



فصلنامه علمی پژوهشی دانش سرمایه‌گذاری  
سال نهم / شماره سی و چهارم / تابستان ۱۳۹۹

## مدل‌سازی ریسک اعتباری مشتریان بانک با استفاده از مدل تحلیل بقا مبتنی بر روش اسپلاین

محمدعلی رستگار

استادیار دانشگاه تربیت مدرس، دانشکده مهندسی صنایع و سیستم‌ها، (نویسنده مسئول)  
ma\_rastegar@modares.ac.ir

مهدی عیدی گوش

دانش آموخته کارشناسی ارشد مهندسی صنایع گرایش سیستم‌های مالی، دانشگاه تربیت مدرس  
eidimahdi@ymail.com

تاریخ دریافت: ۹۶/۱۰/۲۵ تاریخ پذیرش: ۹۷/۰۵/۰۲

### چکیده

امروزه بانک‌های کشور با معضلات جدی به لحاظ نوع دارایی‌هایشان مواجه هستند. از جمله عواملی که منجر به این وضعیت شده‌اند می‌توان به کیفیت بد دارایی‌های بانک‌ها اشاره داشت که علت آن را می‌توان نداشتن سیستم رتبه‌بندی و ارزیابی درست در ریسک اعتباری دانست.

در این پژوهش با استفاده از مدل رگرسیون کاکس و همچنین مدل بقای رگرسیون لجستیک مبتنی بر اسپلاین به پیش‌بینی احتمال نکول در طول زمان پرداخته ایم. برای مدل‌سازی ریسک اعتباری با استفاده از این دو روش از ۱۰ متغیر مربوط به ۲۸۶۱ نفر از مشتریان یک بانک ایرانی استفاده گردیده است. پس از انجام مدل‌سازی با توجه به نتایجی که از مقایسه این دو مدل با استفاده از روش ROC به دست آمد، مدل رگرسیون کاکس با معیار  $AUC=0.799$  از کارایی بالاتری نسبت به مدل بقای رگرسیون لجستیک مبتنی بر اسپلاین با معیار  $AUC=0.746$  برخوردار است.

**واژه‌های کلیدی:** ریسک اعتباری، رگرسیون کاکس، رگرسیون لجستیک مبتنی بر اسپلاین، ROC

## ۱- مقدمه

یکی از مشکلات اساسی شبکه بانکی و موسسات اعتباری در حال حاضر رشد چشمگیر مطالبات معوق، دارایی‌های مشکوک‌الوصول، تسهیلات سررسید گذشته و... است. بانک‌های کشور به دلیل عدم نقد شدن تسهیلاتی که در گذشته پرداخت کرده‌اند، جریان نقدی ورودیشان با مشکل مواجه شده است که این کاهش جریان ورودی نقدینگی باعث ایجاد ریسک نقدینگی در بانک‌ها شده است. دلایل رشد بالای مانده مطالبات معوق را می‌توان ناشی از عوامل درونی و بیرونی شبکه بانکی جستجو نمود. از جمله مهمترین عوامل مرتبط با شبکه بانکی به عملکرد بانک‌ها در نحوه ارزیابی مشتریان خود باز می‌گردد که در این مسیر ارزیابی رتبه اعتباری مشتریان از اهمیت بسیاری برخوردار است. مهمترین ابزاری که بانک‌ها و موسسات اعتباری برای مدیریت و کنترل ریسک اعتباری بدان نیازمند هستند، سیستم رتبه‌بندی اعتباری مشتریان است. دستورالعمل‌های مختلفی از جمله مقررات بازل ۱، بازل ۲ و بازل ۳ استانداردهای جدید بازار جهت مدیریت ریسک اعتباری و کفایت سرمایه برای بانک‌ها را برشمرده است. بانک مرکزی جمهوری اسلامی ایران نیز در راستای کیفیت بخشی به بانک‌ها مقرراتی را برای بانک‌ها در نظر گرفته است. همانطور که اشاره شد یکی از مهمترین معضلاتی که بانک‌ها با آن مواجه هستند نداشتن مدل‌های ارزیابی مشتریان خود به دلیل نبود موسسات رتبه‌بندی در کشور است. لذا بانک‌ها باید از مدل‌های خود برای رتبه‌بندی مشتریان خود استفاده کنند. مدل‌های مختلفی برای اندازه‌گیری ریسک اعتباری معرفی شده‌اند از جمله مدل‌های شبکه عصبی، مدل تحلیل تمایزی، مدل نزدیکترین همسایه، سیستم‌های خبره و ... که اکثر این مدل‌ها به دنبال برآورد احتمال نکول در پایان دوره هستند و تغییرات احتمال نکول در طول زمان را بررسی نمی‌کنند. از جمله مدل‌هایی که به بررسی تغییرات در طول زمان می‌پردازند می‌توان به مدل‌های تحلیل بقا اشاره کرد.

در این پژوهش سعی شده است با معرفی مدل بقای رگرسیون لجستیک مبتنی بر اسپلاین و مدل رگرسیون کاکس که در دسته مدل‌های تحلیل بقا جای می‌گیرند مدلی بهینه به منظور برآورد احتمال نکول مشتریان ارائه شود.

## ۲- بحث و بررسی

## ۲-۱- پیشینه تحقیق

برای اولین بار نارین<sup>۱</sup> (1992) شروع به استفاده از مدل‌های تجزیه و تحلیل بقا کرد. نارین مدل تحلیل بقا را برای اطلاعات وام در بازه زمانی ۲۴ ماه، اعمال کرد که نتایج به دست آمده نشان داد رویکرد تجزیه و تحلیل بقا اطلاعات دقیق‌تر و مرتبط‌تری برای مدیریت اعتبارات نسبت به رویکردهای معمول فراهم می‌کند. از دیگر پژوهش‌های انجام گرفته پژوهش توماس و همکاران (۱۹۹۲) است که در آن مدلی که نارین استفاده کرده بود را با استفاده از مدل شتاب دهنده مدت<sup>۲</sup> برای اطلاعات وام در بازه زمانی ۲۴ ماه اعمال کردند. نتایج به دست آمده از پژوهش نشان دادند که رویکرد تحلیل بقا نسبت به روش‌های متعارف در تصمیم‌گیری در مورد اعتبار در صورتی که نمره اعتباری توسط مدل پشتیبانی شود می‌تواند بهتر عمل کند. در این پژوهش توماس همچنین

مقایسه‌ای بین مدل‌های نیمه پارامتریک وایبول و کاکس با مدل رگرسیون لجستیک انجام داد که نتایج نشان داد مدل نیمه پارامتریک وایبول و کاکس که از مدل‌های تحلیل بقا هستند ابزار مدل‌سازی بهتری هستند. ریکاردو کوا و همکاران (۲۰۰۹) در مقاله‌ای تحت عنوان مدل ریسک اعتبار مصرف کننده از طریق تجزیه و تحلیل بقا از سه روش برای تخمین احتمال نکول استفاده کرده‌اند، در روش اول مدل خطرهای متناسب کاکس استفاده شده است، در دومین روش از مدل‌های خطی تعمیم یافته و سومین روش با مدل رگرسیون غیر پارامتریک تصادفی ساخته شده است. نتایج این پژوهش نشان داد مدل رگرسیون کاکس از عملکرد بهتری برخوردار است.

سازا میلر<sup>۳</sup> (۲۰۱۰) از مدل رگرسیونی که توسط کاکس (۱۹۷۴) ارائه شد برای پیش بینی تاثیر مشخصات وام گیرنده برای احتمال و یا ریسک نکول استفاده کرده است. این مدل یک فرم زمانی ثابت را بر روی ویژگی های مشاهده شده اعمال می‌کند. این مورد ضروری است که بتوان متغیرهای کمکی تغییرات نرخ خطر، بالا یا پایین رفتن آن را در طول عمر یک وام شناسایی کرد.

ماریا استپانوس و همکاران (۲۰۰۲) در مقاله‌ای تحت عنوان روش‌های تحلیل بقا برای اطلاعات وام شخصی روش‌های تجزیه و تحلیل بقا را به اطلاعات شخصی از یک موسسه بزرگ مالی انگلیس که ۵۰ هزار متقاضی وام شخصی را با دوره بازپرداخت ۳۶ ماهه ارائه می‌داد مورد استفاده قرار دادند. نویسندگان این مقاله از باقیمانده های مدل های اسنل-کاکس، مارتینگل، دیویانس<sup>۴</sup> به عنوان ابزار تشخیصی برای بررسی مدل مناسب استفاده کرده‌اند.

توماس و همکاران در سال ۲۰۰۶ پژوهشی با عنوان "مدل‌سازی ریسک اعتباری سبد وام"<sup>۵</sup> انجام دادند. آنها در این پژوهش اطلاعات مربوط به وام‌های مصرفی و همچنین متغیرهای کلان اقتصادی را بر روی مدل کاکس منطبق کردند و به این نتیجه رسیدند که وارد کردن متغیرهای کلان اقتصادی در پیش‌بینی رفتار اعتباری مشتریان تاثیر مثبتی دارد.

سیرانگ لو و همکاران (۲۰۱۶) در پژوهش خود که با عنوان "مدل بقای مبتنی بر اسپلاین برای مدل‌سازی ریسک اعتباری" انجام دادند، از دو روش رگرسیون کاکس و رگرسیون لجستیک مبتنی بر روش اسپلاین به منظور برآورد احتمال نکول مشتریان کارت‌های اعتباری استفاده کردند. آنها برای بررسی کارایی مدل از منحنی ROC استفاده کردند. که در پایان به این نتیجه رسیدند روش رگرسیون اسپلاین مبتنی بر اسپلاین از عملکرد بهتری برخوردار است.

همچنین می‌توان به پژوهشی که کرانی و آقای (۱۳۹۰) به منظور برآورد احتمال نکول مشتریان حقیقی تسهیلات پژوهشی را با عنوان کاربرد نظریه تحلیل بقا در مدیریت ریسک اعتباری دریافت کنندگان تسهیلات انجام دادند. آنها در این پژوهش از متغیر مبلغ وام به عنوان متغیر تبیینی استفاده کردند و به مقایسه روش پیشنهادی خود با روش رگرسیون کاکس پرداختند که در پایان به این نتیجه رسیدند که روش رگرسیون کاکس از عملکرد بهتری نسبت به روش پیشنهادی آنها برخوردار است.

ابراهیمی و دریابر (۱۳۹۱) در مقاله‌ای با عنوان مدیریت ریسک اعتباری در نظام بانکی- رویکرد تحلیل

پوششی داده‌ها و رگرسیون لجستیک و شبکه عصبی به بررسی و مقایسه روش‌های تحلیل پوششی داده‌ها، رگرسیون لجستیک و شبکه عصبی جهت به دست آوردن روشی مناسب به منظور رتبه بندی مشتریان حقوقی بانک پرداخته اند.

پژوهش‌های انجام شده در زمینه تحلیل بقا به اختصار در جدول ۱ آورده شده است:

جدول ۱: پژوهش‌های انجام شده در زمینه تحلیل بقا

نام	سال	موضوع
Narain	۱۹۹۲	استفاده از مدل‌های تحلیل بقا به منظور تجزیه و تحلیل داده‌های طول عمر
Cao at el	۲۰۰۹	استفاده از روش‌های کاکس، مدل‌های خطی و مدل رگرسیون غیر پارامتریک تصادفی
Sarah Miller	۲۰۱۰	استفاده از مدل رگرسیون کاکس به منظور برآورد احتمال نکول
Luo at el	۲۰۱۶	استفاده از مدل بقای رگرسیون لاجیت مبتنی بر اسپلاین

## ۲-۲- ارائه مدل پیشنهادی

در این بخش ابتدا به معرفی مدل رگرسیون کاکس می‌پردازیم و سپس مدل بقای لجستیک مبتنی بر اسپلاین که مدل پیشنهادی در این پژوهش است پرداخته خواهد شد. سیرانگ لو و همکاران (۲۰۱۶) نیز از این مدل استفاده کرده‌اند.

از مدل کاکس بیشتر در مواردی که نیاز به سانسور کردن داده‌ها باشد استفاده می‌شود. این مدل می‌تواند زمان‌های بقا (پیوسته و گسسته) و همچنین ترکیب ریسک‌ها را مدل کند. از آنجایی که در این مدل نیازی به مشخص کردن تابع توزیع زمان‌های بقا نیست، یک مدل نیمه پارامتریک است. از ویژگی‌های این مدل این است که می‌توان متغیرهای وابسته به زمان را نیز وارد مدل کرد که می‌تواند کمک به بهبود مدل پیش‌بینی نماید. تابع احتمال نکول در مدل کاکس به صورت رابطه ۱ تعریف می‌شود:

$$h(t|\vec{x}) = h_0(t)e^{\vec{\beta}\vec{x}} \quad \text{رابطه ۱}$$

که در آن:

$h(t|\vec{x})$ : احتمال اینکه قرض گیرنده در زمان  $t$  نکول کند

$h_0(t)$ : یک تابع پایه برای  $\vec{X} = 0$  که یک برآورد غیر پارامتریک از داده‌ها است

$\vec{\beta}$ : ضرایب متغیرهای کمکی  $X$  است

$$\frac{h(t|\vec{x}_1)}{h(t|\vec{x}_2)} = e^{\vec{\beta}(\vec{x}_1 - \vec{x}_2)} \quad \forall (t > 0) \quad \text{رابطه ۲}$$

همانطور که در رابطه ۲ مشاهده می شود واضح است که نسبت دو خطر مستقل از زمان  $t$  است. این مدل نرخ مخاطره نامیده می شود. در بعضی موارد متغیرهای کمکی می توانند به صورت تابعی از زمان داده شوند مانند متغیرهای کلان اقتصادی، نرخ بیکاری و تغییر نرخ بهره در طول زمان. ترکیب این متغیرهای کلان اقتصادی می تواند به بهبود دقت وصحت پیش بینی کمک کند.

مدل کاکس به کمک متغیرهای وابسته به زمان می تواند به شکل رابطه ۳ در بیاید:

$$h(t | \vec{X}(t)) = h_0(t) e^{\beta \vec{X}(t)} \quad \text{رابطه ۳}$$

مدل بقای لجستیک مبتنی بر اسپلین (SLRS): به منظور توضیح بیشتر این مدل یک نمونه سبد وام با متغیرهای کمکی  $X(t)$ ،  $T$  زمان رویداد که به صورت گسسته،  $D$  که نشانگر سانسور و  $M$  نشان دهنده نوع رویداد است را در نظر بگیرید.  $T$  نشان دهنده مدت زمان از تاریخ شروع تا تاریخ نتیجه رویداد و  $M$  نشان دهنده نتایج مختلفی که می تواند به وجود بیاید ( $M = 1$  نکول اتفاق بیافتد،  $M = 0$  به عنوان وضعیت باز) است. بنابراین  $M$  نشان دهنده ترکیب ریسکها در داده ها است.

توزیع مشترک احتمالی  $(T, M)$  وقوع نتایج وام وابسته به زمان را توصیف می کند، فرض کنید در زمان  $c$  ما داده هایی برای  $n$  داده به صورت زیر مشاهده می کنیم:

$$\left\{ t_i, d_i, m_i, x_i(t_i) \right\}_{i=1}^n$$

$t_i$ : زمان بقا  
 $d_i$ : نماد سانسور  
 $m_i$ : نوع رخداد  
 $x_i(t_i)$ : تابع متغیرهای وابسته

به عنوان مثال اگر وام  $i$  در زمان  $c$  نکول نکرده باشد و وضعیت باز داشته باشد مشخصات آن به این صورت خواهد بود:  $\left\{ t_i = c, d_i = 1, m_i = 0, x_i(t_i) \right\}$ . با این فرض که توزیع مشترک  $(T, M)$  مستقل از زمان مشاهده  $c$  است، نرخ خطر را می توان برای  $(T, M)$  از داده های بقا برآورد کرد.

با توجه به دادهها ما می توانیم توزیع  $(T, M)$  را تعریف کنیم:

$f(t, m)$  تابع چگالی،  $F(t, m)$  تابع تجمعی،  $S(t, m)$  تابع بقا و  $h(t, m)$  تابع خطر تعریف می کنیم.

برای رخداد  $M$  تعریف می‌کنیم:

$$f(t, m) = F(t, m) - F(t-1, m) \quad \text{رابطه ۴}$$

$$F(t, m) = \Pr\{T \leq t, M = m | X(t)\} = \sum_{j=1}^t f(j, m) \quad \text{رابطه ۵}$$

$$S(t, m) = F(\infty, m) - F(t, m) \quad \text{رابطه ۶}$$

$$h(t, m) = \Pr\{T = t, M = m | T \geq t, X(t)\} = \frac{f(t, m)}{1 - F(t-1)} \quad \text{رابطه ۷}$$

ایده اصلی برای مدل بقای رگرسیون لجستیک (LRS) چندان پیچیده نیست. تابع بقای هر وام را می‌توان به مجموعه‌ای از نقاط گسسته در زمان که هر کدام مشاهدات متمایزی هستند شکست. با این مشاهدات ادغام شده ما می‌توانیم برای پیش‌بینی رویداد در هر زمان از رگرسیون لجستیک استفاده کنیم (آلیسون ۲۰۱۰). مدل LRS را می‌توان برای آنالیز داده‌های تحلیل بقا به کار برد. علاوه بر این متغیرهای کمکی وابسته به زمان را نیز می‌توان به راحتی در این مدل جایگذاری کرد. از همه نکات مهمتر این است که می‌توان مدل Splin را در LRS جایگذاری کرد که این مورد انعطاف‌پذیری مدل را نسبت به اثر زمان خطر بیشتر می‌کند. برای استفاده از مدل بقای رگرسیون لجستیک مبتنی بر اسپلاین ما نیاز به گسترش داده‌ها داریم به گونه‌ای که اطلاعات مربوط به هر داده برای هر نقطه زمانی ثبت شده باشند، یعنی داده‌ها تا قبل از توقف وجود داشته باشند. به عنوان مثال در جدول ۲ فرض کنید ۳ وام  $i$ ،  $i+1$ ،  $i+2$  در بازه زمانی ۶ ماه بررسی داشته باشیم. داریم وام  $i$  در دوره ۲ و وام  $i+1$  در دوره ۴ نکول می‌کنند و وام  $i+2$  تا پایان بررسی نکول نکرده است.

جدول ۲: داده‌های مورد استفاده در مدل بقای رگرسیون لجستیک مبتنی بر اسپلاین

متغیرهای وابسته	رویداد (Mi)	زمان (T)	وام (i)
$x_i(1)$	0	1	i
$x_i(2)$	1	2	i
$x_{i+1}(1)$	0	1	i+1
$x_{i+1}(2)$	0	2	i+1
$x_{i+1}(3)$	0	3	i+1
$x_{i+1}(4)$	1	4	i+1
$x_{i+2}(1)$	0	1	i+2
$x_{i+2}(2)$	0	2	i+2
$x_{i+2}(3)$	0	3	i+2

متغیرهای وابسته	رویداد (Mi)	زمان (T)	وام (i)
$\vec{x}_{i+2}(4)$	0	4	i+2
$\vec{x}_{i+2}(5)$	0	5	i+2
$\vec{x}_{i+2}(6)$	0	6	i+2

با توجه به اطلاعات گسترده می‌توان مدل بقای رگرسیون لجستیک مبتنی بر اسپلاین را به صورت رابطه ۸ درآورد:

$$\text{Ln} \left( \frac{h(t | \vec{x}(t))}{1 - h(t | \vec{x}(t))} \right) = \Theta(t, \vec{x}(t), \vec{\xi}) \quad \text{رابطه ۸}$$

که در آن  $h(t | \vec{x}(t))$  احتمال شرطی رخداد نکول در زمان  $t$  با توجه به متغیرهای کمکی  $\vec{x}(t)$  است.

در این پژوهش برای تعیین احتمال نکول از رگرسیون لاجیت استفاده شده است. اگرچه در این روش، برآورد ضرایب از بسیاری جهات شبیه رگرسیون معمولی است اما نحوه برآورد آنها کاملاً متفاوت است. ضرایب مدل لاجیت به روش متداول حداکثر درست‌نمایی برآورد می‌شود. البته قابل ذکر است که اگر در تخمین رگرسیون لجستیک، نسبت شانس  $\frac{p}{1-p}$  قابل محاسبه باشد، می‌توان از روش حداکثر مربعات رگرسیونی استفاده کرد (عرب مازار، ۱۳۶۶). در مدل‌های دودویی<sup>۶</sup>، برای تخمین متغیر وابسته دو حالت (۰، ۱) که مقدار صفر و یک را به خود اختصاص می‌دهد، می‌توان از مدل لاجیت<sup>۷</sup> استفاده نمود. اگر فرض کنیم که  $Y$  متغیر تصادفی با مقادیر صفر و یک باشد، در این صورت احتمال وقوع  $Y$  را می‌توان به صورت روابط زیر در نظر گرفت.

$$P(Y = 1) = h(t | \vec{x}(t)) = \frac{\exp(\Theta(t, \vec{x}(t), \vec{\xi}))}{1 + \exp(\Theta(t, \vec{x}(t), \vec{\xi}))} \quad \text{رابطه ۹}$$

$$P(Y = 0) = (1 - h(t | \vec{x}(t))) = \frac{1}{1 + \exp(\Theta(t, \vec{x}(t), \vec{\xi}))} \quad \text{رابطه ۱۰}$$

تابع پارامتری  $\Theta(t, \vec{x}(t), \vec{\xi})$  نشان‌دهنده اثر زمان و محدودیت‌های وابسته به زمان است. مدل بقای رگرسیون لجستیک مبتنی بر اسپلاین می‌تواند وابستگی بین نتایج وام و پیش‌بینی را نشان دهد.

از آنجایی که شکل تابع خطر می‌تواند غیر خطی و دارای تیزی باشد ما می‌توانیم برای مدل‌سازی آن از مدل زیر استفاده کنیم:

$$\vec{\xi} = \left\{ \vec{\alpha}, \vec{\beta} \right\}$$

$$\Theta(t, \vec{x}(t), \vec{\alpha}, \vec{\beta}) = \alpha_1 x_1(t) + \dots + \alpha_p x_p(t) + S(t, \vec{\beta}) \quad \text{رابطه ۱۱}$$

به طوری که  $\vec{\alpha} = \{\alpha_1, \dots, \alpha_p\}$  بردار پارامترهای متغیر وابسته به زمان و  $\vec{\beta}$  بردار پارامترهای رگرسیون اسپلاین است.

این فرمول هم اثرات خطی زمان و هم اثرات غیر خطی را شامل می‌شود. رگرسیون اسپلاین  $S(t, \vec{\beta})$  به مدل انعطاف پذیری بیشتری نسبت به اثرات مختلف زمانی مانند تیزی روند صاف که شکل تابع خطر را تشکیل می‌دهند می‌دهد.

رگرسیون اسپلاین: از مدل اسپلاین بیشتر برای درون‌یابی استفاده می‌شود. از جمله مزایای این مدل می‌توان به دقت بالای این مدل در درونیابی چند جمله‌ای درجه بالا اشاره کرد که با افزایش درجه مدل میزان دقت و انعطاف‌پذیری این مدل افزایش می‌یابد. از مدل اسپلاین برای تقریب خیلی از موارد مانند شاخص صنعتی داوجونز و تابع تقاضا در بازاریابی استفاده شده است. این توابع به صورت قطعه قطعه از درجه  $n$  تعریف می‌شوند. در این رگرسیون نقاط اتصال بین قطعه‌ها گره نامیده می‌شود. یک رگرسیون اسپلاین متشکل از چند بخش مکعبی (یا درجه دوم) و یا بخش انتهایی خطی است که به یکدیگر متصل می‌شوند. برای به دست آوردن رگرسیون اسپلاین ما به انتخاب درجه تابع و همچنین نقاط گره داریم. رگرسیون اسپلاین را می‌توان با افزودن گره‌های بیشتر، تغییر مکان گره‌ها یا افزایش درجه چند جمله‌ای انعطاف‌پذیرتر کرد. برای یک سبد وام معمولاً نرخ خطر ثابت باقی می‌ماند، با این حال اگر تابع خطر تجربی دارای روند خاصی باشد یا فراز و نشیب باشد می‌توان با استفاده از رگرسیون اسپلاین آن را مدل کرد.

داریم:

$$S(t, \vec{\beta}) = \beta_0 + \sum_{j=1}^k \beta_j \tilde{S}(t, t_j) \quad \text{رابطه ۱۲}$$

تابع اسپلاین از درجه  $d$  و نقاط گره  $(t_i; i=1, 2, \dots, k)$  است که شرط پیوستگی آن مشتق‌پذیری آن است. از علامت  $\ll + \gg$  (به این معنی که اگر مثبت بود اعمال می‌شود و اگر منفی بود صفر در نظر گرفته می‌شود) استفاده می‌کنیم. به منظور بیان کلی از تابع از درجه  $d$  و نقاط گره  $k$  استفاده می‌کنیم.

$$S(t) = \sum_{j=0}^d \beta_{0j} t^j + \sum_{i=1}^k \sum_{j=0}^d \beta_{ij} (t - t_j)_+^j \quad \text{رابطه ۱۳}$$



حال با فرض  $k=3$  و  $d=3$  تابع  $S(t)$  را به دست می‌آوریم:

$$S(t) = \beta_{00} + \beta_{01}t + \beta_{02}t^2 + \beta_{03}t^3 + \beta_{10}(t-t_1)_+^0 + \beta_{11}(t-t_1)_+^1 + \beta_{12}(t-t_1)_+^2 + \beta_{13}(t-t_1)_+^3 + \beta_{20}(t-t_1)_+^0 + \beta_{21}(t-t_1)_+^1 + \beta_{22}(t-t_1)_+^2 + \beta_{23}(t-t_1)_+^3 + \beta_{30}(t-t_1)_+^0 + \beta_{31}(t-t_1)_+^1 + \beta_{32}(t-t_1)_+^2 + \beta_{33}(t-t_1)_+^3 \quad \text{رابطه ۱۴}$$

برای اینکه پیوستگی برقرار باشد باید مشتق اول و دوم و سوم در نقاط اتصال برابر باشد پس داریم:

$$\beta_{10} = \beta_{11} = \beta_{12} = \beta_{20} = \beta_{21} = \beta_{22} = \beta_{30} = \beta_{31} = \beta_{32} = 0 \quad \text{رابطه ۱۵}$$

در نتیجه داریم:

$$S(t) = \beta_{00} + \beta_{01}t + \beta_{02}t^2 + \beta_{03}t^3 + \beta_{13}(t-t_1)_+^3 + \beta_{23}(t-t_1)_+^3 + \beta_{33}(t-t_1)_+^3 \quad \text{رابطه ۱۶}$$

حال با توجه به اینکه می‌دانیم نرخ خطر بعد از مدتی روند ثابتی می‌گیرد، محدودیت‌های زیر را اضافه می‌کنیم:

$$\begin{aligned} \beta_{03} + \beta_{13} + \beta_{23} + \beta_{33} &= 0 \\ \beta_{02} - 3\beta_{13}t_1 + 3\beta_{23}t_2 - 3\beta_{33}t_3 &= 0 \\ \beta_{01} + 3\beta_{13}t_1^2 + 3\beta_{23}t_2^2 - 3\beta_{33}t_3^2 &= 0 \end{aligned}$$

با جایگذاری  $\{\beta_{01}, \beta_{02}, \beta_{03}\}$  در معادله،  $S(t)$  به صورت زیر نوشته می‌شود:

$$S(t) = \beta_{00} + \beta_{13}((t-t_1)_+^3 - t^3 + 3t_1t^2 - 3t_1^2t) + \beta_{23}((t-t_2)_+^3 - t^3 + 3t_2t^2 - 3t_2^2t) + \beta_{33}((t-t_3)_+^3 - t^3 + 3t_3t^2 - 3t_3^2t) \quad \text{رابطه ۱۷}$$

برآورد حداکثر درست‌نمایی: به منظور برآورد ضرایب تابع لاجیت از روش حداکثر درست‌نمایی استفاده می‌کنیم. تابع درست‌نمایی به صورت زیر محاسبه می‌شود:

چنانچه هر مشاهده را یک آزمون برنولی فرض کنیم، در این صورت برای مشاهده  $i$ ام رابطه زیر برقرار خواهد بود:

$$P(Y = y_i) = P_i^{y_i} (1 - p_i)^{1-y_i} \quad (y_i = 0, 1) \quad \text{رابطه ۱۸}$$

که در آن احتمال وقوع پیشامد مورد نظر در مشاهده  $i$ ام و  $y_i$  نیز مقدار متغیر تصادفی است که می‌تواند برحسب مورد، صفر یا یک باشد (یک برای وقوع نکول و صفر برای عدم وقوع نکول). همچنین با فرض اینکه

مشاهده مستقل باشد، در اینصورت تابع درست‌نمایی نیز به شرح رابطه ۱۹ خواهد بود و با جایگذاری  $h(t_i | \vec{x}_i(t_i))$  بجای  $p_i$  در رابطه ۱۸ به رابطه ۲۰ خواهیم رسید.

$$L = \prod_{i=1}^n P_i^{y_i} (1 - p_i)^{1-y_i} \quad \text{رابطه ۱۹}$$

$$L = \prod_{i=1}^n \left( \frac{\exp(\Theta(t, \vec{x}(t), \vec{\xi}))}{1 + \exp(\Theta(t, \vec{x}(t), \vec{\xi}))} \right)^{y_i} \left( \frac{1}{1 + \exp(\Theta(t, \vec{x}(t), \vec{\xi}))} \right)^{(1-y_i)} \quad \text{رابطه ۲۰}$$

سپس با لگاریتم گرفتن طبیعی از رابطه ۲۰ خواهیم داشت:

$$\ln(L) = 1 = \sum_{i=1}^n y_i \ln\left(\frac{\exp(\Theta(t, \vec{x}(t), \vec{\xi}))}{1 + \exp(\Theta(t, \vec{x}(t), \vec{\xi}))}\right) + \sum (1 - y_i) \ln\left(\frac{1}{1 + \exp(\Theta(t, \vec{x}(t), \vec{\xi}))}\right) \quad \text{رابطه ۲۱}$$

بدین ترتیب برآورد ضرایب مستقل (یعنی  $\vec{\xi}$ ) از طریق حداکثر کردن رابطه ۲۱ به دست می‌آید که توسط مشتق‌گیری نسبت به هریک از ضرایب متغیرهای مستقل و مساوی صفر قرار دادن هر یک از مشتق‌ها محاسبه می‌شود. البته بسته‌های نرم‌افزاری برای محاسبه این تابع وجود دارند که می‌توان از آنها استفاده کرد. روش ROC (کرزانوسکی)<sup>۸</sup>: مشخصه عملیاتی گیرنده یک منحنی است، که نرخ مثبت درست، در برابر نرخ مثبت نادرست را به وسیله یک آستانه<sup>۹</sup> رده بندی کننده  $t$  در یک سیستم رده بندی کننده به نمایش می‌گذارد. فرض کنید  $t$  مقدار آستانه‌ای در یک سیستم رده‌بندی کننده باشد، اگر نمره رده‌بندی کننده یک شخص  $s$  بیشتر از  $t$  باشد، به جامعه  $P$  نسبت داده شده و در غیر این صورت به جامعه  $N$  نسبت داده خواهد شد. در این صورت در عمل ۴ حالت می‌تواند اتفاق بیفتد:

احتمال اینکه یک شخص از جامعه  $P$  به درستی رده‌بندی شده باشد. یعنی اگر شخص عضو جامعه  $P$  است، سیستم رده بندی کننده نیز آن را به درستی به جامعه  $P$  نسبت دهد.

$$TP = p(s > t | P),$$

احتمال این که یک شخص از جامعه  $P$  به طور نادرستی رده‌بندی شده باشد. یعنی اگر شخص عضو جامعه  $P$  است، سیستم رده‌بندی کننده آن را به نادرستی به جامعه  $N$  نسبت دهد.

$$FP = p(s > t | N),$$

• احتمال این که یک شخص از جامعه  $N$  به طور درستی رده‌بندی شده باشد.

$$TN = p(s \leq t | N),$$

- احتمال این که یک شخص از جامعه N به طور نادرستی رده‌بندی شده باشد.

$$FN = p(s \leq t | P),$$

تغییرات t و ارزیابی همه ۴ کمیت پیش گفته به طور آشکاری اطلاعات کاملی در ارزیابی عملکرد رده‌بندی کننده می‌دهد؛ اما چون  $TP + FN = 1$  و  $FP + TN = 1$  تمام اطلاعات را نیاز نداریم. همان طور که در تعریف ROC اشاره شد، مشخصه عملیاتی گیرنده، منحنی است که نرخ مثبت درست در برابر نرخ مثبت نادرست را نشان می‌دهد.

نسبت  $\frac{TP}{TP + FN}$  حساسیت و نسبت  $\frac{TN}{TN + FP}$  ویژگی نامیده می‌شود، چنانکه یک رده بندی کننده دارای  $TP = 1$  و  $TN = 1$  باشد، از آن به عنوان یک رده بندی کننده ایده آل نام می‌بریم که به هیچ وجه مرتکب خطا نشده و همیشه دارای مقدار حساسیت و ویژگی یک است. از ترسیم کمیت های یادشده در برابر یکدیگر یک خم نتیجه می‌شود که از آن به عنوان مساحت زیر منحنی (AUC) نام می‌بریم که بیشترین مقدار آن یک است که در حالت رده‌بندی کننده کامل اتفاق می‌افتد و در حالت رده‌بندی کننده تصادفی مقدار ۰٫۵ را اختیار می‌کند. بدیهی است، برای عملکرد خوب سیستم های رده‌بندی به نرخ درست بالا و نرخ نادرست پایین نیاز داریم.

### ۳- تحلیل داده‌ها

در این پژوهش از اطلاعات مربوط به ۲۸۶۱ نفر از مشتریان یک بانک ایرانی استفاده شده است. از اطلاعات ۲۰۰۰ نفر به عنوان داده های پیش‌بینی و از مابقی به عنوان داده‌های پس‌آزمایی استفاده شده است. به منظور برآورد احتمال نکول از عوامل موثر بر ریسک مشتریان اعتباری که در ادامه توضیح داده شدند استفاده شده است.

بر اساس مقررات بازل ۲ سیستم رتبه بندی داخلی بانک باید دو بعد مجزا داشته باشد:

(۱) عوامل ریسکی مربوط به ویژگی‌های قرض گیرنده

(۲) عوامل ریسکی مربوط به اطلاعات تسهیلات و نوع معامله.

در این توافقنامه اشاره می‌شود که بعد اول باید ریسک نکول شدن قرض گیرندگان را تعیین نماید. به عبارت دیگر تسهیلات مختلفی که به یک قرض گیرنده پرداخت می‌شود همگی از یک رتبه اعتباری یکسان برخوردارند. بعد دوم سیستم رتبه بندی داخلی سایر اطلاعات غیر از ویژگی قرض گیرنده نظیر نوع وثیقه، نرخ سود، مدت وام، مبلغ وام، نوع صنعت و سایر عواملی که از یک معامله به معامله دیگر متغیر است را دربر می‌گیرند (پاراگراف ۳۹۸ مقررات بازل ۲). به دلیل محدودیت در ارائه اطلاعات توسط بانک، تنها داده‌های خامی که در این بخش جمع‌آوری گردیده است، عبارت است از: مبلغ وام، سابقه فعالیت در شعبه، نرخ سود، امتیازی که مسئولین مربوط در بانک به وام‌گیرنده اختصاص داده اند، سابقه اشتغال در حرفه مشتری، وضعیت محل اصلی مشتری، حسن شهرت، نوع وثیقه، نسبت بدهی به دارایی و زمان وام. این متغیرها به اختصار نام‌گذاری شده‌اند که در

جدول ۳ آمده است. همانطور که در جدول ۳ مشاهده می‌شود، محدوده داده‌ها با هم تفاوت آشکاری دارد از این رو برای استفاده از این داده‌ها در مدل آن‌ها را به روش خطی تغییر معیار داده و بین (0,1) تعریف می‌کنیم و سپس از آن‌ها استفاده می‌کنیم.

جدول ۳: مقادیر بیشترین، کمترین و میانگین هر متغیر

نام متغیر	نماد اختصاری	MAX	MIN	AVREGE	$\sigma$
زمان (ماه)	T	12	3	6.4225	58.819
سابقه فعالیت در شعبه (ماه)	A	360	1	75.4745	55.75
نرخ سود* ۱۰۰	B	2250	1	77.5805	115.240
امتیاز* ۱۰۰	C	9613	4674	7246.996	30571.996
مبلغ وام (ریال)* ۱۰ <sup>۳</sup>	D	454027	22	8851.813	94.200
سابقه اشتغال در حرفه مشتری	E	600	5	138.7855	34.376
وضعیت محل اصلی مشتری	F	100	25	56.368	14.590
حسن شهرت	G	100	25	80.5535	26.198
نوع وثیقه	H	100	۲۵	44.638	451.307
نسبت بدهی به دارایی	I	6739	6	834.932	4.705
زمان وام (ماه)	J	36	3	6.9365	58.819

### ۳-۱- مدل‌سازی رگرسیون کاکس

پس از مرتب‌سازی داده‌ها به منظور مدل‌سازی، با استفاده از نرم افزار R این مدل بر روی داده‌ها برازش شد که نتایج آن در جدول ۵ آمده است. همچنین با استفاده از روش گام به گام (که در آن از معیار آکائیکه به منظور حذف متغیرهای اضافی استفاده شده است، مقادیر AIC بعد از حذف هر متغیر در جدول ۴ آمده است) متغیرهای اضافی حذف شده و پارامترهای مدل برازش شدند. نتایج این مدل‌سازی در جدول ۵ آمده است.

پس از برآورد ضرایب رگرسیون کاکس احتمالات نکول مربوط به هر زمان را محاسبه کرده و سپس با استفاده از روش ROC مقدار  $AUC=0.799$  برای این تابع به دست آمد (مقدار AUC هر چه به یک نزدیکتر باشد نشان‌دهنده کارایی بهتر مدل است). نمودار تابع خطر  $h(x)$  همراه با نمودار احتمال نکوال واقعی برای داده‌های آزمایش در شکل ۱ رسم شده است.

جدول ۴: ضرایب متغیرهای مدل رگرسیون کاکس

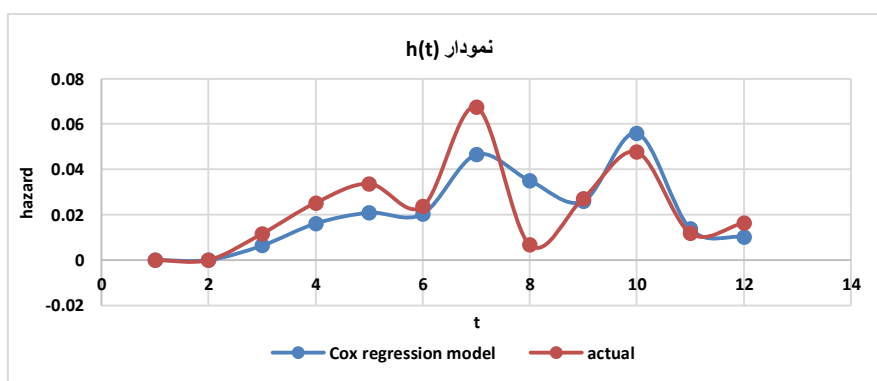
	coef	exp(coef)	se(coef)	z	Pr(> z )	AIC	Level of deleted variable	
A	-6.158	0.00211	1.567	-3.929	$8.54 \times 10^{-5}$			***
B	-0.291	0.74738	19.20	-0.015	0.987	2296.25	1	
C	2.399	11.0207	0.649	3.693	0.00022			***
D	0.537	1.7115	0.849	0.632	0.527	2290.51	5	
E	0.334	1.396	0.592	0.564	0.573	2292.79	3	
F	-0.312	0.731	0.272	-1.147	0.251	2290.51	5	
G	-0.637	0.528	0.429	-1.485	0.1375	2289.48	7	
H	-0.365	0.693	0.2918	-1.252	0.210	2289.99	6	
I	-0.606	0.545	1.223	-0.496	0.620	2294.52	2	
J	3.622	37.424	0.4264	8.495	$< 2e \times 10^{-16}$			***

منبع: داده‌های پژوهش

جدول ۵: نتایج به دست آمده از مدل رگرسیون کاکس

	coef	exp(coef)	se(coef)	z	Pr(> z )	
A	-5.2014	0.005509	1.46527	-3.55	0.000386	***
C	1.656277	5.239767	0.43845	3.778	0.000158	***
J	3.695881	40.28103	0.418206	8.837	$< 2e-16$	***
Concordance= 0.653 (se = 0.024 )						
Rsquare= 0.035 (max possible= 0.692 )						
Likelihood ratio test= 70.75 on 3 df, p=2.998*10 <sup>-15</sup>						
Wald test = 109.6 on 3 df, p=0						
Score (logrank) test = 124 on 3 df, p=0						

منبع: داده‌های پژوهش



شکل ۱: نمودار احتمال نکول در زمان t مدل رگرسیون کاکس

منبع: داده‌های پژوهش

## ۳-۲- بقای رگرسیون لجستیک مبتنی بر اسپلین (SLRS)

مدل اصلی که ما به دنبال مقایسه آن با مدل رگرسیون کاکس هستیم مدل بقای رگرسیون لجستیک مبتنی بر اسپلین است. به منظور برآورد احتمال نکول با استفاده از این روش ما باید نقاط تغییر تابع را پیدا کنیم که برای این کار از تابع احتمال نکول داده های واقعی استفاده می‌کنیم. همانطور که در شکل ۲ مشخص است، نقاط  $t=\{6,7,9,10,12\}$  را برای قسمت اسپلین مورد استفاده قرار می‌دهیم و همچنین نقاط  $k=\{7,9,10\}$  را به عنوان متغیرهای اسپایک در نظر می‌گیریم. در این مدل از اسپلین عادی استفاده شده است زیرا به دلیل استفاده از روش گام به گام نیازی به استفاده از مدل جریمه در اسپلین نیست.

به منظور برآورد تابع احتمال نکول با استفاده از این روش ابتدا با استفاده از نرم‌افزار فورترین داده‌ها را باز کرده به صورتی که مقادیر متغیرها برای تمامی نقاط (زمان‌ها) موجود باشد. سپس با استفاده از تابع  $gml$  در نرم‌افزار R این مدل را اجرا کرده و ضرایب متغیرها را به دست می‌آوریم که نتایج در جدول ۷ آورده شده است. برای بهتر شدن نتایج مانند مدل قبل برای این مدل نیز از روش گام به گام استفاده کرده تا متغیرهای اضافی حذف شوند (مقادیر AIC مربوط به هر مرحله در جدول ۷ آورده شده است). مقادیر به دست آمده برای ضرایب هر یک از متغیرها در جدول ۶ آورده شده است.

همانطور که مشاهده می‌شود تعداد متغیرها پس اجرای روش گام به گام از ۱۸ متغیر به ۹ متغیر کاهش یافته که خود تاثیر بسزایی بر سرعت محاسبات دارد.

پس از برآورد ضرایب مدل بقای رگرسیون لجستیک مبتنی بر اسپلین احتمالات نکول مربوط به هر زمان را محاسبه کرده و سپس با استفاده از روش  $ROC$  مقدار  $AUC=0.746$  برای این تابع به دست آمد. نمودار تابع خطر  $h(x)$  همراه با نمودار احتمال نکوال واقعی برای داده‌های آزمایش در شکل ۳ رسم شده است.

جدول ۶: نتایج بدست آمده از مدل بقای رگرسیون لجستیک مبتنی بر اسپلین

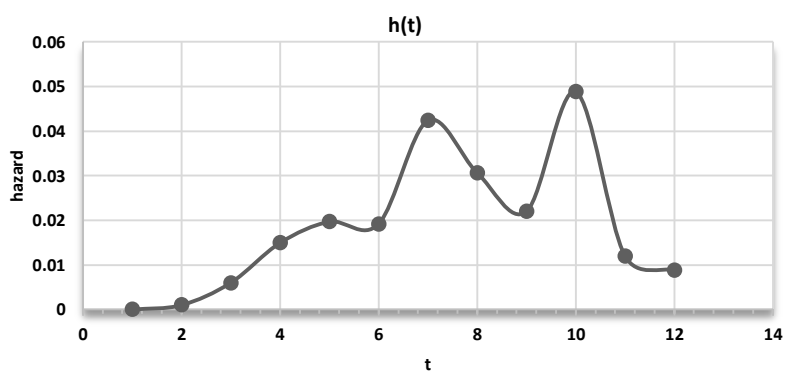
	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-6.3248	0.3714	-17.032	$< 2*10^{-16}$	***
A	-5.3508	1.5026	-3.561	0.000369	***
C	1.7196	0.4507	3.815	0.000136	***
J	3.9129	0.4809	8.137	$4.06*10^{-16}$	***
S6	-91.9583	24.4193	-3.766	0.000166	***
S7	171.2696	41.4625	4.131	$3.62*10^{-5}$	***
S9	-218.138	47.9254	-4.552	$5.32*10^{-6}$	***
S10	128.9505	28.6048	4.508	$6.54*10^{-6}$	***
K9	0.5175	0.3489	1.483	0.138005	
K10	1.5636	0.3186	4.908	$9.20*10^{-7}$	***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
(Dispersion parameter for binomial family taken to be 1)					
Null deviance: 1859.7 on 12844 degrees of freedom					
Residual deviance: 1585.2 on 12835 degrees of freedom					
AIC: 1605.2					

منبع: داده‌های پژوهش

جدول ۷: ضرایب متغیرهای مدل بقای رگرسیون لجستیک مبتنی بر اسپلاین

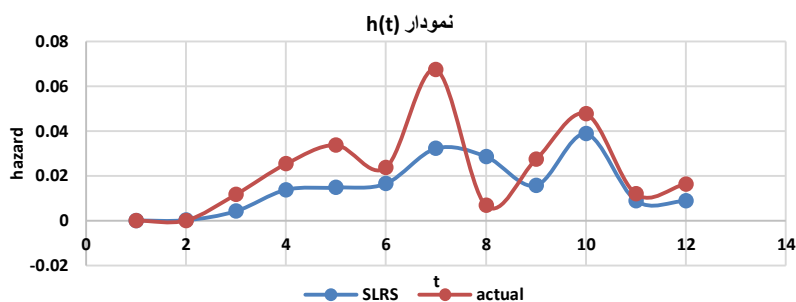
	Estimate	Std. Error	z value	Pr(> z )	AIC	Level of deleted variable	
(Intercept)	-5.8974	0.49535	-11.906	$< 2*10^{-16}$			***
A	-6.28768	1.60421	-3.919	$8.87*10^{-5}$			***
B	0.04637	18.51856	0.003	0.998002	1615.97	1	
C	2.46638	0.66043	3.735	0.000188			***
D	0.56167	0.87058	0.645	0.518823	1607.13	6	
E	0.32834	0.60614	0.542	0.588036	1608.53	5	
F	-0.31887	0.27809	-1.147	0.25153	1606.27	7	
G	-0.66026	0.43989	-1.501	0.133359	1605.18	9	
H	-0.35345	0.29876	-1.183	0.236794	1605.57	8	
I	-0.5662	1.23261	-0.459	0.645979	1610.27	4	
J	3.82303	0.4866	7.857	$3.95*10^{-15}$			***
S6	-80.2675	88.3089	-0.909	0.363382			
S7	141.5307	215.6958	0.656	0.511722			
S9	-146.651	462.852	-0.317	0.751364			
S10	61.8878	409.114	0.151	0.879761			
S12	14.06326	77.19983	0.182	0.855452	1612.05	3	
K7	0.06187	0.8366	0.074	0.941046	1613.98	2	
K9	0.36272	0.99231	0.366	0.714718			
K10	1.52004	0.51178	2.97	0.002977			**

منبع: داده‌های پژوهش



شکل ۲: نمودار احتمال نکول در زمان t (ماه) برای داده‌های آزمایش

منبع: داده‌های پژوهش



شکل ۳: نمودار احتمال نکول در زمان  $t$  (ماه) مدل بقای رگرسیون لجستیک مبتنی بر اسپلاین  
منبع: داده‌های پژوهش

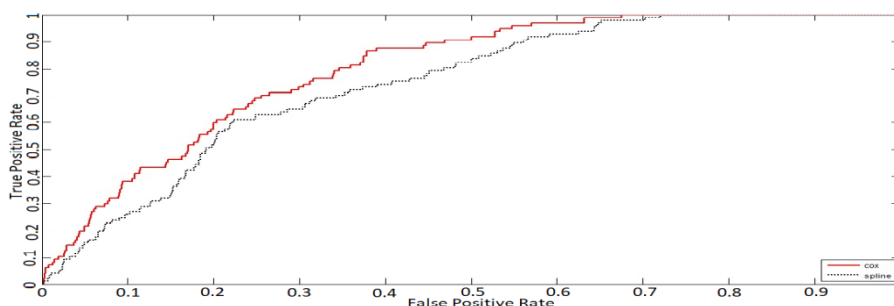
### ۳-۳- مقایسه دو مدل با استفاده از روش ROC

پس از مدل‌سازی احتمال نکول با استفاده از دو مدل رگرسیون کاکس و مدل بقای رگرسیون لجستیک مبتنی بر اسپلاین اکنون به مقایسه این دو مدل با استفاده از روش ROC می‌پردازیم.

در مورد صحت پیش‌بینی مدل‌های مطرح شده در این پژوهش به یک معیار پیش‌بینی نیاز داریم. پس از بررسی‌های بسیار بهترین معیار را روش ROC مشاهده کردیم. این ابزار در متون مالی بسیار کاربرد است، چنانکه افرادی هم چون توماس (۲۰۰۰)، اشتاین (۲۰۰۵)، بلوچلینگر و همکاران (۲۰۰۶)، کائو و همکاران (۲۰۰۹) و دیگران از آن، در این زمینه استفاده کرده‌اند.

پس از آزمایش مدل‌ها با استفاده از روش ROC برای مدل رگرسیون کاکس معیار  $AUC=0.799$  به دست آمد که نسبت به مدل بقای رگرسیون مبتنی بر اسپلاین از کارایی بالاتری برخوردار است. همچنین نمودار ROC این دو مدل را می‌توان در شکل ۴ مشاهده کرد.

همانطور که در شکل ۴ مشاهده می‌شود مدل کاکس از کارایی بالاتری نسبت به مدل پیشنهادی برخوردار است.



شکل ۴: مقایسه نمودار ROC دو مدل رگرسیون کاکس و مدل بقای رگرسیون لجستیک مبتنی بر اسپلاین  
منبع: داده‌های پژوهش



## ۴- نتیجه‌گیری

در این پژوهش برای به دست آوردن احتمال نکول تسهیلات اعطایی به مشتریان یک بانک ایرانی که شامل اطلاعات ۲۸۶۱ مشتری با ۱۰ متغیر است از مدل رگرسیون کاکس و مدل بقای رگرسیون لجستیک مبتنی بر اسپلاین استفاده شده است. اندازه‌گیری احتمال نکول مشتریان موجب بهبود پروسه وام‌دهی بانک‌ها و در نتیجه بهبود کیفیت دارایی بانک‌ها خواهد شد.

نتایج این پژوهش همانطور که در جدول ۷ مشخص است پس از حذف متغیرهای اضافی با استفاده از روش گام به گام (بازگشت به عقب) از میان ۱۰ متغیر اصلی متغیرهای سابقه فعالیت مشتری، نوع وام (۳ماهه،...) و امتیاز هر مشتری به ترتیب با ضرایب ۵.201-، ۱.۶۵۶، ۳.۶۹۵ در مدل رگرسیون کاکس و ضرایب ۱.719، ۳.۹۱ و ۵.350 در مدل بقای رگرسیون لجستیک مبتنی بر اسپلاین بیشترین تاثیر را در احتمال نکول دارد. همانطور که در جدول ۸ مشاهده می‌شود تفاوت معناداری بین مقادیر دو تابع مشاهده نمی‌شود. از جمله دیگر نتایج به دست آمده می‌توان به حذف متغیر نوع وثیقه و وضعیت محل مشتری از لحاظ ملکی یا استیجاری داشت که تاثیر کمی بر احتمال نکول دارند.

جدول ۸: نتایج مدل‌سازی با استفاده از مدل رگرسیون کاکس و مدل بقای رگرسیون لجستیک مبتنی بر اسپلاین

COX Regression		SLRS	
	Estimate		Coef
(Intercept)	-6.324	A	-5.201
A	-5.350	C	1.656
C	1.7196	J	3.695
J	3.9129		
S6	-91.9583		
S7	171.2696		
S9	-218.138		
S10	128.950		
K9	0.517		
K10	1.563		
AUC = 0.799		AUC = 0.746	

منبع: داده‌های پژوهش

پس از به دست آوردن میزان تاثیرگذاری هر متغیر بر احتمال نکول همچنین بعد از مقایسه دو مدل با استفاده از روش ROC چنین استنباط شد که مدل رگرسیون کاکس با مقدار  $AUC=0.799$  نسبت به مدل بقای رگرسیون لجستیک مبتنی بر اسپلاین با مقدار  $AUC=0.746$  عملکرد بهتری نسبت به روش پیشنهادی برخوردار است. این نتیجه نتایج مقاله کرانی را مبنی بر کارا بودن روش رگرسیون کاکس تایید می‌کند ولی مخالف با نتیجه سیرانگ لو و همکاران (۲۰۱۶) در کارایی بهتر مدل بقای رگرسیون لجستیک مبتنی بر اسپلاین نسبت به مدل رگرسیون کاکس است.

### فهرست منابع

- \* حامد کرانی و مولود آقاییپور. (۱۳۹۰). کاربرد نظریه تحلیل بقا در مدیریت ریسک اعتباری دریافت کنندگان تسهیلات؛ مطالعه موردی بانک مسکن. فصلنامه روند، شماره ۶۵، ۱۷۵-۲۰۰.
- \* مرضیه ابراهیمی و عبدالله دریاپر. (۱۳۹۱). مدیریت ریسک اعتباری در نظام بانکی - رویکرد تحلیل پوششی داده ها و رگرسیون لجستیک و شبکه عصبی، فصلنامه دانش سرمایه‌گذاری سال اول / شماره دوم / تابستان ۱۳۹۱.
- \* Narain, B. (1992). Survival analysis and the credit granting decision. In: Thomas L., Crook, J. N. and Edelman, D. B. (eds.). Credit Scoring and Credit Control. OUP: Oxford, 109-121.
- \* Cao, R., Vilar, J.M., Devia, A., Veraverbeke, N., Boucher, J.P. and Beran, J., 2009. Modelling consumer credit risk via survival analysis.
- \* Miller, S., 2010. Risk Factors for Consumer Loan Default: A Censored Quantile Regression Analysis.
- \* Stepanova, M. and Thomas, L., 2002. Survival analysis methods for personal loan data. Operations Research, 50(2), pp.277-289.
- \* Luo, S., Kong, X. and Nie, T., 2016. Spline based survival model for credit risk modeling. European Journal of Operational Research, 253(3), 869-879.
- \* Dietz, K., Gail, M., Krickeberg, K., Samet, J. and Tsiatis, A., 2002. Statistics for Biology and Health. Survival Analysis, Edition Springer.
- \* Malik, M. and Thomas L. (2006). Modeling Credit Risk of Portfolio of Consumer Loans. University of Southampton, School of Management
- \* Stein, R. (2005). The Relationship Between Default Prediction and Lending Profits: Integrating the ROC Analysis and Loan Pricing. Journal of Banking and Finance. Vol. 29, pp 1213-1236.
- \* Blöchliger, A. and Leippold, M. (2006). Economic Benefit of Powerful Credit Scoring. Journal of Banking and Finance. Vol. 30, pp 851-873

### یادداشت‌ها

- <sup>1</sup> Narain  
<sup>2</sup> Conventional life  
<sup>3</sup> Sarah Miller  
<sup>4</sup> Cox-Snell residuals, Martingale, Deviance  
<sup>5</sup> Modeling Credit Risk of Portfolio of Consumer Loans  
<sup>6</sup> Binary  
<sup>7</sup> Logit Model  
<sup>8</sup> Krzanowski (2009)  
<sup>9</sup> Threshold  
<sup>10</sup> Area Under Curve (AUC)