

认罪认罚量刑从宽实效研究

——基于故意伤害罪轻罪的数据解读

王芳 甘叠 刘念

摘要：我国认罪认罚从宽制度设计具有明显的公正优先价值导向。通过构建非线性量刑模型分析177161份故意伤害罪致人轻伤的初审判决书发现，司法实践中表现出明显的“从宽”克制倾向；认罪认罚对轻罪量刑的作用倾向于在“底线”游走，接近量刑情节“坦白”；传统量刑情节仍在习惯性适用，认罪认罚发挥实效的空间不足。在法定主义背景下推动更高层次上公正与效率的统一，可以考虑，一方面建立认罪认罚分级分类制度，适当加大对轻罪的从宽幅度；另一方面扩大认罪认罚从宽幅度的法定空间，在立法上赋予认罪认罚“减轻”情节的地位，破解量刑折扣零和博弈的困局。

关键词：认罪认罚；轻罪；量刑从宽实效；递推最小二乘算法(RLS)

DOI：10.19836/j.cnki.37-1100/c.2022.03.006

认罪认罚从宽制度是我国刑事司法制度的创新性尝试。从设计初衷来看，该制度旨在通过兑现“从宽”的制度承诺激励被告人自愿认罪认罚，推动程序上的繁简分流和司法效率的提升。“从宽”承诺的兑现，在一定程度上决定了该制度的有效性和可持续性；其背后所折射出的司法公正与效率的价值冲突、法检之间刑罚权的冲突与协调^①等问题，也是刑事一体化司法制度改革的痛点难点。实体层面的量刑从宽是认罪认罚从宽制度的核心。围绕认罪认罚对量刑的实效这一主题，我们调研访谈了最高人民法院相关部门，以及山东省J市、Y市、W市等三地市基层法院的刑事审判法官。访谈发现，在司法实践中，不同的案件、不同的办案人员对从宽幅度掌握不尽一致，特别是对于认罪认罚与其他从宽情节同时适用时如何掌握从宽幅度，不少办案人员存在困惑。为此，我们聚焦认罪认罚对量刑的从宽效力问题，构建高度模拟司法实践的非线性量刑模型，利用递推最小二乘算法(RLS)研究认罪认罚对有期徒刑量刑的实际作用，为认罪认罚从宽制度进一步完善提供参考。

一、相关概念的厘清

认罪认罚的“从宽”作用主要表现在实体层面的量刑从宽和程序层面的从宽处理两个方面。关于何为“从宽”，有实体性与程序性、监禁刑与非监禁刑、从宽处理与从宽处罚等诸多视角，学界已有广泛探讨^②。本文重点关注的是实体法层面的量刑从宽问题，为此首先需要梳理清楚若干相关概念。

基金项目：山东大学人文社会科学创新团队资助项目“全面依法治国战略实施中的数据运用与数据治理”。

作者简介：王芳，山东大学数据科学研究院(山东系统与计算法学研究中心)教授，法学博士(济南250100；wangfang226@sdu.edu.cn)；甘叠，中国科学院数学与系统科学研究院博士研究生(北京100190；gandie@amss.ac.cn)；刘念，中国科学院数学与系统科学研究院博士研究生(北京100190；liunian@amss.ac.cn)。

致谢：感谢中国科学院院士郭雷教授对本研究模型与算法深入全面的指导，感谢中国司法大数据研究院为研究提供的数据支持。

^① 龙宗智：《完善认罪认罚从宽制度的关键是控辩平衡》，《环球法律评论》2020年第2期。陈卫东：《认罪认罚从宽制度的理论问题再探讨》，《环球法律评论》2020年第2期。陈明辉：《认罪认罚从宽制度中法检权力的冲突与协调》，《法学》2021年第11期。韩轶：《认罪认罚案件中的控审冲突及其调和》，《法商研究》2021年第2期。

^② 陈瑞华：《“认罪认罚从宽”改革的理论反思——基于刑事速裁运行经验的考察》，《当代法学》2016年第4期。赵恒：《论从宽的正当性基础》，《政治与法律》2017年第11期。周新：《论从宽的幅度》，《法学杂志》2018年第1期。

(一)认罪认罚实体从宽的考量要素

我们关注认罪认罚实体从宽的考量要素基于两个目的:一是需要尽可能全面地梳理可能影响认罪认罚量刑从宽的法定要素,为量化分析模型中变量的设定奠定基础。二是在此基础上厘清认罪认罚的独立价值与实体效力。

《刑事诉讼法》第15条规定:“犯罪嫌疑人、被告人自愿如实供述自己的罪行,承认指控的犯罪事实,愿意接受处罚的,可以依法从宽处理。”从认罪认罚从宽制度的构成要件来看,认罪、认罚是从宽处理的必要条件,但并非充分条件。其“必要但非充分性”体现在两个方面:一是,“可以”意味着从宽是可选择的,而非“必须”;二是,是否从宽以及从宽多少,需要结合社会危害性、诉讼阶段等因素具体来把握。因此,被告人认罪认罚在量刑上从宽幅度的考量要素,至少应当包括认罪、认罚以及从宽幅度的把握三个部分。其中,“认罪”是指犯罪嫌疑人、被告人自愿如实供述自己的罪行,对指控的犯罪事实没有异议。“认罚”是指犯罪嫌疑人、被告人真诚悔罪,愿意接受处罚,考察的重点是犯罪嫌疑人、被告人的悔罪态度和悔罪表现,应当结合退赃退赔、赔偿损失、赔礼道歉、财产刑执行情况等因素考量^①。从定义来看,“认罪”与自首、坦白、当庭自愿认罪等情节存在交集,“认罚”则与退赃退赔、赔偿损失、赔礼道歉、缴纳财产刑保证金等悔罪表现存在交集。从整体来看,又与刑事和解、取得被害人谅解等密切相关。

“从宽幅度的把握”主要体现在《关于适用认罪认罚从宽制度的指导意见》(以下简称《指导意见》)第9条。从第9条的规定来看,“从宽幅度的把握”除考虑上述与“认罪”“认罚”相关的体现悔罪态度、悔罪表现、对查明案件事实的价值和意义等要素外,还包括:(1)社会危害性因素,如初犯、偶犯、过失犯或者未成年犯等社会危害不大的,从宽幅度可以大一些;罪行较重、人身危险性较大的以及累犯、再犯,从宽幅度应当从严把握。(2)罪行严重程度。如对于罪行较轻的,从宽幅度可以大一些。(3)诉讼阶段。如山东省规定,犯罪嫌疑人、被告人自侦查阶段始终认罪认罚的,从宽幅度比自审查起诉阶段认罪认罚的可以增加5%以下^②。但我们发现,除山东省外,其他地区的判决书一般不对诉讼阶段作明确说明,较为普遍的做法是区分“认罪认罚”与“当庭自愿认罪”。

总体来看,除罪行严重程度与诉讼阶段的规定外,认罪认罚从宽所考量要素的最终落脚点似乎并没有明显超出传统量刑影响因素的范围。这也是理论界及司法实务界困惑的主要原因:既然认罪认罚的构成要素与已有量刑制度存在较大交集,认罪认罚对量刑的效力应当如何理解?

(二)认罪认罚从宽的实体效力定位

认罪认罚对量刑的从宽效力的正当性来自《刑事诉讼法》。因此,如何理解认罪认罚从宽这一刑事诉讼基本制度在实体法中的效力定位,即认罪认罚是否可以作为独立量刑情节,是把握量刑从宽问题的基础。有学者认为,《刑事诉讼法》第15条可以作为认罪认罚从宽处罚的依据,认罪认罚应当是自首、坦白之外的一个新的独立的量刑情节^③。相反观点则认为,不能根据《刑事诉讼法》的规定直接对被告人减轻处罚,《刑法》必须及时与认罪认罚从宽制度相衔接,需要在立法上将认罪认罚增加到《刑法》第61条关于量刑原则的一般规定之中^④。对此,2021年7月发布的《关于常见犯罪的量刑指导意见(试行)》(以下简称新版《量刑指导意见》)起到了一定的说明作用。在新版《量刑指导意见》中,认罪认罚作为“常见量刑情节”之十四,“当庭自愿认罪”作为“常见量刑情节”之八,与自首、坦白、赔偿谅解、刑事和解等传统量刑情节并列出现,并且就其对基准刑的减少作用给出了具体幅度。这在一定程度上确认了认罪认罚作为量刑情节的地位,肯定了认罪认罚对于量刑的实体效力。

① 参见《关于适用认罪认罚从宽制度的指导意见》第6条、第7条。

② 参见山东省《关于适用认罪认罚从宽制度办理刑事案件的实施细则(试行)》(鲁检会〔2019〕10号)第40条。

③ 樊崇义:《认罪认罚从宽与自首坦白》,《人民法治》2019年第1期。

④ 周光权:《论刑法与认罪认罚从宽制度的衔接》,《清华法学》2019年第3期。

但司法实践中的困惑仍然存在。前述“认罪认罚实体从宽的考量要素”并没有明显超出传统量刑情节的考查要点,如果仅在这一层面认识认罪认罚从宽制度,则难免陷入“重复评价”的误区。应该如何理解和定位认罪认罚从宽的实体效力?我们认为,认罪认罚从宽制度的价值不能脱离其诉讼法本源。认罪认罚从宽制度作为一项刑事诉讼制度,其目的在于通过兑现“从宽”的制度承诺,提高诉讼效率,在更高层面上实现公正与效率的统一^①。因此,不能脱离诉讼效率价值去理解认罪认罚从宽制度。在这一意义上,认罪认罚从宽制度的独立价值体现在其对实体法与程序法的勾连效力中,换言之,是程序价值的实体呈现。这种通过实体法兑现程序价值的制度设计,体现了刑事一体化的改革思想,与刑事和解制度的设计有着相似之处(自2013年起,若干版《量刑指导意见》都将刑事和解列为“常见量刑情节”,对量刑发挥着实际影响)。因此有学者认为,认罪认罚从宽制度“标志着我国法律继接受被害方与被告方的私力合作模式之后,再一次确立了一种在协商和妥协基础上的公力合作模式”^②。

当然,对于认罪认罚到底是实体法层面的量刑情节还是诉讼法层面的量刑情节、如何认识避免重复评价等问题,仍有不同观点。但基于本文研究的实证立场,我们站在司法实务的角度将认罪认罚视为具有独立作用的量刑情节。这一定位构成了本文研究的前提。基于此,我们可以将认罪认罚设定为独立特征(变量),在量刑模型中进行参数估计。

二、样本数据与相关特征提取

(一)样本数据

本文研究数据来源于中国裁判文书网公开的刑事司法判决书,判决时间为2011年1月至2021年6月。考虑到故意伤害罪为常见犯罪,认罪认罚从宽适用占比较高,且主要分布在致人轻伤的判决中,我们选择故意伤害致人轻伤的初审判决书作为样本,共计177161份,其中认罪认罚的判决书6574份。样本年度分布见表1。

表1 样本年度分布

时间	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
数量	490	1333	5669	24510	26940	26642	26869	25242	23038	15511	917
总计	177161										

(二)特征(要素)提取

我们考虑了前述认罪认罚从宽制度的构成要素,并根据故意伤害罪犯罪构成、量刑制度等知识以及相关实证研究文献,在判决书中提取了相关特征61个。然后根据具体特征的分布情况,删除了分布过分稀疏的特征(频次少于200),形成涉及实体性与程序性情节的7个法定特征、11个酌定特征、4个其他特征。法定特征包括:轻微伤人数、轻伤人数、主犯、从犯、累犯、自首、坦白等。酌定特征包括:认罪认罚、当庭自愿认罪、和解、谅解、积极赔偿、抢救被害人和挽回损失、前科、初犯偶犯、互殴、被害人过错、民间矛盾等。其他特征包括:缓刑、附带民事诉讼、地域(省份)、是否有辩护人等。其中,我们关注的特征为:认罪认罚、当庭自愿认罪、自首、坦白、和解、谅解、积极赔偿。

^① 2016年8月,最高人民法院院长周强向全国人大常委会就试点授权决定草案作说明时指出这一必要性。

^② 陈瑞华:《刑事诉讼的公力合作模式——量刑协商制度在中国的兴起》,《法学论坛》2019年第4期。

三、模型构建与研究思路方法

关于量刑问题的实证分析已经形成了一套相对完备的研究方法。如白建军建立了反映刑量指标权重关系的宣告刑量刑模型^①;王剑波等利用 SW 检验、KW 检验对刑期的平均值或比例初步进行描述性分析并建立了回归模型解读量刑地区差异^②;彭雅丽依托计量分析模型评估了量刑指导意见运行效果^③;吴雨豪通过“准随机的试验”最大化匹配对照组与处理组量刑情节,研究认罪认罚的从宽裁量模式,给出了“混合模式”的结论^④。考虑到刑事司法审判具有空间上的一致性和时间上的关联性,我们尝试在对裁判文书按判决时间排序的基础上,利用递推最小二乘算法(Recursive Least Squares, 缩写为 RLS)估计认罪认罚及相关制度对量刑的实际影响和可能变化。其依据在于,首先,《刑法》《量刑指导意见》在全国范围内具有一致的法定指导作用,各地在量刑上虽然可能存在地域差别,但差别不能超出法定范畴。其次,类案推送系统在全国范围的适用,以及最高人民法院指导案例与典型案例的发布,对全国司法审判均有显著的引导作用。因此,按判决时间进行排序并进行相关分析是合理的。同时,RLS 算法不需独立同分布先验假设,更加符合司法判决这一类社会数据的基本性质;另外,它还可以观察主要特征随时间的变化趋势。

(一)构建模型

根据《刑法》《量刑指导意见》相关规定,量刑过程如下:在定性分析的基础上,结合定量分析,依次确定量刑起点、基准刑和宣告刑。具体而言,量刑起点应根据基本犯罪构成事实在相应的法定刑幅度内确定;再根据其他影响犯罪构成的犯罪数额、犯罪次数、犯罪后果等犯罪事实,在量刑起点的基础上增加刑罚量确定基准刑;然后根据量刑情节调节基准刑,并综合考虑全案情况,依法确定宣告刑。

就故意伤害罪而言,其有期徒刑的量刑起点区间分别为 2 年以下(致 1 人轻伤)、3-5 年(致 1 人重伤)、10-13 年(致人死亡或手段特别残忍)^⑤。我们调研访谈了山东省三地市以及上海市某区基层法院刑事审判庭,发现法官一般在相应量刑幅度内的前三分之一处确定量刑起点。因此,在构建模型时,我们将量刑起点选定在相应量刑幅度的前 1/3。具体的量刑逻辑过程如图 1 所示:

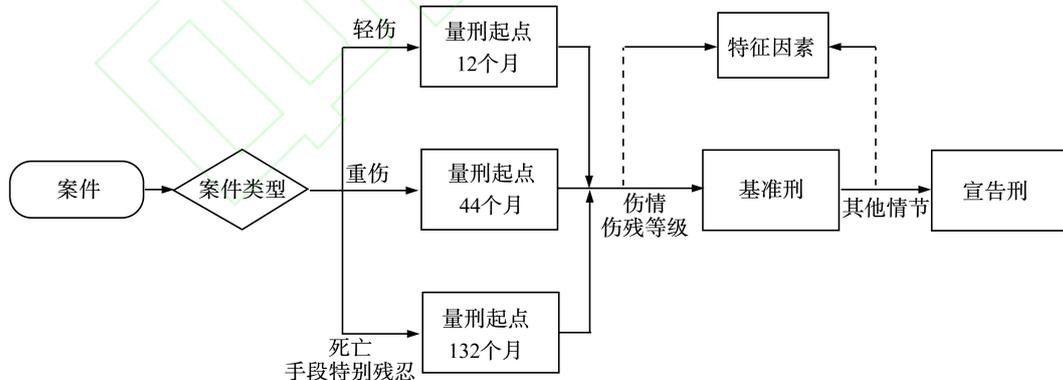


图 1 量刑逻辑过程图

① 白建军:《量刑基准实证研究》,《法学研究》2008 年第 1 期。
 ② 王剑波:《我国受贿罪量刑地区差异问题实证研究》,《中国法学》2016 年第 4 期。
 ③ 彭雅丽:《量刑指导意见的司法实践与重构——以盗窃罪为切入点》,《法学研究》2021 年第 4 期。
 ④ 吴雨豪:《认罪认罚“从宽”裁量模式实证研究——基于部分城市醉酒型危险驾驶罪的定量研究》,《中外法学》2020 年第 5 期。
 ⑤ 参见《关于常见犯罪的量刑指导意见(试行)》(法发〔2021〕21 号、法发〔2017〕7 号)之四“常见犯罪的量刑”中的“(七)故意伤害罪”。

图 1 展示的是故意伤害罪基本的量刑逻辑过程,包括了致人重伤、致人轻伤、致人死亡或手段特别残忍等三类量刑起点类型。在具体的分析中,我们基于量刑逻辑过程,根据样本特点选择“致人轻伤”类型建立以下非线性回归模型^①:

$$y_t = \underbrace{\left(\frac{\text{量刑起点}}{a} + bx_t^1 + cx_t^2 + dx_t^3 \right)}_{\text{基准刑}} \underbrace{(1 + \phi_t^T \vartheta)}_{\text{量刑情节因素的影响}} + e + w_t$$

我们根据判决时间将所有样本按照先后顺序进行排序。模型中 y_t 为第 t 个案件的有期徒刑刑期(单位:月); a 表示量刑起点(单位:月); x_t^1 、 x_t^2 、 x_t^3 分别表示第 t 个案件中额外的重伤、轻伤、轻微伤人数; b 、 c 、 d 分别代表每多造成一个重伤、轻伤、轻微伤所增加的刑期; ϕ_t 表示所选取的量刑特征构成的回归向量; ϑ 表示量刑特征所对应的百分比向量; e 是偏置项,代表其他可能未考虑到的量刑因素; w_t 是可能存在的噪声。 $(\cdot)^T$ 表示向量转置符号。

为了对上述非线性量刑模型进行计算,我们需要对该模型进行等价转换。引入新的回归向量 φ_t 如下:

$$\varphi_t = (1, x_t^1, x_t^2, x_t^3, a \phi_t^T, x_t^1 \phi_t^T, x_t^2 \phi_t^T, x_t^3 \phi_t^T)^T$$

相应的,定义如下未知参数向量:

$$\theta = (e, b, c, d, \vartheta^T, b \vartheta^T, c \vartheta^T, d \vartheta^T)^T$$

则上述非线性量刑模型可以转换为如下形式上的“线性”参数模型:

$$y_t = \varphi_t^T \theta + a + w_t$$

(二)算法及估计精度的可靠性检验

我们希望基于案件文本特征和刑期等数据得到的 $\{\varphi_t, y_t\}$, 利用递推最小二乘算法(RLS)来估计未知参数 $\{e, b, c, d, \vartheta\}$, 并观察主要特征随时间的变化趋势。

使用 RLS 算法,我们不需要假定噪声满足传统的独立同分布假设,但需要假定噪声是鞅差序列,且具有相同的方差。理论上可以证明,只要验证信息矩阵的最小特征值趋于无穷,即 $\lambda_{\min} \left(\sum_{k=1}^t \varphi_k \varphi_k^T \right) \xrightarrow{t \rightarrow \infty} \infty$, 则 RLS 估计值可几乎处处收敛到真值^②。但是对于有限数据而言,一般很难验证上述条件,为此我们需要建立有限数据下的理论结果和可靠性验证方法。

1. 利用 RLS 估计误差的概率不等式检验有限数据下估计的可靠性^③

引入概率不等式如下: $Pr(|\hat{\theta}_t^{(i)} - \theta^{(i)}| < \delta_t^{(i)}) \geq 1 - \gamma^2$

其中 $Pr(\cdot)$ 表示概率, $\hat{\theta}_t^{(i)}$ 表示第 t 个案件对应第 i 个分量的估计值, $\theta^{(i)}$ 为真实参数 θ 的第 i 个分量,误差上界 $\delta_t^{(i)} \triangleq \frac{\sigma_w \sqrt{P_t^{(i)}}}{\gamma} + |e^{(i)} P_t P_0^{-1} (\theta - \theta_0)|$, P_0 和 θ_0 是 RLS 的初值, $P_t^{(i)}$ 表示矩阵 $P_t = (P_0^{-1} + \sum_{k=1}^t \varphi_k \varphi_k^T)^{-1}$ 对角线上的第 i 个元素。为简单起见,本文中取常数 $\gamma = \sqrt{0.1}$, 则估计值误差小于 $\delta_t^{(i)}$ 的概率不小于 0.9。如果通过计算能够验证误差上界小于估计值绝对值,则可以得到一个具有很高可靠性的估计值范围。

估计值误差上界 $\delta_t^{(i)}$ 表示估计值与真实值差的上界,因此我们需要综合观察估计值与估计值误差上界,才能确定量刑特征的实际作用。换言之,量刑特征实际作用的比重应当位于区间(估计值 - $\delta_t^{(i)}$, 估计值 + $\delta_t^{(i)}$)内。利用下面表 2 提供的具体值可以对该区间进行计算。

① 根据《刑法》与《量刑指导意见》相关规定,宣告刑应受法定刑边界限制,因此量刑模型应为饱和非线性模型。但基于本文研究数据的特点(轻伤案件、多数样本有期徒刑刑期不超过饱和函数上界),因此可以暂时不考虑饱和和非线性的影响。

② Lai T. L., Robbins H., Wei C. Z., “Strong Consistency of Least Squares Estimates in Multiple Regression II”, *Journal of Multivariate Analysis*, 1979, 9(3), pp.343-361.

③ 郭雷:《时变随机系统:稳定性与自适应理论》,北京:科学出版社 2022 年,第 155 页。

2.利用数值实验检验估计值的可靠性

我们任意取定实验值 $\bar{\theta}$, 利用提取出来的案件文本特征得到的 φ_i , 以及随机产生的噪声, 生成案件对应的刑期 \bar{y}_i , 再利用数据 $\{\varphi_i, \bar{y}_i\}$ 得到 $\bar{\theta}$ 的 RLS 估计值。

数值试验发现, 估计值与实验值吻合(附录数值实验结果), 说明用本文数据得到的估计值具有很强的可靠性。

四、认罪认罚对轻罪量刑的影响分析

(一)相对其他从宽量刑情节, 认罪认罚对有期徒刑量刑的影响较弱

1.观察量刑特征估计值及误差界

根据以上模型与 RLS 算法, 可以得到全国轻伤模型的估计值与估计值误差上界(参见表 2)。

表 2 量刑特征估计值及误差界(全样本)

特 征	估计值	估计值误差上界 $\delta_i^{(t)}$
偏置项	1.5951***	0.1270
轻伤人数	5.0612***	0.6291
轻微伤人数	1.1782***	0.5039
自首	-0.1132***	0.0064
坦白	-0.0355***	0.0062
和解	-0.0247***	0.0062
谅解	-0.1899***	0.0073
积极赔偿	-0.0768***	0.0086
从犯	-0.1849***	0.0509
缓刑哑变量	0.0029	0.0071
当庭自愿认罪	0.0092***	0.0060
认罪认罚	-0.0289***	0.0139
抢救被害人和挽回损失	0.0964***	0.0597
初犯、偶犯	0.0024	0.0140
是否有辩护人	-0.0167***	0.0102
被害人过错	-0.0292***	0.0097
累犯	0.2019***	0.0138
主犯	0.0480***	0.0424
前科	0.0532***	0.0183
互殴	-0.0194***	0.0115
附带民事诉讼	0.0954***	0.0123
民间矛盾	-0.0247***	0.0095

注:1.我们也对变量进行了 p 值检验, *** 表示 $p < 0.01$, 没有标注表示 $p > 0.1$ 。

2.特征估计值含义:偏置项以及轻伤人数、轻微伤人数的估计值指的是刑期(单位:月);其他量刑情节的估计值对应的是百分比。

3.偏置项估计值较低(1.5951个月),这说明模型能够较好地逼近实际的量刑过程,特征选择也能够反映实际审判过程中的几乎所有主要量刑影响因素。

由表 2 可以看出,“认罪认罚”在全国致人轻伤数据中的估计值为 -2.89% ,估计值误差上界为 1.39% ,说明“认罪认罚”在全国致人轻伤数据中的实际影响以 90% 的概率落在区间 $(-4.28\%, -1.5\%)$,即对量刑从宽的作用为“减轻基准刑”的 1.5% 至 4.28% 。相应地,当庭自愿认罪的实际影响区间为 $(0.32\%, 1.52\%)$,自首为 $(-11.96\%, -10.68\%)$,坦白为 $(-4.17\%, -2.93\%)$,谅解为 $(-19.72\%, -18.26\%)$,积极赔偿为 $(-8.54\%, -6.82\%)$ 。

由于参数估计是多因素共同作用的结果,某一参数的估计结果实际上是综合效应的局部体现。也就是说,认罪认罚特征参数估计值,可以解释为认罪认罚在各种量刑情节的综合作用中发挥的具体效应。因此,虽然我们不能仅凭参数的估计值就给定某一情节的作用,但我们可以看出某一情节的相对位置,即某一情节相对于其他情节的影响力大小;进而在对各情节相对位置的观察中发现我们关注的认罪认罚的实际作用。

基于此,对比上述特征实际影响的区间可以看出,认罪认罚对有期徒刑的量刑实际影响与坦白情节较为接近,但远低于自首、谅解、积极赔偿等从宽情节;当庭自愿认罪对量刑的影响微弱,作用可以忽略。

2. 观察排除缓刑影响后的量刑特征估计值与误差界

考虑到缓刑的适用是认罪认罚从宽的一个重要侧面,为观察缓刑是否影响了认罪认罚对有期徒刑从宽幅度的作用,我们进一步排除了缓刑的判决,得到样本 73822 份,其中认罪认罚 2732 份。分析结果如表 3 所示。

表 3 量刑特征估计值及误差界(排除缓刑的数据)

特征	估计值	估计值误差上界 $\delta_i^{(t)}$
偏置项	2.4564***	0.1900
轻伤人数	5.0704***	0.8253
轻微伤人数	1.0720***	0.6783
自首	-0.1352***	0.0106
坦白	-0.0422***	0.0100
和解	-0.0540***	0.0140
谅解	-0.3342***	0.0126
积极赔偿	-0.0741***	0.0134
从犯	-0.2156***	0.0682
缓刑哑变量	/	/
当庭自愿认罪	0.00460	0.0104
认罪认罚	-0.0396***	0.0240
抢救被害人和挽回损失	0.1098***	0.0841
初犯、偶犯	0.0005	0.0140
是否有辩护人	-0.0418***	0.0102
被害人过错	-0.0401***	0.0097
累犯	0.2316***	0.0138

续表3

特征	估计值	估计值误差上界 $\delta_i^{(t)}$
主犯	0.0594***	0.0424
前科	0.0623***	0.0183
互殴	-0.0374***	0.0115
附带民事诉讼	0.1210***	0.0123
民间矛盾	-0.0307***	0.0095

由表3可以看出,排除缓刑判决后,认罪认罚的作用区间为(-6.3%, -1.5%),即对基准刑的减轻作用为1.5%到6.3%。相较全样本,其实际作用的上界略有提高。但总体来看,仍与坦白情节的实际作用区间比较接近。

我们进一步考察了缓刑适用率,发现样本中认罪认罚案件的缓刑适用率为58.43%;非认罪认罚案件的缓刑适用率为58.31%。二者的缓刑适用率并无明显差异。

(二)认罪认罚从宽制度实施后,传统从宽量刑情节对量刑的影响并未发生明显变化

认罪认罚从宽制度于2018年底正式实施。为观察认罪认罚从宽制度的实施对传统量刑情节的影响,我们进一步给出了所关注的相关特征随时间的变化趋势图(参见图2)。

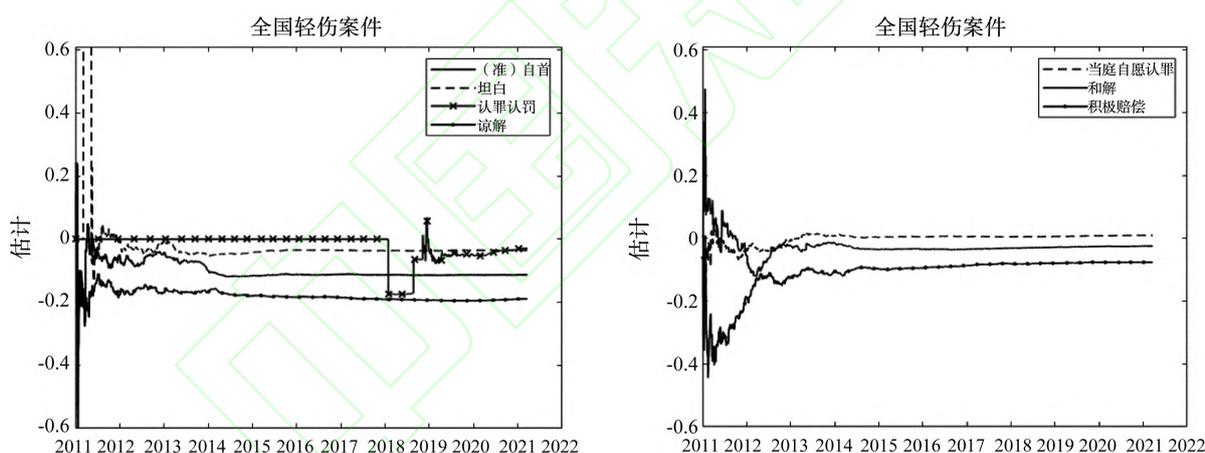


图2 全国致人轻伤案件部分关注变量变化趋势(2011-2021)

注:基于RLS估计算法,前端数据数量小且对估计值的影响较大,导致估计值波动激烈,不具参考意义。随着数据量积累,估计值趋于平稳,此时估计值具有参考意义。

由图2可以看出,排除前端数据干扰后,自首、和解、谅解、积极赔偿等特征在2013年后表现基本稳定。2019年及以后,除认罪认罚和当庭自愿认罪外,自首、坦白、和解、谅解、积极赔偿等传统从宽量刑情节表现非常稳定,未出现异常波动。可见,认罪认罚并未对其他传统从宽量刑情节的实际效力产生明显影响。

五、公正优先的“从宽”克制倾向解读

(一)我国认罪认罚从宽制度改革的价值导向

关于认罪认罚从宽制度的价值,从诉讼角度学界一般认为,认罪认罚从宽制度改革的目的在于破解当前“案多人少”的困局,提高司法效率,“公正为本,效率优先”应当是认罪认罚从宽制度改革的核

心价值取向^①。也有学者从实体法的角度,认为认罪认罚从宽制度不应仅考虑司法效率价值,还应重视“实体性地给予被告人一些优待,使其能享受到一些实体上的利益——即处罚从宽,使案件处理结论获得被告人的实际认同,减少社会对立面”^②。事实上,司法效率与实体从宽的兑现互为表里,相互成全。兑现从宽承诺,一方面能够激励被告人自愿认罪认罚,推动案件繁简分流从而提高司法效率;另一方面也能够消除诉讼对抗,提高息讼率,达到良好的法律效果与社会效果。但是上述价值定位不可避免地公正的价值目标发生纠缠。因此,在认罪认罚从宽制度的立法讨论过程中,如何避免花钱买刑、防止权钱交易、防止量刑畸轻、消除民众担忧,成为争议的焦点和改革者重点关注的问题^③。

2016年8月,最高人民法院院长周强向全国人大常委会就试点授权决定草案作说明时强调:“要更好发挥认罪认罚从宽制度的功能作用,在更高层次上实现公正与效率相统一。”围绕如何正确认识在更高层次上实现公正与效率的统一,最高人民法院副院长沈亮撰文对认罪认罚从宽制度进行了阐释^④。文章提出,从制度定位上看,认罪认罚从宽的根本目的是落实宽严相济、促进司法公正;被告人只是通过认罪认罚争取从宽,而不是就定罪量刑进行讨价还价。从制度性文件的角度来看,《指导意见》明确规定了4条基本原则:宽严相济、罪责刑相适应、证据裁判、公检法配合制约;并强调对从宽的理解是“一般应当从宽”,但“不是一律从宽”。《人民检察院办理认罪认罚案件开展量刑工作指导意见》提出了人民检察院对认罪认罚案件提出量刑建议的5项基本原则,分别为宽严相济、依法建议、客观公正、罪责刑相适应、量刑均衡。这5项基本原则均围绕量刑建议的公正性展开,表现出明显的法定主义特点。

可见,我国认罪认罚从宽的制度设计具有明显的公正优先导向,强调从宽的“法定性”,以“事实清楚,证据确实充分”为基本前提。因此,有学者将这一模式称为“规范控制路径”。而我国认罪认罚从宽制度的司法实践充分回应了这一思路,表现出相对保守的“从宽”克制倾向。

(二)我国认罪认罚从宽制度司法实践中的“从宽”克制倾向

1.认罪认罚的实际作用在“底线”游走,接近量刑情节“坦白”

从前述发现“四(一)”来看,认罪认罚对有期徒刑量刑的影响并非无底线的“讨价还价”,其实际作用接近“坦白”,并且远低于“自首”“谅解”“积极赔偿”。司法实践中对认罪认罚从宽作用的定位或心理评价,更加接近坦白这一具体的量刑情节。

这一方面说明,我国的认罪认罚司法实践绝非“交易”。这一点可以在与美国的辩诉交易的对比中清晰看到。2005年“美国诉布克案”^⑤以前,辩诉交易严重倾向于检察官,而检察官以获得被告人认罪为目标主导量刑协商,导致普通程序案件与交易案件量刑差异巨大。一份基于美国1997-2004年量刑数据的研究分析比较了辩诉交易、无陪审团审判与陪审团审判之间的量刑差异,结果表明陪审团审判的刑期显著高于接受辩诉交易的刑期,在马里兰州这一数值为3.5倍,对于毒品案件这一数值高达7倍^⑥。一份基于美国2010-2020年间提起诉讼的分析论文发现,辩诉方案中的刑期平均占被告原判

① 陈卫东:《认罪认罚从宽制度研究》,《中国法学》2016年第2期。

② 周光权:《论刑法与认罪认罚从宽制度的衔接》,《清华法学》2019年第3期。

③ 《多城市试点刑案认罪认罚从宽 如何防范“花钱买刑”?》, <https://www.chinanews.com.cn/gn/2016/09-03/7992896.shtml>, 访问日期:2022年2月20日。

④ 沈亮:《凝聚共识 推进认罪认罚从宽制度深入有效实施》,《人民法院报》2021年7月22日,第5版。

⑤ *United States v. Booker*, 543 U. S. 220, 2005, <https://suprme.justia.com/cases/federa/us/543/220/>, 访问日期:2022年2月20日。

⑥ Nancy J. K., David A. S., Sara S., Robert R. W., “When Process Affects Punishment: Differences in Sentences after Guilty Plea, Bench Trial, and Jury Trial in Five Guidelines States”, *Columbia Law Review*, 2005, 105, pp.959-1009.

最高量刑的55%,即检方提出的减刑幅度平均为45%^①。这与我国认罪认罚从宽制度实践的差异巨大。

另一方面也说明,我国认罪认罚对实体量刑从宽的适用基本上是在“底线”游走。《指导意见》关于“从宽的幅度”的规定明确指出,认罪认罚的从宽幅度一般应当大于仅有坦白情节时的从宽幅度。这就给出了认罪认罚从宽幅度的“底线”。而在司法实践中,认罪认罚的实际作用与“坦白”非常接近。这似乎表达了一种既不得不从宽、又不愿意太从宽的司法情绪。我们在与刑事审判一线的法官访谈中发现,法官对于认罪认罚对量刑从宽的作用仍存在不少疑虑:“认罪认罚与坦白、自首等制度并无本质区别,如果仅仅因为适用了认罪认罚程序就大幅从宽,是否能经得起历史的考验?”

关于认罪认罚作为量刑情节在刑罚论中的地位,有学者主张应当属于预防刑的范畴,即区别于违法事实和责任事实等影响责任刑的情节,而与犯罪人属性(如前科、累犯等)和犯罪人态度(如自首、立功等)等情节具有同类属性。因此,认罪认罚仅能对最终刑期起到调节作用,但作用有限,在立法上不能夸大这种影响^②。要注意到,“坦白”处于法定从宽量刑情节谱系中效力最低的层级(这一点可以在《量刑指导意见》关于从宽幅度的规定以及相关的实证研究^③中看出)。当认罪认罚与坦白站在同一层级时,即使站在预防刑的角度,目前的司法实践中认罪认罚对量刑的调节作用也处于“底线”位置。

2. 传统量刑情节惯性适用,认罪认罚从宽的实效空间不足

由前述发现“四(二)”可以看出,认罪认罚从宽制度实施后(2019年及以后),其他从宽情节的估计值表现依旧非常稳定,似乎认罪认罚从宽制度的实施并没有明显影响其他从宽量刑情节的实际作用。这一现象或许与传统量刑情节的适用惯性有关。从制度存续时间来看,自首、坦白、刑事和解与赔偿谅解具有明显的制度稳定性。而认罪认罚从宽作为一项创新制度,与其他量刑情节的关系还没有彻底厘清,在适用的权重和优先级上可能存在差异。更为重要的原因或许在于,目前关于认罪认罚的一般性规定并没有给足认罪认罚量刑从宽发挥实效的空间。

一方面,《指导意见》“关于从宽的幅度”的规定给定了认罪认罚实体从宽幅度的基本区间。概括地看,第一,认罪认罚应比坦白作用大,这就给定了认罪认罚从宽幅度的“底线”。第二,多从宽情节共存时,认罪认罚从宽应限定在法定刑幅度内。这就意味着认罪认罚的作用类似“从轻”情节,而弱于“减轻”情节。这就给定了认罪认罚从宽幅度的“上界”。第三,当自首、坦白与认罪认罚竞合时,择一评价。在新版《量刑指导意见》中,不重复评价的范围进一步扩展至当庭自愿认罪、退赔退赃、赔偿谅解、刑事和解、羁押期间表现好等量刑情节,基本涵盖了认罪认罚的构成要素。总体来看,“底线较低,上界不高,且各情节从宽效应不能叠加”。在“底线”与“上界”确定的前提下,“折扣”幅度总量控制,各量刑情节之间的适用是零和博弈。

另一方面,司法实践中认罪认罚与自首、坦白等其他从宽量刑情节大范围大比例同时认定,在不能重复评价的指导原则下,具体情节的适用转化为“戴哪顶帽子”的问题。对此我们进行了数据检验。我们排除了具有自首、坦白两个法定从宽量刑情节的判决后,剩余认罪认罚判决630件;进一步排除赔偿谅解、刑事和解等情节的判决后,认罪认罚独立适用的案件仅剩余不足50件。可见,认罪认罚与传统从宽量刑情节大范围大比例重叠。在“不重复评价”的原则下,要么根据认罪认罚给出一揽子的“从宽”量刑,要么仅以传统从宽量刑情节作出量刑。在后一种情形下,认罪认罚的意义则仅剩下了程序上的简化和司法效率的提升。而在前一种情形下,一揽子从宽也很容易因为超出“底线”与“上界”的区间限制而无法充分兑现。具体来说,新版《量刑指导意见》明确在认罪认罚与其他从宽量刑情节

① Findley K., Amaya C. A. Hatch G., Smith J., “Plea Bargaining in the Shadow of a Retrial: Bargaining Away Innocence”, *University of Wisconsin Legal Studies Research*, Paper No.1726, 2021, <https://ssrn.com/abstract=3944845>, 访问日期: 2022年2月20日。

② 周光权:《论刑法与认罪认罚从宽制度的衔接》,《清华法学》2019年第3期。

③ Wang Fang, Guo Liang, “Do Confessions Contribute to Lenient Punishments in China?” *Tsinghua China Law Review*, 2019, 12(1), pp.57-85.

同时存在时可以给予“一揽子”的从宽处罚。“基础版”规定认罪认罚可以减少基准刑的30%以下;“升级版”则规定“具有自首、重大坦白、退赃退赔、赔偿谅解、刑事和解等情节的,可以减少基准刑的60%以下,犯罪较轻的,可以减少基准刑的60%以上或者依法免除处罚。”但事实上,认罪认罚作为“从轻”情节的定位,限制了上述“一揽子”从宽承诺的兑现。特别是在没有自首等“减轻”情节时,法定刑底线不能突破,60%甚至60%以上的从宽幅度难以实现。

六、兑现制度承诺,实现更高层次上公正与效率的统一

(一)认罪认罚从宽制度承诺兑现不足的深层次原因

认罪认罚从宽制度的酝酿与实施,发端于节约司法成本、提高司法效率的现实需求,争议焦点集中于提高司法效率的同时如何不伤及司法公正的实现。认罪认罚从宽制度实施之初,各界对制度实施的担忧主要表现为两个方面,一是担心协商可能带来过度从宽,“量刑时面对‘下不了手’的难办案件突破实体法的量刑限制”,因此特别强调刑法必须及时与认罪认罚从宽制度相衔接^①。二是担心被追诉人认罪认罚却得不到适当、有效从宽处理,因此特别强调认罪认罚从宽制度改革应致力于落实从宽处罚的优惠,落实对被追诉人的实体权利供给^②。目前来看,第一个方面的担忧已得到充分的重视,但第二个方面的问题有进一步关注的必要。为此,需要我们进一步思考的是,在公正优先的价值定位中,如何充分兑现制度承诺,更好地发挥认罪认罚从宽的制度优势,“在更高层次上实现公正与司法效率的统一”。

从诉讼类型来看,认罪认罚从宽制度具有明显的协同型司法的特点。“要约、承诺的程序要素,以及双方均作一定利益让渡(一方认罪认罚与另一方从宽处置)的实体内容”^③,是认罪认罚从宽制度的当然内涵。从实用性的角度看,兑现制度承诺是确保制度良好运行的基础。兑现承诺-激励认罪认罚-提高司法效率与诉讼满意度-提高司法公信力,这是一个正向激励的循环链条。其中,兑现承诺是核心动力,被追诉人认罪认罚是兑现承诺的对价条件。在这个循环中,“从宽”待遇的兑现程度必然对被追诉人的后续行为产生影响,如果陷入类似“坦白从宽、牢底坐穿”的尴尬局面,司法公信力与制度有效性都将遭受质疑。因此有学者认为,“从宽”承诺的兑现应具有强制性,特别是对于轻罪案件的认罪认罚,从宽承诺兑现的强制性具有正当性基础^④。问题在于,“从宽”承诺的兑现,不应仅关注“是与否”,更应关注“多与少”。目前来看,“多与少”是困扰司法实务工作者的主要问题。

在对一线法官进行访谈的过程中我们发现,关于从宽问题的困惑主要来源于三个方面:一是认罪认罚从宽是一项刑事诉讼制度,而不是《刑法》中的法定量刑情节,如何理解认罪认罚从宽制度在实体法或刑罚制度中的定位?二是根据《刑事诉讼法》和《指导意见》,认罪认罚的从宽不是“应当”,而是“可以”。如何区分把握具体的适用场景?三是认罪认罚与传统的从宽量刑情节之间不可避免地存在交集,在《指导意见》规定的“不重复评价”或综合评价的标准下,认罪认罚对量刑的实际效力如何体现?新版《量刑指导意见》明确将“认罪认罚”“当庭自愿认罪”作为“常见量刑情节”,部分地回应了上述第一个问题,但对于其他两方面的问题并未澄清。

也有学者将目前“从宽”承诺兑现不足的原因概括为控辩合意的去协商化、量刑从宽法庭审理的非实质化、量刑从宽的宽宥窄化三重弊病^⑤。具体而言,第一,我国认罪认罚实际运作中的协商,是一种依检察官法定职权的“引导协商”方式。在这一模式下,公诉方在协商过程中占据主导地位,公诉方

① 周光权:《论刑法与认罪认罚从宽制度的衔接》,《清华法学》2019年第3期。

② 左卫民:《认罪认罚何以从宽:误区与正解——反思效率优先的改革主张》,《法学研究》2017年第3期。

③ 龙宗智:《完善认罪认罚从宽制度的关键是控辩平衡》,《环球法律评论》2020年第2期。

④ 刘泊宁:《认罪认罚从宽制度中司法承诺之考察》,《法学》2020年第12期。

⑤ 陈实:《论认罪认罚案件量刑从宽的刑事一体化实现》,《法学家》2021年第5期。

与被迫诉方的地位很难对等,双方掌握的信息也很难对称。在检察机关一系列考核指标的任务压力下,为达成认罪认罚考核目标,有可能存在过度承诺、量刑隐瞒^①等现象。从最终的量刑建议来看,在控辩双方地位和专业不对等的情况下,量刑是否真的被给予了实质的优惠,很难说清。第二,程序简化带来法庭审理的非实质化,《指导意见》规定人民法院的职责是在审判阶段进行认罪认罚的自愿性、合法性审查。法庭审理主要围绕认罪认罚的自愿性、合法性进行审查,而忽视量刑问题。对于量刑问题,人民法院“一般应当”采纳检察机关提出的量刑建议,仅有“明显不当”时可以建议调整量刑建议(《刑事诉讼法》第201条)。这就损失了一部分对量刑建议进行调整的机会。第三,认罪认罚与自首、坦白等其他从宽量刑情节大范围大比例重叠,在不能重复评价并且折扣幅度总量限制的双重控制下,认罪认罚的“从宽”价值很难独立体现出来。这一点在本文第五部分已有论述。

事实上,正是认罪认罚制度的不完备、从宽适用规则的模糊,成为司法实践中量刑从宽克制的深层次原因,也成为制度承诺兑现不足的深层次心理机制。

(二)在确保公正的前提下实现司法效率的提高

基于上述深层次的矛盾和困难,有学者提出从诉讼程序角度再造认罪认罚案件量刑从宽合意机制,赋予辩方量刑协商的启动权,规范量刑建议的协商性交涉,开放认罪量刑从宽答辩机制,从而改变控方对量刑建议形成的绝对主导现状^②。这对于完善依检察官法定职权的“引导协商”方式具有借鉴意义。但必须看到,“法定化”“公正优先”是我国认罪认罚从宽制度改革一再强调的突出特点,也是我国认罪认罚从宽制度与美国辩诉交易制度的本质区别。在这一前提下,法定规范的引导具有更加重要的意义。明确认罪认罚从宽制度的规则与价值,消除司法者的思想顾虑,是兑现认罪认罚从宽制度承诺,在更高层次上实现公正与司法效率的统一的的基础。基于这一判断,化解认罪认罚从宽制度的效率初衷与公正目标之间的矛盾,在确保公正价值的前提下实现司法效率的提高,可以从以下两个方面入手:

第一,分级分类,建立从宽量刑阶梯。其目标是明确对于轻罪案件的认罪认罚给予更大幅度的从宽。从目前对认罪认罚从宽制度的官方解释来看,对相对轻罪及一些过失犯罪、未成年人犯罪等社会危害不大且案情明了的认罪认罚案件给予较大幅度的从宽,已经形成了较为一致的意见。但对此尚未形成具有法定效力的文件,也缺乏具体的指导。可以考虑在《量刑指导意见》中进一步明确区分具体案件性质、情节和对社会的危害程度,综合权衡从严、从宽因素,做到区别对待、罚当其罪;界定轻罪的范围,并且增加对轻罪案件认罪认罚从宽幅度的规定,引导形成对轻罪案件可以给予更大幅度从宽的法定预期。

第二,调整认罪认罚的定位,扩大认罪认罚从宽幅度的法定空间,赋予其“减轻”情节地位,即可以突破法定刑底线量刑。其目标是化解“不重复评价”可能带来的量刑从宽额度零和博弈的现实困局。目前,认罪认罚作为“从轻”情节的定位,限制了从宽承诺的兑现。新版《量刑指导意见》关于60%甚至60%以上的从宽幅度在现有认罪认罚法定效力定位下难以实现。因此,需要调整认罪认罚对量刑的效力定位,将其从“从轻”情节升级为“减轻”情节,具体可以比照“自首”情节来调整认罪认罚从宽实效的评价预期,使其高于“坦白”情节。

附录 数值试验结果

特征	实验值	估计值	特征	实验值	估计值
偏置项	1.2569	1.2897	认罪认罚	-0.0155	-0.0102
轻伤人数	5.1889	5.1641	抢救被害人和挽回损失	0.0974	0.1012
轻微伤人数	0.9017	1.0614	初犯、偶犯	0.0149	0.0132
自首	-0.1078	-0.1087	是否有辩护人	0.0046	0.0052

① 刘泊宁:《认罪认罚从宽制度中司法承诺之考察》,《法学》2020年第12期。

② 陈实:《论认罪认罚案件量刑从宽的刑事一体化实现》,《法学家》2021年第5期。

续附录

特征	实验值	估计值	特征	实验值	估计值
坦白	-0.0355	-0.0359	被害人过错	-0.0303	-0.0310
和解	-0.2076	-0.2087	累犯	0.1971	0.2007
谅解	-0.2043	-0.2048	主犯	0.1216	0.1325
积极赔偿	-0.0925	-0.0937	前科	0.0226	0.0204
从犯	-0.1721	-0.1422	互殴	-0.0134	-0.0167
缓刑哑变量	-0.0117	-0.0090	附带民事诉讼	0.0939	0.0960
当庭自愿认罪	0.0375	0.0390	民间矛盾	-0.0260	-0.0274

A Study on Impact of Leniency on Sentencing for Suspects Admitting Guilt & Accepting Punishments

—Based on Data of Misdemeanor Crime of Intentional Injury

Wang Fang¹ Gan Die² Liu Nian²

(1.Data Science Institute of Shandong University, Jinan 250100, P.R.China;

2.Academy of Mathematics and Systems Science, Chinese Academy of Science, Beijing 100190, P.R.China)

Abstract: Does suspects admitting guilt & accepting punishments contribute to lenient punishment? This is the key point of a new mechanism—Leniency for Suspects Admitting Guilt & Accepting Punishments. Through analyzing the first—instance judgments of 177,161 intentional assaults causing minor injuries, we find that the effect of the mechanism on the sentence of misdemeanors is close to the “confession”. Further inspections manifest that the implementation of the mechanism did not change the effects of traditional sentencing circumstances for lighter punishment significantly. Research conclusions indicate that the implementation of the mechanism shows a tendency of justice priority. At the same time, compared with the traditional sentencing circumstances for lighter punishment, the application of the new mechanism is conservative for misdemeanors, and sentences are often determined in the lower part of the sentencing interval, which implies that the leniency range or the application of probation in misdemeanors can be appropriately increased, so as to achieve the unity of justice and efficiency. In terms of analysis methods, a non—linear sentence model that highly simulates judicial practice is constructed. Considering the spatial consistency and correlation effect over time of criminal judgments, we use the Recursive Least Squares algorithm to estimate the actual impact and possible changes of the mechanism, and use Error Probability Inequality to test the reliability of estimates under limited data. Our methods have two characteristics: First, the independent and identical distributed assumption of noises is not required, which is more suitable for the nature of social data such as judicial judgments. Secondly, it is possible to observe the trends of main characteristics over time.

Keywords: Leniency for Suspects Admitting Guilt & Accepting Punishments; Misdemeanor; Lenient sentencing; Recursive Least Squares algorithm

[责任编辑:李春明]