

量刑建议的“高采纳率”误区

孙 皓*

摘 要 量刑建议已经日渐成为刑事诉讼格局中控辩审关系的联结纽带,尤其是在适用认罪认罚从宽程序的案件集群内。而检察机关提出的量刑建议能否被法院所采纳,也牵动着各个诉讼权能主体的敏感神经。目前呈现的“高采纳率”是否符合制度改革理想预期,则真正取决于其内在衍生进程。对此,多维度的实证调查有助于客观准确的认知形成,以防过度释放片面盲目的乐观情绪。从现有情况来看,法院对检察机关所提量刑建议的采纳,并不是完全出于法律评价上的相互认同,而一定程度上受到了双方关系融洽与否的影响。究其根源,配合制约原则仍深刻影响着量刑建议的实践运行效果。只有在排除了某些观念误区的前提下,量刑建议才能满足公正与效率之间的均衡诉求。同时,刑事司法权即可被平稳维系于控辩审三方的共同合意。当然,这种理想情境的达成绝非朝夕之功,而势必仰赖程序模型的持续进化。

关键词 量刑建议 认罪认罚从宽 检法关系

引言:应然、实然与必然

不可否认,随着认罪认罚从宽改革的持续推进,我国刑事诉讼法学研究的重心已然呈现某种虹吸效应。^{〔1〕}其主要特质,即针对认罪认罚从宽的理论探索愈发具象化且富于针对性。这中间,又尤以针对量刑建议的学术研究最为突出。总的来说,目前关于量刑建议的理论探索大致可区分为两个维度。一是就其基本内涵的解构,藉由性质、特征等命题的基本立场,最终

* 天津大学法学院副教授。

〔1〕 所谓虹吸效应,社会学意义上系某一区域将其他区域的资源全部吸引过去,从而使得自身相比其他地方更加有吸引力,从而持续并加强该过程的现象。在本文中,其指涉近年来刑事诉讼法学研究以认罪认罚从宽改革作为聚焦热点的现象。

落脚于法律效力层面。^{〔2〕}特别是随着余金平交通肇事案的曝光,由此牵涉出量刑建议背后的检法关系问题,则诱发了更多学者的研究兴趣。^{〔3〕}此外,也有不少学者运用解释学的方法论手段,以《刑事诉讼法》第201条的“一般应当采纳”条款切入,剖析量刑建议的功能范畴,试图为司法实践中某些令人困惑的难题供给破局思路。^{〔4〕}二是针对量刑建议的精准化建构命题,开展相应的自洽性设计。^{〔5〕}此类研究事实上已默认量刑建议的裁判制约力,又将其自身的精准化作为前置要件,论证思路大多以司法实践中幅度刑量刑建议的粗放倾向作为批判对象。深究起来,上述两个维度的研究成果都暗含了一条应然的逻辑脉络,即检察机关提出的量刑建议需要被法院大体接受,且采纳率越高越好。当然,这种预期须诉诸于特定的前提:每一份量刑建议的内容均是准确而适度的。正因为如此,当我们回溯至现实场域,量刑建议的采纳程度及比重早被视为衡量认罪认罚从宽制度实施成效的硬性指标,甚至说是一种“定言令式”亦不为过。^{〔6〕}

循着整体主义的理论推导进路,上述应然逻辑是合乎常识理性的。既然“以审判为中心”的刑事诉讼改革强调庭审实质化的优先位阶,那么其指涉的“诉讼证据质证在法庭、案件事实查明在法庭、诉辩意见发表在法庭、裁判理由形成在法庭”,势必要仰赖于成本投入的充分支持。^{〔7〕}为此,以“认罪认罚从宽”为代表的制度图式,必然要承担起提升诉讼效能、降低资源损耗的目标任务。而量刑建议的提出集中了诉讼参与各方的合意成果,故迅速简约地被法院采纳即构成了效率价值的突出彰显。^{〔8〕}因此,量刑建议的高采纳率便无可厚非地汲取了正当性。从实然角度出发,量刑建议所诉求的这种高度认同似乎正接近极值。在2020年10月由最高人民检察院向全国人大常委会所作的《关于人民检察院适用认罪认罚从宽制度情况的报告》中,法院对检察机关确定刑量刑建议采纳率就显示为89.9%。^{〔9〕}显而易见,这是一种

〔2〕 如陈卫东:“认罪认罚案件量刑建议研究”,《法学研究》2020年第5期,第160—164页;再如熊秋红:“认罪认罚从宽制度中的量刑建议”,《中外法学》2020年第5期,第1177—1178页。

〔3〕 参见龙宗智:“余金平交通肇事案法理重述”,《中国法律评论》2020年第3期,第95页。

〔4〕 参见孙远:“‘一般应当采纳’条款的立法失误及解释论应对”,《法学杂志》2020年第6期,第112—120页;董坤:“认罪认罚案件量刑建议精准化与法院采纳”,《国家检察官学院学报》2020年第3期,第28—38页。

〔5〕 如李振杰:“困境与出路:认罪认罚从宽制度下的量刑建议精准化”,《华东政法大学学报》2021年第1期,第148—152页;再如赵恒:“量刑建议精准化的理论透视”,《法制与社会发展》2020年第2期,第131—135页。

〔6〕 所谓“定言令式”,指涉一种“普遍法则形式”。参见(英)保罗·科利尔:《战争、枪炮与选票》,吴遥译,南京大学出版社2018年版,第47页。

〔7〕 参见陈卫东:“以审判为中心:解读、实现与展望”,《当代法学》2016年第4期,第14页。

〔8〕 参见郭烁:“控辩主导下的‘一般应当’:量刑建议的效力转型”,《国家检察官学院学报》2020年第3期,第20页。

〔9〕 参见徐日丹:《最高人民法院检察长张军作关于人民检察院适用认罪认罚从宽制度情况的报告》,载最高人民法院官网,https://www.spp.gov.cn/spp/zd gz/202010/t20201015_481978.shtml,最后访问日期:2021年2月23日。

异常理想的数据形态,根本不需要任何比照对象的衬托。那么,据此是否可以导出如下推断:当前对于量刑建议的激烈探讨——遑论其精准化建构抑或法律效力之评价——不过是研究者与实务人士追求精益求精的态度体现;即便存在余金平案之类的争执,亦不过沧海一粟,至少在体量上并不占主流,检察院、法院两家在量刑建议可采性的问题上共识远大于分歧。

在作者看来,应然与实然的这种近乎一致并不代表着客观规律的显灵。易言之,应然与实然之间尚需要某种必然性的衔接,方可建构难以颠覆的因果链条。这就如同“通过一个地点不能和停留在那个地方画等号,就是说不能和存在于那个地方画等号”。^{〔10〕}即使某些区域目前实现了100%的采纳率,^{〔11〕}也不能由此断定该地的量刑建议——乃至认罪认罚从宽——之实施已完全契合改革预设的效率价值。这是因为,我们不能认定上述诉讼进程未受任何附加因素的影响,而隐含了部分虚假而扭曲的成分。为了准确厘清这种应然、实然及必然之间的关系线路,澄清量刑建议“高采纳率”可能涵盖的某些误区,还需要优先回归至司法操作环节,以探知此现象背后真正的形塑路径。

一、量刑建议的实施效果评估

眼见为实,耳听为虚。虽然针对量刑建议的实践状况已然形成了不少可靠的调研范例,^{〔12〕}但是专门着眼于高采纳率背后作用机制的素材整理似乎仍属凤毛麟角。为此,作者通过特定实证手段,尝试进一步挖掘量刑建议在诉讼进程中的影响方式。一方面,小范围样本的调查结论可成为佐证官方数据的基础;另一方面,定性与定量方法的有效结合则利于明确某些附加因素的存在与否。

(一)针对庭审直播的观察

依作者之见,对于量刑建议实施效果,目前大致可有三种调查手段廓清其基本面。一是通过裁判文书网收集相关资料,以辨识各类案件的量刑建议采纳状况。这种方法的优势在于聚拢较大的样本量,其具体手段系分析裁判文书中法官涉及量刑建议的说理论证。然而,这种方法难以清晰还原庭审运行全貌,加之相当数量的裁判文书并未过多着墨于量刑建议的审查,故难以为本文论点提供足够的信息支持。二是藉由田野调查方法,利用亲身观察司法实践的机会,近距离体察量刑建议的运行机理,并“从直觉、奇闻异事与相关性中进行推断”。^{〔13〕}此一手段的特质能确保信息挖掘深度,并揭露某些不易察觉的变量因子;但是,其操作效果往往受限于成本投入,且难以达致较大的样本量要求,进而无法企及实证数据的广泛性与代表性。因

〔10〕 (法)萨特:《存在与虚无》,陈宣良等译,生活·读书·新知三联书店2016年版,第271页。

〔11〕 参见陈国庆:“量刑建议的若干问题”,《中国刑事法杂志》2020年第1期,第7页。

〔12〕 参见左卫民:“量刑建议的实践机制:实证研究与理论反思”,《当代法学》2020年第4期,第47页。

〔13〕 (美)艾伦·格伯、唐纳德·格林:《实地试验:设计、分析与解释》,王思琦译,中国人民大学出版社2018年版,第2页。

此,这种模式更适宜在本研究中充当补充手段。三是借助司法公开化与信息化的改革成果,以庭审网络直播作为数据收集的载体,近乎于身临其境地了解量刑建议的提出及讨论,直至法庭作出即时决策。^[14]这种方法某种程度上补齐了前述第一种实证手段的短板,尽管在样本量上或许依旧无法与之相提并论,却客观展示了量刑建议在庭审中的运行表象。此外,其便宜性避免了资源投入的靡费,而竭力促成以较小代价换取有价值信息的研究目标。当然,由于追求直播效果的缘故,诉讼参与者的实际表现多少会带有失真倾向,不过总体上并不妨碍本文研究目标的达成。为此,作者安排了五名法科学生以简单随机抽样方式,^[15]选取了100场一审刑事案件的审判直播进行观察与评估,以期获得有价值的现场信息。

具体来说,上述五名研究人员被要求通过登录“中国庭审公开网”,并以“刑事案件”和“基层法院”作为搜索条件,各自在一个月时间内集中观看20起案件庭审。需要注意的是,每名研究人员选择的观察对象必须是当日开庭审理的直播案件,而不得带有针对性地去回顾过往的视频录像。为避免重复性观看,五名研究人员还被限定了具体的时间范围进行抽样,而其选择的对象均统一系该网站“正在直播”栏目所即时上线的最新庭审视频。鉴于“中国庭审公开网”接入了近乎全国范围内各级法院的直播信号,其数据资源的覆盖广度是毋庸置疑的;而“正在直播”栏目则天然具备了技术层面的随机属性,故其置顶的庭审视频便可视为一种抽样结果。这样一来,研究代表性也一定程度得到了保证。换言之,研究人员抽取的案例本就是随机生成的,在特定时间内基本能够反映国内刑事一审程序中量刑建议运行的普遍情形。考虑到研究投入等外部因素,由此形成的定量结论虽然没有极其庞大的数据信息作为基础,却未见得不能给予我们更多的警醒和提示。

而在观察庭审的过程中,研究人员围绕量刑建议的实操环节,选取其中对法官心证可能产生影响的因素,进行信息记录并制定标准,继而形成庭审视频记录标准表。其内容包括量刑建议形式、参加庭审主体等客观内容,以及研究人员对于庭审直播的满意度评价等主观要素。而研究人员观看的庭审直播案件涉及刑法十类罪名中的五类,分别是:危害公共安全罪、破坏市场经济秩序罪、侵犯公民人身权利、民主权利罪、侵犯财产罪、妨害社会管理秩序罪(见图1)。从案件类型分布来看,随机抽样的结果大致可以映射司法实践的广泛性与全局性,并不至于出现偏废某些典型案件的情状。

[14] 目前的庭审直播基本要求当庭形成裁判结果。参见孙皓:“关于当庭宣判的逆向反思”,《当代法学》2020年第2期,第122页。

[15] 所谓简单随机抽样,又称纯随机抽样,是一种特殊的等概率抽样方法,即总体中每一个体均有同等被选机会,而且样本中每一个体是被单独选出的,它是一个元素抽样。简单随机抽样分为重复抽样和不重复抽样两类。本文中作者所运用的为不重复抽样,即在样本集群中,被选取的庭审案例只能出现一次。其优点在于,进入样本中的个案事实上是机会均等的,且在一定程度上足以全面、客观地呈现刑事司法运行生态。参见袁方主编:《社会研究方法教程》(重排本),北京大学出版社2015年版,第155—157页。

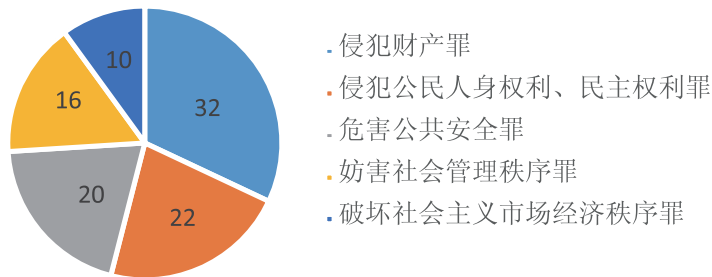


图 1 庭审观摩案件类型分布图(基于 100 起观摩案件的分类)

研究人员在完整观看庭审直播后,综合整体视频情况,对于标准表中控辩双方表达意见是否充分、量刑信息是否精准发现与利用、程序是否透明、控辩双方参与及法官指挥作用的情况、程序满意度等若干方面进行评价。在上述过程中研究人员重点记载了法官采纳量刑建议的基本情况,并填写庭审视频记录评估表(见图 2)。值得注意的是,针对庭审直播的采集并未设定认罪认罚从宽的框架限制,故在作为样本集合的 100 起个案中,被告人不认罪认罚的尚有 3 起,占总量的 3%。这一定程度上暗合了外界对于繁简分流机制的功能预期。而从结果上看,被告人认罪认罚的 97 起个案中,法官最终未采纳量刑建议的状况仅有 2 次,占比 2.06%;大致采纳量刑建议的有 39 起个案,占比 40.21%;^[16]完全采纳量刑建议的 56 起案件则占 57.73%的比重;另外被告人未作认罪认罚处理的案件则达到了 100%的采纳率。由此,量刑建议的采纳率的确处于高位,然而结合量刑建议的法庭讨论周期、主体参与程度等指标信息,却又难以武断认定其运行质量。

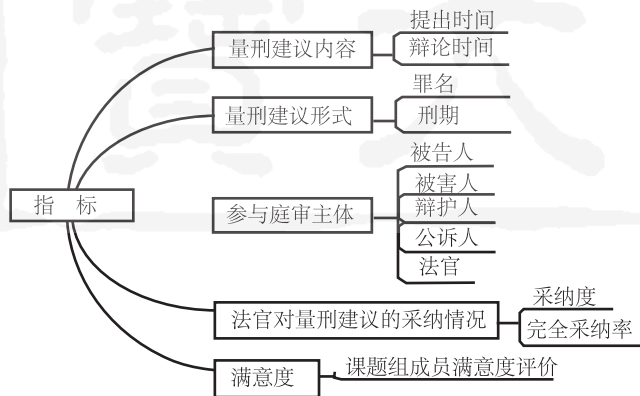


图 2 庭审观摩评估体系图

在法庭辩论阶段,围绕量刑建议之适宜与否的探讨在时长占比的平均值层面仅及于 23%

[16] 这里需要说明的是,实证研究中发现的“大致采纳量刑建议”,指的是检察机关仅提出了幅度刑建议,而法院在此范围内予以认可。这表明司法实践中公诉部门尚未完全满足精准化量刑的理想条件,但是以具体刑期为量刑建议内容的做法已经逐步占据了相对较大的比重。

左右。而众所周知的是,当诉讼各方主体已就犯罪事实及性质等问题达成一致,那么庭审的确认活动本应围绕量刑妥当性展开。^[17] 于是,上述数字呈现的结果显然不能算是理想结局。综合各诉讼主体的参与程度,量刑建议的采纳似乎未同法庭辩论的清晰性建立正相关。在这里,研究人员根据自身的满意度认知为量刑建议的接受程度设定了1—5的量化参数值,数值越大代表法官最终的认可度愈高。同理,法庭辩论围绕量刑建议的讨论充分程度也借用数值进行指代。其结果,再次佐证了量刑建议高采纳率的形成貌似另有隐情(见下表)。尤其是看起来毫无规律的数值频次分布,基本证实了量刑建议尚未在法庭环境中塑造有迹可循的反馈机制,更多时候只不过是象征性地展示于诉讼进程中。纵然不同研究人员会带有主观倾向审视庭审直播内容,难免在认定标准上有所差异,但相近的学识背景及先前共同协商而成的估值方法已最大限度削弱了这些干扰因素。不管怎样,通过对庭审直播的观察,可得出如下结论:量刑建议的高采纳率具备了实然性,却并不能由此推导程序运转的流畅自如与高度理性化。

表1 不同评分值出现频次表(单位:次)

量刑建议 接受程度 / 量刑建议 辩论充分程度	5分	4分	3分	2分	1分
5分	1	5	3	1	0
4分	6	9	5	2	1
3分	11	8	3	5	0
2分	9	2	7	8	1
1分	4	3	2	4	0

还有一点颇为值得关注,即在100起直播案件中,只有36%的被告人拥有辩护人。有辩护人支持的案件中,量刑建议被完全采纳的数量为24件,大致采纳的系14件,而完全不采纳的只有1件;无辩护人的个案集合里,完全采纳量刑建议的系35/61,大致采纳的则为25/61,完全不采纳也是1/61。单从数据比例来看,有无辩护人参与对量刑建议的采纳与否没有产生实质影响。这也间接表明,量刑建议的实施效果与庭审机制的运行之间是相互脱节的。尽管针对被告人的量刑关乎其切身利益,且构成了大部分案件亟待解决的核心事项,但涉及量刑建议的博弈在审判现场根本没有被置于突出地位。

(二)定性材料的补充

那么,究竟应当如何解释前述数据的某些不合理之处呢?单从庭审内容入手恐怕依旧会

[17] 参见李奋飞:“论认罪认罚量刑建议与量刑裁决的良性互动”,《暨南学报》(哲学社会科学版)2020年第12期,第51页。

心存疑惑,而无法探知其背后的隐性机制。既然庭审无法给出直观信息,便只能在庭外寻觅有价值的蛛丝马迹。为此,作者运用前文中提及的第二种实证方法,在一定范围内展开田野调查,力图从司法实务人士提供的某些“轶事”中找寻答案。毕竟,“政策如何得到实施最终取决于那些执行政策的人”。^{〔18〕} 具体而言,作者选择通过当面访谈、电话交流等方式,同若干位相识的检察官、法官及辩护律师进行接触,收集与量刑建议相关的事例作为定性资料,以补充前述量化研究中潜在的信息盲区。最终,20余位司法实务人士向作者分享了自己处理量刑建议过程中的一些经历及心得。其中,很多细节性的“爆料”的确很难藉由在庭审现场进行观察的方式获取,却又未尝不是某些矛盾现象的解释来源。当然,作者此处并不能将所有收集的信息予以重现,却可从中筛选某些最具说服力的例证。其实在经历充分的访谈后,有两则事例具有较强的代表性和说服力,也能够为前述样本调查中的未尽之处供应阐释思路。无独有偶,这两起个例均系检察机关的量刑建议未被直接采纳的情形。然而通过此二案例,作者却无意中窥见量刑建议实现“高采纳率”的某些因由。

第一起案例是由辩护律师张某讲述的。在其为涉嫌猥亵妇女罪的王某代理案件过程中,检察机关在审查起诉阶段启动了认罪认罚从宽。在此期间,经张某努力而求得了被害人一方谅解,并由王某家属及时作出赔付。进而,公诉人提出具体的量刑建议,将缓刑作为王某认罪认罚的从宽对价。对此,无论是被告人抑或被害人均未提异议。然而随着案件进入审判环节,主审法官何某当庭否定了量刑建议,并将刑罚明确限定为实刑。这一过程发生在庭审期间,控辩双方均难以接受。何某给出的理由则是:王某的猥亵情节较为恶劣,给予其缓刑系量刑畸轻,有助长类似行为的嫌疑。尽管法庭辩论环节公诉人积极进行了交涉,试图令法官改弦更张,却最终无济于事。据张某表述,控审双方在法庭上几近于激烈争吵。在法官何某看来,根据最高人民法院对于《刑诉法》第201条中“一般应当采纳”的阐释思路,本案的量刑决策并无不可。^{〔19〕} 而对于这样的结果,作为辩护人的张某也很难理解,甚至建议立法应明文禁止法院随意推翻量刑建议。毕竟在其看来,先前控辩双方做了大量工作以达成认罪认罚从宽,而法官的偏执却使上述所有努力归零。既然连被害人都未作异议,裁判者的态度似乎多此一举。此外,据悉当地检法两家已多次对各类刑事案件的适用法律、刑罚轻重等事项产生争执,且近年来无罪判决数字始终呈增长态势。而为稳妥起见,非认罪案件一般不提量刑建议、认罪认罚案件当庭提出幅度型建议的做法,已被检察机关坚持了相当长的时间。

第二起案例则是作者同检察官李某交流后所获悉,即被告人孙某涉嫌贩卖毒品案。本案在审查起诉环节达成认罪认罚从宽,且孙某签署了具结书。作为办案人的检察官李某基于此

〔18〕 (美)格雷格·伯曼、奥布里·福克斯:《失败启示录:刑事司法改革的美国故事》,何挺译,北京大学出版社2017年版,第118页。

〔19〕 根据《最高人民法院关于适用〈中华人民共和国刑事诉讼法〉的解释》第354条规定:“对量刑建议是否明显不当,应当根据审理认定的犯罪事实、认罪认罚的具体情况,结合相关犯罪的法定刑、类似案件的刑罚适用等作出审查判断”。

在量刑幅度的底部提出建议,给予孙某较大的从宽优惠。而在案件开庭审理前,本案主审法官赵某却致电李某,主张调整孙某的量刑建议,以接近于刑罚幅度的顶格进行处分。李某当即表示不解。法官解释称,孙某令其家属近一段时间里四处托关系、找门路,试图说动法院给予其从轻处罚。由此,法官认定孙某的主观恶性不足以降低刑罚基准,仍须以严厉姿态对待;否则,之前的认罪认罚从宽处置不免被误认为司法腐败的产物。在听过上述理由后,检察官李某当即表示认同,不久就主动调整了量刑建议尺度。同时,李某告知作者在当地检法两家的办案人之间,诸如此类的事先沟通已司空见惯,以避免双方在法庭上产生摩擦,进而影响量刑建议的采纳效果。

(三)检法关系的“晴雨表”

当我们将上面两则事例与庭审直播的数据结论加以串联,就会对时下量刑建议的被采纳情况有更加深刻的认知。“深度的个案研究只要能够容纳足够的时空与关系张力,展示时间过程足够完整与复杂的面貌和机理,的确是可以揭示其所具有的普遍性启发意义的。”^[20]只要控审双方就量刑建议的大致思路或者具体内容有事先协调,且未持任何重大异议,那么后续庭审的确认活动不过是顺理成章走流程。反之,即便检察机关在审前阶段已经完成了所有分内工作,只要未就求刑目标的某些细节展开足够的私下沟通,就极易给量刑建议的落实造成不良影响。上述两则事例的根本区别,其实不在于外化的程序机制,而是体现于检法之间的日常关系。某种程度上,辩护律师张某在审查起诉环节的付出甚至是多于检察官李某的,认罪认罚的操作条件也更适宜;只是基于控方对事先沟通的忽略,法院的反馈可谓大相径庭。正因为如此,从庭审直播中很难发觉量刑建议受到应有的重视。易言之,决定量刑建议可采性的时间节点并未出现在庭审环节,而是在此之前的预先沟通中尘埃落定。之所以辩护人的存在与否没有显而易见成为干预变量,或许亦能从中找到答案。

即便最高层级的检法机关在量刑建议的接受模式上存有或明或暗的分歧,^[21]也很难否认这样的事实:国内绝大部分基层检法单位的相互关系还是比较融洽的。而具象化到办案人层面,有着工作对接需求的法官与检察官之间,出现剑拔弩张局面的概率还是大致低于双方的相互体谅及支持,哪怕后者带有逢场作戏成分。^[22]这样的状况过去在程序倒流、留有余地裁判等问题上映射得十分明显;^[23]而如今,随着量刑建议在认罪认罚从宽框架内对于控审之间资讯流转的诉求,催生了高度依赖检法关系的新场域。某种程度上,也可以将量刑建议的采纳情形视作某一地区检法关系融洽与否的“晴雨表”。当此地的检法模式以相互扶持为主流形态

[20] 吴毅:《小镇喧嚣:一个乡镇政治运作的演绎与阐释》,生活·读书·新知三联书店2018年版,第480页。

[21] 参见杨立新:“对认罪认罚从宽制度中量刑建议问题的思考”,《人民司法》2020年第1期,第9页。

[22] 检察官李某就明确向作者表示:“我们和法官之间只是工作分工不同,即便没有提出量刑建议的问题,日常工作中也需要相互支持、帮助,伤面子与和气的尽量少干,不要相互为难”。

[23] 参见陈瑞华:“留有余地的判决——一种值得反思的司法裁判方式”,《法学论坛》2010年第4期,第26页。

时,量刑建议往往在开庭甚至提起公诉前即在控审间达成共识,只待最终程序确认;当然,在此期间双方可能会及时消弭信息不对称现象,就如上述第二起事例提到的那样,其结果是促成量刑建议符合双方共同预期。倘若本地的检法机关长期存在龃龉,恐怕基于量刑建议的评价就会激生双方的频繁摩擦,而事先的私下协调自然无从谈起。由此,信息不对称所衍生的误解与不信任更易被放大,甚至进一步激化双方的隔阂乃至对立情绪。显然,这与“人情”及“面子”影响下的机构人格化密切相关。^[24]不妨作这样的理解:量刑建议作为评价认罪认罚从宽实施成效的重要指标,却被一只“无形之手”左右;而这只“无形之手”即为检法之间日常的互动模式及协调程度。

二、“高采纳率”的现实风险

或许有人会认为由此生成的“高采纳率”并无不妥。毕竟,随着认罪认罚从宽的脱颖而出,刑事诉讼的程序主导模式已悄然改变。^[25]既然繁简分流的目标达成需要大部分个案经由认罪认罚渠道高效终结,那么检法机关针对量刑建议的庭外协商似乎是大有裨益的,何况其稳定性又能够有所保障。但是,只看结果而不审视过程的做法却遮掩了某些现实风险。倘若只沉湎于外部的繁荣景象而难以察觉风险,不但认罪认罚从宽凸显的效率价值有沦为昙花一现之虞,刑事诉讼的整体均衡性也将面临重大挑战。

(一)“地下暗流”的主导性

所谓的“潜规则”起主导作用,其实早就是中国各个领域中老生常谈的问题了。^[26]而在刑事诉讼系统中,这就如同一条深埋于地下的暗流管道,真正决定了司法权能的走势,实际影响了被追诉人的程序结局。但是在认罪认罚从宽的框架内,这种“地下暗流”的主导力却很容易受忽略。这是因为,此项改革强调的诉讼主体间的积极协商,仿佛本就暗示了鼓励“潜规则”的意向。虽然量刑建议的内容可能为控审主体于规范机制外的渠道所敲定,却满足了简化庭审等程序环节的外部诉求。如果从实用主义的立场切入,既然量刑精准化要以控审机关的共识为基础,那么其实际形成的时间节点或早或晚就无伤大雅。

可问题的关键在于,实用主义立场在此领域并不是适宜的观念基础。这与我们不能赞同刑讯手段获取事实真相在本质上是别无二致的。^[27]或许庭外的私下沟通真的有助于补全诉讼各方的信息盲点,又或许检方的量刑建议本就内嵌了精准意涵而为裁判者事先认可,但是在规范之外的时空环境决定案件结局终究是不妥的。其伤害的是合法诉讼程序的权威性。无论是刑事立法还是司法解释,都为检法两家的单独接触设定了条件及范围;却并未暗许双方可就

[24] 参见翟学伟:《人情、面子与权力的再生产》,北京大学出版社2017年版,第166页。

[25] 参见周新:“论认罪认罚案件量刑建议精准化”,《政治与法律》2021年第1期,第22页。

[26] 参见陈瑞华:“法律程序构建的基本逻辑”,《中国法学》2012年第1期,第71页。

[27] 参见张建伟:《阿Q之死的标本意义》,法律出版社2017年版,第337页。

被告人如何量刑提前达成一致。^[28] 针对量刑建议的评估,是在法庭上需要解决的事项,而不应发生于电话连线中。否则,认罪认罚从宽的法庭审理无异于“走过场”,何不干脆予以取消?易言之,既然被告人已认罪认罚,那么量刑建议的采纳与否反而是庭审应当重点审查的事项,将其置于“地下暗流”有舍本逐末之嫌疑。

除此之外,鼓励这样的私下沟通,更会给正当程序的理念落实带来难以想象的负面损耗。认罪认罚从宽的提出,实际是决策者在资源优化配置思维影响下,对于“以审判为中心”刑事诉讼制度改革的一种有效补充;而后者恰是正当程序理念催生的产物。为了实现刑事司法程序的公正价值,“以审判为中心”主要是围绕庭审实质化来展开设计并落实相应举措的。^[29] 不遑多论,庭审实质化的存续是以排斥自身的流于形式为前提的,故控审权能主体的各司其职至关重要。就目前的发展态势看,庭审实质化的具体样态尚未完全成型且深入人心;反观认罪认罚从宽承载的案件比重却明显占优,意味着私下沟通确认量刑幅度将变相成为主流。同样一批掌控司法权的个体,绝不可能因为案件属性差异而接受迥异的价值观点。一旦认同了“潜规则”,就很难再对明确的规范主义保持敬畏。长此以往,针对庭审实质化的预期值就会持续降低,且日臻覆盖所有类型的案件形态直至付之阙如。我们完全可以将其视作认罪认罚从宽改革衍生的一种“反噬”效应,所付出的牺牲代价即为追求公正的程序价值观。这并非杞人忧天,而早已在司法实践和理论探讨中呈现些许苗头。^[30]

(二)畸形化的效率指标

的确,量刑建议的高采纳率表面上可作为办案效率提升的佐证;然而,在这一语境下的程序运行却未见得符合认罪认罚从宽制度初创时的价值期许。尽管目前量刑建议向量刑裁决的转化进程极少造成办案时间的损耗,却给诉讼效率以外的其他价值实现埋下隐患。如果不能向上述效率指标渗透充分的内省因子,单纯的数量增长只不过反映了一种畸形化的片面发展,其过度膨胀的结果是破坏诉讼体系的总体平衡。这其中,尤以对“先定后审”现象的恣意纵容为甚。

针对庭审实质化的大力提倡,往往是以“案卷笔录中心主义”作为指摘对象的。^[31] 某种程度上,这客观反映了办案人员对于书面审的依赖心理。基于案卷材料的存在,司法权力主体其实更易于接受“不打无准备之仗”的观念偏向。于是,“先定后审”成为一种心照不宣的习惯

[28] 根据《最高人民法院关于适用〈中华人民共和国刑事诉讼法〉的解释》规定,对认罪认罚案件,人民法院经审理认为量刑建议明显不当,或者被告人、辩护人对量刑建议提出异议的,人民检察院可以调整量刑建议。这表明针对量刑建议的妥当与否,应当在其审理后加以评断,而非私下协商决定。此外,尽管根据此一解释,适用速裁程序审理认罪认罚案件,需要调整量刑建议的,应当在庭前或者当庭作出调整,但是作者并不认同速裁程序将量刑建议的审核置于庭审以外的做法。此处的合理解释应为,对于量刑建议异常明显地呈现畸轻畸重,法官在阅卷后才可于庭前要求检察机关作出反馈。

[29] 参见汪海燕:“论刑事庭审实质化”,《中国社会科学》2015年第2期,第104页。

[30] 参见佘化强、余韵洁:“审判中心主义与卷宗制度的前世今生”,《法学家》2020年第6期,第115页。

[31] 参见沈德咏:“论以审判为中心的诉讼制度改革”,《中国法学》2015年第3期,第13页。

定势。“行动由根深蒂固的习惯决定。”^{〔32〕}按照改革蓝图的预设,这种行为惯性本应随着相关从业者的观念进步而逐渐式微;相关思路改变的基础就在于“开庭日”能够成功取代文牍主义的地位,即成为衡量案件信息的主导载体。^{〔33〕}即便大部分案件被分流至简化处置的程序范畴内,也不宜破坏这一基本立场。然而,当前对于繁简分流的过度痴迷导致盲目追逐效率指标,进而反向巩固了办案个体对于书面审的依赖度。于是,“先定后审”再度大行其道,且影响范畴恐怕不逊于1979年刑诉法实施期间。量刑建议的内容没有成为法庭审查的焦点,只能说明针对被告人如何刑罚的“剧本”早已完成。特别是当个案需要以速裁抑或简易程序终结,不仅意味着庭审周期的大幅缩减,更表明现场宣判构成了法官决策的不二选择。基于办案主体时的素能局限,为了确保裁断过程与结论的妥当性,针对量刑建议的提前交涉便水到渠成。可见,大多数情况下庭审活动又沦为一场例行公事的表演,而并未承载决策场域的实质功能。在这里,我们不禁要问:既然认罪认罚从宽的实践探索是以推崇庭审中心逻辑为前提的,却又为何成为“先定后审”的助力驱动?这仿佛形成了一种难以名状的悖论关系。

此外,这些量刑建议的高采纳情形是否真的触发了诉讼效率提升的“加速键”呢?单看量化数据或许会对此有坚定不移的信心;但“评估概率是一种有用和理性的处理不确定性的方式;当新的信息注入时,人们应随之刷新概率的评估;并且,新信息对于人们最终决策的影响则有赖于人们的验前几率”。^{〔34〕}对于司法权而言,衡量资源配置是否合理的指标从不是孤立和有限的。尤其是案件办结的快节奏,更不代表权力运行的内在自洽。藉由作者对司法实务人员的调查情况看,针对量刑建议的提前商定并不等同于其自身工作负荷的减缓。相反,当人情因素夹杂其中,程序平台以外的权力关系更容易呈现不确定性与复杂性,进而促使检法机关投入较多无谓的隐性成本。更不容忽视的是,控辩关系基于此类状况的愈演愈烈在程序地位上已明显弱于控审关系。

(三)控辩关系的持续失衡

在不少人看来,认罪认罚从宽的大规模兑现从根本上变更了中国刑事诉讼制度的模式划分,使得“协商(或者可以称为‘合意’)一对抗”的二元轮廓逐步清晰起来。^{〔35〕}但实际上,学理层面的“协商关系”更多时候指代的本是控辩之间。^{〔36〕}而现实状况却并非如此,且控辩失衡现象在量刑建议引发的权力/权利关系互动中甚至有进一步加剧的风险。这是因为,任何协商模式的建立势必要以参与双方或者多方主体的地位平等为基础。否则,合意的达成就会多少内嵌着强势一方的主导意志。从这个角度看,在检察机关拟定及提出量刑建议的期间内,被迫

〔32〕 郑戈:“韦伯论西方法律的独特性”,载李猛主编:《韦伯:法律与价值》,上海人民出版社2001年版,第65页。

〔33〕 参见孙皓:“司法文牍主义与开庭日——关于刑事办案模式的实验性研究”,《环球法律评论》2018年第2期,第124页。

〔34〕 (美)理查德·A.波斯纳:《证据法的经济分析》,徐昕、徐昀译,中国法制出版社2004年版,第7页。

〔35〕 参见王新清:“合意式刑事诉讼论”,《法学研究》2020年第6期,第150页。

〔36〕 参见陈瑞华:“论协商性的程序正义”,《比较法研究》2020年第6期,第2页。

诉人并未掌握太多话语权特别是争辩机会。当辩护权依旧明显缺位而值班律师又难以填补相应的防御漏洞时,基于信息不对等所造成的被迫接受,几乎不可避免地构成了量刑建议的实质驱动力。即便辩论律师积极参与其中,其担纲的角色也非谈判桌前的协商者,而是提供意见的辅助者。基于相关的法律规范文本,检察机关的“听取意见”其实不过是一种行为姿态,而很难影响诉讼走向的实质结果。^[37]这并不是否认辩护律师在量刑建议形成过程中的建树,即使部分从业者的确主张认罪认罚从宽框架下自身的市场空间受挤压;^[38]但很显然,检察机关在此一过程中享有完全的主动权,可自行裁量辩护方意见的可接受度。在这样的关系模式下,倘若主办案件的检察官一意追求认罪认罚从宽的适用比率,那么处于弱势的被追诉者很难有效维护自身利益,毕竟相关的规范平台本就偏狭。这也是认罪认罚从宽与域外辩诉交易机制的一大本质差异。

对于辩诉交易而言,检察官与嫌疑人之间的交涉是决定程序走向的关键;^[39]而在认罪认罚的现有实践中,控审主体能否对量刑建议排除分歧则占据了核心要津。或许基于立法决策的理想设计,控辩协商真的需要与量刑建议的采纳建立起正相关;但司法实践呈现的镜像,却明显增强了检法之间的协商强度。换言之,目前影响认罪认罚从宽个案结局的,更多时候是控审合意而非控辩协商。而且,当控审主体为保证量刑建议的采纳率而加强了私下沟通时,辩护方的诉讼角色就容易显得进一步模糊化。至此,控辩平衡自然不能成为量刑建议被采纳的前提要件。其危害结果,在于认罪认罚从宽事实上鼓励了将被追诉人置于次要地位的心理倾向。如果再结合某些指标导向,“人”在程序中似乎又成为达到量化业绩的客观手段,而非权利目标本身。

三、误区生成的诱因

这样看来,时下量刑建议所附着的高采纳率,并不意味着刑事诉讼整体效益及质量的提升,反倒还蛰伏着不可小觑的危险因子。诚然,不能否认高采纳率依然是一种值得推崇的量化结果,但至少在作者看来,其形成过程应当体现自然演化的动态规律,以遏制人为干扰造成的片面性。为了使量刑建议的运行重回正轨,务必要厘清哪些因素诱使其陷入规律之外的误区中,进而对症下药以恢复理想层面的可欲性。^[40]

[37] 参见林喜芬:“论量刑建议制度的规范结构与模式——从《刑事诉讼法》到《指导意见》”,《中国刑事法杂志》2020年第1期,第6页。

[38] 作者在开展田野调查时,不少律师表示认罪认罚从宽的大力推广,加之值班律师制度的建立,一定程度上压缩了刑事辩护的市场空间。

[39] 参见(美)安吉娜·J.戴维斯:《专横的正义:美国检察官的权力》,李昌林、陈川陵译,中国法制出版社2012年版,第18页。

[40] 所谓可欲性,指涉了个体与制度之间的一种期待关系。参见金观涛、刘青峰:《中国思想史十讲》(上卷),法律出版社2018年版,第112页。

(一)改革逻辑变换

其实,量刑建议作为新生事物出现在刑事司法活动中,并非伴生于时下的认罪认罚从宽制度,而发轫自十多年前的量刑规范化改革。若要追溯其在实践领域的试点探索,应落脚于2008年左右;而有章可循的文本呈现可归结至2010年由“两高三部”联合会签的《关于规范量刑程序若干问题的意见(试行)》。^[41]作者之所以要强调此一历史渊源,意在彰显量刑建议背后的改革逻辑已然发生变换。如果依旧循着量刑规范化的设计初衷体察当下的制度模型,就不免会产生某些似是而非的偏颇理解,进而影响认罪认罚从宽改革期待的价值取向。对于量刑规范化改革而言,主导性的推动力量来源于法院系统,其肇始立场为纠正量刑法裁量的有失均衡;或者可以说是改造重定罪、轻量刑的决策习惯。为此,改革方案的重点聚集在了实体标准的设计上。^[42]易言之,来自法院的改革决策试图通过量刑标准的指南性规范以遏制裁量权的恣意倾向。在这一过程中,量刑建议构成了维系控审意见的程序纽带。也就是说,量刑建议的存续价值体现于检法之间对实体裁判标准能否作同一化认知。尤其是“估堆法”作为一种计算模式遭到摒弃后,量刑建议可以为法官的内心确信活动供给参照系,以凸显裁量进路的规范化。^[43]由此,针对量刑建议的需求被系属于实体标准;相较之下,检察机关对其并无任何急迫意愿,反倒是法院的期待更多。这就不难理解,为何直到2012年刑诉法修订都未赋予检察机关涉及量刑建议的强制义务。某种程度上,量刑规范化改革内嵌的“实体先于程序”特质弱化了各方在此方面的积极性。

而随着认罪认罚从宽改革的勃兴,关于量刑建议的驱动来源发生了主体上的转换。检察机关作为这一改革的主导,对量刑建议的依赖度骤然提升。这是因为,认罪认罚从宽更加偏重于程序模式的改造,核心立足点反映在了诉讼各方的主观合意上。于是,量刑建议成为串联控辩以及控审关系的重要载体,继而有了不可或缺的存续价值。一方面,控辩之间能否就定罪处刑的减让达成共识,最终会藉由量刑建议的书面内容得以兑现;另一方面,控审之间关于认罪认罚的自愿性及合法性评估,将量刑建议的采纳与否作为结论标志。不难发现,倘若缺失了量刑建议,认罪认罚从宽在个案层面的运行将无法成立。此时,检察机关更希望法院能够接受己方的建议内容,以确保认罪认罚从宽完成程序意义上的闭合。反而是法院对待量刑建议的态度更容易趋于消极:一者,此项改革的责任承担其实无涉自身,多少有事不关己的意味;再者,量刑规范化影响下的思维惯性持续放大了实体把关的心理暗示。用一句话足以贴切形容检法机关看待量刑建议的上述差别,即“攻守之势异也”。

值得注意的是,量刑规范化改革的推动进程始终以实体裁量基准的完善作为催化机制,特

[41] 参见苏彩霞、崔仕绣:“中国量刑规范化改革发展研究——立足域外经验的考察”,《湖北大学学报》(哲学社会科学版)2019年第1期,第126页。

[42] 参见李玉萍:“量刑规范化,十年磨一剑”,载《人民法院报》2018年10月16日,第2版。

[43] 参见赵阳:“从‘估堆’量刑到规范量刑”,载《法制日报》2011年8月15日,第2版。

别是针对迥异案由类型的情状细化。^{〔44〕}从常理判断,随着实体上的量刑指南趋向不断细化,程序维度控辩之间就从宽减让的博弈空间反倒会愈发狭窄。如果认罪认罚从宽的实施进程中并未呈现上述现象,则说明要么量刑规范化的效果不及预期,要么就是控审主体为了特定目标达成而在实操中相互迁就。现在看来,这两种诱因或许都一定程度存在。关于量刑精准的呼吁本就暴露了规范化改革在实体维度上的短板,似乎其业绩无法裨益于认罪认罚从宽的实施。^{〔45〕}另外,当改革逻辑发生背景意义上的变换,而实体与程序在诉求层面尚未完全一致之时,检察机关难免——基于强烈且亟待的“客观需求”——主动迎合法院的量刑规范化立场,确保能够实现自身建议的被采纳率。^{〔46〕}与此同时,控审之间的纵横捭阖构成了认罪认罚从宽能否达致的主流环节,被迫诉方的利益关切在两大改革的交汇耦合中逐渐冲淡。更多时候,检察机关会出于科层顶端的压力传导,不得不率先作出向裁判权靠拢的姿态,这才是前述所谓“客观需求”的真正来源。

(二)外部压力传导

在中国,每项司法改革举措几乎都要具体分包给某个权力部门来承担主责。如前所述,认罪认罚从宽在归口管理上是隶属于检察机关的。早在此项改革的试点阶段,个案适用比重的衡量就被纳入到各级检察机关的业绩考核指标体系中。尤其是对基层单位而言,相关的警戒性要求犹如一柄利剑高悬在头顶,迫使其不得不加以重视。^{〔47〕}不难推测,隐匿于司法绩效考核背后的支配性力量又一如既往地干预了量刑建议的权能运行图式。^{〔48〕}正是在其强劲的压力传导之下,契合决策意志的“暗通款曲”才真正演化为普遍性的程序模式基础。

“来自各方面的经验都显示,纯粹官僚型的行政组织——即一元化领导的官僚制——由技术的观点来看,可能会获得最高的效率。”^{〔49〕}而针对认罪认罚从宽的探索,恰好依托了科层

〔44〕 例如在2017年4月,最高人民法院对《关于常见犯罪的量刑指导意见》作了修改,进一步完善量刑方法,规范故意伤害、抢劫、盗窃、毒品犯罪等十五个罪名的量刑。同时,进一步扩大量刑规范化范围,将当时多发易发、与群众生命财产安全密切相关的危险驾驶、非法集资、信用卡诈骗等八个罪名纳入规范。

〔45〕 其实在作者看来,量刑规范化改革本就存在创设初衷上的疑点。一方面,事无巨细地列举现实中需要予以考量的情节几乎是不可能完成的任务,遑论还要以此为基础设定具象化的量刑幅度。另一方面,即便真的能够创制出此类指南性文本,法官势必会沦为徒有其表的权力工具,陷入机械僵化的思维定式,进而令个案判决丧失经验尺度。

〔46〕 在2021年7月,结合认罪认罚从宽的实施,最高人民法院会同最高人民检察院以通知的形式联合印发了《关于常见犯罪的量刑指导意见(试行)》,共同明确了量刑的指导原则、量刑的基本方法、常见量刑情节的适用和常见犯罪的量刑,将23种常见犯罪判处有期徒刑的案件纳入规范范围,同时规范罚金、缓刑的适用。从中不难看出,对于实体层面的量刑标准,检法之间已呈现趋同化的利益诉求。

〔47〕 举例来说,根据作者调研得知,在某市检察系统,每年年底的考评依据相关业绩将18个区县单位区分为四档,即“4—5—5—4”的基本格局。按照其考核办法,倘若基层检察院在下一年度的认罪认罚从宽适用比率无法企及全市平均水平,将自动无缘前两档,遑论其他成绩如何突出。

〔48〕 参见朱桐辉:“绩效考核与刑事司法环境之辩——G省X县检察院、司法局归来所思”,《刑事法评论》2007年第2期,第273页。

〔49〕 (德)马克斯·韦伯:《支配的类型》,康乐等译,广西师范大学出版社2016年版,第311页。

组织的上令下从之关系范式。为了保证认罪认罚从宽的实施效果,量刑建议要尽可能地被法院照单全收。否则,检察权自不必大费周章与裁判者争执量刑建议的可采性。这期间的症结,在于无论是侦查权抑或审判主体,都不存有必须达成从宽合意的强烈动机;而检察机关则不然,量刑建议的采纳数据是利益攸关的。而颇为吊诡的是,决定这一量化指标的权限掌握在法院手中。“上层官员理解普遍性;下层官员理解具体性。”^[50]对于承办具体案件的检察官而言,不能提升量刑建议的采纳率就意味着被冠之以“办事不力”的负面评价。其结果,自身的政治前途及经济收益都面临着责任归咎后的减灭风险。是故,随着科层指令的逐级传导,理论维度上量刑建议的意涵合理性会被滥觞于体系上端的权威力所遮蔽乃至扭曲。

为了满足认罪认罚从宽的期待效果,既然无法独立决定量刑建议的程序命运,那么寻求来自裁判权的谅解与支持就本能地成为公诉权的不二选择。如此一来,与其说量刑建议代表了控辩之间的合意结果,倒不如承认检法意见的互相吸收乃至迁就,才是认罪认罚从宽显现蓬勃生机的关键。但是,这样的境况与此一改革的初始立场形成了分歧。因为可期待的程序效能提升,本应见诸于被追诉人权利得以充分保障的前提,并且将量刑减让与程序简化分别置于对价关系的两端。遗憾的是,源自科层上端的权力干预未见得遵守方法论维度关于价值中立的要求,而仰赖于主动寻求符合自身目标预设的证据。继而,不切实际的政绩观异化为压倒一切的驱动源泉。^[51]对于被追诉人,检察官完全可以利用信息不对称及双方的实力悬殊关系,强行获取被追诉者的具结表态,同时催生符合自愿性、合法性等形式要求的量刑建议内容。而一旦面对法院的审查,控诉权又能适度放低身段,通过单方面的妥协退让换取量刑建议在程序维度上的“通行证”。显见不争的是,在上述行为逻辑的循环往复中,量刑活动演绎的终极结论未必是最适宜于被追诉人的。

(三) 配合制约原则的投射

毋庸置疑,“分工负责、互相配合、互相制约”的原则立场依旧统辖着刑事诉讼的权力运行格局。^[52]即便认罪认罚从宽的适用比重已占据了绝大多数的案件份额,公检法之间的相互配合、制约关系却从未受到颠覆。相反,正是由于配合制约原则强有力的影响,量刑建议的提出及接受才能呈现如今这般的居高不下。当前量刑建议的可采性程度很难充分映衬刑事司法的精密化指向,倒可以视作配合制约原则在认罪认罚从宽领域内投射。这一原则指引下的阶段论之程序构造,本就强调了公权力的关系模式,而对其余权利个体是持漠视态度的。

见诸于自身的利益动机,检察机关当然期望法院能将采纳量刑建议的做法固化为常态。但是在流水作业式的诉讼体系中,法院身处程序运转的最后环节,所担负的责任是最为关键

[50] (美)米尔伊安·R.达玛什卡:《司法和国家权力的多种面孔:比较视野中的法律程序》,郑戈译,中国政法大学出版社2015年版,第27页。

[51] 参见(法)阿兰·佩雷菲特:《官僚主义的弊害》,孟鞠如、李直、吕彤邻译,东方出版社2014年版,第390页。

[52] 参见韩大元、于文豪:“法院、检察院和公安机关的宪法关系”,《法学研究》2011年第3期,第3页。

的。故而,“以审判为中心”的权力立场才能具备不容置喙的正当性。^[53]进一步论,法院需要对检察机关在审查起诉阶段形成的裁判倾向进行审查把关,由此彰显了自身具备的制约职责。这也解释了,各级法院为何普遍侧重于理解阐述所谓“一般应当采纳”条款中的“一般”,而非“应当”。无论量刑建议的形成过程如何体现司法资源的优化配置,或者怎样充分代表了被迫诉人的自愿态度,都不构成法院放宽审查把关标准的合理根据。基于制约义务的职责所系,法院必须站在自身秉持的观点立场上,对量刑建议的适当性作全方位评估。只要同裁判者的惯有立场相左,不管其形塑进程多么合乎认罪认罚从宽的改革要领,都难免会遭到否定性排斥。

为了避免制约原则影响下的检法关系导向扯皮推诿,对于配合理念的贯彻无疑起到了缓解、中和效应。由此不难看出,配合制约原则天然具备二元属性,且其内部关系的此消彼长主导着刑事诉讼的动态平衡。具象化到量刑建议的问题上,为了防止法院过度仰赖制约手段,公诉权势必要在前置的时序阶段竭力履行配合义务。因而,双方关于量刑幅度的提前沟通即使出离于程序轮廓,却多少算得上情有可原。毕竟,相互之间的私下通气可被视为加强协作且表达尊重的信号。其实,目前认罪认罚从宽在推进过程中呈现的控审互动,本质上不过是这种配合原则的一项变种形式,尚难称作司法合意的达成。更何况,辩护权的缺位更顺应了配合制约原则的运行机理,而非预期的繁简分流体系需要倚仗的“讨价还价”。鉴于法院掌握了审判阶段的终极决策权限,检察机关意欲确保量刑建议的可采性,就势必诉诸相互配合的元规则;即便此类预先接触的结果将导致建议尺度的调整,也是优于由此产生冲突乃至损及相关量化指标的最佳选项。

四、回归合理形态的关键

经过前述一系列的调查及论证,作者试图阐释这样的观点:时下的量刑建议采纳率尽管已经居高不下,却不能为认罪认罚从宽机制——乃至刑事司法权能——的高效运转起到充分背书作用。相反,一味追逐高采纳的结果只会模糊外界对于量刑建议的准确认知,尤其可能造成诉讼结构的持续失衡状态。那么,究竟怎样才能使量刑建议的程序功能回归至合理形态呢?在这里,作者恐怕尚不能给出相对微观的完善进路,但是从避免进一步陷入错误的角度,却可提供些许有助于正本清源的意见或思路。

(一) 正视检法之间的制约关系

首先需明确的是,法院提出不同于检察机关的量刑意见在当前的司法语境下不仅是正常的,甚至可以说是必须且亟待的。如果公诉权的量刑倾向被裁判者不加甄别地予以吸收,即便提升了诉讼效率,也会给司法公正价值的实现制造了巨大隐患,进而诱发得不偿失的“副作用”。殊不知,近年来中国司法改革的原始动机是最大限度消灭冤错案件,确保公平正义渗透到具体个例中。这一点具有无可超越的优位属性,无论认罪认罚从宽激生的效率提升作用如

[53] 参见陈光中、魏晓娜:“推进以审判为中心的诉讼制度改革”,《中国法律》2015年第1期,第9页。

何突出,也难与之相提并论。^[54]正如前文所述,配合制约原则构成了中国刑事司法权顺畅推进的源动力,由此形成的层控式特质左右了不同权力主体的关系模式。^[55]即使认罪认罚从宽的渗入导致大部分案件的处理方式变化,也未根本触动“配合”与“制约”这两大程序行为主题。其实,配合制约关系的存续本就内嵌了效率与公正的价值诉求,至少从理论设计上是可将此作为周延性来源的。一方面,不同机关的配合协助促成各司其职的理想样态,强化了案件在流转衔接上的紧密度;另一方面,诉讼程序的层层推进有赖于逐级审查把关的制约效应,以竭力防止个别化偏见演变为错谬裁判。对于法院而言,其掌控的审判权可谓处在最后一道制约“屏障”,显然是不容有失的。反之,检察权针对法院的制约权力同样可以起到积极作用,特别是进一步检视量刑裁决的妥当性。在这样的相互制约关系中,被追诉人的命运或许不至于被检法中的任何一方完全操纵。

而随着司法责任制体系的不断成型,这种制约关系的存在更显得弥足珍贵。毫无疑问,当员额制以及权力清单的产生满足了权责一致化趋向,检察官在个案中的话语权自然与日俱增;认罪认罚从宽的广泛适用则进一步强化了上述态势。这表明,员额检察官事实上将成为量刑尺度的真正决策者。一个国家刑事司法的风格很大程度上是取决于该国司法人员素质的。^[56]显然,我国的司法权力单元尚难以同责任制改革的预期目标完全匹配。由此,针对权力释放的制约与监督尤为关键。否则,检察官在指标体系的压力下很容易落入恣意滥权的惯性,从而迫使被追诉人违心接受量刑建议的内容,甚至连认罪认罚从宽的达成都不免是“城下之盟”。对此,指望检察权的自我约束无疑是一种自我安慰,所谓的量刑精准化也不会有助于拓宽控辩之间的合意空间,唯有来自法院的外部审查才堪当实质有效的屏障。至少在当前的司法格局中,法院就量刑建议提出不同意见并非坏事;检察机关完全没有必要将前者的采纳态度视作自身工作质量的评价基准。或许域外的检察权基于便宜主义立场正愈发呈现权力扩张的性格特征,且越来越多地独立决策被追诉人的刑罚后果,^[57]但服膺于本土司法构造及资源配置现实却是变革的前置要件。故此,在权力主体间保持配合制约关系的主旋律下,量刑建议的约束效力其实不具备人们预期的讨论价值,貌似联结了实然与应然性的高采纳率现状更缺失了必然性基础。

(二)从“刻意”到“真意”

一言以蔽之,目下司法实践中呈现的量刑建议高采纳率带有相当程度的欺骗性,本质上是控审主体之间相互迁就、忍让的结果,并不能真实反映认罪认罚从宽在个案中的合意属性。当然,这不是说量刑建议的采纳率处在低位状态才合乎情理。至少,检法之间在情节认定、法律

[54] 参见陈卫东:“新一轮司法改革的重点与展望”,《中国法律》2015年第1期,第14页。

[55] 参见李奋飞:“我国刑事诉讼制度持续发展因子探析”,《法商研究》2016年第5期,第110页。

[56] 参见(德)托马斯·魏根特:《德国刑事诉讼法原理》,江溯等译,中国法制出版社2021年版,第130页。

[57] 参见(瑞)古尔蒂斯·里恩:《美国和欧洲的检察官——瑞士、法国和德国的比较分析》,王新玥、陈涛等译,法律出版社2019年版,第227—228页。

适用等方面的认识分歧应当被最大限度地展示与接纳。控审权力分别代表着截然不同的观点立场,绝不可能在短期内迅速形成一致性看法。这意味着,即便各方没有部门利益上的私心,关于量刑建议的争执势必是客观存在的。由此,检察机关刻意追逐量刑建议的采纳比重不但显得多此一举,也有强人所难之嫌。正是这种心态上的偏颇导致了认罪认罚从宽的适用被检法关系所绑定,而过于在意表象的政绩观驱使难免束缚了公诉权能的决策自由。从根本上,检察机关内部的科层僚属体系不宜将量刑建议的采纳情况作为一项显性衡量指标。其结果,就是具体负责办案的检察官为实现形式的目标,而不得已强行向法院的观点靠拢。

在上述靠拢行为的具体表现形式上,私下的沟通协调成为无法回避的潜规则。如此一来,司法权配置的非人格化特征受到削弱,个人关系更容易渗入量刑裁断过程。在作者看来,即使量刑建议未被法院采纳,也不是什么不堪设想的后果。这其实只是心理态度问题。检法之间应当尽可能减少程序框架外的实体交涉,以保证庭审作为决策场域的存续意义。这不仅关乎刑事司法的结构性稳定,更会影响辩护权的覆盖范畴。通过架空法庭审理这一正当程序的方式催生量刑建议的高采纳率,只能令权力关系陷入紊乱与畸形状态。对此,釜底抽薪的做法是摒弃意图炫耀此一量化指标的通例,真正还权于具象化的司法办案单元,使后者凭借自身的专业认知及内心确信进行量刑决策的博弈。果真如此,畸高的数据指标或许会有所回落,却足以印证认罪认罚从宽的被接受度。

任何量化指标都不免因外部因素的变化呈现不时波动,关键在于评估者能否秉持客观中立的科学精神。“权力说到底并不是一个物质性的事实,而是一个使人们肯服从的舆论状态。”^[58]真正的高采纳率绝不是刻意为之的产物,而应当归属于控审之间长期磨合的结果,是一种“真意”的本能反映。换言之,法院采纳检察机关的量刑建议,应当源自双方在法律观点上的自然契合;而此种一致性系二者于较长周期内反复互动方可生成。这样的结果必然历经漫长回合的交锋,藉由熟能生巧的量刑累积而得以巩固,且基本上无捷径可言。虽然很多人均寄希望于精确化的量刑指南来实现“短平快”的检法合意,但其落实难度之大是断难忽视的。规范层面囊括全部量刑情节是不切实际的奢望,而将个体裁量的向度限制在完全由文本设定的刑罚选项内更难免流于荒诞。形象比喻之下,未来关于量刑建议的普遍采信应当出自检法之间的“自由恋爱”,而非强制性的“包办婚姻”。归根结底,理想的高采纳率只能是一种经验范式的外化样态。

(三)改良适宜的决策平台

尽管量刑建议的提出与认罪认罚从宽的适用之间——至少从法律规范层面上看——并非是绝对的一一对应;但不可否认的是,在大部分个案中,二者确实存在着共生关系。故而,裁判权针对量刑建议的考量,还要立足于认罪认罚从宽的背景因素。既然前文中已提及高采纳率的达成有赖于检法之间的默契度培育,那么这样的理想图式又是如何形塑而成的呢?在作者

[58] (英)弗里德里希·奥古斯特·冯·哈耶克:《自由宪章》,杨玉生、冯兴元等译,中国社会科学出版社2016年版,第276页。

看来,双方就量刑决策的观念靠拢应该是基于专业主义立场的,而绝对排斥部门利益或者私人关系的竞合。着眼于目前的规范文本预期,认罪认罚从宽案件的审查焦点聚集在了被追诉人的自愿性以及程序运作的合法性上,并且试图将法院的职责导向确认模式。^[59]正因为如此,检法之间生出嫌隙的盖然性放大,二者的关系好坏就异化为左右量刑活动走向的要旨。基于时下的司法环境,仅围绕自愿性与合法性来确保认罪认罚从宽的成立与否,尚不足以保障效率价值实现的必要条件,即案件办理质量的可靠性。故而,此类案件的审查对象应当着重补充一项子要素——量刑建议的适当性。易言之,当法院接收了检察机关移送的、符合认罪认罚从宽要求的个案后,随即应利用庭审这一事实调查平台按照如下步骤开展审核工作。

首先,裁判者要借助口头交流以确认被追诉人知晓相应的法律后果,继而防范认罪认罚的表态源于强迫或者欺骗。其次,以具结书的内容为考察对象,分别从形式与实质角度辨识控辩之间合意达成的真实性与合法性,避免逾越法律条件的程序简化与刑罚从宽。最后,法院需要重点审视量刑建议的适当性,即利用庭审机制的原有框架,具体结合案情事实及证据现状,在认可犯罪性质的前提下决策量刑建议的合理性与可采性。即,针对量刑建议的把关与审查将转换为法庭调查或辩论的重点问题,遑论程序模式的选择为简易程序抑或者速裁程序。这意味着,即使法院适用的审判程序被化约至极简,关于量刑建议的审查也要依口头质证及辩论的原则加以展开。所有的决策活动都不能游离于公开程序之外,追求资源配置最优的认罪认罚从宽更须建立在庭审实质化的基础上,而不是绕开正当渠道去另起炉灶。同时,当量刑建议的采纳与否完全付诸于公开庭审的模式,私下预先的沟通就会最大限度降低。当然,这要求检法机关都将庭审视为真正解决问题的场所,而不是一个用于满足形式主义要求的表演空间。^[60]此外,一旦针对量刑建议的评价被限于专门场域,辩护权就获得了扩充话语权的契机。由此,控辩审之间的良性互动便可给予足够期待,检法关系对于最终量刑决策的影响会趋向弱勢。其结果,从量刑建议的提出到采纳均析出自诉讼各方主体的真实意志及相应的专业判断,而将其他干扰变量逐步剔除。在此基础上形成的高采纳率,才是客观佐证司法权顺畅流转的有力根据。

五、余论:关于未来的两种可能

“发展是一个发现的过程,不是一个建筑的过程。”^[61]如果不能对目前超高比率的量刑建议采纳有着客观清晰的认知,就不免陷入盲目乐观的误区,而忽略了隐含于制度运行环节的某些结构性障碍。伴随着认罪认罚从宽不断介入现有刑事诉讼格局的进程,未来的程序机理可

[59] 参见李奋飞:“论‘确认式庭审’——以认罪认罚从宽制度的入法为契机”,《国家检察官学院学报》2020年第3期,第39页。

[60] 参见孙皓,见前注[14],第122页。

[61] 徐爱国:《法学的圣殿:西方法律思想与法学流派》,中国法制出版社2016年版,第456页。

能会呈现两种较为危险的倾向,值得予以更多关注与警醒。一者,由于繁简分流态势的初步体现,大部分案件都不再依托正当程序机制,事实上令“先定后审”现象恢复了普遍性。然而,庭审实质化却并未步入精密化轨道,原有的“走过场”状况依旧根深蒂固。于是,“以审判为中心”的目标期望日益被认罪认罚从宽的“大水漫灌”所挤压乃至吞噬。二者,因为配合制约原则对于刑事诉讼构造的深远影响,控辩之间的合意状态始终被置于次要位置,而公权力主体的相互谅解或者扯皮仍然是程序运行的主基调。其结果,认罪认罚从宽的实质内核逐渐被上述权力关系模型架空,进而演变成为一种僵化固定的诉讼手续。在此一过程中,辩护权会呈现日益萎缩的边缘样貌,而被追诉人不免陷于客体化境地,直至非主动接受公检法的程序安排及实体认定。这样一来,认罪认罚从宽与理想状态就有着较大差异,抑或说是落入名存实亡的危险境地。当然,或许两种情形未来均未发生;又或者,最糟糕的境遇便是二者都难以避免。无论如何,过于自信与忽视风险的态度都是不足取的。只要有问题的苗头滋生,就须及时跳脱出认识误区且付诸相应的对策。由此不难发现,探讨量刑建议的高采纳率虽然比较微观,却称得上是举足轻重的关键命题。

Abstract: Sentencing recommendation has gradually become the link between all parties in criminal proceedings, especially under the system of pleading guilty and accepting punishments with leniency. Whether the sentencing recommendation generated from prosecution can be accepted by the court attracts the attention of all related litigation participants. The current “high adoption rate” may meet the expectation of mechanism reform depends on its internal derivative process. In this regard, various empirical investigation helps us to form objective and accurate cognition to prevent blind optimism. Based on the current situation, the large number of court’s adoption of sentencing recommendation is not entirely out of professional recognition, but more or less affected by the relationship between Prosecution and Judge. The principle of coordination and restriction in criminal procedure determines the practical effect of sentencing recommendation. Only when any misunderstandings were eliminated, it would satisfy the balance between justice and efficiency. At the same time, the criminal justice process will be maintained smoothly in the consensus of all parties involved. This ideal situation is obviously difficult to achieve in the short term, but we should rely on the continuous advancement of system construction.

Key Words: Sentencing Recommendation; Pleading Guilty and Accepting Punishments with Leniency; Relationship between Prosecutors and Judges

(责任编辑:车 浩)