

文章编号:1003-207(2011)06-0010-05

# 信用风险中回收率分布的双 Beta 模型

王国栋<sup>1</sup>, 詹原瑞<sup>2</sup>

(1. 天津商业大学经济学院, 天津 300134; 2. 天津大学管理学院, 天津 300072)

**摘要:**本文研究了随机回收率的分布,建立了回收率的双 Beta 分布密度模型,它具有双峰分布的特点,这与 Moody 公司的最新研究相吻合,弥补了现有回收率分布模型均为单峰的不足。利用基于数论网格的序贯优化算法对所建模型的参数做出了估计,借助于核密度估计的工具,进行了实证分析,结果表明双 Beta 模型的拟合误差很小,远小于 Beta 模型的误差,它是表示回收率理想的模型。最后给出了抽取双 Beta 分布随机数的方法。

**关键词:**回收率;双 Beta 模型;核密度估计;双峰分布;序贯优化算法

**中图分类号:**F832 **文献标识码:**A

## 1 引言

信用风险是银行等金融机构面临的最主要风险。违约损失率是构成信用风险的基本要素,它指违约发生后,资产损失部分占违约暴露的比率,度量了违约事件中债权人损失的严重程度。回收率是违约后资产回收部分占违约暴露的比率,等于 1 减违约损失率,经常用回收率代替违约损失率进行讨论。与信用风险的另一基本要素违约率不同,对违约损失率或回收率的研究近几年才开始,目前对它们了解和研究得还不充分,相关研究仍处于深入过程中。我国银行业将逐步实施新巴塞尔资本协议,该协议要求对银行资产的违约率和违约损失率进行准确的估计,在此基础上计算资本要求,所以研究违约损失率或回收率的估计方法对新巴塞尔资本协议在我国顺利实施具有重要意义。

目前有关回收率的研究主要集中于回收率的影响因素、回收率与违约率的相关性、回收率的分布模型等方面。与违约率相比,回收率的影响因素更复杂。其中具代表性的是 2002 年 Moody 公司公布的 Losscale 模型。另外,Altman 和 Brady 等(2005)<sup>[1]</sup>,Grunert 和 Weber(2008)<sup>[2]</sup>,国内叶晓可等(2006)<sup>[3]</sup>,

汪办兴(2007)<sup>[4]</sup>,代太山等(2008)<sup>[5]</sup>利用相关数据资料研究了债券或贷款回收率的影响因素。

在回收率和违约率关系方面,很多研究都证实了违约率和违约损失率具有正相关性,如文献<sup>[6-8]</sup>。Altman 和 Brady 等(2005)<sup>[1]</sup>,Hu 和 Perraudin(2002)<sup>[6]</sup>,Rosch 和 Scheule(2005)<sup>[9]</sup>等都发现如果忽略违约率和回收率之间的相关关系,将会低估组合的损失,从而使得组合的资本要求也被低估。

由于多种因素的影响,回收率具有随机性。目前已经建立了多种回收率分布模型,其中最常用的是 Beta 模型。另外,Dullmann 和 Trapp(2004)<sup>[7]</sup>,Rosch 和 Scheule(2005)<sup>[9]</sup>讨论了 logit-正态分布模型,Frye(2000)<sup>[10]</sup>建立了回收率的正态分布模型,Pykhtin(2003)<sup>[11]</sup>提出 log-正态分布模型,Andersen 和 Sidenius(2004)<sup>[12]</sup>讨论了 probit-正态模型。Dullmann 和 Trapp(2004)<sup>[7]</sup>对正态模型、log-正态模型和 logit-正态模型一起做了研究。作者利用标准普尔公司 1982—1999 年间回收率数据,通过计算两种正态性检验即 Shapiro-检验和 Jarque-Bera-检验的 p-值,发现 log-正态模型拟合回收率方面的表现不如另外两种模型。

准确刻画回收率的随机性对于回收率的估计,以及确定单个资产和信用组合的潜在损失都有重要意义。Servigny 和 Renault(2004)<sup>[13]</sup>研究发现,如果忽略回收率的随机性而将其看作常数,常常会使得风险被低估。本文主要研究回收率的分布模型,在所有回收率的分布模型中,Beta 分布最常用,这是因为 Beta 分布有许多优点适合拟合回收率,如 Beta 分布随机变量定义在 $[0, 1]$ 区间上,符合回收

收稿日期:2009-12-20;修订日期:2011-08-20

基金项目:国家自然科学基金资助项目(70573076);国家社会科学基金资助项目(11BGL072)

作者简介:王国栋(1978-),男(汉族),山东济宁人,天津商业大学经济学院,讲师,研究方向:金融风险、金融工程。

率一般在 0% 到 100% 之间取值的特点; Beta 分布密度函数有两个参数可以调节, 能使得概率密度曲线表现出不同的形状, 如钟型(对称型)、U 型、J 型、左偏型、右偏型等, 见图 1<sup>[14]</sup>, 非常灵活; Beta 分布参数估计比较容易, 通过样本均值和方差就能作出估计等。许多著名的信用风险模型, 如 J. P. Morgan 公司开发的信用计量模型 CreditMetrics, KMV 公司开发的组合经理模型 Portfolio Manager 以及 Moody 公司的 Losscale 模型都假设回收率是服从 Beta 分布的。

Moody 公司最近利用自己的贷款和债券回收率数据库研究发现: 回收率的分布密度一般呈现双峰的特征, 实际回收率主要分布为: 要么较小为 20% 左右, 要么较大为 80% 左右, 见图 2<sup>[15]</sup>。而无论参数怎么变化, Beta 分布密度都不会是双峰的。Renault 和 Scaillet(2004)<sup>[14]</sup> 利用标准普尔 PMD 数据库的 1981—1999 年间违约债券数据, 分优先级和行业用 Beta 分布拟合了回收率, 通过与核密度估计作比较发现用 Beta 分布表示回收率的分布准确度较低。另外, 上述其它模型也都是单峰分布的, 所以它们一定程度上都不能准确地拟合回收率。为了弥补现有模型都不是双峰分布的不足, 本文建立双 Beta 分布模型用来表示随机回收率的分布, 它具有 Beta 分布的优点, 同时还具有双峰特征, 根据现实回收率数据进行拟合, 发现双 Beta 分布模型误差很小, 明显优于 Beta 分布模型。

## 2 双 Beta 分布模型及参数估计

Beta 分布密度函数为:

$$K_1(x; \alpha, \beta) = \begin{cases} \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, & 0 \leq x \leq 1 \\ 0, & \text{其它} \end{cases} \quad (1)$$

其中  $B(\alpha, \beta)$  是 Beta 函数。随着参数  $\alpha, \beta$  取值的不同, Beta 密度曲线可以表现出多种形状, 如图 1<sup>[14]</sup> 所示, 但这些曲线都不是双峰的, 为此建立双 Beta 分布模型。

双 Beta 分布模型为:

$$K_2(x; \alpha_1, \beta_1, \alpha_2, \beta_2, \rho) = \begin{cases} \rho \frac{1}{B(\alpha_1, \beta_1)} x^{\alpha_1-1} (1-x)^{\beta_1-1} \\ + (1-\rho) \frac{1}{B(\alpha_2, \beta_2)} x^{\alpha_2-1} (1-x)^{\beta_2-1}, & 0 \leq x \leq 1 \\ 0, & \text{其它} \end{cases} \quad (2)$$

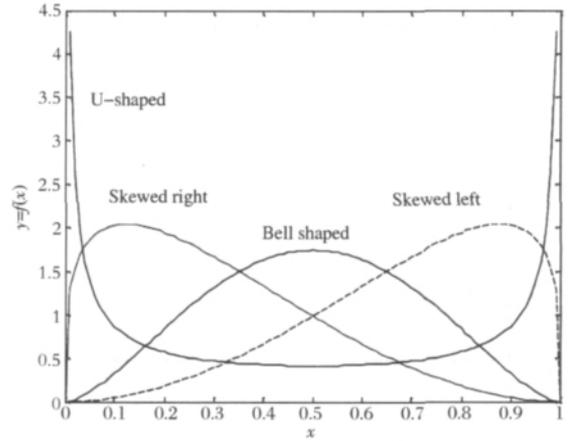


图 1 多种形状的 Beta 分布密度曲线

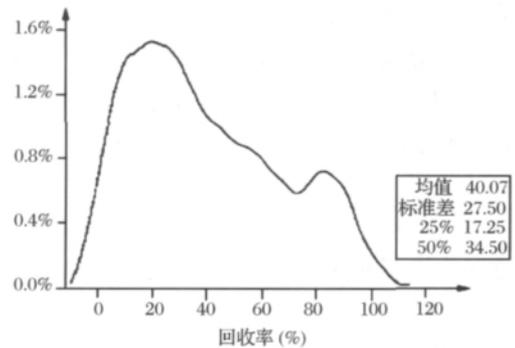


图 2 Moody 公司 1970—2002 年第二季度间所有债券和贷款回收率

其中权重  $\rho$  也是需要估计的参数,  $0 \leq \rho \leq 1$ 。  $K_2$  由两个不同的 Beta 密度函数加权得到, 是非负的, 在 0 到 1 之间的积分值等于 1, 也是概率密度函数。当  $\rho$  等于 0 或 1 时,  $K_2$  退化为 Beta 密度函数, Beta 模型可看作双 Beta 模型的特例, 所以  $K_2$  具有 Beta 模型的优点, 此外它还具有双峰特征, 例如  $\alpha_1 = 4, \beta_1 = 10, \alpha_2 = 8, \beta_2 = 3, \rho = 0.65$  时, 双 Beta 密度曲线如图 3 所示。

下面对双 Beta 模型的参数进行估计。作对数似然函数:

$$LL(x_1, \dots, x_n; \alpha_1, \beta_1, \alpha_2, \beta_2, \rho) = \sum_{i=1}^n \ln \left[ \frac{\rho}{B(\alpha_1, \beta_1)} x_i^{\alpha_1-1} (1-x_i)^{\beta_1-1} + \frac{1-\rho}{B(\alpha_2, \beta_2)} x_i^{\alpha_2-1} (1-x_i)^{\beta_2-1} \right] \quad (3)$$

求极大似然估计常用的数值算法是 Newton-Raphson 迭代法及其推广算法, 但这种算法当参数个数增加时会变得很复杂, 而且还可能会出现迭代不收敛或仅收敛到局部极大值的情况。本文采用另外一种方法, 即基于数论网格(NT-net)的序贯优化

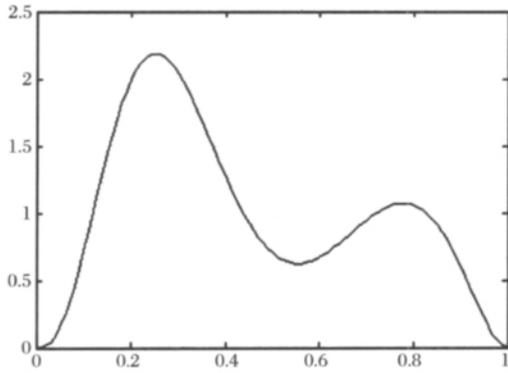


图3 双 Beta 密度曲线

算法<sup>[16,17]</sup>。它不需对数似然函数关于参数的偏导数组成的方程组,而直接在参数空间的某些均匀散布点集中搜索求似然函数的极大值,即使需要估计参数个数增加,估计也不会变得更复杂,对多个参数做估计具有明显优势。假设我们要估计  $s$  个参数,首先确定这  $s$  个参数估计值的大致范围为  $s$  维矩形:

$$[a, b] = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_s, b_s]$$

在  $s$  维矩形里面找参数的极大似然估计值点。

步骤如下:

(1) 初始:令  $t = 1, D^{(1)} = [a, b], a^{(1)} = a, b^{(1)} = b$ 。

(2) 产生 NT-net:用某一数论方法产生  $n_t$  个点的点集  $P^{(t)}$  使其在  $[a^{(t)}, b^{(t)}]$  上均匀散布。

(3) 计算新的近似值:求  $\alpha^{(t)} \in P^{(t)} \cup \alpha^{(t-1)}$  使得

$$LL(\alpha^{(t)}) \geq LL(x), \forall x \in P^{(t)} \cup \alpha^{(t-1)}$$

此处  $\alpha^{(0)}$  为空集,  $\alpha^{(t)} = (\alpha_1^{(t)}, \alpha_2^{(t)}, \dots, \alpha_s^{(t)})$  是当前极大似然估计最好的近似值。

(4) 停止准则:令  $c^{(t)} = (b^{(t)} - a^{(t)})/2$ , 若  $\max c^{(t)} < \delta$ , 这里  $\delta$  为预先给定的很小的正数,则  $D^{(t)}$  已足够小,  $\alpha^{(t)}$  被接受为近似估计值,停止。若  $\max c^{(t)} \geq \delta$ , 进入下一步。

(5) 收缩区域:定义一个新的区域  $D^{(t+1)} = [a^{(t+1)}, b^{(t+1)}]$  如下:

$$a_i^{(t+1)} = \max(\alpha_i^{(t)} - \gamma c_i^{(t)}, a_i)$$

$$b_i^{(t+1)} = \min(\alpha_i^{(t)} + \gamma c_i^{(t)}, b_i), i = 1, \dots, s$$

其中  $\gamma$  为预先定义的收缩比。令  $t = t + 1$  回到步骤 2。

序贯优化算法求参数的极大似然估计时需要参数的一个初始值,可以把参数的矩估计值作为初始值,因为一般情况下矩估计值和极大似然估计值在

参数空间中的距离不会太远,而且求矩估计值也相对容易。

### 3 核密度估计

为了比较双 Beta 模型和 Beta 模型的精确度,需要借助核密度估计。核密度估计是一种概率密度的非参数估计方法,它是由 Rosenblatt(1956)<sup>[18]</sup> 和 Parzen(1962)<sup>[19]</sup> 首先提出并研究的,这种重要方法是古老的直方图方法的自然发展<sup>[20]</sup>。核密度估计不需预先假定分布密度的具体形式,在一定意义下,它随抽取样本个数的增加而逐渐逼近真实的概率密度。核密度估计有多种类型<sup>[21]</sup>,如 Beta 核密度估计、标准 Gauss 核密度估计、转换 Gauss 核密度估计等,可以利用其中较简单的 Beta 核密度估计。

首先 Beta 核函数定义为  $K_1(u; x/b + 1, (1 - x)/b + 1), x \in [0, 1]$ ,  $K_1$  为 Beta 密度函数,如(1)式定义。 $b$  称为带宽,满足  $n \rightarrow \infty$  时,  $b \rightarrow 0$ 。设  $X_1, \dots, X_n$  是取自总体密度  $f(x)$  的样本, Beta 核密度定义为:

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_1(X_i; x/b + 1, (1 - x)/b + 1) \tag{4}$$

Chen(1999)<sup>[22]</sup>证明了 Beta 核估计具有均方相合性,即  $\lim_{n \rightarrow \infty} E[\hat{f}_n(x) - f(x)]^2 = 0$ 。用 Beta 核密度估计近似代替回收率的真实分布密度,以它为标准比较 Beta 分布模型和双 Beta 分布模型。可以从数值上比较两个模型拟合误差的大小,定义积分均方差为:

$$ISE_i = \int_0^1 (K_i(x) - \hat{f}_n(x))^2 dx, i = 1, 2 \tag{5}$$

### 4 实证分析

由于很难获得国内贷款或债券回收率的公开数据,我们利用 Altman 和 Kishore(1996)<sup>[23]</sup> 收集的 1971—1995 年间部分违约债券回收率数据对 Beta 模型和双 Beta 模型分别进行检验。数据描述如表 1 所示:

表 1 回收率数据的统计描述

	个数	均值	标准差	中位数
回收率	73	0.4123	0.1726	0.3778

首先用序贯优化算法对两个模型的参数进行极大似然估计,估计结果见表 2 和表 3:

表 2 Beta 模型的参数估计结果

$\alpha$	$\beta$	$\max(LL)$
1.9487	2.4549	12.2710

表 3 双 Beta 模型的参数估计结果

$\alpha_1$	$\beta_1$	$\alpha_2$	$\beta_2$	$\rho$	$\max(LL)$
1.2347	0.8290	4.3432	6.8669	0.1852	31.0584

然后用 Beta 核密度模型对回收率进行拟合,按照文献<sup>[14]</sup>的做法,带宽取为  $b = \text{回收率数据标准差} \times n^{-2/5}$ 。将 Beta 模型、双 Beta 模型和 Beta 核密度模型拟合的回收率概率密度曲线绘制成图,如图 4 所示:

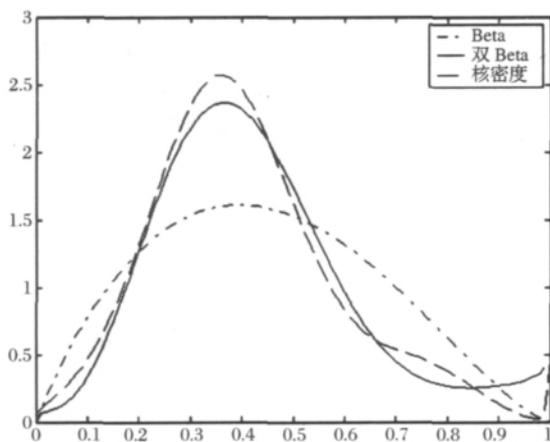


图 4 三种模型拟合的回收率概率密度曲线

通过与 Beta 核密度对比不难发现,双 Beta 模型的拟合结果明显好于 Beta 模型。尤其双 Beta 模型能够捕捉回收率分布尾部的变化,而 Beta 模型做不到这一点。进一步利用 6 个节点的 Gauss-Legendre 数值积分公式,求得拟合的误差为:

$$ISE_1 = 0.2747, ISE_2 = 0.0195$$

双 Beta 模型的拟合误差很小,而 Beta 模型拟合误差为双 Beta 模型的 14 倍,双 Beta 模型是表示回收率的理想模型。

## 5 双 Beta 模型的随机抽样

在信用风险管理和度量的时候,常常需要从预知分布密度中抽取回收率随机数,最后给出双 Beta 模型随机抽样方法。对于一些常见分布如正态分布, Beta 分布等,在工具软件如 Matlab 中有专门的随机数生成器函数,抽取这些分布的随机数比较容易。然而对于双 Beta 分布,从中抽取随机数并没有

这么直接的办法,可以按照下面的合成方,保证抽取的随机数服从双 Beta 分布<sup>[24]</sup>:

(1) 根据上面(2)式将双 Beta 分布密度表示为  $K_2 = \rho g_1(x) + (1-\rho)g_2(x)$ , 其中  $g_1(x), g_2(x)$  表示 Beta 密度函数。

(2) 由  $[0, 1]$  区间上的均匀分布密度  $U(0, 1)$  中抽取随机数  $u$ 。

(3) 如果  $u \leq \rho$ , 由  $g_1(x)$  中抽取随机数  $x$ ; 如果  $u > \rho$ , 由  $g_2(x)$  中抽取随机数  $x$ 。

(4)  $x$  就是服从双 Beta 分布的随机数。

## 6 结语

Moody 公司最新研究结果表明回收率一般具有双峰分布,常见的回收率分布模型如正态、logit-正态、log-正态、probit-正态,以及最常用的 Beta 模型都是单峰分布的。本文建立了双 Beta 分布模型,它以 Beta 模型为特例,具有 Beta 模型的优点,此外它还是双峰分布的,弥补了现有回收率模型都是单峰的不足。用基于数论网络的序贯优化算法对模型参数进行了估计,这种方法不需对参数求偏导数和解方程组,对较多参数的估计具有明显优势,克服了参数个数增加给估计带来的困难。论文借助 Beta 核密度进行实证分析,结果表明双 Beta 模型拟合的积分均方误差很小,远小于 Beta 模型的误差,它是表示回收率的理想模型。最后给出了双 Beta 模型随机抽样方法。本文给出的模型可以方便地用于贷款和债券等回收率的度量和管理中。

## 参考文献:

- [1] Altman, E., Brady, B., Resti, A. etc. The link between default and recovery rates: Theory, empirical evidence, and implications[J]. Journal of Business, 2005, 78(6): 2203-2227.
- [2] Grunert, J., Weber, M.. Recovery rates of commercial lending: empirical evidence for German companies[J]. Journal of Banking and Finance, 2008, 33(3): 505-513.
- [3] 叶晓可, 刘海龙. 银行不良贷款违约损失率结构特征研究[J]. 上海管理科学, 2006, (6): 12-15.
- [4] 汪办兴. 中国银行贷款违约损失率影响因素的实证分析[J]. 经济评论, 2007, (3): 90-93.
- [5] 代太山, 陈敏, 杨晓光. 不同担保类型之下违约损失率的结构特征: 针对中国的实证[J]. 南方经济, 2008, (8): 28-39.
- [6] Hu, Y., Perraudin, W.. The dependence of recovery rates and defaults[R]. Working paper, Birkbeck College, 2002.

- [7] Dullmann, K., Trapp, M.. Systematic risk in recovery rates—an empirical analysis of U. S. corporate credit exposures[R]. Working paper, 2004, February.
- [8] Gupton, G., Stein, R.. Losscalc V2: dynamic prediction of LGD[R]. Working paper, Moody's KMV, January 2005.
- [9] Rosch, D., Scheule, H.. A multi-factor approach for systematic default and recovery risk[J]. The Journal of Fixed Income, 2005, 15(2):63—75.
- [10] Frye, J.. Depressing recoveries[J]. Risk, 2000, 11: 108—111.
- [11] Pykhtin, M.. Unexpected recovery risk[J]. Risk, 2003, 8:74—78.
- [12] Andersen, L., Sidenius, J.. Extension to the Gaussian copula: Random recovery and random factor loadings [J]. Journal of Credit Risk, 2004, 1(1): 29—70.
- [13] Servigny, A., Renault, O.. Measuring and Managing Credit Risk[M]. McGraw-Hill, 2004.
- [14] Renault, O., Scaillet, O.. On the way to recovery: A nonparametric bias free estimation of recovery rate densities[J]. Journal of Banking and Finance, 2004, 28: 2915—2931.
- [15] 詹原瑞. 银行内部评级的方法与实践[M]. 北京: 中国金融出版社, 2009.
- [16] 方开泰, 王元. 数论方法在统计中的应用[M]. 北京: 科学出版社, 1996.
- [17] 张亚静. Beta 分布参数的极大似然估计[J]. 山西大学学报(自科版), 1998, 21(2): 113—118.
- [18] Rosenblatt, M.. Remarks on some nonparametric estimates of a density function[J]. The Annals of Mathematical Statistics, 1956, 27: 832—837.
- [19] Parzen, E.. On estimation of a probability density function and mode[J]. The Annals of Mathematical Statistics, 1962, 33: 1065—1076.
- [20] 陈希孺, 方兆本, 李国英, 等. 非参数统计[M]. 上海: 上海科学技术出版社, 1989.
- [21] Gouriéroux, C., Monfort, A.. Nonconsistency of the Beta kernel estimator for recovery rate distribution [R]. Working paper, University of Toronto, 2006, 1—3.
- [22] Chen, S.. Beta kernel estimators for density functions [J]. Computational Statistics & Data Analysis, 1999, 31: 131—145.
- [23] Altman, E., Kishore, V.. Almost everything you wanted to know about recoveries on defaulted bonds[J]. Financial Analysts Journal, 1996, Nov/Dec, 57—64.
- [24] 茆诗松, 王静龙, 濮晓龙. 高等数理统计[M]. 北京: 高等教育出版社, 2006.
- [25] 陈暮紫, 黄意球, 陈敏, 杨晓光. 基于 logistic 模型族的不良贷款极端零回收强度模型研究[J]. 中国管理科学, 2011, (2): 16—23.
- [26] 马宇超, 陈暮紫, 陈敏, 杨晓光. 大规模同质不良贷款资产组合的回收率估计方法[J]. 管理科学学报, 2009, (2): 86—96.

### Double Beta Distribution Model of Recovery Rate in Credit Risk

WANG Guo-dong<sup>1</sup>, ZHAN Ruan-rui<sup>2</sup>

(1. School of Economics, Tianjin University of Commerce, Tianjin 300134, China;

2. School of Management, Tianjin University, Tianjin 300072, China)

**Abstract:** In this paper, the distribution of random recovery rate is studied, and double Beta distribution density model is established. This model has a characteristic of two peaks, which meets the new findings of Moody company, and improves the result that the present models all have only one peak. By using the sequential optimization algorithm based the NT-net to estimate parameters and the tools of kernel density estimation, the demonstration study is made. The result shows that the fitting error of double Beta model is quite little and far less than that of Beta model, so double Beta model is an ideal model to denote the distribution of recovery rate. The method is given to draw the random numbers from double Beta model at the end of the paper.

**Key words:** recovery rate; double Beta distribution; kernel density estimation; two peak distribution; sequential optimization algorithm