

外审评议与作者回复（一）

首先万分感谢审稿专家在百忙之中对本文提出极具建设性的修改意见。为了能够更好地理解审稿专家提出的建设性修改意见，我们在对论文进行认真修改的同时，对审稿专家提出的意见进行逐一回复，不当之处还请审稿专家继续批评指正。论文修改稿修改部分用红色字体标出。感谢审稿专家对论文改进提供的帮助！以下我们将逐条回答审稿专家的意见（黑色字体为审稿专家意见，蓝色字体为意见回复）：

1.在理论模型部分，基本设定存在问题，按照的作者的逻辑，评级机构接收到的信号为 θ ，给出的评级为 r ， ε 为准确评估的概率，则 $pr(r = g|\theta = g) = pr(r = b|\theta = b) = \varepsilon$ ，而信号好坏不是简单的 ε 与 $1 - \varepsilon$ ，如果作者想参考 Bolton（2012）的做法，引入信号的准确性，则还需引入一个企业风险的真实类型。

修改说明：感谢审稿专家的意见，原稿的理论模型中确实缺少了企业风险真实类型的因素。在修改稿中，我们在原稿模型的基础上加入企业真实类型的因素，对模型进行了补充，并对假设 1 进行了补充和完善，具体修改如下。

本文采用一个简单的模型来发展研究假设，本文模型基于 Bolton et al.（2012）、Opp et al.（2013）、周香芸等（2019）的博弈模型发展而来。

市场中有一家评级机构、多个债券发行人（发债企业）、一个评级监管部门。评级机构能接收到发债企业的质量信息，发债企业构建多种债券向投资者获取多期融资。企业真实质量较好的企业发行好的债券，企业真实质量较差的企业发行差的债券。评级机构（N）根据收集的信息对发行债券进行质量评估。评级机构获得信息质量 $\theta \in (g, b)$ ， g 代表好信号， b 代表差信号。评级机构获取信息的评级成本为 c ，评级成本与评级机构投入资源（R）、获取信息难易度（H）相关。投入资源越多，信息越难获取，评级成本越高。因此，评级成本为 $c(R, H)$ 。评级机构公布评级结果 $r \in (g, b)$ ，若评级机构发布较高评级，则收取费用为 W_g ；此时评级机构期望效用为 $S_1(W_g)$ 。假定 $S_1'(W_g) > 0$ ， $S_1''(W_g) < 0$ 。若评级机构发布较低评级，则收取费用为 W_b ，此时评级机构期望效用为 $S_2(W_b)$ 。假定 $S_2'(W_b) > 0$ ， $S_2''(W_b) < 0$ 。这意味着收费的期望效用的收益函数 $S(W)$ 一阶导数大于0，二阶导数小于0，符合边际收益递减的原则。投资者根据评级机构披露信息，选择投资企业债券。为了保证评级质量，相关部门对评级行业进行监管，对虚假披露给予相应惩罚。监管强度为 M ，评级机构过高评级被查处的概率为 P_{bg} ；查处后带来的声誉损失为 $-L_1(W_g)$ ，假设 $L_1'(W_g) > 0$ ， $L_1''(W_g) > 0$ 。假设评级机构过低评级被查处的概率为 P_{gb} ；查处后带来的声誉损失为 $-L_2(W_b)$ ，假设 $L_2'(W_b) > 0$ ， $L_2''(W_b) > 0$ 。这意味着收费的声誉惩罚函数 $L(W)$ 一阶导数大于0，二阶导数大于0，符合边际成本递增的原则。显然 M 越大， P_{bg} 和 P_{gb} 都越大。 S_1 、 S_2 、 L_1 、 L_2 均为连续函数。假设投资者对评级较高的债券购买概率为 p_g ，对评级较低的债券购买概率为 p_b 。 ε 表示评级机构接收企业债券信息并准确评估该债券信用等级的概率也就是信号的准确性。 β 代表当发债企业为好企业时，评级机构披露高评级债券的概率； γ 代表当发债企业为差企业时，评级机构披露低评级债券的概率。

图 1 为信用评级的博弈模型：

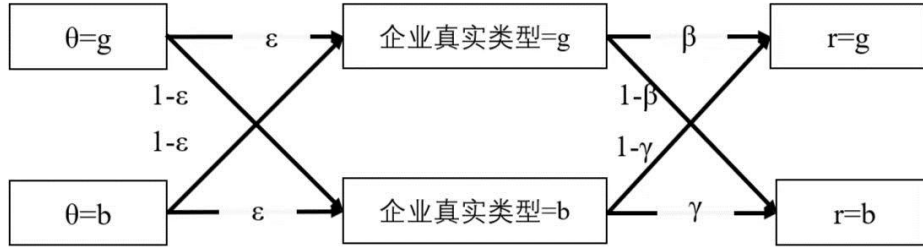


图 1 信用评级博弈模型

表 1 为不同博弈状态下的评级机构收益情况。

表 1 不同博弈状态下的评级机构收益情况

不同博弈状态下的评级机构收益情形	发生概率	评级机构收益
$\theta=g, r=g$	$\varepsilon \times \beta + (1-\varepsilon) \times (1-\gamma)$	$P_g \times S_1(W_g)$
$\theta=g, r=b$	$\varepsilon \times (1-\beta) + (1-\varepsilon) \times \gamma$	$P_b \times S_2(W_b)$
$\theta=b, r=g$	$(1-\varepsilon) \times \beta + \varepsilon \times (1-\gamma)$	$P_g \times S_1(W_g)$
$\theta=b, r=b$	$(1-\varepsilon) \times (1-\beta) + \varepsilon \times \gamma$	$P_b \times S_2(W_b)$

当同时考虑收益、成本和声誉损失时，各种情形下评级机构对应的效用为：

$$\pi_{gg} = [\varepsilon \cdot \beta + (1-\varepsilon) \cdot (1-\gamma)] \cdot p_g \cdot S_1(W_g) - c(R,H)$$

$$\pi_{gb} = [\varepsilon \cdot (1-\beta) + (1-\varepsilon) \cdot \gamma] \cdot p_b \cdot S_2(W_b) - c(R,H) - P_{gb} \cdot L_2(W_b)$$

$$\pi_{bg} = [(1-\varepsilon) \cdot \beta + \varepsilon \cdot (1-\gamma)] \cdot p_g \cdot S_1(W_g) - c(R,H) - P_{bg} \cdot L_1(W_g)$$

$$\pi_{bb} = [(1-\varepsilon) \cdot (1-\beta) + \varepsilon \cdot \gamma] \cdot p_b \cdot S_2(W_b) - c(R,H)$$

本文关注 $\pi_{bg} - \pi_{bb}$ 是否大于 0，如果大于 0，说明评级机构会倾向于过度评级，即存在过度评级的现象。如果小于等于 0，说明评级机构会倾向正常评级。

因此：

$$\begin{aligned} \pi_{bg} - \pi_{bb} &= [(1-\varepsilon) \cdot \beta + \varepsilon \cdot (1-\gamma)] \cdot p_g \cdot S_1(W_g) - P_{bg} \cdot L_1(W_g) - [(1-\varepsilon) \cdot (1-\beta) + \varepsilon \cdot \gamma] \cdot p_b \cdot S_2(W_b) \\ &= [(1-\varepsilon) \cdot \beta + \varepsilon \cdot (1-\gamma)] \cdot p_g \cdot S_1(W_g) - [(1-\varepsilon) \cdot (1-\beta) + \varepsilon \cdot \gamma] \cdot p_b \cdot S_2(W_b) - P_{bg} \cdot L_1(W_g) \end{aligned}$$

一般情况下， $0 < \varepsilon < 1$ ，故 $0 < (1-\varepsilon) < 1$ ；类似的， $0 < \beta < 1$ ， $0 < (1-\beta) < 1$ ； $0 < \gamma < 1$ ， $0 < (1-\gamma) < 1$ 。由上面条件，可知 $S_1'(W_g) > 0$ ， $S_1''(W_g) < 0$ ； $L_1'(W_g) > 0$ ， $L_1''(W_g) > 0$ ； $P_{bg} > 0$ 。

$$\frac{\partial(\pi_{bg} - \pi_{bb})}{\partial W_g} = [(1-\varepsilon) \cdot \beta + \varepsilon \cdot (1-\gamma)] \cdot p_g \cdot S_1'(W_g) - P_{bg} \cdot L_1'(W_g)$$

$$\frac{\partial(\pi_{bg} - \pi_{bb})^2}{\partial^2 W_g} = [(1-\varepsilon) \cdot \beta + \varepsilon \cdot (1-\gamma)] \cdot p_g \cdot S_1''(W_g) - P_{bg} \cdot L_1''(W_g) < 0$$

所以 $\pi_{bg} - \pi_{bb}$ 关于 W_g 是一个先递增后递减的函数，极值点为使得 $[(1-\varepsilon) \cdot \beta + \varepsilon \cdot (1-\gamma)] \cdot p_g \cdot S_1'(W_g) = P_{bg} \cdot L_1'(W_g)$ 的 W_g^* 。这也就说明了评级机构随着收费的增加，过度评级一开始会增加，然后会减少。

$$\frac{\partial(\pi_{bg} - \pi_{bb})}{\partial P_{bg}} = -L_1(W_g) < 0.$$

类似于审计意见或法律意见，信用评级意见也存在很大程度的专业判断，合理水平的认定在不同机构之间存在一定差异。更重要的是，审计意见有参考的标准（审计准则和财

务报告准则），而信用评级并没有规范或者信用水平没有相对统一的标准；虽然信用评级意见有市场公认的合理范围或者水平，但只要评级意见不是特别离谱，信用评级意见是很难被证明不合理的。另外，监管机构可以依靠其专业监管团队或者其他专家，对审计意见或者财务报告做复核性评判，以说明审计意见的合理性；但评级意见在复核性评判中更加困难，因为每家评级机构都有不同的信用评级模型，信用评级水平可以是在一定区间范围的；相对于审计意见的非黑即白（标准意见和非标意见），信用评级意见更具有操纵空间和灵活性。评级机构会综合考虑成本与收益，机会主义地配合被评级企业给予其预期的信用评级水平，即评级意见购买。激励评级机构配合被评级企业的评级意见购买行为的因素包括更高的报酬（评级收费）、更高的市场占有率、维护和客户的关系等，而更高的报酬是最直接，也是最重要的一个因素。在成本一定的情况下，评级收费越高，评级机构的机会主义动机就越强。

但是，信用评级机构劳务付出、声誉补偿、潜在损失等都会随着评级收费的提升而获得补偿，评级机构会对较高的评级收费更为重视。随着收费的增加，客户对评级机构的重要性也在增加。这种重要客户一旦评级出现问题，评级机构面临的成本将更大，包括负面新闻导致更多市场关注带来的隐性成本以及因高额收费导致的更严厉的处罚等。因此，随着客户重要性的提升或者评级收费占评级机构收入较大比重，评级机构反而会更加谨慎。而且，不论是老客户还是新客户，较高的评级收费，尤其是超过评级付出的合理成本以及声誉溢价之上的费用，往往意味着需要满足客户的评级需求，即给予客户超过其合理评级的评级水平。这部分超额的收费越高，意味着超过合理评级的程度就越大，对于评级机构而言风险就越大。一旦该债券出现违约等严重事件，会影响到评级机构的声誉以及由于监管处罚带来的其他损失。因此，虽然较高的收费对于评级机构具有很大的诱惑力，但是在更高的预期损失面前，尤其是在监管处罚极为严厉的情况下，越高的评级收费对评级机构而言反而更多地是预期损失超过收益，评级机构会变得更加谨慎，降低对于发债企业的评级迎合，减少过度评级行为。因此，评级收费与过度评级之间呈现先增后减的关系。因此，本文提出假设 1：

假设 1：评级收费与过度评级呈倒 U 型关系，即随着评级收费的提高，过度评级首先呈上升趋势；当评级收费达到一定水平后，过度评级呈现下降趋势。

2. 针对实验设计。

第一，以评级费用为例，如果想说明评级费用对信用评级的影响，应是企业基本面等信息不变，评级费用随机变化，从而引起信用评级的变化。合理的实验设计，应是每个参与者相当于一个评级机构，评级机构间有不同的收费标准，这一点可以通过随机抽取一组收费数据实现，然而论文中参与者只是抽取被评级企业顺序变化，但对于同一企业，所有参与者的信息一致，即使给出评级差异，也并非来源于评级费用变化，而是参与者个人对违约概率预期的差异。评级监管的实验设计也是同理，核心在于企业基本面一致时，自变量的随机变化。

第二，真实的评级类似一级价格拍卖过程，评级机构给出评级，发行人可以选择最高评级进行购买，因此并非评价准确有收益，而是给出最高评级赢得业务才有收益，而评级债券违约才会有声誉损失，每轮只有一人获得收益，全程没有赢得评级的参与者收益为 0。但按照现在的设计，所有参与者并不是在一轮中对同一企业给出评级，与真实评级机构的收益激励不同，参与者的决策行为也无法很好地模拟出评级机构真实行为。

第三，参与者的预期影响，在真实的评级市场中，评级机构可能面临着一个无限期博弈，实验中博弈次数为 10，作者需要考虑博弈次数是否会影响参与者的行为决策，并给出相应说明。

第四，参与者是否知道违约概率公式，如果参与者并不清楚违约概率评判标准，最终评级结果则包含了两部分，一是评级费用引起的评级虚高动机，二是评级机构判断失误，如果与理论一致，评级机构知道准确接收到 b 信号，但仍有动机给出 g 评级，则参与者应能准确计算违约概率。

修改说明：感谢审稿专家的意见，以下针对四个问题逐一回复。

第一个问题。原稿的实验设计中，如 A/B/C 公司之间的对比，检验同样的评级收费水平（收费略有差异但总体水平一样）下监管强度改变产生的效果，D/E 公司与 A/B/C 公司之间的对比，检验了同样的监管强度下不同评级收费水平产生的效果。A-E 公司基本面是相同的（总资产规模、负债率、净资产收益率、经营活动现金流入净额与净利润的倍数、过去 3 年销售收入复合增长率、净利润复合增长率、发债期限、发债金额、发债时间点），即使被试者根据财务数据进行风险评估后对不同公司的违约概率预期会有一定的差异，但实验中也给到被试者的 A-E 公司合理的主体信用评级水平均为 AA+ 级，意味着被试者对 A-E 公司的违约概率判断是差不多的，因为其合理的信用评级水平是同样的 AA+ 级。因此，参与者给出评级差异理论上来源于评级费用的变化和监管的变化。同样的，F/G/H 公司之间的对比，检验同样的评级收费水平（收费略有差异但总体水平一样）下监管强度改变产生的效果，I/J 公司与 F/G/H 公司之间的对比，检验了同样的监管强度下不同评级收费水平产生的效果。F-J 公司基本面是相同的（总资产规模、负债率、净资产收益率、经营活动现金流入净额与净利润的倍数、过去 3 年销售收入复合增长率、净利润复合增长率、发债期限、发债金额、发债时间点），实验中给到被试者的这些公司合理的主体信用评级水平均为 B+ 级。因此，参与者给出评级差异理论上来源于评级费用的变化和监管的变化。

另外，被试者为会计专业硕士学生，其拥有企业风险评估和信用评级相关知识背景（本科财务管理课程融资部分学习的内容），因此其具有一定的风险判断相关专业知识和基础。相同基本面和同样信用评级水平的评级对象，其违约概率不应该有较大差别，或者说其违约概率基本是一样的。综上，被试者对于 A-E（或者 F-J）公司给出不同决策不会是因为其违约概率的差异（因为合理的评级水平是一样的，意味着违约概率基本也是一样的），而是因为评级收费水平或者监管强度不同造成的。

第二个问题。如审稿人指出真实的评级类似一级价格拍卖过程，评级机构给出评级，发行人可以选择最高评级进行购买，也可能选择最低价格的，也可能选择发行人认为合适的评级。也就是说，并非给出最高评级才会赢得业务，才有收益。否则市场上就会出现整体上评级高得离谱的信用评级，或者出现低价揽客获得很多客户的情况，而市场上评级机构也有一定的风险管理和客户选择，发债企业也并非都认为最高的评级是最合理的和有用的。因此，各个评级水平或收费水平下的评级都是有可能获得客户认可的。

而且，现实里一些评级并非拍卖过程，而是企业的询价过程。评级机构在获得评级业务后会根据企业的基本面信息（包括企业的性质、行业、财务数据、发债情况）以及公司治理等其他指标后，对企业的风险情况进行评估，并给出企业合理的信用评级。之后企业与评级机构之间进行博弈，最终可能出现以下情况：①沟通之后获得双方达成的结果，如实验中的选择。②企业不满意目前评级，选择另一家评级机构进行评级，新的评级机构同样根据企业的基本面信息（包括企业的性质、行业、财务数据、发债情况）以及公司治理等其他指标对企业的风险情况进行评估，给出自己的评级结果，并对比企业之前获得的信用评级水平，与企业进行沟通博弈，最终可能达成或者没有达成评级结果。达成就如①的情况，没有达成就可能出现企业在市场里重新选择评级机构，继而重复以上的循环。最终发债企业会在市场中找到合适的评级机构并最终达成一致的评级意见，即评级机构妥协（出现评级意见购买行为）或者发债企业妥协（没有评级意见购买行为），即实验中的选择。当然发债企业也可能找不同的两家评级机构出具两个不同的评级意见（多重评级）。考虑到

实验中很难设计企业二次选择评级机构，而且多重评级与单一评级又有一定的差异，因此实验在初始设计时就进行了简化，即评级机构是发债企业最终的唯一选择，不存在之前的多次选择和多重选择问题。

第三个问题。考虑到被测试者的风险厌恶程度以及对政策的预期会与实验的进度有一定的关系，即被测试对象存在学习曲线（Learning），因此，回归中采用实验的轮次表征学习曲线，若实验为第一轮，则 Learning 为 1，若为第十轮，则 Learning 为 10。回归中对此进行了控制，结果并没有显著影响。

另外，根据审稿人提出的问题，本文也分别采用了 4 轮/8 轮博弈样本进行检验，结果与主检验一致，结论依旧成立。修改稿在稳健性检验中进行了补充，简洁起见未在正文中报告回归结果，具体回归结果如下。

表 4 轮/8 轮博弈样本回归结果

变量	Resrating	
	前 4 轮实验结果	前 8 轮实验结果
	(1)	(2)
Fee	74.118*** (4.57)	78.514*** (3.59)
Feesq	-61.745*** (-3.66)	-60.611*** (-2.85)
Reguforce	-4.021*** (-3.05)	-3.486*** (-4.12)
Wealth	-1.610 (-1.64)	-6.011** (-2.44)
StandRating	-0.044 (-0.23)	0.111 (1.50)
Learning	-2.856*** (-2.68)	0.794* (1.66)
Individual	控制	控制
N	144	288
Pseudo R ²	0.9504	0.9324

注：被解释变量为过度评级ResRating；模型为Ologit模型；***p<0.01,**p<0.05,*p<0.1；括号内为Z值。

第四个问题。本文的实验是将企业的基本面信息（包括企业的性质、行业、财务数据、发债情况）都给到被试者，让其对公司的风险进行初步的判断。考虑到公司治理等指标对风险评估也有一定的影响，但因素太多会使得被试者不好判断，因此所有案例公司都只是考虑企业的基本面信息，而将例如公司治理等因素忽略或者说统一。基于以上信息，被试者可以做出对于发债企业的风险评估决策，即初步的信用评级水平。之后给出的合理信用评级可以辅助被试者对其风险评估逻辑和合理性进行判断和修正。参与者虽然不知道违约概率公式，但是根据合理评级能够对违约概率有一定的了解以及合理的判断。因为被试者为会计专业硕士学生，其拥有企业风险评估和信用评级相关知识背景（本科财务管理课程融资部分学习的内容），因此其具有一定的风险判断相关专业知识和基础。相同基本面和同样信用评级水平的评级对象，其违约概率不应该有较大差别，或者说其违约概率基本是一样的。综上，被试者对于 A-E（或者 F-J）公司给出不同决策不会是因为其违约概率的差异（因为合理的评级水平是一样的，意味着违约概率基本也是一样的），而是因为评级收费水平或者监管强度不同造成的。而且，实验中为了减少对于违约概率不同的估算导致风险判断上差异较大，实验中仅存在两种合理评级和违约概率，参与者并不会有很大的判断失误

和干扰。

以上是我们对审稿专家提出问题的回复和修改，再次感谢审稿专家具有建设性的全面建议。由于我们自身能力上的不足和客观条件的局限，上述修改和说明可能还存在一些不尽如人意的地方，敬请您继续批评指正。

论文作者

外审评议与作者回复（二）

首先万分感谢审稿专家在百忙之中对本文提出的极具建设性的修改意见。为了能够更好地理解审稿专家提出的建设性修改意见，我们在对论文进行认真逐一修改的同时，对审稿专家所提出的意见进行逐一回复，不当之处还请审稿专家继续批评指正。论文修改稿修改部分用红色字体标出。感谢审稿专家对论文改进提供的帮助！以下我们将逐条回答审稿专家的意见（黑色字体为审稿专家意见，蓝色字体为意见回复）。

1.论文单独检验评级收费、评级监管对信用评级的影响，两个自变量缺乏关联，从论文题目看应该是评级收费和评级监管对信用评级的交互影响，既然两者没有联系为什么放在一个实验里检验，不符合实验研究的范式。

修改说明：感谢审稿专家指出评级收费和评级监管的交互影响以及论文对此阐述不足的问题。原稿的实验中对此进行了研究设计，如 A/B/C 公司之间的对比，检验同样的评级收费水平下监管强度改变产生的效果，D/E 公司与 A/B/C 公司之间的对比，检验了同样的监管强度下不同评级收费水平产生的效果，而 F-J 公司与 A-E 公司之间的对比，则检验了收费水平和评级监管对评级质量的交互影响，但我们之前对其阐述不够具体和清晰，因此修改稿在实证检验中对这部分交互影响的结果的分析和说明也进行了补充完善，参考表 9 的补充检验，具体如下。

监管的加强是否也影响了评级机构在评级意见购买中的风险判断，即影响到评级收费对评级合理性的作用上？对此，本文进一步进行了检验，结果见表 9。首先，本文对全样本、无监管样本和有监管样本分别进行回归，比较评级收费的拐点，以此来评价监管和评级收费给评级合理性带来的交互影响。

表 9 评级收费与监管强度的交叉影响

变量	Resrating				
	全部样本	无监管样本	有监管样本	全部样本	全部样本
	(1)	(2)	(3)	(4)	(5)
Fee	45.535*** (9.30)	76.065*** (2.71)	52.496*** (4.23)	78.910** (2.20)	95.053 (1.27)
Feesq	-35.831*** (-7.63)	-64.277** (-2.43)	-33.249*** (-2.63)	-69.004** (-2.09)	-87.963 (-1.13)
Reguforce				-3.424** (-2.17)	-3.284** (-2.10)
Reguforce×Fee				8.403* (1.72)	-0.373 (-0.04)
Reguforce×Feesq					24.746 (0.69)
Wealth	-4.352*** (-3.40)	-13.340** (-2.17)	-5.658 (-1.58)	-3.975* (-1.94)	-4.113* (-1.88)
StandRating	0.070 (1.57)	0.214** (2.22)	0.105 (1.47)	0.165* (1.96)	0.179 (1.53)
Learning	0.631*** (2.74)	1.896* (1.87)	0.877 (1.36)	0.479 (1.44)	0.498 (1.45)
Gender	0.573	11.434***	0.334	7.926*	0.426

	(0.97)	(3.20)	(0.33)	(1.75)	(0.56)
Age	2.684***	3.760**	0.331	3.488*	3.961***
	(3.89)	(2.01)	(0.31)	(1.73)	(4.52)
Individual	控制	控制	控制	控制	控制
N	360	360	360	360	360
Pseudo R ²	0.9045	0.9214	0.9268	0.9257	0.9262

注：被解释变量为过度评级ResRating；模型为Ologit模型；***p<0.01,**p<0.05,*p<0.1；括号内为Z值。

在前三列的回归中，评级收费 Fee 和 Feesq 的回归系数都显著，且与之前结果一致，此时评级收费 Fee 与过度评级倒 U 型关系的拐点同样在第一象限，进一步支持了假设 1。而第（1）列全样本中的拐点为 0.64（即 $45.535 / (2 \times 35.831)$ ），即随着评级收费的自然对数 lFee 达到 0.64 后，评级收费越高，过度评级会下降。第（2）列为对没有监管措施的样本（Reguforce=0）进行的回归，也就是说即使出现债券违约（不论是否存在评级虚高），信用评级机构也不会受到监管的任何监管措施。第（2）列无监管措施样本的拐点为 0.59（即 $76.065 / (2 \times 64.277)$ ）。第（3）列为有监管措施的样本（Reguforce>0）进行的回归，只要出现债券违约且存在评级虚高，信用评级机构就会受到监管措施，这种情况下的拐点为 0.79（即 $52.496 / (2 \times 33.249)$ ）。也就是说，全样本回归的拐点较无监管措施样本的拐点更大。这意味着随着监管强度的提高，评级收费 Fee 与过度评级的倒 U 型关系进一步右移，即拐点向右平移。本文认为可能的原因在于：相对于没有监管的情况，由于存在较强的监管，评级机构会要求更高的评级收费从而补偿过度评级可能带来的损失。换言之，在存在较强的监管情况下，与没有监管相比，同样的评级收费可能无法达到评级机构对预期损失的补偿预期，因此就不太可能出现预期中的过度评级。以上结果表明，信用评级中较高的监管不仅会直接遏制过度评级现象，也会造成评级机构更加关注预期损失，降低评级意见购买的可能。

再次感谢审稿专家的这一意见，原稿中我们对评级收费和评级监管两者的交互影响的逻辑和检验均阐述得不够清楚，修改稿中我们在实验设计以及实验结果的实证分析中都予以了补充，审稿专家建议的这一分析大大加强了本文的研究深度。

2. 理论逻辑不清楚，作者构造了一个简单的模型说明评级收费和评级监管对信用评级的影响，需要进一步的理论阐释，为什么是倒 U 型关系。

修改说明：感谢审稿专家指出本文在理论层面阐述该倒 U 型关系时的存在的不足。修改稿中在理论分析部分对倒 U 型关系进行了加强阐述。具体参见理论分析与研究假设。

类似于审计意见或法律意见，信用评级意见也存在很大程度的专业判断，合理水平的认定在不同机构之间存在一定差异。更重要的是，审计意见有参考的标准（审计准则和财务报告准则），而信用评级并没有规范或者信用水平没有相对统一的标准；虽然信用评级意见有市场公认的合理范围或者水平，但只要评级意见不是特别离谱，信用评级意见是很难被证明不合理的。另外，监管机构可以依靠其专业监管团队或者其他专家，对审计意见或者财务报告做复核性评判，以说明审计意见的合理性；但评级意见在复核性评判中更加困难，因为每家评级机构都有不同的信用评级模型，信用评级水平可以是在一定区间范围的；相对于审计意见的非黑即白（标准意见和非标意见），信用评级意见更具有操纵空间和灵活性。评级机构会综合考虑成本与收益，机会主义地配合被评级企业给予其预期的信用评级水平，即评级意见购买。激励评级机构配合被评级企业的评级意见购买行为的因素包括更高的报酬（评级收费）、更高的市场占有率、维护和客户的关系等，而更高的报酬是最直接，也是最重要的一个因素。在成本一定的情况下，评级收费越高，评级机构的机会主义动机

就越强。

但是，信用评级机构劳务付出、声誉补偿、潜在损失等都会随着评级收费的提升而获得补偿，评级机构会对较高的评级收费更为重视。随着收费的增加，客户对评级机构的重要性也在增加。这种重要客户一旦评级出现问题，评级机构面临的成本将更大，包括负面新闻导致更多市场关注带来的隐性成本以及因高额收费导致的更严厉的处罚等。因此，随着客户重要性的提升或者评级收费占评级机构收入较大比重，评级机构反而会更加谨慎。而且，不论是老客户还是新客户，较高的评级收费，尤其是超过评级付出的合理成本以及声誉溢价之上的费用，往往意味着需要满足客户的评级需求，即给予客户超过其合理评级的评级水平。这部分超额的收费越高，意味着超过合理评级的程度就越大，对于评级机构而言风险就越大。一旦该债券出现违约等严重事件，会影响到评级机构的声誉以及由于监管处罚带来的其他损失。因此，虽然较高的收费对于评级机构具有很大的诱惑力，但是在更高的预期损失面前，尤其是在监管处罚极为严厉的情况下，越高的评级收费对评级机构而言反而更多地是预期损失超过收益，评级机构会变得更加谨慎，降低对于发债企业的评级迎合，减少过度评级行为。因此，评级收费与过度评级之间呈现先增后减的关系。

假设 1：评级收费与过度评级呈倒 U 型关系，即随着评级收费的提高，过度评级首先呈上升趋势；当评级收费达到一定水平后，过度评级呈现下降趋势。

3.实验设计存在明显缺陷。实验材料中在给出公司的基本信息后，直接给出公司债券的合理评级，这种操作不是基于现实，也不合理，具有很强的引导性。因此，导致实验结果的说服力差。

修改说明：感谢审稿专家指出的关于实验材料和实验设定的问题，原文中我们对这方面的阐述不够清晰。本文的实验材料中的确是直接给出了公司债券的合理评级，然后按照评级费用给出了相应的评级。但是，实验并不是一上来就直接给出这样的陈述和选择，而是将企业的基本面信息（包括企业的性质、行业、财务数据、发债情况）都给到被试者，让其对公司的风险进行初步的判断。考虑到公司治理等指标对风险评估也有一定的影响，但因素太多会使得被试者不好判断，因此所有案例公司都只是考虑企业的基本面信息，而将例如公司治理等因素忽略或者说统一。基于以上信息，被试者可以做出对于发债企业的风险评估决策，即初步的信用评级水平。之后给出的合理信用评级可以辅助被试者对其风险评估逻辑和合理性进行判断和修正。实验中不同企业的基本面有一定差异，其他方面相同，这样被试者就可以从这些基本面的差异来对发债企业的风险作判断。案例公司的信息一开始进行陈述的，上述信息都是在每个评级决策之前（具体参见附注中的企业信息、信用评级报酬以及监管情况），从而在一定程度上减少评级购买的引导性。从另一个角度来说，如果案例公司的信用评级只是给出基本面信息，让被试者自行给出评级水平，那么每个人的风险评价标准不一样，对于同样的企业给出的信用评级就会不同，从而影响到实验设计的目的（同样风险的发债企业，评级机构对于评级收费的风险收益判断）。因此实验设计中在给出被试者企业的基本面信息（包括企业的性质、行业、财务数据、发债情况）后，给出了一个合理的信用评级水平，从而减少这方面的影响。而且，案例公司之间的可比性就更强，可以聚焦实验设计的目的，主要研究评级机构对评级收费的风险收益判断。

现实中，评级机构在获得评级业务后会根据企业的基本面信息（包括企业的性质、行业、财务数据、发债情况）以及公司治理等其他指标后，对企业的风险情况进行评估，并给出企业合理的信用评级。之后企业与评级机构之间进行博弈，最终可能出现以下情况：①沟通之后获得双方达成的结果，如实验中的选择。②企业不满意目前评级，选择另一家评级机构进行评级，新的评级机构同样根据企业的基本面信息（包括企业的性质、行业、财务数据、发债情况）以及公司治理等其他指标对企业的风险情况进行评估，给出自己的

评级结果，并对比企业之前获得的信用评级水平，与企业进行沟通博弈，最终可能达成或者没有达成评级结果。达成就如①的情况，没有达成就可能出现企业在市场里重新选择评级机构，继而重复以上的循环。最终发债企业会在市场中找到合适的评级机构并最终达成一致的评级意见，即评级机构妥协（出现评级意见购买行为）或者发债企业妥协（没有评级意见购买行为），即实验中的选择。当然发债企业也可能找不同的两家评级机构出具两个不同的评级意见（多重评级）。考虑到实验中很难设计企业二次选择评级机构，而且多重评级与单一评级又有一定的差异，因此实验在初始设计时就进行了简化，即评级机构是发债企业最终的唯一选择，不存在之前的多次选择和多重选择问题。

以上是我们对审稿专家提出问题的回复和修改，再次感谢审稿专家具有建设性的全面建议。由于我们自身能力上的不足和客观条件的局限，上述修改和说明可能还存在一些不尽如人意的地方，敬请您继续批评指正。再次衷心感谢您！祝老师工作顺利、身体健康！

论文作者

评级收费、评级监管与信用评级合理性

——基于实验研究的证据（附录）

附录 1 文献综述

信用评级作为市场中重要的信息中介，能够减少投资者与发行人之间的信息不对称，通过不同水平的信用评级对外揭示发行人不同的风险等级，对债券融资成本产生直接影响（Ahmed et al., 2002; Kisgen & Strahan, 2010）。针对我国债券市场信用评级经济后果的研究也发现，中国债券市场的信用评级有着显著的融资成本效应（何平和金梦，2010；王雄元和张春强，2013；寇宗来 等，2015；顾小龙 等，2017）。然而，正因为信用评级会对债券能否顺利发行及其融资成本产生显著影响，发行人有强烈动机去获得高评级（Graham & Harvey, 2001）；而作为评级机构，在“发行人付费”模式下，出于对利益的追求，也会迎合发行方的需求，为其出具虚高的评级（Becker & Milbourn, 2011; Jiang et al., 2012; 林晚发等，2017；寇宗来 等，2020；吴育辉 等，2020），导致“评级购买”（Rating Shopping）现象的出现（Bolton et al., 2012）。

因此，学术界对如何抑制低质量评级、提升评级质量进行了大量研究。评级市场的竞争、声誉机制及监管被认为有可能会抑制评级的虚高，提升评级质量；然而，有关研究成果并不尽然。不同于产品市场的竞争会带来产品质量的提升，评级市场的竞争被发现往往会导致虚高评级的产生或恶化这一现象（Skreta & Veldkamp, 2009; Becker & Milbourn, 2011; Bolton et al., 2012）。这一结果并不意外，因为面临激烈的评级市场竞争，为了争取到业务并获得更高的利益，评级机构有动机去迎合发行人对更高评级的需求，从而导致竞争机制的失效。针对我国债券市场的研究也发现市场竞争对评级质量具有负面影响，即竞争会导致或加剧评级的虚高（詹明君和邢贺，2016；邢天才 等，2016；徐晓萍 等，2018）。周宏 等（2013）通过博弈模型分析也发现，评级机构数量越多，机构与发债企业串谋的风险越大。2010 年引入的“投资人付费”评级模式也没有办法改善整体的评级质量，如寇宗来 等（2020）发现面对“投资人付费”评级机构中债资信的低评级，“发行人付费”的评级机构不但不跟随其评级，反而会调高拟发债企业的后续评级进行对冲。那么声誉机制能起作用吗？Mathis et al.（2009）认为声誉无法约束评级公司；Heski & Shapiro（2012）认为声誉机制的作用受到企业所处经济周期的影响，在经济繁荣周期下，声誉机制无法抑制低质量的评级。Bolton et al.（2012）也认为在经济繁荣时，声誉损失较小，评级机构更有可能出具低质量评级。在针对我国债券市场的研究中，詹明君和邢贺（2016）、邢天才 等（2016）都发现评级机构的声誉机制对评级质量有正面影响。由此可见，关于声誉机制对评级质量的影响，相关结论并不一致。除此之外，外部监管对评级质量的影响，结论也不一致，如 Cheng & Neamtiu（2009）发现自《萨班斯-奥克斯利法案》通过后，信用评级机构提高了评级的准确度，强化监管提高了评级质量。而 Dimitrov et al.（2015）针对《多德-弗兰克法案》公布后对信用评级质量影响的实证研究表明，该法案增加了信用评级机构的法律和监管成本，这一成本的增加产生了逆向的效果。Goel & Thakor（2015）也认为，增加诉讼和监管成本是一把双刃剑，既有可能使评级机构提供更有信息含量的评级，也有可能使其模糊其评级。但邓博文和曹廷贵（2016）发现，通过加强监管能够改善评级质量。

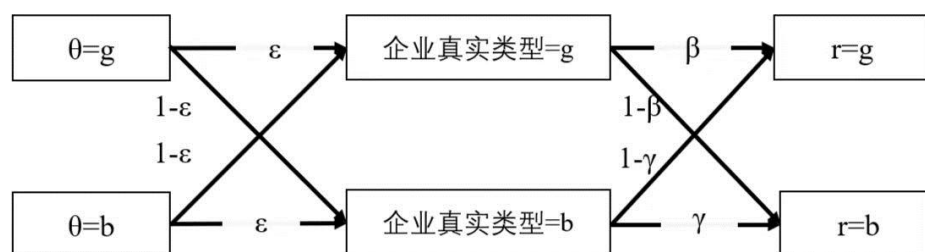
附录 2 理论模型

本文采用一个简单的模型来发展研究假设，本文模型基于 Bolton et al.（2012）、Opp et

al. (2013)、周香芸 等 (2019) 研究的博弈模型发展而来。

市场中有一家评级机构、多个债券发行人（发债企业）、一个评级监管部门。评级机构能接收到发债企业的质量的信息，发债企业构建多种债券向投资者获取多期融资。企业真实质量较好的企业发行违约可能性较低的债券，企业真实质量较差的企业发行违约可能性较高的债券。评级机构（N）根据收集的信息对发行债券进行质量评估。评级机构获得信息质量 $\theta \in (g, b)$ ， g 代表好信号， b 代表差信号。评级机构获取信息的评级成本为 c ，评级成本与评级机构投入资源（ R ）、获取信息难易度（ H ）相关。投入资源越多，信息越难获取，评级成本越高。因此，评级成本为 $c(R, H)$ 。评级机构公布评级结果 $r \in (g, b)$ ，若评级机构发布较高评级，则收取费用为 W_g ；此时评级机构期望效用为 $S_1(W_g)$ 。假定 $S_1'(W_g) > 0$ ， $S_1''(W_g) < 0$ 。若评级机构发布较低评级，则收取费用为 W_b ，此时评级机构期望效用为 $S_2(W_b)$ 。假定 $S_2'(W_b) > 0$ ， $S_2''(W_b) < 0$ 。这意味着收费的期望效用的收益函数 $S(W)$ 一阶导数大于 0，二阶导数小于 0，符合边际收益递减的原则。投资者根据评级机构披露信息，选择投资企业债券。为了保证评级质量，相关部门对评级行业进行监管，对虚假披露给予相应惩罚。监管强度为 M ，评级机构过高评级被查处的概率为 P_{bg} ；查处后带来的声誉损失为 $-L_1(W_g)$ ，假设 $L_1'(W_g) > 0$ ， $L_1''(W_g) > 0$ 。假设评级机构过低评级被查处的概率为 P_{gb} ；查处后带来的声誉损失为 $-L_2(W_b)$ ，假设 $L_2'(W_b) > 0$ ， $L_2''(W_b) > 0$ 。这意味着收费的声誉惩罚函数 $L(W)$ 一阶导数大于 0，二阶导数大于 0，符合边际成本递增的原则。显然 M 越大， P_{bg} 和 P_{gb} 都越大。 S_1 、 S_2 、 L_1 、 L_2 均为连续函数。假设投资者对评级较高的债券购买概率为 p_g ，对评级较低的债券购买概率为 p_b 。 ε 表示评级机构接收企业债券信息并准确评估该债券信用等级也就是信号的准确性。 β 代表当发债企业为好企业时，评级机构披露高评级债券的概率； γ 代表当发债企业为差企业时，评级机构披露低评级债券的概率。

附图 1 为信用评级的博弈模型。



附图 1 信用评级博弈模型

附表 1 为不同博弈状态下的评级机构收益情况表。

附表 1 不同博弈状态下的评级机构收益情况

不同博弈状态下的评级机构收益情形	发生概率	评级机构收益
$\theta=g, r=g$	$\varepsilon \times \beta + (1-\varepsilon) \times (1-\gamma)$	$P_g \times S_1(W_g)$
$\theta=g, r=b$	$\varepsilon \times (1-\beta) + (1-\varepsilon) \times \gamma$	$P_b \times S_2(W_b)$
$\theta=b, r=g$	$(1-\varepsilon) \times \beta + \varepsilon \times (1-\gamma)$	$P_g \times S_1(W_g)$
$\theta=b, r=b$	$(1-\varepsilon) \times (1-\beta) + \varepsilon \times \gamma$	$P_b \times S_2(W_b)$

当同时考虑收益、成本和声誉损失时，各种情形下评级机构对应的效用为：

$$\pi_{gg} = [\varepsilon \cdot \beta + (1-\varepsilon) \cdot (1-\gamma)] \cdot p_g \cdot S_1(W_g) - c(R, H)$$

$$\pi_{gb} = [\varepsilon \cdot (1-\beta) + (1-\varepsilon) \cdot \gamma] \cdot p_b \cdot S_2(W_b) - c(R, H) - P_{gb} \cdot L_2(W_b)$$

$$\pi_{bg} = [(1-\varepsilon) \cdot \beta + \varepsilon \cdot (1-\gamma)] \cdot p_g \cdot S_1(W_g) - c(R,H) - P_{bg} \cdot L_1(W_g)$$

$$\pi_{bb} = [(1-\varepsilon) \cdot (1-\beta) + \varepsilon \cdot \gamma] \cdot p_b \cdot S_2(W_b) - c(R,H)$$

本文关注 $\pi_{bg} - \pi_{bb}$ 是否大于 0，如果大于 0，说明评级机构会倾向于过度评级，即存在过度评级的现象。如果小于等于 0，说明评级机构会倾向正常评级。

因此：

$$\begin{aligned} \pi_{bg} - \pi_{bb} &= [(1-\varepsilon) \cdot \beta + \varepsilon \cdot (1-\gamma)] \cdot p_g \cdot S_1(W_g) - P_{bg} \cdot L_1(W_g) - [(1-\varepsilon) \cdot (1-\beta) + \varepsilon \cdot \gamma] \cdot p_b \cdot S_2(W_b) \\ &= [(1-\varepsilon) \cdot \beta + \varepsilon \cdot (1-\gamma)] \cdot p_g \cdot S_1(W_g) - [(1-\varepsilon) \cdot (1-\beta) + \varepsilon \cdot \gamma] \cdot p_b \cdot S_2(W_b) - P_{bg} \cdot L_1(W_g) \end{aligned}$$

一般情况下， $0 < \varepsilon < 1$ ，故 $0 < (1-\varepsilon) < 1$ ；类似的， $0 < \beta < 1$ ， $0 < (1-\beta) < 1$ ； $0 < \gamma < 1$ ， $0 < (1-\gamma) < 1$ 。由上面条件，可知 $S_1'(W_g) > 0$ ， $S_1''(W_g) < 0$ ； $L_1'(W_g) > 0$ ， $L_1''(W_g) > 0$ ； $P_{bg} > 0$ 。

$$\frac{\partial(\pi_{bg} - \pi_{bb})}{\partial W_g} = [(1-\varepsilon) \cdot \beta + \varepsilon \cdot (1-\gamma)] \cdot p_g \cdot S_1'(W_g) - P_{bg} \cdot L_1'(W_g)$$

$$\frac{\partial(\pi_{bg} - \pi_{bb})^2}{\partial^2 W_g} = [(1-\varepsilon) \cdot \beta + \varepsilon \cdot (1-\gamma)] \cdot p_g \cdot S_1''(W_g) - P_{bg} \cdot L_1''(W_g) < 0$$

所以 $\pi_{bg} - \pi_{bb}$ 关于 W_g 是一个先递增后递减的函数，极值点为使得 $[(1-\varepsilon) \cdot \beta + \varepsilon \cdot (1-\gamma)] \cdot p_g \cdot S_1'(W_g) = P_{bg} \cdot L_1'(W_g)$ 的 W_g^* 。这也就说明了评级机构随着收费的增加，过度评级一开始会增加，然后会减少。

$$\frac{\partial(\pi_{bg} - \pi_{bb})}{\partial P_{bg}} = -L_1(W_g) < 0.$$

基于此，可以得到本文的假设 1：

假设 1：评级收费与过度评级呈倒 U 型关系，即随着评级收费的提高，过度评级首先呈上升趋势；当评级收费达到一定水平后，过度评级呈现下降趋势。

模型中 $\pi_{bg} - \pi_{bb}$ 关于 P_{bg} 是一个单调递减函数，因此随着 P_{bg} 的增加，即监管强度的增加，过度评级的可能性会下降。基于此，可以得到本文的假设 2：

假设 2：随着监管强度的提高，过度评级现象减少。

附录 3 实验设计

本实验目的在于探究监管对评级机构风险判断的影响。实验建立在如下前提假设之上：

- (1) 存在业界可以达成共识的合理的信用评级模型和信用评级水平；
- (2) 评级机构的专业胜任能力足以识别被评级单位的任何信用风险和漏洞；
- (3) 所有被评级单位的非财务信息基本相同；
- (4) 对每个被评级单位耗费的评级成本相同；

本实验在学校行为实验室进行，被试同时进行实验。实验对象为国内某重点大学工商管理学院 MPAcc 专业硕士学生。相比普通学生，这些 MPAcc 专业硕士学生有较强的会计财务知识基础，了解债券评级机构的内容和职责；其中一些 MPAcc 专业硕士学生也有一定的相关行业的从业或实习经验。因此相对其他学生来说，其认知偏差带来的对实证结果的干扰较少。用 MPAcc 专业硕士学生作为被试对象也是会计领域实验研究中一种较为普遍的实验对象选择，他们被认为是公司财务主管、审计师以及非职业投资者等角色的合适替代者（Libby et al., 2002；张继勋 等，2010；张继勋 等，2016；He et al., 2019；张继勋 等，2019）。

实验为情景模拟实验，被试在实验中扮演评级机构高层的角色，有权决定对被评级单位出具不同水平的信用等级。实验为连续实验，共分为 10 轮，每轮结束后产生的后果将影响下一轮实验，即本轮获得报酬、被查处后的罚款金额不清零，累积到下一轮，直到 10 轮实验结束后计算总报酬。被试在实验中最终获得的报酬与被试实际得到的被试费挂钩，提高实验的真实感，增强了被试参与实验的体验感和积极性。实验形式为上机操作，已根据实验流

程设计软件，在每名被试进行实验前，主试告知其操作方法，引导被试进行实验。具体实验流程如附表 2 所示。

附表 2 实验流程

顺序	实验程序
第 1 轮	
1	被试对象进入评级机构的角色，形成实验之前需要形成上述 4 条信念
2	向被试对象提供一家被评级公司的背景信息和报酬（随机抽取 A~J 中的一家）
3	被试对象根据所提供信息作出评级意见的决策
4	根据本轮债券违约概率 λ_1 ，告知被试对象被评级企业是否违约，监管机构是否约谈和/或处罚金额
第 2-9 轮	
1	随机抽取剩余公司中的一家，向被试对象提供该公司的背景信息和报酬
2	被试对象根据所提供信息作出评级意见的决策 ¹
3	根据每轮债券违约概率 $\lambda_i(i=2,3,4\ldots 9)$ ，告知被试对象被评级企业是否违约，监管机构是否约谈和/或处罚金额
第 10 轮	
1	步骤同以上几轮
2	根据被试对象得到的报酬和被惩罚的金额计算其最终获得的收益

注：每轮的违约概率 $\lambda=1-(0.3*\text{总资产规模}+0.1*\text{净资产收益率}+0.2*\text{经营活动现金流入金额占净利润的比例}+0.1*\text{销售收入增长率}+0.1*\text{净利润增长率}-0.2*\text{资产负债率})$ ；计算机程序会计算违约概率，每轮实验中违约是随机的。

实验中 10 家公司的具体信息见附录 7——实验公司情况部分。附表 3 对以上 A~J 这 10 家被评级公司的财务情况进行了汇总。前 5 家公司（A~E）的财务状况相对较好，后 5 家公司（F~J）的财务状况相对较差。债券的发行规模和发行期限相同，从而减少发行规模和债务期限对信用评级的影响。

附表 3 被评级公司财务情况

公司	资产规模	负债率	净资产收益率	现金流占利润比例	销售增长率	净利润增长率	发行规模	发行期限
A	201	16	17	115	32	29	10	5
B	200	15	18	120	30	28	10	5
C	201	16	17	110	35	30	10	5
D	200	16	17	118	32	28	10	5
E	199	15	18	115	33	29	10	5
F	201	52	6.3	10	3	5	10	5

¹ 现实情况中，评级机构在获得评级业务后会根据企业的基本面信息（包括企业的性质、行业、财务数据、发债情况）以及公司治理等其他指标后，对企业的风险情况进行评估，并给出企业合理的信用评级。之后企业与评级机构之间进行博弈，最终可能出现以下情况：①沟通之后获得双方达成的结果，如实验中的选择。②企业不满意目前评级，选择另一家评级机构进行评级，新的评级机构同样根据企业的基本面信息（包括企业的性质、行业、财务数据、发债情况）以及公司治理等其他指标对企业的风险情况进行评估，给出自己的评级结果，并对比企业之前获得的信用评级水平，与企业进行沟通博弈，最终可能达成或者没有达成评级结果。达成就如①的情况，没有达成就可能企业出现企业在市场里重新选择评级机构，继而重复以上的循环。最终发债企业会在市场中找到合适的评级机构并最终达成一致的评级意见，即评级机构妥协（出现评级意见购买行为）或者发债企业妥协（没有评级意见购买行为），即实验中的选择。当然发债企业也可能找不同的两家评级机构出具两个不同的评级意见（多重评级）。考虑到实验中很难设计企业二次选择评级机构，而且多重评级与单一评级又有一定的差异，因此实验设计进行了简化，即评级机构是发债企业最终的唯一选择，不存在之前的多次选择和多重选择问题。

G	202	50	6	11	2	5.5	10	5
H	200	49	6.2	9	2.5	4.8	10	5
I	199	50	6.4	10	2.8	4.9	10	5
J	201	51	6.5	10	3	5	10	5

附表 4 汇总了以上 A~J 这 10 家被评级公司的合理评级、违约概率、不同评级水平下的评级付费（评级公司的评级收费）以及在出现违约情况下的预期监管方式（预期约谈、预期反馈或者没有监管）。基于附表 3 的财务状况，前 5 家公司（A~E）合理评级为 AA+，违约概率在 12%左右；后 5 家公司（F~J）合理评级为 B+，违约概率在 46%左右²。以上设定用于比较和检验不同违约概率（合理评级）下的过度评级情况。

评级收费围绕合理评级设定了 5 档，4 档高于合理评级，1 档低于合理评级。高于合理评级的为过度评级或高估评级，低于合理评级的为评级低估。考虑到有些评级机构可能更加重视声誉以及更谨慎稳健，因此设定了评级低估的选项。以上设定可以根据实际评级与合理评级的差异度量过度评级，从而更加合理地检验信用评级合理性问题。

A/D/F/I 这四家公司如果债券出现违约时，评级公司不会受到任何监管处罚；B/G 这两家公司如果债券出现违约时，评级公司仅会受到监管机构的约谈，提交评级资料并进行整改；C/E/H/J 这四家公司如果债券出现违约时，评级公司不仅会被监管机构约谈，同时还会处以一定的罚金。以上设定用于检验不同的监管方式（没有监管、预期约谈、预期约谈并罚款）对于过度评级的影响。

A/B/C 这三家公司在不同评级水平下的评级收费大致相同，而 D/E 这两家公司则与 A/B/C 形成对比，其他信息基本相同，但评级收费有差异。后 5 家公司（F~J）的设定规则和逻辑同前 5 家公司（A~E）。以上设定用于检验不同评级收费对过度评级的影响³。

附表 4 被评级公司信用评级、评级收费及预期监管方式

公司	合理 评级	违约 概率	评级收费										预期 约谈	预期 罚款
			AAA+	AAA	AAA-	AA+	AA	BB+	BB	BB-	B+	B		
A	AA+	12.1	39	31	19	15	9						0	0
B	AA+	11.4	40	30	20	15	9						1	0
C	AA+	12.7	41	32	22	15	10						1	1
D	AA+	12.7	54	39	29	15	9						0	0
E	AA+	12.3	55	40	30	15	8						1	1
F	B+	46.67						40	30	21	15	8	0	0
G	B+	45.85						41	31	22	15	9	1	0
H	B+	46.65						40	30	20	15	10	1	1
I	B+	46.89						55	41	30	15	9	0	0
J	B+	46.45						54	42	31	15	8	1	1

附录 4 关键变量详细定义

ResRating 为被评级企业与评级机构之间的评级意见购买行为，采用实际评级水平与合理信用评级之间的差异表征，该指标为正意味着过度评级（存在评级意见购买），指标为零

² 实务中，AA+和 B+的债券违约率并没有这么高。考虑到实验中需要有一定的违约率，因此设定了较高的违约率。不过 AA+和 B+两类债券违约率的差异相对合理。

³ A/B/C 公司之间的对比，检验同样的评级收费水平下监管强度改变产生的效果，D/E 公司与 A/B/C 公司之间的对比，检验了同样的监管强度下不同评级收费水平产生的效果，而 F~J 公司与 A~E 公司之间的对比，则检验了收费水平、评级监管的交互影响。

意味着合理评级（不存在评级意见购买），指标为负意味着评级低估⁴。信用评级水平按照 Afonso 等量化方法对微调式信用等级（AAA+、AAA、AAA-、AA+、AA、AA-、A+、A、A-、BBB+、BBB、BBB-、BB+、BB、BB-、B+、B、B-、CCC、CC、C）依次赋值，其中最低级 C 赋值为 1，CC 赋值为 2，依次递增，最高级 AAA+ 赋值为 21。若合理的信用评级水平(StandRating)为 AA+，评级机构给出的信用评级水平(RealRating)为 AAA，则 ResRating 等于 20（AAA 对应 20）减去 18（AA+ 对应 18），即过度评级为 2。稳健性检验中，本文也采用评级机构实际给出的信用评级作为被解释变量进行检验。

评级收费类似于审计收费，收费在很大程度上取决于工作量以及承担的风险。本文在实验中对工作量以及基本风险做了相应控制和处理，因此评级机构给出的评级收费实际上相当于因承担额外风险而收取的费用⁵。Fee 为评级机构给出信用评级水平后收到的评级费用，采用实验中的实际给出的评级水平对应的评级收费表征。为检验评级收费与过度评级之间的非线性关系，本文加入了 Fee 的二次项 Feesq。为降低变量的共线性问题，回归中采用评级收费取自然对数，并对其进行中心化。稳健性检验中，本文也采用了过度收费替换评级收费重新检验，过度收费即评级机构实际收取的评级收费减去该企业合理评级应当收取的评级收费。

Regulation 为评级监管变量。本文采用监管强度（Reguforce）表征监管部门对信用评级机构在债券出现违约后的监管措施，0 表示公司在债券存续期间出现违约时，信用评级机构不会被监管机构警告或者处罚⁶；1 表示公司在债券存续期间出现违约时，信用评级机构会被监管约谈，要求提交评级资料，并进行整改；2 表示公司在债券存续期间出现违约时，信用评级机构会被监管约谈，要求提交评级资料，进行整改，没收评级收费并处以 3 倍罚金⁷。该指标越大，意味着监管强度越高，监管越严格。为了更好的区分不同监管措施的效果，本文在稳健性检验中首先采用哑变量定义监管方式进行检验。根据实验设定，本文将监管方式划分为监管约谈 Talk、监管处罚 TalkPenalty 以及无处罚 Noaction。无处罚 Noaction 为哑变量，1 表示公司在债券存续期间出现违约时，信用评级机构不会被监管机构警告或者处罚，

⁴ 在实验过程中合理评级的确定已经控制了相关的定性和定量信息；进一步的，由于我们在实验过程中已经告知了相应公司的合理评级，因此被试对象主要需要作出决策的并不是判断该公司的合理评级是否合适，而是基于实验给出的合理评级以及对应收费信息作为参照基准点，需要决定基于不同评级收费是否给出相应的评级。

⁵ 评级收费包含了两部分内容，一部分是评级的劳务，另一部分是购买评级的费用。在本文实验设计过程中，我们对评级劳务费用这部分进行了控制，即控制了公司的规模和债券的规模，也就是评级公司的规模、债券的发行规模和发行期限基本上是接近的；在此基础上，本文将公司分为了较好的公司（ABCDE）和较差的公司（FGHIJ）两类，这两类公司的其他财务特征如负债率、净资产收益率、现金流占利润比例、销售增长率、净利润增长率也分别是接近的。上述两点也就基本上控制了评级劳务费用的一致性。另外，本文在实验过程中，也告诉了被试者该公司的合理评级水平以及该水平对应下的合理收费。并且，不管是财务状况较好的公司还是财务状况较差的公司，其合理评级下对应的合理评级收费均是 15 万元。因此通过上述实验过程的控制，我们已经对评级的合理劳务费用进行了控制，目前收取的评级费用的差异，主要就是基于购买评级的费用。

⁶ 在诸如美国、欧洲等海外成熟市场，甚至在中国香港地区，企业发债出现违约，并不一定会被监管处罚。投资者通常基于法律诉讼的途径寻求赔偿，监管机构则较少对评级机构处罚。如 2018 年 8 月，穆迪同意支付 1 500 万美元来解决涉及美国住宅抵押贷款支持证券（RMBS）评级模型的内部控制失败指控，并将聘请独立顾问来评估和改善其内部控制。另外，穆迪还同意支付 125 万美元了解信贷评级符号的指控，并审查其有关评级符号的政策，程序和内部控制。因此本文参考国外市场的监管机构较少直接进行行政管制的现状，在实验中设定有如果发生违约，评级机构不会被监管处罚的情景。

⁷ 关于处罚金额，我国刑法关于罚金刑的数额，采取了三种立法模式：倍比罚金刑、限额罚金刑、无限额罚金制。刑法第一百八十条、第二百二十五条规定“并处或者单处违法所得一倍以上五倍以下罚金”。而行政处罚方面，法律、法规、规章对某类违法行为罚款没有最高限额规定的，而且每个地方规定的行政处罚的数额本身就不一样，行政处罚的数额是需要结合案件本身决定的。现实中，如证监会等监管机构给出的行政处罚往往为“违法所得”的 1~5 倍，顶格为 5 倍。本实验中有处罚金额的情境中，对于处罚金额的设定“没一罚三”，即本次信用评级收费加三倍信用评级收费，贴近监管机构通行的做法。感谢匿名审稿人指出这一问题，让本文对于罚金的设定更为合理。

否则为 0；监管约谈 *Talk* 为哑变量，1 表示被评级企业在债券存续期间出现违约时信用评级机构被监管约谈，要求提交评级资料 and 进行整改，否则为 0；监管处罚 *TalkPanelty* 为哑变量，1 表示公司在债券存续期间出现违约时，信用评级机构会被监管约谈，要求提交评级资料，进行整改，如果发现评级虚高，则评级机构会被没收违约所得且处以 3 倍罚金，否则为 0。稳健性检验中，本文也采用了预期处罚金额 *ExpPenalty* 衡量监管。预期处罚金额为若被评级企业在债券存续期间出现违约，预期监管机构对评级机构进行的罚款金额，回归中采用（预期罚款金额+1）的自然对数。

附录 5 稳健性检验

1. 监管方式衡量评级监管

附表 5 从监管方式角度衡量监管，区分了不同监管方式对过度评级的影响。第（1）列和第（2）列分别检验不同的监管方式（监管谈话 *Talk* 与监管谈话加处罚 *TalkPanelty*），第（3）列综合了三种方式进行比较。在所有回归中，信用评级收费 *Fee* 的系数显著为正，评级收费二次项 *Feesq* 的系数显著为负，同正文表 8 的回归，即评级收费与过度评级之间呈倒 U 型关系，支持了本文的假设 1。

不论是在单独回归的第（1）列，还是在综合回归的第（3）列，监管约谈（*Talk*）的回归系数都显著为正，意味着相对于没有监管处罚的情况，监管约谈反而会进一步刺激过度评级。可能的原因在于虽然监管部门采用了约谈的方式，在债券出现违约的情况下对评级机构进行监管，要求其整改，但约谈并没有更进一步的实质性罚金或者其他禁止营业等措施，对评级机构而言没有起到约束的作用，反而会令其认为违规成本很低，刺激评级收费对其的诱惑，从而刺激过度评级⁸。但是，如果监管部门除了约谈之外，还有更加严厉的惩罚措施，如高额罚金，那么就会对评级机构构成较大的威慑，从而有效抑制过度评级。这表现在监管约谈并处罚（*TalkPanelty*）的回归系数在第（2）列和第（3）列中都高度显著为负，即存在监管约谈并处罚的措施能够降低过度评级问题。这一结果更进一步支持了假设 2。其他变量的回归情况基本与正文表 8 一致，篇幅原因不再累述。

总之，附表 5 的结果更进一步表明：不同的监管方式对抑制过度评级的效果不同，简单的监管约谈不仅不会起到预期效果，反而会刺激过度评级；而存在实质性处罚（罚金）能够有效抑制过度评级，提高信用评级的合理性。

附表 5 评级收费、监管方式与过度评级

变量	Resrating		
	(1)	(2)	(3)
<i>Fee</i>	60.597*** (4.24)	89.794*** (2.78)	84.003*** (3.32)
<i>Feesq</i>	-50.691*** (-3.26)	-74.956** (-2.35)	-73.236*** (-2.79)
<i>Talk</i>	4.943*** (3.84)		5.254*** (3.18)

⁸ 监管约谈、检查等是监管的重要手段之一，通过约谈、检查可能会发现违规问题，从而带来后续更严厉惩处措施；即使不涉及违规，也有应对监管的成本、后续整改成本等。随着近年来监管改善，如人民银行等五部委《关于促进债券市场信用评级行业健康发展的通知》中提到，“加强对信用评级机构评级质量和全流程作业合规情况的检查，对评级区分度明显低于行业平均水平、跟踪评级滞后、大跨度调整级别、更换信用评级机构后上调评级结果等情形进行重点关注”。为了避免被监管约谈和检查，评级机构会控制此类情况的出现，以及由此带来的合规风险。因此本文中发现的约谈对过度评级没有约束的情形可能会发生改变。这一方面指出未来的监管约谈要配合后续的整改措施督查，另一方面也是本文目前实验设计的局限性之一。

TalkPenalty		-5.217*** (-3.14)	-4.103*** (-2.83)
Wealth	-4.813*** (-3.22)	-3.415 (-1.47)	-3.469* (-1.84)
StandRating	0.047 (0.91)	0.159** (2.30)	0.124** (1.96)
Learning	0.703*** (2.77)	0.405 (1.03)	0.438 (1.24)
个人固定效应	控制	控制	控制
N	360	360	360
Pseudo R ²	0.9250	0.9359	0.9413

注：被解释变量为过度评级ResRating；模型为Ologit模型；***p<0.01,**p<0.05,*p<0.1；括号内为Z值。

2. 预期处罚金额表征评级监管

附表 6 进一步从预期处罚金额角度衡量监管强度，对假设重新进行检验。第（1）列为采用全部样本的回归结果。信用评级收费 Fee 的系数显著为正，评级收费二次项 Feesq 的系数显著为负，同表 8 的回归，即评级收费与过度评级之间呈倒 U 型关系，支持了本文的假设 1。预期处罚金额 ExpPenalty 的回归系数在 0.01 水平上显著为负，即预期处罚金额越高，过度评级越不明显。也就是说，预期处罚越严格（在监管约谈基础上增加实质性罚金），越会对评级机构产生更高的威慑，从而抑制其机会主义动机，从而降低过度评级，提升信用评级的合理性。

第（2）列剔除了评级低估样本进行回归⁹，信用评级收费 Fee 的系数显著为正，评级收费二次项 Feesq 的系数显著为负，预期处罚金额 ExpPenalty 的回归系数显著为负，同第（1）列回归，进一步支持了假设 1 和假设 2。

第（3）列剔除了违约样本，仅采用未违约样本进行回归。同样的，信用评级收费 Fee 的系数显著为正，评级收费二次项 Feesq 的系数显著为负，预期处罚金额 ExpPenalty 的回归系数显著为负，同第（1）列和第（2）列，结果进一步支持了假设 1 和假设 2。

总之，附表 6 的结果更进一步表明：实质性的高额罚金对抑制过度评级起到了显著作用，预期处罚金额越高，越会加大评级机构机会主义行为的违约成本，从而降低其机会主义动机，提高信用评级的合理性。

附表 6 评级收费、处罚金额与过度评级

变量	Resrating		
	全部样本	剔除评级低估样本	剔除违约样本
	(1)	(2)	(3)
Fee	87.992*** (2.84)	87.994*** (2.84)	81.176** (2.50)
Feesq	-71.918** (-2.37)	-71.920** (-2.37)	-58.493** (-2.11)
ExpPenalty	-1.052*** (-3.10)	-1.052*** (-3.10)	-1.135** (-2.28)
Wealth	-3.415 (-1.49)	-3.415 (-1.49)	-2.825 (-0.79)

⁹ 仅采用过度评级样本进行回归的结果与目前结果基本一致。

StandRating	0.159** (2.31)	0.159** (2.31)	0.211* (1.67)
Learning	0.412 (1.06)	0.412 (1.06)	0.405 (0.68)
个人固定效应	控制	控制	控制
N	360	356	263
Pseudo R ²	0.9353	0.9320	0.9362

注：被解释变量为过度评级ResRating；模型为Ologit模型；***p<0.01,**p<0.05,*p<0.1；括号内为Z值。

3. 评级收费、监管强度与信用评级

考虑到实验中人为设定了一个合理的信用评级水平，而现实中不同评级机构对于风险判断不同，给出的合理评级可能也会有差异，因此稳健性检验中本文没有采用过度评级作为被解释变量，而直接采用评级机构实际给出的信用评级作为被解释变量，用数据回归的方式检验合理性问题。结果见附表 7 的前 3 列，第（1）列采用监管强度（Reguforce），第（2）列采用不同的监管方式（Talk、TalkPanelty），第（3）列采用预期处罚金额（ExpPanelty）。

评级收费变量Fee的回归系数在三列回归中都高度显著为正，而收费二次方项Feesq的系数也都高度显著为负，意味着评级收费与过度评级之间呈倒U型关系。这与采用过度评级ResRating的结果一致，假设 1 进一步得到验证。监管强度（Reguforce）的回归系数显著为负，结果同正文表 8；监管方式Talk的系数显著为正，而TalkPanelty的系数显著为负，结果同附表 5；预期处罚金额（ExpPanelty）的回归系数显著为负，结果同附表 6。总之，假设 2 进一步得到验证。

4. 过度收费、监管强度与过度评级

鉴于信用评级收费刺激过度评级的逻辑，在本质上是过度收费刺激了过度评级，因此稳健性检验中本文也采用了信用评级过度收费替换评级收费，重新进行检验。结果见附表 7 的后 3 列，同样的第（4）列采用监管强度（Reguforce），第（5）列采用不同的监管方式（Talk、TalkPanelty），第（6）列采用预期处罚金额（ExpPanelty）。

过度评级收费变量ResFee的回归系数在三列回归中都高度显著为正，而收费二次方项ResFeesq的系数也基本高度显著为负，意味着过度评级收费与过度评级之间呈倒U型关系。这与采用评级收费Fee的结果一致，假设 1 进一步得到验证。监管强度（Reguforce）、监管方式Talk、TalkPanelty以及预期处罚金额（ExpPanelty）的回归系数都高度显著，结果与之前的一致。总之，假设 2 进一步得到验证。

附表 7 稳健性检验——不同变量度量方式

变量	Rating			ResRating		
	(1)	(2)	(3)	(4)	(5)	(6)
Fee	63.921*** (3.86)	91.390*** (3.79)	88.193*** (3.04)			
Feesq	-46.437** (-2.45)	-80.651*** (-3.13)	-71.928** (-2.37)			
ResFee				48.533*** (3.62)	72.523*** (3.25)	74.138*** (2.69)
ResFeesq				-18.116 (-1.02)	-45.969*** (-2.70)	-43.451** (-2.10)
Reguforce	-1.727*** (-2.68)			-1.732** (-2.52)		

Talk	5.969***			5.203***		
	(3.66)			(3.22)		
TalkPanelty	-4.101***			-4.102***		
	(-2.98)			(-2.82)		
ExpPanelty		-1.052***			-1.050***	
		(-3.13)			(-3.07)	
Wealth	-3.896**	-3.384*	-3.407	-3.944**	-3.435*	-3.373
	(-2.05)	(-1.78)	(-1.45)	(-2.04)	(-1.84)	(-1.49)
StandRating	16.460***	23.971	21.420***	0.143*	0.121*	0.156**
	(5.34)	(1.57)	(3.64)	(1.84)	(1.94)	(2.28)
Learning	0.512	0.423	0.410	0.521	0.433	0.406
	(1.58)	(1.20)	(1.04)	(1.59)	(1.24)	(1.05)
个人固定效应	控制	控制	控制	控制	控制	控制
N	360	360	360	360	360	360
Pseudo R ²	0.9467	0.9619	0.9577	0.9184	0.9414	0.9353

注：前 3 列的被解释变量为实际给出的信用评级Rating，后 3 列的被解释变量为过度评级ResRating；模型为Ologit模型；***p<0.01,**p<0.05,*p<0.1；括号内为Z值。

5. 剔除中途破产样本

一旦企业发行的债券出现违约，而评级机构被查出存在过度评级的情况，评级机构会被监管机构给予实质性的金钱上的处罚，即“没一罚三”，导致可能在某轮实验过程中被测试者的初始财富不足罚金，出现“负初始财富”的情况。而在这种情况下被测试者可能出现更加激进冒险的选择，导致结果出现一定偏差。在实验中，只出现了 2 例“负初始财富”的情况。剔除这 2 例样本，回归结果与之前一致，支持了本文的假设 1 和假设 2。本文也进一步将初次出现“负初始财富”之后的样本全部删除，共删除样本 13 个，回归结果也与之前一致，支持了本文的假设 1 和假设 2¹⁰。

6. 区分高违约率和低违约率样本

本文的实验中，A~E这五家公司的违约率大致相同，而F~J这五家公司的违约率大致相同，但远高于A~E的五家，有可能人为设定的较高的违约率导致结果的偏差，本文分别对高违约率和低违约率的样本分别进行回归。而不论是在高违约率样本，还是低违约率样本中，评级收费变量Fee、收费二次方项Feesq、监管强度（Reguforce）、监管方式Talk、TalkPanelty以及预期处罚金额（ExpPanelty）的回归系数也都符合预期，与之前结果一致，支持本文的假设 1 和假设 2。

7. 参与者预期的影响

在真实的评级市场中，评级机构可能面临着一个无限期博弈。不过，本实验中博弈次数为 10，博弈次数可能会影响参与者的行为决策。因此，本文也采用了 4 次和 8 次的博弈次数样本进行检验。而不论是 4 次博弈样本，还是 8 次博弈样本，评级收费变量 Fee、收费二次方项 Feesq、监管强度（Reguforce）的回归系数都符合预期，与之前结果一致，支持本文的假设 1 和假设 2。

附录 6 进一步检验：债券违约后的评级收费与过度评级

附表 8 对比了上一轮实验中债券是否出现违约（PreDefault：哑变量，上一轮出现违约，则 PreDefault 为 1，否则为 0）、是否实际被监管约谈（PreRealTalk：哑变量，上一轮出现违约后，评级机构实际被监管进行约谈的则 PreRealTalk 为 1，否则为 0）、是否实际被监管处

¹⁰ 篇幅原因未在正文中报告结果。

罚（PreRealTalkPanelty：哑变量，上一轮出现违约后，评级机构实际被监管进行处罚的，则PreRealTalkPanelty 为 1，否则为 0）的情况下，本轮的过度评级情况。全部 360 个信用评级中，有 263 例未违约，97 例债券违约。

上一轮未违约情况下（PreDefault=0）的过度评级，其均值和中位数都没有显著高于上一轮违约情况（PreDefault=0）的过度评级，表明上一轮违约对于本轮是否过度评级没有明显的影响，上一轮违约没有导致评级机构在之后的评级中更加谨慎。

上一轮债券出现违约，评级机构被监管部门实际进行了约谈并整改但没有罚款（PreRealTalk=1）的情况出现了 52 次，而这种情况下本轮评级机构的过度评级的均值和中位数依旧没有显著高于没有被监管约谈的情况，意味着表明上一轮即使被监管机构约谈并要求整改，并没有对本轮是否过度评级产生明显的影响，即上一轮监管约谈没有使得评级机构在之后的评级中更加谨慎。

上一轮债券出现违约，评级机构被监管部门实际进行约谈并进行罚款（PreRealTalkPanelty=1）的情况出现了 6 次，而这种情况下本轮评级机构的过度评级的均值和中位数也依旧没有显著高于没有被监管处罚的情况，意味着表明上一轮即使被监管机构约谈并处罚，并没有对本轮是否过度评级产生明显的影响，即上一轮监管处罚也没有使得评级机构在之后的评级中更加谨慎。

附表 8 债券违约与过度评级¹¹

变量	类型	PreDefault=0 (272)	PreDefault=1 (88)	Diff	t/z
ResRating	Mean	1.779	1.693	0.086	0.532
	Median	2	2	0	0.484
		PreRealTalk=0 (308)	PreRealTalk=1 (52)		
ResRating	Mean	1.743	1.846	-0.102	-0.518
	Median	2	3	-1	-0.668
		PreRealTalkPanelty=0 (354)	PreRealTalkPanelty=1 (6)		
ResRating	Mean	1.751	2.166	-0.415	-0.764
	Median	2	2.5	-0.5	-0.715

注：***p<0.01,**p<0.05,*p<0.1。

附表 9 对上一轮是否违约（PreDefault）、是否实际被监管约谈（PreRealTalk）以及是否实际被监管处罚（PreRealTalkPanelty）对本轮信用评级合理性的影响进行了回归检验。第（1）列单独采用上一轮违约变量 PreDefault，未考虑本轮评级收费和预期处罚的影响。PreDefault 的系数不显著，意味着上一轮违约没有显著影响抑制下一轮信用评级机构的过度评级，这与附表 8 的均值检验结果一致。加入评级收费和监管处罚措施（预期处罚金额）的第（2）列回归中，PreDefault 的系数依旧不显著。

第（3）列和第（4）列加入了上一轮的实际处罚方式，即上一轮实际监管约谈（PreRealTalk）或者监管约谈加罚金（PreRealTalkPanelty）。结果显示：上一轮监管约谈 PreRealTalk 和上一轮监管约谈加罚金 PreRealTalkPanelty 的系数也都不显著。表明之前的监管措施对于后续的信用评级的影响并不明显，影响最大的也就是本轮的评级收费和预期处罚金额。

第（5）列和第（6）列综合了上一轮是否违约（PreDefault）以及上一轮的实际处罚方式，结果与之前第（1）~（4）列的结果一致。总之，附表 9 的结果表明：上一轮违约以及

¹¹ 剔除第一轮样本（即没有上一轮是否违约的样本），结果与目前附表 8 结果一致。

违约后的监管方式对之后的信用评级都没有产生明显的影响。对评级风险判断影响最大的依旧是评级收费和预期的处罚金额。

附表 9 评级收费、监管处罚与过度评级-基于债券违约后比较¹²

变量	Resrating					
	(1)	(2)	(3)	(4)	(5)	(6)
PreDefault	-0.125 (-0.49)	0.269 (0.34)			-0.265 (-0.77)	0.902 (0.41)
PreRealTalk			0.065 (0.18)	-0.210 (-0.27)	0.293 (0.61)	-1.100 (-0.49)
PreRealTalkPenalty			-0.004 (-0.47)	0.010 (0.67)	-0.004 (-0.44)	0.009 (0.64)
Fee		87.844*** (2.83)		89.633*** (2.80)		84.731*** (2.87)
Feesq		-72.217** (-2.33)		-71.948** (-2.28)		-67.560** (-2.34)
ExpPenalty		-1.052*** (-3.07)		-1.083*** (-3.18)		-1.057*** (-3.13)
Wealth	-1.834*** (-4.02)	-3.319 (-1.46)	-1.911*** (-3.92)	-2.942 (-1.20)	-1.864*** (-3.83)	-3.004 (-1.24)
StandRating	0.051*** (2.92)	0.157** (2.30)	0.050*** (2.83)	0.170** (2.45)	0.050*** (2.85)	0.165** (2.27)
Learning	0.232*** (2.84)	0.395 (1.00)	0.245*** (2.79)	0.331 (0.77)	0.237*** (2.72)	0.354 (0.84)
个人固定效应	控制	控制	控制	控制	控制	控制
N	360	360	360	360	360	360
Pseudo R ²	0.0965	0.9354	0.0964	0.9355	0.0970	0.9358

注：被解释变量为过度评级ResRating；模型为Ologit模型；***p<0.01,**p<0.05,*p<0.1；括号内为Z值。

附表 8 的均值检验以及附表 9 的回归分析都表明之前的监管措施对于后续的信用评级的影响并不明显。可能的原因可以用赌徒谬误理论来解释：人们认为上一轮已经抛过硬币正面，下一轮就不容易抛正面。评级机构在决策时也会如此，认为即使上一轮被处罚过，后面被处罚的概率就小。也就是说，评级机构会机会主义地认为再次被监管机构发现的概率不会变大，甚至可能变小，因此并没有降低其激进的做法，即提供过度评级服务，从而导致之前的监管措施对本期的评级决策影响不明显。这一结果与黄小琳 等（2017）一致，可能的原因在于现行制度环境尤其是监管强度较轻，不足以从实质上促使评级机构改变其过度评级的动机。只有更加严厉的评级监管措施，如被监管机构禁止从业，或许能够较好地约束评级机构的激进行为。如 2014 年“11 超日债”打破中国债券市场的刚性兑付后，监管机构并没有对评级机构表现出严格的监管态度和措施。而 2018 年 8 月某公国际资信评估有限公司被暂停银行间债券市场相关业务及整改期间不得承接新的证券评级业务，证监会也宣布暂停某公国际证券评级业务一年¹³，监管态度和措施才严格起来。因此在未来，随着监管强度的增加，之前监管措施与本期评级决策之间的关系可能会发生改变。

¹² 剔除第一轮样本（即没有上一轮是否违约的样本），回归结果与目前附表 9 结果一致。

¹³ 北京证监局依据《暂行办法》拟责令大公国际限期整改，期限一年，整改期间不得承接新的证券评级业务，更换不符合条件的高级管理人员。

附录 7 实验公司情况

实验中涉及的公司信息、信用评级报酬以及监管情况如下所示。

A 控股集团有限公司成立于 2003 年，于 2011 年在深圳证券交易所注册上市。公司主要经营范围为煤炭的洗选、加工；矿产品的开发与经营；开展煤炭、铁路、电力经营的配套服务；能源与环保技术开发与利用、技术转让、技术咨询、技术服务；进出口业务；化工产品、化工材料、建筑材料、机械设备的销售。该公司 2018 年末总资产为 201 百万元，负债率 16%，净资产收益率为 17%，经营活动现金流入净额为净利润的 115%。过去 3 年销售收入复合增长率为 32%，净利润复合增长率为 29%。公司计划于 2019 年发行 10 亿元 5 年期的公司债。该公司聘请你所在的信用评级机构对其主体进行信用评级，该公司合理的主体信用评级水平为 AA+级。公司愿意为 AAA+级付出 39 万元评级费用，为 AAA 级付出 31 万元评级费用，为 AAA-付出 19 万元评级费用，为 AA+付出 15 万元评级费用，为 AA-付出 9 万元评级费。即使公司在债券存续期间出现违约，你所在的信用评级机构不会被监管机构警告或者处罚。

B 股份有限公司成立于 2003 年，于 2014 年在上海证券交易所注册上市，是国家重点支持的高新技术企业，控股股东为地级市国资委，主要经营范围为复合材料、电子绝缘材料、覆铜板材料的生产、销售、技术开发及技术服务，经营进出口业务，自有房屋租赁。该公司 2018 年末总资产为 200 百万元，负债率 15%，净资产收益率为 18%，经营活动现金流入净额为净利润的 120%。过去 3 年销售收入复合增长率为 30%，净利润复合增长率为 28%。公司计划于 2019 年发行 10 亿元 5 年期的公司债。该公司聘请你所在的信用评级机构对其主体进行信用评级，该公司合理的主体信用评级水平为 AA+级。公司愿意为 AAA+级付出 40 万元评级费用，为 AAA 级付出 30 万元评级费用，为 AAA-级付出 20 万元评级费用，为 AA+付出 15 万元评级费用，为 AA 付出 9 万元评级费用。如果公司在债券存续期间出现违约，则你所在的信用评级机构会被监管机构约谈，要求提交评级资料，并进行整改。

C 股份有限公司成立于 2002 年，由某国资控股公司等股东共同投资，于 2010 年在深圳证券交易所注册上市，公司主要经营范围为无尘无菌净化系统，设备及其周边机电、仪控产品的生产、组装，并提供设计、咨询、调试、维修服务。该公司 2018 年末总资产为 201 百万元，负债率 16%，净资产收益率为 17%，经营活动现金流入净额为净利润的 110%。过去 3 年销售收入复合增长率为 35%，净利润复合增长率为 30%。公司计划于 2019 年发行 10 亿元 5 年期的公司债。该公司聘请你所在的信用评级机构对其主体进行信用评级，该公司合理的主体信用评级水平为 AA+级。公司愿意为 AAA+级付出 41 万元评级费用，为 AAA 级付出 32 万元评级费用，为 AAA-级付出 22 万元评级费用，为 AA+付出 15 万元评级费用，为 AA 付出 10 万元评级费用。如果公司在债券存续期间出现违约，则你所在的信用评级机构会被监管约谈，要求提交评级资料，进行整改；如果发现评级虚高，则评级机构会被没收违约所得且处以 3 倍罚金。

D 控股集团有限公司成立于 2005 年，于 2010 年在上海证券交易所注册上市。公司主要经营范围为商品染料的生产，燃料及助剂、化工产品的技术开发、技术转让、技术服务及销售，以及上述产品的原辅材料、机械设备以及零配件、包装产品的销售。该公司 2018 年末总资产为 200 百万元，负债率 16%，净资产收益率为 17%，经营活动现金流入净额为净利润的 118%。过去 3 年销售收入复合增长率为 32%，净利润复合增长率为 28%。公司计划于 2019 年发行 10 亿元 5 年期的公司债。该公司聘请你所在的信用评级机构对其主体进行信用评级，该公司合理的主体信用评级水平为 AA+级。公司愿意为 AAA+级付出 54 万元评级费用，为 AAA 级付出 39 万元评级费用，为 AAA-级付出 29 万元评级费用，为 AA+付出 15 万元评级费用，为 AA 付出 9 万元评级费用。即使公司在债券存续期间出现违约，你所在的信用评级机构不会被监管机构警告或者处罚。

E 股份有限公司成立于 2002 年，于 2010 年在深圳证券交易所注册上市，公司主要经营范围为机械零部件的生产销售；汽车、内燃机、压缩机及摩托车零部件的生产销售、机床设备及配件的生产销售。该公司 2018 年末总资产为 199 百万元，负债率 15%，净资产收益率为 18%，经营活动现金流入净额为净利润的 115%。过去 3 年销售收入复合增长率为 33%，净利润复合增长率为 29%。公司计划于 2019 年发行 10 亿元 5 年期的公司债。该公司聘请你所在的信用评级机构对其主体进行信用评级，该公司合理的主体信用评级水平为 AA+级。公司愿意为 AAA+级付出 55 万元评级费用，为 AAA 级付出 40 万元评级费用，为 AAA-级付出 30 万元评级费用，为 AA+付出 15 万元评级费用，为 AA 付出 8 万元评级费用。如果公司在债券存续期间出现违约，则你所在的信用评级机构会被监管约谈，要求提交评级资料，进行整改；如果发现评级虚高，则评级机构会被没收违约所得且处以 3 倍罚金。

F 股份有限公司成立于 2002 年，于 2012 年在深圳证券交易所注册上市，是一家专注于新材料研发、生产和销售的高新技术产业，公司经营范围为 EVA 太阳能电池胶膜、太阳能电池背板、热熔胶膜的生产、销售，新材料、新设备、新能源的技术开发、进出口业务。该公司 2018 年末总资产为 201 百万元，负债率 52%，净资产收益率为 6.3%，经营活动现金流入净额为净利润的 10%。过去 3 年销售收入复合增长率为 3%，净利润复合增长率为 5%。公司计划于 2019 年发行 10 亿元 5 年期的公司债。该公司聘请你所在的信用评级机构对其主体进行信用评级，该公司合理的主体信用评级水平为 B+级。公司愿意为 BB+级付出 40 万元评级费用，为 BB 级付出 30 万元评级费用，为 BB-级付出 21 万元评级费用，为 B+级付出 15 万元评级费用，为 B 级付出 8 万元评级费用。即使公司在债券存续期间出现违约，你所在的信用评级机构不会被监管机构警告或者处罚。

G 股份有限公司成立于 2001 年，于 2011 年在上海证券交易所注册上市，主要经营范围是计算机软件开发、软件技术咨询服务、网络游戏创意策划、研发制作及商业化运营。该公司 2018 年末总资产为 202 百万元，负债率 50%，净资产收益率为 6%，经营活动现金流入净额为净利润的 11%。过去 3 年销售收入复合增长率为 2%，净利润复合增长率为 5.5%。公司计划于 2019 年发行 10 亿元 5 年期的公司债。该公司聘请你所在的信用评级机构对其主体进行信用评级，该公司合理的主体信用评级水平为 B+级。公司愿意为 BB+级付出 41 万元评级费用，为 BB 级付出 31 万元评级费用，为 BB-付出 22 万元评级费用，为 B+付出 15 万元评级费用，为 B 付出 9 万元评级费用。如果公司在债券存续期间出现违约，则你所在的信用评级机构会被监管约谈，要求提交评级资料，并进行整改。

H 股份有限公司成立于 2003 年，于 2012 年在深圳证券交易所注册上市，主要经营范围是通信及计算机软硬件技术开发、生产、销售；通信及计算机网络工程技术咨询、技术培训。该公司 2018 年末总资产为 200 百万元，负债率 49%，净资产收益率为 6.2%，经营活动现金流入净额为净利润的 9%。过去 3 年销售收入复合增长率为 2.5%，净利润复合增长率为 4.8%。公司计划于 2019 年发行 10 亿元 5 年期的公司债。该公司聘请你所在的信用评级机构对其主体进行信用评级，该公司合理的主体信用评级水平为 B+级。公司愿意为 BB+级付出 40 万元评级费用，为 BB 级付出 30 万元评级费用，为 BB-付出 20 万元评级费用，为 B+付出 15 万元评级费用，为 B 付出 10 万元评级费用。如果公司在债券存续期间出现违约，则你所在的信用评级机构会被监管约谈，要求提交评级资料，进行整改；如果发现评级虚高，则评级机构会被没收违约所得且处以 3 倍罚金。

I 股份有限公司成立于 2001 年，于 2010 年在上海证券交易所注册上市，主营舰船、舰船配套产品、海洋工程及装备、能源装备、交通装备和环保装备等机械电子设备的设计、研制。该公司 2018 年末总资产为 199 百万元，负债率 50%，净资产收益率为 6.4%，经营活动现金流入净额为净利润的 10%。过去 3 年销售收入复合增长率为 3%，净利润复合增长率为 5%。公司计划于 2019 年发行 10 亿元 5 年期的公司债。该公司聘请你所在的信用评级机构

对其主体进行信用评级。该公司合理的主体信用评级水平为 B+级。公司愿意为 BB+级付出 55 万元评级费用，为 BB 级付出 41 万元评级费用，为 BB-付出 30 万元评级费用，为 B+付出 15 万元评级费用，为 B 付出 9 万元评级费用。即使公司在债券存续期间出现违约，你所在的信用评级机构不会被监管机构警告或者处罚。

J 股份有限公司成立于 2001 年，于 2010 年在深圳证券交易所注册上市，主要经营范围是氢电池、锂电池等电池业务、新能源的技术开发、进出口业务。该公司 2018 年末总资产为 201 百万元，负债率 51%，净资产收益率为 6.5%，经营活动现金流入净额为净利润的 10%。过去 3 年销售收入复合增长率为 3%，净利润复合增长率为 5%。公司计划于 2019 年发行 10 亿元 5 年期的公司债。该公司聘请你所在的信用评级机构对其主体进行信用评级，该公司合理的主体信用评级水平为 B+级。公司愿意为 BB+级付出 54 万元评级费用，为 BB 级付出 42 万元评级费用，为 BB-付出 31 万元评级费用，为 B+付出 15 万元评级费用，为 B 付出 8 万元评级费用。如果公司在债券存续期间出现违约，则你所在的信用评级机构会被监管约谈，要求提交评级资料，进行整改；如果发现评级虚高，则评级机构会被没收违约所得且处以 3 倍罚金。

参考文献

- [1] 邓博文,曹廷贵.信用评级行业的监管与评级质量[J].国际金融研究,2016(03):40-50.
- [2] 顾小龙,施燕平,辛宇.风险承担与公司债券融资成本:基于信用评级的策略调整视角[J].财经研究,2017,43(10):134-145.
- [3] 何平,金梦.信用评级在中国债券市场的影响力[J].金融研究,2010(04):15-28.
- [4] 黄小琳,朱松,陈关亨.债券违约对涉事信用评级机构的影响——基于中国信用债市场违约事件的分析[J].金融研究,2017(03):130-144.
- [5] 寇宗来,盘宇章,刘学悦.中国的信用评级真的影响发债成本吗?[J].金融研究,2015(10):81-98.
- [6] 寇宗来,千茜倩,陈关亨.跟随还是对冲:发行人付费评级机构如何应对中债资信的低评级?[J].管理世界,2020,36(09):26-39.
- [7] 林晚发,何剑波,周畅,等.“投资者付费”模式对“发行人付费”模式评级的影响:基于中债资信评级的实验证据[J].会计研究,2017(09):62-68+97.
- [8] 王雄元,张春强.声誉机制、信用评级与中期票据融资成本[J].金融研究,2013(08):150-164.
- [9] 吴育辉,翟玲玲,张润楠等.“投资人付费”vs.“发行人付费”:谁的信用评级质量更高?[J].金融研究,2020(01):130-149.
- [10] 邢天才,詹明君,王文钢.评级机构竞争、声誉与债券信用评级质量[J].财经问题研究,2016(06):66-71.
- [11] 徐晓萍,阮永锋,刘音露.市场竞争降低评级质量了吗——基于新进入评级机构的实证研究[J].财贸经济,2018,39(11):96-111.
- [12] 詹明君,邢贺.评级机构竞争、声誉对债项评级的非对称影响[J].东北财经大学学报,2016(06):66-73.
- [13] 张继勋,蔡闫东,刘文欢.标准审计报告改进、管理层和审计人员的关系与管理层沟通意愿——一项实验证据[J].审计研究,2016(03):77-83.
- [14] 张继勋,倪古强,张广冬.关键审计事项的结论性评价与投资者的投资判断[J].会计研究,2019(07):90-96.
- [15] 张继勋,孙鹏,周冉.签字注册会计师轮换、会计师事务所轮换与审计谈判[J].当代财经,2010(03):109-119.
- [16] 周宏,温笑天,夏剑超等.评级机构数量选择对企业债券信用风险监管的影响——基于评级机构与发债企业串谋行为的博弈分析[J].会计研究,2013(08):76-80+97.
- [17] 周宏,温笑天,夏剑超等.评级机构数量选择对企业债券信用风险监管的影响——基于评级机构与发债企业串谋行为的博弈分析[J].会计研究,2013(08):76-80+97.
- [18] 周香芸,田益祥,李立等.新评级监管对中国企业债券信用评级的声誉效应——基于新评级选购博弈模

型[J].金融季刊,2019,13(01):151-175.

- [19] Afonso A, Gomes P, Rother P. What “Hides” Behind Sovereign Debt Ratings?[R]. European Central Bank Working Paper Series, 2007, 711: 1-31.
- [20] Ahmed A S, Billings B K, Morton R M, Stanford-Harris M. The Role of Accounting Conservatism in Mitigating Bondholder-shareholder Conflicts over Dividend Policy and in Reducing Debt Costs[J]. Accounting Review, 2002, 77(4): 867-890.
- [21] Becker B, Milbourn T. How Did Increased Competition Affect Credit Ratings?[J]. Journal of Financial Economics, 2011, 101(3): 493-514.
- [22] Bolton P, Freixas X, Shapiro J. The Credit Rating Game[J]. Journal of Finance, 2012, 67(1): 85-111.
- [23] Cheng M, Neamtiu M. An Empirical Analysis of Changes in Credit Rating Properties: Timeliness, Accuracy and Volatility. Journal of Accounting and Economics, 2009, 47(1-2): 108-130.
- [24] Dimitrov V, Palia D, Tang L. Impact of the Dodd-Frank Act on Credit Ratings[J]. Journal of Financial Economics, 2015, 115(3): 505-520.
- [25] Geol A M, Thakor A V. Information Reliability and Welfare: A Theory of Coarse Credit Ratings[J]. Journal of Financial Economics, 2015, 115(3): 541-557.
- [26] Graham J R, Harvey C R. The Theory and Practice of Corporate Finance: Evidence from the Field[J]. Journal of Financial Economics, 2001, (60): 187-243.
- [27] He Y, Tan H, Yeo F, Zhang J. When Do Qualitative Risk Disclosures Backfire? The Effects of a Mismatch in Hedge Disclosure Formats on Investors’ Judgments[J]. Contemporary Accounting Research, 2019,36(4): 2093-2112.
- [28] Heski I, Shapiro J. Rating Quality over the Business Cycle[J]. Journal of Financial Economics, 2012, 108(01): 62-78.
- [29] Jiang J, Stanford M H, Xie Y. Does It Matter Who Pays for Bond Ratings? Historical Evidence[J]. Journal of Financial Economics, 2012, 105(3): 607-621.
- [30] Kisgen D J, Strahan P E. Do Regulations Based on Credit Ratings Affect a Firm’s Cost of Capital?[J]. Review of Financial Studies, 2010, 23(12): 4324-4347.
- [31] Libby R, Bloomfield R, Nelson M W. Experimental Research in Financial Accounting[J]. Accounting, Organizations and Society, 2002, 27(8): 775-810.
- [32] Mathis J, McAndrews J, Rochet J. Rating the Raters: Are Reputation Concerns Powerful Enough to Discipline Rating Agencies?[J]. Journal of Monetary Economics, 2009, 56(5): 657-674.
- [33] Opp C C, Opp M M, Harris M. Rating Agencies in the Face of Regulation[J]. Journal of Financial Economics, 2013, 108(1): 46-61.
- [34] Skreta V, Veldkamp L. Ratings Shopping and Asset Complexity: A Theory of Rating Inflation[J]. Journal of Monetary Economics, 2009, 56(5): 678-695.