

论文 NO. 2017 年 8

发表时间：2017 年 8 月 15 日

信用风险的 GLMMs 建模分析

胡志浩 卜永强

摘要： 如何精确计量信用风险一直是理论界和实务部门的难点和热点问题。本文使用广义线性混合模型对信用风险进行建模分析，将影响违约概率的可观测因素和不可观测因素分别用固定效应和随机效应表示，根据需要随机效应可扩展为多个因子。研究表明，模型具有较好的延展性，宏观经济变量作为可观测变量无法全部解释违约率的异质性，随机效应可以更好地捕捉违约率的异质性，行业因素对违约概率的影响比宏观经济变量显著。

Abstract: How to accurately measure the credit risk is always a hard and hot issue in the research and administrative institutions. This paper is to model and analyze the credit risks by the generalized linear mixed models (GLMMs). The observable and

unobservable factors, which affect credit default probability, are expressed as fixed effects and random effects respectively, and the random effects can be expanded into multiple variables as required. The result shows that model can be extended with more variables and the heterogeneity of default probability can't be totally explained by the macroeconomic covariates, but can be well captured by the random effects. It is also found that the impacts from industrial sectors on the credit default probability are more significant than those from the macroeconomic variables.

Key words: Credit Risk; GLMMs(Generalized Linear Mixed Models) ; PD(Probability of Default)

声明：国家金融与发展实验室论文发表实验室工作人员的研究成果，以利于开展学术交流与研讨。论文内容仅代表作者个人学术观点。如需引用，请注明来源为《国家金融与发展实验室论文》。

一、引言

信用风险，也称违约风险，是金融行业所面临的三大最重要的风险之一。如何精确计量信用风险一直是理论界和实务部门的难点和热点问题。过去几十年，从最初的古典分析方法到现代度量模型，信用风险度量模型不断演进。Altman(1968)作为古典分析方法的典型，最先提出了将信用风险量化管理的观点，提出 Z 型评分模型，对不同客户进行打分卡评分的模式，来分析公司或者行业的信用风险差异性。Merton(1974)将企业价值理论与期权定价理论共同引入到债券的信用风险研究中来，结合 Black 和 Scholes 关于期权定价的研究，开创了信用风险研究的新纪元。此后，信用风险定价领域大量的理论和实证研究如雨后春笋般涌现出来。经过 20 多年的理论和实践，在 20 世纪 90 年代中期，银行和咨询机构由于业务发展需要研发了四大现代度量信用风险模型，主要包括 KMV 公司(现已被穆迪合并)开发的信用监测模型(Credit Monitor Model, 1993)，J. P. Morgan 以 VaR 为基础开发的信用度量技术(CreditMetrics)，麦肯锡公司(Wilson and McKinsey)开发的信贷组合分析模型(Credit Portfolio View, 1995)和瑞士信贷第一波士顿的信用风险附加法(CreditRisk + 模型, 1997)等。巴塞尔 III 允许银行使用复杂的统计模型进行信用风险建模，这大大激发了各家银行开发自己信用风险模型的兴趣。

目前，由于数据和技术等原因，国内大部分银行测度信用风险主要应用古典分析方法和现代度量信用风险模型，更复杂的模型很少涉及。本文应用广义线性混合模型(generalized linear mixed models,

简称 GLMMs) 对信用风险进行建模分析, 为商业机构测度信用风险提供借鉴和参考。Clayton (1996) [1] 进一步推广了 1972 年 Nelder 提出的广义线性模型, 考虑变量的随机效应后提出广义线性混合模型 (GLMMs)。McNeil 和 Wendin (2007) [2] 运用潜在因子方程和贝叶斯方法测度了几个固定效应和随机效应模型, 发现评级因素随着时间周期变动对违约概率产生影响。Stefanescu、Tunaru 和 Turnbull (2009) [3] 建立了一个信用评级流程模型来模拟评级变化模式并用贝叶斯方法做参数估计。Grzybowska、Karwanski 和 Orłowski (2012) [4] 使用 GLMM 对信用风险建模中的信用迁移矩阵进行分析。从国内文献来看, GLMMs 在金融领域的应用主要集中于风险投资、非寿险精算以及概率风险评价中, 如孟生旺和邱子真 (2016) [5] 在多水平费率因子厘定中使用广义线性混合模型进行估计。贺宝龙和唐湘晋 (2016) [6] 将广义线性混合模型应用到信度理论中。谢远涛和毛羽 (2016) [7] 利用广义线性混合模型估计非寿险准备金。但在信用风险领域, 与国际上相关应用较多不同的是, 无论是 GLM 还是 GLMMs, 国内研究几乎没有涉及, 这也是本文应用这个方法测度信用风险的初衷。由于信用风险数据在可获取性方面不如市场风险数据, 国内公开评级数据违约样本量太小, 银行内部具有足够样本进行建模, 但由于商业秘密的原因而不能公开获得。为了验证 GLMMs 对信用风险测度的有效性, 本文选取公开的标准普尔美国信用违约数据进行实证研究, 以期为信用风险测度提供新的思路和方法。

二、广义线性混合模型和违约风险应用

(一) 理论模型

违约风险可以通过宏观经济变量或潜在变量观察到, 给定这些因素, 假设每个公司的违约事件都是独立的, 可以使用伯努利混合模型 (Bernoulli mixture models) 对违约风险进行建模。

1. 伯努利混合模型。给定 p 维随机向量, 其中 $p < m$, $\Psi = (\Psi_1, \dots, \Psi_p)'$, 随机向量 $Y = (Y_1, \dots, Y_m)'$ 服从带有因子向量 Ψ 的伯努利混合分布。如果存在函数 $p_i : R^p \rightarrow [0, 1]$ $1 \leq i \leq m$, 在给定 Ψ 的条件下, 违约指标 Y 是一个独立的伯努利随机变量, 且 $P(Y_i = 1 | \Psi = \phi) = p_i(\phi)$ 。

对于 $y = (y_1, \dots, y_m)'$ 在 $\{0, 1\}^m$ 则有:

$$P(Y = y | \Psi = \phi) = \prod_{i=1}^m p_i(\phi)^{y_i} (1 - p_i(\phi))^{1-y_i}$$

其中, 违约指标 Y 的无条件分布是通过向量 Ψ 的积分而得到。

$$P(Y_1 = y_1, \dots, Y_m = y_m) = \int_{R^p} \prod_{i=1}^m p_i(\psi)^{y_i} (1 - p_i(\psi))^{1-y_i} dG(\psi)$$

其中, $y_1, \dots, y_m \in \{0, 1\}$, G 是 R^p 上的分布函数。

公司价值模型可以用混合模型来表示。假设在每个时间段末, 临界值 X_i (莫顿模型中的公司价值取对数 $\ln V_T$) 低于 d_i (莫顿模型中的负债取对数 $\ln B$) 时, 债务人 i 违约。KMV /CreditMetrics 就是其中两种使用莫顿结构的模型。这两个模型可以用混合模型来表示, KMV /CreditMetrics 是因子向量为 Ψ 的伯努利混合模型。给定 Ψ , $\varepsilon_1, \dots, \varepsilon_m$, 违约是独立的。考虑一个多期模型, 广义潜变量因子模型 (general latent variable factor model) 可以表示为:

$$X_{it} = b'_{it}F_t + \varepsilon_{it} \quad (1)$$

其中, $F_t \sim N_p(0, \Omega)$, $\varepsilon_{it} \sim N(0, 1 - \beta_{it})$, $\beta_{it} = \text{var}(b'_{it}F_t) = b'_{it}\Omega b_{it}$ 。 X_{it} 可以表示为莫顿模型中的 $\ln V_{it}$ 。在信用风险模型中, 违约发生与否, 由指示变量 $Y_{it} = I_{(x_{it} \leq d_{it})}$ 来决定, 条件违约概率可以表示为:

$$P(Y_{it} = 1 | F_t) = P(\varepsilon_{it} \leq d_{it} - b'_{it}F_t | F_t) = \Phi\left(\frac{d_{it}}{\sqrt{1 - \beta_{it}}} - \frac{b'_{it}F_t}{\sqrt{1 - \beta_{it}}}\right)$$

无条件违约概率可以表示为 $p_{it} = \Phi(d_{it})$ 。

决定组合信用风险的三个最重要的驱动因子是违约概率(PD)、违约损失率(LGD)和违约相关性。公司 i 和 j 之间的资产相关性由以下公式给出:

$$\rho(X_i, X_j) = \text{cov}(X_i, X_j) = E(X_i, X_j) = b'_{it}\Omega b_{jt}$$

实际模型的资产相关性在下文中具体展示。

2. GLMMs 的混合模型。在单周期模型中, 假设条件违约概率 $p_i(\Psi)$ 服从: $p_i(\Psi) = h(\mu + \beta'x_i + \Psi)$, 其中, h 是连接函数, 向量 x_i 包含了诸如公司特定信息或者行业信息, β 和 μ 是模型参数, 随机变量 $\Psi \sim N(0, \sigma^2)$ 。

该模型可以拓展为多期模型以便于计算不同时期的违约概率。考虑一系列混合变量 Ψ_1, \dots, Ψ_n 在每个时期 $t = 1, \dots, n$ 产生的违约依赖性。公司 i 在时期 t 的违约指标 $Y_{t,i}$ 服从概率为 $p_{t,i}(\Psi_t)$ 的伯努利分布, 给定 Ψ_t 则有: $p_{t,i}(\Psi_t) = h(\mu + x'_{t,i}\beta + \Psi_t)$, 其中, $\Psi_t \sim N(0, \sigma^2)$ 和 $x_{t,i}$ 为公司 i 在时期 t 的协变量。

(二) 计量模型

新巴塞尔协定允许银行使用内部评级模型对信用风险进行建模, 金融机构开始建立信用风险统计模型, 用于经济资本计量管理和风险管理。巴塞尔公布的方法显示, 渐进单风险因素模型是内评法的基础模型, 模型主要假设信用风险由单一的系统性风险因子引发, 信用资产组合之间相关性

为 0, 这意味着所有信用资产组合只受共同的宏观经济因素影响, 不受行业或者区域等集中度影响。很明显, 这两点在现实中是不可能满足的, 信用资产组合不可能仅受单一系统风险因子影响, 信用风险组合也不可能完全分散, 一般都存在集中度的情况。因此, 如何克服此模型的缺陷, 成为学术界和业界讨论的热点问题。GLMMs 巧妙地把系统风险因子分解为可观测的固定效应和不可观测的随机效应, 进而捕捉违约率的异质性。我们从简单的单因子模型开始建模, 并逐步

引入其他因子来进一步拓展信用风险模型。下面是四个递进的信用风险模型：

1. 模型 I：平均相关结构 (Equicorrelation) 的单因子模型。假设违约概率只和债务人在时间段 t 的信用评级 $r(i, t)$ 有关，并且 b 在所有时间段对所有债务人保持不变，可以得出：

$$P(Y_{it} = 1 | F_t) = \Phi\left(\frac{\Phi^{-1}(p_{r(i,t)})}{\sqrt{1-b^2}} - \frac{bF_t}{\sqrt{1-b^2}}\right)$$

其中, b 是共同因素负荷量 (common loading), 资产相关性为 $\rho = b^2$ 。这个模型可以用简单的 GLMM 形式表示：

$$P(Y_{it} = 1 | \Psi_t) = \Phi(\gamma_{r(i,t)} + \Psi_t)$$

其中, γ_r 是固定的评级效应, $\Psi_t \sim N(0, \sigma^2)$ 是时间段 t 上的随机效应。参数存在以下关系：

$$\rho = b^2 = \frac{\sigma^2}{1 + \sigma^2}; p_r = \Phi\left(\frac{\gamma_r}{\sqrt{1 + \sigma^2}}\right)$$

2. 模型 II：带有宏观经济协变量模型。信用事件由系统风险和非系统性风险共同引发, 给定系统性风险, 违约事件发生时非系统性风险之间是相互独立的, 这就是条件独立性假设。对于多笔债务的违约情况, 需要考虑一个企业违约引起另一个企业也违约的可能性, 即违约相关性。利用 GLMMs 来对信用风险建模时, 宏观周期可以作为固定效应, 而行业等因素可以作为随机效应出现在模型中。

由于模型 I 只考虑评级和时间, 通过增加可观测的宏观经济协变量 z_t 对模型 I 进行扩展。这里违约概率不仅仅和债务人在时间段 t 的信用评级 $r(i, t)$ 有关, 同时也和宏观经济变量 z_t 有关。在相同的时间段, 相同的评级债务人的违约概率相同。这个模型可以表示为：

$$P(Y_{it} = 1 | \Psi_t) = \Phi(\gamma_{r(i,t)} + \eta z_t + \Psi_t)$$

其中, η 相比较模型 I 来说, 是个额外的变量。在时间段 t , 隐含资产相关性为：

$$\rho = \frac{\sigma^2}{1 + \sigma^2}$$

3. 模型 III: 带有行业随机效应模型。在 GLMMs 中, 可以允许多个变量作为固定效应和随机变量, 除了宏观经济变量之外, 行业、区域、个体因素等均可以被拓展到模型中。现在我们考虑 (1) 的一个特例, $bit = bes(i, t)$, 其中, e_j 表示单位向量, $s(i, t)$ 表示行业或者区域。假设有 p 个行业以及 Ω 是参数为 ρ 的对称平均相关矩阵。

由于 $\beta_{it} = \text{var}(b'_{it}F_t) = b^2$, 我们仍旧假设违约概率只和信用评级 $r(i, t)$ 有关:

$$P(Y_{it} = 1 | F_t) = \Phi\left(\frac{\Phi^{-1}(p_{r(i,t)}) - bF_{it(i,t)}}{\sqrt{1-b^2}}\right)$$

这个模型的 GLMM 结构可以表示为: $P(Y_{it} = 1 | \Psi_t) = \Phi(\gamma_{r(i,t)} + \Psi_{it(i,t)})$ 。

其中, $\Psi_t = (\Psi_{t1}, \dots, \Psi_{tp})' \sim N_p(0, \Sigma)$, 其中 Σ 对角线元素为 $\sigma^2 + \tau^2$ 以及非对角线元素 σ^2 。综合上述两个模型, 可以得到:

$$\sigma^2 + \tau^2 = \text{var}(\Psi_{it}) = \frac{b^2}{1-b^2}; \sigma^2 = \text{cov}(\Psi_{it}, \Psi_{it'}) = \frac{\rho b^2}{1-b^2}$$

$$\text{其中, } b^2 = \frac{\sigma^2 + \tau^2}{1 + \sigma^2 + \tau^2}, \rho = \frac{\sigma^2}{1 + \sigma^2}, \rho_r = \Phi\left(\frac{\gamma_r}{\sqrt{1 + \sigma^2 + \tau^2}}\right)。$$

b^2 是隐含行业内资产相关性, ρ 是行业之间资产相关系数。

4. 模型 IV: 带有行业随机效应和宏观经济协变量模型。通过增加观测的宏观经济协变量 z_t , 可以进一步扩展模型 III。模型 IV 可以表示为:

$$P(Y_{it} = 1 | \Psi_t) = \Phi(\gamma_{r(i,t)} + \eta z_t + \Psi_{it(i,t)})$$

其中, $\Psi_t = (\Psi_{t1}, \dots, \Psi_{tp})' \sim N_p(0, \Sigma)$ 。这里, Σ 的对角线元素为 $\sigma^2 + \tau^2$, 非对角线元素为 σ^2 。综合上述两个模型, 可以得到:

$$\sigma^2 + \tau^2 = \text{var}(\Psi_{it}) = \frac{b^2}{1-b^2}; \sigma^2 = \text{cov}(\Psi_{it}, \Psi_{it'}) = \frac{\rho b^2}{1-b^2}$$

$$\text{其中, } b^2 = \frac{\sigma^2 + \tau^2}{1 + \sigma^2 + \tau^2}, \rho = \frac{\sigma^2}{\sigma^2 + 1}。$$

给定宏观经济协变量, b^2 是隐含行业内资产相关系数, 而 ρ 是行业间相关系数。

三、实证研究

GLMMs 的参数估计可以采用惩罚伪似然估计、边际伪似然估计、拉普拉斯、自适应高斯求积和贝叶斯、马尔科夫蒙特卡洛 (MCMC) 方法。拉普拉斯和自适应高斯求积在 Breslow 和 Clayto(1993) [8] 中有详细介绍, 其中有研究表明 PQL 结果好于 MQL。基于 MCMC 和 Gibbs Sampling 算法的贝叶斯估计方法也是常用手段之一, 贝叶斯

方法能提高估计精度，特别是对于处理低频率事件，原因在于其允许使用模型参数的主观先验分布。刘乐平和袁卫(2002) [9] 对现代贝叶斯方法在精算学中的应用做过相应讨论。实证研究表明，PQL 方法的有效性与贝叶斯方法比较接近，具体计算过程中有很多软件可以实现，比如 WinBugs、R 和 C 语言等。考虑到运行速度的便捷性，本文使用 R 语言中的 glmm 的惩罚伪似然估计方法来进行参数估计。

(一) 数据选择和统计分析

本文采用了标准普尔的数据库子集 CreditPro 6.6，样本来自美国 13 个行业的 6897 个债务人，共有 23 年违约数据。图 1 是经验违约概率，显示不同评级类别具有类似的信用违约周期，即不同评级在相同的时间段上有着类似的违约高峰期，而在其他相同时间段则违约率相对较低。

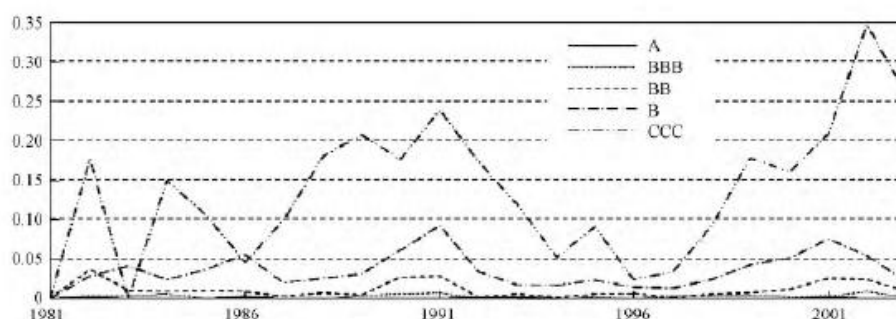


图 1 标准普尔 23 年的经验违约概率

在数据分析中，评级级别包括以下： $K = \{ CCC, B, BB, BBB, A \}$ ，我们将实际评级中的 $k +$ 、 k 、 $k -$ 合并为 k ，另外把 CCC、CC 和 C 级别合并为 CCC 级别。由于 AAA 和 AA 级别违约数据很少，这里没有考虑此两种级别，原因在于国内外部评级 AA 及以上违约数量极少，所以暂时违约概率建模存在一定困难。

(二) 宏观经济协变量选择

无论在学术界还是业界,存在几个常用的反应宏观经济周期变化的经济指标,这些指标可以解释违约概率中的周期效应。Chava、Stefanescu 和 Turnbull (2008) [10]使用 S&P 500(spret1) , GDP 和芝加哥联邦储备银行全国活动指数(CFNAI) 作为宏观经济解释变量,其中 CFNAI 也被 McNeil 和 Wendin (2007) [2] 使用。本文中,使用了包括芝加哥联邦储备银行全国活动指数在内的多个宏观经济协变量,主要包括 CFNAIMA3、SP500 一年的收益率和 GDP。根据违约模型 II 得出以下结果:

从表 1 可以看出, CFNAIMA3 和 spret1 的参数都很显著,但是 GDP 的结果并不显著。结果表明,使用 CFNAIMA3 和 spret1 比使用 GDP 更好,因为 CFNAIMA3 的 P 值更小,CFNAIMA3 比 spret1 更能很好地解释美国经济,它是解释信贷品质变化最重要的变量,在后续分析中将采用 CFNAIMA3 作为可观测因子。芝加哥联储国内活动指数(CFNAI) 是以 85 个现存美国经济活动为基础的月度加权平均指标,基本上是衡量美国经济活动最权威的指标之一,CFNAIMA3 是 CFNAI(用来追踪经济扩张和收缩的) 3 个月的移动平均数,CFNAI 指标与经济扩张和收缩时刻保持一致。

表 1 模型 II 带有宏观经济协变量 CFNAIMA3、spret1 和 GDP 的比较分析

	γ_A	γ_{BBB}	γ_{BB}	γ_B	γ_{CoC}	η_1	η_2	η_3
均值	-3.5406	-3.1306	-2.6952	-2.1739	-1.2465	-0.1643	-0.4319	0.0049
标准差	(0.1275)	(0.0970)	(0.0849)	(0.0775)	(0.0810)	(0.0686)	(0.1939)	(0.0175)
P 值	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0210	0.0312	0.7809

(三) 实证结果

本文采用惩罚伪似然估计进行参数估计,所有 4 个违约模型的实证结果如表 2 所示。违约事件发生的数量与所处的经济状态相关,

而宏观经济协变量是反映经济状态并用以捕捉经济周期的变化趋势。在本文中 CFNAIMA3 作为宏观经济协变量反应违约的可能性，结果显示 CFNAIMA3 的系数均为负，意味着随着 CFNAIMA3 的增长，违约的可能性减小。这里违约概率由 $P(Y_{it} = 1 | \Psi_t) = \Phi(\gamma_r(i) + \eta z_t + \Psi_t)$ 得到。

表 2 4 个违约风险模型实证结果

		γ_A	γ_{BBB}	γ_{BB}	γ_B	γ_{CCC}	η	σ	τ
模型 I	均值	-3.5456	-3.1352	-2.7002	-2.1817	-1.2517	—	0.2163	—
	标准差	(0.1062)	(0.069)	(0.0518)	(0.0394)	(0.0461)	—	—	—
	概率	0.0002	0.0011	0.0042	0.0165	0.1106	—	—	—
模型 II	均值	-3.5646	-3.1538	-2.7185	-2.1982	-1.2694	-0.1765	0.1681	—
	标准差	(0.1063)	(0.0675)	(0.0488)	(0.0346)	(0.0423)	(0.0426)	—	—
	p 值	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0001	—	—
模型 III	均值	-3.5806	-3.1774	-2.7606	-2.2445	-1.2834	0.1945	0.2571	—
	标准差	(0.0936)	(0.0626)	(0.0484)	(0.0391)	(0.0437)	—	—	—
	p 值	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	—	—	—
模型 IV	均值	-3.5959	-3.1929	-2.7760	-2.2589	-1.2992	-0.1616	0.1497	0.2549
	标准差	(0.0924)	(0.0604)	(0.0453)	(0.0348)	(0.0400)	(0.0434)	—	—
	p 值	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.0006	—	—

图 2 上方函数图展示了关于模型 I 中不可观察因素 $\{(\Psi_t) : t = 1, \dots, T\}$ (实线) 以及在模型 II 中的可观察因素 $\{(\Psi_t, z_t \eta) : t = 1, \dots, T\}$ (虚线)。下方函数图比较了模型 I (实线) 中不可观察因素和模型 II 的不可观察因素 (虚线)。可以看出， $\{(\Psi_t, z_t \eta) : t = 1, \dots, T\}$ 和 $\{(\Psi_t) : t = 1, \dots, T\}$ 这两条线似乎具有类似轨迹但并不完全相同，这说明 z_t 不能完全捕捉到违约信息。下方函数图比较了关于模型 I 和 II 不可观察因素，尽管它们的路径非常相似，但是可以观察到当 z_t 的拟合效果很好时， Ψ_t 的方差明显减少，标准差由模型 I 中的 0.2163 降到模型 II 中的 0.1681。说明模型 II 相对于模型 I 而言统计

意义上更好。

针对不能被可观测变量解释的因素，GLMMs 允许通过形式的变化，来获取更多的随机效应。在模型 III 中，添加了行业因素作为随机效应，将 CreditPro 公司分为 6 个行业。如表格 3 所示，加入行业因素后的模型，其最大对数似然值是 - 1818，而基本模型中的最大对数似然值为 - 1913. 3。根据实证结果计算，行业内部隐含资产相关性为 9. 41% ，而与之相对的跨行业资产相关性仅为 3. 42% 。无行业特定随机效应的模型总体隐含资产相关性 4. 47% 。这说明在某一偿债主体违约的情况下，相同行业内的公司更容易违约。模型 III 中加入行业随机效应能更好地捕捉不同行业的异质性。

模型 IV 在模型 III 的基础上加入宏观经济协变量 CFNAIMA3，结果也是显著的。这也说明行业因素比宏观周期因素对违约的影响更大，验证了现实中为了避免违约事件发生而对行业做出的选择性。

4 个模型属于层层递进关系，表 3 给出了统计变量 logLik、AIC 和 BIC。模型 I 和模型 II 的最大对数似然函数是相近的，而行业随机效应大幅增加了最大对数似然函数 logLik 从 - 1919. 3 和 -1925. 6 增加到 - 1818. 0 和 - 1817. 8，模型 III 和模型 IV 中 logLik 只增加了 0. 2 到 - 1817. 8。这个结果表明，行业随机效应可以帮助我们捕捉到无法仅靠观察宏观经济变量来解释的现象。在违约活动中，随机行业影响的贡献超过宏观经济协变量。

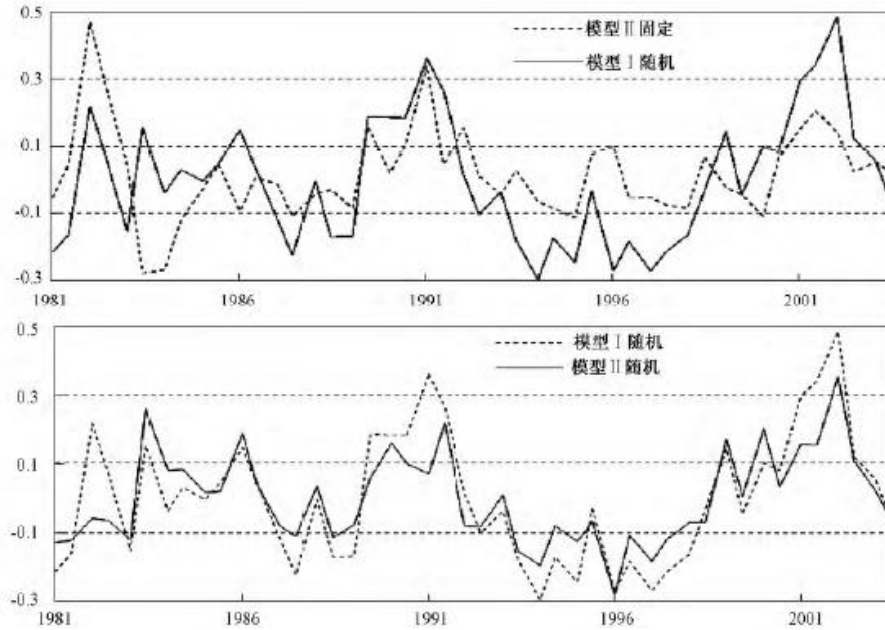


图 2 模型 I 和模型 II 系统性风险因子对比图

表 3 4 个模型的 LogLik, AIC 和 BIC 指标

模型	AIC	BIC	logLik
IV	3653.6	3700.6	-1817.8
III	3652.0	3693.8	-1818.0
II	3867.2	3909.0	-1925.6
I	3840.6	3877.2	-1913.3

四、结论

本文应用 GLMMs 模型对信用风险进行建模，并使用标准普尔数据进行了实证研究，通过模型分析可以得出以下结论：

(一) GLMMs 模型的可延展性

广义线性混合模型 (GLMMs) 巧妙地把系统风险因子分解为可观测的固定效应和不可观测的随机效应来捕捉违约率的异质性。该模型可以根据实际情况扩展到多变量模型，而且随机效应可以扩展到时间和区域等因子，用以捕获时间异质性以及跨越国家 (区域) 或行业的异质性。在国内具体应用该模型完全可以加入行业以及区域两个随机效应以捕捉行业和区域的异质性。由于经济结构和经济发展水平和程度不一致，行业与区域的因素差异很大，该模型正好可以解释信用风

险建模中我们最关心的几个因素。

（二）违约概率异质性捕捉

违约风险根据宏观经济协变量的变化而变动，宏观经济是违约概率变化的一个重要因素，在本文使用标准普尔违约数据的实证研究中发现 CFNAIMA3 是用来描述信用违约的已观察到的最优宏观经济协变量。随机效应的方差随着宏观经济变量变大而下降，意味着隐含资产相关性也随之下降。然而，宏观经济变量作为可观测变量无法全部解释违约率的异质性，借助广义线性混合模型中的随机效应有助于有效捕捉到违约率的异质性。

（三）关注行业内部资产相关性带来的风险

实证研究结果显示，考虑了特定行业的潜在因素，行业内部隐含资产相关性增加了，这表明信用风险模型应该考虑到行业之间的异质性特征。巴塞尔协定中关于集中度风险也提到过类似问题。在“资产荒”的背景下，金融企业为了避险，不约而同把资产投放到少数几个行业中，这将带来过高的违约相关性和集中度风险。

值得说明的是，由于信用违约数据的不可获得性，我们难以应用 GLMMs 计算我国有关金融机构信用风险情况，但是相关人员完全可以应用这个模型和数值计算方法，结合银行等金融机构的内部数据进行实际测算，以验证模型的可靠性和适应性。

参考文献

- [1] Clayton D. Generalized linear mixed models [J]. In: GilksW, RichardsonS, SpiegelhalterD. (Eds.), Markov Chain Monte Carlo in Practice [M]. London: Chapman & Hall, 1996. 275 — 301.
- [2] McNeil AJ, J Wendin. Bayesian inference for generalized linear mixed models of portfolio credit risk [J]. Journal of Empirical Finance 2007(14) : 131 — 149.
- [3] Stefanescu C, Tunaru R, TurnbullS. The credit rating process and estimation of transition probabilities: A Bayesian approach [J]. Journal of Empirical Finance, 2009(16) : 216 — 234.
- [4] Grzybowska U, Karwanski M, Orłowski A. Examples of Migration Matrices Models and their Performance in Credit Risk Analysis [J]. Acta Physica Polonica A, 2012(121) : B40 — B46.
- [5] 孟生旺, 邱子真. 混合效应模型及其在非寿险费率厘定中的应用 [J]. 数理统计与管理, 2016(1) : 154 — 161.
- [6] 贺宝龙, 唐湘晋. 广义线性混合模型在信度理论中的应用 [J]. 金融经济, 2008(20) .
- [7] 谢远涛, 毛羽. 基于进展时间和操作时间的两步广义线性混合模型非寿险准备金估计 [J]. 保险研究, 2016(11) : 68 — 77.
- [8] BreslowN, D Clayton. Approximate inference in generalized linear mixed models [J]. Journal of American Statistical Association, 1993 (88) : 9 — 25.
- [9] 刘乐平, 袁卫. 现代 Bayes 方法在精算学中的应用及展望 [J]. 统计研究, 2002(8) : 45 — 49.
- [10] Chava S, C Stefanescu, S Turnbull. Modelling expected loss [A]. Working paper, Bauer College of Business, University of Houston, 2008.

作者简介

胡志浩, 男, 2006年毕业于中国社会科学院研究生院, 获金融学博士学位, 现任中国社会科学院金融研究所副研究员、国家金融与发展实验室全球经济与金融研究中心主任。研究方向为国际金融市场分析。

卜永强, 男, 2014年毕业于英国赫瑞瓦特大学精算系, 获统计学博士学位, 现就职于中国平安人寿保险股份有限公司。研究方向为信用风险建模、市场风险建模和金融工程。