: Estimating the effects of criminal sanctions on crime rates*. National Academy of Sciences Washington, DC.

Blusmtein, A. (1986). *Criminal careers and quot;career criminals,quot;: Volume i* (A. Blumstein, J. Cohen, J. A. Roth, & C. A. Visher, Eds.). The National Academies Press. https://doi.org/10.17226/922

Braga, A. A., & Cook, P. J. (2023). *Policing gun violence: Strategic reforms for controlling our most pressing crime problem*. Oxford University Press.

Braga, A. A., & Kennedy, D. M. (2021). *A framework for addressing violence and serious crime: Focused deterrence, legitimacy, and prevention*. Cambridge University Press.

Braga, A. A., Kennedy, D. M., Waring, E. J., & Piehl, A. M. (2017). Problem-oriented policing, deterrence, and youth violence: An evaluation of boston's operation ceasefire. In *Gangs* (pp. 513–543). Routledge.

Braga, A. A., MacDonald, J. M., & Tita, G. (2024, February). Defund the police?: Considerations for reducing gang violence. In D. C. Pyrooz, J. A. Densley, J. Leverso, D. C. Pyrooz, J. A. Densley, & J. Leverso (Eds.), *The oxford handbook of gangs and society*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197618158.013.41

Braga, A. A., & Piza, E. (2023, April). Serious violence in boston: Then and now.

Braga, A. A., Turchan, B. S., & Weisburd, D. L. (2026). Focused deterrence can reduce crime: A systematic review of randomized controlled trials and quasi-experiments. *Criminology & Public Policy*.

Chalfin, A., LaForest, M., & Kaplan, J. (2021). Can precision policing reduce gun violence? evidence from "gang takedowns" in new york city. *Journal of Policy Analysis and Management*, *40*(4), 1047–1082.

Chalfin, A., & McCrary, J. (2017). Criminal deterrence: A review of the literature. *Journal of Economic Literature*, *55*(1), 5–48.

Charette, Y., & Papachristos, A. V. (2017). The network dynamics of co-offending careers. *Social Networks*, *51*, 3–13.

Clarke, R. V. (1995). Situational crime prevention. *Crime and justice*, *19*, 91–150.

Cook, P. J. (1979). The clearance rate as a measure of criminal justice system effectiveness. *Journal of Public Economics*, *11*(1), 135–142.

Cook, P. J. (1980). Research in criminal deterrence: Laying the groundwork for the second decade. *Crime and Justice*, *2*, 211–268. Retrieved June 18, 2026, from http://www.jstor.org/stable/1147415

Cook, P. J. (1986). The demand and supply of criminal opportunities. *Crime and Justice*, 1–27.

Cook, P. J., Braga, A. A., Turchan, B. S., & Barao, L. M. (2019). Why do gun murders have a higher clearance rate than gunshot assaults? *Criminology & Public Policy*, *18*(3), 525–551.

Davis, A. J. (2018). The progressive prosecutor: An imperative for criminal justice reform. *Fordham L. Rev. Online*, *87*, 8.

Durlauf, S. N., & Nagin, D. S. (2011). Imprisonment and crime: Can both be reduced? *Criminology & Public Policy*, *10*(1), 13–54.

Epstein, R. A. (2012). *The theory of gambling and statistical logic*. Academic Press.

Farrington, D. P. (1986). Age and crime. *Crime and Justice*, *7*, 189–250.

Fryer, D. (2020). Race, reform, & progressive prosecution. *The Journal of Criminal Law and Criminology (1973-)*, *110*(4), 769–802.

Green, B., Horel, T., & Papachristos, A. V. (2017). Modeling contagion through social networks to explain and predict gunshot violence in chicago, 2006 to 2014. *JAMA internal medicine*, *177*(3), 326–333.

Harmon, R. A. (2016). Why arrest. *Mich. L. Rev.*, *115*, 307.

Heller, S. B., Jakubowski, B., Jelveh, Z., & Kapustin, M. (2026). Machine learning can predict shooting victimization well enough to help prevent it. *The Review of Economics and Statistics*, 1–16. https://doi.org/10.1162/rest_a_01519

Hogan, T. P. (2022). De-prosecution and death: A synthetic control analysis of the impact of de-prosecution on homicides. *Criminology & Public Policy*, *21*(3), 489–534.

Jensen, J. L. W. V. (1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta mathematica*, *30*(1), 175–193.

Kennedy, D. M., Kleiman, M. A., & Braga, A. A. (2017). Beyond deterrence: Strategies of focus and fairness. In *Handbook of crime prevention and community safety* (pp. 157–182). Routledge.

Kennedy, D. M., Piehl, A. M., & Braga, A. A. (1996). Youth violence in boston: Gun markets, serious youth offenders, and a use-reduction strategy. *Law and Contemporary Problems*, *59*(1), 147–196.

Lattimore, P. K., Linster, R. L., & MacDonald, J. M. (1997). Risk of death among serious young offenders. *Journal of Research in Crime and Delinquency*, *34*(2), 187–209.

Lauritsen, J. L., Sampson, R. J., & Laub, J. H. (1991). The link between offending and victimization among adolescents. *Criminology*, *29*(2), 265–292.

Leovy, J. (2015). *Ghettoside: A true story of murder in america*. One World.

Mitchell, O., & Petersen, N. (2025). The rise of progressive prosecutors in the united states: Politics, prospects, and perils. *Annual Review of Criminology*, *8*, 459–484. https://doi.org/10.1146/annurev-criminol-111523-122427

Mueller-Smith, M., & T. Schnepel, K. (2021). Diversion in the criminal justice system. *The Review of Economic Studies*, *88*(2), 883–936. https://doi.org/10.1093/restud/rdaa030

Nagin, D. S. (2013). Deterrence in the twenty-first century. *Crime and Justice*, *42*(1), 199–263.

Nagin, D. S., Solow, R. M., & Lum, C. (2015). Deterrence, criminal opportunities, and police. *Criminology*, *53*(1), 74–100.

Nieto, A., Davies, T., & Borrion, H. (2022). "offending with the accomplices of my accomplices": Evidence and implications regarding triadic closure in co-offending networks. *Social Networks*, *70*, 325–333.

Papachristos, A. V., Braga, A. A., & Hureau, D. M. (2012). Social networks and the risk of gunshot injury. *Journal of Urban Health*, *89*(6), 992–1003.

Piquero, A. R., Farrington, D. P., & Blumstein, A. (2003). The criminal career paradigm. *Crime and Justice*, *30*, 359–506.

Raphael, S., & Stoll, M. A. (2013). *Why are so many americans in prison?* Russell Sage Foundation.

Reaves, B. A. (2013). Felony defendants in large urban counties, 2009—statistical tables. *Washington, DC: US Department of Justice*.

Shen, Y., Bushway, S. D., Sorensen, L. C., & Smith, H. L. (2020). Locking up my generation: Cohort differences in prison spells over the life course. *Criminology*, *58*(4), 645–677.

Sherman, L. W. (2007). The power few: Experimental criminology and the reduction of harm: The 2006 joan mccord prize lecture. *Journal of experimental criminology*, *3*(4), 299–321.

Spelman, W. (1993). *Criminal incapacitation*. Springer Science & Business Media.

Spelman, W., & Kramer, S. (1990). *Repeat offender programs for law enforcement*. Police Executive Research Forum Washington, DC.

Taleb, N. N. (2020). Statistical consequences of fat tails: Real world preasymptotics, epistemology, and applications. *arXiv preprint arXiv:2001.10488*.

Thomas, K. J., & Pogarsky, G. (2026). Recency and decay effects in sanction risk updating. *Criminology*.

Tracy, M., Braga, A. A., & Papachristos, A. V. (2016). The transmission of gun and other weapon-involved violence within social networks. *Epidemiologic Reviews*, *38*(1), 70–86.

Travis, J., Western, B., & Redburn, F. S. (2014). *The growth of incarceration in the united states: Exploring causes and consequences*. National Academy of Sciences.

Wolfgang, M. E. (1958). *Patterns in criminal homicide*. University of Pennsylvania Press.

Appendix A: Model Parameters and Descriptive Tables

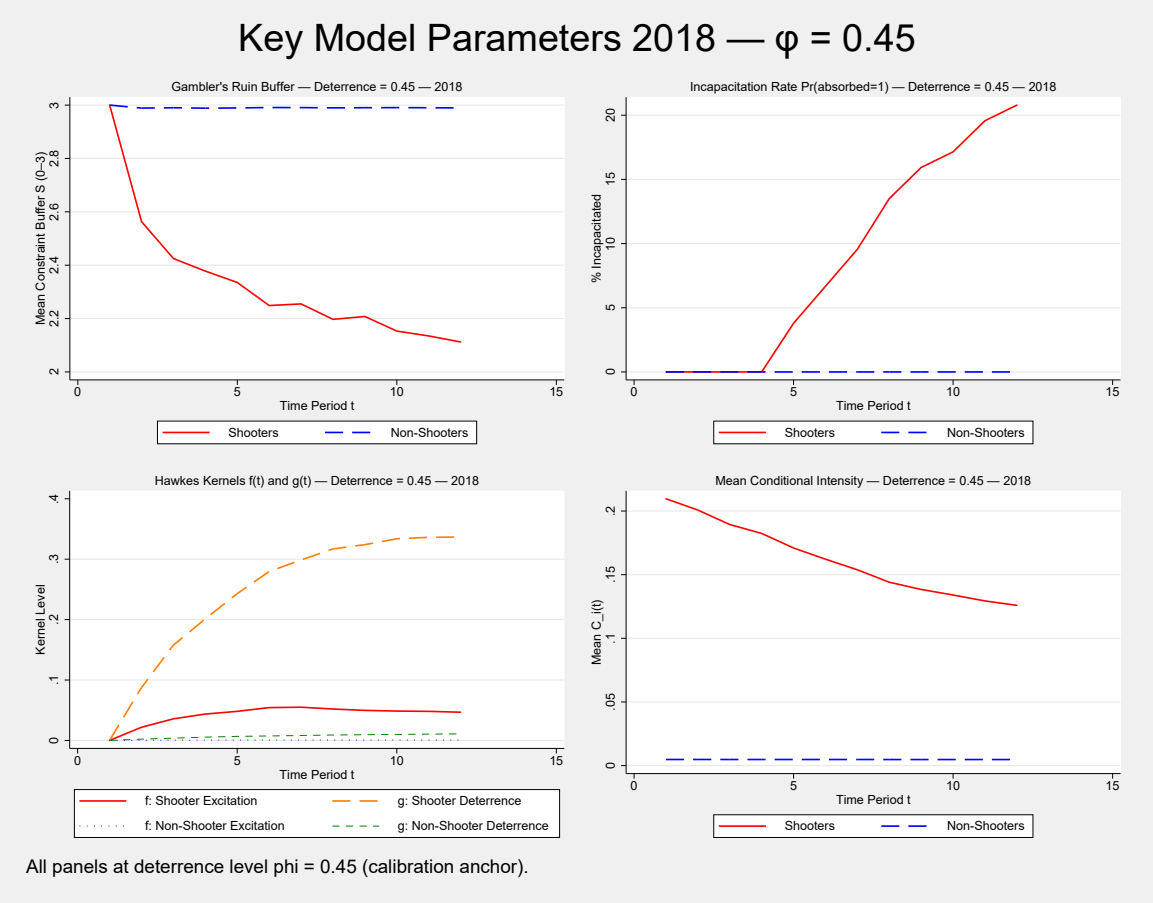


Figure A1: Key model parameters, 2018 ($\phi = 0.45$). Panels show the Gambler's Ruin buffer $S_i(t)$, incapacitation rate, Hawkes kernels $f_i(t)$ and $g_i(t)$, and mean conditional intensity $C_i(t)$, each by latent type (shooter vs. non-shooter).

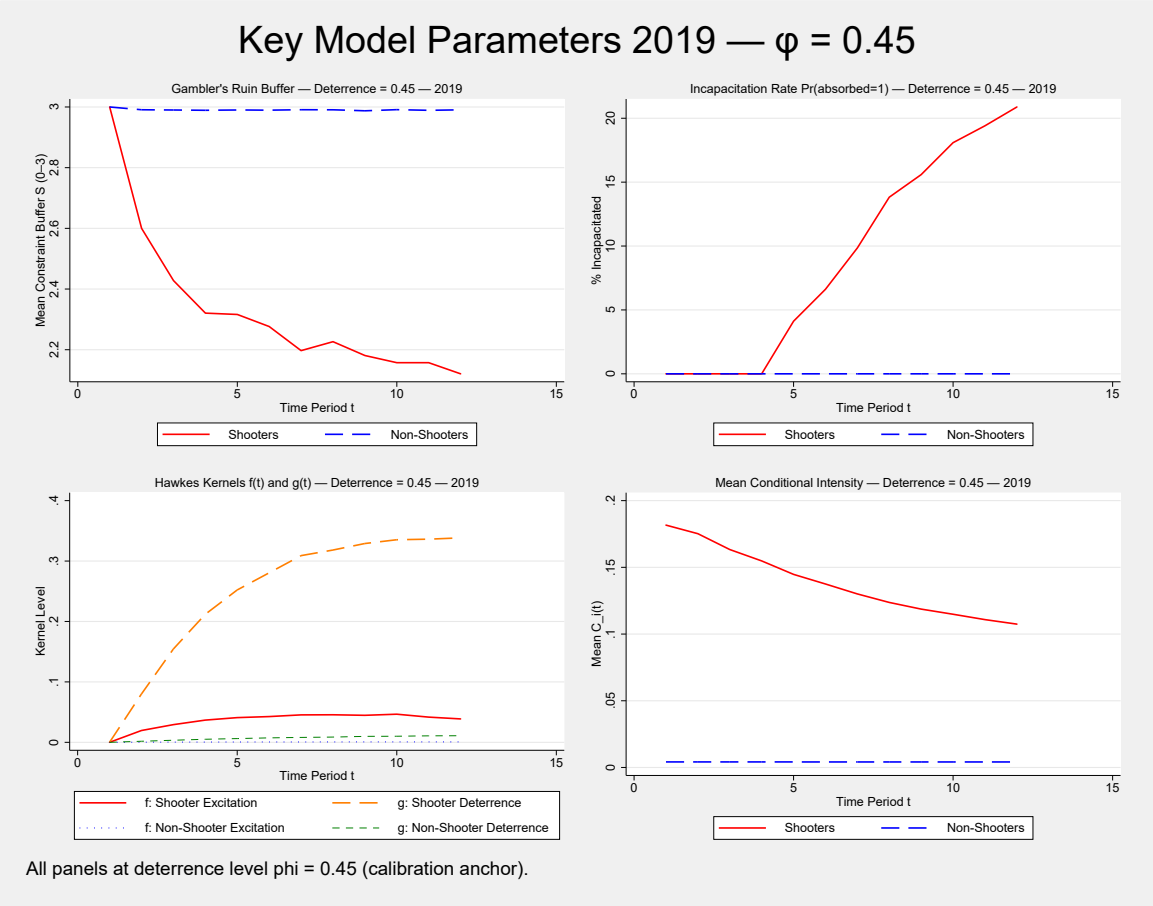


Figure A2: Key model parameters, 2019 ($\phi = 0.45$). Panels show the Gambler's Ruin buffer $S_i(t)$, incapacitation rate, Hawkes kernels $f_i(t)$ and $g_i(t)$, and mean conditional intensity $C_i(t)$, each by latent type (shooter vs. non-shooter).

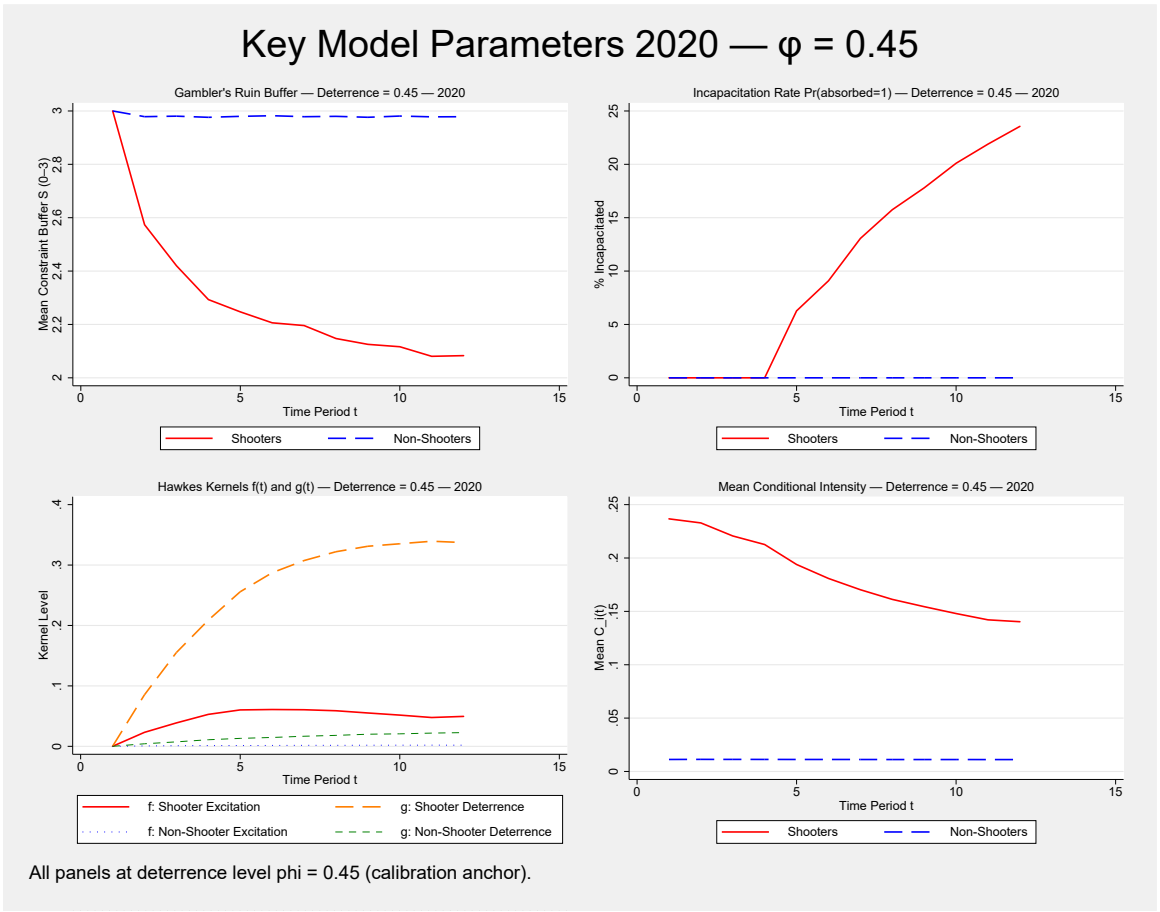


Figure A3: Key model parameters, 2020 ($\phi = 0.45$). Panels show the Gambler's Ruin buffer $S_i(t)$, incapacitation rate, Hawkes kernels $f_i(t)$ and $g_i(t)$, and mean conditional intensity $C_i(t)$, each by latent type (shooter vs. non-shooter).

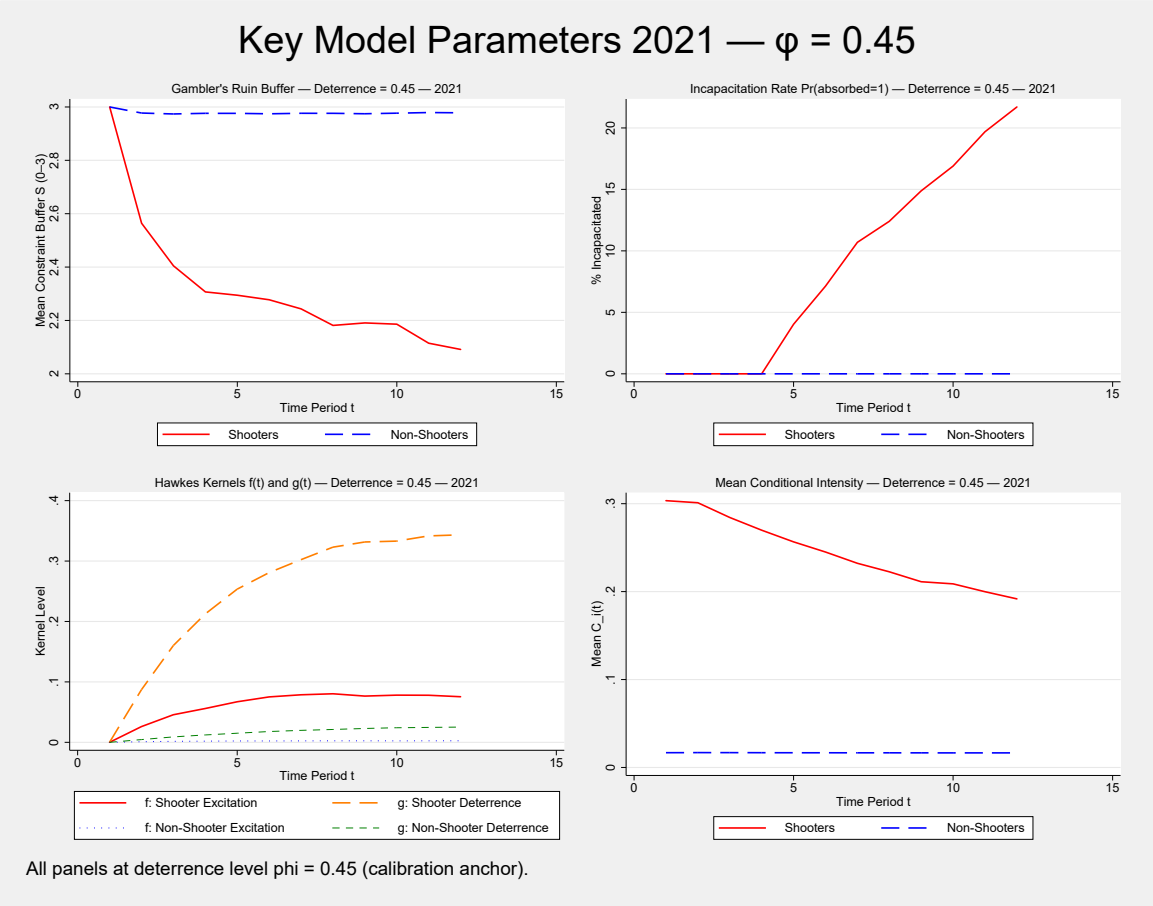


Figure A4: Key model parameters, 2021 ($\phi = 0.45$). Panels show the Gambler's Ruin buffer $S_i(t)$, incapacitation rate, Hawkes kernels $f_i(t)$ and $g_i(t)$, and mean conditional intensity $C_i(t)$, each by latent type (shooter vs. non-shooter).

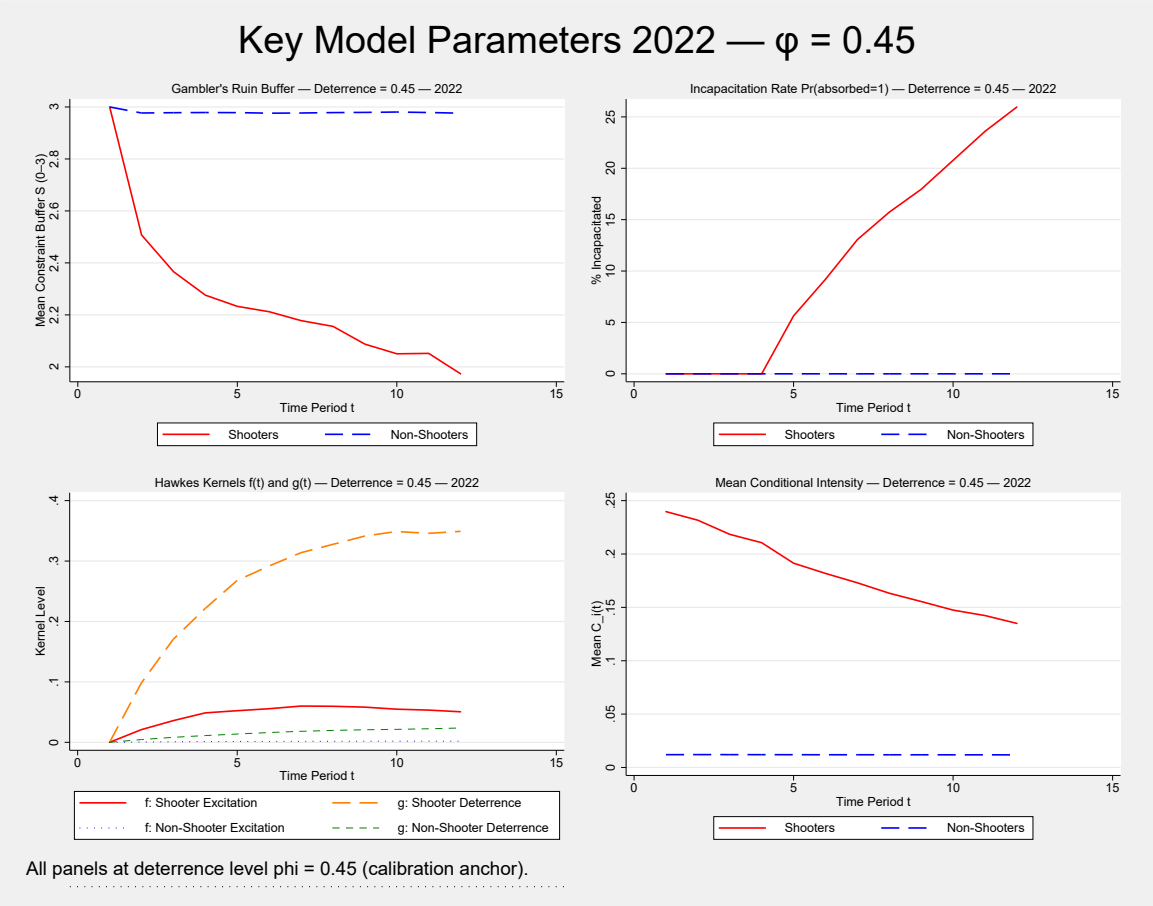


Figure A5: Key model parameters, 2022 ($\phi = 0.45$). Panels show the Gambler's Ruin buffer $S_i(t)$, incapacitation rate, Hawkes kernels $f_i(t)$ and $g_i(t)$, and mean conditional intensity $C_i(t)$, each by latent type (shooter vs. non-shooter).

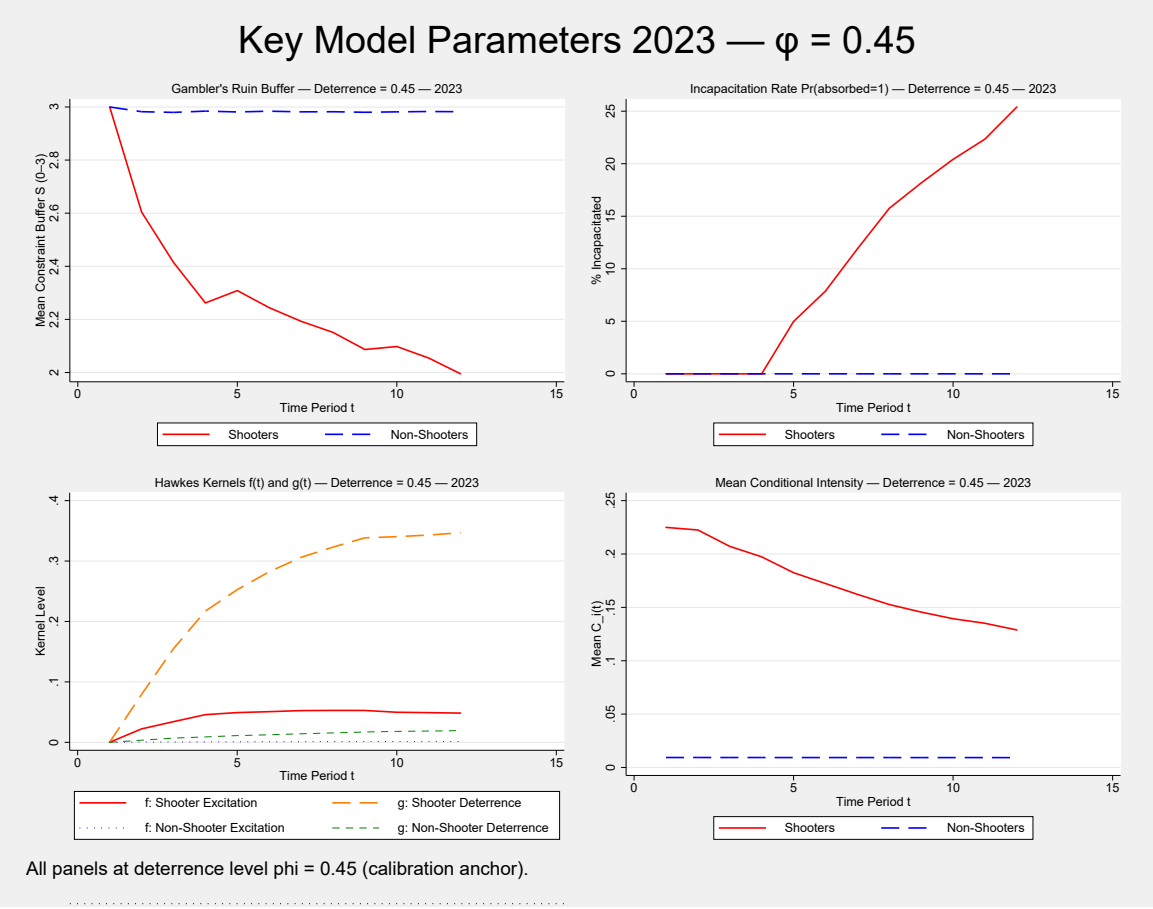


Figure A6: Key model parameters, 2023 ($\phi = 0.45$). Panels show the Gambler's Ruin buffer $S_i(t)$, incapacitation rate, Hawkes kernels $f_i(t)$ and $g_i(t)$, and mean conditional intensity $C_i(t)$, each by latent type (shooter vs. non-shooter).

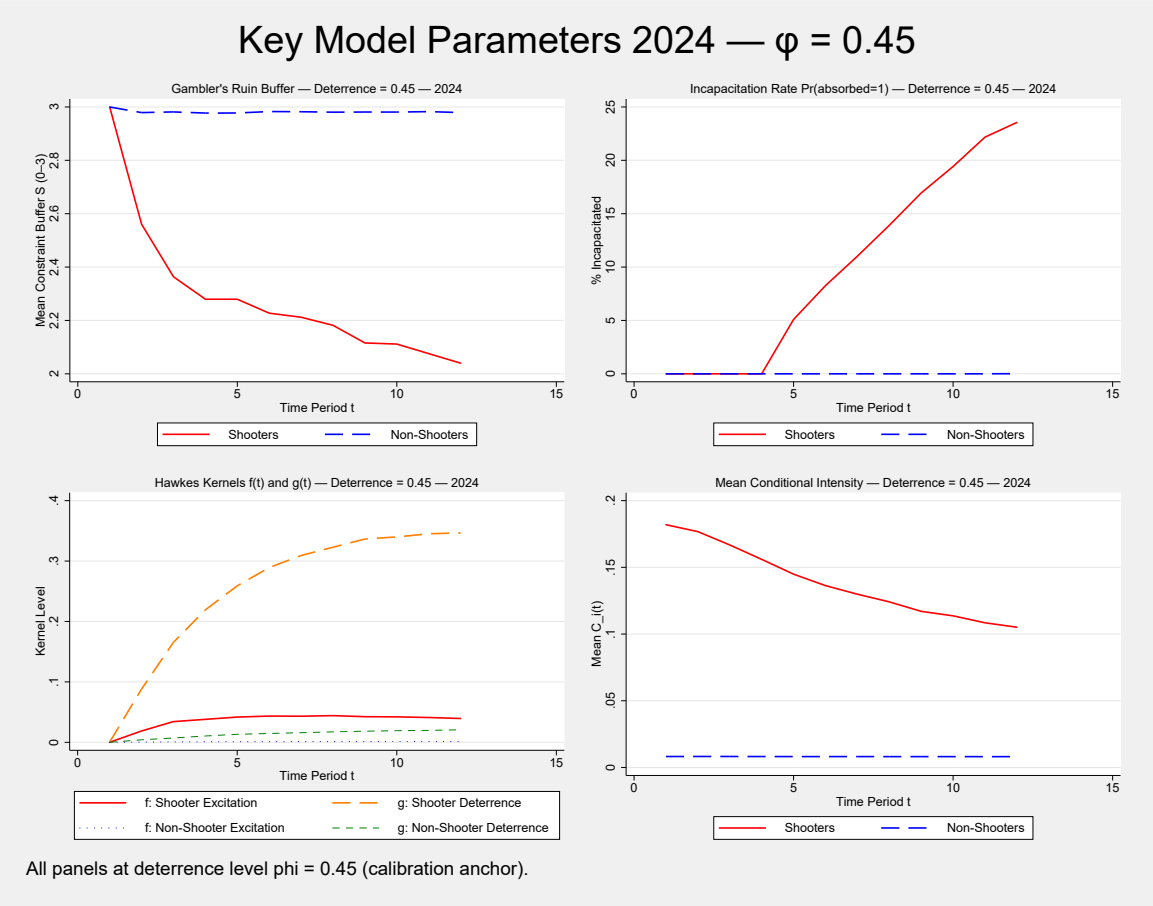


Figure A7: Key model parameters, 2024 ($\phi = 0.45$). Panels show the Gambler's Ruin buffer $S_i(t)$, incapacitation rate, Hawkes kernels $f_i(t)$ and $g_i(t)$, and mean conditional intensity $C_i(t)$, each by latent type (shooter vs. non-shooter).

Table A1: Simulated Shootings by Latent Type, Deterrence Scenario, and Year

Year	Scenario	Non-Shooters	Shooters	Total	Actual
2018	$\phi_0 = 0.15$	728	1,615	2,343	
	$\phi_0 = 0.25$	710	1,485	2,195	
	$\phi_0 = 0.35$	726	1,377	2,103	
	$\phi_0 = 0.45$	752	1,291	2,043	
	$\phi_0 = 0.55$	738	1,123	1,861	
	Actual				*2,504*
2019	$\phi_0 = 0.15$	631	1,441	2,072	
	$\phi_0 = 0.25$	658	1,287	1,945	
	$\phi_0 = 0.35$	720	1,266	1,986	
	$\phi_0 = 0.45$	698	1,147	1,845	
	$\phi_0 = 0.55$	634	987	1,621	
	Actual				*2,221*
2020	$\phi_0 = 0.15$	1,047	2,283	3,330	
	$\phi_0 = 0.25$	1,031	2,113	3,144	
	$\phi_0 = 0.35$	953	1,919	2,872	
	$\phi_0 = 0.45$	1,048	1,705	2,753	
	$\phi_0 = 0.55$	1,049	1,478	2,527	
	Actual				*3,368*
2021	$\phi_0 = 0.15$	1,064	2,351	3,415	
	$\phi_0 = 0.25$	1,095	2,260	3,355	
	$\phi_0 = 0.35$	1,099	2,054	3,153	
	$\phi_0 = 0.45$	1,070	1,887	2,957	
	$\phi_0 = 0.55$	1,064	1,683	2,747	
	Actual				*3,637*
2022	$\phi_0 = 0.15$	851	1,940	2,791	
	$\phi_0 = 0.25$	864	1,878	2,742	
	$\phi_0 = 0.35$	913	1,697	2,610	
	$\phi_0 = 0.45$	862	1,446	2,308	
	$\phi_0 = 0.55$	893	1,335	2,228	
	Actual				*2,950*
2023	$\phi_0 = 0.15$	752	1,785	2,537	
	$\phi_0 = 0.25$	728	1,590	2,318	
	$\phi_0 = 0.35$	785	1,483	2,268	
	$\phi_0 = 0.45$	678	1,252	1,930	
	$\phi_0 = 0.55$	718	1,158	1,876	
	Actual				*2,540*
2024	$\phi_0 = 0.15$	736	1,585	2,321	
	$\phi_0 = 0.25$	736	1,475	2,211	
	$\phi_0 = 0.35$	737	1,391	2,128	
	$\phi_0 = 0.45$	742	1,205	1,947	
	$\phi_0 = 0.55$	758	1,102	1,860	
	Actual				*2,376*

Table A2: Shooting Clearance Rates by Year and Victimization Type

	Homicide		Non-Homicide		Overall	
Year	Incidents	Rate	Incidents	Rate	Incidents	Rate
2018	614	27.85%	1,960	9.85%	2,574	14.14%
2019	534	25.66%	1,746	10.25%	2,280	13.86%
2020	827	32.16%	2,620	7.10%	3,447	13.11%
2021	834	23.74%	2,852	5.86%	3,686	9.90%
2022	743	29.34%	2,246	7.52%	2,989	12.95%
2023	644	26.24%	1,935	8.42%	2,579	12.87%
2024	611	25.70%	1,803	10.43%	2,414	14.29%
Total	4,807	27.38%	15,162	8.21%	19,969	12.83%

Notes: Annual arrest rates for shootings based on authors' calculations from Chicago data.

Table A3: Simulated Shootings by Latent Type, Deterrence Scenario, and Year

Year	ϕ_0	Non-Shooters	Shooters	Total	Actual	95% CI
2018	0.106	729	1,806	**2,535**	*2504*	[2,436, 2,634]
2019	0.114	634	1,451	**2,085**	*2221*	[1,996, 2,174]
2020	0.141	1,045	2,259	**3,304**	*3368*	[3,191, 3,417]
2021	0.138	1,064	2,363	**3,427**	*3637*	[3,312, 3,542]
2022	0.131	863	1,928	**2,791**	*2950*	[2,687, 2,895]
2023	0.099	750	1,824	**2,574**	*2540*	[2,475, 2,673]
2024	0.129	726	1,572	**2,298**	*2376*	[2,204, 2,392]

Appendix B: Stata Replication Code

This appendix reproduces the Stata programs used to generate the simulation results reported in the paper, ordered to follow the structure of the paper. Code that produced diagnostic figures or summary tables not shown in the paper has been removed, so that the listings below can be copied and run as-is to reproduce the paper's key tables and figures. Section B.1 reproduces the Jensen's inequality figure (`jensen_inequality_two.png`). Section B.2 reproduces the gamma and geometric toy-simulation figures (`Gamma_2020.pdf` and `Geometric_mean2020.pdf`). Section B.3 reproduces the full model: the table of simulated shootings by latent type, deterrence scenario, and year (`table_shootings_all_years.tex`; the 2021 panel is reported in the paper), the simulated-versus-actual comparison (`fig_simulated_vs_actual_shootings.pdf` and `cross_year_comparison.rtf`), the monthly trajectory figure (`combined_trajectories.pdf`), and the yearly model-parameter snapshots (`fig_yearly_snapshot_2021.pdf`). The network map figure (`network_maps_lines_combined.png`) is constructed from administrative network data and is not produced by these simulation programs. Three small additions, each flagged with a comment in the code, align output filenames with those referenced in the paper. Set the working directory at the top of each program before running. An R version of the code is available upon request and produces the same estimates with only small differences based on internal randomization differences between programs.

B.1 Jensen's Inequality Figure

This program draws the convex shooting-response curve with two secant chords, illustrating how enforcement pullback widens the Jensen gap.

```
* ============================================================
* jensen_inequality_two_chords.do
* Jensen's Inequality: Shooting Response and Deterrence
*
* Compares two enforcement regimes:
*   - Orange: baseline (x1b=0.22 to x2=0.78)
*   - Red: more enforcement pullback (x1p=0.06 to x2=0.78)
*
* Shows how increasing enforcement pullback shifts the chord
* upward and widens the Jensen gap.
*
* OUTPUT: jensen_inequality_two.pdf (and .png)
* ============================================================
cd "<project directory>"

* PARAMETERS
local alpha = 2.2         // decay rate: f(x) = exp(-alpha*x)
local x1_original = 0.22  // low enforcement (baseline)
local x1_extreme  = 0.06  // enforcement pullback (more severe)
local x2 = 0.78           // full enforcement (same for both)
local xbar_original = 0.50  // E[X]: midpoint of baseline range
local xbar_extreme = 0.42   // E[X]: midpoint of extreme range
local xtitle_txt "Consequences for illegal gun carry (enforcement intensity)"
local ytitle_txt "Shooting response"
local x1b_lbl    `""x{subscript:1b}" "(baseline)""'
local x1p_lbl    `""x{subscript:1p}" "(pullback)""'
local x2_lbl     `""x{subscript:2}" "(full enforcement)""'

* STEP 1: CURVE, SECANTS, AND SHADED-GAP DATA
clear
set obs 200
gen x       = 0.02 + (_n - 1) * (0.94 / 199)
gen y_curve = exp(-`alpha' * x)

* Secant 1 (baseline): chord from x1_original to x2
gen y_chord_original = exp(-`alpha'*`x1_original') + ///
                       (exp(-`alpha'*`x2') - exp(-`alpha'*`x1_original')) * ///
```

```
                      (x - `x1_original') / (`x2' - `x1_original') ///
                      if x >= `x1_original' & x <= `x2'

* Secant 2 (extreme): chord from x1_extreme to x2
gen y_chord_extreme = exp(-`alpha'*`x1_extreme') + ///
                      (exp(-`alpha'*`x2') - exp(-`alpha'*`x1_extreme')) * ///
                      (x - `x1_extreme') / (`x2' - `x1_extreme') ///
                      if x >= `x1_extreme' & x <= `x2'

* Shaded Jensen gap 1 (baseline)
gen y_gap1_top = y_chord_original
gen y_gap1_bot = y_curve if x >= `x1_original' & x <= `x2'

* Shaded Jensen gap 2 (extreme)
gen y_gap2_top = y_chord_extreme
gen y_gap2_bot = y_curve if x >= `x1_extreme' & x <= `x2'

* STEP 2: KEY POINT ROWS
local n0 = _N
set obs `=`n0' + 12'

* Orange dots at x1_original, x2, and E[f(X)] on original secant
gen y_orange = .
replace x        = `x1_original'   in `=`n0'+1'
replace y_orange = exp(-`alpha'*`x1_original') in `=`n0'+1'

replace x        = `x2'   in `=`n0'+2'
replace y_orange = exp(-`alpha'*`x2') in `=`n0'+2'

replace x        = `xbar_original' in `=`n0'+3'
replace y_orange = (exp(-`alpha'*`x1_original') + exp(-`alpha'*`x2')) / 2 in `=`n0'+3'

* Red dots at x1_extreme, x2, and E[f(X)] on extreme secant
gen y_red = .
replace x      = `x1_extreme'   in `=`n0'+4'
replace y_red  = exp(-`alpha'*`x1_extreme') in `=`n0'+4'

replace x      = `x2'   in `=`n0'+5'
replace y_red  = exp(-`alpha'*`x2') in `=`n0'+5'

replace x      = `xbar_extreme' in `=`n0'+6'
replace y_red  = (exp(-`alpha'*`x1_extreme') + exp(-`alpha'*`x2')) / 2 in `=`n0'+6'

* Blue dot at f(E[X]) on the curve (original baseline)
gen y_blue = .
replace x      = `xbar_original' in `=`n0'+7'
replace y_blue = exp(-`alpha'*`xbar_original') in `=`n0'+7'

* Arrow 1 (original): from f(E[X]) up to E[f(X)]
gen x_arr1    = .
gen y_arr1_lo = .
gen y_arr1_hi = .
replace x_arr1    = `xbar_original' in `=`n0'+8'
replace y_arr1_lo = exp(-`alpha'*`xbar_original') in `=`n0'+8'
replace y_arr1_hi = (exp(-`alpha'*`x1_original') + exp(-`alpha'*`x2')) / 2 in `=`n0'+8'

* Arrow 2 (extreme): from f(E[X]) up to E[f(X)]
gen x_arr2    = .
gen y_arr2_lo = .
gen y_arr2_hi = .
replace x_arr2    = `xbar_extreme' in `=`n0'+9'
replace y_arr2_lo = exp(-`alpha'*`xbar_extreme') in `=`n0'+9'
replace y_arr2_hi = (exp(-`alpha'*`x1_extreme') + exp(-`alpha'*`x2')) / 2 in `=`n0'+9'

* Invisible points carrying inequality captions
gen y_ineq_lbl1 = .
replace x           = `xbar_original' + 0.12 in `=`n0'+10'
replace y_ineq_lbl1 = (exp(-`alpha'*`xbar_original') + ///
                      (exp(-`alpha'*`x1_original') + exp(-`alpha'*`x2')) / 2) / 2 + 0.03 ///
                      in `=`n0'+10'

gen y_ineq_lbl2 = .
```

```
replace x           = `xbar_extreme' + 0.12 in `=`n0'+11'
replace y_ineq_lbl2 = (exp(-`alpha'*`xbar_extreme') + ///
                      (exp(-`alpha'*`x1_extreme') + exp(-`alpha'*`x2')) / 2) / 2 + 0.05 ///
                      in `=`n0'+11'

* STEP 3: TEXT LABELS
gen lbl_orange = ""
replace lbl_orange = "E[f(X)]" in `=`n0'+3'

gen lbl_red = ""
replace lbl_red = "E[f(X)]" in `=`n0'+6'

gen lbl_blue = ""
replace lbl_blue = "f(E[X])" in `=`n0'+7'

gen lbl_ineq1 = ""
replace lbl_ineq1 = "E[X]={Baseline}: f(E[X]) {&le} E[f(X)]" in `=`n0'+10'

gen lbl_ineq2 = ""
replace lbl_ineq2 = "E[X]={Pullback}: f(E[X]) {&le} E[f(X)]" in `=`n0'+11'

* STEP 4: FIGURE
twoway ///
    /// 1. Shaded Jensen gap 1 (dark purple)
    (rarea y_gap1_top y_gap1_bot x if x >= `x1_original' & x <= `x2', ///
        color(purple%50) lwidth(none)) ///
    /// 2. Shaded Jensen gap 2 (light red)
    (rarea y_gap2_top y_gap2_bot x if x >= `x1_extreme' & x <= `x2', ///
        color(red%10) lwidth(none)) ///
    /// 3. Convex decreasing shooting response curve
    (line y_curve x, lcolor(navy) lwidth(vthick)) ///
    /// 4. Secant 1 (orange), baseline
    (line y_chord_original x, lcolor(orange) lpattern(dash) lwidth(thick)) ///
    /// 5. Secant 2 (red), more pullback
    (line y_chord_extreme x, lcolor(red) lpattern(dash) lwidth(thick)) ///
    /// 6. Orange dots at x1_original, x2, and E[f(X)]
    (scatter y_orange x in `=`n0'+1'/`=`n0'+3', ///
        msymbol(circle) mcolor(orange) msize(medium) ///
        mlabel(lbl_orange) mlabcolor(orange) ///
        mlabposition(3) mlabsize(small)) ///
    /// 7. Red dots at x1_extreme, x2, and E[f(X)]
    (scatter y_red x in `=`n0'+4'/`=`n0'+6', ///
        msymbol(circle) mcolor(red) msize(medium) ///
        mlabel(lbl_red) mlabcolor(red) ///
        mlabposition(3) mlabsize(small)) ///
    /// 8. Blue dot at f(E[X]) on the curve
    (scatter y_blue x in `=`n0'+7'/`=`n0'+7', ///
        msymbol(circle) mcolor(navy) msize(medium) ///
        mlabel(lbl_blue) mlabcolor(navy) ///
        mlabposition(9) mlabsize(small)) ///
    /// 9. Arrow 1 (original)
    (pcbarrow y_arr1_lo x_arr1 y_arr1_hi x_arr1, ///
        lcolor(orange) mcolor(orange) ///
        lwidth(medthick) msize(small)) ///
    /// 10. Arrow 2 (extreme)
    (pcbarrow y_arr2_lo x_arr2 y_arr2_hi x_arr2, ///
        lcolor(red) mcolor(red) ///
        lwidth(medthick) msize(small)) ///
    /// 11. Inequality captions
    (scatter y_ineq_lbl1 x in `=`n0'+10'/`=`n0'+10', ///
        msymbol(none) ///
        mlabel(lbl_ineq1) mlabcolor(orange) ///
        mlabposition(3) mlabsize(small)) ///
    (scatter y_ineq_lbl2 x in `=`n0'+11'/`=`n0'+11', ///
        msymbol(none) ///
        mlabel(lbl_ineq2) mlabcolor(red) ///
        mlabposition(3) mlabsize(small)) ///
    , ///
    xline(`xbar_original' `xbar_extreme', lpattern(dot) lcolor(gs1) lwidth(medthin)) ///
    xlabel(`x1_extreme'  `"`x1p_lbl'"' ///
           `x1_original' `"`x1b_lbl'"' ///
           `x2'          `"`x2_lbl'"', ///
```

```
          noticks labsize(small) labgap(small)) ///
    ylabel(0 "0" 1 "1", nogrid labsize(vsmall) angle(0)) ///
    xtitle("`xtitle_txt'", size(small)) ///
    ytitle("`ytitle_txt'", size(small)) ///
    legend(order(3 "f(x): shooting response (convex)" ///
                 4 "Secant (orange, baseline: x{subscript:1b} to x{subscript:2})" ///
                 5 "Secant (red, pullback: x{subscript:1p} to x{subscript:2})" ///
                 1 "Jensen's gap (baseline)" 2 "Jensen's gap (pullback)") ///
          pos(1) ring(0) cols(1) size(vsmall) ///
          region(lcolor(gs10))) ///
    scheme(s1color) ///
    xsize(10) ysize(7)

graph export "jensen_inequality_two.pdf", replace
graph export "jensen_inequality_two.png", replace width(2000)
```

B.2 Gamma and Geometric Toy Simulations

This program generates the two toy-simulation figures for the 2020 cohort ($N = 781$ offenders): heterogeneity in criminal propensity under a gamma distribution, and offenses before first deterrence under a geometric distribution.

```
*==============================================================================
* SECTION 1: GAMMA DISTRIBUTION - Variation in Criminal Propensity
* Models heterogeneity in offending using Gamma(alpha=2, theta), where
* alpha (shape) controls skewness/concentration and theta (scale) controls
* spread. Increasing theta stretches the distribution, producing more
* high-rate offenders in the right tail.
* Simulation based on N=781 offenders (2020 analytic sample).
* OUTPUT: Gamma_2020.pdf (figure used in the paper)
*==============================================================================
clear
set obs 781
set seed 123456

* Stata's rgamma(shape, scale) uses scale theta; mean = alpha x theta
gen gamma1 = rgamma(2, 1.0)    // Gamma(2, 1.0): expected mean = 2.0
gen gamma2 = rgamma(2, 1.3)    // Gamma(2, 1.3): expected mean = 2.6

label var gamma1 "Gamma(a=2, t=1.0): baseline"
label var gamma2 "Gamma(a=2, t=1.3): 30% scale increase"

* Compute totals as scalars to avoid gsort corruption
quietly sum gamma1
scalar totalg1 = r(sum)
quietly sum gamma2
scalar totalg2 = r(sum)

display "Total gamma1 (theta=1.0): " totalg1
display "Total gamma2 (theta=1.3): " totalg2
display "Difference:               " totalg2 - totalg1
display "% difference:             " (totalg2 - totalg1) / totalg1 * 100

sum gamma1 gamma2

twoway (histogram gamma1, freq) ///
       (histogram gamma2, freq color(red%30)), ///
       title("Relative Variation in Criminal Propensity by Scale Parameter {&theta}", size(medium)) ///
       xtitle("Criminal Propensity ({&gamma})") ///
       ytitle("Frequency") ///
       legend(order(1 "{&theta}=1.0 (baseline)" ///
                    2 "{&theta}=1.3 (30% scale increase)")) ///
       note("N=781 offenders. {&Gamma}({&alpha}=2, {&theta}) distribution." ///
            "Increasing {&theta} from 1.0 to 1.3 raises the expected mean from 2.0 to 2.6 and shifts mass" ///
            "toward higher propensity, stretching the right tail.") ///
       scheme(sj)
graph export "Gamma_2020.pdf", replace   // filename referenced in main.tex

```

```
*==============================================================================
* SECTION 2: GEOMETRIC DISTRIBUTION - Offenses Before First Deterrence
*
* Models the number of offenses an offender commits before being deterred
* for the first time. Generated via inverse transform sampling from U(0,1).
*
* Formula:  outcome = int(ln(u) / ln(p_offense)) + 1,  u ~ Uniform(0,1)
*
* Parameterization:
*   p_offense = probability of committing an offense in a given period
*   p_deter   = 1 - p_offense = probability of being deterred (stopping)
*   Mean E[Y] = 1/p_deter;  Var[Y] = (1 - p_deter)/p_deter^2
*
* Parameters used:
*   p_deter  : 0.25  0.35  0.45  0.55
*   p_offense: 0.75  0.65  0.55  0.45
*   E[Y]     : 4.0   2.9   2.2   1.8
*
* Lower deterrence -> higher p_offense -> longer right tail
* OUTPUT: Geometric_mean2020.pdf (figure used in the paper)
*==============================================================================
clear
set obs 781
set seed 1234

* Generate uniform random variable for inverse transform sampling
gen u = uniform()

* Generate geometric outcomes for each deterrence probability
foreach p_d of numlist 25 35 45 55 {

    local p_deter   = `p_d' / 100
    local p_offense = 1 - `p_deter'

    local pd_str = string(`p_deter',   "%4.2f")
    local po_str = string(`p_offense', "%4.2f")
    local mean   = string(1 / `p_deter', "%4.1f")
    local var    = string((1 - `p_deter') / (`p_deter'^2), "%4.2f")

    * Geometric draw: offenses until first deterrence
    gen outcome_`p_d' = int(ln(u) / ln(`p_offense')) + 1
    label var outcome_`p_d' ///
        "p_deter=`pd_str' | p_offense=`po_str' | E[Y]=`mean' | Var=`var'"
}

* Summary statistics
sum outcome*

* Overlay histogram: all four deterrence levels
twoway ///
    (histogram outcome_25, percent color(blue%60)   lwidth(medium)) ///
    (histogram outcome_35, percent color(red%50)    lwidth(medium)) ///
    (histogram outcome_45, percent color(green%40)  lwidth(medium)) ///
    (histogram outcome_55, percent color(orange%30) lwidth(medium)) ///
    , xscale(range(1 .)) ///
    xlabel(1(2)30) ///
    title("Simulation of 781 Offenders by Deterrence Probability", size(medium)) ///
    subtitle("Lower deterrence probability -> longer right tail") ///
    xtitle("Offenses before first deterrence ({&gamma})") ///
    ytitle("Percent") ///
    legend(order( ///
        1 "p{subscript:deter}=0.25 E[Y]=4.0" ///
        2 "p{subscript:deter}=0.35 E[Y]=2.9" ///
        3 "p{subscript:deter}=0.45 E[Y]=2.2" ///
        4 "p{subscript:deter}=0.55 E[Y]=1.8") ///
        ring(0) pos(2) cols(1)) ///
    note("Geometric distribution; p_offense = prob. of offending per period;" ///
         " E[Y] = 1/p_deter;  Var[Y] = (1 - p_deter) / p_deter^2") scheme(sj)
graph export "Geometric_mean2020.pdf", replace   // filename referenced in main.tex
```

B.3 Gambler's Ruin Point Process Self-Exciting Mixture Model

This program simulates the four-component model of shooting intensity (finite mixture, Hawkes self-excitation, deterrence memory, and gambler's ruin barrier) for Chicago's 2018–2024 at-risk population under five deterrence scenarios, $\phi_0 \in \{0.15, 0.25, 0.35, 0.45, 0.55\}$. Requires the community-contributed `estout` package (`ssc install estout`).

```
* ============================================================
* GAMBLER'S RUIN POINT PROCESS SELF-EXCITING MIXTURE MODEL
* ============================================================
* Simulates gun violence in a social network as a self-exciting
* (Hawkes) point process subject to a gambler's ruin
* absorbing-barrier constraint. Individuals are drawn from a
* two-component finite mixture: a small high-risk "shooter"
* class and a large low-risk background class.
*
* MIXTURE DISTRIBUTION
*   z_i = 1 (shooter)     with probability pi = S/N
*   z_i = 0 (non-shooter) with probability 1 - pi
*
* CONDITIONAL INTENSITY (HAWKES POINT PROCESS)
*   C_i(t) = lambda_base_i
*            * exp[ theta_k + mu_i + eps_it
*                   + f_i(t) - g_i(t)
*                   - lambda_base_i * (1 / (S_i(t) + 1)) ]
*            * (1 + A_i(t))^beta_net
*   where:
*     theta_k ~ N(0, 0.02)  network fixed effect
*     mu_i    ~ N(0, 0.01)  individual fixed effect
*     eps_it  ~ N(0, 0.01)  idiosyncratic shock
*     f_i(t)                self-excitation kernel
*     g_i(t)                deterrence kernel
*     S_i(t)                gambler's ruin buffer
*     A_i(t)                mean network shooting rate at t-1
*     beta_net = 0.40       network contagion elasticity
*
* SELF-EXCITATION KERNEL
*   f_i(t) = (1 - omega1)*f_i(t-1) + R1_z*shoot_i(t-1)
*   omega1 = 0.30; R1_z = 0.10 (shooter), 0.05 (non-shooter)
*
* DETERRENCE KERNEL
*   g_i(t) = (1 - omega2)*g_i(t-1) + R2*1{dS_i(t-1) = -1}
*   omega2 = 0.15; R2 = 0.20
*
* GAMBLER'S RUIN BUFFER
*   S_i(t) in {0,...,M=3}; S = 0 is absorbing (incapacitation)
*   p_down = min(phi_k * C/mu_base_s, 1)
*
* SCENARIOS: phi_k in {0.15, 0.25, 0.35, 0.45, 0.55}
* ============================================================

cd "<project directory>"

clear all
set seed 2024

* ============================================================
* FIXED MODEL SCALARS (constant across all years)
* ============================================================
scalar omega1   = 0.30   // self-excitation decay
scalar omega2   = 0.15   // deterrence decay (slower)
scalar R1_s     = 0.10   // shooter excitation jump
scalar R1_ns    = 0.05   // non-shooter excitation jump
scalar R2       = 0.20   // deterrence jump per arrest
scalar beta_net = 0.40   // network contagion elasticity

* Baseline intensities recalculated each year:
*   mu_base_s  = (total_shoot * 0.70) / (S     * 12)
*   mu_base_ns = (total_shoot * 0.30) / ((N-S) * 12)

```

```
local T 12   // time periods
local M  3   // absorbing barrier

* ============================================================
* OUTER LOOP: YEARS 2018-2024
* ============================================================
tempfile allresults
save `allresults', emptyok replace   // cross-year comparison

tempfile allsanity
save `allsanity', emptyok replace    // stacked all-years table

foreach yr of numlist 2018 2019 2020 2021 2022 2023 2024 {

    * Re-seed RNG at the start of each year for reproducibility
    set seed `yr'

    * Year-specific population parameters
    if `yr' == 2018  local N 13677
    if `yr' == 2018  local G  3752
    if `yr' == 2018  local S   659
    if `yr' == 2019  local N 13975
    if `yr' == 2019  local G  3779
    if `yr' == 2019  local S   680
    if `yr' == 2020  local N  8228
    if `yr' == 2020  local G  3468
    if `yr' == 2020  local S   781
    if `yr' == 2021  local N  6032
    if `yr' == 2021  local G  3084
    if `yr' == 2021  local S   645
    if `yr' == 2022  local N  6800
    if `yr' == 2022  local G  3096
    if `yr' == 2022  local S   674
    if `yr' == 2023  local N  7402
    if `yr' == 2023  local G  3371
    if `yr' == 2023  local S   622
    if `yr' == 2024  local N  7932
    if `yr' == 2024  local G  3555
    if `yr' == 2024  local S   726

    local net_size = round(`N' / `G')
    local Nns = `N' - `S'

    * Observed shootings (70% shooter / 30% non-shooter split)
    if `yr' == 2018  local total_shoot 2504
    if `yr' == 2019  local total_shoot 2221
    if `yr' == 2020  local total_shoot 3368
    if `yr' == 2021  local total_shoot 3637
    if `yr' == 2022  local total_shoot 2950
    if `yr' == 2023  local total_shoot 2540
    if `yr' == 2024  local total_shoot 2376

    scalar mu_base_s  = (`total_shoot' * 0.70) / (`S'    * 12)
    scalar mu_base_ns = (`total_shoot' * 0.30) / (`Nns'  * 12)

    display _newline "========================================="
    display "YEAR `yr':  N=`N'  G=`G'  S=`S'  N-S=`Nns'"
    display "========================================="

    * ============================================================
    * SECTION 2: SCENARIO LOOP -- vary deterrence baseline phi_g
    * ============================================================
    tempfile results
    clear
    save `results', emptyok replace

    foreach phi0 in 0.15 0.25 0.35 0.45 0.55 {

        display _newline "  phi_g = `phi0'"
        clear

        * 2A. Cross-sectional setup
```

```
        set obs `N'
        gen id = _n
        gen network = mod(id - 1, `G') + 1

        * 2B. Mixture assignment: exactly S latent shooters
        gen sort_rand = runiform()
        sort sort_rand
        gen shooter = (_n <= `S')
        drop sort_rand
        sort id

        * 2C. Individual-level latent traits (before expand)
        gen mu_i        = rnormal(0, 0.01)
        gen lambda_base = cond(shooter == 1, mu_base_s, mu_base_ns)

        * 2D. Network-level latent traits (before expand)
        preserve
            keep network
            duplicates drop network, force
            sort network
            gen theta_g = rnormal(0, 0.02)
            gen phi_g   = `phi0' + runiform() * 0.005
            tempfile netfe
            save `netfe'
        restore
        merge m:1 network using `netfe', nogenerate

        * 2E. Panel expansion: N*T observations
        expand `T'
        bysort id (network): gen t = _n
        gen eps = rnormal(0, 0.01)

        * 2F. State variable initialization
        gen buf = .
        bysort id (t): replace buf = `M' if _n == 1
        gen buf_traj = .
        gen absorbed = 0
        gen f = 0
        gen g = 0

        * 2G. Outcome and transition variables
        gen inv_buf      = .
        gen C            = .
        gen shoot        = .
        gen p_down       = .
        gen u            = .
        gen dS           = .
        gen net_prev     = 0
        gen net_sum_prev = 0
        gen net_effect   = 0

        * ============================================================
        * SECTION 3: MAIN TIME LOOP  t = 1, ..., T
        * ============================================================
        forvalues tt = 1/`T' {

            * 3A. Gambler's ruin constraint term: inv_buf = 1/(S+1)
            replace inv_buf = 1 / (buf + 1) if t == `tt'

            * 3B. Network contagion term A_i(t)
            if `tt' == 1 {
                replace net_prev     = 0
                replace net_sum_prev = 0
                replace net_effect   = 0 if t == `tt'
            }
            else {
                replace net_prev = 0
                replace net_prev = shoot if t == `=`tt' - 1'
                bysort network: egen net_sum_temp = total(net_prev)
                replace net_sum_prev = net_sum_temp
                drop net_sum_temp
                replace net_effect = net_sum_prev / `net_size' if t == `tt'
```

```
            replace net_effect = 0 if missing(net_effect) & t == `tt'
        }

        * 3C. Hawkes kernel updates f_i(t) and g_i(t)
        if `tt' > 1 {
            bysort id (t): replace f = ///
                (1 - omega1) * f[_n-1] + ///
                cond(shooter == 1, R1_s, R1_ns) * shoot[_n-1] ///
                if t == `tt'

            bysort id (t): replace g = ///
                (1 - omega2) * g[_n-1] + ///
                R2 * cond(dS[_n-1] == -1, 1, 0) ///
                if t == `tt'
        }

        * 3D. Conditional intensity C_i(t)
        if `tt' == 1 {
            replace C = lambda_base ///
                * exp(theta_g + mu_i + eps ///
                      - lambda_base * inv_buf) ///
                * (1 + net_effect) ^ beta_net ///
                if t == `tt'
        }
        else {
            replace C = lambda_base ///
                * exp(theta_g + mu_i + eps + f - g ///
                      - lambda_base * inv_buf) ///
                * (1 + net_effect) ^ beta_net ///
                if t == `tt'
        }
        replace C = 0   if absorbed == 1 & t == `tt'
        replace C = max(C, 0)  if t == `tt'
        replace C = min(C, 10) if t == `tt'

        * 3E. Realized shooting events
        replace shoot = rpoisson(C) if t == `tt'
        replace shoot = 0 if absorbed == 1 & t == `tt'

        * 3F. Gambler's ruin transition
        replace p_down = min(phi_g * (C / mu_base_s), 1) if t == `tt'
        replace p_down = 0 if absorbed == 1 & t == `tt'

        replace u  = runiform() if t == `tt'
        replace dS = cond(u < p_down, -1, 1) if t == `tt'
        replace dS = 0 if absorbed == 1 & t == `tt'

        replace buf_traj = buf if t == `tt'

        * 3G. Buffer update for next period
        if `tt' < `T' {
            bysort id (t): replace buf = ///
                buf[_n-1] + dS[_n-1] if t == `tt' + 1
            replace buf = 0   if buf < 0   & t == `tt' + 1
            replace buf = `M' if buf > `M' & t == `tt' + 1
            bysort id (t): replace absorbed = ///
                cond(buf[_n-1] == 0, 1, absorbed[_n-1]) ///
                if t == `tt' + 1
        }
    }
    * End time loop

    gen scenario_phi0 = `phi0'
    gen year = `yr'

    append using `results'
    save `results', replace
}
* End scenario loop

* ============================================================
* SECTION 4: LOAD AND LABEL COMBINED RESULTS -- YEAR `yr'
```

```
* ============================================================
use `results', clear

gen phi_label = ""
replace phi_label = "Deterrence = 0.15" if abs(scenario_phi0 - 0.15) < 0.001
replace phi_label = "Deterrence = 0.25" if abs(scenario_phi0 - 0.25) < 0.001
replace phi_label = "Deterrence = 0.35" if abs(scenario_phi0 - 0.35) < 0.001
replace phi_label = "Deterrence = 0.45" if abs(scenario_phi0 - 0.45) < 0.001
replace phi_label = "Deterrence = 0.55" if abs(scenario_phi0 - 0.55) < 0.001

* ============================================================
* SECTION 5: FIGURES USED IN THE PAPER -- YEAR `yr'
* (diagnostic figures from the original program are omitted)
* ============================================================

* Yearly trajectory panel (feeds combined_trajectories.pdf)
preserve
collapse (sum) total_shoot=shoot, by(t phi_label)
twoway ///
    (line total_shoot t if phi_label == "Deterrence = 0.15", lcolor(purple) lpattern(solid))      ///
    (line total_shoot t if phi_label == "Deterrence = 0.25", lcolor(red)    lpattern(dash))       ///
    (line total_shoot t if phi_label == "Deterrence = 0.35", lcolor(orange) lpattern(dot))        ///
    (line total_shoot t if phi_label == "Deterrence = 0.45", lcolor(green)  lpattern(shortdash))  ///
    (line total_shoot t if phi_label == "Deterrence = 0.55", lcolor(blue)   lpattern(longdash)),  ///
    legend(order(1 "Deterrence = 0.15" 2 "Deterrence = 0.25" ///
                 3 "Deterrence = 0.35" 4 "Deterrence = 0.45" 5 "Deterrence = 0.55")) ///
    xtitle("Time Period t", size(medsmall)) ytitle("Total Shootings", size(medsmall)) ///
    title("Total Shootings Over Time -- `yr'", size(medsmall))
graph save "fig3_total_shoot_time_`yr'.gph", replace
restore

* The four panels below are produced only at phi = 0.45,
* the calibration anchor used in the paper's yearly
* model-parameter snapshot figure (Section 10).

* Mean gambler's ruin buffer S_i(t) over time
preserve
collapse (mean) mean_buf=buf, by(t shooter phi_label)
foreach p in 0.45 {
    local label = "Deterrence = `p'"
    local pfx = subinstr("`p'", ".", "_", .)
    twoway ///
        (line mean_buf t if phi_label == "`label'" & shooter == 1, ///
            lcolor(red)  lpattern(solid) lwidth(medthick)) ///
        (line mean_buf t if phi_label == "`label'" & shooter == 0, ///
            lcolor(blue) lpattern(dash)  lwidth(medthick)), ///
        legend(order(1 "Shooters" 2 "Non-Shooters")) ///
        xtitle("Time Period t", size(medsmall)) ytitle("Mean Constraint Buffer S (0-3)", size(medsmall)) ///
        title("Gambler's Ruin Buffer -- `label' -- `yr'", size(medsmall)) ///
        yline(0, lcolor(black) lpattern(dot))
    graph save "fig5_buffer_phi_`pfx'_`yr'.gph", replace
}
restore

* Incapacitation (absorption) rate over time
preserve
collapse (mean) pct_absorbed=absorbed, by(t shooter phi_label)
replace pct_absorbed = pct_absorbed * 100
foreach p in 0.45 {
    local label = "Deterrence = `p'"
    local pfx = subinstr("`p'", ".", "_", .)
    twoway ///
        (line pct_absorbed t if phi_label == "`label'" & shooter == 1, ///
            lcolor(red)  lpattern(solid) lwidth(medthick)) ///
        (line pct_absorbed t if phi_label == "`label'" & shooter == 0, ///
            lcolor(blue) lpattern(dash)  lwidth(medthick)), ///
        legend(order(1 "Shooters" 2 "Non-Shooters")) ///
        xtitle("Time Period t", size(medsmall)) ytitle("% Incapacitated", size(medsmall)) ///
        title("Incapacitation Rate Pr(absorbed=1) -- `label' -- `yr'", size(medsmall))
    graph save "fig6_absorption_phi_`pfx'_`yr'.gph", replace
}
restore
```

```
352
353     * Hawkes kernel paths f_i(t) and g_i(t)
354     preserve
355     collapse (mean) mean_f=f (mean) mean_g=g, by(t shooter phi_label)
356     foreach p in 0.45 {
357         local label = "Deterrence = `p'"
358         local pfx = subinstr("`p'", ".", "_", .)
359         twoway ///
360             (line mean_f t if phi_label == "`label'" & shooter == 1, ///
361                 lcolor(red)    lpattern(solid)     lwidth(medthick)) ///
362             (line mean_g t if phi_label == "`label'" & shooter == 1, ///
363                 lcolor(orange) lpattern(dash)      lwidth(medthick)) ///
364             (line mean_f t if phi_label == "`label'" & shooter == 0, ///
365                 lcolor(blue)   lpattern(dot)       lwidth(thin))     ///
366             (line mean_g t if phi_label == "`label'" & shooter == 0, ///
367                 lcolor(green)  lpattern(shortdash) lwidth(thin)),    ///
368             legend(order(1 "f: Shooter Excitation" 2 "g: Shooter Deterrence" ///
369                          3 "f: Non-Shooter Excitation" 4 "g: Non-Shooter Deterrence")) ///
370             xtitle("Time Period t", size(medsmall)) ytitle("Kernel Level", size(medsmall)) ///
371             title("Hawkes Kernels f(t) and g(t) -- `label' -- `yr'", size(medsmall))
372         graph save "fig7_kernels_phi_`pfx'_`yr'.gph", replace
373     }
374     restore
375
376     * Mean conditional intensity C_i(t) by type
377     preserve
378     collapse (mean) mean_C=C, by(t shooter phi_label)
379     foreach p in 0.45 {
380         local label = "Deterrence = `p'"
381         local pfx = subinstr("`p'", ".", "_", .)
382         twoway ///
383             (line mean_C t if phi_label == "`label'" & shooter == 1, ///
384                 lcolor(red)  lpattern(solid) lwidth(medthick)) ///
385             (line mean_C t if phi_label == "`label'" & shooter == 0, ///
386                 lcolor(blue) lpattern(dash)  lwidth(medthick)), ///
387             legend(order(1 "Shooters" 2 "Non-Shooters")) ///
388             xtitle("Time Period t", size(medsmall)) ytitle("Mean C_i(t)", size(medsmall)) ///
389             title("Mean Conditional Intensity -- `label' -- `yr'", size(medsmall))
390         graph save "fig9_intensity_phi_`pfx'_`yr'.gph", replace
391     }
392     restore
393
394     * ============================================================
395     * SECTION 7: ACCUMULATE YEAR-LEVEL TOTALS
396     * ============================================================
397
398     * Block 1: shooter/non-shooter totals by phi (feeds Section 9)
399     preserve
400         collapse (sum) shootings=shoot, by(phi_label shooter)
401         list phi_label shooter shootings, sepby(phi_label)
402         gen year   = `yr'
403         gen actual = `total_shoot'
404         append using `allsanity'
405         save `allsanity', replace
406     restore
407
408     * Block 2: year totals by scenario (feeds Section 8)
409     preserve
410         collapse (sum) sim_shootings=shoot, by(scenario_phi0)
411         gen phi_int = round(scenario_phi0 * 100)   // 15, 25, 35, 45, 55
412         gen year    = `yr'
413         gen actual  = `total_shoot'
414         drop scenario_phi0
415         append using `allresults'
416         save `allresults', replace
417     restore
418
419     display _newline "Year `yr' complete."
420 }
421 * End year loop
422
423 * ============================================================
```

```
* COMBINED TRAJECTORIES FIGURE (combined_trajectories.pdf)
* Combines the seven yearly trajectory panels into the single
* figure shown in the paper.
* ============================================================
graph combine ///
    "fig3_total_shoot_time_2018.gph" "fig3_total_shoot_time_2019.gph" ///
    "fig3_total_shoot_time_2020.gph" "fig3_total_shoot_time_2021.gph" ///
    "fig3_total_shoot_time_2022.gph" "fig3_total_shoot_time_2023.gph" ///
    "fig3_total_shoot_time_2024.gph" ///
    , cols(2) l1title("Total Shootings") b1title("Month") ///
    title("Monthly Shooting Trajectories by Deterrence Scenario", size(medsmall)) ///
    iscale(0.45) xsize(12) ysize(14) scheme(sj)
graph export "combined_trajectories.pdf", replace

* ============================================================
* SECTION 8: CROSS-YEAR COMPARISON -- SIMULATED vs ACTUAL
* ============================================================
use `allresults', clear

reshape wide sim_shootings, i(year actual) j(phi_int)

rename sim_shootings15 phi015
rename sim_shootings25 phi025
rename sim_shootings35 phi035
rename sim_shootings45 phi045
rename sim_shootings55 phi055

label var year    "Year"
label var actual  "Actual Shootings"
label var phi015  "Simulated (phi=0.15)"
label var phi025  "Simulated (phi=0.25)"
label var phi035  "Simulated (phi=0.35)"
label var phi045  "Simulated (phi=0.45)"
label var phi055  "Simulated (phi=0.55)"

sort year

* Table: actual vs simulated shootings by year and scenario
mkmat actual phi015 phi025 phi035 phi045 phi055, matrix(comp) rownames(year)

esttab matrix(comp, fmt(%9.0f)) using "cross_year_comparison.rtf", ///
    replace title("Simulated vs Actual Shootings by Year and Deterrence Scenario") ///
    mlabels("Actual" "phi=0.15" "phi=0.25" "phi=0.35" "phi=0.45" "phi=0.55") ///
    collabels(none) noobs

twoway ///
    (line actual year, lcolor(black)  lwidth(thick)    lpattern(solid))     ///
    (line phi015 year, lcolor(purple) lwidth(medium)   lpattern(dash))      ///
    (line phi025 year, lcolor(red)    lwidth(medium)   lpattern(shortdash)) ///
    (line phi035 year, lcolor(orange) lwidth(medium)   lpattern(dot))       ///
    (line phi045 year, lcolor(green)  lwidth(medium)   lpattern(longdash))  ///
    (line phi055 year, lcolor(blue)   lwidth(medium)   lpattern(dash_dot))  ///
    , ///
    xlabel(2018/2024, angle(45)) ///
    xtitle("Year", size(medsmall)) ytitle("Total Shootings", size(medsmall)) ///
    title("Actual vs Simulated Shootings by Deterrence Scenario") ///
    legend(order(1 "Actual" 2 "Deterrence=0.15" 3 "Deterrence=0.25" ///
                 4 "Deterrence=0.35" 5 "Deterrence=0.45" 6 "Deterrence=0.55") ///
        ring(0) pos(1) cols(1)) ///
    scheme(sj)
graph export "fig_simulated_vs_actual_shootings.pdf", replace

* ============================================================
* SECTION 9: STACKED ALL-YEARS TABLE -- LaTeX output
* Produces table_shootings_all_years.tex with one panel per
* year: five phi-scenario rows plus an Actual row.
* Format: Year | Scenario | Non-Shooters | Shooters | Total | Actual
* (the 2021 panel appears in the paper)
* ============================================================
use `allsanity', clear

* Reshape long -> wide: shooter 0 and 1 become separate columns
```

```
reshape wide shootings, i(year phi_label actual) j(shooter)
rename shootings0 ns
rename shootings1 sh
gen total = ns + sh

* Ordering variable so phi scenarios sort 0.15 < 0.25 < ... < 0.55
gen phi_order = .
replace phi_order = 1 if strpos(phi_label, "0.15")
replace phi_order = 2 if strpos(phi_label, "0.25")
replace phi_order = 3 if strpos(phi_label, "0.35")
replace phi_order = 4 if strpos(phi_label, "0.45")
replace phi_order = 5 if strpos(phi_label, "0.55")
sort year phi_order

local ds = char(36)   // dollar sign, avoids macro expansion in file write

file open fh using "table_shootings_all_years.tex", write replace

file write fh "\begin{table}[h]" _n
file write fh "\centering" _n
file write fh "\caption{Simulated Shootings by Latent Type, Deterrence Scenario, and Year}" _n
file write fh "\label{tab:shootings_all_years}" _n
file write fh "\begin{tabular}{llrrrr}" _n
file write fh "\toprule" _n
file write fh "Year & Scenario & Non-Shooters & Shooters & Total & Actual \\" _n
file write fh "\midrule" _n

local phi_vals "0.15 0.25 0.35 0.45 0.55"

foreach yr of numlist 2018/2024 {

    quietly sum actual if year == `yr'
    local act : display %12.0fc r(mean)
    local act = strtrim("`act'")

    local first 1
    forvalues p = 1/5 {

        quietly sum ns    if year == `yr' & phi_order == `p'
        local nsv : display %9.0fc r(mean)
        local nsv = strtrim("`nsv'")

        quietly sum sh    if year == `yr' & phi_order == `p'
        local shv : display %9.0fc r(mean)
        local shv = strtrim("`shv'")

        quietly sum total if year == `yr' & phi_order == `p'
        local tot : display %9.0fc r(mean)
        local tot = strtrim("`tot'")

        local phi : word `p' of `phi_vals'
        local plab "`ds'\phi_0=`phi'`ds'"

        if `first' {
            file write fh "`yr' & `plab' & `nsv' & `shv' & `tot' & \\" _n
            local first 0
        }
        else {
            file write fh "     & `plab' & `nsv' & `shv' & `tot' & \\" _n
        }
    }

    file write fh "     & \textit{Actual} & & & & \textit{`act'} \\" _n

    if `yr' < 2024 {
        file write fh "\midrule" _n
    }
}

file write fh "\bottomrule" _n
file write fh "\end{tabular}" _n
file write fh "\end{table}" _n
```

```
file close fh

display _newline "LaTeX table saved: table_shootings_all_years.tex"

* ============================================================
* SECTION 10: YEAR-BY-YEAR SNAPSHOT -- Key Model Parameters
* (phi = 0.45; the 2021 snapshot appears in the paper)
* ============================================================
foreach yr of numlist 2018 2019 2020 2021 2022 2023 2024 {

    graph combine                                                 ///
        "fig5_buffer_phi_0_45_`yr'.gph"                           ///
        "fig6_absorption_phi_0_45_`yr'.gph"                       ///
        "fig7_kernels_phi_0_45_`yr'.gph"                          ///
        "fig9_intensity_phi_0_45_`yr'.gph"                        ///
        , cols(2) rows(2)                                          ///
        title("Key Model Parameters `yr' -- phi = 0.45",           ///
              size(medlarge) color(black))                         ///
        note("All panels at deterrence level phi = 0.45 (calibration anchor).", ///
             size(vsmall))                                         ///
        xsize(14) ysize(11) iscale(0.40)                          ///
        scheme(sj)

    graph save   "fig_yearly_snapshot_`yr'.gph", replace
    graph export "fig_yearly_snapshot_`yr'.pdf", replace
    display _newline "Year `yr' snapshot saved."
}
```