

带有辅助信息对数变换模型的违约预报

王莫明¹, 周勇^{2*}

1. 上海财经大学统计与管理学院, 上海 200433;

2. 华东师范大学统计交叉科学研究院和统计学院, 统计与数据科学前沿理论及应用教育部重点实验室, 上海 200062

E-mail: wangmoming0902@163.com, yzhou@amss.ac.cn

收稿日期: 2018-09-08; 接受日期: 2019-02-21; 网络出版日期: 2019-08-01; * 通信作者

国家自然科学基金 (批准号: 71331006)、国家自然科学基金重大研究计划 (批准号: 91546202) 和上海财经大学研究生创新计划项目科研创新基金 (批准号: CXJJ-2018-411) 资助项目

摘要 本文考虑对数变换的逻辑模型以刻画不同的违约概率曲线, 研究如何将辅助信息加入到模型的估计中以提高违约估计的稳定性和效率. 通过非参数经验似然, 提出模型参数统计推断方法, 并推导估计的相合性和渐近正态性. 从理论上证明添加了辅助信息的估计的有效性, 并且模拟表明该方法能够很好地提升估计的效率, 另外也通过模拟讨论辅助信息的影响. 将所提出的方法应用于 ST (special treatment) 股票的数据, 实证结果表明, 加入了辅助信息的参数估计更加有效.

关键词 违约 辅助信息 信用风险 经验似然 对数变换

MSC (2010) 主题分类 62F10, 62G20, 62P05

1 引言

一直以来, 违约预报和信用风险都是投资者、研究者和各监管机构的重点研究对象. 市场的监管者利用违约预报来对市场上的企业进行监管; 市场的参与者通过违约模型来调整投资策略; 行业的研究者通过对企业违约的预测决定企业的评级; 学者意图通过模型找到并解释企业违约与各类经济因素之间的关系. 另一方面, 企业违约也是部分金融衍生品 (如信用互换) 定价的重要因素. 企业的违约不仅会给投资者造成巨大的损失, 还会扰乱市场秩序从而阻碍经济健康发展, 因此, 准确的违约预报对金融市场有着重要的作用.

衡量企业违约有许多有用的风险指标, 如衡量极端损失的在险价值 (value at risk) 和衡量期望风险的预期不足 (expected shortfall), 相关的研究可以参见文献 [1, 2]. 相对于在险价值、预期不足等间接指标, 利用违约概率研究违约现象更加直接, 应用更加广泛. 大量文献围绕违约的可能性开展研究. Altman^[3] 提出了基于五个经济指标的 Z 得分公式模型方法来评价企业违约. 随后又在文献 [4] 中对模型进行修正, 进一步提出了 ZETA 模型, 该模型至今仍受众多金融机构的青睐. Altman 和 Saunders^[5]

英文引用格式: Wang M M, Zhou Y. Default forecast with auxiliary information using a logarithmic transformation model (in Chinese). *Sci Sin Math*, 2021, 51: 513–534, doi: 10.1360/SCM-2018-0622

对具体的企业违约状况进行了讨论. Horrigan^[6] 和 Misund^[7] 也研究过类似的问题. 以上的研究并没有具体给出协变量与违约概率之间显示表达式, 一些学者从量化模型的角度进行了研究. Olhson^[8] 利用假设条件更加宽松的逻辑分析对违约概率进行了预测. 另外, Tserng 等^[9] 通过建立逻辑模型来研究企业的破产概率. 对比了仅依靠财务比率的单因素逻辑模型与加入市场因素如变现能力 (liquidity)、杠杆作用 (leverage)、活动性 (activity) 和收益率 (profitability) 后的多因素模型, 结果表明, 加入市场因素后能够较好地提高预测的准确性. 另外一些文献借助其他的参数模型来刻画违约概率 (参见文献 [10-13]). 也有部分学者利用非参数模型来研究违约概率, 例如, Hong 等^[14] 利用支撑向量集作为分类器, 来研究破产预报. Weißbach 和 Walter^[15] 通过构造卡方统计量来检验强度函数是否有跳跃. Tian 等^[16] 考虑了变量筛选的问题并利用 LASSO (least absolute shrinkage and selection operator) 来选择对破产预报有显著影响的变量. Pinches 和 Mingo^[17] 以及 Pogue 和 Soldofsky^[18] 讨论了债券评级的问题. 也有部分文献考虑了公司之间的交互作用. 周勇等^[19] 在简约模型下, 应用多种生存分析方法对违约概率进行了估计, 最后得出了综合因素的动态性和交互作用的估计. 谢尚宇等^[20] 考虑了违约公司的传染性, 研究了金融危机下, 带有传染机制的违约概率的估计问题.

在实际应用中, 研究者通常会掌握一些关于违约概率的辅助信息, 这样的辅助信息可能来源于之前研究的结论, 例如, 之前曾全面研究过某些经济指标 (如杠杆率、收盘价) 对违约的影响; 也有可能来源于其他的研究机构的一些概括性信息, 如不同的盈利能力下企业违约的频率的统计; 也有可能是来源于一些日常的总结性工作, 如债券交易员日常的统计数据或研究所的研究报告等. 这样的信息是非常大量的. 以 Wind 数据库为例, 每天的研究报告就高达上千篇, 而此类研究报告中往往包含了大量的经验的在不同情形下的条件违约概率. 一般来说, 这样的辅助信息往往不包含具体的个体数据, 而是以总结性的数据形式出现, 如均值、比例等, 大多都可以转化为条件概率的形式. 如何利用这些辅助信息提高研究结果是本文的重点研究内容.

虽然市场的企业很多, 但真正违约的个体较少, 得到的关于违约的信息也就较少, 故而利用辅助信息来提高估计的效率具有重要的意义. 在大数据时代的今天, 单个数据处理中心可能很难处理如此大量的数据. 一个简单的想法便是将数据分成两份, 将其中一份用作核心研究, 将另一份的信息通过一定的方式加以整合. 不同数据中心之间由于数据存储、隐私保护等原因, 共享的数据往往是一些简要的数据, 本文将研究如何利用这些简要的数据对估计进行提升. 辅助信息在生存分析、荟萃分析 (meta-analysis) 等领域有成功的应用. Qin 等^[21] 曾利用辅助信息在逻辑模型下来增强检验的势. Tarima 和 Pavlov^[22] 也通过辅助信息来提高经验分布函数、经验特征函数和风险函数的估计. Huang 等^[23]、Chen 和 Qin^[24]、Hellerstein 和 Imbens^[25] 以及 Liu 等^[26,27] 均做过类似研究. 研究结果表明辅助信息能够有效地提高估计的效率. 本文致力于利用辅助信息来改进违约概率的估计, 利用估计方程将辅助信息加入到参数估计中以提高估计的效率; 另外, 通过理论结果和实证分析还可以看到辅助信息可以将原本不可估的参数估计出来, 以得到各影响因素与违约概率的解析表达式. 考虑到违约的个体比例较少, 本文利用病例 - 对照 (case-control) 抽样, 即分别在违约个体中抽取样本和在未违约个体中抽取样本. 这样的研究框架可以调节违约个体与未违约个体数量比例的严重失衡, 解决经济研究中回顾性数据类型的问题. 在金融市场中, 可能会出现当经济指标处于一定范围内违约的概率很小, 在超过某个值时, 违约概率剧增; 又或是一直处于较缓慢的状态, 如极其稀有事件. 本文在逻辑模型的基础上引入对数变化, 通过调节参数以捕捉上述现象. 对数变化是一个单调增变化, 并不改变各企业的评价指标下对应违约风险的相对大小, 但却能调节不同取值下违约概率的绝对距离, 从而使得曲线更加陡峭或者更加平缓. 调节参数的大小决定了违约概率曲线的陡峭程度, 使违约概率增长幅度最大的点是监管机构控制的最有效节点, 违约概率曲度的刻画不仅对理解企业违约与各协变量之间的关系有积极作用, 还对相

关政策有参考价值. 模拟的结果表明, 引入的参数可以刻画原本逻辑模型无法刻画的数据特征. 另外, 基于建立的模型, 可以给出企业的信用得分, 其取值是各变量的线性组合.

本文余下部分的组织如下. 第 2 节介绍研究的模型和估计的方法, 展示如何将辅助信息应用到估计中并从理论上给出带有辅助信息的估计的渐近正态性和有效性. 第 3 节给出与理论对应的模拟结果, 并进一步讨论方法的稳健性和辅助信息对估计结果的影响. 第 4 节将提出的方法应用于 ST 股票的数据. 结论在第 5 节中给出. 附录 A 给出定理的相关证明.

2 模型与理论结果

本节讨论对数变换的逻辑模型及其性质. 在该模型的基础上, 将辅助信息应用到病例 - 对照抽样得到的样本中. 本节给出提出的估计的渐近正态性, 并比较带辅助信息的估计和不带辅助信息的估计的方差.

2.1 模型

对于给定的研究时刻, 令 $D = 1$ 表示企业的违约状态, $D = 0$ 表示非违约状态, \mathbf{X} 和 \mathbf{Y} 表示风险因素. 将风险因素分为两部分的原因是, 实际中, 我们得到的辅助信息可能只是关于部分风险因素而非所有风险因素的, 故而引入不同的记号以区分具有辅助信息的自变量 \mathbf{X} 和不具有辅助信息的自变量 \mathbf{Y} . 在风险管理中, 我们感兴趣的是给定协变量下的违约概率, 即 $\Pr(D = 1 | \mathbf{X}, \mathbf{Y})$. 与违约概率这一绝对数相比, 我们同时也关注违约概率与未违约概率之比, 即 $\Pr(D = 1 | \mathbf{X}, \mathbf{Y}) / \Pr(D \neq 1 | \mathbf{X}, \mathbf{Y})$ 这一相对数指标, 即常用的优势比 (odds). 当优势比大于 1 时, 企业违约概率大于不违约概率. 一个常用的基于优势比的模型便是逻辑模型. 众所周知, 逻辑模型具有如下形式:

$$\frac{\Pr(D = 1 | \mathbf{X}, \mathbf{Y})}{1 - \Pr(D = 1 | \mathbf{X}, \mathbf{Y})} = e^{\alpha + \mathbf{Z}^T \boldsymbol{\beta}},$$

其中 $\mathbf{Z} = (\mathbf{X}^T, \mathbf{Y}^T)^T$, \mathbf{T} 表示矩阵或向量的转置. 为更好地刻画当风险因素在超过某一临界值后违约风险剧增或缓慢上升的现象, 本文引入带有参数的转换函数 G_λ , 使得转换后的优势比满足逻辑模型, 即

$$G_\lambda \left(\frac{\Pr(D = 1 | \mathbf{X}, \mathbf{Y})}{1 - \Pr(D = 1 | \mathbf{X}, \mathbf{Y})} \right) = e^{\alpha + \mathbf{Z}^T \boldsymbol{\beta}}, \quad (2.1)$$

其中 $G_\lambda(y) = \frac{\log(\lambda y + 1)}{\lambda}$, $\lambda \neq 0$, 并且令 $G_0(y) = y$, 即 λ 趋于 0 的极限. 当 $\lambda = 0$ 时, 模型即逻辑模型. 调节参数 λ 能够调节真实的优势比的曲率以使调节后的模型满足传统的逻辑模型. Li 等^[28] 也曾通过引入调节参数以得到更灵活的连接函数. Chen 等^[29] 及 Zeng 和 Lin^[30] 也研究过类似的转换问题.

图 1 给出在单一协变量 x 下, 不同 λ 对应的优势比曲线. 由图形可以看出, 当 $\lambda > 0$ 时, 模型的优势比曲线比逻辑模型更加陡峭, 不同的 λ 的取值对应的陡峭程度不同. 当 $\lambda < 0$ 时, 对应的优势比曲线相对于逻辑模型更加平缓. 前文中曾提及到, 虽然市场上的企业很多, 但违约的个体却较少. 为了减小随机抽样中违约个体与非违约个体之间比例的严重失调, 并考虑到金融数据往往已经发生, 本文采取回顾性研究中的病例 - 对照抽样方法, 即分别在违约和未违约的个体中进行抽样, 并在此基础上进行进一步的理论研究. 通过简单抽样, 我们得到来源于同一总体的样本 $(D, \mathbf{X}, \mathbf{Y})$, 而通过病例 - 对照抽样的样本可能来自于两个总体, 即 $(\mathbf{X}, \mathbf{Y} | D = 1)$ 和 $(\mathbf{X}, \mathbf{Y} | D = 0)$, 有偏抽样的相关讨论可以参

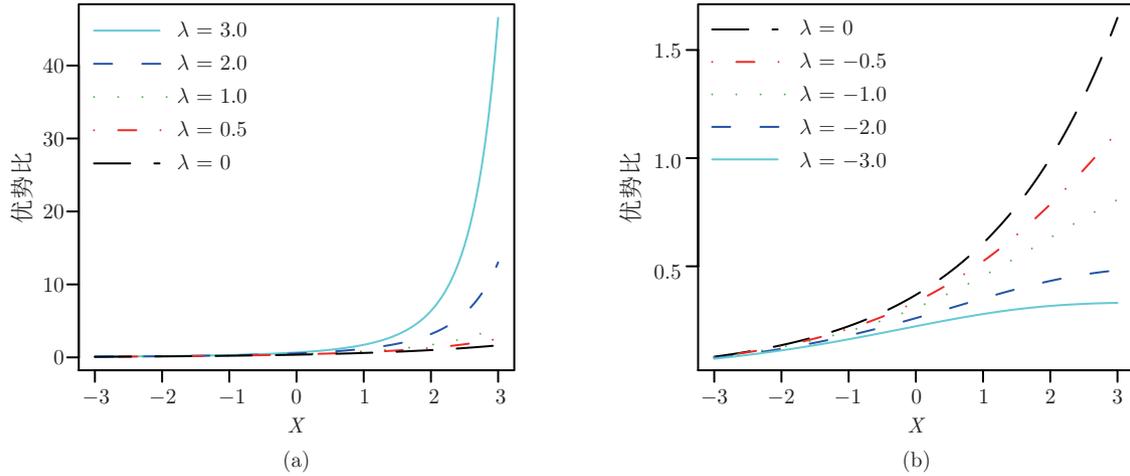


图 1 (网络版彩图) 单一自变量下, 不同调节参数下对应的优势比曲线, 其中 $\lambda = 0$ 为逻辑模型, 参数设置为 $\alpha = -1, \beta = 0.5$

见文献 [31, 32]. 下面将违约的组称为事件组, 将未违约的组称为控制组, n_1 和 n_0 分别表示事件组和控制组抽样的样本量. 由此可以得到样本的对数似然函数为

$$\prod_{i=1}^{n_1} f(\mathbf{x}_i, \mathbf{y}_i | D = 1) \prod_{j=1}^{n_0} f(\mathbf{x}_j, \mathbf{y}_j | D = 0). \quad (2.2)$$

注意到, 在模型 (2.1) 下, 协变量与违约概率之间具有如下关系:

$$\Pr(D = 1 | \mathbf{X}, \mathbf{Y}) = \begin{cases} 1 - \frac{1}{1 + \frac{1}{\lambda} \{\exp[\lambda \exp(\alpha + \beta^T \mathbf{Z})] - 1\}}, & \lambda \neq 0, \\ 1 - \frac{1}{1 + \exp(\alpha + \beta^T \mathbf{Z})}, & \lambda = 0. \end{cases} \quad (2.3)$$

利用 Bayes 公式, 可以得到

$$f(\mathbf{x}, \mathbf{y} | D = 1) = w(\mathbf{Z} | \lambda^*, \alpha^*, \beta) f(\mathbf{x}, \mathbf{y} | D = 0), \quad (2.4)$$

其中 $\pi = \Pr(D = 1)$, $k = \pi / (1 - \pi)$, $\lambda^* = k\lambda$, $\alpha^* = \alpha - \log k$,

$$w(\mathbf{Z} | \lambda^*, \alpha^*, \beta) = \begin{cases} \frac{1}{\lambda^*} \{\exp[\lambda^* \exp(\alpha^* + \mathbf{Z}^T \beta)] - 1\}, & \lambda^* \neq 0, \\ \exp(\alpha^* + \mathbf{Z}^T \beta), & \lambda^* = 0. \end{cases}$$

注意到, 对于使用病例 - 对照抽样得到的样本, 损失了对总体违约率的不可识别性. 在实际应用中, 该参数可以很简单地获得较好的估计, 如频率估计, 故而并不会对研究造成实质性阻碍. 另外, 在后文中我们可以发现, 添加了辅助信息以后, 违约概率 π 或等价的参数 k 是可识别的, 故而在具有辅助信息的情形下, 可以直接得到 (2.3) 的具体表达式. 为了简化记号, 令 $\theta_2 = (\lambda^*, \alpha^*, \beta^T)^T$, 则对数似然函数 (2.2) 可以简化为

$$l(\theta_2) = \sum_{i=1}^n D_i \log w(Z_i | \theta_2) + \sum_{i=1}^n \log p_i, \quad (2.5)$$

其中 $n = n_1 + n_0$ 为总的样本量, $D_i = 1$ ($i = 1, \dots, n_1$), $D_i = 0$ ($i = n_1 + 1, \dots, n$), $p_i = dF_0(\mathbf{X}_i, \mathbf{Y}_i)$, F_0 是对照组下 (\mathbf{X}, \mathbf{Y}) 的累积分布函数在指定点的取值.

2.2 辅助信息及大样本性质

假设拥有协变量 \mathbf{X} , 如公司类型、杠杆水平和收盘价格等与违约相关的因素的辅助信息. 假定辅助信息具有形式

$$\Pr(D = 1 | \mathbf{X} \in A) = h(A), \quad (2.6)$$

其中 A 是已知的区域. A 与对应的 $h(A)$ 的取值即为辅助信息. 形如 (2.6) 的总结性信息在市场上是常见的, 例如, 从业者日常统计工作中往往包含了不同评级、杠杆率、收盘价、市值下公司的违约概率. 另外, 众多的研究报告往往也以不同类别、不同等级等变量作为划分来统计企业的违约情形. 大部分的研究结果也可以转化为 (2.6) 的形式. 这样的数据相对粗糙, 占用的空间较小, 利于数据的存储和不同中心之间数据交互. 若 $A = \Omega_{\mathbf{X}}$ 为变量 \mathbf{X} 的所有可能取值构成的空间, 则实际给出的辅助信息为 $\pi = \Pr(D = 1)$, 即整个市场的违约概率. 为了使用辅助信息, 类似于文献 [21], 我们将辅助信息转化为一系列的估计方程, 再设法将估计方程融入到参数估计中. 利用 Bayes 公式, 可以将 (2.6) 转化为

$$E_1\{I_{(\mathbf{X} \in A)}\} = \frac{1 - \pi}{\pi} \frac{h(A)}{1 - h(A)} E_0\{I_{(\mathbf{X} \in A)}\},$$

其中 E_i 表示在 $D = i$ 下求期望. 利用 (2.4), 上式可以转化为

$$E_0\left\{I_{(\mathbf{X} \in A)} \left[w(\mathbf{Z} | \lambda^*, \alpha^*, \beta) - \frac{1 - \pi}{\pi} \frac{h(A)}{1 - h(A)} \right] \right\} = 0.$$

令 $\boldsymbol{\theta} = (k, \boldsymbol{\theta}_2^T)^T = (k, \lambda^*, \alpha^*, \beta^T)^T$, 假设拥有 L 个形如 (2.6) 的辅助信息方程, 即对 $\Omega_{\mathbf{X}}$ 的 L 个互不相交的子集 A_1, \dots, A_L , 有对应的 $h(A_1), \dots, h(A_L)$. 令

$$\boldsymbol{\psi}(\mathbf{Z} | \boldsymbol{\theta}) = (\psi^1(\mathbf{Z} | \boldsymbol{\theta}), \dots, \psi^L(\mathbf{Z} | \boldsymbol{\theta}))^T,$$

其中

$$\psi^i(\mathbf{Z} | \boldsymbol{\theta}) = I_{(\mathbf{X} \in A_i)} \left\{ k w(\mathbf{Z} | \boldsymbol{\theta}_2) - \frac{h(A_i)}{1 - h(A_i)} \right\}, \quad i = 1, 2, \dots, L, \quad (2.7)$$

则有

$$E_0\{\boldsymbol{\psi}(\mathbf{Z} | \boldsymbol{\theta})\} = 0.$$

注意到 $\boldsymbol{\psi}(\mathbf{Z} | \boldsymbol{\theta})$ 各参数是可识别的, 故而在有辅助信息下, 参数的估计为在满足限制

$$p_i \geq 0, \quad \sum_{i=1}^n p_i = 1, \quad \sum_{i=1}^n p_i \{w(\mathbf{Z}_i | \boldsymbol{\theta}_2) - 1\} = 0 \quad \text{以及} \quad \sum_{i=1}^n p_i \boldsymbol{\psi}(\mathbf{Z}_i | \boldsymbol{\theta}) = 0$$

下极大化对数似然函数

$$l(\boldsymbol{\theta}) = \sum_{i=1}^n D_i \log w(\mathbf{Z}_i | \boldsymbol{\theta}) + \sum_{i=1}^n \log p_i \quad (2.8)$$

对应的取值. 第一个与第二个限制来源于 p_i 作为分布函数的增量的性质. 第三个限制是由两组中抽取样本之间满足密度比模型的关系所决定. 最后一个限制来源于辅助信息. 在没有辅助信息时, 对 $\boldsymbol{\theta}_2$ 的

估计即满足前三个限制条件下求对数似然函数的最大值. 一个求解上面优化问题可能的方法就是使用统计学中的非参数经验似然方法. 最近, Qin 和 Lawless^[33] 将经验似然^[34] 和无偏估计方程结合进行了研究. 类似地, 我们将沿用 Qin 和 Lawless^[33] 的方法对 (2.8) 在满足相应限制条件 (2.7) 下进行求解. 对没有辅助信息的估计也可以采用类似的方法所得到. 利用 Lagrange 乘子法, 可以得到 p_i 的估计为

$$\hat{p}_i = n^{-1}[1 + \xi[w(\mathbf{Z}_i | \boldsymbol{\theta}) - 1] + \mathbf{t}^T \psi(\mathbf{Z}_i | \boldsymbol{\theta})]^{-1}.$$

故而关于 $\boldsymbol{\theta}$ 的截面似然函数为

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n D_i \log w(\mathbf{Z}_i | \boldsymbol{\theta}) - \sum_{i=1}^n \log[1 + \xi(w(\mathbf{Z}_i | \boldsymbol{\theta}) - 1) + \mathbf{t}^T \psi(\mathbf{Z}_i | \boldsymbol{\theta})],$$

其中对应的 Lagrange 乘子 ξ 和 \mathbf{t} 由下面的方程组所决定:

$$\begin{aligned} \sum_{i=1}^n \frac{w(\mathbf{Z}_i | \boldsymbol{\theta}) - 1}{1 + \xi[w(\mathbf{Z}_i | \boldsymbol{\theta}) - 1] + \mathbf{t}^T \psi(\mathbf{Z}_i | \boldsymbol{\theta})} &= 0, \\ \sum_{i=1}^n \frac{\psi(\mathbf{Z}_i | \boldsymbol{\theta})}{1 + \xi[w(\mathbf{Z}_i | \boldsymbol{\theta}) - 1] + \mathbf{t}^T \psi(\mathbf{Z}_i | \boldsymbol{\theta})} &= 0. \end{aligned}$$

不妨令 $\boldsymbol{\theta}$ 的真值为 $\boldsymbol{\theta}_0$. 可以证明在有辅助信息下 ξ 的估计 $\hat{\xi}$ 和在没有辅助信息下 ξ 的估计 $\tilde{\xi}$ 取值均为 n_1/n , 详见附录 A. 令

$$r = \frac{n_1}{n_0}, \quad g(\mathbf{Z} | \boldsymbol{\theta}) = (1 + r) \frac{\psi(\mathbf{Z} | \boldsymbol{\theta})}{\eta(\mathbf{Z} | \boldsymbol{\theta})}, \quad \eta(\mathbf{Z} | \boldsymbol{\theta}) = 1 + rw(\mathbf{Z} | \boldsymbol{\theta}).$$

对于提出的带有辅助信息的估计 $\hat{\boldsymbol{\theta}}$, 我们有如下的定理和推论.

定理 2.1 假设 $E[gg^T/\eta^2 | \boldsymbol{\theta}_0]$ 是正定的. $\partial g(\mathbf{Z} | \boldsymbol{\theta})/\partial \boldsymbol{\theta}$ 在真值 $\boldsymbol{\theta}_0$ 附近连续, $\|\partial g(\mathbf{Z} | \boldsymbol{\theta})/\partial \boldsymbol{\theta}\|$ 以及 $\|g(\mathbf{Z}, \boldsymbol{\theta})\|^3$ 在这个邻域内被可积函数 $G(\mathbf{Z})$ 控制 (在两个组中均有). 另外, $E[\partial g(\mathbf{Z} | \boldsymbol{\theta})/\partial \boldsymbol{\theta} | \boldsymbol{\theta}_0]$ 的秩为 L . 则含有辅助信息的估计 $\hat{\boldsymbol{\theta}}$ 以及不含有辅助信息的估计 $\tilde{\boldsymbol{\theta}}$, 随着 $n \rightarrow \infty$, 在真值的 $n^{-1/3}$ 邻域内以概率 1 有解.

以上的定理表明, 带有辅助信息的估计和不带辅助信息的估计均是相合的, 下面的定理给出带有辅助信息下参数估计的渐近正态性和方差的具体表达式.

定理 2.2 假设 r 是一个常数, 随着 $n \rightarrow \infty$, 基于辅助信息得到的估计具有如下的渐近结果:

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \rightarrow N(0, \boldsymbol{\Sigma}),$$

其中

$$\boldsymbol{\Sigma} = \mathbf{J}^{-1} - \frac{(1+r)^2}{r} (-k_0, -\lambda_0^*, 1, \mathbf{0}_p^T)^T (-k_0, -\lambda_0^*, 1, \mathbf{0}_p^T),$$

$\mathbf{J} = \mathbf{U}\mathbf{V}^{-1}\mathbf{U}^T$, \mathbf{V} 和 \mathbf{U} 的具体表达式在 (A.11) 和 (A.15) 中给出.

由定理 2.2 可见, 参数的渐近方差具有三明治方差的形式. 基于上述两个定理, 可以得到以下的推论, 该推论比较了带辅助信息下参数估计和不带辅助信息下参数估计的方差.

推论 2.1 令 $\tilde{\boldsymbol{\theta}}_2$ 是不带辅助信息下基于病例 - 对照样本的估计结果, $\boldsymbol{\theta}_{20} = (\lambda_0^*, \alpha_0^*, \boldsymbol{\beta}_0^T)^T$ 为 $\tilde{\boldsymbol{\theta}}_2$ 的真值, $\hat{\boldsymbol{\theta}}_2 = (\hat{\lambda}^*, \hat{\alpha}^*, \hat{\boldsymbol{\beta}}^T)^T$ 为带有辅助信息下的估计. 在定理 2.2 的假设条件下, 有

- (1) 当辅助信息个数 $L = 1$ 时, $\sqrt{n}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{20})$ 与 $\sqrt{n}(\tilde{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{20})$ 具有相同的渐近方差;
- (2) 当辅助信息个数 $L > 1$ 时, $\sqrt{n}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{20})$ 与 $\sqrt{n}(\tilde{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{20})$ 渐近方差的差距是非正定矩阵;
- (3) 当辅助信息个数 $L > 1$ 时, $\sqrt{n}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_{20})$ 的方差不会随着估计方程个数的减小而减小.

上述推论比较了带有辅助信息和不带有辅助信息的估计的方差矩阵. 理论结果表明, 二者之差是非正定矩阵, 这意味着带有辅助信息的估计的方差可能“小于”不带辅助信息的方差. 在模拟部分和实证研究部分, 我们将以具体的数据来展示具体的提升效果.

注 2.1 注意到当估计具有辅助信息时, 即 $L \geq 1$, 参数 k 是可识别的, 即总体的违约概率 π 是可识别的, 这一点在没有辅助信息下是不能做到的, 故而在没有辅助信息下, 不能直接得到模型的显式表达式 (2.3). 另外, 当 $L > 1$ 时, 方程的个数大于未知参数的个数, 多余的辅助信息将用于提高估计参数的效率. 但由于辅助信息是关于违约概率与部分协变量 \mathbf{X} 的而非 \mathbf{Y} 的, 故而效率的提升主要体现在 \mathbf{X} 对应的系数上, 在模拟部分我们也可以看到这一点. 更详细的讨论可以参见文献 [21]. 辅助信息的个数 L 大于 1 意味着辅助信息中包含了某些风险因素处于 L 个不同状态下总体的违约率的信息. L 越大表明得到的辅助信息对某些协变量与违约的关系刻画得越全面 (当 L 趋于无穷时, 辅助信息趋于协变量 \mathbf{X} 给定下是否违约的条件分布), 倾向于得到更有效的估计.

注 2.2 估计的理论方差的推导要求辅助信息 $h(A)$ 的取值是准确的并且是已知的. 然而, 实际应用中, 该值可能是基于一定方法估计得到的, 对应的理论方差不再满足定理. 对于模拟部分, 我们基于估计的辅助信息探讨了辅助信息质量对估计结果的影响. 模拟结果表明, 当辅助信息的数据来源样本量较小 (约 2 倍于当前研究数据量) 时, 辅助信息的准确性较低, 估计的方差仍有所下降, 但不具有辅助信息的变量对应的参数估计的偏差可能增大, 从均方误差来看, 误差仍旧可控, 本文提出的估计方法仍可用. 过于不精确的辅助信息会对算法提出一定的挑战, 相关的理论和修偏的方法本文不再深入, 将在后续的工作中进行补充. 当辅助信息来源的样本量较大 (约 10 倍于当前研究数据量) 时, 估计结果的提升效果明显, 与基于真实的辅助信息的估计结果类似.

3 模拟

本节首先给出一些与定理假设相符合的模拟结果, 以对比加入了辅助信息和未加入辅助信息的估计效率. 其次, 通过模拟讨论辅助信息的各种来源和质量对估计结果的影响. 模型的稳定性与逻辑模型比较也在本节给出. 考虑二元的变量, 分别从密度函数为 $f(x, y | D = 1)$ 的总体中抽取样本量为 n_1 的样本 (x_i, y_i) , $i = 1, \dots, n_1$, 从密度函数为 $f(x, y | D = 0)$ 的总体中抽取样本量为 n_0 的样本 (x_i, y_i) , $i = n_1 + 1, \dots, n = n_1 + n_0$. (x, y) 的无条件分布满足正态分布, 具有以下的均值和协方差结构:

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N \left(\begin{pmatrix} 0.5 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{xy} \\ \rho_{xy} & 1 \end{pmatrix} \right), \quad (3.1)$$

其中 ρ_{xy} 的取值在具体模拟中给出. 我们感兴趣的是, 在给定协变量下, 因变量等于 1 的概率即

$$\Pr(D = 1 | x, y) = 1 - \frac{1}{1 + \frac{1}{\lambda} \{ \exp[\lambda \exp(\alpha + \beta_1 x + \beta_2 y)] - 1 \}}.$$

本文选取参数 $\alpha = -2.5$, $\lambda = 1.5$, $\beta_1 = -1$, $\beta_2 = 0.5$. 考察与第一个变量 X 相关的辅助信息, 即假设我们知道在风险因素 X 落入一系列具有形式 $(a, b]$ 的区间对应的 $D = 1$ 的概率. 类似于文献 [21], 选取四个区间分别是 $(-\infty, -1.67\sigma_x + \mu_x]$ 、 $(-1.67\sigma_x + \mu_x, \mu_x]$ 、 $(\mu_x, 1.67\sigma_x + \mu_x]$ 和 $(1.67\sigma_x + \mu_x, \infty)$, 其中 σ_x 和 μ_x 分别是变量 X 的均值和标准差, 而 ± 1.67 是标准正态分布的上下四分之一分位数. 在上述模型下, 本文采取拒绝抽样法, 在违约的总体和未违约的总体中分别抽取 50 和 2,000 个样本.

这样的设计是考虑到实际的金融市场中发生违约的企业占比较小. 在 ρ_{xy} 取值不同的情形下, 参数估计结果见表 1. 表 1 展示了不同方法的估计结果的偏差 (BIAS)、方差 (SSD)、均方误差 (MSE)、估计方差 (ESD)、正态的 95% 置信区间的覆盖概率 (CP) 以及辅助信息与未带辅助信息下估计的均方误差的比值 (MSER). 由表 1 的结果可以看到, 带有辅助信息的估计会比不带辅助信息的估计与真实值之间距离更近, 这表明加入辅助信息以后, 估计的偏差减小, 准确度提高. 从估计的方差来看, 带有辅助信息的估计的方差明显小于不带辅助信息的方差. 参数 β_2 估计的精度提升要小于 β_1 的提升, 这源于辅助信息与 X 相关, 故辅助信息中包含了更多对 β_1 估计的信息. 另外, 模拟结果表明, 在不包含辅助信息下, 估计的效果更容易受自变量之间的相关性影响.

当转换函数中的调节参数 $\lambda = 0$ 时, 模型 (2.3) 退化为传统逻辑模型. 本文发现, (2.3) 确实能刻画传统逻辑模型不能刻画的数据特征. 另外, 由于 λ 是否为 0 影响估计的程序, 本文指出, 在实际应用中, 直接用 λ 不为 0 的模型对数据进行拟合即可, 即本文提出的估计程序对模型中部分参数的误判具有稳健性. 为了说明上述的情形, 本文在表 2 中给出两种情形下带有辅助信息的估计结果. 情形 1: 真实数据来源于满足 $\alpha = -2.5, \beta_1 = -1, \beta_2 = 0.5$ 以及 (3.1) 中 $\rho_{xy} = -0.5$ 的逻辑模型. 利用病例 - 对照抽样, 分别从违约个体和未违约个体中抽取 $n_1 = 50$ 和 $n_0 = 2000$ 个样本. 针对得到的数据, 利用转换的逻辑模型 (即 (2.3)) 对参数进行估计. 辅助信息的形式同表 1. 情形 2: 假设数据来自于满足 (3.1)

表 1 自变量相关系数不同下, 带有辅助信息和不带辅助信息估计结果比较

	方法	参数	λ^*	α^*	β_1	β_2		
$\rho_{xy} = -0.5$	带辅助信息	TRUE	0.267	-0.774	-1.000	0.500		
		BIAS	0.020	-0.015	0.005	0.001		
		SSD	0.103	0.215	0.075	0.128		
		ESD	0.101	0.204	0.072	0.120		
		MSE	0.011	0.046	0.006	0.016		
		CP	0.957	0.948	0.949	0.936		
	不带辅助信息	BIAS	0.553	-0.129	0.082	-0.045		
		SSD	0.878	0.242	0.260	0.159		
		MSE	1.077	0.075	0.074	0.027		
		MSER	0.010	0.613	0.081	0.593		
		$\rho_{xy} = 0$	带辅助信息	TRUE	0.228	-0.615	-1.000	0.500
				BIAS	0.020	-0.015	0.002	0.002
				SSD	0.125	0.186	0.054	0.123
				ESD	0.123	0.176	0.055	0.114
MSE	0.016			0.035	0.003	0.015		
CP	0.952			0.933	0.955	0.934		
不带辅助信息	BIAS		0.665	-0.143	0.116	-0.057		
	MSER		0.012	0.361	0.036	0.417		

注: TRUE 指参数的真实值, BIAS 指多次估计值的平均值与真实值之间的差, SSD 是多次模拟估计值的真实方差, ESD 是基于矩估计的理论方差, CP 是正态的 95% 置信区间覆盖概率, MSE 是均方误差, MSER 是辅助信息与未带辅助信息下均方误差的比值.

表 2 逻辑模型与转化逻辑模型估计效果比较. 情形 1 下真实数据来自于逻辑模型, 情形 2 下真实数据来自于转换参数 $\lambda = 1.5$ 的模型

	参数	λ^*	α^*	β_1	β_2
情形 1	TRUE	0.000	-0.557	-1.000	0.500
	BIAS	0.007	-0.010	0.003	0.003
	SSD	0.036	0.269	0.084	0.173
	MSE	0.001	0.072	0.007	0.030
情形 2	TRUE	0.267	-0.774	-1.000	0.500
	BIAS	-	-0.173	-0.104	0.221
	SSD	-	0.562	0.139	0.384
	MSE	-	0.345	0.030	0.196

的模型, 对应的参数为 $\lambda = 1.5$, $\alpha = -2.5$, $\beta_1 = -1$, $\beta_2 = 0.5$ 并且 $(X, Y)^T$ 仍满足 (3.1), 其中 $\rho_{xy} = -0.5$. 同样给定表 1 中的辅助信息形式, 利用逻辑模型对违约个体中抽取的 50 个样本和未违约个体中抽取的 2,000 个样本进行分析从而得到参数估计结果. 最终在两种情形下的估计结果见表 2. 估计结果表明, 转换的逻辑模型能够很好地拟合来自于逻辑模型的数据, 而逻辑模型并不能准确地估计出转换模型的参数. 这表明转换的模型能够刻画逻辑模型不能刻画的一部分数据特征. 同时, 从表 1 和 2 也看出, 加入辅助信息模型参数估计都比没有辅助信息的渐近方差和 MSE 更小, 从均方误差比 MSER 可以看出, 增加辅助信息可以大大提高估计的效率, 甚至有本质的提高, 效率可以达到 90%, 即使辅助信息不包含的风险因素其系数估计的效率也有很大的提高, 提高的效率可达 50% 左右.

在实际应用中, 辅助信息往往并不如定理中的一样是完全准确的, 实际应用中的辅助信息可能是基于一些数据的估计值, 本文通过模拟说明该方法的实用性. 假设辅助信息来源于独立数据源并包含了抽样数据源两种情形. 考虑表 1 中的辅助信息来自于样本量为 N 的简单抽样数据源, 并以频率作为辅助信息的估计, 即

$$\hat{h}(a, b) = \frac{\#\{i; D_i = 1, a < X_i \leq b\}}{\#\{i; a < X_i \leq b\}},$$

其中 $\#A$ 表示集合 A 中的元素个数. 理论上需要 $N/(n_1 + n_0)$ 趋于无穷可以达到基于真实的辅助信息的效果, 本节将讨论 $N/(n_1 + n_0)$ 较小时和一般大时提出方法的表现. 对于给定的估计, 基于不同的样本量 N , 给出辅助信息, 对辅助信息重复给出 100 次, 每给定辅助信息, 重复 1,000 次模拟得到估计的偏差和标准差, 最后求得 100 次平均的估计偏差、标准差以及对应的均方误差, 见表 3. 注意到 $n = n_1 + n_0 = 2050$, $N = 5000$ 时, N/n 略大于 2, 比值较小. 从表 3 的分析结果可以看出, 即使辅助信息的来源的样本量较小, 参数估计的方差仍旧能够有效地减小, 但是估计的偏差会有所上升, 从均方误差来看, 相对于不带有辅助信息的估计, 大部分的估计效果仍在可接受的范围之内, 有部分参数 (如 λ^*) 的估计效果明显提升, 在实际应用中具有一定的可操作性. 从模拟结果还可以看出, 在对辅助信息的准确性把握并不大时, 辅助信息应当包含所选数据. 辅助信息质量较低时, 可以采取一定的方法进行修偏, 具体的方法与理论将在我们未来的工作中进行讨论. 模拟的结果还表明, 当 $N/(n_1 + n_0)$ 一般大时, 本文提出的估计相对于不带辅助信息的估计在偏差和方差上都有明显的提升, 估计的效果与基于真实的辅助信息的估计效果相当, 故而在实际应用中有较好的应用价值.

表 3 基于估计的辅助信息的估计结果

N	参数	独立数据源				包含抽样数据			
		λ^*	α^*	β_1	β_2	λ^*	α^*	β_1	β_2
5,000	TRUE	0.267	-0.774	-1.000	0.500	0.267	-0.774	-1.000	0.500
	BIAS	-0.040	-0.285	-0.002	0.202	-0.026	-0.281	0.003	0.193
	SSD	0.124	0.208	0.070	0.121	0.096	0.185	0.061	0.106
	MSE	0.017	0.124	0.005	0.055	0.010	0.113	0.004	0.048
10,000	BIAS	-0.011	-0.198	0.009	0.134	0.014	-0.190	0.025	0.119
	SSD	0.127	0.205	0.070	0.120	0.121	0.192	0.066	0.112
	MSE	0.016	0.081	0.005	0.032	0.015	0.073	0.005	0.027
25,000	BIAS	0.010	-0.118	0.014	0.073	-0.008	-0.095	0.001	0.067
	SSD	0.125	0.206	0.072	0.121	0.115	0.207	0.071	0.121
	MSE	0.016	0.056	0.005	0.020	0.013	0.052	0.005	0.019

4 实证分析

由于我国的金融市场破产机制的独特性,企业破产具有一定的延迟性,破产数据的获取也存在一定的难度.当企业出现财务风险时,对应的股票能被及时地判定成ST股票.ST股票即为特别处理股票,1998年4月22日,沪深交易所宣布,将对财务状况或其他状况出现异常的上市公司股票交易进行特别处理,本文将股票成为ST股票作为破产的信号,将ST股票视为“违约”状态,研究各变量与股票是否为ST股票的相关关系.依据Ding等^[10]和Campbell等^[35]的研究结果,企业的财务状态主要由盈利能力、杠杆、短期流动性、市场账面价值比例、超额收益、公司大小、股票价格和波动影响.考虑到盈利能力、杠杆率、流动性水平仅能在报告期有数据,时效性较差,故而本文从Wind数据库中提取2018年6月1日的alpha值(基于数据提取日前半年)、波动率(基于数据提取日前半年)、股票收盘价、总市值和市净率(基于最新报告期)作为自变量研究各变量与股票是否为ST股的关系.鉴于本文关注自变量极端值对是否违约的影响,在处理异常值时,本文以上下四分位数作为参考点,将超过上四分之一分位数加上5倍上下四分位数之差,或低于下四分之一分位数减去5倍上下四分位数之差的值视为异常值并对价格进行对数处理.本文以全部上市A股作为研究对象,依据股票在2018年6月1日各自变量的取值情况判断一周后即2018年6月8日股票是否为ST股票,各自变量的描述性统计结果见表4.从表4中可以看出,非ST股票的alpha值、对数价格以及对数市值高于ST股票相对应的取值.而在市净率的指标上,非ST股票高于ST的取值,但对应的标准差却更小,主要是由ST股票中含有部分净利润为负值的企业造成.在波动率上,ST股票的平均波动率大于非ST股票的波动率,标准差估计相差不大.

将第*i*家公司在2018年6月8日对应的股票是ST股票记为 $D_i = 1$,否则记为 $D_i = 0$,并利用模型(2.3)对数据进行拟合.研究的辅助信息来源于过去14天不同股票收盘价对应的ST股票率.具体为,先计算对数收盘价处于 $(a_i, b_i]$ 的股票中ST股票的个数在2018年5月18日到2018年6月1日市场上所有上市公司中的比例(每月最后一日不统计市场规模),并以其平均值作为辅助信息加入到模型中,其中 $a_1 = -b_4 = -\infty$, $a_i = b_{i-1}$ 为对数价格的下*i*/4分位数, $i = 2, 3, 4$.由于样本量较小,故本文选取所有样本作为估计的样本,并对自变量标准化后进行估计.鉴于本文没有给出不带辅助信息下估计结果的方差和推论2.1的第一个结论,本文在不加证明的基础上利用 $L = 1$ 的带有辅助信息

表 4 2018 年 6 月 8 日市场 A 股 ST 股票和非 ST 股票在 2018 年 6 月 1 日各变量的描述性统计

样本量	变量	alpha 值	波动率	对数收盘价	对数市值	市净率	
59	ST 股票	最小值	-4.630	0.000	0.199	20.816	-10.199
		中位数	-1.534	6.387	1.413	21.666	2.577
		均值	-1.595	6.705	1.459	21.780	2.981
		最大值	1.736	14.838	3.417	24.399	14.345
		标准差	1.276	2.389	0.629	0.700	4.828
3367	非 ST 股票	最小值	-5.036	0.000	-0.538	19.954	-6.225
		中位数	-0.186	5.368	2.383	22.473	2.739
		均值	-0.091	5.736	2.444	22.695	3.382
		最大值	4.847	19.869	6.614	28.329	16.924
		标准差	0.902	2.419	0.760	1.007	2.359

表 5 ST 股票数据估计结果

	变量	λ^*	α^*	alpha 值	波动率	对数收盘价	对数市值	市净率
带有辅助信息	估计	-0.016	-2.801	-1.727	0.768	-2.161	0.491	0.877
	ESD	0.005	0.412	0.243	0.209	0.246	0.192	0.172
	估计/ESD	0.311	0.147	0.141	0.272	0.114	0.390	0.196
	P 值	0.001	0.000	0.000	0.000	0.000	0.010	0.000
不带有辅助信息	估计	-0.011	-2.600	-1.016	0.441	-1.794	-1.295	0.350
	ESD	0.005	0.440	0.185	0.195	0.293	0.302	0.175
	估计/ESD	0.465	0.169	0.182	0.441	0.163	0.233	0.501
	P 值	0.031	0.000	0.000	0.023	0.000	0.000	0.046

的方差估计对不带辅助信息的估计的方差进行估计. 估计结果见表 5. 由于 $\hat{\lambda} = \hat{\lambda}^*/\hat{k} \approx -0.86$, 可以看到, 该 odds 曲线相对于逻辑模型的 odds 曲线更加平缓. 由表 5 中的结果可以得到, 加入辅助信息后, alpha 值对应的方差稍大的原因是其估计值较大. 从 P 值来看, 加入辅助信息后, alpha 值对应的 P 值仍有所下降. 除总市值的 P 值外, 带有辅助信息的估计的 P 值均小于不带辅助信息的 P 值. 辅助信息使得估计的效率有所提升.

5 结论

本文利用对数变换模型扩展了传统的逻辑模型, 将辅助信息融入到估计中, 以提高对违约预报的估计的精确性, 从而提高估计的效率; 在病例 - 对照抽样下给出了估计的渐近性质并从理论上证明了辅助信息有助于提高估计的效率. 模拟结果表明提升的效果明显, 本文所提出的模型能够刻画一些逻辑模型不能刻画的数据特征, 模型的鲁棒性较好. 实际数据的分析结果也进一步说明了方法的实用性.

然而本文仍旧存在一些可以进一步探讨的空间, 将在未来的工作中进一步解决. 首先, 各变量应该具有时变性, 因此可以将模型进一步扩展, 在此情形下, 需要类似于 Cox 模型, 从风险函数的角度出发进行建模并开展相关的讨论. 另外, 变量选择和估计在本文中相互独立, 可以考虑结合一些变量选

择的方法, 如 LASSO 和 SCAD (smoothly clipped absolute deviation) 等, 在模型估计的同时进行变量选择.

致谢 衷心地感谢各位审稿人给出的宝贵的意见.

参考文献

- 1 Liu X Q, Zhou Y. Nonparametric estimation and comparative analyses of ES in risk measure with applications (in Chinese). *Syst Eng Theory Pract*, 2011, 31: 631–642 [刘晓倩, 周勇. 金融风险管理中 ES 度量的非参数方法的比较及其应用. *系统工程理论与实践*, 2011, 31: 631–642]
- 2 Liu X Q, Zhou Y. Efficiency of the smoothed VaR estimator in financial risk management (in Chinese). *Acta Math Appl Sin*, 2011, 34: 752–768 [刘晓倩, 周勇. 金融风险管理中 VaR 度量的光滑估计效率. *应用数学学报*, 2011, 34: 752–768]
- 3 Altman E I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J Finance*, 1968, 23: 589–609
- 4 Altman E I, Haldeman R G, Narayanan P. ZETATM analysis: A new model to identify bankruptcy risk of corporations. *J Bank Finance*, 1977, 1: 29–54
- 5 Altman E I, Saunders A. Credit risk measurement: Developments over the last 20 years. *J Bank Finance*, 1997, 21: 1721–1742
- 6 Horrigan J O. The determination of long-term credit standing with financial ratios. *J Account Res*, 1966, 4: 44–62
- 7 Misund B. Financial ratios and prediction on corporate bankruptcy in the Atlantic salmon industry. *Aquac Econ Manag*, 2017, 21: 241–260
- 8 Ohlson J A. Financial ratios and the probabilistic prediction of bankruptcy. *J Account Res*, 1980, 18: 109–131
- 9 Tserng H P, Chen P C, Huang W H, et al. Prediction of default probability for construction firms using the Logit model. *J Civ Eng Manag*, 2014, 20: 247–255
- 10 Ding A A, Tian S N, Yu Y, et al. A class of discrete transformation survival models with application to default probability prediction. *J Amer Statist Assoc*, 2012, 107: 990–1003
- 11 Artzner P, Delbaen F, Eber J M, et al. Coherent measures of risk. *Math Finance*, 1999, 9: 203–228
- 12 Dar A A, Anuradha N. Probability default in Black Scholes formula: A qualitative study. *J Bus Econom Dev*, 2017, 2: 99–106
- 13 Leow M, Crook J. The stability of survival model parameter estimates for predicting the probability of default: Empirical evidence over the credit crisis. *European J Oper Res*, 2016, 249: 457–464
- 14 Hong B, Xie X, Guo H. Research on probability of default prediction based on loan company's credit fund trading behaviours. *Int J Simul Process Model*, 2014, 9: 240–245
- 15 Weißbach R, Walter R. A likelihood ratio test for stationarity of rating transitions. *J Econometrics*, 2010, 155: 188–194
- 16 Tian S, Yu Y, Guo H. Variable selection and corporate bankruptcy forecasts. *J Bank Finance*, 2015, 52: 89–100
- 17 Pinches G E, Mingo K A. A multivariate analysis of industrial bond ratings. *J Finance*, 1973, 28: 1–18
- 18 Pogue T F, Soldofsky R M. What's in a bond rating. *J Financial Quant Anal*, 1969, 4: 201–228
- 19 Zhou Y, Xie S Y, Yuan Y. Statistical inference of default probability in credit risk models (in Chinese). *Syst Eng Theory Pract*, 2008, 8: 206–214 [周勇, 谢尚宇, 袁媛. 信用违约风险模型中违约概率的统计推断. *系统工程理论与实践*, 2008, 8: 206–214]
- 20 Xie S Y, Wang S Y, Zhou Y. Default prediction with credit contagion under financial crisis (in Chinese). *J Manag Sci China*, 2011, 14: 1–12 [谢尚宇, 汪寿阳, 周勇. 金融危机下带传染效应的违约预报. *管理科学学报*, 2011, 14: 1–12]
- 21 Qin J, Zhang H, Li P, et al. Using covariate-specific disease prevalence information to increase the power of case-control studies. *Biometrika*, 2015, 102: 169–180
- 22 Tarima S, Pavlov D. Using auxiliary information in statistical function estimation. *ESAIM Probab Stat*, 2006, 10: 11–23
- 23 Huang C Y, Qin J, Tsai H T. Efficient estimation of the Cox model with auxiliary subgroup survival information. *J Amer Statist Assoc*, 2016, 111: 787–799
- 24 Chen J, Qin J. Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika*, 1993, 80: 107–116
- 25 Hellerstein J K, Imbens G W. Imposing moment restrictions from auxiliary data by weighting. *Rev Econ Stat*, 1999, 81: 1–14
- 26 Liu X, Jiang H, Zhou Y. Local empirical likelihood inference for varying-coefficient density-ratio models based on case-control data. *J Amer Statist Assoc*, 2014, 109: 635–646

- 27 Liu X, Liu P, Zhou Y. Distribution estimation with auxiliary information for missing data. *J Statist Plann Inference*, 2011, 141: 711–724
- 28 Li D, Wang X, Lin L, et al. Flexible link functions in nonparametric binary regression with Gaussian process priors. *Biometrics*, 2016, 72: 707–719
- 29 Chen K, Jin Z, Ying Z. Semiparametric analysis of transformation models with censored data. *Biometrika*, 2002, 89: 659–668
- 30 Zeng D, Lin D Y. Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*, 2006, 93: 627–640
- 31 Qin J. Empirical likelihood in biased sample problems. *Ann Statist*, 1993, 21: 1182–1196
- 32 Qin J. Inferences for case-control and semiparametric two-sample density ratio models. *Biometrika*, 1998, 85: 619–630
- 33 Qin J, Lawless J. Empirical likelihood and general estimating equations. *Ann Statist*, 1994, 22: 300–325
- 34 Owen A. Empirical likelihood ratio confidence regions. *Ann Statist*, 1990, 18: 90–120
- 35 Campbell J Y, Hilscher J, Szilagyi J. In search of distress risk. *J Finance*, 2008, 63: 2899–2939

附录 A 定理的证明

本节将沿用正文中所定义的符号, 在不混淆的基础上, 不再将向量加粗表示. 另外简记符号 $w_i = w(Z_i | \theta)$, $\eta_i = 1 + rw_i$, $\psi = \psi(Z | \theta)$, $w = w(Z | \theta)$, $w_0 = w(Z | \theta_0)$, $\eta_0 = 1 + rw(Z | \theta_0)$, $g = \frac{\psi}{1 + \xi_0(w-1)} = \frac{\psi}{\eta}(1+r)$, $g_i(\theta_0) = g(Z_i | \theta_0)$, $c = (\theta^T, t^T)^T$. 假设参数 c 的真值为 $c_0 = (\theta_0, 0)^T$. 由以上定义的符号, 可以得到

$$1 + \xi_0(w_i - 1) = \frac{1}{n}(n + n_1 w_i - n_1) = \frac{1}{n}(n_1 w_i - n_0) = \frac{n_0}{n}(r w_i + 1) = \frac{1}{1+r} \eta_i,$$

其中

$$n_1 = \frac{r}{1+r}n, \quad n_0 = \frac{1}{1+r}n.$$

在给出定理的证明之前, 我们先给出几个引理.

引理 A.1 若 $(w(Z_i | \theta) - 1, \psi(Z_i | \theta))^T$ ($i = 1, \dots, n$) 构成的区域包含了零向量, 则 $t(\theta)$ 和 $\xi(\theta)$ 是关于 θ 的连续可导函数.

引理 A.1 的证明参见文献 [33].

引理 A.2 若 r 是一个参数, $\hat{\xi}$ 是有辅助信息下 ξ 的估计值, $\tilde{\xi}$ 为不带辅助信息的估计值, 则该估计是相合的并且有 $\hat{\xi} = \tilde{\xi} = \xi_0 = n_1/n$.

证明 记截面似然函数 $\ell(\theta) = \ell(\theta, \xi, t)$, 其中 ξ 和 t 由一阶导数 $\partial \ell(\theta, \xi, t) / \partial \xi = 0$ 和 $\partial \ell(\theta, \xi, t) / \partial t = 0$ 所决定, 并且 $\ell(\theta, \xi, t)$ 的定义如下:

$$\ell(\theta, t, \xi) = \sum_{i=1}^n D_i \log w(Z_i | \theta_2) - \sum_{i=1}^n \log \{1 + \xi[w(Z_i | \theta_2) - 1] + t^T \psi(Z_i | \theta)\}.$$

对各参数求导, 可得

$$\frac{\partial \ell(\theta, \xi, t)}{\partial k} = - \sum_{i=1}^n \frac{t^T \nabla_k \psi_i}{1 + \xi(w_i - 1) + t^T \psi_i}, \quad (\text{A.1})$$

$$\frac{\partial \ell(\theta, \xi, t)}{\partial \lambda^*} = \sum_{i=1}^n D_i \frac{\nabla_{\lambda^*} w_i}{w_i} - \sum_{i=1}^n \frac{\xi \nabla_{\lambda^*} w_i + t^T \nabla_{\lambda^*} \psi_i}{1 + \xi(w_i - 1) + t^T \psi_i}, \quad (\text{A.2})$$

$$\frac{\partial \ell(\theta, \xi, t)}{\partial \alpha^*} = \sum_{i=1}^n D_i \frac{\nabla_{\alpha^*} w_i}{w_i} - \sum_{i=1}^n \frac{\xi \nabla_{\alpha^*} w_i + t^T \nabla_{\alpha^*} \psi_i}{1 + \xi(w_i - 1) + t^T \psi_i}, \quad (\text{A.3})$$

$$\frac{\partial \ell(\theta, \xi, t)}{\partial \beta} = \sum_{i=1}^n D_i \frac{\nabla_{\beta} w_i}{w_i} - \sum_{i=1}^n \frac{\xi \nabla_{\beta} w_i + \nabla_{\beta} \psi_i^T t}{1 + \xi(w_i - 1) + t^T \psi_i}, \quad (\text{A.4})$$

$$\frac{\partial \ell(\theta, \xi, t)}{\partial \xi} = - \sum_{i=1}^n \frac{w_i - 1}{1 + \xi(w_i - 1) + t^T \psi_i}, \quad (\text{A.5})$$

$$\frac{\partial \ell(\theta, \xi, t)}{\partial t} = - \sum_{i=1}^n \frac{\psi_i}{1 + \xi(w_i - 1) + t^T \psi_i}. \quad (\text{A.6})$$

通过计算可以得到对于任意的 i , 有

$$-\lambda^* \nabla_{\lambda^*} w_i + \nabla_{\alpha^*} w_i = w_i, \quad -\lambda^* \nabla_{\lambda^*} \psi_i + \nabla_{\alpha^*} \psi_i = k \nabla_k \psi_i.$$

令 (A.1)–(A.3) 和 (A.5) 分别等于 0, 并结合 $-\lambda^*$ 倍 (A.2) 加上 (A.3) 等于 0 和 $-\xi$ 倍 (A.5) 加上 (A.6) 等于 0, 即可得到 $\hat{\xi} = n_1/n \equiv \xi_0$, 同理可得 $\tilde{\xi} = n_1/n = \xi_0$. \square

引理 A.3 假设 $h(\cdot)$ 是关于 Z 的可测函数, 则

$$\mathbb{E}_1 h(Z) = \mathbb{E}_0 \{w_0 h(Z)\},$$

并且有

$$\mathbb{E} \sum_{i=1}^n h(Z_i) = \frac{n}{1+r} \mathbb{E}_0 \{\eta_0 h(Z)\}.$$

引理 A.3 的证明参见文献 [21, 引理 1].

引理 A.4 若定理 2.1 的假设成立, 则当 $\|\theta - \theta_0\| \leq n^{-1/3}$ 时,

$$t = \left[\frac{1}{n} \sum_{i=1}^n g(Z_i | \theta) g^T(Z_i | \theta) \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n g(Z_i | \theta) \right] + o(n^{-1/3}),$$

其中 $\|\cdot\|$ 表示 Euclid 范数.

证明 首先给出

$$\frac{1}{n} \sum_{i=1}^n g(Z_i | \theta_0) = O_p(n^{-1/2}). \quad (\text{A.7})$$

由

$$\begin{aligned} \sum_{i=1}^n \mathbb{E} g(Z_i | \theta_0) &= n_1 \mathbb{E}_1 g(Z | \theta_0) + n_0 \mathbb{E}_0 g(Z | \theta_0) \\ &= n \frac{r}{1+r} \mathbb{E}_0 w(Z | \theta_0) \frac{\psi(Z | \theta_0)}{\eta(Z | \theta_0)} (1+r) + n \frac{1}{1+r} \mathbb{E}_0 \frac{\psi(Z | \theta_0)}{\eta(Z | \theta_0)} (1+r) \\ &= n \mathbb{E}_0 \psi(\theta_0) = 0, \end{aligned}$$

可以得到

$$\sum_{i=1}^n g(Z_i | \theta_0) = \sum_{i=1}^n [g(Z_i | \theta_0) - \mathbb{E} g(Z_i | \theta_0)].$$

由此可得

$$\text{var} \left(\sum_{i=1}^n g(Z_i | \theta_0) \right) = n_1 \text{var}_1 g(Z | \theta_0) + n_0 \text{var}_0 g(Z | \theta_0)$$

$$\begin{aligned}
 &= n_1 \left[\mathbb{E}_1 \left(\frac{\psi\psi^T}{\eta^2} (1+r)^2 \right) - \mathbb{E}_1 \frac{\psi}{\eta} \mathbb{E}_1 \frac{\psi^T}{\eta} (1+r)^2 \right] \\
 &\quad + n_0 \left[\mathbb{E}_0 \left(\frac{\psi\psi^T}{\eta^2} (1+r)^2 \right) - \mathbb{E}_0 \frac{\psi}{\eta} \mathbb{E}_0 \frac{\psi^T}{\eta} (1+r)^2 \right] \\
 &= n(1+r) \mathbb{E}_0 \frac{\psi\psi^T}{\eta} - n(1+r)^2 \mathbb{E}_0 \frac{\psi}{\eta} \mathbb{E}_0 \frac{\psi^T}{\eta}.
 \end{aligned}$$

注意到

$$\begin{aligned}
 &B_n^{-2} \sum_{i=1}^n \mathbb{E}[g(Z_i | \theta_0) - \mathbb{E}g(Z_i | \theta_0)]^2 I_{|g(Z_i|\theta_0) - \mathbb{E}g(Z_i|\theta_0)| > \epsilon B_n} \\
 &= \frac{r}{1+r} \mathbb{E}_1 [g(Z | \theta_0) - \mathbb{E}g(Z | \theta_0)]^2 I_{|g(Z|\theta_0) - \mathbb{E}_1 g(Z|\theta_0)| > \epsilon B_n} \\
 &\quad + \frac{1}{1+r} \mathbb{E}_0 [g(Z | \theta_0) - \mathbb{E}_0 g(Z | \theta_0)]^2 I_{|g(Z|\theta_0) - \mathbb{E}_0 g(Z|\theta_0)| > \epsilon B_n},
 \end{aligned}$$

由控制收敛定理可得, Lindeberg 中心极限定理条件满足, 故 (A.7) 成立. 故而,

$$\begin{aligned}
 \frac{1}{n} \sum_{i=1}^n e^T g(Z_i | \theta) &= \frac{1}{n} \sum_{i=1}^n \left[e^T g(Z_i | \theta_0) + e^T \frac{\partial g(Z_i | \theta)}{\partial \theta} \Big|_{\theta_0} un^{-1/3} + o(n^{-1/3}) \right] \\
 &= O_p(n^{-1/2}) + O_p(n^{-1/3}) + o(n^{-1/3}) = O_p(n^{-1/3}).
 \end{aligned}$$

注意到 $t = t(\theta)$ 由方程 $\sum_{i=1}^n \frac{\psi_i}{1 + \xi(w_i - 1) + t^T \psi_i} = 0$ 决定, 故而可以得到

$$\sum_{i=1}^n \frac{\psi_i(\theta)}{1 + \xi(w_i(\theta) - 1) + t^T \psi_i} = \sum_{i=1}^n \frac{g_i(\theta)}{1 + t^T g_i(\theta)} = 0. \tag{A.8}$$

令 $t = \|t\|e$, 由 (A.8) 可以得到

$$\sum_{i=1}^n e^T g_i(\theta) \left(1 - \frac{g_i^T(\theta)t}{1 + t^T g_i(\theta)} \right) = 0.$$

因此,

$$\begin{aligned}
 \sum_{i=1}^n e^T g_i(\theta) &= \sum_{i=1}^n \frac{e^T g_i(\theta) g_i^T(\theta) e \|t\|}{1 + t^T g_i(\theta)} \geq \sum_{i=1}^n \frac{e^T g_i(\theta) g_i^T(\theta) e \|t\|}{1 + \max_{1 \leq i \leq n} \|t^T g_i(\theta)\|} \\
 &\geq \sum_{i=1}^n \frac{e^T g_i(\theta) g_i^T(\theta) e \|t\|}{1 + \|t\| \max_{1 \leq i \leq n} |e^T g_i(\theta)|},
 \end{aligned}$$

其中应用了 $1 + t^T g_i(\theta) > 0$. 类似于文献 [34, 引理 3], 可以得到

$$\begin{aligned}
 \max_{1 \leq i \leq n} |e^T g_i(\theta)| &= \max \left\{ \max_{1 \leq i \leq n_1} |e^T g_i(\theta)|, \max_{n_1+1 \leq i \leq n} |e^T g_i(\theta)| \right\} \\
 &= \max \{ o(n_1^{1/3}), o(n_0^{1/3}) \} = o(n^{1/3}).
 \end{aligned} \tag{A.9}$$

经过简单的运算可以得到

$$\frac{1}{n} \sum_{i=1}^n e^T g_i(\theta) \left(1 + \|t\| \max_{1 \leq i \leq n} |e^T g_i(\theta)| \right) \geq e^T \left(\sum_{i=1}^n \frac{1}{n} g_i(\theta) g_i^T(\theta) \right) e \|t\|,$$

即

$$\frac{1}{n} \sum_{i=1}^n e^{\mathbf{T} g_i(\theta)} \geq e^{\mathbf{T} \left[\left(\sum_{i=1}^n \frac{1}{n} g_i(\theta) g_i^{\mathbf{T}}(\theta) \right) e - \frac{1}{n} \sum_{i=1}^n e^{\mathbf{T} g_i(\theta)} \max_{1 \leq i \leq n} |e^{\mathbf{T} g_i(\theta)}| \right]} \|t\|.$$

基于以上的论证, 可以得到

$$O_p(n^{-1/3}) = \|t\| \{O_p(1) - O_p(n^{-1/3})o_p(n^{1/3})\},$$

故而有

$$\|t\| = O_p(n^{-1/3}).$$

对 (A.8) 进行 Taylor 展开并令 $Y_i = t^{\mathbf{T}} g_i$, 可得

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^n g_i(\theta) (1 - Y_i + Y_i^2 + o(Y_i^2)) \\ &= \frac{1}{n} \sum_{i=1}^n g_i(\theta) - \frac{1}{n} \sum_{i=1}^n g_i(\theta) g_i^{\mathbf{T}}(\theta) t + \frac{1}{n} \sum_{i=1}^n g_i(\theta) Y_i^2 (1 + o_p(1)). \end{aligned}$$

对于上式的第四项, 类似于文献 [34] 和 (A.9), 可以得到以下等式:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \|g_i(\theta)\|^3 \|t\|^2 &= o(n^{1/3}) O_p(n^{-2/3}) = o_p(n^{-1/3}), \\ t(\theta) &= \left[\frac{1}{n} \sum_{i=1}^n g_i(\theta) g_i^{\mathbf{T}}(\theta) \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n g_i(\theta) \right] + o_p(n^{-1/3}). \end{aligned}$$

证毕. □

引理 A.5 关于 $g_i(\theta_0)$ ($i = 1, \dots, n$) 的重对数律成立, 即

$$\frac{1}{n} \sum_{i=1}^n g_i(\theta_0) = O_p(n^{-1/2} [\log \log(n)]^{1/2}).$$

证明 由样本中病例组和对照组的样本量比值恒定, 以及引理 A.3, 并在病例组和对照组中分别使用重对数定律, 可以得到

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n g_i(\theta_0) &= \frac{n_1}{n} \frac{1}{n_1} \sum_{i=1}^{n_1} g(Z_i | \theta_0) + \frac{n_0}{n} \frac{1}{n_0} \sum_{i=n_1+1}^n g(Z_i | \theta_0) \\ &= \frac{n_1}{n} \frac{1}{n_1} [n_1 E_1 g(\theta_0) + O_p(n^{1/2} [\log \log(n)]^{1/2})] \\ &\quad + \frac{n_0}{n} \frac{1}{n_0} [n_0 E_0 g(\theta_0) + O_p(n^{1/2} [\log \log(n)]^{1/2})] \\ &= 0 + n^{-1} O_p(n^{1/2} [\log \log(n)]^{1/2}) \\ &= O_p(n^{-1/2} [\log \log(n)]^{1/2}). \end{aligned}$$

证毕. □

附录 A.1 定理 2.1 的证明

注意到对数似然函数具有形式

$$\ell(\theta) = \sum_{i=1}^{n_1} \log w_i - \sum_{i=1}^n \log[1 + \xi_0(w_i - 1) + t^T \psi_i].$$

可以轻易地将对数似然函数分为两部分, 记为 $\ell(\theta) = \ell_2(\theta) - \ell_1(\theta)$, 其中

$$\begin{aligned} \ell_1(\theta, t(\theta)) &= \sum_{i=1}^n \log[1 + t^T g_i], \\ \ell_2(\theta) &= \sum_{i=1}^{n_1} \log w_i - \sum_{i=1}^n \log(1 + \xi_0(w_i - 1)). \end{aligned}$$

类似于文献 [33], 当 $\|\theta - \theta_0\| < n^{-1/3}$ 时, 可以得到对于足够大的 n , 有

$$\ell_1(\theta) > \ell_1(\theta_0).$$

对于第二部分 $\ell_2(\theta)$, 利用引理 A.3, 可以得到

$$\begin{aligned} \mathbb{E} \left[\frac{\partial \ell_2(\theta)}{\partial \theta} \right] \Big|_{\theta_0} &= n_1 \mathbb{E}_0 \left(w \frac{\nabla_{\theta} w}{w} \right) \Big|_{\theta_0} - \frac{n}{1+r} \mathbb{E}_0 \left(\eta \frac{r \nabla_{\theta} w}{\eta} \right) \\ &= \frac{rn}{1+r} \mathbb{E}_0(\nabla_{\theta} w) \Big|_{\theta_0} - \frac{rn}{1+r} \mathbb{E}_0(\nabla_{\theta} w) \Big|_{\theta_0} = 0, \end{aligned}$$

以及

$$\begin{aligned} \mathbb{E} \left[\frac{\partial^2 \ell_2(\theta)}{\partial \theta \partial \theta^T} \right] \Big|_{\theta^*} &= \frac{rn}{1+r} \mathbb{E}_0 \left(\frac{\nabla_{\theta \theta^T} w w - \nabla_{\theta} w \nabla_{\theta^T} w}{w} \right) \Big|_{\theta^*} - \frac{n}{1+r} \mathbb{E}_0 \left(\frac{r \nabla_{\theta \theta^T} w \eta - r^2 \nabla_{\theta} w \nabla_{\theta^T} w}{\eta} \right) \Big|_{\theta^*} \\ &= \frac{rn}{1+r} \mathbb{E}_0 \left(\frac{-\nabla_{\theta} w \nabla_{\theta^T} w}{w_0} \right) \Big|_{\theta^*} + \frac{n}{1+r} \mathbb{E}_0 \left(\frac{r^2 \nabla_{\theta} w \nabla_{\theta^T} w}{\eta} \right) \Big|_{\theta^*} \\ &= -\frac{nr}{1+r} \mathbb{E}_0 \left(\frac{\nabla_{\theta} w \nabla_{\theta^T} w}{w \eta} \right) \Big|_{\theta^*}. \end{aligned}$$

假设 $\theta = \theta_0 + un^{-1/3}$, 当 n 足够大时, 有

$$\begin{aligned} \ell_2(\theta) &= \ell_2(\theta_0) + \frac{\partial \ell_2(\theta)}{\partial \theta} \Big|_{\theta_0} un^{-1/3} + \frac{1}{2} u^T \frac{\partial^2 \ell_2(\theta)}{\partial \theta \partial \theta^T} \Big|_{\theta^*} un^{-2/3} \\ &= \ell_2(\theta_0) + (O(n \log \log n)^{1/2} un^{-1/3}) - \frac{rn}{2(1+r)} u^T \left(\mathbb{E}_0 \left(\frac{\nabla_{\theta} w \nabla_{\theta^T} w}{w \eta} \right) \Big|_{\theta^*} + o_p(1) \right) un^{-2/3} \\ &< \ell_2(\theta_0). \end{aligned} \tag{A.10}$$

由此可以得到, 对于足够大的 n , 有

$$\ell(\theta) \leq \ell(\theta_0).$$

由于 $\ell(\theta)$ 在闭区域 $\|\theta - \theta_0\| \leq n^{-1/3}$ 内是关于 θ 的连续函数, 因此, $\ell(\theta)$ 在球内取得极大值点. 由 $\tilde{\theta}_2 = \arg \max_{\theta_2} \ell_2(\theta_2)$ 以及 (A.10) 可以得到 $\tilde{\theta}_2$ 是极大化 $\ell_2(\theta)$ 的最值点. 故而定理 2.1 得证.

附录 A.2 定理 2.2 的证明

在正式开始定理 2.2 的证明之前, 首先给出以下的事实和引理. 记截面似然函数为 $\ell(\theta) = \ell(\theta, \xi, t)$, 则 Lagrange 乘子 ξ 和 t 是方程组 $\partial\ell(\theta, \xi, t)/\partial\xi = 0$ 和 $\partial\ell(\theta, \xi, t)/\partial t = 0$ 的解, 其中 $\ell(\theta, \xi, t)$ 的定义如下:

$$\ell(\theta, t, \xi) = \sum_{i=1}^n D_i \log w(Z_i | \theta_2) - \sum_{i=1}^n \log\{1 + \xi[w(Z_i | \theta_2) - 1] + t^\top \psi(Z_i | \theta)\}.$$

引理 A.6 对于截面似然函数的二次导数 $\frac{\partial^2 \ell(\theta, \xi, t)}{\partial c \partial c^\top}$, 有

$$H := \frac{1}{n} \mathbb{E} \left. \frac{\partial^2 \ell(\theta, \xi, t)}{\partial c \partial c^\top} \right|_{\theta_0, \xi_0, 0} = \begin{pmatrix} 0 & 0 & 0 & 0 & -H_{15} \\ 0 & -H_{22} & -H_{23} & -H_{24} & -H_{25} \\ 0 & -H_{32} & -H_{33} & -H_{34} & -H_{35} \\ 0 & -H_{42} & -H_{43} & -H_{44} & -H_{45} \\ -H_{51} & -H_{52} & -H_{53} & -H_{54} & H_{55} \end{pmatrix},$$

其中

$$H_{15} = H_{51}^\top = \mathbb{E}_0 \nabla_k \psi_i^\top |_{\theta=\theta_0},$$

$$H_{22} = \frac{r}{1+r} \mathbb{E}_0 \left[\frac{(\nabla_{\lambda^*} w |_{\theta_0})^2}{w_0} - \frac{r(\nabla_{\lambda^*} w |_{\theta_0})^2}{\eta_0} \right] = \frac{r}{1+r} \mathbb{E}_0 \left[\frac{(\nabla_{\lambda^*} w |_{\theta_0})^2}{w_0 \eta_0} \right],$$

$$H_{23} = H_{32} = \frac{r}{1+r} \mathbb{E}_0 \left[\frac{\nabla_{\lambda^*} w |_{\theta_0} \nabla_{\alpha^*} w |_{\theta_0}}{w_0} - \frac{r \nabla_{\lambda^*} w |_{\theta_0} \nabla_{\alpha^*} w |_{\theta_0}}{\eta_0} \right] = \frac{r}{1+r} \mathbb{E}_0 \left[\frac{\nabla_{\lambda^*} w |_{\theta_0} \nabla_{\alpha^*} w |_{\theta_0}}{w_0 \eta_0} \right],$$

$$H_{24} = H_{42}^\top = \frac{r}{1+r} \mathbb{E}_0 \left[\frac{\nabla_{\lambda^*} w |_{\theta_0} \nabla_{\beta^\top} w |_{\theta_0}}{w_0} - \frac{r \nabla_{\lambda^*} w |_{\theta_0} \nabla_{\beta^\top} w |_{\theta_0}}{\eta_0} \right] = \frac{r}{1+r} \mathbb{E}_0 \left[\frac{\nabla_{\lambda^*} w |_{\theta_0} \nabla_{\beta^\top} w |_{\theta_0}}{w_0 \eta_0} \right],$$

$$H_{25} = H_{52}^\top = \mathbb{E}_0 \left[\nabla_{\lambda^*} \psi^\top |_{\theta_0} - \frac{r(\nabla_{\lambda^*} w |_{\theta_0}) \psi_0^\top}{\eta_0} \right],$$

$$H_{33} = \frac{r}{1+r} \mathbb{E}_0 \left[\frac{(\nabla_{\alpha^*} w |_{\theta_0})^2}{w_0} - \frac{r(\nabla_{\alpha^*} w |_{\theta_0})^2}{\eta_0} \right] = \frac{r}{1+r} \mathbb{E}_0 \left[\frac{(\nabla_{\alpha^*} w |_{\theta_0})^2}{w_0 \eta_0} \right],$$

$$H_{34} = H_{43}^\top = \frac{r}{1+r} \mathbb{E}_0 \left[\frac{\nabla_{\alpha^*} w |_{\theta_0} \nabla_{\beta^\top} w |_{\theta_0}}{w_0} - \frac{r \nabla_{\alpha^*} w |_{\theta_0} \nabla_{\beta^\top} w |_{\theta_0}}{\eta_0} \right] = \frac{r}{1+r} \mathbb{E}_0 \left[\frac{\nabla_{\alpha^*} w |_{\theta_0} \nabla_{\beta^\top} w |_{\theta_0}}{w_0 \eta_0} \right],$$

$$H_{35} = H_{53}^\top = \mathbb{E}_0 \left[\nabla_{\alpha^*} \psi^\top |_{\theta_0} - \frac{r(\nabla_{\alpha^*} w |_{\theta_0}) \psi_0^\top}{\eta_0} \right],$$

$$H_{44} = \frac{r}{1+r} \mathbb{E}_0 \left[\frac{\nabla_{\beta} w |_{\theta_0} \nabla_{\beta^\top} w |_{\theta_0}}{w_0} - \frac{r \nabla_{\beta} w |_{\theta_0} \nabla_{\beta^\top} w |_{\theta_0}}{\eta_0} \right] = \frac{r}{1+r} \mathbb{E}_0 \left[\frac{\nabla_{\beta} w |_{\theta_0} \nabla_{\beta^\top} w |_{\theta_0}}{w_0 \eta_0} \right],$$

$$H_{45} = H_{54}^\top = \mathbb{E}_0 \left[\nabla_{\beta} \psi^\top |_{\theta_0} - \frac{r(\nabla_{\beta} w |_{\theta_0}) \psi_0^\top}{\eta_0} \right],$$

$$H_{55} = (1+r) \mathbb{E}_0 \frac{\psi_0 \psi_0^\top}{\eta_0}.$$

证明 仅以 H_{25} 为例说明, 其他类似可得. 类似于引理 A.5 的证明, 利用引理 A.3 的结论, 以及样本中病例组和对照组样本量的关系, 可以得到

$$-\frac{1}{n} \mathbb{E} \frac{\partial^2 \ell(\theta_0, \xi_0, 0)}{\partial \lambda^* \partial t^\top} = \mathbb{E} \left(\frac{1}{n} (1+r) \sum_{i=1}^n \frac{\eta_i \nabla_{\lambda^*} \psi_i^\top - r \nabla_{\lambda^*} w_i \psi_i^\top}{\eta_i^2} \right) \Big|_{\theta=\theta_0}$$

$$\begin{aligned}
 &= E_0 \left[\frac{\eta_0 \nabla_{\lambda^*} \psi^T |_{\theta_0} - r \nabla_{\lambda^*} w |_{\theta_0} \psi_0^T}{\eta_0} \right] \\
 &= E_0 \left[\nabla_{\lambda^*} \psi^T |_{\theta_0} - \frac{r \nabla_{\lambda^*} w |_{\theta_0} \psi_0^T}{\eta_0} \right].
 \end{aligned}$$

证毕. □

引理 A.7 令 $(0, S_n^T) := \frac{\partial \ell(\theta, \xi, t)}{\partial c} |_{\theta_0, 0, \xi_0} = (0, S_{n2}, S_{n3}, S_{n4}^T, S_{n5}^T)^T$, 可以验证

$$E(S_n) = 0,$$

并且有

$$\frac{1}{n} \text{var}(S_n) = V - \frac{(1+r)^2}{r} vv^T,$$

其中

$$\begin{aligned}
 v &= \left(\frac{r}{1+r} E_0 \left(\frac{\nabla_{\lambda^*} w |_{\theta_0}}{\eta_0} \right), \frac{r}{1+r} E_0 \left(\frac{\nabla_{\alpha^*} w |_{\theta_0}}{\eta_0} \right), \frac{r}{1+r} E_0 \left(\frac{\nabla_{\beta^*} w |_{\theta_0}}{\eta_0} \right)^T, E_0 \left(\frac{\psi_0}{\eta_0} \right)^T \right)^T, \\
 V &= \begin{pmatrix} H_{22} & H_{23} & H_{24} & 0 \\ H_{32} & H_{33} & H_{34} & 0 \\ H_{42} & H_{43} & H_{44} & 0 \\ 0 & 0 & 0 & H_{55} \end{pmatrix}.
 \end{aligned} \tag{A.11}$$

证明 由引理 A.3 可以轻易得到 $E(S_n) = 0$. 下面以 $\text{var}(S_{n5})$ 为例给出证明:

$$\begin{aligned}
 \text{var}(S_{n5}) &= (1+r)^2 E_0 \frac{[n_1 w_0 + n_0] \psi_0 \psi_0^T}{\eta_0} - (1+r)^2 E_0 \left[n_1 \frac{1}{r^2} + n_0 \right] E_0 \frac{\psi_0}{\eta_0} E_0 \frac{\psi_0^T}{\eta_0} \\
 &= (1+r) n E_0 \frac{\psi_0 \psi_0^T}{\eta_0} - \frac{(1+r)^2}{r} n E_0 \frac{\psi_0}{\eta_0} E_0 \frac{\psi_0^T}{\eta_0} \\
 &= H_{55} - \frac{(1+r)^2}{r} E_0 \frac{\psi_0}{\eta_0} E_0 \frac{\psi_0^T}{\eta_0}.
 \end{aligned}$$

证毕. □

接下来继续定理 2.2 的证明. 我们已经证明了 $\hat{\theta} = \theta_0 + O_p(n^{-1/3})$, $\hat{t} = O_p(n^{-1/3})$ 以及 Lagrange 乘子 \hat{t} 满足

$$\frac{\partial \ell(\hat{\theta}, \hat{\xi}, \hat{t})}{\partial c} = 0.$$

由 $\hat{\xi} = \xi_0$, 可以得到

$$\frac{\partial \ell(\hat{\theta}, \xi_0, \hat{t})}{\partial c} = 0.$$

在 c_0 处对 $\frac{1}{n} \frac{\partial \ell(\hat{\theta}, \xi_0, \hat{t})}{\partial c}$ 进行 Taylor 展开, 可以得到

$$0 = \frac{1}{n} \frac{\partial \ell(\theta_0, \xi_0, 0)}{\partial c} + \frac{1}{n} \frac{\partial^2 \ell(\theta_0, \xi_0, 0)}{\partial c \partial c^T} (c - c_0) + o_p(\delta_n),$$

其中 $\delta_n = \|\hat{\theta} - \theta\| + \|\hat{t}\|$. 由大数定理可以得到

$$\frac{1}{n} \frac{\partial^2 \ell(\theta_0, \xi_0, 0)}{\partial c \partial c^T} \xrightarrow{p} E \left\{ \frac{\partial^2 \ell(\theta_0, \xi_0, 0)}{\partial c \partial c^T} \right\}, \quad \frac{1}{n} \frac{\partial \ell(\theta_0, \xi_0, 0)}{\partial c} = O_p(n^{-1/2}).$$

由此可以得到 $\delta_n = O_p(n^{-1/2})$, 即 $\|\hat{\theta} - \theta\| = \|\hat{t}\| = O_p(n^{-1/2})$. 故而可以得到

$$-\frac{1}{n} \begin{pmatrix} 0 \\ S_{n2} \\ S_{n3} \\ S_{n4} \\ S_{n5} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & -H_{15} \\ 0 & -H_{22} & -H_{23} & -H_{24} & -H_{25} \\ 0 & -H_{32} & -H_{33} & -H_{34} & -H_{35} \\ 0 & -H_{42} & -H_{43} & -H_{44} & -H_{45} \\ -H_{51} & -H_{52} & -H_{53} & -H_{54} & H_{55} \end{pmatrix} \begin{pmatrix} \hat{\theta} - \theta_0 \\ \hat{t} - t \end{pmatrix} + o_p(n^{-1/2}). \quad (\text{A.12})$$

由 (A.12), 可以得到

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -H_{22} & -H_{23} & -H_{24} \\ 0 & -H_{32} & -H_{33} & -H_{34} \\ 0 & -H_{42} & -H_{43} & -H_{44} \end{pmatrix} (\hat{\theta} - \theta_0) - \begin{pmatrix} H_{15} \\ H_{25} \\ H_{35} \\ H_{45} \end{pmatrix} \hat{t} = -\frac{1}{n} \begin{pmatrix} 0 \\ S_{n2} \\ S_{n3} \\ S_{n4} \end{pmatrix} + o_p(n^{-1/2}), \quad (\text{A.13})$$

以及

$$-(H_{51}, H_{52}, H_{53}, H_{54})(\hat{\theta} - \theta_0) + H_{55}\hat{t} = -\frac{1}{n}S_{n5} + o_p(n^{-1/2}).$$

故而可以得到

$$\hat{t} = H_{55}^{-1}(H_{51}, H_{52}, H_{53}, H_{54})(\hat{\theta} - \theta_0) - H_{55}^{-1} \frac{S_{n5}}{n} + o_p(n^{-1/2}). \quad (\text{A.14})$$

令

$$U = \begin{pmatrix} 0 & 0 & 0 & H_{15} \\ H_{22} & H_{23} & H_{24} & H_{25} \\ H_{32} & H_{33} & H_{34} & H_{35} \\ H_{42} & H_{43} & H_{44} & H_{45} \end{pmatrix}, \quad (\text{A.15})$$

以及 $J = UV^{-1}U^T$. 将 (A.14) 代入 (A.13) 可以得到

$$J(\hat{\theta} - \theta_0) = UV^{-1} \left(\frac{1}{n} S_n \right) + o_p(n^{-1/2}). \quad (\text{A.16})$$

故而可以得到 $n^{1/2}(\hat{\theta} - \theta_0) \rightarrow N(0, \Sigma)$, 其中

$$\Sigma = J^{-1} - \frac{(1+r)^2}{r} WW^T,$$

并且有 $W = J^{-1}UV^{-1}v$. 另外, 注意到

$$v = (-\lambda_0^*H_{22} + H_{32}, -\lambda_0^*H_{23} + H_{33}, -\lambda_0^*H_{24} + H_{34}, -\lambda_0^*H_{25} + H_{35} - k_0H_{15})^T.$$

因此,

$$UV^{-1}v = \begin{pmatrix} -k_0H_{15}H_{55}^{-1}H_{51} - \lambda_0^*H_{15}H_{55}^{-1}H_{52} + H_{15}H_{55}^{-1}H_{53} \\ -\lambda_0^*H_{22} + H_{23} - k_0H_{25}H_{55}^{-1}H_{51} - \lambda_0^*H_{25}H_{55}^{-1}H_{52} + H_{25}H_{55}^{-1}H_{53} \\ -\lambda_0^*H_{32} + H_{33} - k_0H_{35}H_{55}^{-1}H_{51} - \lambda_0^*H_{35}H_{55}^{-1}H_{52} + H_{35}H_{55}^{-1}H_{53} \\ -\lambda_0^*H_{42} + H_{43} - k_0H_{45}H_{55}^{-1}H_{51} - \lambda_0^*H_{45}H_{55}^{-1}H_{52} + H_{45}H_{55}^{-1}H_{53} \end{pmatrix} = J \begin{pmatrix} -k_0 \\ -\lambda_0^* \\ 1 \\ \mathbf{0}_p \end{pmatrix}.$$

故而渐近方差的表达式可以简化为

$$\Sigma = J^{-1} - \frac{(1+r)^2}{r} (-k_0, -\lambda_0^*, 1, \mathbf{0}_p^T)^T (-k_0, -\lambda_0^*, 1, \mathbf{0}_p^T).$$

由此定理 2.2 证毕.

附录 A.3 推论 2.1 的证明

由 $\tilde{\theta}_2$ 的定义可以得到 $\tilde{\theta}_2 = \arg \max_{\theta_2} \ell_2(\theta_2)$. 在 θ_{20} 处进行 Taylor 展开可以得到

$$0 = \frac{1}{n} \frac{\partial \ell_2(\theta_{20}, \xi_0)}{\partial \theta} + \frac{1}{n} \frac{\partial^2 \ell_2(\theta_{20}, \xi_0)}{\partial \theta \partial \theta^T} (\theta - \theta_0) + o_p(\delta_n^T).$$

注意到

$$\begin{aligned} \frac{\partial \ell_2(\theta_{20}, \xi_0)}{\partial \theta} &= \frac{\partial \ell(\theta_0, \xi_0, 0)}{\partial \theta}, \\ \frac{\partial^2 \ell_2(\theta_{20}, \xi_0)}{\partial \theta \partial \theta^T} &= \frac{\partial^2 \ell(\theta_0, \xi_0, 0)}{\partial \theta \partial \theta^T}. \end{aligned}$$

类似于之前的证明可以得到 $\delta_n^T = O_p(n^{-1/2})$, 并且有

$$-\frac{1}{n} \begin{pmatrix} S_{n2} \\ S_{n3} \\ S_{n4} \end{pmatrix} = \begin{pmatrix} -H_{22} & -H_{23} & -H_{24} \\ -H_{32} & -H_{33} & -H_{34} \\ -H_{42} & -H_{43} & -H_{44} \end{pmatrix} (\tilde{\theta}_2 - \theta_{20}) + o_p(n^{-1/2}).$$

当 $L = 1$ 时, U 为方阵. 故而, (A.16) 可以简化为

$$U^T(\hat{\theta} - \theta_0) = \left(\frac{1}{n} S_n \right) + o_p(n^{-1/2}).$$

由 U 、 S_n 和 $\theta = (k, \theta_2^T)^T$ 的定义可以得到

$$\frac{1}{n} \begin{pmatrix} S_{n2} \\ S_{n3} \\ S_{n4} \end{pmatrix} = \begin{pmatrix} H_{22} & H_{23} & H_{24} \\ H_{32} & H_{33} & H_{34} \\ H_{42} & H_{43} & H_{44} \end{pmatrix} (\hat{\theta}_2 - \theta_{20}) + o_p(n^{-1/2}).$$

故而, $\sqrt{n}(\hat{\theta}_2 - \theta_{20})$ 和 $\sqrt{n}(\tilde{\theta}_2 - \theta_{20})$ 的渐近方差相同. 对于推论剩下的部分, 令 U_m 、 V_m 、 J_m 、 Σ_m 和 v_m 表示由 $\psi(Z | \theta)$ 的前 m 个估计方程得到的 U 、 V 、 J 、 Σ 和 v , $1 < m \leq L$. 故而只需要证明

$$\Sigma_m \leq \Sigma_{m-1}.$$

由文献 [21], 可以得到

$$V_m^{-1} \geq \begin{pmatrix} V_{m-1}^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

由此可得

$$J_m = U_m V_m^{-1} U_m^T \geq (U_{m-1}, u_m) \begin{pmatrix} V_{m-1}^{-1} & 0 \\ 0 & 0 \end{pmatrix} (U_{m-1}, u_m)^T = U_{m-1} V_{m-1}^{-1} U_{m-1}^T = J_{m-1}.$$

故而,

$$\Sigma_m \leq \Sigma_{m-1}.$$

由此推论 2.1 证毕.

Default forecast with auxiliary information using a logarithmic transformation model

Moming Wang & Yong Zhou

Abstract This paper studies how to integrate auxiliary information into the estimation procedure to improve the stability and efficiency of estimators for default forecast and introduces a log-transformation logic model to characterize different default probability curves. In this paper, we establish the consistency and asymptotic normality of the estimators and prove the efficiency of the proposed estimators with auxiliary information. Simulation results show that the proposed method can improve the efficiency of estimation and the influence of auxiliary information is discussed. We apply the proposed method to the data of ST (special treatment) stocks, and the empirical results show that the parameter estimation with auxiliary information is more effective.

Keywords default, auxiliary information, credit risk, empirical likelihood, log-transformation

MSC(2010) 62F10, 62G20, 62P05

doi: 10.1360/SCM-2018-0622