دوفصلنامه اقتصاد پولی، مالی(دانش وتوسعه سابق) دوره جدید، سال بیست وچهارم، شمارهٔ۱۳، بهار و تابستان ۱۳۹۶

مطالعهٔ عوامل اثرگذار بر ریسک اعتباری مشتریان بانکی با استفاده از مدل‌های ناپارامتریک و شبه پارامتریک تحلیل بقا

ناهید رجب زاده مغانی[۱]
دانشجوی دکتری علوم اقتصادی دانشگاه فردوسی مشهد
محمدرضا لطفعلی پور[۲]
استاد گروه اقتصاد دانشگاه فردوسی مشهد
احمد سیفی[۳]
استادیار گروه اقتصاد دانشگاه فردوسی مشهد
مصطفی رزمخواه[۴]
دانشیار گروه اقتصاد دانشگاه فردوسی مشهد

تاریخ دریافت: ۱۳۹۴/۹/۲۸ تاریخ پذیرش:۱۳۹۵/۳/۵

چکیده

ریسک اعتباری یکی از ریسک‌های بسیار مهم در صنعت بانکداری است؛ بنابراین شناسایی عوامل اثرگذار بر آن همواره مورد توجه بانک‌ها بوده است. این مطالعه با هدف شناسایی عوامل اثرگذار بر ریسک اعتباری با رویکرد زمان تا وقوع قصور مشتریان انجام گرفته است. بدین‌منظور از یک نمونه تصادفی شامل ۵۳۱۶ نفر از مشتریان که در بازهٔ زمانی ۱۳۸۸-۱۳۹۳ از بانک مسکن وام گرفته‌اند استفاده شده است. مطالعهٔ

---

۱ - rajabzadehnahid@gmail.com

۲ - نویسنده مسئول: Lotfalipour@um.ac.ir

۳ -spring05@ferdowsi.um.ac.ir

4 - razmkhah_m@um.ac.ir

DOI: 10.22067/pm.v24i13.52294

حاضر با استفاده از مدل‌های مرسوم تحلیل بقا شامل مدل ناپارامتریک کاپلان-میر و مدل شبه پارامتریک خطرات متناسب کاکس به شناسایی عوامل اثرگذار بر ریسک قصور مشتریان پرداخته است. نتایج مدل ناپارامتریک پس از طبقه‌بندی متغیرها با توجه به میزان فراوانی و روش استپانوا و توماس (۲۰۰۲) نشان داد که متغیرهایی همچون مبلغ وام، تعداد اقساط، تعداد فرزند، تحصیلات، سن، نوع شغل و عنوان شغلی بر منحنی‌های تابع احتمال بقا و تابع نرخ خطر تأثیرگذارند. همچنین ورود همزمان متغیرها در رگرسیون کاکس حاکی از وجود اثر معنی‌دار متغیرهای مبلغ وام، تعداد اقساط، تحصیلات، سن، نوع شغل و وضعیت تأهل بر ریسک قصور مشتریان می‌باشد. آزمون هرل و لی نیز حاکی از برقراری فرض خطرات متناسب برای تمامی متغیرها بوده و لذا تفسیر نتایج حاصل از مدل کاکس قابل اطمینان است. علاوه بر این، ابزار تشخیص مانند باقیماندهٔ کاکس-اسنل مناسب بودن مدل را تأیید نمود. لذا بانک‌ها در تصمیم‌گیری اعطای وام به مشتریان باید به این متغیرها توجه لازم را داشته باشند.

کلیدواژه‌ها: ریسک اعتباری، زمان تا وقوع قصور، تحلیل بقا، مدل کاکس، مدل کاپلان-میر.
طبقه بندی JEL: E51 و G2

مقدمه

نظام بانکی یکی از ارکان اصلی هر نظام اقتصادی است. بانک‌ها و مؤسسات مالی و اعتباری نقش به سزایی در اجرای سیاست‌های پولی ایفا می‌کنند؛ بنابراین عملکرد صحیح و اصولی آن‌ها می‌تواند به رشد و شکوفایی اقتصادی جامعه کمک قابل توجهی نماید.

بسیس (Bessis, 2002) ریسک‌های اصلی بانک را به ریسک اعتباری، نقدینگی، نرخ بهره و عملیاتی طبقه‌بندی می‌کند. طبق نظر سینکی (Sinkey, 1992) نیز ریسک‌های اصلی بانک به سه دستهٔ ریسک‌های اعتباری، نرخ بهره و نقدینگی تقسیم می‌شوند. اما از آنجایی‌که اعطای تسهیلات یکی از اصلی‌ترین موارد مصارف بانکی است، از بین ریسک‌هایی که بانک‌ها با آن مواجه هستند، ریسک اعتباری که همان احتمال عدم بازپرداخت وام‌های مشتریان است، از مسائل بسیار حیاتی در صنعت بانکداری به شمار می‌رود (Arabmazar & Ruintan, 2006). این ریسک زمانی اتفاق می‌افتد که دریافت‌کنندگان اعتبارات تمایل یا توانایی بازپرداخت بدهی خود را نداشته باشند که از آن به عنوان ریسک قصور یاد می‌شود و مترادف با ریسک اعتباری است (Asli, 2011).

یکی از مهم‌ترین چالش‌هایی که نظام بانکی کشورهای مختلف از جمله ایران با آن مواجه

است، افزایش مطالبات سررسید گذشته و وصول نشده است. با توجه به نقش محوری بانک‌ها در اقتصاد ایران، این حجم گستردهٔ مطالبات معوق می‌تواند آسیب‌های جدی به اقتصاد کشور وارد کند. اهمیت این حجم بالای مطالبات معوق تا آنجا است که در برنامهٔ چهارم و پنجم توسعه به تمهیداتی به منظور جلوگیری از افزایش و در صورت امکان کاهش مطالبات معوق با تأکید بر نقش بانک مرکزی اشاره شده است.

یکی از دلایل مهم پیدایش ریسک اعتباری، وجود اطلاعات نامتقارن بین بانک و مشتری است. از آنجا که در بازار وام واعتبارات اطلاعات مشتری از وضعیت خود دقیق‌تر و کامل‌تر از اطلاعات بانک است، نامتقارن بودن اطلاعات، نهایتاً به زیان بانک تمام می‌شود. در بازار وام، مسأله اطلاعات نامتقارن ممکن است پیش یا پس از بسته شدن قرارداد رخ دهد. در این زمینه دو نوع اطلاعات نامتقارن وجود دارد. نوع اول، مسأله انتخاب بد[۱] می‌باشد که پیش از بسته شدن قرارداد اتفاق می افتد و به موقعیتی اشاره دارد که بانک به دلیل عدم شناخت کافی نسبت به مشتریان، با کسانی قرارداد می‌بندد که بدحساب بوده و از عهدهٔ بازپرداخت اصل و فرع وام برنمی‌آیند. نوع دوم اطلاعات نامتقارن، مسأله مخاطرات اخلاقی[۲] است که پس از بسته شدن قرارداد اتفاق می‌افتد. در این وضعیت مشتری تسهیلات را در پروژهٔ تحت مفاد قرارداد مصرف نمی‌کند و در پروژهٔ ریسکی‌تر به کار می‌گیرد.

بانک‌ها همواره تلاش می‌کنند با استفاده از ابزارهای موجود، چاره‌ای برای این عدم تقارن اطلاعات و در پی آن کاهش ریسک اعتباری بیابند. از ابزارهای مورد نظر بانک می‌توان به تعیین نرخ بهره بهینه جهت قیمت‌گذاری دقیق وام‌ها، تعیین حجم مناسب وثیقه و جیره‌بندی اعتبارات اشاره نمود.اما نکته‌ای که وجود دارد این است که بانک برای اینکه بتواند به طور صحیح و اصولی از این ابزارها در جهت کاهش عدم تقارن اطلاعات استفاده نماید، باید تا حد امکان شناخت مناسبی از مشتریان خود به دست آورد. یکی از راه‌های کسب شناخت نسبت به مشتریان، استفاده از روش‌های کمی مانند اعتبارسنجی است.

۱- Adverse selection

۲- Moral hazard

در این مطالعه، با استفاده از مدل‌های ناپارامتریک و شبه پارامتریک تحلیل بقا[1] به بررسی این موضوع پرداخته می‌شود که کدامیک از پارامترهای فردی و مربوط به وام می‌توانند بر زمان تا وقوع قصور (بقای مشتریان) تأثیرگذار باشند. لازم به ذکر است از زمانیکه، در صنعت بانکداری به اهمیت شناسایی مشتریان و پیش‌بینی عملکرد آن‌ها، پی برده شد، پژوهش‌های کاربردی متعددی برای برآورد احتمال قصور مشتریان انجام شده است که با استفاده از روش‌های رایج همچون رگرسیون لجستیک، شبکه‌های عصبی، درخت تصمیم‌گیری و ... احتمال خوش حساب یا بد حساب بودن مشتری را برآورد می‌نمایند. این مطالعات به زمان وقوع قصور و در نتیجه بقای مشتریان توجهی نداشته‌اند. شاید بتوان عنوان نمود مطالعاتی که با رویکرد جدید زمان تا وقوع قصور مشتریان به مدلسازی پرداخته باشند بسیار محدود می‌باشند و در داخل کشور نیز این مدل-سازی‌ها تاکنون مورد توجه قرار نگرفته است. این مطالعه با رویکردی متفاوت به تحلیل و بررسی عوامل مؤثر بر زمان تا وقوع قصور مشتریان می‌پردازد.

این تحقیق در پنج بخش سازمان یافته است. مقدمه حاضر بخش اول مقاله را تشکیل می‌دهد. بخش دوم به بیان ادبیات موضوع، پیشینه تحقیق و مطالعات مرتبط با رویکرد زمان تا وقوع قصور می‌پردازد. در بخش سوم روش تحقیق و معرفی مدل‌های مورد استفاده در روش تحلیل بقا که شامل مدل ناپارامتریک و شبه پارامتریک است، تبیین می‌شود. بخش چهارم برآورد مدل‌ها و تفسیر نتایج با استفاده از مدل‌های رایج ارائه و در بخش پنجم و پایانی جمع‌بندی بیان می‌شود.

ادبیات موضوع و پیشینه تحقیق

تاریخچهٔ اعتبار بسیار طولانی بوده و به ۲۰۰۰ سال قبل از میلاد مسیح برمی‌گردد. شاید بتوان گفت پدیده قرضدادن و قرض گرفتن دارای قدمتی به اندازه تجارت و بازگانی است (Thomas et al., 2002). علی‌رغم تاریخچهٔ طولانی اعتبار که به ۲۰۰۰ سال قبل از میلاد مسیح برمی‌گردد، تاریخچهٔ اعتبارسنجی بسیار کوتاه است و تنها به ۶ دهه پیش می‌رسد (Abdou & Pointon, 2011). آغازگر اعتبارسنجی، توسعه کارت‌های اعتباری بود. با ظهور کارت‌های اعتباری، بانک‌ها

۱- Survival analysis

و مؤسسات مالی و اعتباری به اهمیت اعتبارسنجی پی بردند و به مرور و با رشد روزافزون متقاضیان کارت‌های اعتباری، استفاده از روش‌های پیشرفته اعتبارسنجی ضروری به نظر می‌رسید؛ چراکه طبق گزارش‌های میرز و فورگی (Myers & Forgy, 1963)، با استفاده از این روش‌ها نرخ‌های قصور تا ۵۰ درصد و حتی بیشتر کاهش یافت.

به طور کلی اعتبارسنجی را می‌توان مجموعه‌ای از مدل‌های تصمیم‌گیری و تکنیک‌های مربوطه تعریف نمود که به قرض‌دهندگان در تصمیم‌گیری در مورد اعطای اعتبار به مشتریان کمک می‌کند. این تکنیک‌ها در تصمیم‌گیری در مورد اینکه به چه کسی اعتبار تعلق گیرد، چه میزان اعتبار باید داده شود و چه استراتژی‌های عملیاتی باید به کار گرفته شود تا سود قرض-دهندگان را افزایش دهد، کمک می‌کند.

یکی از شاخه‌های اعتبارسنجی، درخواست سنجی[1] است که تلاش می‌کند با استفاده از اطلاعات موجود در فرم تقاضای مشتریان پیش از اعطای وام، پر ریسک و یا کم ریسک بودن مشتری را مشخص نماید. تمامی مدل‌های اعتبارسنجی سنتی یک ایده اساسی دارند. اولین گام طبقه‌بندی نمونه‌ای از مشتریان گذشتهٔ بانک به دو طبقه خوش حساب و بد حساب براساس اطلاعات بازپرداختی آن‌هاست. گام بعدی برقراری ارتباط بین ویژگی‌های مشتریان و وضعیت قصور مشتری است. سپس با مدل‌سازی عملکرد وام‌های قبلی، عملکرد آتی وام‌هایی با ویژگی‌های مشابه را پیش‌بینی می‌نمایند. در این شرایط بانک‌ها می‌توانند احتمال قصور یک مشتری تازه وارد را براساس ویژگی‌های فردی‌اش برآورد نمایند. پس از درخواست سنجی، بانک می‌تواند به مشتریان کم ریسک وام اعطا کند و مشتریان پر ریسک را به کلی رد کند یا برای مشتریان پر ریسک متناسب با میزان ریسک پیش‌بینی شده، نرخ بهره را بالا ببرد و یا وثیقهٔ بیشتری از آن‌ها دریافت کند.

اما در دهه‌های اخیر بانک‌ها به این نتیجه رسیده‌اند که اعتبارسنجی مشتریان قبل از اعطای وام ضروری و لازم است اما کافی نیست. بدین منظور باید پس از اعطای وام، رفتار مشتریان و پویایی آن‌ها به دقت مورد بررسی قرار گیرد تا بتوان به طور دقیق‌تر احتمال قصور مشتریان در آینده را

۱- Application scoring

برآورد کرد. در واقع، در رفتارسنجی از اطلاعات بازپرداختی مشتریان برای پیش‌بینی احتمال قصور آن‌ها در دوره‌های زمانی آینده استفاده می‌شود. در این شرایط بانک بعد از درخواست-سنجی و تعیین اینکه به کدامیک از مشتریان باید وام اعطا شود، به وسیلهٔ رفتارسنجی[1] مشتریان خود قادر خواهد بود پیش از وقوع قصور، اقدامات پیشگیرانهٔ لازم را به عمل آورد.

اخیراً نیز انقلابی در کشورهای پیشرفته در حوزه اعتبارسنجی اتفاق افتاده است و آن این است که بانک‌ها به دنبال شناسایی سودآورترین مشتریان هستند. این هدف کسب سود از پیش‌بینی احتمال قصور برای آن‌ها مهم‌تر است. لذا بانک‌ها علاوه بر استفاده از روش‌های درخواست‌سنجی و رفتارسنجی، مایل به استفاده از روش‌های سنجش سود[2] هستند. براین اساس برای بانک، نه تنها این امر که آیا مشتری قصور می کند یا خیر اهمیت دارد، بلکه اطلاع از اینکه قصور چه زمانی اتفاق می‌افتد نیز از اهمیت ویژه‌ای برخوردار است. بنابراین یک سؤال مهم کلیدی و بسیار مشکل-تر برای قرض‌دهندگان این است که چه زمانی مشتریان قصور می کنند. برای پاسخ به این سؤال، محقق به جای برآورد احتمال قصور در دوره زمانی مشخص، به دنبال برآورد توزیع احتمال طی زمان است. سیستم اعتبارسنجی که در پی پاسخ به این سؤال مهم است مزیت‌های قابل توجهی در برابر دیگر سیستم‌های اعتبارسنجی دارد. این مزیت‌ها به صورت زیر بیان می‌شوند:

۱- استفاده از این گونه مدل‌سازی‌ها بانک را قادر می‌سازد، در مدتی که مشتری با بانک در ارتباط است، سود خود را محاسبه نماید؛ چرا که میزان سودآوری یک مشتری برای بانک نه تنها به احتمال قصور مشتری، بلکه به این امر که قصور چه زمانی اتفاق می‌افتد، بستگی دارد (Baesens et al., 2005).

۲- بانک می‌تواند در برخورد با یک مشتری از این مدل‌سازی‌ها به نحوی استفاده کند که زمینهٔ سودآوری خود را فراهم کند. به عنوان مثال بانک با توجه به تحلیل بقای مشتریان خود، مدت بازپرداخت وام را به طور بهینه تعیین نماید.

۳- بانک با استفاده از این مدل‌سازی‌ها می‌تواند سطوح قصور مشتریان طی زمان را برآورد کند

۱- Behavioral scoring

۲- Profit scoring

(Baesens et al., 2004). این امر چشم‌انداز روشن‌تری از میزان منابع بانک در آینده ایجاد می‌کند. براین اساس بانک می‌تواند نرخ ذخایر اضافی و آزاد خود را به طور دقیق‌تر تعیین نماید.

با توجه به مزیت‌های مهم فوق‌الذکر در این مطالعه از روش تحلیل بقا برای مدل‌سازی زمان تا وقوع قصور استفاده می‌شود.تحلیل بقا نیز به خودی خود چندین مزیت نسبت به روش‌های مرسوم و سنتی اعتبارسنجی دارد که در ادامه تلاش می‌شود به آن‌ها پرداخته شود. در تحلیل بقا به جای روش‌های مرسوم طبقه‌بندی ایستا، علاوه بر اطلاعاتی که متقاضیان هنگام درخواست وام در اختیار بانک قرار می‌دهند، رفتار بازپرداختی آنان در زمان‌های مختلف، که کاملاً پویا است، نیز مشاهده و ثبت می‌شود و همین امر قوت تحلیل بقا را در مقابل روش‌های دیگر نشان می‌دهد؛ چرا که از اطلاعات کامل‌تری نسبت به روش‌های دیگر اعتبارسنجی استفاده می‌نماید. مدل‌های تحلیل بقا به محقق اجازه می‌دهد نه تنها احتمال قصور، بلکه زمان تا وقوع قصور را مدل‌سازی نماید. این روش اعتبارسنجی که در آن محقق به دنبال برآورد مدت زمانی است که طول می‌کشد تا مشتری قصور کند، دقیقاً همان هدفی است که مدل‌های تحلیل بقا برای آن طراحی شده است. لذا استفاده از این روش می‌تواند منجر به کارا شدن سیستم اعتبارسنجی بانک‌ها گردد. یکی دیگر از مزیت‌های روش تحلیل بقا مشارکت داده‌های سانسور شده در مدل می‌باشد. در مطالعهٔ حاضر سانسور به این معنی می‌باشد که زمان رخداد (قصور) قابل مشاهده نیست. در یک مجموعهٔ ایده‌آل از داده‌ها باید زمان‌های شروع و پایان تمامی تسهیلات گرفته شده در دسترس باشند. در این صورت زمان تا وقوع قصور (زمان بقا) مشتریان به طور کامل تعیین می‌شود. اما در داده‌های واقعی، برخی از مشتریان وجود دارند که زمان پایان پرداخت آن‌ها در دسترس نیست. به چنین داده‌هایی سانسور شده از راست می‌گویند (Man, 2014). اگر مشاهدات سانسور شده از مجموعه داده‌ها حذف شوند، تعداد مشاهدات به طور قابل توجهی کاهش می‌یابد.

تحلیل بقا یک تکنیک آماری است که در آن متغیر مورد علاقه، زمان وقوع یک رخداد است. تحلیل بقا در زمینه‌های مختلفی از جمله، پزشکی و مهندسی کاربرد دارد (Glasson, 2007). اخیراً این روش در صنعت بانکداری در کشورهای پیشرفته برای برآورد زمان تا وقوع قصور مورد استفاده قرار می‌گیرد. زمان رخداد در این گونه مطالعات زمان لازم تا شکست (مرگ) است. از آنجا که این روش در ابتدا برای مطالعات مرگ و میر به کار برده می‌شد و اصلاً بدین منظور

طراحی شده بود، به همین جهت نام تجزیه و تحلیل داده‌های بقا بر آن نهاده شده است.

تحلیل بقا در اعتبارسنجی توسط ناراین (Narain, 1992) معرفی شد و سپس توسط باناسیک و دیگران (Bnasik et al., 1999) گسترش یافت. ناراین با استفاده از مجموعهٔ داده‌ها برای ۱۲۴۲ وام‌گیرنده بین سال‌های ۱۹۸۶-۱۹۸۸ دریافت که این روش تعداد قصور را در هر فاصله زمانی به خوبی برآورد می‌کند و نشان داد که تصمیمات اعطای اعتبار با استفاده از روش‌های تحلیل بقا در مقایسه با رگرسیون‌های دیگر، بهبود می‌یابد (Thomas, 2009). باناسیک و دیگران با استفاده از داده‌های ۵۰۰۰۰ وام شخصی طی سال‌های ۱۹۹۴-۱۹۹۷، در یک مؤسسه مالی مهم در انگلستان، روش‌های ناپارامتریک و پارامتریک تحلیل بقا را با رگرسیون لجستیک مقایسه نمودند. آن‌ها دریافتند که روش تحلیل بقا قابل رقابت با رگرسیون لجستیک بوده و در برخی مواقع، نسبت به آن برتری دارد.بعد از آن‌ها، مطالعات دیگری در این حوزه انجام گرفتند که به عنوان نمونه به چند مورد از آن‌ها اشاره می‌شود.

مونتورد (Monteverde, 2000)، در مقاله‌ای با عنوان «مدیریت ریسک قصور وام‌های دانشجویان» به تحلیل بقای وام‌های تحصیلی با استفاده از رگرسیون خطرات متناسب کاکس[1] می‌پردازد. او با استفاده از وام‌های دانشجویان حقوق طی دهه ۱۹۹۰، نشان می‌دهد که متغیرهایی مانند ویژگی‌های فردی دانشجویان، شهرت دانشگاه، منطقه جغرافیایی دانشگاه و نرخ بیکاری در حوزه وکالت، بر ریسک قصور اثر معنی‌داری دارند.

استپانوا و توماس (Stepanova & Thomas, 2002)، در مقاله‌ای با عنوان «مدل‌های تحلیل بقا برای داده‌های وام شخصی» به ارائه مدل اعتبارسنجی می‌پردازد که علاوه بر واقعه قصور، واقعه پرداخت پیش از موعد را نیز در نظر می‌گیرند. آن‌ها با استفاده از اطلاعات ویژگی‌های فردی ۵۰۰۰۰ مشتری در یک مؤسسه مالی مهم در انگلستان و اطلاعات بازپرداختی آن‌ها طی یک دورهٔ ۳۶ ماهه، مدل خطرات متناسب کاکس را به کار می‌گیرند. در مطالعهٔ آن‌ها، ویژگی‌های فردی مشتریان با استفاده از یک روش جدید طبقه‌بندی می‌شود و علاوه براین، چندین روش تشخیص برای بررسی دقت مدل مورد استفاده قرار می‌گیرد. همچنین آن‌ها برای افزایش قدرت پیش‌بینی،

۱- Cox Proportional Hazard Regression

ورود متغیرهای فردی وابسته به زمان را به مدل پیشنهاد می‌دهند.

کائو و دیگران (Cao et al., 2009)، در مطالعه‌ای با عنوان «مدل‌سازی ریسک اعتباری با استفاده از تحلیل بقا» احتمال قصور مشتریان را به روشی متفاوت از روش‌های رایج اعتبارسنجی، برآورد می‌نمایند. آن‌ها اطلاعات وام ۲۵۰۰۰ مشتریکه از بانک اسپانیا از جولای ۲۰۰۴ تا نوامبر ۲۰۰۶ درخواست وام نموده‌اند، را برای مدل‌سازی به کار می‌گیرند. در مطالعهٔ آن‌ها، با استفاده از ۳ روش تحلیل بقا شامل مدل رگرسیون کاکس، مدل‌های خطی تعمیم‌یافته[۱] و مدل ناپارامتریک کرنل[۲]، تابع توزیع شرطی زمان وقوع قصور برآورد می‌شود. نتایج نشان می‌دهد که از بین روش‌های مورد استفاده، مدل ناپارامتریک مدل سازگاری می‌باشد.

من (Man, 2014) در رساله دکتری خود با عنوان «تحلیل بقا در اعتبارسنجی: چارچوبی برای برآورد احتمال قصور» به تخمین تابع بقای مشتریان طی زمان می‌پردازد. در مطالعهٔ فوق کارایی تحلیل بقا با رگرسیون لجستیک مقایسه می‌شود. نتایج نشان می‌دهد که از نظر شاخص‌های آماری، استفاده از روش تحلیل بقا مناسب‌تر است. زیرا استفاده از تحلیل بقا مزیت‌های قابل توجهی نسبت به رگرسیون لجستیک دارد. او بیان می‌کند که یکی از مزیت‌های روش تحلیل بقا این است که مشاهدات کمتری حذف می‌شوند و برای برآورد احتمال بقا، از کل مجموعه داده‌ها استفاده می‌شود. حال آنکه در روش رگرسیون لجستیک احتمال بقا برای فواصل زمانی ثابت برآورد می‌شود.

از دیگر مطالعات صورت گرفته در زمینهٔ تحلیل بقا می‌توان به پژوهش‌های کارلینگ و دیگران (Carling et al., 1998)، رزباچ (Roszbach, 2004)، گلنان و نیجرو (Glennon & Nigro, 2005)، آلن و رز (Allen & Rose, 2006)، بابا و گکو (Baba & Goko, 2006)، سارلیجا و دیگران (Sarlija et al., 2006)، مالیکو توماس (Malik & Thomas, 2006)، بران و جاجا (Beran & Djaidja, 2007) و تانگ و دیگران (Tong et al., 2012) اشاره نمود.

روش تحقیق

۱- Generalized linear models

۲- Nonparametric kernel estimation

تحلیل بقا، زمان تا وقوع رخداد را مورد بررسی قرار می‌دهد (Kalbfleisch & Prentice, 1980). در ابتدا کاپلان و میر (Kaplan- Meier, 1958)، برآوردگری را ایجاد نمودند که به دنبال برآورد تابع بقا[۱] برای داده‌های سانسور شده و سانسور نشده بود. سپس کاکس (Cox, 1972) توابع خطر متناسب[۲] را مطرح نمود تا بتوان از طریق آن، بین ویژگی‌های فردی افراد و زمان تا وقوع رخداد مورد علاقه (در اینجا قصور) ارتباط برقرار نماید.

در این مطالعه، متغیر تصادفی T زمان تا وقوع قصور یا همان زمان بقا[۳] برای یک مشتری می‌باشد. به عبارت دیگر، این متغیر طول مدت زمان تا رسیدن به وضعیت قصور برای هر مشتری را نشان می‌دهد. معمولاً در داده‌های مربوط به وام و اعتبارات، اطلاعات بازپرداختی مشتریان به صورت ماهانه نشان داده می‌شوند. براساس تعاریف رایج در صنعت بانکداری، اگر مشتری۳ ماه متوالی و یا بیشتر تأخیر در بازپرداخت اقساط داشته باشد، آنگاه این مشتری به عنوان فرد قصور کرده، تلقی می‌شود. این تعریف نیز توسط کمیته بازل (Basel) مورد تأیید قرار گرفته است (Belloti & Crook, 2013).

اگر زمان تا وقوع قصور مشتریان با متغیر تصادفی Tنشان داده شود، حال احتمال قصور مشتری قبل از زمان t ($P_B(t,X)$) به صورت زیر است:

$$P_B(t,X) = Pr\{T < t\} \qquad (۱)$$

لازم به ذکر است که احتمال فوق به برخی ویژگی‌های فردی مشتریان که با X نشان داده می‌شوند نیز می‌تواند بستگی داشته باشد. می‌توان $P_B(t,X)$ تابع توزیع احتمال متغیر تصادفی T و $P'_B(t,X)$ تابع چگالی متناظر با آن نیز تعریف نمود.

براین اساس احتمال قصور در فاصله $[t, t+\delta t)$ به صورت زیر خواهد بود:

$$Pr\{t < T \le t+\delta t, X\} = P'_B(t,X)\delta t \qquad (۲)$$

تابع احتمال بقا که در آن قصور بعد از زمان t اتفاق می‌افتد ($P_G(t,X)$)، به صورت رابطهٔ ۳

۱- Survival Function

۲- Proportional Hazard Functions

۳- Survival time

نشان داده شده است.

$$P_G(t,X) = 1 - P_B(t,X) = Pr\{T \geq t\} = \int_t^{\infty} P'_B(u,X)du \quad (۳)$$

تابع نرخ خطر یکی دیگر از مفاهیم پایه‌ای تحلیل بقا است و این را نشان می‌دهد که اگر مشتری در فاصلهٔ $(0,t]$ قصور نکرده باشد، نرخ قصور در فاصلهٔ $(t,t+\delta t]$ چه قدر است. این تابع احتمال شرطی به صورت زیر تعریف می‌شود.

$$h(t,X) = \Pr\{t < T \leq t+\delta t : T > t; X\} \triangleq \frac{P'_B(t,X)}{1-P_B(t,X)} \quad (۴)$$

$$h(t,X) = -\frac{d}{dt}Ln\big(1-P_B(t,X)\big) = -\frac{d}{dt}Ln(P_G(t,X))$$

تابع نرخ خطر تجمعی نیز به صورت $H(t,X) \triangleq \int_0^t h(u,X)du$ تعریف می‌شود. بنابراین می‌توان با استفاده از رابطهٔ (۴)، تابع بقا را برحسب تابع نرخ خطر تجمعی به صورت زیر تعریف نمود.

$$P_G(t,X) = 1 - P_B(t,X) = e^{-H(t,X)} \quad (۵)$$

علاوه بر این می‌توان نوشت:

$$P'_B(t,X) = h(t,X)e^{-H(t,X)} \quad (۶)$$

در روش تحلیل بقا، مدل‌های مختلفی برای برآورد تابع بقا وجود دارند که در ادامه تلاش می‌شود رایج‌ترین مدل‌های ناپارامتریک و شبه پارامتریک تبیین شوند. از میان مدل‌های ناپارامتریک، مدل کاپلان–میر و از میان مدل‌های شبه پارامتریک مدل خطرات متناسب کاکس مورد استفاده قرار می‌گیرند.

مدل کاپلان–میر

مزیت استفاده از این روش این است که داده‌های سانسور شده را به حساب می‌آورد و به محقق اجازه می‌دهد که از تمامی اطلاعات موجود (سانسور شده و سانسور نشده) برای برآورد تابع بقا استفاده نماید.

در اینجا متغیر D_i متغیر سانسور در نظر گرفته می‌شود که اگر برابر با یک باشد، یعنی اطلاعات

مشتری کامل و زمان وقوع قصور مشخص است و اگر برابر با صفر باشد یعنی قصور هنوز اتفاق نیفتاده و بنابراین اطلاعات سانسور شده است. لذا برای هر مشتری t_i و D_i جداگانه‌ای وجود خواهد داشت. حال زمان بقای مشاهده شده برای مشتریان، به صورت صعودی از کوچک به بزرگ مرتب می‌شوند. در رابطهٔ زیر، در زمان t_j، تعداد d_j نفر قصور (با $D_i = 1$) نموده و تعداد c_j نفر سانسور (با $D_i = 0$) شده‌اند. حال تابع احتمال بقا به صورت زیر تعریف می‌شود:

$$\hat{P}_G(t) = \prod_{t_{(j)}<t} \frac{(n_j - d_j)}{(n_j)} \quad (۷)$$

در این رابطه n_j به صورت زیر قابل محاسبه است:

$$n_j = n - \sum_{j=0}^{j-1} (d_j + c_j) \quad (۸)$$

در برآورد گر کاپلان–میر، می‌توان جامعه مورد مطالعه را براساس معیارهایی مانند نرخ بهره، میزان وام و ... تقسیم نمود و با رسم برآوردگر هر گروه دریافت که احتمال بقای کدامیک بیشتر است. این‌گونه تحلیل‌ها، ابزار بسیار مفیدی برای تحلیل بقای مشتریان بانکی می‌باشد. تابع برآوردی احتمال بقا با استفاده از روش کاپلان–میر به صورت پله‌ای می‌باشد.

مزیت مدل ناپارامتری کاپلان–میر این است که محاسبه و تفسیر آن نسبتاً ساده می‌باشد. این روش، یک برآوردگر توضیحی بوده و می‌تواند تابع بقا یک جمعیت را برآورد کند. با تقسیم‌بندی جامعه براساس معیارهای مختلف می‌توان تحلیل‌های بقای متفاوتی برای هر گروه از مشتریان داشت که این امر ابزار مفیدی را در اختیار محقق قرار می‌دهد. اما یکی از معایب اصلی این برآوردگر این است که در این روش میزان اثرگذاری متغیرهای توضیحی بر زمان تا وقوع قصور مورد توجه قرار نمی‌گیرند.

مدل خطرات متناسب کاکس

گاهی اوقات محقق به دنبال این است که داده‌ها را با جزئیات بیشتری بررسی نماید. در تحلیل بقا می‌توان با استفاده از روش‌های شبه پارامتریک، اثر ویژگی‌های فردی مشتریان را بر احتمال اینکه چه زمانی مشتریان قصور می‌کنند، بررسی نمود. یکی از مدل‌های مشهور شبه پارامتریک،

مدل خطرات متناسب[۱] است که توسط کاکس (Cox, 1972) مطرح شد.

در این مدل فرض می‌شود که اگر برداری از متغیرهای توضیحی X وجود داشته باشد، یعنی $X = (X_1, X_2, ..., X_p)$ باشد، آنگاه نرخ خطر در زمان t به صورت زیر قابل نمایش است (Kleinbaum & Klein, 2005):

$$h(t, X) = e^{WX} h_0(t) \quad (۹)$$

این مدل نرخ مخاطره را برآورد می‌نماید و از دو قسمت تابع نرخ خطر پایه یا مبنا و یک ترکیب خطی از مجموعه p متغیر مستقل تشکیل شده است. در این فرمول، $h_0(t)$، تابع خطر مبنا نامیده می‌شود. اگر تمامی متغیرهای توضیحی X برابر با صفر در نظر گرفته شوند، آنگاه نرخ خطر برابر با تابع خطر مبنا خواهد شد. پارامترهای اثر نیز نشان می‌دهند که چگونه خطر در واکنش به متغیرهای توضیحی تغییر می‌کند.

اگر t_j، D_j و $X_j = (X_{j1}, \dots, X_{jP})$ به ترتیب زمان‌های شکست (یا سانسور)، متغیر سانسور (اگر برابر با یک باشد قصور اتفاق افتاده و اگر برابر با صفر باشد مشتری سانسور شده است) و ویژگی‌های مشتری j در نمونه باشد. حال زمان‌های بقا به صورت $t_1 \le t_2 \le \cdots \le t_n$ مرتب و مجموعه در خطر به صورت $R(i) = (i, i+1, i+2, \dots, n)$ تعریف می‌شود. مجموعه در خطر، مجموعه‌ای از مشتریان هستند که قبل از زمان t_i (زمانی که در آن قصور اتفاق افتاده و $D_i = 1$ است) هنوز در نمونه وجود دارند. حال می‌توان پارامتر W را با استفاده از روش حداکثر راست-نمایی جزئی برآورد نمود.

$$L(W) = \prod_{i:D_i=1} \frac{exp\{W.X_i\}}{\sum_{j \epsilon R\{i\}} exp\{W.X_j\}} \quad (۱۰)$$

در مدل کاکس هیچ فرض خاصی در مورد تابع مبنا وجود ندارد. همچنین این مدل کاملاً انعطاف‌پذیر بوده و در بسیاری از موارد انتخاب مناسبی می‌باشد (Man, 2014). یکی از دلایل متداول شدن این مدل با وجود نامشخص بودن تابع خطر مبنا، برآوردهای درست و منطقی برای ضرایب رگرسیونی، است. این مدل با استفاده از حداقل فرض‌ها می‌تواند منحنی‌های بقا و ضرایب

۱- Proportional hazard model

مدل را برآورد نماید. لذا در این مطالعه علاوه بر مدل ناپارامتریک کاپلان-میر از مدل کاکس نیز برای مدل‌سازی زمان تا وقوع قصور استفاده می‌شود.

برآورد مدل

پیش‌پردازش داده‌ها و اطلاعات مورد نیاز

در این مطالعه، اطلاعات مربوط به وام‌های مشتریان و ویژگی‌های فردی آن‌ها از سرپرستی بانک مسکن در استان خراسان رضوی اخذ شده است. بانک اطلاعات مربوط به ۵۳۱۶ نفر همراه با اطلاعات بازپرداختی آن‌ها از سال ۱۳۸۸ تا پایان سال ۱۳۹۳ را در اختیار محقق قرار داد. این اطلاعات مربوط به ۴۵ شعبه از بانک مسکن در سطح شهر مشهد می‌باشد. داده‌هایی که بانک مسکن حاضر به ارائهٔ آن‌ها گردید، شامل اطلاعات مربوط به ویژگی‌های فردی مشتریان، اطلاعات مربوط به وام و اطلاعات مربوط به تاریخ‌های بازپرداخت اقساط مشتریان می‌باشد. ویژگی‌های فردی مشتریان شامل سن، وضعیت تأهل، تعداد فرزند، تحصیلات (از بیسواد تا دکتری)، نوع شغل (حقیقی دولتی، حقیقی خصوصی، حقوقی دولتی و حقوقی خصوصی) و شرح شغل (بیکار، کارمند، کارگر، استاد دانشگاه و ...) می‌باشد. ویژگی‌های مربوط به وام شامل مبلغ تسهیلات، نرخ بهره، تعداد کل اقساط می‌باشد.

منطق محاسبه ماه‌های تأخیر براساس استانداردهای مؤسسه رتبه بندی aaarrrr d & rrrr 's می‌باشد.براساس تعاریف رایج در صنعت بانکداری، اگر مشتری ۳ ماه متوالی و یا بیشتر تأخیر در بازپرداخت اقساط داشته باشد، آنگاه این مشتری به عنوان فرد قصور کرده، تلقی می‌شود. این تعریف نیز توسط کمیته بازل مورد تأیید قرار گرفته است. حال مشتری که بیش از سه ماه تأخیر در بازپرداخت اقساط داشته باشد به عنوان فرد قصور نموده تلقی و مدت زمان تا وقوع قصور به عنوان طول بقای مشتری در نظر گرفته می‌شود. همچنین مشتریانی که به دلایلی مانند رسیدن به پایان مطالعه و تسویه حساب پیش از موعد، قصور برای آن‌ها اتفاق نیفتاده است، پس از محاسبه طول بقای آن‌ها به عنوان سانسور در نظر گرفته می‌شوند. باید عنوان نمود هم از نظر تئوری و استانداردهای رایج بانکداری و هم از نظر بکار بستن ابزار این تحقیق باید تعریف مشخصی از قصور را ارائه نمود. براین اساس حجم داده‌های سانسور کاهش یافته و درصد قابل ملاحظه‌ای از مشتریان به وضعیت قصور می‌رسند. این امر به دلایل آماری کاملاً منطقی به نظر می‌رسد. زیرا

حجم بالای داده‌های سانسور، کارایی مدل را کاهش می‌دهد. در این مطالعه، به دلیل حجم بالای داده‌ها و همچنین افزایش دقت پردازش آن‌ها، برای محاسبهٔ ماه‌های تأخیر هر مشتری از نرم‌افزار Matlab 2014 استفاده شده است.

تحلیل ویژگی‌های فردی و ویژگی‌های مربوط به وام در بقای مشتریان و ریسک قصور با استفاده از روش کاپلان-میر

یکی از اهداف این مطالعه برآورد تابع بقا و تابع نرخ خطر برای گروههای مختلف از نظر ویژگی‌های فردی و ویژگی‌های مربوط به وام می‌باشد. به عبارت دیگر هدف، برآورد ریسک قصور و زمان تا وقوع قصور برای هر یک از گروه‌های مورد نظر می‌باشد. در این پژوهش براساس داده‌های موجود، مشتریان از نظر ویژگی‌های مربوط به وام طبقه‌بندی و سپس به تحلیل بقای هر گروه از مشتریان پرداخته شود. در ادامه همین روند برای ویژگی‌های فردی صورت می‌گیرد. هدف این بخش شناسایی عواملی است که بر روی زمان تا وقوع قصور اثر می‌گذارند.

اما قبل از برآورد تابع بقا و تابع نرخ خطر برای گروه‌های مختلف باید دسته‌بندی‌ها به نحو صحیحی صورت پذیرد. یکی از موارد مهم برای چگونگی ورود متغیرها در مدل، طبقه‌بندی متغیرهای پیوسته مانند سن و متغیرهای گسسته مانند نوع شغل به گروه‌ها یا دسته‌های مختلف است. طبقه‌بندی متغیرهای پیوسته به دلیل عدم یکنواختی ریسک در این متغیرها صورت می‌گیرد (Thomas, 2009). به عنوان مثال بانک‌ها بسته به گروه‌های مختلف سنی می‌توانند استراتژی‌های مختلفی را اتخاذ نمایند و ریسک مشتریان خود را بهتر پیش‌بینی نمایند. دلیل گروه‌بندی متغیرهای گسسته همچون شغل نیز وجود پاسخ‌های گوناگون و زیاد می‌باشد. در این شرایط ممکن است نمونهٔ کافی برای هر پاسخ مشخص وجود نداشته باشد و لذا تحلیل اعتبارسنجی دچار ضعف شود. بر این اساس، طبقه‌بندی متغیرهای گسسته (که جواب‌های گسسته برای آن وجود دارد) و متغیرهای پیوسته (که مجموعه نامحدودی از جواب‌ها برای آن وجود دارد و الزاماً به صورت بازه-

ای صورت می‌گیرد) ضروری به نظر می‌رسد. لذا برای اینکه سیستم اعتبارسنجی قوی[۱] باشد، در ابتدا باید با تقسیم‌بندی هر متغیر به گروه‌های مختلف، آن متغیر را طبقه‌بندی نمود (Thomas et al., 2002).

همان‌طور که عنوان شد، به دلیل فراوانی کم برخی پاسخ‌ها برای متغیرهای فردی و متغیرهای مربوط به وام که بانک در مجموعه داده‌های خود در اختیار محقق قرار داد ابتدا لازم است متغیرهای مورد نظر (متغیرهای فردی و متغیرهای مربوط به وام) به گروه‌های بزرگتری تقسیم شوند. به عنوان مثال افراد بالای ۸۰ سال و یا کسانی که مدرک فوق دکترا دارند بسیار کم و در حد ۳ و ۴ نفر می‌باشد. لذا برای انجام اعتبارسنجی قوی، ترکیب این گروه‌ها در دسته‌های بزرگتر قدم اول است. پس از بررسی توزیع فراوانی و انجام دادن دسته‌بندی‌های اولیه، بایستی اطمینان حاصل نمود که این دسته‌بندی‌ها جهت برآورد تابع نرخ خطر و تابع بقا مناسب است.

در روش‌های سنتی پیدا نمودن دسته‌ها و یا گروه‌های مناسب براساس نسبت خوب-بد[۲] صورت می‌گیرد (Stepanova & Thomas, 2002). اما نکته‌ای که در اینجا باید به آن توجه نمود این است که در تمامی روش‌های سنتی طبقه‌بندی مانند روش نسبت خوب-بد، انتخاب یک افق زمانی بسیار کلیدی است. در این روش‌ها اگر قصور قبل از یک افق زمانی خاص اتفاق بیافتد (مثلاً ۱۲ ماه) مشتری بد و اگر قصور بعد از این افق زمانی اتفاق بیافتد و یا هرگز اتفاق نیافتد مشتری خوب تلقی می‌شود. از آنجا که روش تحلیل بقا زمان تا وقوع قصور را در نظر می‌گیرد و قائل به افق زمانی نیست، لذا در هیچ‌یک از محاسبات، نباید افق زمانی در نظر گرفته شود. برای انجام این کار استپانوا و توماس (Stepanova & Thomas, 2002) طبقه‌بندی به روش تحلیل بقا را پیشنهاد می‌دهند. مراحل انجام طبقه‌بندی متغیرها به صورت زیر است:

۱) ویژگی‌های فردی به دسته‌های مختلف تقسیم می‌شوند؛ به طوری‌که برای متغیرهای پیوسته تا ۲۰ دسته بازه‌ای و برای متغیرهای گسسته به تعداد پاسخ‌ها دسته یا گروه ایجاد می‌شود.

۲) برای هر دسته یک متغیرایجاد می‌شود.

۱- robust

۲- Good-bad ratio

۳) مدل خطرات متناسب کاکس با استفاده از این متغیرها برآورد می‌شود.
۴) گروه‌ها یا دسته‌های نهایی براساس مشابهت ضرایب برآورد شده انتخاب می‌شوند.

برای انجام مرحلهٔ ۱ ابتدا دسته‌بندی‌های اولیه براساس نمودار توزیع فراوانی صورت می‌گیرد. نمودار توزیع فراوانی مربوط به تمامی متغیرها در پیوست ۱ آورده شده است. در نهایت با استفاده از روش تحلیل بقا متغیرها به دسته‌های نهایی براساس درجه ریسک آن‌ها طبقه‌بندی می‌شوند.

براساس اطلاعات اتخاذ شده از بانک، برای هر مشتری ۳ متغیر مهم مبلغ کل وام، تعداد اقساط و نرخ بهره وجود دارند. در نمونهٔ مورد بررسی مبلغ کل وام در بازه‌ای بین ۵۰ میلیون تا ۲۷۲ میلیون ریال می‌باشد. متغیر نرخ بهره دو عدد ۱۱ درصد و ۱۳ درصد و متغیر تعداد اقساط سه عدد ۱۲۰، ۱۴۴ و ۱۸۰ قسط را به خود اختصاص داده‌اند. براساس نمودارفراوانی متغیرهای مربوط به وام، متغیرهای نرخ بهره و تعداد وام به دلیل اینکه نسبتاً حجم مناسبی از مشاهده در هر طبقه وجود دارد دسته بندی به همان صورت خام باقی ماند. اما برای متغیر مبلغ وام، دسته بندی اولیه به صورت دسته صفر از ۵۰ میلیون تا ۸۰ میلیون ریال، دسته یک از ۸۰ میلیون ریال تا ۱۱۰ میلیون ریال، دسته دو از ۱۱۰ میلیون ریال تا ۱۴۰ میلیون ریال و الی آخر صورت گرفت که در مجموع هشت دسته برای مبلغ وام اختصاص داده شد.

به دلیل پراکندگی و فراوانی کم داده‌ها دسته‌های صفر تا سه (۵۰ میلیون تا ۱۷۰ میلیون ریال) در یک طبقه، دسته ۴ (۱۷۰ تا ۲۰۰ میلیون ریال) در یک طبقه و دسته پنج تا هفت (۲۰۰ میلیون ریال به بالا) در یک طبقه قرار گرفتند. بنابراین براساس طبقه‌بندی جدید ۵۰ تا ۱۷۰ میلیون در دسته صفر، ۱۷۰ تا ۲۰۰ میلیون در دسته یک و ۲۰۰ میلیون به بالا در دسته دو قرار می‌گیرند. همچنین متغیر نرخ بهره به دو دسته صفر (۱۱ درصد) و دسته یک (۱۳ درصد) و متغیر تعداد اقساط به سه دسته صفر (۱۲۰ قسط)، دسته یک (۱۴۴ قسط) و دسته دو (۱۸۰ قسط) طبقه‌بندی می‌شوند.

حال با استفاده از روش استپانوا و توماس (Stepanova & Thomas, 2002) مراحل ۲ تا ۴ برای دسته‌بندی‌های نهایی انجام می‌گردد. برای ۳ متغیر نرخ بهره، تعداد اقساط و مبلغ کل وام، مدل کاکس به طور جداگانه بر روی دسته‌های مختلف هر متغیر برآورد می‌شود. نرم افزار Stata در برآورد ضرایب مربوط به هر طبقه، دستهٔ صفر را به عنوان سطح مبنا در نظر می‌گیرد و ضرایب مربوط به دسته‌های دیگر را نسبت به دستهٔ مبنا برآورد می‌نماید. براین اساس می‌توان نرخ خطر یا احتمال بقا دسته‌های دیگر را با دستهٔ مبنا مقایسه نمود. در برآورد مدل، ضریب دستهٔ مبنا صفر در

نظر گرفته می‌شود (در جدول نتایج برآورد حذف می‌شود) و ضریب برای دسته‌های دیگر نسبت به مبنا برآورد می‌شود. برآورد مدل کاکس برای متغیرهای مربوط به وام به صورت جدول ۱ می‌باشد.

نتایج حاصل از جدول ۱ نشان می‌دهد ضرایب مربوط به دسته اول و دوم مربوط به مبلغ وام نسبتاً مشابه بوده و لذا این دو دسته با هم ترکیب می‌شوند[۱]. لذا دسته بندی نهایی برای مبلغ وام به صورت دو طبقه می‌باشد. طبقه صفر از ۵۰ تا ۱۷۰ میلیون ریال و طبقه یک از ۱۷۰ میلیون ریال به بالا می‌باشد. ضریب متغیر نرخ بهره برای دسته یک اختلاف معنی‌داری با صفر ندارد و لذا تابع نرخ خطر و تابع بقا برای نرخ بهره ۱۱ درصد و نرخ بهره ۱۳ درصد متفاوت نیست. برای متغیر تعداد اقساط نیز ضرایب به طور معنی‌داری از هم متفاوت بوده و لذا همین سه دسته مد نظر قرار می‌گیرند.

جدول (۱): برآورد مدل کاکس برای دسته‌بندی نهایی ویژگی‌های مربوط به وام

متغیر	طبقه	ضریب برآوردی	سطح معنی داری
مبلغ وام	دسته اول (۱۷۰-۲۰۰ میلیون ریال)	-۰/۲۵	۰/۰۰
	دسته دوم (۲۰۰ میلیون به بالا)	-۰/۳۸	۰/۰۰
نرخ بهره	دسته اول (۱۳ درصد)	-۰/۰۷	۰/۱۴
تعداد اقساط	دسته اول (۱۴۴ قسط)	۰/۳۸	۰/۰۰
	دسته دوم (۱۸۰ قسط)	۰/۵۷	۰/۰۰

منبع: محاسبات تحقیق

حال به برآورد تابع بقا برای متغیرهای مربوط به وام در دسته‌های مختلف پرداخته می‌شود. این برآوردها با استفاده از روش ناپارامتریک کاپلان-میر صورت می گیرد. همچنین برای آزمون اینکه آیا منحنی‌های بقا به طور معنی‌داری با هم اختلاف دارند یا خیر از آزمون مشهور لگاریتمی رتبه‌ای[۲] استفاده می‌شود. فرضیه صفر این آزمون برابری توابع بقا و فرضیه مقابل عدم برابری توابع بقا برای گروه‌های مختلف می‌باشد.

۱- لازم به ذکر است نتایج حاصل از آزمون لگاریتمی رتبه‌ای نشان داد که ضرایب مربوطه اختلاف معنی‌داری با هم ندارند.

۲- Log-rank statistic

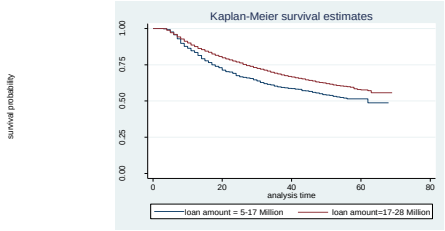
نتایج حاصل از جدول ۲ نشان می‌دهد که فرض صفر یعنی عدم اختلاف معنی‌دار بین توابع بقا برای دو متغیر مبلغ وام و تعداد اقساط را نمی‌توان پذیرفت. طبق نمودار ۱، مشتریانی که وام‌های بالای ۱۷ میلیون تومان می‌گیرند نسبت به مشتریانی که کمتر از ۱۷ میلیون تومان می‌گیرند بقای بیشتری دارند و بنابراین ریسک قصور کمتری دارند. اگرچه این نتیجه خلاف انتظار می‌باشد؛ چراکه انتظار می‌رود با افزایش مبلغ وام و به دنبال آن افزایش مبلغ اقساط بقای مشتریان کاهش یابد. اما با بررسی جزئی‌تر داده‌ها به نظر می‌رسد وام‌های با رقم بالا عمدتاً به کسانی داده می‌شود که توانایی مالی بهتری دارند و لذا رخداد قصور برای این دسته از افراد دیرتر اتفاق می‌افتد. ضمن اینکه بانک‌ها در اعطای وام‌های با رقم بالا به مشتریان محتاط‌تر می‌باشند. نمودار ۲ نشان می‌دهد با افزایش تعداد اقساط، احتمال بقا کاهش و ریسک قصور افزایش می‌یابد. به عبارتی دیگر هرچه طول دورهٔ بازپراخت افزایش یابد احتمال اینکه مشتریان سه ماه متوالی اقساط خود را پرداخت نکنند افزایش می‌یابد.

همان‌طور که نتایج حاصل از جدول ۲ نشان می‌دهد دسته‌بندی براساس نرخ بهره بی معنی بوده و تابع بقا برای دو گروه نرخ بهره ۱۱ درصد و نرخ بهره ۱۳ درصد اختلاف معنی‌داری با هم ندارند. این نتیجه در روش استپانوا و توماس (Stepanova & Thomas, 2002) نیز بدست آمد و آنجا نیز پیشنهاد شد دسته‌بندی براساس نرخ بهره صورت نگیرد؛ چراکه ضرایب مدل برای هر دو نرخ بهره مشابه بودند. این امر به دلیل اختلاف ناچیز نرخ بهره‌ها طبیعی به نظر می‌رسد. چنانچه بانک مورد نظر این مطالعه طیفی از نرخ بهره‌ها را برای مشتریان وضع می‌نمود، نتایج و تحلیل‌های مناسب‌تری می‌توانست بدست آید. برآورد توابع خطر با استفاده از روش کاپلان میر در پیوست ۲ آورده شده است.

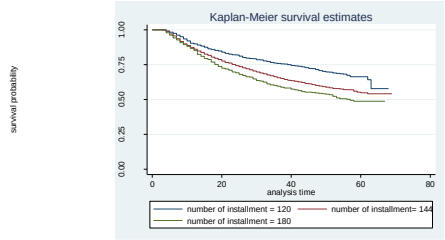
جدول (۲): آزمون لگاریتمی-رتبه‌ای برای متغیرهای مربوط به وام

متغیر	آماره کای دو	prob
مبلغ وام	۲۲/۷۳	۰/۰۰
تعداد اقساط	۵۴/۷۸	۰/۰۰
نرخ بهره	۲/۱۶	۰/۱۴

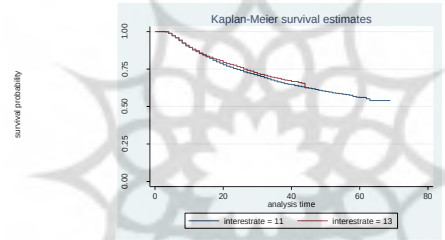
منبع: محاسبات تحقیق



نمودار (۱): تابع بقا برای متغیر مبلغ وام



نمودار (۲): تابع بقا برای متغیر تعداد اقساط



نمودار (۳): تابع بقا برای متغیر نرخ بهره

حال به بررسی اثر متغیرهای فردی بر تابع بقا و نرخ خطر پرداخته می‌شود. براساس اطلاعات بانک سن مشتریان در بازهٔ بین ۱۸ تا ۸۴ سال قرار دارد. ابتدا سن به دسته‌های کوچک‌تر تقسیم می‌شود و سپس با رسم نمودار فراوانی، این دسته‌ها با هم ترکیب می‌شوند. براین اساس متغیر سن به دستهٔ صفر ۱۸–۲۳ سال، دستهٔ یک ۲۳–۲۸ سال، دستهٔ دو ۲۸–۳۳ و الی آخر دسته‌بندی می‌شود. متغیر وضعیت تأهل نیز دو کد صفر و یک را به خود اختصاص داده است؛ به‌طوریکه کد صفر مجرد بودن و کد یک متأهل بودن را نشان می‌دهد. متغیر تعداد فرزند در بازهٔ بین صفر تا ۱۰ فرزند قرار دارد متغیر سطح تحصیلات ۹ کد (بیسواد، سیکل، سوم راهنمایی، دیپلم، فوق دیپلم، لیسانس، فوق لیسانس، دکتری، فوق دکتری) را به خود اختصاص داده است.نوع شغل در ۴ کد قرار داده شده است که کد ۱ شغل‌های حقیقی دولتی، کد ۲ شغل‌های حقیقی خصوصی، کد ۳ شغل‌های

حقوقی دولتی و کد ۴ شغل‌های حقوقی خصوصی را نشان می‌دهند. برای متغیر شرح شغل، کدهای ۱۱ تا ۲۴ مربوط به شغل‌های به ترتیب کارمند، کارگر، نظامی، معلم، کارکنان نهادها، استاد دانشگاه، پزشک، مهندس و وکیل، بازرگان، کشاورز، صنعتکار، فروشنده و بازنشسته می‌باشند.لازم به ذکر است طبق اطلاعات اخذ شده از بانک، برای تعداد زیادی از مشتریان عنوان شغلی به طور کامل مشخص نبود و لذا این مشتریان از تحلیل حذف شدند. از بین شغل‌های موجود ۱۱۳۶ نفر کارمند، ۱۷۲ نفر کارگر، ۴۳۳ نفر معلم، ۱۳۸ نفر مهندس و ۲۰۶ نفر بازنشسته می‌باشند. لذا مابقی شغل‌ها به دلیل فراوانی بسیار کمی که دارند در تحلیل تابع بقا و تابع نرخ خطر حذف شدند.

حال به دسته‌بندی براساس نمودار توزیع فراوانی پرداخته می‌شود. براساس نمودار توزیع فراوانی سن، به دلیل فراوانی کم داده‌ها در دسته صفر و یک، این گروه‌های سنی با هم ترکیب می‌شوند. یعنی گروه سنی ۱۸–۲۸ سال در یک طبقه قرار می‌گیرند. همچنین گروه‌های سنی در دسته‌های ۷ تا ۱۳ به دلیل فراوانی کم در یک طبقه قرار می‌گیرند. براین اساس ۷ طبقه سنی برای وارد نمودن این متغیر در مدل کاکس وجود دارد؛ که شامل دسته صفر ۱۸–۲۸ سال، دسته یک ۲۸–۳۳ سال، دسته دو ۳۳–۳۸ سال، دسته سه ۳۸–۴۳ سال، دسته چهار ۴۳–۴۸ سال، دسته پنج ۴۸–۵۳ سال و دسته شش از ۵۳ سال به بالا می‌باشد. در مورد متغیر وضعیت تأهل، به دلیل اینکه در اطلاعات خام همین دو دسته وجود داشت ترجیح داده شد تحلیل‌ها براساس همین دو گروه صورت گیرد. براساس نمودار توزیع فراوانی تعداد فرزند، تعداد ۳ فرزند بیشتر در یک دسته قرار داده شد. بر این اساس ۴ دسته برای فرزند وجود دارد که دسته‌های صفر، یک و دو به ترتیب شامل ۰، ۱ و ۲ فرزند و دسته سه برای از ۳ فرزند بالاتر می‌باشند. براساس توزیع فراوانی متغیر تحصیلات افراد زیر سیکل تعداد بسیار کمی دارند لذا این گروه با افراد دارای دیپلم در یک طبقه قرار داده شدند. همچنین به دلیل فراوانی کم افراد از لیسانس بالاتر، این گروه با افراد دارای فوق دیپلم در یک گروه قرار گرفتند. لذا برای متغیر تحصیلات دو دسته وجود دارد. دسته صفر از بیسواد تا دیپلم و دسته یک از فوق دیپلم به بالا. نمودار توزیع فراوانی نوع شغل نیز نشان می‌دهد فراوانی کدهای ۳ و ۴ بسیار کم است. از آنجا که این کدها به ترتیب مربوط به شغل‌های حقوقی دولتی و حقوقی خصوصی است این کدها با کدهای ۱ (حقیقی دولتی) و ۲ (حقیقی خصوصی) ترکیب شده‌اند. براین اساس کدهای یک و سه در یک گروه قرار گرفتند که نشان‌دهندهٔ شغل‌های دولتی

است. همچنین کدهای ۲ و ۴ در یک گروه قرار گرفتند که نشان‌دهندهٔ شغل‌های خصوصی است. برای متغیر شرح شغل نیز پس از حذف مشتریان با شغل‌های فراوانی کم، ۵ شغل در نظر گرفته می‌شوند. در اینجا کد صفر کارمند، کد یک کارگر، کد دو معلم، کد سه مهندس و کد ۴ بازنشسته و را نشان می‌دهد.

در نهایت براساس مدل کاکس، ضرایب مشابه مربوط به دسته‌های مختلف در طبقات نهایی قرار می‌گیرند. برای این کار همانند قبل و براساس روش استپانوا و توماس (Stepanova & Thomas, 2002) به طبقه‌بندی نهایی پرداخته می‌شود. همان‌طور که قبلاً عنوان شد در جدول نتایج دستهٔ صفر به عنوان سطح مبنا در نظر گرفته می‌شود و ضرایب مربوط به دسته‌های دیگر نسبت به دستهٔ مبنا برآورد می‌شود.

جدول (۳): برآورد مدل کاکس برای دسته‌بندی نهایی ویژگی‌های فردی مشتریان

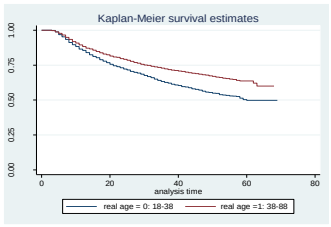
متغیر	طبقه	ضریب برآوردی	سطح معنی‌داری
سن	دسته اول: ۲۸-۳۳ سال	۰/۰۱	۰/۸۲
	دسته دوم: ۳۳-۳۸ سال	-۰/۰۳	۰/۶۵
	دسته سوم: ۳۸-۴۳ سال	-۰/۳۱	۰/۰۰
	دسته چهارم: ۴۳-۴۸ سال	-۰/۳۹	۰/۰۰
	دسته پنجم: ۴۸-۵۳ سال	-۰/۴۲	۰/۰۰
	دسته ششم: ۵۳ سال به بالا	-۰/۳۹	۰/۰۰
وضعیت تأهل	کد یک: متأهل	-۰/۰۰۳	۰/۹۵
تعداد فرزند	کد یک: یک فرزند	۰/۰۳	۰/۶۳
	کد دو: دو فرزند	-۰/۱۹	۰/۰۰
	کد سه: از سه فرزند به بالا	-۰/۱۵	۰/۰۴
آموزش	کد یک: فوق دیپلم به بالا	-۰/۱۸	۰/۰۰
نوع شغل	کد یک: شغل‌های خصوصی	۰/۳۸	۰/۰۰
شرح شغل	کد یک: کارگر	۰/۱۹	۰/۱۳
	کد دو: معلم	-۰/۴۲	۰/۰۰
	کد سه:مهندس	۰/۲۴	۰/۰۸
	کد چهار: بازنشسته	-۰/۵۲	۰/۰۰

منبع: محاسبات تحقیق

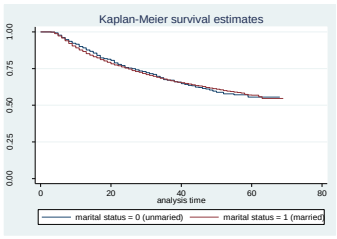
نتایج حاصل از جدول ۳ نشان می‌دهد که برای متغیر سن، ضرایب دستهٔ اول و دوم اختلاف

معنی‌داری با صفر ندارند و لذا با دسته صفر (۱۸-۲۸) ترکیب می‌شوند. همچنین دسته‌های سنی ۳ تا ۶ نیز به دلیل مشابهت ضرایب در یک گروه قرار می‌گیرند. لذا برای متغیر سن دو دسته نهایی صفر (۱۸-۳۸ سال) و یک (۳۸ سال به بالا) در نظر گرفته می‌شود. ضریب متغیر وضعیت تأهل نیز برای کد یک (متأهل) نشان می‌دهد که این ضریب اختلاف معنی‌داری با صفر ندارد و لذا اختلافی بین تابع نرخ خطر و بقا برای دو گروه متأهل و مجرد وجود ندارد (در ادامه این نتیجه براساس آزمون لگاریتم رتبه‌ای نیز تأیید می‌شود). ضریب متغیر تعداد فرزند نیز نشان می‌دهد که ضریب کد یک (تعداد یک فرزند) اختلاف معنی‌داری با صفر ندارد و لذا افراد دارای یک فرزند با افرادی که هیچ فرزندی ندارند (دسته صفر) در یک دسته قرار داده می‌شوند. افراد دارای ۲ فرزند و از سه فرزند به بالا نیز به دلیل مشابه بودن ضرایب آن‌ها در یک گروه قرار می‌گیرند. لذا متغیر تعداد فرزند نهایتاً به دو دسته تقسیم می‌شوند. دسته صفر شامل تعداد صفر یا یک فرزند و دسته یک شامل از دو فرزند به بالا، می‌باشد. ضرایب متغیر آموزش برای فوق دیپلم به بالا و متغیر نوع شغل برای شغل‌های خصوصی نشان می‌دهد که دسته‌بندی‌های مناسبی صورت گرفته است. لذا متغیر آموزش دارای دو دسته نهایی صفر (بیسواد تا دیپلم) و دسته یک (فوق دیپلم به بالا) و متغیر نوع شغل شامل دو دسته نهایی صفر (شغل‌های دولتی) و دسته یک (شغل‌های خصوصی) می‌باشند. ضرایب متغیر شرح شغل نشان می‌دهد که ضریب کد یک و کد سه اختلاف معنی‌داری با صفر ندارند و لذا این کدها با کد صفر ترکیب می‌شود. همچنین ضرایب کد دو و چهار نیز به هم شبیه بوده و لذا در یک گروه قرار می‌گیرند. بنابراین ۲ کد برای شرح شغل به وجود می‌آیند که عبارتند از کد صفر، شامل کارمند، مهندس و کارگر، کد یک، شامل معلم و بازنشسته.

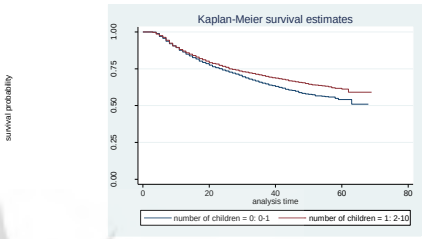
حال به برآورد تابع بقا برای متغیرهای فردی مشتریان در دسته‌های مختلف پرداخته می‌شود. همانند قبل این برآوردها با استفاده از روش ناپارامتریک کاپلان-میر صورت می‌گیرد و برای آزمون اینکه آیا منحنی‌های بقا به طور معنی‌داری با هم اختلاف دارند یا خیر از آزمون‌های لگاریتمی رتبه‌ای استفاده می‌شود.



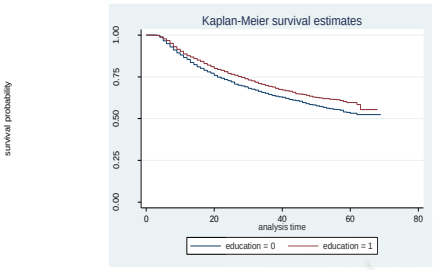
نمودار (٤): تابع بقا برای متغیر سن



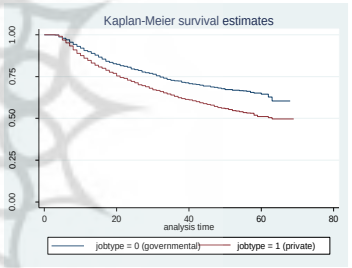
نمودار (٥): تابع بقا برای متغیر وضعیت تأهل



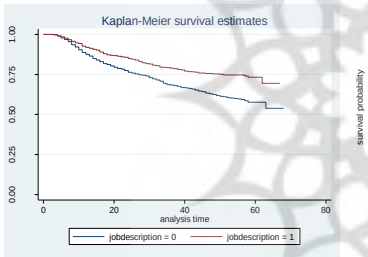
نمودار (٦): تابع بقا برای متغیر تعداد فرزند



نمودار (٧): تابع بقا برای تحصیلات



نمودار(٨): تابع بقا برای نوع شغل



نمودار (٩): تابع بقا برای شرح شغل

جدول ۴ نتایج حاصل از آزمون لگاریتمی رتبه‌ای را برای متغیرهای فردی نشان می‌دهد. فرض صفر این آزمون به معنای این است که توابع بقا بدست آمده از روش کاپلان-میر برای گروه‌های مختلف با هم اختلاف معنی‌داری ندارند؛ حال آنکه فرض یک نشان‌دهنده وجود اختلاف معنی‌دار بین توابع بقا را نشان می‌دهد.

جدول (٤): آزمون لگاریتمی-رتبه‌ای برای متغیرهای فردی مشتریان

متغیر	آماره کای دو	prob
سن	۶۲/۵۰	۰/۰۰

وضعیت تأهل	۰/۰۰	۰/۹۵
تعداد فرزند	۱۳/۸۶	۰/۰۰
سطح تحصیلات	۱۴/۷۶	۰/۰۰
نوع شغل	۵۳/۵۲	۰/۰۰
شرح شغل	۳۰/۳۴	۰/۰۰

منبع: محاسبات تحقیق

همان‌طور که جدول فوق نشان می‌دهد توابع بقا مربوط به دسته‌بندی‌های انجام شده برای هر یک از متغیرهای فردی به غیر از وضعیت تأهل، اختلاف معنی‌داری با هم دارند. نمودار ۵ تابع بقا و تابع نرخ خطر را برای متغیر وضعیت تأهل نشان می‌دهند. همان‌طور که مشاهده می‌شود منحنی‌های بقا بسیار به هم نزدیک هستند. این نتیجه توسط آزمون لگاریتمی-رتبه‌ای در جدول ۴ نیز تأیید می‌شود. براین اساس تابع بقا برای دو گروه متأهل و مجرد اختلاف معنی‌داری با هم ندارند.

براساس نمودار ۴ مشتریانی که سن آن‌ها در بازهٔ ۱۸ تا ۳۸ است نسبت به کسانی که سن آن‌ها ۳۸ سال به بالا است، بقای کمتر و ریسک قصور بیشتری دارند. همچنین براساس نمودار ۶ مشتریان با هیچ یا یک فرزند احتمال بقای کمتر و ریسک قصور بیشتر نسبت به مشتریان با دو فرزند به بالا هستند. نمودار ۷ نیز نشان می‌دهند مشتریان با تحصیلات فوق دیپلم به بالا، احتمال بقای بیشتر و ریسک قصور کمتری نسبت به مشتریان با تحصیلات دیپلم و زیر دیپلم دارند. نمودار ۸ نیز حاکی از آن است که مشتریانی دارای شغل‌های دولتی نسبت به مشتریان دارای شغل‌های خصوصی احتمال بقای بیشتر و در نتیجه ریسک قصور کمتری دارند. نمودار ۹ نیز نشان می‌دهند مشتریانی که معلم و یا بازنشسته هستند (کد یک) نسبت به مشتریانی که کارمند، مهندس و یا کارگر می‌باشند احتمال بقای بیشتر و ریسک قصوری کمتری دارند.

برآورد مدل با استفاده از مدل خطرات متناسب کاکس

مدل خطرات متناسب کاکس تلاش می‌کند اثر متغیرهای توضیحی را بر روی بقای مشتریان به طور همزمان برآورد نماید. این مدل اجازه محاسبه تابع ریسک (تابع نرخ خطر) را نیز می‌دهد.

در آنالیز چند متغیری با استفاده از رگرسیون گام به گام [۱]می‌توان، بهترین مجموعه از متغیرهای پیش‌بینی کننده را شناسایی نمود. رایج‌ترین فرآیندهای رگرسیونی مورد استفاده در تحلیل بقا حذف رو به عقب[۲] و انتخاب رو به جلو[۳] می‌باشند. آنچه واضح است نتیجه حاصل از انتخاب هر یک از روش‌ها یکسان می‌باشد. در این پژوهش از هر دو روش برای آنالیز چند متغیری استفاده شده است. اما به دلیل یکسان بودن نتیجهٔ آنها، تنها برآوردهای حاصل از روش حذف رو به عقب با استفاده از نرم افزار Stata در جدول ۵ گزارش شده است. در این تکنیک ابتدا تمامی متغیرها وارد مدل می‌شوند و سپس براساس آزمون معنی‌داری، متغیرهایی که کمترین میزان قدرت پیش-بینی را براساس p-value دارند حذف می‌شوند. متغیرهایی با بیشترین میزان p-valueکه بالاتر از سطح معنی‌داری (به طور معمول ۰/۰۵) هستند، از مدل حذف می‌شوند و دوباره مدل با متغیرهای باقی‌مانده برآورد می‌شود. این فرآیند تا زمانی که تمامی متغیرها معنی‌دار شوند ادامه پیدا می‌کند.

لازم به ذکر است جهت انجام آنالیز چند متغیری، تمامی متغیرها وارد مدل می‌شوند تا میزان اثرگذاری هر یک از عوامل بر بقا و نرخ خطر مشتریان با کنترل نمودن سایر متغیرها بررسی گردند. براین اساس متغیر مبلغ وام، تعداد اقساط، نرخ بهره، سن، وضعیت تأهل، تعداد فرزندان، آموزش و نوع شغل وارد مدل کاکس می‌شوند.

جدول (۵): برآورد ضرایب متغیرهای مستقل و نرخ خطر با استفاده از مدل کاکس

متغیر	ضریب	انحراف معیار	سطح معنی‌داری	نرخ خطر (فاصله اطمینان ۹۵٪)
مقدار وام	-۰/۰۵	۰/۰۰۹	۰/۰۰	۰/۹۵ (۰/۹۳/۹۶،۰)
نوع شغل	۰/۲۸	۰/۰۶۳	۰/۰۰	۱/۳۲ (۱/۱۷/۵۰،۱)
تعداد اقساط	۰/۳۲	۰/۰۴۹	۰/۰۰	۱/۳۷ (۱/۲۵/۵۱،۱)
سن	-۰/۱۰	۰/۰۱۵	۰/۰۰	۰/۸۹ (۰/۸۷/۹۲،۰)
وضعیت تأهل	۰/۲۷	۰/۰۹۸	۰/۰۰۶	۱/۳۱ (۱/۰۸/۵۹،۱)
تحصیلات	-۰/۰۷	۰/۰۲۵	۰/۰۰۴	۰/۹۲ (۰/۸۸/۹۷،۰)

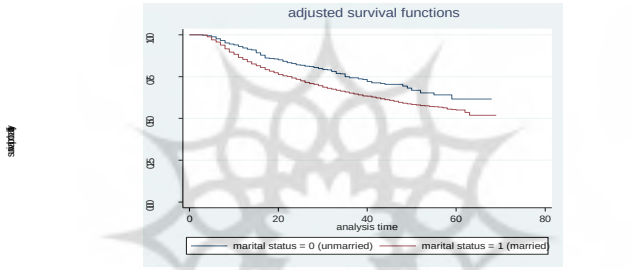
منبع: محاسبات تحقیق

۱- Stepwise Regression

۲- Backward Elimination

۳- Forward Selection

در روش شبه پارامتریک کاکس، وقتی تمامی متغیرها در کنار یکدیگر وارد مدل می‌شوند، ضریب متغیر تعداد فرزند معنی‌دار نبوده و لذا از مدل حذف شده است. اما وقتی متغیر وضعیت تأهل وارد مدل گردید، ضریب آن معنی‌دار شد. یعنی برخلاف روش ناپارامتریک کاپلان-میر، در این مدل، با ورود و کنترل سایر متغیرها، ضریب تعداد فرزند بی‌معنی و ضریب وضعیت تأهل معنی‌دار می‌باشد. لازم به ذکر است اثرگزاری سایر متغیرها در مدل شبه پارامتریک کاکس و مدل ناپارامتریک کاپلان-میر مشابه هستند. نمودار ۱۰ تابع بقای تعدیل شده برای متغیر وضعیت تأهل که درآن اثر سایر متغیرها تعدیل شده است را نشان می‌دهد. براین اساس احتمال بقای مشتریان متأهل نسبت به مشتریان مجرد کمتر و لذا ریسک قصور آن‌ها بیشتر است.



نمودار (۱۰): تابع بقای تعدیل شده برای متغیر وضعیت تأهل

حال به تفسیر ضرایب و نرخ خطر پرداخته می‌شود. ضریب منفی حاکی از آن است که با افزایش متغیر مستقل، قصور دیرتر اتفاق می‌افتد؛ به عبارتی دیگر با افزایش متغیر مستقل، طول بقای مشتریان افزایش و ریسک قصور (نرخ خطر) کاهش می‌یابد. همچنین ضریب مثبت حاکی از آن است که با افزایش متغیر مستقل، قصور زودتر اتفاق می‌افتد؛ به عبارتی دیگر می‌توان گفت با افزایش متغیر مستقل، زمان بقای مشتریان کاهش و ریسک قصور (نرخ خطر) افزایش می‌یابد. از آنجا که نرخ خطر برای هر متغیر برابر با $e^{\text{coefficient}}$ مربوطه می‌باشد، اگر ضریب منفی باشد به معنای این است که نرخ خطر کمتر از یک و اگر ضریب مثبت باشد به معنای این است که نرخ خطر بزرگتر از یک می‌باشد.

همان‌طور که نتایج جدول ۵ نشان می‌دهد با افزایش مقدار وام، زمان بقا افزایش می‌یابد که این نتیجه در آنالیز تک متغیری نیز بدست آمد. همچنین با افزایش هر واحد افزایش در مبلغ وام،

ریسک قصور به اندازهٔ (۱-۰/۹۵) یعنی ۵ درصد کاهش می‌یابد. همچنین ضریب نوع شغل نیز نشان می‌دهد مشتریان دارای شغل‌های خصوصی نسبت به شغل‌های دولتی بقای کمتری دارند و ریسک قصور آن‌ها ۳۲ درصد (۱-۱/۳۲) بیشتر است. مقدار ضریب و نرخ خطر برآوردی برای متغیر تعداد اقساط نیز نشان می‌دهد با افزایش دوره بازپرداخت، طول بقای مشتریان کاهش و ریسک قصور تقریباً ۳۷ درصد (۱-۱/۳۷) افزایش می‌یابد. ضریب متغیر سن نیز نشان می‌دهد با افزایش سن، طول دورهٔ بقا افزایش و ریسک قصور ۱۱ درصد (۱-۰/۸۹) کاهش می‌یابد. ضریب متغیر وضعیت تأهل نیز نشان می‌دهد که متأهل‌ها نسبت به مجردها بقای کمتر و در نتیجه ریسک قصور بیشتری دارند؛ میزان ریسک قصور متأهل‌ها ۱/۳۱ برابر ریسک قصور مجردها می‌باشد. علاوه براین، ضریب متغیر تحصیلات حاکی از آن است که با افزایش تحصیلات طول دوره بقای مشتریان افزایش و ریسک قصور به میزان ۸ درصد (۱-۰/۹۲) کاهش می‌یابد. براساس روش رگرسیون گام به گام، متغیر تعداد فرزند و نرخ بهره به دلیل اینکه p-value آن‌ها به ترتیب ۰/۵۷ و ۰/۰۷ بود، از مدل حذف شده‌اند.

ارزیابی فرض خطرات متناسب

یکی از پیش‌فرض‌های اساسی مدل شبه پارامتریک کاکس، فرض خطرات متناسب است. این مدل فرض می‌کند که نسبت خطر برای ۲ فرد یا گروه با ویژگی‌های متفاوت طی زمان ثابت است. یکی از آزمون‌های مورد استفاده برای ارزیابی فرض خطرات متناسب، آزمون هرل و لی (Harrel & Lee, 1986) است که براساس باقیمانده‌های شوئنفلد[1] فرض خطرات متناسب را بررسی می‌کند. این روش به دلیل اینکه برای ارزیابی فرض خطرات متناسب، آماره آزمون و p-value برای هر یک از متغیرهای مستقل ارائه می‌کند، بسیار مورد استفاده قرار می‌گیرد. با استفاده از این روش می‌توان تصمیمات عینی در مورد برقراری فرض خطرات متناسب اتخاذ نمود. اجرای این آزمون دارای سه گام است. ۱) اجرای مدل خطرات متناسب کاکس و بدست آوردن باقیمانده‌های شوئنفلد برای هر متغیر پیش‌بینی کننده. ۲) رتبه‌بندی زمان‌های رخداد قصور و ۳) آزمون همبستگی

۱- Schoenfeld residuals

بین باقیمانده‌ها و زمان‌های رخداد مرتب شده. براین اساس فرضیه صفر این آزمون عدم همبستگی بین باقیمانده‌ها و زمان بقا (و یا زمان‌های رخداد مرتب شده) می‌باشد که به صورت $H_0: \rho = 0$ نمایش داده می‌شود (Kleinbaum& Klein, 2005).

در این آزمون، رد فرضیه صفر به معنای این است که فرض خطرات متناسب را نمی‌توان پذیرفت. لذا اگر p-value آزمون کوچکتر از ۰/۰۵ باشد، فرض خطرات متناسب را نمی‌توان پذیرفت و اگر بزرگتر از ۰/۰۵ باشد، شواهد کافی برای رد فرضیه صفر وجود ندارد و لذا فرض خطرات متناسب برقرار می‌باشد. نتایج این آزمون در جدول ۶ آورده شده است.

جدول (٦): آزمون فرض خطرات متناسب برای متغیرهای مستقل

متغیر	rho	Chi2	p-value
مقدار وام	۰/۰۳۹	۲/۵۱	۰/۱۱
نوع شغل	-۰/۰۰۰۷	۰/۰۰	۰/۹۷
تعداد اقساط	-۰/۰۱	۰/۳۴	۰/۵۵
سن	-۰/۰۴	۳/۵۹	۰/۰۶
وضعیت تأهل	-۰/۰۳۴	۱/۹۷	۰/۱۶
آموزش	۰/۰۲۰	۰/۷۱	۰/۳۹

منبع: محاسبات تحقیق

همان‌طور که نتایج جدول نشان می‌دهد، p-value برای تمامی متغیرها بزرگتر از ۰/۰۵ بوده و لذا فرض خطرات متناسب برای هیچکدام متغیرهای موجود در مدل را نمی‌توان رد نمود.

آزمون مناسب بودن مدل:

برای اطمینان از درستی تفسیر نتایج مدل و بررسی مناسب بودن مدل، از انواع مختلفی از باقی-مانده‌ها استفاده می‌شود. این باقی‌مانده‌ها تحت عنوان ابزارهای تشخیصی نیز شناخته می‌شوند. در روش‌های رگرسیون‌های خطی، باقیمانده‌ها به سادگی تفاوت مقادیر مشاهده شده و مقادیر پیش-بینی شده برای متغیر وابسته را منعکس می‌کنند. اگر نمودار باقیمانده‌ها به طور تصادفی در اطراف صفر پراکنده شده باشند به معنای برازش مناسب مدل است. اما زمانیکه مشاهدات سانسور شده در مدل وجود داشته باشد و تنها از روش حداکثر راست‌نمایی برای برآورد مدل خطرات متناسب استفاده شود، مفهوم رایج باقیمانده قابل کاربرد نیست (Lee & Wang, 2003). چند روش برای

محاسبه باقیمانده در روش تحلیل بقا وجود دارد که عبارتند از باقیمانده کاکس-اسنل[۱]، باقیمانده مارتینگل[۲] و باقیمانده شوئنفلد. به دلیل رایج بودن باقیمانده کاکس اسنل و جلوگیری از طولانی شدن مباحث در ادامه این روش توضیح داده می‌شود.

باقیمانده کاکس-اسنل: کاکس و اسنل (Cox & Snell, 1968) این باقیمانده را به صورت زیر تعریف نمودند (Stepanova & Thomas, 2002).

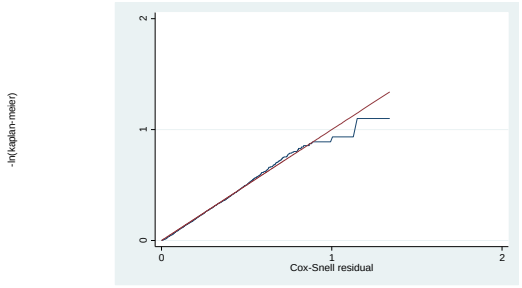
$$r_{C_i} = \exp(\hat{\beta} x_i)\, \hat{H}_0(t_i) = \hat{H}_i(t_i) = -log\hat{S}_i(t_i) \qquad (۱۱)$$

در این رابطه $\hat{H}_0(t_i)$ خطر مبنای تجمعی برآورد شده، $\hat{H}_i(t_i)$ خطر تجمعی برآورد شده برای i امین فرد در زمان t_i و $\hat{S}_i(t_i)$ تابع بقای برآورد شده برای i امین فرد در زمان t_i می‌باشد. می‌توان نشان داد که $\log S(t)$- از توزیع نمایی تبعیت می‌کند (Collett, 1994). بنابراین باید این امر را بررسی نمود که آیا باقیمانده‌ها می‌توانند از توزیع نمایی پیروی کنند یا خیر؟ جهت انجام این کار ابتدا برآورد تابع بقا به روش کاپلان-میر برای این باقی‌مانده‌ها به دست می‌آید و تابع خطر تجمعی این برآورد ($-log\hat{S}(r_{C_i})$) محاسبه می گردد. سپس نمودار این تابع خطر نسبت به باقیمانده کاکس-اسنل (r_{C_i}) رسم می‌شود. هرچه نمودار به دست آمده به خط نیمساز نزدیک باشد (شیب یک و عرض از مبدأ صفر) مدل برازش شده مناسب خواهد بود. این روش باقیمانده برای ارزیابی کلی مدل به کار می‌رود. نمودار ۱۱ نمودار باقیمانده کاکس-اسنل را برای مدل برآورد شده به روش خطرات متناسب کاکس نشان می‌دهد.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

۱- Cox-Snell residuals

۲- Martingale residual



نمودار (۱۱): باقیمانده کاکس-اسنل

همان‌طور که نمودار باقیمانده کاکس-اسنل نشان می‌دهد، باقیمانده به خط ۴۵ درجه بسیار نزدیک می‌باشد و لذا مدل کاکس به طور مناسبی به داده‌ها برازش داده شده است.

جمع‌بندی:

بانک‌ها و مؤسسات مالی و اعتباری به دلیل نوع فعالیت‌های خود با ریسک‌های زیادی از جمله ریسک‌های اعتباری، نرخ بهره، نقدینگی و ... مواجه هستند. بانک‌ها می‌توانند با اعمال روش‌ها و ابزارهایی آن‌ها را تحت کنترل درآورند. در میان ریسک‌هایی که بانک‌ها را تهدید می‌کند، ریسک اعتباری به دلیل اینکه عمده مصارف بانکی اعطای وام و اعتبار است، مهم‌ترین ریسک به شمار می‌رود. لذا بانک‌ها بایستی نگرش خود را در زمینه اعطای وام علمی نموده و بتواند با بررسی صحیح و همه‌جانبهٔ درخواست‌های مشتریان، زمینه لازم را برای موفقیت نظام بانکی و تضمین بازگشت سرمایه و سود مورد انتظار آن‌ها فراهم نماید. تحلیل بقای مشتریان به دلیل اینکه به بررسی این موضوع می پردازد که پارامترهای فردی و مربوط به وام چه اثری بر زمان تا وقوع قصور مشتریان دارد بسیار مورد توجه می‌باشد.

در این مطالعه به بررسی اثر متغیرهای فردی و متغیرهای مربوط به وام بر ریسک قصور (تابع نرخ خطر) و بقای مشتریان پرداخته شد. بدین‌منظور از یک نمونه تصادفی ۵۳۱۶ نفر از مشتریان بانک مسکن استان خراسان رضوی که طی سال‌های ۱۳۸۸-۱۳۹۳ وام گرفته‌اند استفاده گردید. جهت رسیدن به هدف مطالعه، روش‌های ناپارامتریک کاپلان میر و شبه پارامتریک خطرات متناسب کاکس، که از روش‌های رایج در زمینه اعتبارسنجی مشتریان در حوزه تحلیل بقا می‌باشند،

به کار گرفته شد. در ابتدا با استفاده از تحلیل تک متغیری کاپلان-میر به دسته‌بندی هریک از متغیرها با توجه به میزان فراوانی و روش پیشنهادی استپانوا و توماس (۲۰۰۲) پرداخته و سپس منحنی‌های تابع بقا و نرخ خطر در مقابل زمان تا وقوع قصور برای هر یک از متغیرهای در دسترس ترسیم شدند. نتایج حاصل از این مدل نشان داد که افرادی که بالای ۱۷۰ میلیون ریال وام می‌گیرند نسبت به افرادی که کمتر از ۱۷۰ میلیون ریال وام می‌گیرند بقای بیشتر و در نتیجه ریسک قصور کمتری دارند. این امر به دلایلی مانند گرفتن وثیقه‌های بالاتر و توجه بیشتر به شغل و درآمد افراد متقاضی رقم وام‌های بیشتر می‌تواند منطقی به نظر برسد. همچنین با افزایش طول دوره بازپرداخت (از ۱۲۰ به ۱۴۴ و ۱۸۰ قسط) احتمال بقا کاهش و ریسک قصور افزایش می‌یابد. همچنین مشتریان در بازه سنی ۱۸-۳۸ سال نسبت به مشتریان ۳۸ سال به بالا احتمال بقای کمتر و ریسک قصور بیشتری دارند. نتایج حاصل از برآوردگر کاپلان-میر مشتریان با دو فرزند و بیشتر نسبت به مشتریانی که حداکثر یک فرزند دارند، احتمال بقای بیشتر و ریسک قصور کمتری دارند. همچنین افراد دارای تحصیلات دیپلم و زیر دیپلم نسبت به مشتریان دارای تحصیلات بالاتر احتمال بقای کمتر و ریسک قصور بیشتری دارند. علاوه بر این شغل‌های دولتی نسبت به شغل‌های خصوصی ریسک قصور کمتر و احتمال بقای بیشتری دارند. همچنین معلمین و بازنشسته‌ها نسبت به سایرین ریسک قصور کمتر و احتمال بقای بیشتری در سیستم بانکی دارا می‌باشند.

در این مطالعه علاوه بر آنالیز تک متغیری به روش ناپارامتریک، به منظور میزان تأثیرگذاری هر یک از متغیرها بر ریسک قصور، تمامی متغیرها به طور همزمان وارد مدل شبه پارامتریک کاکس شدند. در این صورت اثر هریک از متغیرها با کنترل نمودن سایر متغیرها به طور دقیق‌تر بررسی می‌شود. برآورد این مدل حاکی از عدم وجود رابطهٔ معنی‌دار بین نرخ بهره و تعداد فرزندانبر ریسک قصور مشتریان است. لازم به ذکر است عدم وجود رابطه معنی‌دار بین نرخ بهره و وضعیت قصور مشتریان به دلیل این است که نرخ بهره وام‌ها در دو دسته ۱۱ و ۱۳ درصد قرار دارند و لذا به دلیل نزدیک بودن نرخ بهره‌ها به همدیگر این نتیجه دور از انتظار نیست. اگر طیف گسترده‌ای از نرخ بهره‌ها وجود داشته باشد می‌توانست نتیجه دیگری در این خصوص بدست آید. نتایج حاصل از برآورد نشان داد که با افزایش یک واحد در مبلغ وام ۵ درصد از ریسک قصور کاسته می‌شود. همچنین با افزایش سن ۱۱ درصد ریسک قصور کاهش می‌یابد. علاوه بر این، با افزایش هر مقطع تحصیلی ۸ درصد از ریسک قصور کاسته می‌شود. همچنین با افزایش طول دوره

بازپرداخت ریسک قصور ۳۷ درصد افزایش می‌یابد. مشتریان دارای شغل‌های خصوصی نسبت به شغل‌های دولتی ۳۲ درصد ریسک قصور بیشتری دارند و متأهل بودن منجر به افزایش ۳۱ درصدی در ریسک قصور مشتریان می‌شود. آزمون هرل و لی نیز نشان داد که فرض کلیدی خطرات متناسب برای مدل کاکس برقرار می‌باشد و لذا با اطمینان می‌توان نتایج حاصل از این مدل را تفسیر نمود. ابزارهای تشخیص همچون باقیمانده‌های کاکس-اسنل صحت مدل را تأیید نمود. لذا بانک‌ها جهت کاهش ریسک اعتباری و تصمیم‌گیری در مورد متقاضیان وام بایداین متغیرها را با دقت بیشتری مد نظر قرار دهند.

References

[1] Abdou, H &Pointon, J. (2011). Credit Scoring,Statistical Techniques and Evaluation Criteria: a Review of the Liteature, *Intelligence Systems in Accounting, Finance and Management*, 18(2,3), 59-88.

[2] Allen, L. N. & Rose, L. C. (2006). Financial survival analysis of defaulted debtors. *Journal of Operational Research Society*, 57, 630-636.

[3] Arabmazar, A. & Ruintan, P. (2006). Effective Factors on Credit Risk of Bank Customers (Case Study of Keshavarzi Bank). *Journal of Economic literature*, 3(6), 45-80. (In Persian)

[4] Asli, SH. (2011). *Credit Risk Management with Regards to Patterns of Loan Payment in Other Countries*. Research Administration and Risk Control of Sepah Bank. (In Persian)

[5] Baba, N. &Goko, H. (2006). Survival analysis of hedge funds, Bank of Japan. *Working Papers Series*, No.06-E-05.

[6] Baesens, B., Van gestel, T., Stepanova, M. & Van den Poel, D. (2005). Neural Network Survival Analysis for personal Loan Data. *Journal of the Operational Research Society*, 56(9), 1089-1098.

[7] Banasik, J., Crook, J. & Thomas, L.C. (1999).Not if but When will Borrowers Default. *The journal of the operational research*, 50(12), 1185-1190.

[8] Basel Committee on Banking Supervision. (2006). ‘Itt rrttt illll Convergence of Capital measurement and Capital Standards: A Revised aaamkkkkk Cmmrr kkkiii vV Vrr ii ’’’ .

[9] Bellotti, T. & Crook, J. (2013). Forecasting and stress testing credit card default using dynamic models. *International Journal of Forecasting*, 29(4), 563-574.

[10] Beran, J. &Djaidja, A.K. (2007). Credit risk modeling based on survival analysis with inmunes. *Statistical Methodology*, 4(3), 251-276.

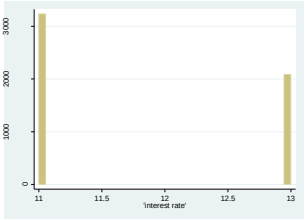
[11] Cao, R., Vilar, J.M. &Devia, A. (2009). Modeling Consumer Credit Risk via Survival Analysis. *Sort*, 33(1), 3-30.

[12] Carling, K., Jacobson, T. &Roszbach, K. (1998). Duration of consumer loans and bank lending policy: dormancy versus default risk. *Working Paper Series in Economics and Finance*. Stockholm School of Economics, 280.
[13] Glasson, S. (2007). *Censored Regression Techniques for Credit Scoring*, Ph.D Thesis, RMIT University, Australia, Melbourne.
[14] Glennon, D. &Nigro, P. (2005). Measuring the default risk of small business loans: a survival analysis approach. *Journal of Money, Credit and Banking*, 37(5), 923-947.
[15] Kalbfleisch, J., & Prentice, R. (1980). *The statistical analysis of failure time data.* NewYork: Wiley.
[16] Kleinbaum, D.G. & Klein, M. (2005). *Survival Analysis, a Self-Learning Text* (Second Ed.). Springer.
[17] Lee, E.T., & Wang, J.W. (2003). *Statistical Methods for Survival Data Analysis* (Third Ed.). John Wiley & Sons
[18] Malik, M &Thomas L.C. (2006). Modeling credit risk of portfolio of consumer loans. University of Southampton. *School of Management Working Paper Series*, No. CORMSIS-07-12.
[19] Man, R. (2014). *Survival Analysis in Credit Scoring: a Framework for PD Estimation*, Ph.D thesis, University of Twente, Netherlands, Enschede.
[20] Monteverde, K. (2000). Managing Student Loan Default risk: Evidence from a Privately Guaranteed Portfolio. *Research in Higher Education*, 41(3), 331-352.
[21] Myers, J.H. &Forgy, E.W. (1963). The Development of Numerical Credit Evaluation Systems. *Journal of the American Statistical Association*, 58(303), 799-806.
[22] Roszbach, K. (2004). Bank lending policy, credit scoring and the survival of loans. Review of Economics and Statistics, 86(4), 946-958.
[23] Sarlija, N., Bensic, M. &Zekic-Susac, M. (2006). Modeling Customer Revolving credit Scoring Using Logistic Regression, Survival Analysis and Neural Networks. Paper presented at the 7th WSEAS international Conference on Neural Networks.
[24] Stepanova, M. &Thomas, L.C. (2002). Survival Analysis Methods for Personal Loan Data. *Opertion Research*, 50(2), 277-289.
[25] Thomas, L.C. (2009). *Consumer Credit Models: Pricing, Profit, and Portfolios* (first Ed.). Oxford University Press.
[26] Thomas, L.C., Edelman, D.B., & Crook, J.N. (2002). *Credit Scoring and Applications*. Philadelphia: Society for Industrial and Applied Mathematics.
[27] Tong, E.N., Mues, C. & Thomas, L.C. (2012). Mixture cure models in credit scoring: If and when borrowers default. *European Journal of Operational Research*, 218(1), 132-139.

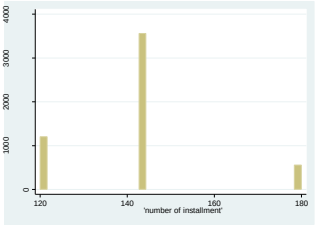
پیوست

پیوست ۱- نمودار توزیع فراوانی متغیرهای مورد بررسی

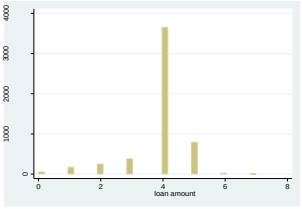
نمودار ۱- توزیع فراوانی نرخ بهره



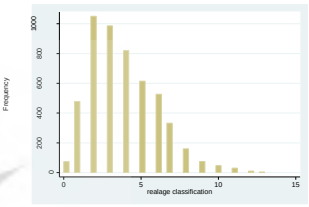
نمودار ۲- توزیع فراوانی تعداد اقساط



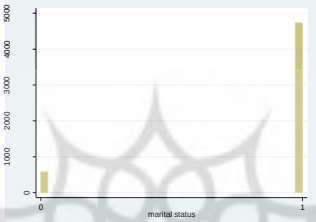
نمودار ۳- توزیع فراوانی دسته‌های مبلغ وام



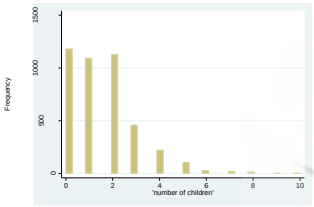
نمودار ٤- توزیع فراوانی دسته‌های سنی



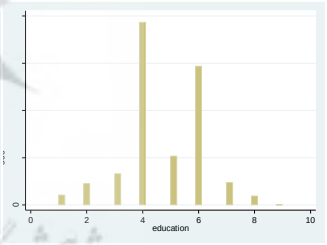
نمودار ٥- توزیع فراوانی تأهل



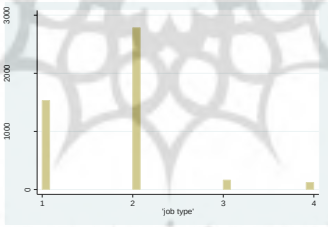
نمودار ٦- توزیع فراوانی تعداد فرزند



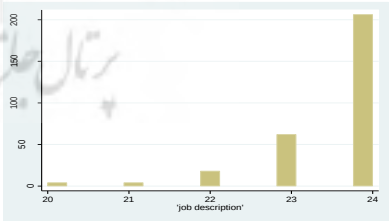
نمودار ۷- توزیع فراوانی کدهای تحصیلی



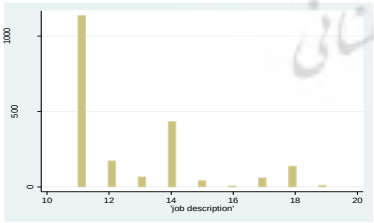
نمودار ۸- توزیع فراوانی کدهای شغلی



نمودار ۹- توزیع فراوانی عنوان شغل (کدهای ۱۱-۱۹)

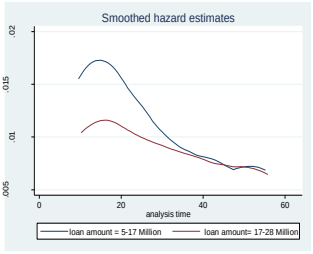


نمودار ۱۰- توزیع فراوانی عنوان شغل (کدهای ۲۰-۲٤)

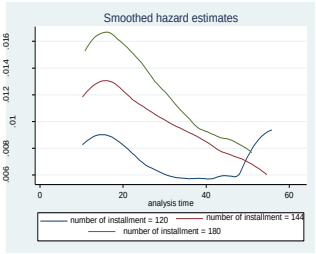


پیوست ۲- توابع نرخ خطر برآوردی به روش ناپارامتریک کاپلان-میر

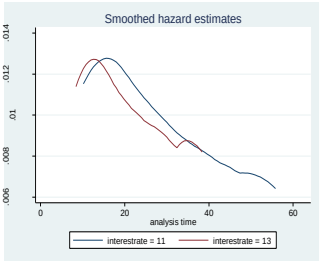
نمودار ۱- تابع نرخ خطر مبلغ وام



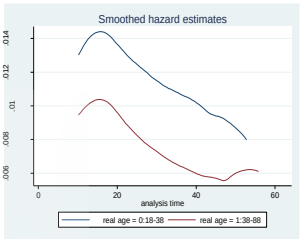
نمودار۲- تابع نرخ خطر تعداد اقساط



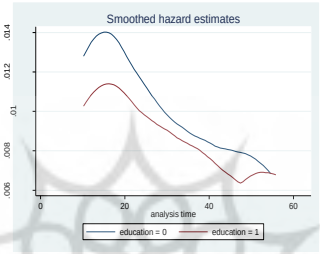
نمودار ۳- تابع نرخ خطر نرخ بهره



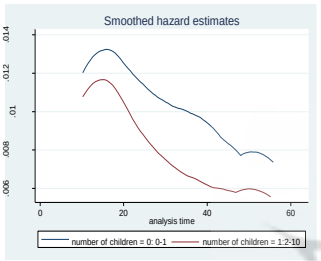
نمودار ٤- تابع نرخ خطر سن



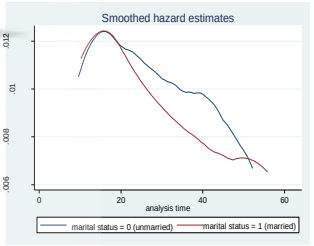
نمودار ٥- تابع نرخ خطر تحصیلات



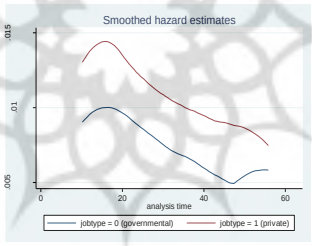
نمودار ٦- تابع نرخ خطر تعداد فرزند



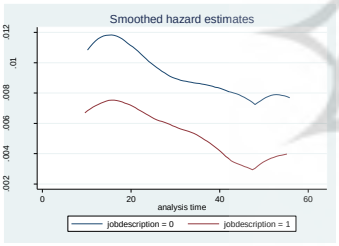
نمودار ۷- تابع نرخ خطر تأهل



نمودار ۸- تابع نرخ خطر نوع شغل



نمودار ۹- تابع نرخ خطر شرح شغل



پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی