

廉政学研究 2019 年第 1 辑（总第 3 辑）

第 69 ~ 91 页

© SSAP, 2019

# 腐败现状大样本评估的三个维度

——基于 2016 年 8133 件贪污贿赂案件起诉书的  
MFT 模型分析

金鸿浩\*

**摘要：**腐败测量是廉政学研究的重要基础。随着司法文书大样本数据的公开，以及数据分析工具的快速升级，利用客观分析法，对我国腐败现状的大样本评估成为可能。如何构建新型腐败评估的框架，成为一个值得研究的前沿问题。基于 2016 年公开的 8133 件贪污贿赂案件（10711 人）起诉书的大数据分析，本文探索建立了金额、次数、时间的三维度分析模型，对我国腐败现状及其相关因素进行了一次横断面研究，有助于标本兼治，服务“全面从严治党”的精准决策。

**关键词：**腐败测量；大数据分析；数据画像；贪污贿赂

## 一 绪论

腐败现状的评估测量是廉政学研究的基础性工作，也是科学推动

---

\* 作者简介：金鸿浩，1990 年生，国家检察官学院讲师，清华大学法学院博士后、助理研究员，主要研究方向为腐败预防、网络治理、智慧司法。

“全面从严治党”向纵深发展的重要前提。对当前“夺取反腐败斗争压倒性胜利”工作任务而言,在新的起点上深入推进反腐败斗争,首先要准确、全面、科学地评估我国当前腐败犯罪的总体状况,如果不清楚我国腐败犯罪的地域分布、罪名分布,不掌握哪些犯罪主体特征会影响犯罪行为 and 犯罪结果,那么反腐败仍然属于粗放式、经验式反腐范畴,无法科学配置反腐败资源,实现标本兼治和高度廉洁。正如习近平总书记所说,“没有危机感和紧迫感,看不到问题和症结所在,那危险就不远了”,“要对症下药,亡羊补牢,未为晚矣”。

当前,学术界对腐败现状评估主要可以分为客观测量法、主观评价法两类。客观测量法的优点是根据统计数据和指标分析,可以较为精确地掌握腐败惩治现状,但由于腐败行为的隐秘性和惩治腐败的滞后性,该方法无法测量腐败增量和腐败黑数。主观评价法可以帮助执政者、研究者实时掌握公众或特定对象的腐败感知程度,但是评价的主观性较强,受抽样方式和样本差异的影响较大。正如法国巴黎政策研究院伊夫·梅尼教授所说的,“对腐败的真实或假定程度的判断更多的是凭感觉、感受,而不是对腐败现象做精确的测度”(伊夫·梅尼,1997)。

随着公开信息渠道的增加,客观测量法也衍生出三种不同评估测量方式。

第一种是案件统计法,数据主要来自中央纪委、国家监察委员会、最高人民检察院和地方纪检监察机关、检察机关的工作报告和年鉴,也有部分数据来自审判机关、国务院新闻办、原国家预防腐败局和审计署。早期廉政评估研究中,客观测量多采取该类方法。例如,何增科(1999,2003)根据历年《最高人民检察院工作报告》中检察机关查处贪污贿赂等腐败案件总数进行分析,发现转轨时期的制度结构和制度安排缺失是腐败行为流行的原因之一,依靠体制改革和制度创新的反腐新路正在形成。王传利(2001)对1988~1999年全国纪检监察机关处分干部情况和1990~1997年检察机关查处的若干类案件情况进行分析,发现腐败频度存在周期波动,20世纪90年代腐败现象日益严重,腐败

的扩展期增长，衰减期缩短。倪星、王立京（2013）进行了类似分析，将1980~2002年的中国腐败波动分为四个周期，并对腐败黑数和经济后果进行估测，推测1980~2002年腐败总经济损失为5000亿元，平均每年250亿元。郭剑鸣（2009）对2008年国家审计署审计报告进行分析，发现存在违规的部门占受审部门数的61.2%，违规资金达50.7亿元，认为预算失规现象一定程度上孵化了腐败行为。罗先觉（2010）、盛宏文和魏娜（2013）、赵秉志（2014）、刘冠君和朱立恒（2015）等也使用纪检机关、检察机关统计数据做了类似研究。

第二种是基于媒体报道的腐败案件指标统计。媒体对中高级领导干部中已经发现的腐败问题，伴随着纪检监察部门、司法机关的立案、查处、判决，多有详细报道或深度调查。该方法最早由过勇（2003）提出，其基于《检察日报》和人民网报道，搜集了1978~2002年82个省部级贪腐要案，研究发现20世纪80年代末90年代初是腐败高发期，20世纪90年代后腐败犯罪潜伏期增长，该阶段的高官腐败案件主要通过其他案件牵连而暴露。过勇（2006）使用类似方法进一步研究，基于媒体报道建立了经济转轨时期（1978~2004年）的省部级、厅级594个典型案例，证明自由化、分权化、私有化、全球化等的经济转轨一定程度滋长了腐败，贷款审批、干部提拔、工程建设成为主要的腐败类型。公婷等（2012）基于《检察日报》2000~2009年公开报道的2802个腐败案例，发现政府采购、工程承包、土地流转、干部提拔等领域成为腐败重灾区；中高级干部是受贿案高发群体，中低级干部是贪污案高发群体。乔德福（2013）对1983~2012年地级市“一把手”岗位被查处的142个腐败案例进行分析，发现涉嫌受贿罪97人、巨额财产来源不明罪22人，平均潜伏期10.5年，任前腐败的占40%。何家弘、徐月笛（2016）统计了十八大以来的中纪委已调查涉嫌刑事犯罪的93名省部级以上官员，研究发现涉嫌受贿罪的77人，主要集中在工程建设、土地开发和房地产建设、矿产开发和人事腐败等领域，不同程度存在官商、帮派、家族腐败利益链。

第三种是基于司法文书的腐败案件指标统计。2013年开始,法院、检察院等司法机关加大了案件文书公开力度,基于公开文书的大数据统计分析成为可能。刘启君(2013)统计了1978~2012年11752份贪污受贿判决书,研究发现犯罪人员年龄均值40.9岁,男性占89.2%,腐败交易次数平均3.55次/人,腐败均值26.8786万元、中位数6000元,我国腐败交易活动总体呈逐渐上升之势,国有企业腐败最为严重,省部级以上官员腐败交易强度最高,正职官员腐败程度普遍比副职官员严重,与职级高低不完全对应。褚红丽、魏建(2016)搜集了2014年法院公开的受贿罪一审判决文书3743份,研究发现样本平均受贿金额为35.6万元,86.48%发生在城市,53.22%有自首情节,9.86%有立功情节,90.3%有退赃行为,7.13%有索贿情节,8.9%属于共同犯罪,东部地区比中西部地区判决更轻,但惩罚没有显著的城乡差异。李辉(2017)统计了2014~2015年7315份裁判文书,比较了贪污和受贿两种腐败类型在地区、领域、级别、职位等方面的系统差异,研究发现东部沿海地区受贿案占比较高,中西部和东北部地区贪污案占比较高,金融领域、房地产和城市规划领域涉案金额较为突出。李辉(2018)利用上述样本进一步研究发现,贪污贿赂犯罪金额存在结构化特征,东部经济发达地区的涉案金额要高于中西部地区,腐败官员的行政级别越高,平均涉案金额也越高。崔晶晶和邓晓梅(2014)、申纯和胡桑(2017)、文宏和杜菲菲(2017)等分别应用裁判文书数据对工程领域、科研领域、乡镇管理的腐败现象进行了统计。

近年来,基于司法机关公开文书的腐败案件指标统计的研究逐步增多,但是分析维度大多局限在腐败犯罪金额方面,集中在腐败犯罪金额的描述统计、相关分析等方面。但一方面,腐败犯罪金额只是其中一个维度,案例数据中本身具备十分丰富的潜在指标有待挖掘和评估;另一方面,腐败犯罪金额也受到其他指标的影响,单独统计腐败犯罪金额指标,比较不同类型对象的腐败金额,客观性、准确性不强。笔者希冀于基于现有公开文书的体例和通常内容,探索一个多维度的大样本评估框

架，从而为我国当前腐败犯罪进行“数据画像”。

## 二 本文的数据工作

笔者对2016年全国检察机关公开的贪污贿赂罪（主要针对受贿罪、贪污罪、挪用公款罪、介绍贿赂罪、利用影响力受贿罪等罪名）起诉书进行实证分析。主要基于Python程序配置网络爬虫软件，对“人民检察院案件信息公开网”的“法律文书公开”栏目中“起诉书”板块（<http://www.ajxxgk.jcy.cn/html/zjxflws/>）同时满足下述三个条件的文书进行爬取：①文件类别为公开文书中的起诉书；②起诉罪名为《刑法》分则第八章贪污贿赂罪的贪污罪、挪用公款罪、受贿罪、利用影响力受贿罪、介绍贿赂罪、巨额财产来源不明罪、私分国有资产罪等（本文没有统计行贿罪和单位犯罪）；③公开时间为2016年1月1日至2016年12月31日。

通过上述方式共搜集到2016年人民检察院案件信息公开网公开的贪污贿赂罪起诉书9588份，通过计算机辅助和人工核对，剔除与贪污贿赂罪无关的文书，共筛查出文书8133份，涉及10711人。根据《2017年中国统计年鉴》数据，2016年我国人民检察院决定起诉贪污贿赂案件为22202件，涉及29640人。本研究采纳的公开起诉书为8133份，涉及10711人，占到当年贪污贿赂犯罪全部起诉书数量的36.63%、起诉人数的36.14%。本文也是现有研究中，起诉书样本的年均数量最大的腐败实证研究。

研究样本数据保存在8133个文本中，总计1283.34万字。传统的人力无法完成编码过程，笔者首先通过计算机软件，对文本要素的特征字段进行信息抽取。在技术原理上，主要综合文本检索、正则表达式匹配、朴素贝叶斯算法等方法，由计算机对文书中所包含的特征字段信息进行抽取，同时通过命名实体识别和实体关系抽取等技术，完成了对所有案例信息点的抽取。例如，提取移送审查起诉时间的正则表达式为：

pattern1 = r"\d + 年\d \* 月\d \* 日\w \* [ (移送| 进入) 审查| 告知],  
pattern2 = r"同年\d \* 月\d \* 日\w \* (移送| 进入| 审查| 告知)。

机器筛查的初次结果并不理想, 准确率在 50% ~ 70%。为了解决这一问题, 笔者和编程团队采取了机器学习的方式, 先对 1200 份起诉书进行机器学习, 使机器筛查准确率逐步提升。先后由编程团队负责对上述起诉书进行了 7 次筛查, 经人工抽验, 准确率逐步提升, 最终准确率在 85% 左右, 此时进一步通过机器学习提升准确率的边际成本较高, 有一些问题, 机器很难准确识别。笔者又组建了人工复核小组, 对照文书爬取的文本和机器筛选的 CSV 表格 (后转化为 EXCEL 表格) 进行复核。

在明确编码过程与标准后, 笔者根据人工校验的数据, 录入 Stata 软件完成统计工作, 并借助大数据可视化工具从多维度展现和分析数据。

### 三 M (Money): 腐败现状评估的金额维度

#### (一) 分罪名犯罪金额分布的描述统计

在 2016 年公开的贪污贿赂罪被告人 10711 人中, 有 10611 人具有犯罪金额信息, 占总样本的 99.07%。其中, 贪污、受贿、挪用公款三项人数较多的罪名合计 77.2799 亿元人民币。

具体情况如下。①贪污罪有 5311 人已知明确的犯罪金额, 合计 14.3642 亿元。最小值为 0 (多为贪污犯罪的共犯), 最大值为 3884.37 万元。人均贪污金额 27.0461 万元, 中位数为 5.5 万元。②受贿罪有 4429 人可知明确的犯罪金额, 犯罪金额总计 32.2745 亿元。最小值为 2000 元 (受贿犯罪的共犯), 最大值为 1.1432 亿元。人均受贿 72.85 万元, 中位数为 18 万元。③挪用公款罪有 1502 人可知明确的犯罪金额, 犯罪金额总计 30.6411 亿元。最小值为 3000 元 (挪用公款罪的共

犯), 最大值为 1.3 亿元。人均挪用公款 204.00 万元, 中位数为 49 万元, 众数为 10 万元。

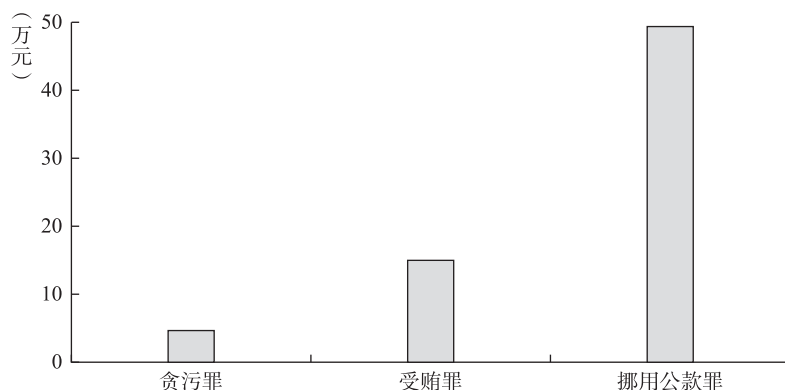


图 1 2016 年贪污罪、受贿罪、挪用公款罪公开起诉书的犯罪金额中位数

## (二) 分罪名犯罪金额分布的回归分析

模型主要考察了犯罪主体因素 (性别、初犯年龄、学历、职务、工作单位性质)、犯罪时空因素 (犯罪持续时间、异省犯罪、区域差异、城乡差异、地域经济差异)、犯罪其他特征因素 (犯罪次数、犯罪人数) 分布对受贿金额、贪污金额、挪用公款金额的影响, 模型在 0.01 水平显著, 调整后的  $R^2$  分别为 0.162、0.292、0.077。

表 1 受贿犯罪金额对数值的 OLS 回归分析

	模型 (1a) ln 受贿金额	模型 (1b) ln 贪污金额	模型 (1c) ln 挪用公款金额
女性	0.074 (0.760)	-0.000 (-0.003)	0.313 ** (2.201)
初犯年龄	0.016 *** (4.885)	-0.015 *** (-4.547)	-0.002 (-0.242)
高等教育	0.458 *** (6.100)	0.400 *** (6.427)	-0.219 * (-1.778)

续表

	模型 (1a) ln 受贿金额	模型 (1b) ln 贪污金额	模型 (1c) ln 挪用公款金额
职务	-0.008 (-0.173)	-0.094* (-1.729)	0.160 (1.355)
党政机关	0.085* (1.861)	0.180** (2.268)	-0.255* (-1.680)
ln 犯罪持续时间	0.362*** (10.096)	-0.019 (-0.500)	-0.019 (-0.215)
异省	0.283*** (3.343)	0.315** (2.161)	0.131 (0.398)
东部	-0.104* (-1.928)	0.094 (1.414)	0.008 (0.062)
县域	-0.335*** (-7.252)	-0.559*** (-9.080)	-0.056 (-0.445)
ln 所在地级市人均 GDP	0.160*** (3.094)	0.108** (2.126)	0.185 (1.504)
ln 犯罪次数	0.177*** (7.005)	0.267*** (9.396)	0.352*** (5.738)
犯罪人数	-0.144*** (-4.075)	-0.178*** (-14.324)	0.142* (1.745)
_cons	-0.355 (-0.601)	1.720*** (2.962)	1.560 (1.121)
N	2664	2120	578
Adj. R - Square	0.162	0.292	0.077

括号中的数字是稳健的标准误差,\*\*\* $p < 0.01$ ,\*\* $p < 0.05$ ,\* $p < 0.1$ 。

根据模型 (1a)、模型 (1b)、模型 (1c) 可以得出如下初步结论。

第一,性别对受贿罪、贪污罪金额没有显著影响,但是对挪用公款犯罪有显著影响。回归结果显示:在控制其他变量的前提下,女性犯罪嫌疑人相比男性犯罪嫌疑人,挪用公款金额会多 31.3% (在 0.05 水平显著)。

第二,初犯年龄越大,受贿罪和贪污罪犯罪金额越大。回归结果显



示：在控制其他变量的前提下，初次犯罪年龄每增加1岁，受贿犯罪金额就会上升1.6%（在0.01水平显著）。在控制其他变量的前提下，初次犯罪年龄每增加1岁，贪污犯罪金额就会下降1.5%（在0.01水平显著）。

第三，学历越高，受贿犯罪和贪污犯罪金额越大，挪用公款犯罪金额越小。回归结果显示：在控制其他变量的前提下，受过高等教育的犯罪嫌疑人相比没有受过高等教育的，受贿犯罪金额会多45.8%（在0.01水平显著），贪污犯罪金额会多40.0%（在0.01水平显著），挪用公款金额会少21.9%（在0.1水平显著）。

第四，具有职务的犯罪嫌疑人，贪污犯罪金额相对较少，受贿、挪用公款犯罪没有显著影响。回归结果显示：在控制其他变量的前提下，有职务的犯罪嫌疑人相比没有职务的，贪污犯罪金额会少9.4%（在0.1水平显著）。

第五，党政机关的犯罪嫌疑人，受贿犯罪、贪污犯罪金额相对较多，挪用公款犯罪金额相对较少。回归结果显示：在控制其他变量的前提下，在党政机关工作的犯罪嫌疑人相比在其他部门工作的犯罪嫌疑人，受贿金额会多8.5%（在0.05水平显著），贪污犯罪金额会多18.0%（在0.05水平显著），挪用公款金额会少25.5%（在0.1水平显著）。

第六，犯罪持续时间越长，受贿犯罪金额越多，但对贪污罪、挪用公款犯罪没有显著影响。回归结果显示：在控制其他变量的前提下，犯罪持续时间每延续1%，受贿犯罪金额会多0.4%（在0.01水平显著）。

第七，异省犯罪嫌疑人，受贿、贪污犯罪金额相对较多，但对挪用公款犯罪没有显著影响。回归结果显示：在控制其他变量的前提下，异省犯罪嫌疑人相比本省籍贯，受贿犯罪金额会多28.3%（在0.01水平显著），贪污犯罪金额会多31.5%（在0.05水平显著）。

第八，在我国东部地区工作的犯罪嫌疑人，受贿犯罪金额相对较少，但是贪污犯罪、挪用公款犯罪没有显著变化。回归结果显示：在控制其他变量的前提下，在东部地区工作的犯罪嫌疑人相比中西部地区，

受贿犯罪金额会少 10.4% (在 0.1 水平显著)。

第九, 在县域工作的犯罪嫌疑人, 受贿犯罪、贪污犯罪金额相对较少, 但是对挪用公款犯罪没有显著影响。回归结果显示: 在控制其他变量的前提下, 在县域工作的犯罪嫌疑人相比在其他行政区域工作的犯罪嫌疑人, 受贿犯罪金额会少 33.5% (在 0.01 水平显著), 贪污犯罪金额会少 55.9% (在 0.01 水平显著)。

第十, 所在地级市人均 GDP 越高, 受贿犯罪、贪污金额越多, 但是对挪用公款犯罪没有显著影响。回归结果显示: 在控制其他变量的前提下, 所在地级市人均 GDP 每增加 1%, 受贿犯罪金额就会上涨 0.2% (在 0.01 水平显著), 贪污犯罪金额就会上涨 0.1% (在 0.05 水平显著)。

第十一, 犯罪次数越多, 受贿犯罪、贪污犯罪、挪用公款犯罪金额越多。回归结果显示: 在控制其他变量的前提下, 犯罪次数每增加 1%, 受贿犯罪金额就会上涨 0.2% (在 0.01 水平显著), 贪污犯罪金额就会上涨 0.3% (在 0.01 水平显著), 挪用公款犯罪金额就会上涨 0.4% (在 0.01 水平显著)。

第十二, 犯罪人数越多, 受贿犯罪、贪污犯罪金额越少, 挪用公款犯罪金额越多。回归结果显示: 在控制其他变量的前提下, 犯罪人数每增加 1 人, 受贿犯罪金额就会下降 14.4% (在 0.01 水平显著), 贪污犯罪金额就会下降 17.8% (在 0.01 水平显著), 挪用公款犯罪金额会上升 14.2% (在 0.1 水平显著)。

#### 四 F (Frequency): 腐败现状评估的次数维度

##### (一) 分罪名犯罪次数分布的描述统计

犯罪次数在不同起诉书中表述不一。第一种是采取总分结构, 指明在一定时期内, 犯罪嫌疑人共贪污 (受贿、挪用公款) N 次; 第二

种是在起诉书正文中依次罗列每笔犯罪事实；第三种则指明在一定期间内犯罪嫌疑人持续犯罪，犯罪金额合计 N 元。对于前两种表述方式可以计算犯罪次数，第三种表述方式无法准确计算犯罪次数。在总样本中，有 8879 个犯罪嫌疑人的起诉书采取前两种表述方式，占 82.89%。

在有效样本中，贪污贿赂犯罪次数中位数是 3 次，众数是 1 次（2506 人），占有效样本的 28.2%。贪污贿赂 2~9 次的占有效样本的 50.6%（其中，贪污贿赂 2 次的 1296 人，占有效样本的 14.6%；贪污贿赂 3 次的 872 人，占有效样本的 9.8%；贪污贿赂 4 次的 641 人，占有效样本的 7.2%；贪污贿赂 5 次的 493 人，占有效样本的 4.6%，依次递减）；贪污贿赂 10~99 次的占有效样本的 20.6%；贪污贿赂 100 次及以上的占有效样本的 6%。

## （二）分罪名犯罪次数分布的回归分析

模型（2）主要考察了犯罪主体特征（性别、初犯年龄、学历、职务、工作单位性质）、犯罪时空特征（犯罪持续时间、异省犯罪、区域差异、城乡差异、地域经济差异）、犯罪综合特征（犯罪人数）因素对三类腐败犯罪次数的影响。模型（2a）考察对受贿犯罪次数的影响，调整后的  $R^2$  为 0.230；模型（2b）考察对贪污犯罪次数的影响，调整后的  $R^2$  为 0.167；模型（2c）考察对挪用公款犯罪次数的影响，调整后的  $R^2$  为 0.041。三个模型均在 0.01 水平显著。

表 2 贪污、受贿、挪用公款犯罪次数对数值的 OLS 回归分析

	模型（2a） ln 受贿犯罪次数	模型（2b） ln 贪污犯罪次数	模型（2c） ln 挪用公款犯罪次数
女性	-0.201 *** (-2.600)	0.079 (1.100)	0.094 (0.849)
初犯年龄	0.002 (0.936)	-0.003 (-1.000)	-0.001 (-0.243)

续表

	模型 (2a) ln 受贿犯罪次数	模型 (2b) ln 贪污犯罪次数	模型 (2c) ln 挪用公款犯罪次数
高等教育	0.116 ** (2.097)	0.279 *** (5.561)	-0.036 (-0.386)
职务	0.151 *** (3.902)	-0.051 (-1.200)	-0.092 (-1.066)
党政机关	0.146 *** (3.971)	0.305 *** (4.418)	-0.068 (-0.640)
ln 犯罪持续时间	0.598 *** (24.247)	0.396 *** (13.617)	0.307 (5.033)
异省	-0.053 (-0.818)	-0.113 (-1.110)	-0.258 (-1.313)
东部	0.142 *** (3.442)	-0.038 (-0.750)	-0.004 (-0.041)
县域	0.036 (0.984)	-0.113 ** (-2.410)	0.055 (0.609)
ln 所在地级市人均 GDP	-0.090 ** (-2.349)	0.179 *** (4.178)	0.122 (1.340)
犯罪人数	-0.088 *** (-3.108)	-0.060 *** (-5.273)	-0.065 (-1.316)
_cons	1.770 *** (3.967)	-0.857 * (-1.720)	-0.190 (-0.195)
N	2665	2144	578
Adj. R - Square	0.230	0.167	0.041

括号中的数字是稳健的标准误差, \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ 。

根据模型 (2a)、模型 (2b)、模型 (2c) 可以得出如下初步结论。

第一, 女性犯罪嫌疑人, 受贿犯罪次数比较少, 但对贪污犯罪、挪用公款犯罪没有显著影响。回归结果显示: 在控制其他变量的前提下, 女性犯罪嫌疑人相比男性犯罪嫌疑人, 受贿次数会少 20.1% (在 0.01 水平显著)。

第二, 学历越高, 受贿犯罪、贪污犯罪次数越多, 但对挪用公款犯

罪没有显著影响。回归结果显示：在控制其他变量的前提下，受过高等教育的犯罪嫌疑人相比没有受过高等教育的，受贿犯罪次数会多 11.6%（在 0.05 水平显著），贪污犯罪次数会多 27.9%（在 0.01 水平显著）。

第三，具有职务的犯罪嫌疑人，受贿犯罪次数相对较多，但对贪污犯罪、挪用公款犯罪没有显著影响。回归结果显示：在控制其他变量的前提下，有职务的犯罪嫌疑人相比没有职务的，受贿犯罪次数会多 15.1%（在 0.01 水平显著）。

第四，党政机关的犯罪嫌疑人，受贿犯罪、贪污犯罪次数相对较多，但对挪用公款犯罪没有显著影响。回归结果显示：在控制其他变量的前提下，在党政机关工作的犯罪嫌疑人相比其他部门的犯罪嫌疑人，受贿犯罪次数会多 14.6%（在 0.01 水平显著），贪污犯罪次数会多 30.5%（在 0.01 水平显著）。

第五，犯罪持续时间越长，受贿犯罪、贪污犯罪次数越多，但对挪用公款犯罪没有显著影响。回归结果显示：在控制其他变量的前提下，犯罪持续时间每延续 1%，受贿犯罪次数会多 0.6%（在 0.01 水平显著），贪污犯罪次数会多 0.4%（在 0.01 水平显著）。

第六，在我国东部地区工作的犯罪嫌疑人，受贿犯罪次数相对较多，但对贪污犯罪、挪用公款犯罪没有显著影响。回归结果显示：在控制其他变量的前提下，在东部地区工作的犯罪嫌疑人相比中西部地区，受贿犯罪次数会多 14.2%（在 0.01 水平显著）。

第七，在县域工作的犯罪嫌疑人，贪污犯罪次数相对较多，但对受贿犯罪、挪用公款犯罪没有显著影响。回归结果显示：在控制其他变量的前提下，在县域工作的犯罪嫌疑人相比在其他行政区域工作的犯罪嫌疑人，贪污犯罪次数会多 11.3%（在 0.05 水平显著）。

第八，所在地级市人均 GDP 越高，受贿犯罪次数越少，贪污犯罪次数越多，但对挪用公款犯罪没有显著影响。回归结果显示：在控制其他变量的前提下，所在地级市人均 GDP 每增加 1%，受贿犯罪次数就会

下降 0.1% (在 0.05 水平显著), 贪污犯罪次数会增加 0.2% (在 0.01 水平显著)。

第九, 犯罪人数越多, 受贿犯罪、贪污犯罪次数越少, 但对挪用公款犯罪没有显著影响。回归结果显示: 在控制其他变量的前提下, 犯罪人数每增加 1 人, 受贿犯罪次数会减少 8.8% (在 0.01 水平显著), 贪污犯罪次数会减少 6.0% (在 0.01 水平显著)。

## 五 T (Time): 腐败现状评估的时间维度

### (一) 分罪名犯罪潜伏期分布的描述统计

犯罪潜伏期 = 案发时间 (侦查机关立案时间) - 最后一次犯罪时间。在统计时, 对于部分无法获得侦查机关立案时间的案例以起诉时间代替, 因为刑法规定了侦查时限, 所以通常侦查机关立案时间与起诉时间相差 0~1 年。

在 2016 年贪污贿赂犯罪公开起诉书中, 有 10665 人可以计算出其犯罪潜伏期。最小值是 0 年 (101 人), 最大值是 32 年 (1 人), 平均值为 5.66 年, 中位数为 5 年, 众数为 4 年。28.4% 的犯罪嫌疑人犯罪

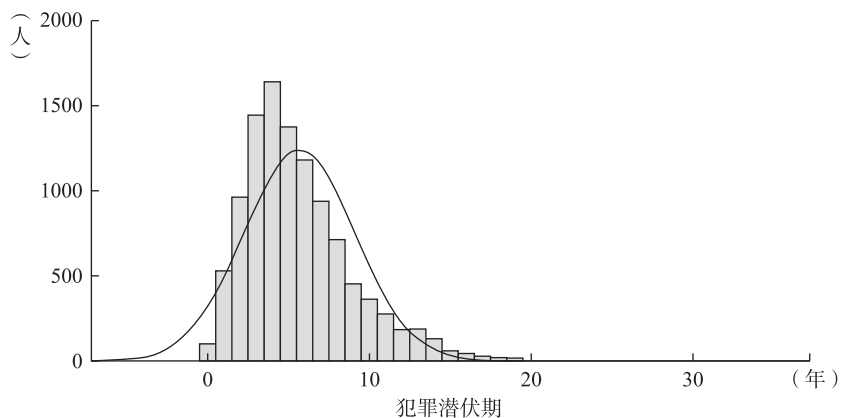


图 2 贪污贿赂罪犯罪嫌疑人的犯罪潜伏期分布直方图

潜伏期在3年（含）以内，56.6%的犯罪嫌疑人犯罪潜伏期在5年（含）以内。但是也可以看到，犯罪潜伏期在10年以上的仍有12.6%，在20年以上的有0.3%。

## （二）分罪名犯罪潜伏期分布的回归分析

模型（3）主要考察了犯罪主体特征（性别、初犯年龄、学历、职务、工作单位性质）、犯罪时空特征（异省犯罪、区域差异、城乡差异、地域经济差异）、犯罪综合特征（犯罪人数）因素对三类腐败犯罪潜伏期的影响。模型（3a）主要考察对受贿犯罪潜伏期的影响，调整后的 $R^2$ 为0.054；模型（3b）考察对贪污犯罪潜伏期的影响，调整后的 $R^2$ 为0.017；模型（3c）考察对挪用公款犯罪潜伏期的影响，调整后的 $R^2$ 为0.016。三个模型均在0.01水平显著。

表3 贪污、受贿、挪用公款犯罪潜伏期对数值的OLS回归分析

	模型（3a） ln 受贿犯罪潜伏期	模型（3b） ln 贪污犯罪潜伏期	模型（3c） ln 挪用公款犯罪潜伏期
女性	0.005 (0.127)	-0.055* (-1.774)	-0.001 (-0.015)
初犯年龄	-0.009*** (-5.967)	-0.006*** (-4.744)	-0.008*** (-3.167)
高等教育	0.136*** (4.396)	-0.030 (-1.334)	0.001 (0.019)
职务	0.159*** (7.060)	0.094*** (4.879)	-0.016 (-0.357)
党政机关	-0.014 (-0.640)	0.034 (1.114)	0.034 (0.613)
异省	-0.055 (-1.450)	-0.071 (-1.352)	-0.083 (-0.726)
东部	0.048** (2.025)	0.044** (2.023)	-0.024 (-0.527)

续表

	模型 (3a) ln 受贿犯罪潜伏期	模型 (3b) ln 贪污犯罪潜伏期	模型 (3c) ln 挪用公款犯罪潜伏期
县域	-0.038 * (-1.780)	0.046 ** (2.160)	-0.005 (-0.124)
ln 所在地级市人均 GDP	0.022 (0.964)	-0.056 *** (-2.908)	0.125 *** (2.920)
犯罪人数	-0.050 *** (-6.258)	-0.007 (-1.388)	0.069 *** (3.168)
_cons	1.665 *** (6.235)	2.341 *** (10.687)	0.206 (0.426)
N	3155	3942	1073
Adj. R - Square	0.054	0.017	0.016

括号中的数字是稳健的标准误差, \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ 。

根据模型 (3a)、模型 (3b)、模型 (3c) 可以得出如下初步结论。

第一, 女性犯罪嫌疑人, 贪污犯罪潜伏期相对较短, 但是对受贿犯罪、挪用公款犯罪没有显著影响。回归结果显示: 在控制其他变量的前提下, 女性犯罪嫌疑人相比男性犯罪嫌疑人, 贪污犯罪潜伏期会短 5.5% (在 0.01 水平显著)。

第二, 初犯年龄越大, 受贿、贪污、挪用公款犯罪潜伏期均会缩短。回归结果显示: 在控制其他变量的前提下, 初次犯罪年龄每增加 1 岁, 受贿犯罪潜伏期会短 0.9% (在 0.01 水平显著), 贪污犯罪潜伏期会短 0.6% (在 0.01 水平显著), 挪用公款犯罪潜伏期会短 0.8% (在 0.01 水平显著)。

第三, 受过高等教育的犯罪嫌疑人, 受贿犯罪潜伏期相对较长, 但是对贪污犯罪、挪用公款犯罪没有显著影响。回归结果显示: 在控制其他变量的前提下, 受过高等教育的犯罪嫌疑人相比没有受过高等教育的, 受贿犯罪潜伏期会长 13.6% (在 0.01 水平显著)。

第四, 具有职务的犯罪嫌疑人, 受贿、贪污犯罪潜伏期相对较长,



但是对挪用公款犯罪没有显著影响。回归结果显示：在控制其他变量的前提下，具有职务的犯罪嫌疑人相比没有职务的，受贿犯罪潜伏期会长 15.9%（在 0.01 水平显著），贪污犯罪潜伏期会长 9.4%（在 0.01 水平显著）。

第五，东部地区的犯罪嫌疑人，受贿、贪污犯罪潜伏期相对较长，但是对挪用公款犯罪没有显著影响。回归结果显示：在控制其他变量的前提下，东部地区的犯罪嫌疑人相比中西部地区，受贿犯罪潜伏期会长 4.8%（在 0.05 水平显著），贪污犯罪潜伏期会长 4.4%（在 0.05 水平显著）。

第六，在县域工作的犯罪嫌疑人，受贿犯罪潜伏期相对较短，贪污犯罪潜伏期相对较长，但是对挪用公款犯罪没有显著影响。回归结果显示：在控制其他变量的前提下，县域的犯罪嫌疑人相比其他行政区域犯罪嫌疑人，受贿犯罪潜伏期会短 3.8%（在 0.1 水平显著），贪污犯罪潜伏期会长 4.6%（在 0.05 水平显著）。

第七，犯罪人数越多，受贿犯罪潜伏期越短，挪用公款犯罪潜伏期越长，但是对贪污犯罪没有显著影响。回归结果显示：在控制其他变量的前提下，犯罪人数每增加 1 人，受贿犯罪潜伏期会缩短 5.0%（在 0.01 水平显著），挪用公款犯罪潜伏期会延长 6.9%（在 0.01 水平显著）。

## 六 腐败现状的多维度交叉评估

腐败现状的评估不仅可以单一维度，在未来更存在多维度交叉的可能性。例如选取腐败评估的金额维度和频度维度，构建频度 - 金额 FM 模型进行分析，可以用于评估不同地区的腐败犯罪现状差异。以天津为例，按照 FM 二分法分类，可以将腐败犯罪分为低频小额犯罪、高频小额犯罪、低频大额犯罪、高频大额犯罪四类，如表 4 所示。其中第三类低频大额犯罪和第四类高频大额犯罪比例均明显高于全国平均比例，需要引起相关职能部门重视。

表 4 天津市 2016 年公开起诉书中受贿犯罪类型占比与  
全国平均情况比较

单位: %

	低频小额犯罪	高频小额犯罪	低频大额犯罪	高频大额犯罪
天津	18.18	9.09	36.36	36.36
全国平均	54.84	26.33	8.33	10.50

为了方便计算,可以将 F 变量、M 变量的标准化得分 z 值进行运算,即  $z_{FM} = z_M + z_F$ ,并对全国 31 个省级行政区受贿情况,用 Stata 软件进行聚类分析。根据聚类分析结果,分为三类:第一类是受贿情况特别突出地区(5 个地区),包括山西、天津、辽宁、河北、重庆;第二类是受贿情况较为突出地区(7 个地区),包括江苏、福建、湖南、浙江、新疆、安徽、四川;第三类是其他地区(19 个地区),包括上海、江西、广西、云南、山东、湖北、内蒙古、河南、海南、贵州、甘肃、北京、陕西、吉林等。山西、天津以及辽宁、河北、重庆实际上属于腐败金额、频率均相对较为突出的省份。这与我们的主观认知基本一致,山西、辽宁均为中央指出的政治生态恶化、出现“塌方式腐败”的地区。重庆市委原书记薄熙来和孙政才、辽宁省委原书记王珉、河北省委原书记周本顺及天津市委原代书记、市长黄兴国等省级行政区主要负责人均纷纷“落马”。

## 七 政策建议

建设新时代“廉洁中国”是一项刻不容缓的系统工程,每一项顶层设计都会或多或少影响到反腐败攻坚战的成败。决心与民心、毅力与能力、惩治与预防、审慎与迅速,缺一不可。如果说在转型时期,下出的一步重要“先手棋”,是可以依靠“纪检铁军”的“高压反腐”快速突破局面,通过从严治标为治本赢得时间和民心,那么,在反腐败进入“深水区”,增强反腐败斗争的现代化、科学化、精细化水平,就成

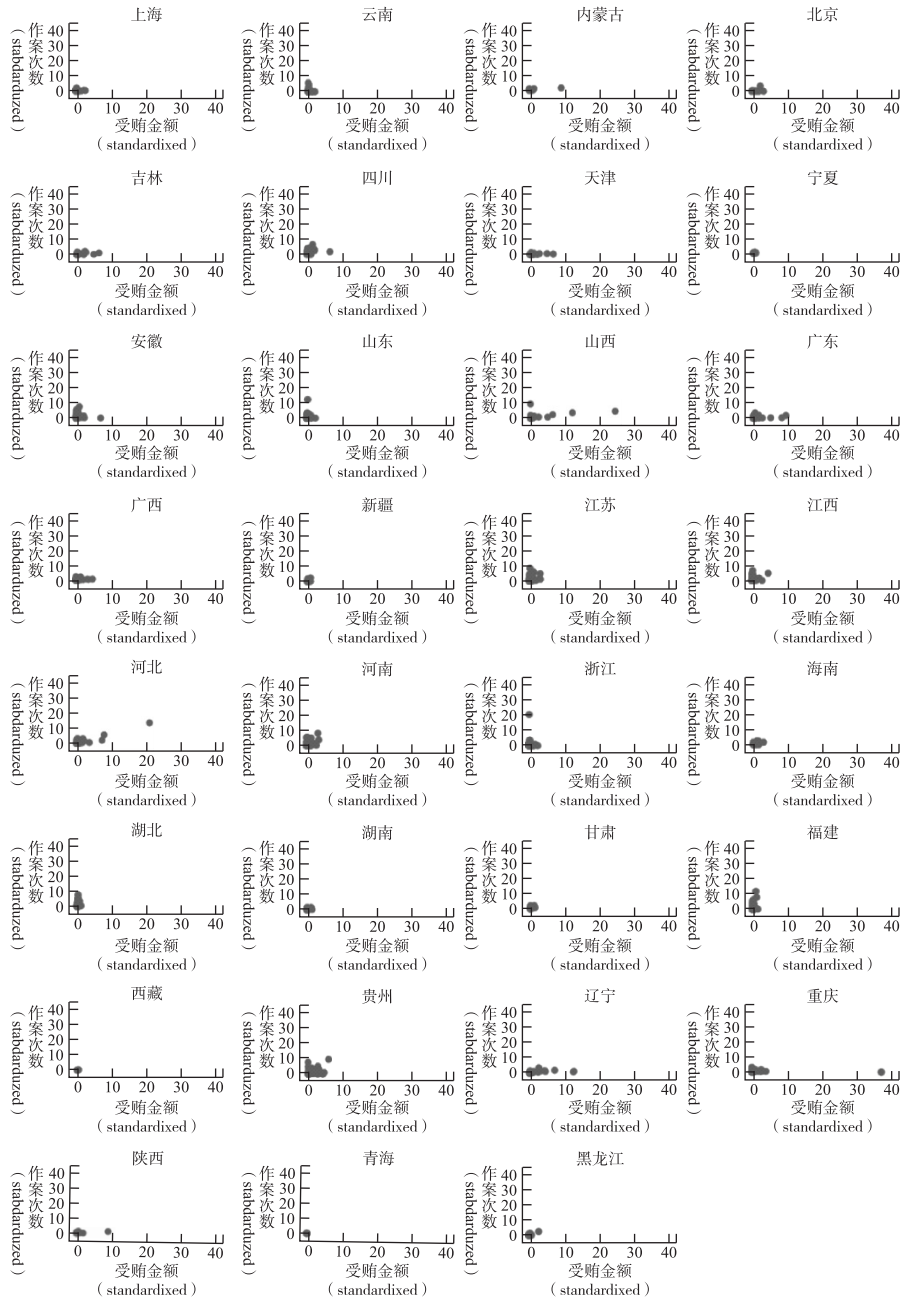


图3 全国31个省级行政单位受贿犯罪FM模型(标准化)散点图

为不二之选。

基于大数据的多维度分析,笔者建议在反腐败科学化过程中,应当注重以下问题。

### (一) 腐败惩治重心应坚持高层和基层并重,重点打击“群众身边的腐败”

2014 年党中央首次提出“全面推进从严治党”战略以来,反腐败斗争逐步取得压倒性胜利的第一阶段。十八大期间,中央纪委五年共立案审查省军级以上及中管干部 440 人。但是“打虎行动”的反腐红利正在消减,由于高级领导干部离群众距离较远,在最初的“震惊”之后,公众一定程度存在“把‘打虎’当故事”的看客心态,反腐败关注度有所下降,老百姓获得感也部分减少。笔者建议,十九大之后,在“全面推进从严治党向纵深发展”的第二阶段,应当在高层反腐“打虎猎狐”的同时,加大基层反腐力度。基层自治组织腐败的监管过去存在一定“盲区”,导致 2016 年贪污贿赂罪公开起诉书中,28.7% 发生在村民委员会、社区委员会,并且“增量”犯罪占到绝大多数,2016 年发现的基层自治组织 54.3% 的受贿行为、44% 的贪污行为、52.5% 的挪用公款行为发生在 2012 年(含)之后。建议在推进“三大攻坚战”的同时,重点加强基层党风廉政建设,加大基层反腐败资源投入,肃清“群众身边的腐败”的恶劣影响,以进一步巩固党在农村等基层的执政基础,获取更多人民群众对反腐败斗争的支持和参与。

### (二) 根据省情市情特征,差异化设置地域反腐败任务重点

我国腐败犯罪的地域差异较大。总体上东部地区犯罪金额、犯罪次数均高于中西部地区。在 2016 年公开的贪污贿赂罪起诉书中,我国东部地区贪污、受贿、挪用公款平均金额分别是中西部地区的 2.10 倍、1.10 倍、1.80 倍,犯罪次数平均数是中西部地区的 1.19 倍。在犯罪类型上,我国华北、东北、西北地区以贪污犯罪为主,东南、西南地区以

受贿犯罪为主。在各省犯罪类型上，也截然不同，例如 2016 年公开的贪污贿赂罪起诉书中，陕西省腐败犯罪中贪污罪占比达到 75.6%、受贿罪占比 18.00%；相反，海南省腐败犯罪中受贿罪占比达到 68.6%、贪污罪占比 29.75%。前者以贪污为主，后者以受贿为主。不同省份腐败犯罪特征有所不同，因此建议各地的反腐败工作也应当因地制宜，差异化设置反腐目标，配置反腐败资源。例如，按照犯罪类型差异，在河北、山西、内蒙古、陕西、甘肃、宁夏、辽宁、吉林、黑龙江等贪污犯罪为主的省域，建议纪检监察机关联合审计部门、财政部门，重点加强辖区内党政机关、国有企事业单位、基层自治组织单位财务情况的检查力度，并采取有针对性的财务管理、审计、监督等预防措施。而上海、江苏、浙江、福建、海南、重庆等受贿犯罪为主的省域，建议纪检监察机关加强受贿犯罪举报宣传力度，重拳打击行贿犯罪和“围猎”现象，提升行贿犯罪的经济成本，并有针对性地选取受贿现象突出的行业进行专项整治。

### （三）重视科学化廉政教育，针对性采取腐败预防宣传措施

建议对不同学历受众选择差异化的预防宣传措施。在犯罪手段上，高学历被告人更偏向于受贿犯罪（86.9% 受过高等教育）和独自作案（76.39%），低学历被告人更偏向于贪污犯罪（58.3% 未受过高等教育）和多人作案（56.87%）。在犯罪金额、犯罪次数上，高学历被告人贪污金额和次数分别是低学历被告人的 3.13 倍、1.74 倍；高学历被告人受贿金额和次数分别是低学历被告人的 2.53 倍、1.52 倍。回归分析中，在控制其他变量的前提下，受过高等教育的犯罪嫌疑人，受贿、贪污犯罪金额分别会多 45.8%、40.0%，受贿次数、贪污次数分别会多 11.6% 和 27.9%。不同学历被告人的文化程度不同，大概率上其犯罪方式也不同，若想提高腐败预防水平和宣传效果，需要有针对性地提出差异化的宣传警示教育方案。例如，一方面，针对低学历被告人，主要在村民委员会、社区委员会等基层自治组织发放手册或组织宣传教

育,用漫画、戏剧、纪录片、忏悔录等图文方式和口语化传播,重点警示教育贪污犯罪、挪用公款犯罪的危害,并有针对性地加强村(社区)财务管理措施;另一方面,针对高学历被告人,主要在党政机关、国有企事业单位进行廉政教育,可以采取政策宣讲、以案说法、行业廉政预防措施建议等方式,帮助其计算利弊得失、认知腐败犯罪风险,重点对受贿犯罪及其危害进行说明。

### 参考文献

- 褚红丽、魏建,2016,《腐败惩罚的边际递减及地区差异:基于腐败金额的实证分析》,《广东财经大学学报》第3期。
- 崔晶晶、邓晓梅,2014,《工程项目腐败交易阶段模型的建立与量测》,《工程管理学报》第1期。
- 公婷、吴木銮,2012,《我国2000-2009年腐败案例研究报告——基于2800余个报道案例的分析》,《社会学研究》第4期。
- 郭剑鸣,2009,《“预算国家”:一项防治腐败的基本制度供给——基于2008年国家审计署审计公告的思考》,《人文杂志》第3期。
- 国家统计局,2017,《中国统计年鉴2017》,中国统计出版社。
- 过勇,2006,《经济转轨滋生腐败机会的微观机制研究——从594个腐败要案中得出的结论》,《经济社会体制比较》第5期。
- 过勇,2013,《中国高官腐败的特点和变化趋势研究(1978—2002)》,载《国情报告》第6卷,社会科学文献出版社。
- 何家弘、徐月笛,2016,《腐败利益链的成因与阻断——十八大后落马高官贪腐案的实证分析》,《政法论坛》第3期。
- 何增科,1999,《中国转型期的腐败与反腐败问题研究:一种制度分析》,《马克思主义与现实》第5期。
- 何增科,2003,《中国转型期腐败和反腐败问题研究(上篇)》,《经济社会体制比较》第1期。
- 金鸿浩、高红玲,2017,《2016年中国反腐倡廉传播与舆情研究报告》,《廉政文化研究》第3期。
- 李辉,2017,《腐败的两幅面孔:基于7000个司法裁判文书数据的描述分析》,《理论与改革》第5期。
- 李辉,2018,《贪污受贿案涉案金额的结构特征——基于司法判决书数据的初步分析》,《复旦学报(社会科学版)》第5期。
- 刘冠君、朱立恒,2015,《遏制腐败需提升腐败惩治力度》,《科学社会主义》第5期。
- 刘启君,2013,《改革开放以来中国腐败状况实证分析》,《政治学研究》第6期。

- 罗先觉, 2010, 《从最高检工作报告看近三十年的反腐败工作》, 《国家检察官学院学报》第1期。
- 倪星、王立京, 2013, 《中国腐败现状的测量与腐败后果的估算》, 《江汉论坛》第10期。
- 乔德福, 2013, 《改革开放以来市“一把手”腐败案例研究报告——基于142个市“一把手”腐败案例分析》, 《理论与改革》第5期。
- 申纯、胡桑, 2017, 《科研腐败成因分析及防治对策研究——基于20个案例的统计分析》, 《长沙理工大学学报》(社会科学版)第4期。
- 盛宏文、魏娜, 2013, 《司法腐败的基本特征及其预防对策——基于1990—2010年相关统计数据的实证分析》, 《重庆工商大学学报》(社会科学版)第4期。
- 王传利, 2001, 《1990年至1999年中国社会的腐败频度分析》, 《政治学研究》第1期。
- 文宏、杜菲菲, 2017, 《镇街一把手腐败行为的演化机理与治理对策——基于2009—2017年镇街腐败案例的内容分析》, 《湘潭大学学报》(哲学社会科学版)第5期。
- 赵秉志, 2014, 《中国反腐败刑事法治的若干重大现实问题研究》, 《法学评论》第3期。