

UNIVERSIDAD NACIONAL DE INGENIERÍA
FACULTAD DE INGENIERÍA ECONÓMICA, ESTADÍSTICA
Y CIENCIAS SOCIALES



TESIS

**“UTILIZACIÓN DE MODELOS DE REGRESIÓN DE COX PARA LA ESTIMACIÓN DE
PROBABILIDADES DE INCUMPLIMIENTO EN UNA CARTERA DE CRÉDITOS
MINORISTA”**

**PARA OBTENER EL GRADO ACADÉMICO DE MAESTRO EN CIENCIAS EN
ECONOMETRÍA BANCARIA Y FINANCIERA**

ELABORADO POR:
VÍCTOR FERNANDO MARISCAL CUETO
ORCID: 0009-0008-3629-2515

ASESOR:
MG. OSCAR BENITO TOLEDO GUERRERO
ORCID: 0009-0005-0493-3531
LIMA – PERÚ
2023

DEDICATORIA

PARA MIS PADRES, PILAR DE
LO QUE LOGRÉ Y A MI HIJA
POR SER LA INSPIRACIÓN
PARA CONSEGUIRLO

RESUMEN

Esta investigación plantea la técnica econométrica conocida como el modelo de riesgos proporcionales de Cox como alternativa al método de regresiones logísticas comúnmente utilizado en la industria financiera peruana. El modelo de riesgos proporcionales de Cox aplicado de forma complementaria junto a la metodología de perdidas esperadas de las normas internacionales de contabilidad sobre instrumentos financieros NIIF 9; con la finalidad de brindar mejores estimaciones de pérdidas esperadas que son validadas con la observación de las pérdidas efectivamente materializadas en la cartera de créditos analizada.

Palabras claves:

Modelo de riesgos proporcionales de COX, perdidas esperadas, instrumentos financiero NIIF 9, estimaciones de perdidas esperadas.

ABSTRACT

This research proposes the econometric technique known as the Cox proportional hazardsmodel as an alternative to the logistic regression method commonly used in the Peruvianfinancial industry. This proportional hazards model, due to its unique characteristics, is applied in a complementary manner together with the expected loss methodology of the new international accounting standards on financial instruments IFRS9., in order to provide better estimates of expected losses that are validated by observing the losses actually materialized in the analyzed loan portfolio.

Key Words

Proportional Hazard Models, Expected Losses, IFRS9

INDICE

RESUMEN	iii
ABSTRACT	iv
LISTA DE TABLAS	vii
LISTA DE GRÁFICAS.....	viii
GLOSARIO DE TÉRMINOS	ix
INTRODUCCION.....	x
CAPITULO I: PLANTEAMIENTO DEL PROBLEMA.....	1
1.1 Descripción de la situación problemática:.....	1
1.2 Formulación del Problema	1
1.2.1 Problema General:.....	1
1.2.2 Problemas Específicos.....	1
1.3 Objetivos de la investigación.....	2
1.3.1 Objetivo general	2
1.3.2 Objetivos Específicos	2
1.4 Justificación, alcance y limitaciones de la investigación.....	2
CAPITULO II: MARCO TEÓRICO – CONCEPTUAL	4
2.1 Antecedentes de la investigación:	4
2.2 Principales diferenciadores de la presente investigación respecto a las investigaciones previas revisadas.....	7
2.3 Teorías Básicas.....	8
2.3.1 Conceptos básicos de riesgo de crédito	8
2.3.2 Conceptos teóricos del análisis de supervivencia.....	15
2.4 Marco Conceptual	19
2.4.1 Modelos de supervivencia – Regresión de Cox	19
2.4.2 Supuestos del modelo de regresión de cox.....	21
2.4.3 Modelo de regresión de Cox con variables dependientes del Tiempo	24
2.4.4 Modelo de regresión logística	24
2.4.5 Estimación de pérdidas esperadas por riesgo de crédito	25
CAPITULO III: DIAGNOSTICO	28
3.1 Población, muestras de desarrollo y validación.....	28
3.1.1 Definición de la población y la muestra:.....	28

3.2 Definición de la variable objetivo	29
3.3 Variables inicialmente analizadas	30
3.4 Pruebas univariadas de Capacidad de predicción de las variables analizadas.....	31
3.5 Modelo de Regresión Logística.....	33
3.5.1 Forma funcional del modelo regresión logística	33
3.5.2 Indicadores de capacidad de predicción	35
3.6 Modelo de Regresión de Cox	36
3.6.1 Estimador de Kaplan Meier.....	36
3.6.2 Coeficientes del modelo de regresión de Cox.....	38
3.6.3 Indicadores de capacidad de predicción	39
3.6.3 Análisis de los supuestos de riesgos proporcionales	39
3.6.4 Análisis del supuesto de linealidad	42
3.7 Análisis comparativo de pérdidas esperadas	42
3.7.1 Resultado de pérdidas esperadas estimadas bajo modelo de regresión logística	43
3.7.2 Resultado de pérdidas esperadas estimadas bajo modelo de regresión de Cox.....	44
3.8 Estimación del número de incumplimientos durante la vida de las operaciones de crédito	46
CAPITULO IV: PROPUESTAS	48
CONCLUSIONES Y RECOMENDACIONES	49
REFERENCIAS BIBLIOGRÁFICAS	50
ANEXOS	51

LISTA DE TABLAS

TABLA 1:	Resumen de investigaciones relacionadas con métodos de supervivencia .	6
TABLA 2:	Distribución de las muestras de desarrollo y validación	28
TABLA 3:	Variables explicativas incluidas en la muestra de desarrollo inicial.....	30
TABLA 4:	Coeficientes del modelo de regresión logística.....	34
TABLA 5:	Forma Funcional de modelo de regresión de Cox.....	38
TABLA 6:	Análisis de residuos de Shoenfeld – Modelo de Cox.....	40
TABLA 7:	Calibración de la probabilidad de Incumplimiento	43
TABLA 8:	Resumen agregado de Pérdida esperada y observada para el período analizado bajo método de regresiones logísticas.....	43
TABLA 9:	Resumen agregado de Pérdida esperada y observada para el período analizado bajo método de regresiones de Cox	45
TABLA 10:	Comparativo Pérdida Esperada bajo modelos de Regresión Logística y de Cox comparados con la pérdida efectivamente materializada	46

LISTA DE GRÁFICAS

GRÁFICO 1:	Pérdidas esperadas y no esperadas.....	10
GRÁFICO 2:	Funciones de supervivencia teóricas.....	16
GRÁFICO 3:	Funciones de supervivencia empíricas	16
GRÁFICO 4:	Censura derecha de datos.....	18
GRÁFICO 5:	Representación de un gráfico Log – Log para evaluación de supuesto de riesgos proporcionales.....	22
GRÁFICO 6:	Función logística normalizada	25
GRÁFICO 7:	Tasas de migración según rango de atraso para mes 6 de madurez.	29
GRÁFICO 8:	Tasas de migración según rango de atraso para mes 12 de madurez	30
GRÁFICO 9:	Variables según su valor de información.....	32
GRÁFICO 10:	Indicadores de capacidad de predicción del modelo de regresión logística en muestras de desarrollo y validación	35
GRÁFICO 11:	Análisis de validación cruzada regresión logística	36
GRÁFICO 12:	Gráfico de riesgo acumulativo basado en estimador Kaplan Meier	36
GRÁFICO 13:	Gráfico de riesgo acumulativo basado en estimador Kaplan Meier - segmentado por nivel de Ingresos y perfil laboral	37
GRÁFICO 14:	Indicadores de capacidad de predicción del modelo de regresión de Cox en muestras de desarrollo y validación	39
GRÁFICO 15:	Gráfico Log-Log de la variable “Diferencia Entidades”	40
GRÁFICO 16:	Gráfico Residuos de Schoenfeld de variable Ingresos versus el tiempo	41
GRÁFICO 17:	Gráfico de los residuos de Schoenfeld para la variable “Diferencia de Entidades” respecto al tiempo.....	42
GRÁFICO 18:	Pérdidas observadas y esperadas según período de desembolso bajo método de regresiones de Cox	44
GRÁFICO 19:	Perdida real y brechas versus modelo de cox y regresión logística.	46
GRÁFICO 20:	Número de incumplimientos pronosticados y estimados bajo el modelo de regresión de Cox para cada punto temporal de las operaciones de crédito	47

GLOSARIO DE TÉRMINOS

Riesgo de crédito: Probabilidad de que una contraparte crediticia no se encuentre en la capacidad de cumplir con sus obligaciones en el tiempo y/o forma pactados inicialmente.

Reservas de Capital regulatorio: Capital exigido por el órgano regulador a las entidades financieras con el fin de asegurar un nivel de solvencia mínimo en función de los riesgos asumidos por la entidad financiera en el desarrollo de sus actividades de intermediación

Provisiones por riesgo de crédito: Reservas contables registradas en el estado de resultados de las entidades financieras con el fin de anticipar y/o reconocer las pérdidas por el incumplimiento de pago de los deudores de su cartera de créditos

Registro Crediticio Consolidado (RCC): Base de datos que contiene el registro detallado de los saldos deudores de los clientes del sistema financiero y que son de uso exclusivo de las entidades financieras reguladas

Índice de Confianza Empresarial: Índice calculado por el Banco Central de Reserva que como resultado de encuestas a los principales líderes empresariales del país, mantiene una escala de 1 a 100 siendo por encima de 100 el tramo “optimista” y por debajo, el tramo “pesimista”.

IFRS9(International Financial Reporting Standards 9): Estándar internacional de reporte financiero para la contabilización de reservas sobre exposiciones sujetas a riesgo de crédito.

ROC(Receiving Operating Curve): mide la capacidad de discriminación de un modelo probabilístico de respuesta dicotómica

KS(Kolmogorov- Smirnoff), mide la máxima distancia entre las distribuciones de dos categorías y es utilizado como un indicador de capacidad de discriminación de modelos probabilísticos.

INTRODUCCION

Se considera importante que las entidades financieras puedan explorar métodos de estimación de probabilidades de incumplimiento que se encuentren acordes con los lineamientos de la Normas Internacionales de Información Financiera (NIIF 9), ello en línea con los cambios regulatorios a nivel internacional que buscan estimar las pérdidas potenciales del portafolio crediticio a lo largo de su ciclo de vida. Esto con la finalidad que puedan tener una mejor estimación de las pérdidas esperadas de sus portafolios crediticios, lo cual tendrá efectos positivos tanto en la solvencia de las entidades financieras como en los consumidores al tener mejores estimaciones de Pricing ajustado a su Perfil de Riesgo.

La adecuada cuantificación del riesgo de crédito en entidades financieras es uno de las preocupaciones más relevantes para los diferentes grupos de interés en los mercados de intermediación financiera. Tanto entidades financieras como reguladores tienen un especial interés en mejorar permanentemente las estimaciones de riesgo de las exposiciones sujetas a riesgo de crédito. Mejoras en la exactitud y capacidad prospectiva de las metodologías de pérdidas esperadas por riesgo de crédito pueden generar diversas externalidades positivas en los mercados de intermediación financiera.

El análisis de supervivencia es una herramienta cuyo uso se ha expandido a las aplicaciones asociadas a la medición del riesgo de crédito, dado que permite incluir en la modelación del riesgo de crédito a series de tiempo macroeconómicas. Entre las investigaciones aplicadas es posible destacar la realizada por Bellotti y Crook (2009) y Gandy (2012) que exploran la hipótesis de que la probabilidad de incumplimiento se encuentra afectada por las condiciones asociadas al entorno económico.

El estudio de Stapanova & Thomas(2002), puntualiza la necesidad de las entidades financieras de estimar con mayor exactitud no solo la probabilidad de default de los clientes de las entidades financieras, sino también su nivel de rentabilidad esperado y para ello identifican como necesario realizar una estimación del tiempo en el que se espera caigan en default.

En la presente investigación, se desarrolla tanto un modelo de regresión logística como un modelo de riesgos proporcionales de Cox, basado en métodos de supervivencia, sobre un portafolio de créditos consumo no revolvente desembolsados por los principales bancos del Perú entre los años 2018a y 2019. La estimación de pérdidas esperadas se realiza tanto bajo la definición de Basilea II

como de la NIIF 9 y se analizan diversos indicadores de capacidad de predicción y eficiencia en la estimación respecto a las pérdidas por riesgo de crédito efectivamente materializadas en el desarrollo de dichas operaciones de crédito.

CAPITULO I: PLANTEAMIENTO DEL PROBLEMA

1.1 Descripción de la situación problemática:

Los modelos de probabilidad con respuesta dicotómica han demostrado ser de alta utilidad en la cuantificación del riesgo de crédito en entidades financieras, sin embargo, éstos no contemplan la temporalidad del evento de interés en la determinación de la severidad de las pérdidas esperadas por estas entidades; pudiendo generar ello estimaciones sesgadas de pérdidas, reservas por riesgo de crédito y rentabilidad esperada en una cartera de créditos.

1.2 Formulación del Problema

1.2.1 Problema General:

Las estimaciones de probabilidad de incumplimiento utilizadas por las entidades financieras no consideran el tiempo al incumplimiento como un parámetro importante que afecta la rentabilidad de las operaciones crediticias, generando sesgos en las pérdidas esperadas y en la rentabilidad esperada. Esperamos resolver este problema a través del uso de Modelos de Regresión de Cox, en combinación con la metodología señalada en las NIIF 9 para pérdidas esperadas. Esperamos que ello contribuya a mejorar las estimaciones de riesgo de crédito, el Pricing ajustado a Riesgo y en última instancia la solvencia de las entidades financieras.

1.2.2 Problemas Específicos

1. Siendo que la temporalidad al evento de incumplimiento tiene un impacto importante en el volumen de las pérdidas por riesgo de crédito en una cartera de créditos, se podrían generar sesgos en la estimación de dichas pérdidas debido a que estos modelos no consideran la estimación de la probabilidad de ocurrencia del evento de interés en cada punto temporal de la vida de las operaciones de crédito.

2. Las estimaciones de probabilidad de incumplimiento mediante métodos como la regresión logística no permiten la inclusión de data censurada o truncada, generando modelos con posibles sesgos muestrales.

1.3 Objetivos de la investigación

1.3.1 Objetivo general

Realizar un análisis comparativo de la eficacia de la utilización de modelos basados en métodos de supervivencia para la estimación de pérdidas por riesgo de crédito en un portafolio de créditos minoristas, respecto a modelos de regresión logística, convencionalmente utilizados en la industria bancaria.

1.3.2 Objetivos Específicos

1. Conocer y entender la teoría de modelos de supervivencia, en específico del modelo de regresión de Cox, aplicado a la estimación del riesgo de crédito en una cartera de créditos.
2. Realizar un análisis comparativo de eficacia en la estimación de pérdidas por riesgo de crédito, entre el método de regresión logística y el método de regresión de Cox.

1.4 Justificación, alcance y limitaciones de la investigación

Es importante contar con métodos que permitan estimar la probabilidad de un evento de incumplimiento en cada punto temporal de la vida de las operaciones de crédito ya que ello representaría potenciales ventajas en los siguientes aspectos:

- a. Ayudarían a mejorar las estimaciones de rentabilidad y riesgo por cada operación de crédito, dando paso asignaciones más eficientes de tasas de interés, en beneficio de los consumidores y de la rentabilidad y solvencia de los intermediarios financieros.
- b. Dado que se espera una mejor estimación de pérdidas potenciales por riesgo de crédito y por tanto las reservas que las entidades deben realizar para cubrirlas, la aplicación de estos métodos ayudaría a fortalecer la solvencia de las mismas.

- c. Adecuaría las estimaciones de pérdidas esperadas por riesgo de crédito a la nueva normativa internacional relacionada a la valorización y reservas por riesgo de crédito sobre instrumentos financieros (IFRS9)

Para lograr el objetivo general de la investigación, se realizará un análisis comparativo de la precisión en la estimación de pérdidas por riesgo de crédito entre la regresión logística como método convencional y la regresión de riesgos proporcionales de Cox como método alternativo, aplicadas sobre operaciones de crédito consumo no revolvente otorgadas por los principales bancos del Perú.

Como exclusión, la presente investigación no abarca la estimación de impactos financieros, solvencia o de asignación de tasa de interés resultantes de la utilización de las metodologías de estimación de probabilidad de incumplimiento que son materia de estudio.

CAPITULO II: MARCO TEÓRICO – CONCEPTUAL

2.1 Antecedentes de la investigación:

La estimación de las probabilidades de incumplimiento ha sido uno de los tópicos más estudiados por los estudios financieros aplicados en los últimos 40 años desde el primer modelo Z - Altman por Altman (1968). A la fecha existen muchas metodologías que han superado la capacidad de discriminación de éste primer modelo como las regresiones logísticas, regresiones discriminantes, modelos de aprendizaje automático y redes neuronales.

En la presente investigación se plantea el uso del modelo de riesgo proporcional de Cox, como una alternativa a los anteriormente mencionados; este modelo fue planteado por primera vez por Narain (1992) y posteriormente fue desarrollado también por Carling et al. (1998), Stepanova y Thomas (2002), Roszbach (2003), Glennon y Nigro (2005), Allen y Rose (2006), Baba y Goko (2006), Malik y Thomas (2006), Beran y Djadja (2007), entre otros.

El denominador de estas investigaciones es el uso de técnicas de regresión paramétricas o semi paramétricas para modelar el tiempo hasta la ocurrencia del evento de interés; tales como modelos de Weibull y el modelo de riesgos proporcionales de Cox (Cao et al, 2009)

Una de las investigaciones aplicadas sobre éste punto más completas, es la de Stepanova y Thomas (1999), aplicada a un grupo de operaciones de crédito de consumo a personas naturales en el Reino Unido, en éste se propone:

- (1) La utilización de un método alternativo para el agrupamiento y discretización de las covariables, que tome en cuenta la censura y no sólo el ratio Bueno / Malo como suele aplicarse en análisis de regresión logística,
- (2) La utilización de técnicas alternativas para el análisis de residuos en métodos de supervivencia y
- (3) La utilización de covariables dependientes del tiempo, que ayuden a superar las restricciones de riesgos proporcionales del modelo original de Cox

Las conclusiones del estudio, al realizar la comparación de los modelos de supervivencia y los modelos de regresión logística es que las diferencias en capacidad de discriminación son casi indistinguibles.

En el estudio de CAO et al (2012), se aplica un análisis comparativo de tres métodos distintos para la estimación de la probabilidad de incumplimiento: (1) Modelos de riesgos proporcionales de Cox, (2) Modelos lineales generalizados y (3) Estimadores no paramétricos.

El estudio fue llevado a cabo sobre un grupo de operaciones de crédito consumo de un banco español. En el mismo, se concluye la capacidad de predicción del modelo de Cox similar a la del modelo de estimadores no paramétricos, y muy superior al modelo de modelos lineales generalizados.

El estudio llevado a cabo por Man (2014), se plantean procedimientos alternativos, a saber:

- (1) Para el agrupamiento de variables predictoras se utilizaron funciones de supervivencia en reemplazo de las tasas de incumplimiento utilizadas en las regresiones logísticas y
- (2) Transformaciones de variables basados en funciones de supervivencia, este procedimiento es llamado por el mismo autor como la transformación *logrank*.

Similar a la investigación de Stepanova y Thomas(1999), al ser comparados con métricas como ROC y KS, no se encuentran diferencias significativas entre el modelo de regresión logística y el modelo de Cox planteado por el investigador.

A nivel local, se ha identificado como única referencia bibliográfica aplicada a riesgos de crédito la investigación realizada por Uriarte & Agüero (2017) en la cual se aplicaron modelos de regresión de cox con variables dependientes del tiempo sobre un portafolio de créditos a medianas y grandes empresas, haciendo uso de variables propias de las empresas y variables macroeconómicas. El estudio que tiene como objetivo la estimación de un modelo de probabilidad de default a través de la regresión de Cox, se limita a evaluar la significancia estadística de las variables así como de la presentación de las funciones de riesgo y supervivencia por cada una de ellas. Según los autores, entre los motivos por los cuales los modelos de supervivencia se han vuelto populares en la cuantificación de riesgo de crédito destacan:

- (1) Es un modelo semi paramétrico, dado que no necesita especificar la forma funcional del tiempo de supervivencia o de falla, por lo tanto los resultados serán consistentes y robustos y

- (2) Permite incorporar variables dependientes de tiempo, que luego pueden mejorar el poder explicativo y la predicción.

El estudio, a través de la verificación del supuesto de riesgos proporcionales y la significancia individual de las variables analizadas, concluye la importancia de la inclusión de variables dependientes del tiempo y encuentra que para la muestra analizada de medianas y grandes empresas en el Perú, las variables más importantes en explicar el incumplimiento de pagos son: el sector económico, tamaño de la empresa, saldo inicial, porcentaje de deuda en Moneda extranjera, porcentaje de garantías e índice de confianza empresarial; esta última como única variable estadísticamente significativa *proxy* del entorno macroeconómico y dependiente del tiempo.

En el siguiente cuadro, extraído de Dirick et al. (2016), se muestra un *benchmark* de los diferentes estudios relacionados a la aplicación de modelos de supervivencia en la cuantificación del riesgo de crédito:

TABLA 1: Resumen de investigaciones relacionadas con métodos de supervivencia

Investigación	aFT/ Paramétrico	COX PH	AFT(*)/ COX PH + Extensiones	No par am étrico	Tipo de Exposición evaluada	Tamaño de Muestra	# variables input	Criterios de evaluación
Narain(1992)	X				Compañías indias listadas	1242	7	Ninguno
Basinek et al(1999)	X	X	X			50000	>7	Curva ROC
Stepanova & Thomas(2002)		X	X		Préstamos de consumo	50000	16	Curva ROC
Belloti & Crook(2009)			X		Tarjetas de crédito	200 000	>11	Costo de un caso mal clasificado
Cao et al(2012)	X	X		X	Créditos personales de consumo	25 000	1	Curva ROC

Tong et al(2012)		X			Créditos de Co ns u m o	27 527	14	Curva ROC, KS
Zhang & Thomas (2012)	X	X	X		Comparaci ón de LGD de Créditos personales	27 000	21	Error en la predicción del tiempo al default

(*) AFT: Accelerated Failure Time Models: son modelos de supervivencia totalmente paramétricos en los que las variables explicativas actúan como factores de aceleración para acelerar o ralentizar el proceso de supervivencia en comparación con la función de supervivencia inicial.

2.2 Principales diferenciadores de la presente investigación respecto a las investigaciones previas revisadas

En nuestro entendimiento, en cada uno de las investigaciones previas revisadas, no se ha reparado en considerar que una de las principales ventajas del modelo de regresión de Cox, es la de realizar una estimación del riesgo de incumplimiento en cada momento temporal de la vida de las operaciones de crédito en estudio.

Esta característica fundamental, en conjugación con las metodologías de estimación de perdidas esperadas bajo las nuevas normas de valorización de instrumentos financieros - IFRS9, podrían permitir realizar estimaciones más precisas de las pérdidas potenciales de dichas operaciones al considerar el perfil de amortización en el tiempo de las operaciones de crédito.

De esta forma, todas las investigaciones previas revisadas basan su análisis comparativo en indicadores estándar de capacidad de predicción como GINI o KS. La presente investigación busca medir si el modelo de riesgos proporcionales de Cox permite una estimación más exacta de las pérdidas efectivamente materializadas a través de la conjugación de:

- (1) La consideración del perfil de amortización de las operaciones de crédito
- (2) La estimación de probabilidades de incumplimiento en cada momento temporal de las operaciones característica exclusiva del modelo de Cox,

2.3 Teorías Básicas

2.3.1 Conceptos básicos de riesgo de crédito

a) Riesgo de Crédito y probabilidad de incumplimiento

El concepto de riesgo de crédito es central en la investigación y éste se define, según el comité de Basilea para la supervisión bancaria, como la probabilidad de que una contraparte crediticia no se encuentre en la capacidad de cumplir con sus obligaciones en el tiempo y/o forma pactados inicialmente.

El riesgo de crédito es el principal riesgo al que los intermediarios financieros tradicionales están expuestos y para su medición utilizan su medida más básica, la probabilidad de incumplimiento, como una medida de la probabilidad de que una contraparte incumpla con sus obligaciones en un tiempo determinado y tomando en consideración que debe existir una definición previa del mismo evento de incumplimiento.

b) Pérdidas dado el incumplimiento (*Loss Given Default*)

La pérdida dado el incumplimiento es otro parámetro necesario para el cálculo de la pérdida esperada y se define como la porción de la exposición al momento del incumplimiento que no se espera recuperar después de una ventana de tiempo definida para la aplicación de acciones de recuperación.

Estudios académicos proveen 3 tipos de estimación de la LGD :

- (1) LGD de mercado (*Market LGD*): Observado de precios de mercado de bonos en incumplimiento de pagos. Para bonos listados en mercados centralizados, las agencias de clasificación de riesgo reportan el precio de mercado de bonos en incumplimiento de pagos un mes posterior al default.
- (2) LGD de mercado implícito (*Implied market LGD*): LGD estimado de precios de bonos de alto riesgo (no en incumplimiento) usando modelos teóricos de valuación de activos
- (3) LGD estructural (*Workout LGD*): EL LGD observado después de un proceso de recuperación , incluye todos los costos de este proceso, así como la actualización de flujos de recuperación a una tasa de interés que suele ser la misma de la exposición en default

El *workout LGD* es el método *ad hoc* en cuanto a préstamos bancarios se refiere.

Según como indica Shuermann(2002), para la determinación del *workout LGD*, las entidades financieras deben determinar exactamente la fecha que las exposiciones cayeron en incumplimiento, los saldos de deuda al momento del incumplimiento y finalmente los saldos recuperados, sus fechas de recuperación junto con los costos en que incurrió durante el proceso.

Todos estos flujos deberán ser descontados a una tasa de interés adecuadamente definida. Lo mencionado se puede resumir en la siguiente forma de estimación:

$$LGD = 1 - RR$$

Donde $RR = Recovery\ Rate$ y está definido por:

$$RR = \frac{\sum_{i=1}^n FC_i - CR}{(1 + i)^n}$$

Donde n es la cantidad de períodos transcurridos desde la primera vez en la que el crédito/exposición se identificó en estado de incumplimiento

FC_i: Los flujos de recuperación efectiva en cada uno de los períodos

CR: Los costos de recuperación incurridos en el proceso de recuperación

i : La tasa de interés a la cual se actualizan los flujos y costos de recuperación, puede aplicarse la misma tasa de interés activa de la operación en incumplimiento como una medida del costo de oportunidad de los flujos no recuperados a tiempo.

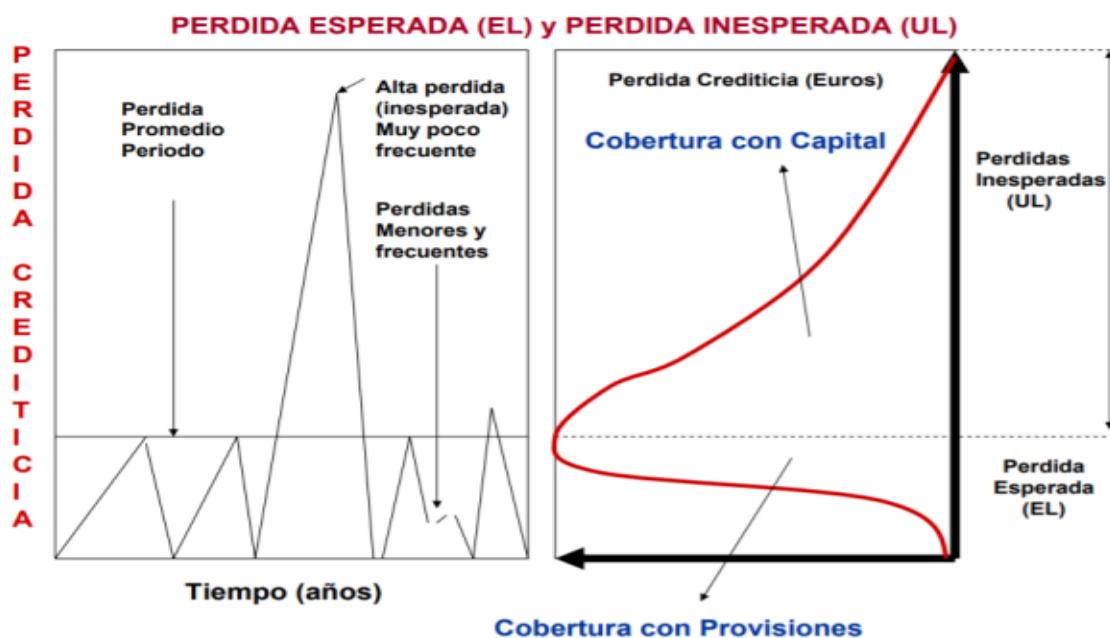
c) Pérdida esperada y no esperada por riesgo de crédito

Las entidades financieras están expuestas a pérdidas esperadas y pérdidas no esperadas por sus actividades de intermediación financiera. De ésta forma, las pérdidas esperadas son las pérdidas en las que securrirá “en promedio” por el otorgamiento de financiamiento, la cual es cargada en los estados de resultados de las entidades financieras en el rubro conocido como provisiones por incobrabilidad de créditos; dado esto último, se considera a éste concepto como un costo más dentro del negocio de intermediación que llevan a cabo las entidades financieras.

Por otro lado, las pérdidas no esperadas representan la desviación alrededor de la media (representada por la pérdida esperada) y es estimado de acuerdo a un nivel de confianza definido que indica que la pérdida incurrida no podrá ser mayor en un nivel α de ocurrencias de cada 100 (nivel de confianza definido). Dicha pérdida es cubierta con requerimientos de capital hacia las entidades financieras.

Las entidades financieras planifican su capital en función del grado de calificación deseado, de esta forma un nivel de capital que permita absorber las pérdidas por riesgo de crédito en 99 de cada 100 veces correspondería a una entidad financiera con clasificación de riesgo AAA.

GRÁFICO 1: Pérdidas esperadas y no esperadas



Fuente y elaboración: Superintendencia de Banca, Seguros y AFP

d) Provisiones para el incumplimiento de créditos

Las entidades financieras “reservan” parte de sus márgenes financieros con el fin de cobertura a las potenciales pérdidas por riesgo de crédito a las que se encuentran expuestas por los préstamos que otorgan. En la práctica y aplicado a la contabilidad de las entidades financieras, se trata de una deducción de los ingresos de las entidades financieras que reducen directamente la utilidad generada por las mismas.

Las reglas bajo las cuales se calculan las provisiones por riesgo de crédito, son brindadas por el ente regulador del sistema financiero en cada país y en el caso peruano por la Superintendencia de Banca y Seguros a través del reglamento de clasificación del deudor y constitución de provisiones, resolución 11356-2008, que establece las reglas para la constitución de éstas provisiones, principalmente en función de los días de atraso de los deudores.

Por tanto, se puede decir que, si bien el fin de las provisiones es establecer una cobertura razonable ante pérdidas potenciales, su forma de cálculo para el caso peruano es de carácter reactivo ante el evento de morosidad de los deudores y no prospectivo pues no se basa en cálculo de probabilidades de incumplimiento por deudor. Siendo que, en otros países como España, se ha podido establecer marcos normativos para la estimación de provisiones basados en calificaciones internas

e) Capital económico para riesgo de crédito

Es la reserva de capital que deben realizar las entidades financieras para cubrir las pérdidas no esperadas producto de las exposiciones afectas a riesgo de crédito, con un nivel de confianza (α) previamente determinado.

El nivel de confianza define la proporción de veces de cada 100 ocurrencias en las cuales se espera que las pérdidas por riesgo de crédito no superarán el umbral asociado. Su determinación está en función de la calificación de solvencia a la cual aspira la entidad financiera. De esta forma, a mayor nivel de confianza, se espera un mayor nivel de pérdidas potenciales y por tanto un mayor nivel de capital económico requerido y en última instancia un nivel de calificación de riesgo sobre la entidad de mayor calidad.

Sus formas de cálculo pueden ser diversas, desde un método paramétrico simple hasta métodos avanzados como Credit Risk+™ (Credit Suisse, 1997), Credit Metrics™ (JP Morgan, 1997), entre otros. El detalle de estas metodologías se encuentra fuera del alcance de revisión de esta investigación, debido a la que la estimación de capital por riesgo de crédito no es parte de los objetivos de la misma.

f) Capital regulatorio para riesgo de crédito

Es el nivel de capital que las entidades deben reservar bajo las normas establecidas por el regulador del sistema financiero de cada país, en el caso peruano, la Superintendencia de Banca, Seguros y AFP regula este requerimiento a través del reglamento para el Requerimiento de Patrimonio Efectivo por riesgo de crédito, resolución 14354-2009.

Los requerimientos de capital brindados por el regulador peruano, si bien tropicalizados y adecuados a las particularidades del sistema financiero peruano, toman como base las recomendaciones brindadas por el comité de Basilea y sus acuerdos de capital Basilea II y Basilea III.

Dentro de esta norma, existen tres métodos de cálculo de capital regulatorio bajo la norma mencionada:

- **Método Estándar:** Basado en ponderaciones predefinidas por el regulador, aplica ponderaciones a las carteras de créditos sin un criterio de riesgo prospectivo y no basado en cálculos de probabilidades de incumplimiento. Las entidades reservarán más capital por aquellas operaciones que el regulador haya definido con una sobre ponderación respecto a otras.

- **Método básico de clasificaciones internas:** En el cual el componente de PD es calculado por la entidad regulada y el componente LGD es definido por el regulador, según el tipo de exposición. Así mismo este método está solo permitido para exposiciones no minoristas

- **Método avanzado de clasificaciones internas:** En el cual el componente de PD y LGD es calculado por la entidad regulada. Este método es aplicable a todo tipo de exposiciones y permite una mucha mayor granularidad y sensibilidad al riesgo en la reserva de capital por las exposiciones afectas a riesgo de crédito en el balance de las entidades financieras.

La fórmula para el cálculo de capital regulatorio para riesgo de crédito bajo métodos de clasificaciones internas viene dado por:

$$K = LGD x \left[N \left[1 - \rho^{-0.5} x N^{-1}(PD) + \left(\frac{\rho}{1-\rho} \right)^{0.5} X N^{-1}(0.999) \right] - PD \right] X \left(\frac{1 + (M - 2.5)x b}{1 - 1.5x b} \right) \quad (1)$$

Donde:

LGD: Loss Given Default

PD: Probability of Default

M: Madurez

ρ : Correlación de default entre activos que viene dada por:

$$\rho = 0.12 x \frac{(1 - e^{-50xPD})}{(1 - e^{-50})} + 0.24 X \left[1 - \frac{e^{(-50xPD)}}{1 - e^{-50}} \right]$$

A pesar de que varias entidades financieras en el país, principalmente bancos, tienen los recursos tecnológicos, estructura y el conocimiento técnico para la reserva de capital por alguno de los métodos basados en clasificaciones internas, a la fecha de culminación de esta investigación, no existe una entidad que haya optado por estos métodos.

Las razones por la cuales la migración no se ha dado en ningún caso, pueden ser diversas, a continuación, se ensayan algunas de éstas:

- Los requerimientos de capital bajo métodos basados en calificaciones internas estarían sujetos a un mayor grado de incertidumbre para los accionistas de lo que sería bajo métodos básicos, por tanto, la planificación de capital sería un ejercicio más complejo y de mayor incertidumbre
- Los ejercicios de impacto como los llevados a cabo por Superintendencia (Javier Poggi, 2006), arrojaban requerimientos de capital incluso mayores en métodos por clasificaciones internas versus métodos básicos.

De esta forma, si bien varias entidades del sistema financiero peruano, cuentan con las capacidades para una migración a métodos de clasificaciones internas, utilizan los modelos subyacentes principalmente para gestión y seguimiento interno.

g) Normas Internacionales de Contabilidad NIIF⁹

Las nuevas normas internacionales para la valorización y reconocimiento de instrumentos financieros que entró en vigor el primero de enero del 2018, reemplazan el anterior estándar conocido como el IAS 39. Los principales impactos de esta nueva norma vienen dados por:

- (1) Mayor volatilidad de la cuenta de resultados.** La NIIF 9 aumenta el riesgo de que se tengan que medir más activos a valor razonable con cambios en el valor razonable reconocidos en resultados a medida que surgen.
- (2) Reconocimiento anticipado de pérdidas por deterioro de cuentas por cobrar y préstamos, incluidas las cuentas por cobrar comerciales.** Las entidades tendrán que comenzar a prever posibles pérdidas crediticias futuras en el primer período de presentación de informes en el que un préstamo se registra en los libros, incluso si es muy probable que el activo sea totalmente cobrable.

(3) Importantes nuevos requisitos de divulgación: los más afectados pueden necesitar nuevos sistemas y procesos para recopilar los datos necesarios.

Dada la naturaleza del presente estudio, nos enfocamos en el segundo impacto identificado, relacionado con un reconocimiento más temprano de las pérdidas por deterioro de cuentas por cobrar y préstamos. IFRS9, adopta el enfoque de “pérdidas esperadas” como la suma ponderada de la probabilidad de ocurrencia de default en cada uno de los períodos del préstamo multiplicados por el flujo de efectivo esperado para dicho período.

Bajo el modelo de “pérdida crediticia esperada”, una entidad calcula la provisión para pérdidas crediticias considerando sobre una base de descuento los déficits de efectivo en los que incurría en varios escenarios de incumplimiento para períodos futuros prescritos y multiplicando los déficits por la probabilidad de que ocurra cada escenario. La asignación es la suma de estos resultados ponderados por probabilidad. Debido a que cada préstamo y cuenta por cobrar conlleva algún riesgo de incumplimiento, cada uno de esos activos tiene una pérdida esperada asociada, desde el momento de su originación o adquisición. Las pérdidas crediticias esperadas representan posibles resultados ponderados por la probabilidad de que ocurran, representan medidas del riesgo crediticio de un activo.

IFRS 9 establece no uno, sino tres enfoques separados para medir y reconocer las pérdidas crediticias esperadas:

- Un enfoque general que se aplica a todos los préstamos y cuentas por cobrar que no son elegibles para los otros enfoques;
- Un enfoque simplificado que se requiere para ciertas cuentas por cobrar comerciales y los denominados “activos contractuales según la NIIF 15” y, por lo demás, es opcional para estos activos y cuentas por cobrar por arrendamientos.
- Un “enfoque ajustado al crédito” que se aplica a los préstamos que tienen deterioro crediticio en el reconocimiento inicial (por ejemplo, préstamos adquiridos con un gran descuento debido a su riesgo crediticio).

Un factor distintivo entre los enfoques es si la provisión para pérdidas crediticias esperadas en cualquier fecha del balance se calcula considerando posibles incumplimientos solo para los próximos 12 meses (“ECL de 12 meses”), o para toda la vida restante del activo (“ECL de por vida”). Para aquellas entidades que

solo tienen cuentas por cobrar a corto plazo de menos de un año de duración, el enfoque simplificado y general probablemente tendría poca diferencia práctica.

En todos los casos, la provisión y cualquier cambio en ella se reconocen reconociendo las ganancias y pérdidas por deterioro en resultados.

Así mismo, diferencia el cálculo en etapas según el nivel de riesgo de las operaciones clasificadas:

STAGE 1: Operaciones con un riesgo identificado “bajo”, sobre las cuales el cálculo de la pérdida esperada se realiza sobre los siguientes 12 meses en el ciclo de vida de las operaciones de crédito.

STAGE 2: Operaciones con un riesgo identificado “medio”, sobre las cuales el cálculo de la pérdida esperadas se realiza sobre todo el ciclo de vida restante de las operaciones de crédito

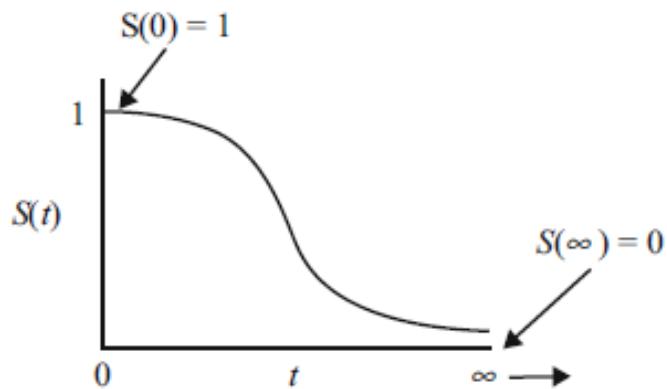
STAGE 3: Operaciones de crédito identificadas como “default”, sobre las cuales el cálculo de la pérdida esperada es solo la multiplicación de la exposición por el porcentaje pérdida dado el incumplimiento(LGD)

2.3.2 Conceptos teóricos del análisis de supervivencia

2.3.2.1 Forma teórica de la función de supervivencia

La metodología aplicada se basa en la teoría de modelos de riesgo proporcionales de Cox, el cual realiza estimaciones semiparamétricas en la línea base de riesgo $h_0(t)$ debido a que esta función no asume ninguna distribución de probabilidad, a diferencia de la expresión $e^{\sum_{i=1}^p B_i X_i}$ la cual asume una función exponencial con el fin de obtener resultados no negativos.

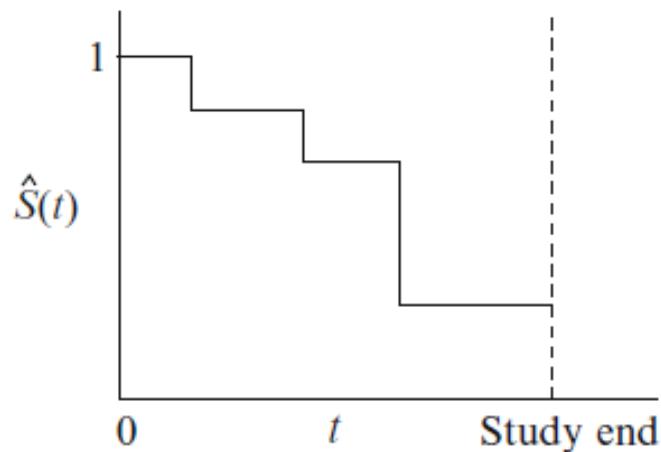
GRÁFICO 2: Funciones de supervivencia teóricas



- No crecientes
- Al momento $t = 0$, $S(t) = S(0) = 1$; esto es al inicio del estudio, dado que aun no se ha dado ningun evento, la probabilidad de sobrevivir en el momento 0 es 100%
- Siempre que $t = \infty$, $S(t) = S(\infty) = 0$, que indica que, teóricamente, si es que el tiempo tiende al infinito, la probabilidad de supervivencia es cero.

2.3.2.2 Forma empírica de la función de supervivencia

GRÁFICO 3: Funciones de supervivencia empíricas



La forma funcional empírica muestra una forma estratificada y no suavizada como en las funciones teóricas, y ello debido a factores como: (1) El período de estudio

es finito, (2) Pueden existir riesgos competidores, (3) Es posible que no todos los sujetos de estudio lleguen en algún momento a presentar el evento (observaciones censuradas).

2.3.2.3 Función de supervivencia no paramétrica de Kaplan Meier

En la presente investigación, no se conoce la distribución de la probabilidad del tiempo hasta la ocurrencia del evento, por tanto, se utiliza el estimado producto límite de Kaplan -Meier, siendo éste ampliamente utilizado en otras investigaciones.

El estimador no paramétrico de Kaplan – Meier propuesto por Kaplan – Meier (Colosimo y Giolo, 2010), para estimar la función de supervivencia, es una adaptación de la función de distribución empírica, que, en ausencia de censuras, es definida como el cociente entre el número de observaciones que no fallaron hasta el tiempo “t” y el número total de observaciones del estudio. El estimador se define como:

$$\hat{S}(t) = \sum_{j:t_j < t} \left(\frac{n_j - d_j}{n_j} \right)$$

Donde:

d_j = Número de ocurrencias del evento en t

n_j = Número de individuos en riesgo en el tiempo t ; es decir aquellos que no presentaron el evento y no fueron censurados hasta el instante inmediatamente anterior a t_j

2.3.2.4 La función de riesgo

$$h(t) = \lim_{\Delta t \sim 0} \frac{P(t < T < t + \Delta t | T \geq t)}{\Delta t} \quad (2)$$

En términos empíricos, la función de riesgo, estima el riesgo instantáneo por unidad de tiempo de que el evento ocurra, dado que el individuo ha sobrevivido más allá del momento t .

2.3.2.5 Datos Censurados

Una de las razones de la popularidad de los métodos de supervivencia es su capacidad para lidiar con datos censurados, fenómeno muy común en los estudios socio demográficos, económicos y epidemiológicos.

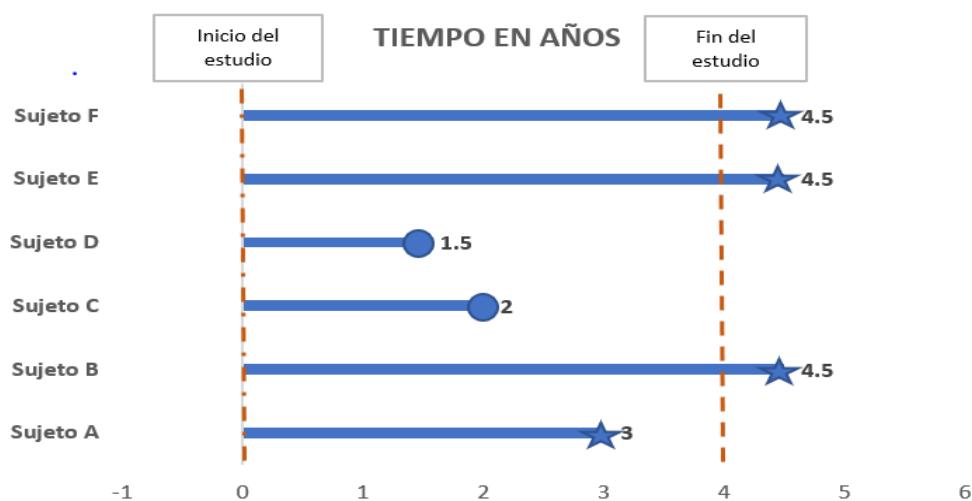
La censura ocurre cuando no se conoce el tiempo exacto de supervivencia del individuo analizado. En el caso de data de créditos, la censura puede ocurrir por dos motivos:

1. El individuo sujeto de estudio prepaga su deuda antes de finalizada la ventana temporal de estudio ó
2. El individuo sujeto de estudio culmina el estudio sin que haya presentado el evento de interés, en este caso no haya caído en incumplimiento de pagos

El tipo de censura que acabamos de mencionar es censura derecha y es el tipo de censura más común en los análisis de supervivencia.

A continuación, se muestra un gráfico que resume los ejemplos de censura en la data de análisis

GRÁFICO 4: Censura derecha de datos



Los sujetos B, C, D, E y F en el gráfico 4 muestran censura derecha, dado que concluyen el período de estudio sin haber experimentado el evento de interés ó dejan de ser observables antes de que el período de observación concluya.

2.4 Marco Conceptual

Uno de los objetivos específicos de la tesis es la comparación de dos metodologías para el modelamiento de la Probabilidad de Incumplimiento: modelos logit y modelos de supervivencia - Regresión de Cox

Considerando que los modelos de regresión logística ya han sido ampliamente estudiados y explicados, se enfatizará en las propiedades teóricas del modelo de Cox.

2.4.1 Modelos de supervivencia – Regresión de Cox

Siendo T una variable aleatoria que representa el tiempo hasta la materialización del evento de interés. La forma de describir la distribución de T es a través de la función de riesgo, *hazard function* en inglés, definida como:

$$h(t) = \lim_{\delta \rightarrow 0} \left\{ \frac{P(t \leq T < t + \delta | T \geq t)}{\delta t} \right\} \quad (3)$$

Esta es la probabilidad que el evento de interés sea observado en el individuo de estudio al momento t condicionado a que el individuo se mantenga activo dentro del estudio hasta ese momento.

Asumamos ahora que se cuenta con diversas métricas y características del individuo al momento del inicio del estudio, dichas covariables estarían agrupadas en el vector x y se busca evaluar la relación entre estas covariables y la distribución del tiempo al evento de interés. Para ello Cox(1972) propuso el siguiente modelo:

$$h(t; x) = h_0(t) e^{(x\beta)} \quad (4)$$

Donde:

$h_0(t)$ = La línea base de riesgo, m depende exclusivamente del tiempo

$e^{\sum_{i=1}^p \beta_i X_i}$ = Es la expresión que envuelve al vector de covariables X_i y que son independientes del tiempo

Al modelo propuesto por Cox, se le conoce como de riesgos proporcionales debido al supuesto de que el riesgo del individuo con covariables x es proporcional a una línea de riesgo base desconocida $h_0(t)$

Así mismo, Cox (1972) demostró que β puede ser estimado sin tener ningún conocimiento de $h_0(t)$ a través de la maximización de la función de verosimilitud:

$$L(\beta) = \prod_{i=1}^k \frac{\exp(x_i\beta)}{\sum_{l \in R(t(i))} \exp(X_l\beta)} \quad (5)$$

Los métodos de riesgos proporcionales asumen que las funciones de riesgo son continuas, sin embargo, los eventos de interés, como el default en las operaciones de crédito se almacenan en las bases de datos de las entidades financieras, típicamente de forma mensual, por tanto, muchos eventos son almacenados en un mismo período t . La función de probabilidad exacta tiene que incluir todos los posibles ordenamientos de fallos vinculados (Kalbfleisch y Prentice 1980), y por lo tanto es muy difícil su estimación debido a aspectos computacionales.

Cox (1972) desarrolló una aproximación a la ecuación (4) a través de un modelo logístico discreto:

$$\frac{h(t; x)}{1 - h(t; x)} = e^{(x\beta)} \frac{h_0(t)}{1 - h_0(t)} \quad (6)$$

Donde

$$h(t; x) = P(t \leq T \leq t + 1 | T \geq t) \quad (7)$$

Para demostrar que la ecuación (6) se reduce a la ecuación (4) cuando el tiempo es continuo, notar que la forma de función de riesgo discreta sería:

$$h(t; x)\delta t = P(t \leq T \leq t + \delta t | T \geq t) \quad (8)$$

La ecuación 6 se convierte en:

$$\frac{h(t; x)\delta t}{1 - h(t; x)\delta t} = e^{(x\beta)} \frac{h_0(t)\delta t}{1 - h_0(t)\delta t} \quad (9)$$

Y tomando el límite mientras $\lim_{\delta t \rightarrow 0}$ nos da como resultado la ecuación (4).

2.4.2 Supuestos del modelo de regresión de cox

a) Respecto a la presencia de data censurada

Los modelos analíticos basados en métodos de supervivencia como la regresión de Cox, basan sus estimaciones en un supuesto importante de independencia de los datos censurados; que postula que, dentro de cualquier sub grupo de interés, los sujetos de estudio que muestran censura dentro del período de estudio son representativos de todos los sujetos dentro de ése sub grupo que aún se mantienen en riesgo al momento “t” respecto a su experiencia de supervivencia.

b) Respecto a la línea base de riesgo

Una de las razones más importantes de la popularidad del modelo de regresión de Cox es que la función base de riesgo $h_0(t)$ no asume ninguna función de probabilidad específica; por tanto, si el modelo generador de datos real sigue una distribución Weibull, el modelo de Cox seguirá aproximadamente una distribución Weibull. De forma similar sería si el modelo real sigue una función Exponencial. De esta forma se logran estimadores robustos y consistentes.

El modelo arroja estimados razonablemente buenos de coeficientes de regresión, razones de riesgo y ajustes de curvas de supervivencia pueden ser obtenidos de una amplia variedad de situaciones de data.

c) Respecto a los riesgos proporcionales

El modelo de Cox asume que el vector de covariables $x\beta$ mantiene un impacto constante a través de la curva de supervivencia (línea base de riesgo $h_0(t)$), el cumplimiento de éste supuesto puede ser validado a través de tres métodos: (1) Método Gráfico, (2) Procedimientos de pruebas bondad de ajuste y (3) Un procedimiento que involucra el uso de variables dependientes del tiempo.

El modelo de riesgos proporcionales de Cox asume que el ratio de riesgo h_0 es constante, incluso si se comparan dos vectores de predictores X.

Equivalentemente, esto significa que el riesgo para un individuo es proporcional al riesgo de otro individuo, donde la proporcionalidad constante es independiente del tiempo, de la siguiente forma:

$$\frac{\hat{h}(t, X^*)}{\hat{h}(t, X)} = \hat{\theta}, \text{ constante a través de } t \quad (8)$$

Por tanto: $\hat{h}(t, X^*) = \hat{\theta} \hat{h}(t, X)$

c.1. Análisis gráfico

Denominado método “log-log”, se trata del método más popular para la evaluación del supuesto de riesgos proporcionales.

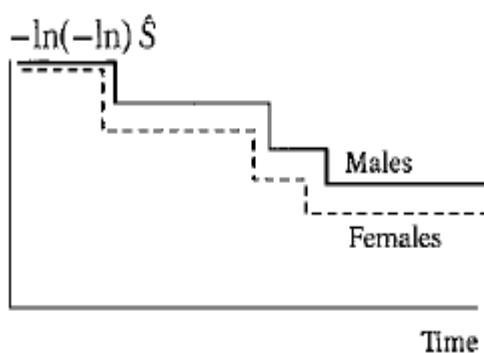
La curva log- log de supervivencia resulta de aplicar el logaritmo natural a la función de supervivencia dos veces. Matemáticamente se escribe una curva log-log como

$$-\ln(-\ln \hat{S})$$

Notar que el logaritmo de una probabilidad como \hat{S} es siempre un número negativo, debido a que solo podemos tomar logaritmos de valores no negativos, debemos negativizar antes de tomar el segundo logaritmo. El valor de $-\ln(-\ln \hat{S})$ puede tomar valores positivos o negativos, siendo cualquiera de los dos aceptables, porque no estamos tomando un tercer logaritmo.

Si se cumple la hipótesis de riesgos proporcionales estos gráficos deben ser visualmente paralelos para las diferentes clases de las covariables del modelo por separado (Boj del Val, 2017).

GRÁFICO 5: Representación de un gráfico Log – Log para evaluación de supuesto de riesgos proporcionales



Fuente: *Survival Analysis. A self learning text book(2012)*

c.2. Análisis de los residuos de Schoenfeld

Los residuos de Schoenfeld se calculan para cada predictor y para cada individuo. Si por ejemplo tenemos dos predictores en el modelo, para cada sujeto que presente el evento de interés tendremos dos residuos de Schoenfeld, los cuales se definen de la siguiente manera:

$$R_i = \delta_i \left(\frac{\sum_{l \in R(t(i))} X_{lj} \exp(\hat{\beta} X'_l)}{\sum_{l \in R(t(i))} \exp(\hat{\beta} X'_l)} \right)$$

Para $i = 1, 2, \dots, n$, y $j = 1, \dots, p$, siendo $\delta_i = 1$ si el sujeto no está censurado y $\delta_i = 1$ si el sujeto está censurado. Por tanto, los residuos de Schoenfeld se calculan solo para individuos no censurados.

Para cada predictor en particular, se cumplirá la hipótesis de proporcionalidad de los riesgos si es que sus residuos de Schoenfeld no se encuentran correlacionados con el tiempo.

El procedimiento de cálculo es el siguiente: Estimar los residuos de Schoenfeld del modelo de Cox en estudio; ordenar los tiempos de fallo etiquetando el orden a cada tiempo; calcular la correlación entre los residuos y la variable de orden creada. Y realizamos el contraste de si para cada covariable por separado. En el caso de aceptar la hipótesis nula de que la correlación es cero, se cumplirá la hipótesis de riesgos proporcionales para el predictor correspondiente. Con lo que, tal y como ya habíamos comentado, nos interesará que los p-valores del contraste sean elevados para aceptar la hipótesis de proporcionalidad (por encima de 0.05).

d) Respecto a la linealidad de los coeficientes

Usualmente se asume que las covariables de carácter continuo mantienen una forma lineal, sin embargo, éste supuesto debe ser verificado. Para ello se utiliza el análisis de los residuos de Martingala definidos como:

$$r_{Mi} = \delta_i - r_{Ci}$$

Puede ser interpretado como la diferencia entre el número de fallas observada para un individuo en el intervalo $(0, t_i)$ denotado por δ_i y el número esperado de fallas acorde con el modelo r_{Ci} . r_{Mi} puede ser graficado versus el tiempo a la ocurrencia

del evento de interés. Idealmente no debería mostrar ningún patrón para verificar la idoneidad del modelo (Stepanova & Cox, 1999)

Es importante señalar que este análisis es adecuado cuando se tratan de variables continuas, en el caso de covariables discretas, la verificación del supuesto de linealidad de la forma funcional de los coeficientes no es necesaria.

2.4.3 Modelo de regresión de Cox con variables dependientes del Tiempo

A pesar de que, esta extensión del modelo de Cox no es parte de la presente investigación, se presenta un breve resumen del mismo.

El supuesto de riesgos proporcionales explicado en la sección anterior implica que las covariables del modelo de Cox sólo son medidas al inicio del estudio. Sin embargo, podrían existir variables que pueden tomar valores distintos entre el inicio del estudio, el desarrollo y finalización del mismo. Por tanto si es que alguna de las variables o el modelo en su conjunto no cumple con el supuesto de riesgos proporcionales, es una opción utilizar el modelo de regresión de Cox estratificado.

Para lidiar con esta complicación, el modelo de regresión de Cox que considera covariables que pueden variar en el tiempo está definido por:

$$h(t/x(t)) = h_0(t)e^{[X(t)\beta]} \quad (10)$$

Su estimación se da a través del logaritmo de la función de verosimilitud parcial, siendo estos asintóticamente normales bajo condiciones regulares. De esta forma, se pueden utilizar las estadísticas de Wald y la de la razón de verosimilitud para realizar inferencia sobre los parámetros del modelo de regresión de Cox con covariables dependientes del tiempo (Uriarte y Agüero, 2017)

La utilización del modelo de Cox extendido no ha sido considerada dentro de ésta investigación debido a: (i) La estructura de la base de datos evaluada y (ii) Los resultados de las pruebas de riesgos proporcionales no arrojaron la necesidad de aplicarlo.

2.4.4 Modelo de regresión logística

El modelo de regresión logística se estima a través de la función logística definida de la siguiente forma:

$$P(y = 1|X) = \frac{1}{1 + exp^{-x\hat{B}}}$$

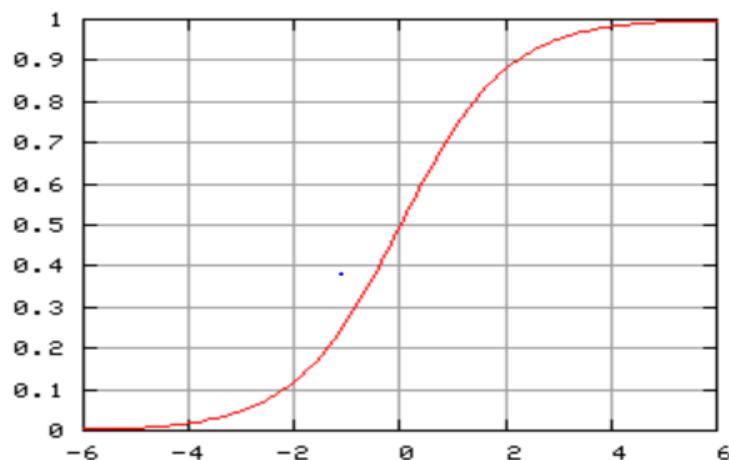
En la cual se busca estimar la probabilidad de ocurrencia del evento de interés ($y=1$) dado un vector de covariables X .

Los estimadores \hat{B} se optimizan a través de la maximización de logaritmo de verosimilitud definido por:

$$L(\alpha, \beta|Y, X) = \prod_{i=1}^n P(Y = y_i|X_i)f(x_i) \propto \prod_{i=1}^n P(Y = y_i|X_i)$$

La popularidad del método, ampliamente utilizado por las entidades financieras, radica en su interpretación directa, simplicidad de implementación, así como bondades de la distribución de probabilidad logística acumulativa con una forma sigmoidal graficada como:

GRÁFICO 6: Función logística normalizada



2.4.5 Estimación de pérdidas esperadas por riesgo de crédito

Bajo estimación por regresión logística

Bajo el enfoque de Internal Rating Based (IRB) de Basilea II, los parámetros para la estimación de las pérdidas esperadas requieren de la definición de los tres parámetros de riesgo: PD, LGD y EAD

PD: Probabilidad de Default, calibrada a 12 meses como la frecuencia de incumplimiento en buckets de riesgo homogéneos

LGD: Loss Given Default, para fines de la presente investigación se asumirá un valor fijo de 100%

EAD: La exposición **monetaria**, en el caso de la presente investigación se asume como el monto desembolsado

De esta forma, la estimación de perdidas esperadas es la combinación lineal de los tres parámetros mencionados

$$PE = PD \times LGD \times EAD \quad (11)$$

Nótese que la combinación lineal de los tres parámetros implica la ausencia de correlación entre los mismos, sin embargo, existen diversos estudios que podrían demostrar la correlación positiva entre estos dos parámetros, exasperando los niveles de pérdida estimados de forma lineal a través de (11). Esto se daría a través de la presencia de una variable latente que asociaría a ambos parámetros como, por ejemplo, algunas relacionadas al ambiente económico general

Bajo estimación por métodos de supervivencia – regresión de cox

Bajo las normas de IFRS9, la estimación de la pérdida esperada se diferencia según el grado de deterioro del portafolio en tres *stages* de riesgo

STAGE 1: Con atrasos menores a 30 días

Se realiza una estimación de pérdida esperada a 12 meses, con una estimación de la PD en cada punto temporal y de acuerdo al perfil de amortización de la cartera, trayendo a valor presente los flujos de pérdida esperada a la tasa de interés de la propia operación.

$$PE = \sum_{t=1}^{12} \frac{PD_t \times LGD \times EAD_t}{(1 + TiE)^t} \quad (12)$$

STAGE 2: Con atrasos entre 31 y 90 días

Se realiza una estimación de pérdida esperada “*lifetime*”, con una estimación de la PD en cada punto temporal y de acuerdo al perfil de amortización de la cartera, trayendo a valor presente los flujos de pérdida esperada a la tasa de interés de la propia operación.

$$PE = \sum_{t=k}^{LT-k} \frac{PD_t \times LGD_t \times EAD_t}{(1 + TiE)^t} \quad (13)$$

STAGE 3: Con atrasos mayores a 90 días

Se realiza la mejor estimación posible de la pérdida esperada, considerando que son operaciones en default

$$PE = ELBE(t_{def}) \times Saldo_{t_{def}} \quad (14)$$

Donde:

PD_t es la probabilidad de hacer default entre t_{Ref} y t_{Ref+12} o entre t y t+12 en el cálculo lifetime

LGD_t es la pérdida dado el incumplimiento en los próximos 12 meses o entre t y t+12 en el cálculo lifetime

EAD_t es el saldo con el que incumple en los próximos 12 meses o entre t y t+12 en el cálculo lifetime

ELBE (Expected Loss Best Estimate) Indica el % del saldo actual (no EAD) que se perderá en función del tiempo que la cuenta lleva en default

TiE es la tasa de interés efectiva de la operación, teniendo en cuenta todos los costos financieros.

Por definición la cartera analizada en el presente estudio, al tratarse de nuevos desembolsos tendrá un tratamiento tipo Stage 1.

CAPITULO III: DIAGNOSTICO

3.1 Población, muestras de desarrollo y validación

3.1.1 Definición de la población y la muestra:

Personas Naturales atendidas con operaciones de crédito tipo consumo no revolvente entre los períodos Septiembre 2018 y Febrero 2019.

El modelo se desarrolló con una partición de la población en dos submuestras, una de desarrollo y otra de validación, con la siguiente representatividad:

TABLA 2: Distribución de las muestras de desarrollo y validación

PERIODO DE DESEMBOLO	REGISTROS	REPRESENTATIVIDAD
2018_09	24 306	17%
2018_10	25 291	18%
2018_11	21 763	15%
2018_12	22 040	15%
2019_01	25 283	18%
2019_02	24 238	17%
TOTALES	142 921	100%

Para fines del desarrollo, tanto de la metodología logística como la de regresión de Cox, se tomará de forma aleatoria una muestra del 70% para el desarrollo y 30% para el proceso de validación.

a. Exclusiones de la base de datos

Registros con comportamiento insuficiente

Se excluyeron sujetos de estudio que se consideraron “Insuficientes” en la observación de su comportamiento, en otras palabras, que mostraron una censura en etapas muy tempranas de su desarrollo. Esta censura temprana implica que el sujeto de estudio dejó de ser observable en los primeros dos meses en la ventana de comportamiento definida.

Sujetos de estudio retirados: 17 200

Registros con comportamiento indeterminado

Se excluyeron sujetos de estudio que se consideraron “Indeterminados” en su comportamiento, debido a que alcanzaron alturas de morosidad que se acercan al comportamiento de un registro que materializó el evento de interés definido en el numeral 3.2; entiéndase días de mora alcanzados entre 51 y 60 días

Sujetos de estudio retirados: 1 358

3.2 Definición de la variable objetivo

Se realizó una revisión de tasas de migración entre estadios de morosidad en 2 diferentes períodos de maduración (meses post desembolso) en la población analizada, encontrando que, tanto para los períodos 6 y 12 meses posteriores al desembolso, el tramo de morosidad mínimo en el cual se observa cómo más probable la migración hacia tramos de morosidad superiores es a partir de 60 días de morosidad.

Esta definición coincide con el estándar que la industria normalmente utiliza para el desarrollo de sus modelos de predicción de probabilidades de incumplimiento en el segmento de préstamos consumo no revolvente.

GRÁFICO 7: Tasas de migración según rango de atraso para mes 6 de madurez

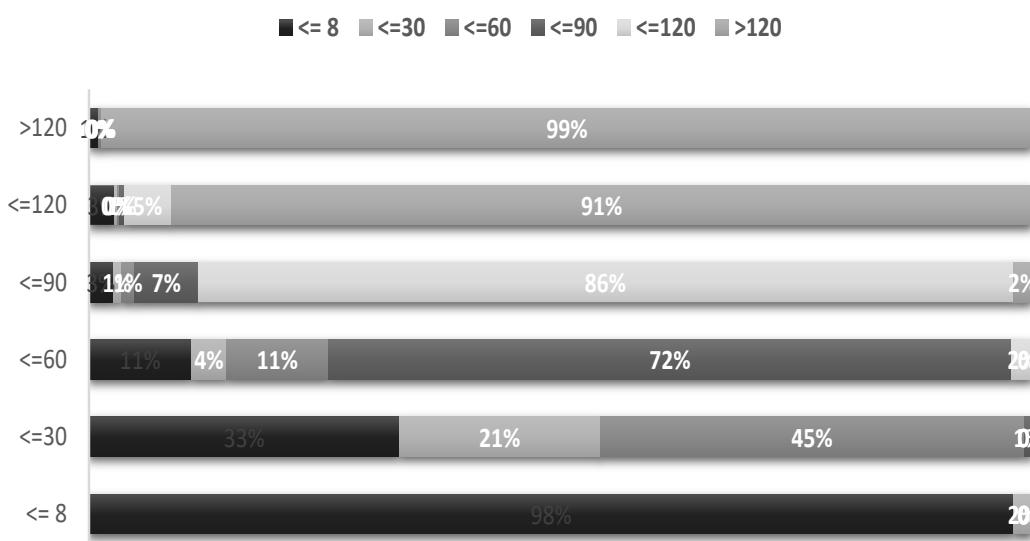
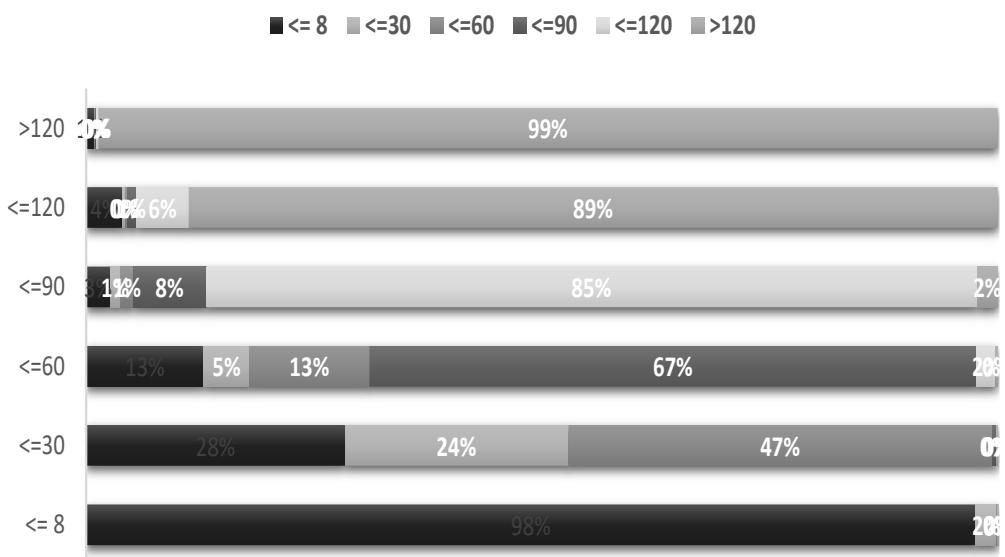


GRÁFICO 8: Tasas de migración según rango de atraso para mes 12 de madurez



3.3 Variables inicialmente analizadas

En el análisis inicial, se incluyeron las siguientes variables, agrupadas según su naturaleza:

TABLA 3: Variables explicativas incluidas en la muestra de desarrollo inicial

GRUPO	NOMBRE VARIABLE	DETALLE
Comportamiento de pago previo en sistema financiero	Calificación SBS consolidada	A la fecha de desembolso y calificación más deteriorada en los 12 meses previos
	Porcentaje de calificación normal	A la fecha de desembolso y calificación más deteriorada en los 12 meses previos
	Días de mora en cualquiera de sus créditos en el sistema financiero	A la fecha de desembolso y el mayor registro de días de mora en los últimos 12 meses
	Saldos deudores en mora 30, 60 y 90 días	A la fecha de desembolso y el mayor registro de saldos en mora en los últimos 12 meses
	Saldos deudores vencidos contables y saldos castigados	

		A la fecha de desembolso y el mayor registro de saldos en mora en los últimos 12 meses
Comportamiento transaccional en el sistema financiero	Entidades acreedoras en el SF	A la fecha de desembolso y el mínimo registro observado en los últimos 12 meses
	Saldos por Disposición de efectivo en Tarjetas de crédito	A la fecha de desembolso y el mínimo registro observado en los últimos 12 meses
	Línea total en tarjetas de crédito	A la fecha de desembolso y el máximo registro observado en los últimos 12 meses
	Utilización de líneas de crédito. El cociente entre Líneas utilizadas / Línea Total	A la fecha de desembolso y el máximo registro observado en los últimos 12 meses
	Variación de saldos deudores en los últimos 12 meses. El cociente entre: Saldo deudor máximo últimos 12 meses-/ Saldo deudor actual – 1	A la fecha de desembolso
	Relación deuda RCC / Ingreso	Deuda RCC observable a la fecha de desembolso
Socio económicos y demográficos	Perfil laboral	Dependiente / Independiente
	Ingresos	Ingresos como dependiente
	Edad	A la fecha de desembolso

3.4 Pruebas univariadas de Capacidad de predicción de las variables analizadas

Es importante señalar el tratamiento que se aplicó a las variables continuas que forman parte del modelo. Con el fin de conocer su capacidad de predicción individual, así como buscar una mayor estabilidad en los modelos tanto de regresión logística como de Cox, se optó por utilizar una técnica ampliamente utilizada en la industria que busca conocer y maximizar su capacidad de discriminación individual a través de la discretización de las variables continuas y del posterior cálculo de los estadísticos *Weight of evidence* e *Information Value*.

El estadístico *Information Value* está definido como:

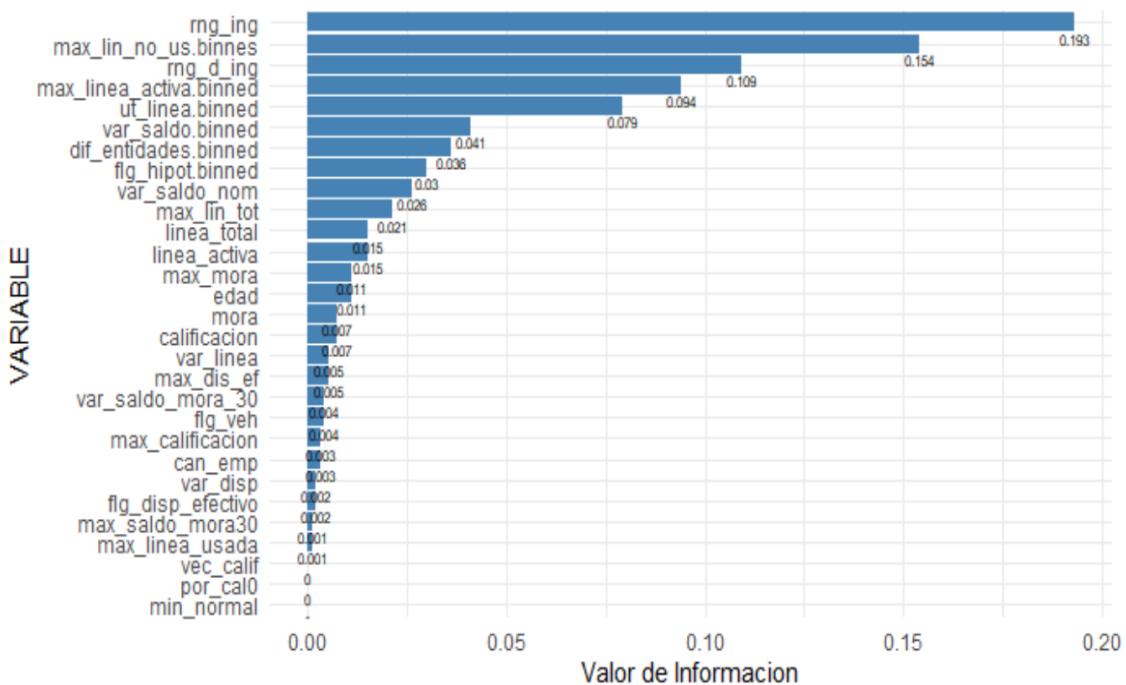
$$IV = \sum_{i=1}^n (\theta_{bi} - \theta_{mi}) * WOE_i$$

Donde θ_{bi} , θ_{mi} son la proporción de clientes buenos y malos, respectivamente del tramo i y WOE_i es el estadístico *Weight of evidence* del tramo i que se define como:

$$WOE_i = \log \left(\frac{\theta_{bi}}{\theta_{mi}} \right)$$

Las variables que fueron parte del análisis bivariado de capacidad de discriminación y que fueron sometidas al tratamiento indicado en el párrafo anterior se presentan en el gráfico 3 ordenadas de mayor a menor poder de discriminación.

GRÁFICO 9: Variables según su valor de información



Para fines del desarrollo tanto del modelo de regresión logística como el de regresión de Cox, se utilizarán sólo las variables que superen un valor de información mayor a 2%, esto siguiendo las mejores prácticas en el desarrollo de modelos (BBVA, 2010), que para éste caso son:

- (1) Interacción Ingresos mensuales y perfil laboral (Dependiente – Independiente)

- (2) Máxima línea de tarjetas de crédito no usada en últimos 12 meses
- (3) Cociente deuda RCC último mes/ Ingresos
- (4) Máxima línea de tarjetas de crédito activas en últimos 12 meses
- (5) Cociente de utilización de líneas de crédito en último RCC observado al momento del desembolso
- (6) Cociente de la variación porcentual de los saldos deudores en el RCC del último mes observado respecto al máximo de los 12 meses previos
- (7) Variación del número de entidades reportantes en RCC del último mes observado respecto al máximo de los 12 meses previos
- (8) Indicador de deuda hipotecaria vigente en último RCC observado
- (9) Variación de saldo deudor en términos absolutos en el RCC del último mes observado respecto al máximo de los 12 meses previos
- (10) Línea Total observada en el último RCC a la fecha de desembolso

3.5 Modelo de Regresión Logística

Siendo el objetivo principal del trabajo de investigación es la verificación de la eficacia de la regresión de Cox como método alternativo a la regresión logística, es necesario desarrollar el segundo con la finalidad de obtener resultados comparativos en la estimación de pérdidas esperadas. Los resultados de la estimación del modelo de regresión logística se muestran a continuación

3.5.1 Forma funcional del modelo regresión logística

Se utilizó el algoritmo *Stepwise (Backward y Forward)* con el fin de definir las variables que añaden información relevante al modelo de regresión logística, se observa un alto grado de significancia de las variables utilizadas en el modelo, cuyo resumen se muestra a continuación en la tabla 4:

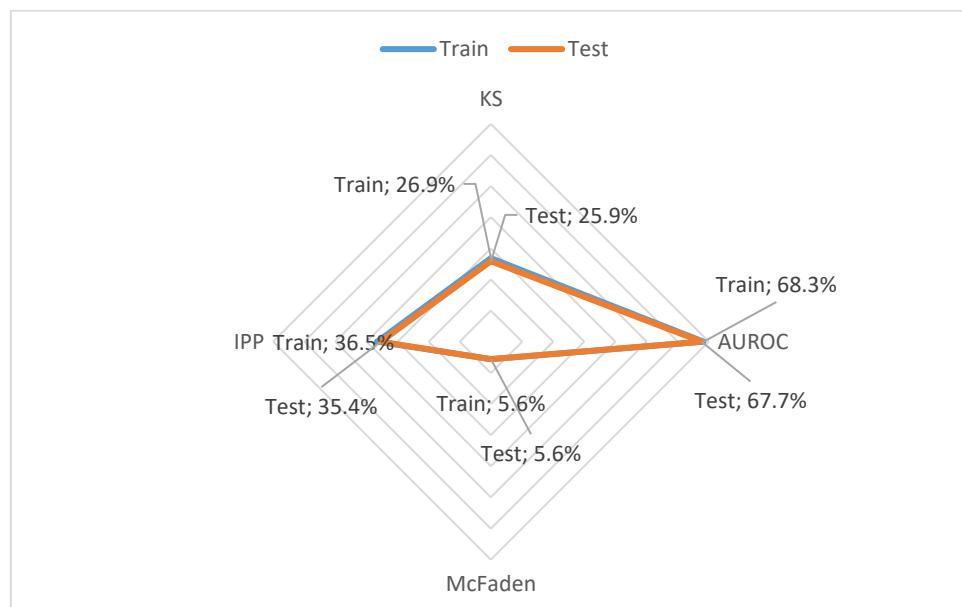
TABLA 4: Coeficientes del modelo de regresión logística

VARIABLE	Descripción Categoría	Coeficientes(B)	Error	P value	Exp(B)-1	Interpretación	Exp(B)-1	
	Intercept	Intercepto	-	4.61	0.16	-	0.990021	
INGRESOS DEMOSTRADOS	rng_ing<=1500	Ingresos <= 1 500	0.99	0.14	-	1.702507	Incrementa el riesgo de default en 170% respecto a cat.	
	rng_ing<=3000	Ingresos >1500 y <=3000	0.76	0.14	-	1.146052	Incrementa el riesgo de default en 114% respecto a cat.	
	rng_ing<=4000	Ingresos >3000 y <=4000	0.64	0.15	-	0.891556	Incrementa el riesgo de default en 89% respecto a cat.	
	rng_ing<=6000	Ingresos >4000 y <=6000	0.27	0.16	0.08	0.314755	Incrementa el riesgo de default en 31% respecto a cat.	
	rng_ing<=930	Ingresos<= 930	1.25	0.14	-	2.487901	Incrementa el riesgo de default en 248% respecto a cat.	
	rng_ing>10 000	Ingresos>10 000	-	12.22	54.31	0.82 -	0.999995	Reduce el riesgo de default en 99% respecto a cat.
	rng_ingIndependiente	Independientes- Ingreso no observado	1.51	0.13	-	3.527002	Incrementa el riesgo de default en 150% respecto a cat.	
Categoría Referencia: Ingreso>6 000 y <=10 000								
MAXIMO VALOR DE LINEA DE CREDITO NO UTILIZADA EN LOS ULTIMOS 12 MESES	max_lin_no_us.binnes<=1200	Línea Máxima no usada <= 1 200	0.20	0.04	-	0.223187	Incrementa el riesgo de default en 22% respecto a cat.	
	max_lin_no_us.binnes<=20000	Línea Máxima no usada <= 20 000	-	0.30	0.04	-	0.261704	Reduce el riesgo de default en -26% respecto a cat.
	max_lin_no_us.binnes<=80000	Línea Máxima no usada <= 80 000	-	0.61	0.04	-	0.454733	Reduce el riesgo de default en -45% respecto a cat.
	max_lin_no_us.binnes>80000	Línea Máxima no usada > 80 000	-	1.03	0.11	-	0.643261	Reduce el riesgo de default en -64% respecto a cat.
	max_lin_no_us.binnesSin Lineas	Sin líneas de crédito registradas	0.63	0.06	-	0.885871	Incrementa el riesgo de default en 88% respecto a cat.	
Categoría Referencia: Maxima Línea no utilizada > 2 000 y <= 11 01								
PORCENTAJE DE UTILIZACION DE LINEAS DE CREDITO EN ULTIMO MES	ut_linea.binned<=70%	Utilización de líneas <= 70%	0.38	0.06	-	0.463733	Incrementa el riesgo de default en 46% respecto a cat.	
	ut_linea.binned>70%	Utilización de líneas > 70%	0.81	0.07	-	1.238867	Incrementa el riesgo de default en 123% respecto a cat.	
	ut_linea.binnedSin Lineas	Sin líneas registradas	-	0.02	0.08	0.81 -	0.018428	Reduce el riesgo de default en -2% respecto a cat.
Categoría Referencia: Utilización de líneas >0% y <= 4%								
VARIACION DE ENTIDADES ACREEDORAS DESDE EL MINIMO REGISTRADO EN LOS ULTIMOS 12 MESES	dif_entidades.binned>1 entidad	Variación de entidades >1	0.49	0.04	-	0.640039	Incrementa el riesgo de default en 46% respecto a cat.	
	dif_entidades.binned1 entidad	Variación de Entidades = 1	0.24	0.03	-	0.266415	Incrementa el riesgo de default en 26% respecto a cat.	
Categoría Referencia: Diferencia de entidades <=0								
VARIACION PORCENTUAL DE SALDO DEUDOR EN EL SISTEMA FINANCIERO. RESPECTO AL MINIMO DE LOS ULTIMOS 12 MESES	var_saldo.binned>0% y <=100%	Variación de saldo >0% y <=100%	0.15	0.04	-	0.159849	Incrementa el riesgo de default en 16% respecto a cat.	
	var_saldo.binned>100%	Variación de saldo >=100%	0.31	0.04	-	0.356991	Incrementa el riesgo de default en 35% respecto a cat.	
	var_saldo.binnedSin deuda previa	Sin deudas previas en SF	0.24	0.08	-	0.274444	Incrementa el riesgo de default en 27% respecto a cat.	
Categoría de referencia: Variación de saldo Negativa								
INDICADOR DE CONTAR CON CREDITO HIPOTECARIO(1:SI, 0:	flg_hipot.binnedsin Hipoteca	Categoría = 0; No tiene hipoteca Categoría referenci = 1; Tiene Hipoteca	0.27	0.08	-	0.304083	Incrementa el riesgo de default en 30% respecto a cat.	

3.5.2 Indicadores de capacidad de predicción

Según se indicó en el numeral 3.1, se definió de forma aleatoria una muestra de desarrollo y validación equivalentes al 70% y 30% de la data analizada respectivamente.

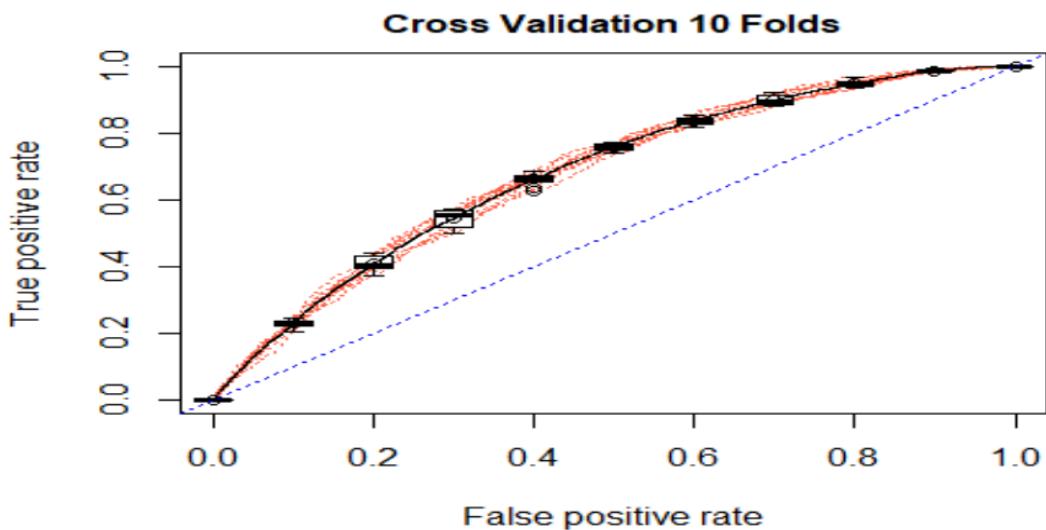
GRÁFICO 10: Indicadores de capacidad de predicción del modelo de regresión logística en muestras de desarrollo y validación



Como se puede apreciar, no existe una pérdida de capacidad de predicción significativa al aplicar el modelo entre las muestras de desarrollo y validación, se puede observar que el indicador de AUROC se encuentra muy cercano a los valores recomendados por Hosmer y Lemeshow (1989) en su libro *Logistic Regression*.

Finalmente, al aplicar un análisis de validación cruzada, se demostró que el modelo muestra estabilidad en sus indicadores de capacidad de predicción

GRÁFICO 11: Análisis de validación cruzada regresión logística

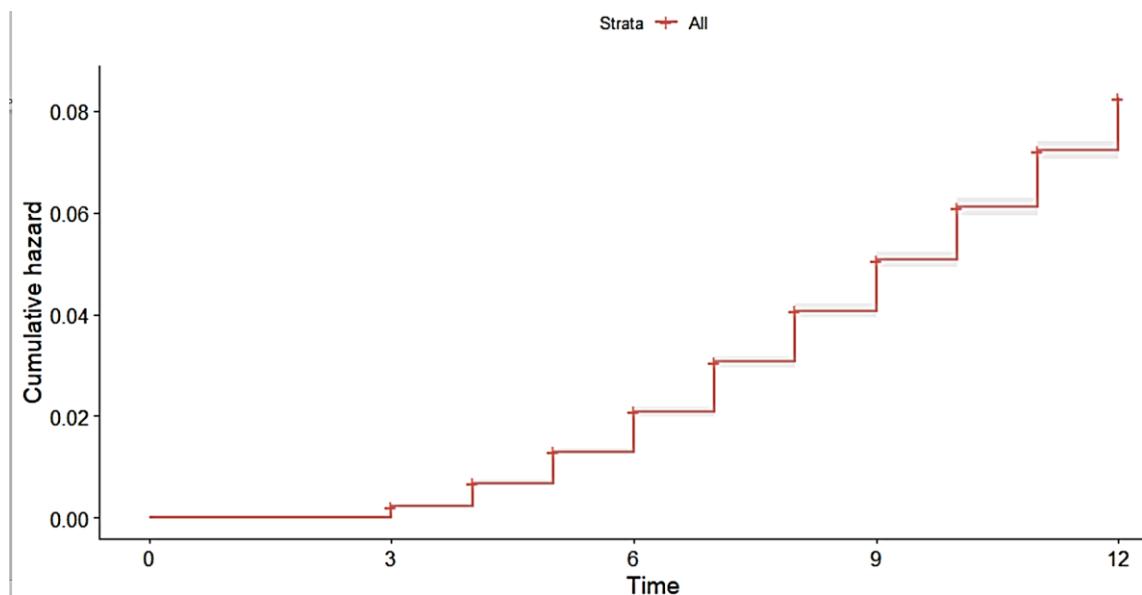


3.6 Modelo de Regresión de Cox

3.6.1 Estimador de Kaplan Meier

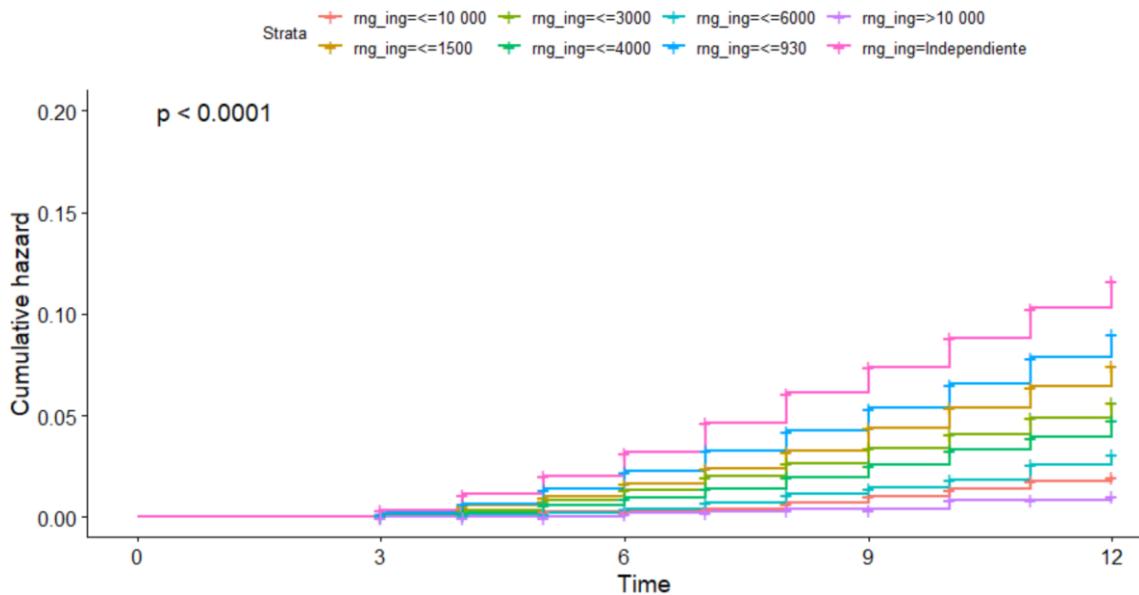
Se muestra el gráfico de riesgo acumulado (“*cumulative hazard*”) general de la muestra de desarrollo sin ninguna estratificación. En él se observa que el riesgo de caer en incumplimiento se acelera a partir del mes 6 de maduración

GRÁFICO 12: Gráfico de riesgo acumulativo basado en estimador Kaplan Meier



Así mismo, el mismo estimador de Kaplan Meier, se muestra segmentado, junto con el estadístico p-value, basados en pruebas de Chi cuadrado con el fin de validar si las diferencias en las probabilidades de supervivencia son estadísticamente significativas entre grupos, siendo que p-values por debajo del 0.05 son evidencia de curvas de riesgo cumulativo diferenciadas entre los estratos de la variable.

GRÁFICO 13: Gráfico de riesgo acumulativo basado en estimador Kaplan Meier - segmentado por nivel de Ingresos y perfil laboral



En anexo 1 se muestran los gráficos de riesgo acumulativo basados en el estimador de Kaplan-Meier de las demás covariables incluidas en el estudio.

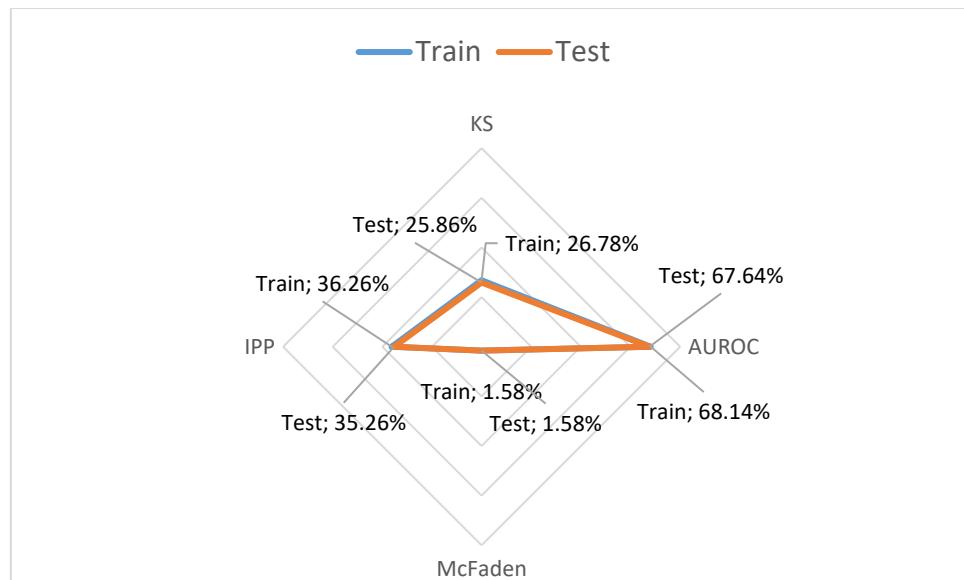
3.6.2 Coeficientes del modelo de regresión de Cox

TABLA 5: Forma Funcional de modelo de regresión de Cox

VARIABLE	Descripción Categoría	Coeficientes(B)	se(B)	P Value	Exp(B)	Hazard Ratio	Interpretación Hazard Ratio
INGRESOS DEMOSTRADOS	rng_ing<=1500	Ingresos <= 1 500	0.95	0.14	-	2.591	El rango de ingresos incrementa el riesgo en 2.5 veces respecto a la categoría de ingresos d
	rng_ing<=3000	Ingresos >1500 y <=3000	0.73	0.14	-	2.077	El rango de ingresos incrementa el riesgo en 2.081 veces respecto a la categoría de ingresos
	rng_ing<=4000	Ingresos >3000 y <=4000	0.64	0.15	-	1.897	El rango de ingresos incrementa el riesgo en 1.9 veces respecto a la categoría de ingresos d
	rng_ing<=6000	Ingresos >4000 y <=6000	0.29	0.16	-	1.337	El rango de ingresos incrementa el riesgo en 1.3 veces respecto a la categoría de ingresos d
	rng_ing<=930	Ingresos<= 930	1.19	0.14	-	3.297	El rango de ingresos incrementa el riesgo en 3.3 veces respecto a la categoría de ingresos d
	rng_ing>10 000	Ingresos>10 000	-	0.52	0.30	0.594	El rango de ingresos reduce el riesgo en 41% respecto a la categoría de ingresos de referen
	rng_ingIndependiente	Independientes- Ingreso no	1.44	0.13	-	4.234	El rango de ingresos incrementa el riesgo en 4.3 veces respecto a la categoría de ingresos d
Categoría Referencia: Ingresos>6 000 y <=10 000							
MAXIMO VALOR DE LINEA DE CREDITO NO UTILIZADA EN LOS ULTIMOS 12 MESES	max_lin_no_us.binnes<=1	Línea Máxima no usada <= 1	0.18	0.04	-	1.194	El rango de linea no usada incrementa el riesgo en 1.19 veces respecto al rango de linea no u
	max_lin_no_us.binnes<=2	Línea Máxima no usada <= 2-	0.26	0.04	-	0.770	El rango de linea no usada reduce el riesgo en 23% respecto al rango de linea no usada de r
	max_lin_no_us.binnes<=8	Línea Maxima no usada <= 8-	0.54	0.04	-	0.585	El rango de linea no usada reduce el riesgo en 42% respecto al rango de linea no usada de r
	max_lin_no_us.binnes>80	Línea Maxima no usada > 80 -	0.93	0.11	-	0.397	El rango de linea no usada reduce el riesgo en 61% respecto al rango de linea no usada de r
	max_lin_no_us.binnesSin	Sin líneas de crédito regi:	0.57	0.06	-	1.771	El rango de linea no usada reduce el riesgo en 4% respecto al rango de linea no usada de r
	Referencia: Maxima Línea no utilizada > 2 000 y						
PORCENTAJE DE UTILIZACION DE LINEAS DE CREDITO EN ULTIMO MES	ut_linea.binned<=70%	Utilización de líneas < = :	0.332	0.06	-	1.394	El rango de utilizacion de lineas incrementa el riesgo e 1,4 veces respecto al rango de util
	ut_linea.binned>70%	Utilización de líneas > 70%	0.699	0.06	-	2.012	El rango de utilizacion de lineas incrementa el riesgo e 2 veces respecto al rango de utiliz
	ut_linea.binnedSin Linea	Sin líneas registradas	-	0.036	0.07	0.965	Clientes sin líneas registradas, reducen el riesgo en 4% respecto al rango de utilización de
VARIACION DE ENTIDADES ACREDITADAS DESDE EL MINIMO REGISTRADO EN LOS ULTIMOS 12 MESES	dif_entidades.binned>1	variación de entidades >1	0.465	0.04	-	1.592	El rango de variación de entidades incrementa el riesgo en 1,59 veces respecto al rango de r
	dif_entidades.binned1	variación de Entidades = 1	0.224	0.03	-	1.251	El rango de variación de entidades incrementa el riesgo en 1,25 veces respecto al rango de r
	Categoría Referencia: Diferencia de entidades <=						
VARIACION PORCENTUAL DE SALDO DEUDOR EN EL SISTEMA FINANCIERO, RESPECTO AL MINIMO DE LOS ULTIMOS 12 MESES	var_saldo.binned>0% y	variación de saldo >0% y <	0.141	0.04	-	1.151	El rango de variación de saldo incrementa el riesgo en 1.15 veces respecto a variación negat
	var_saldo.binned>100%	variación de saldo >=100%	0.299	0.04	-	1.349	El rango de variación de saldo incrementa el riesgo en 1.35 veces respecto a variación negat
	var_saldo.binnedSin deud	Sin deudas previas en SF	0.279	0.08	-	1.322	Clientes sin deudas previas incrementan el riesgo en 1.32 veces respecto a aquellos con vari
INDICADOR DE CONTAR CON CREDITO HIPOTECARIO(1:SI, 0: NO)	flg_hipot.binnedsin Hip	Categoría = 0; No tiene hi	0.25	0.08	-	1.280	Clients sin créditos hipotecarios, incrementan el riesgo en 1.28 veces respercto a los que
	Categoría referencia = 1; Tiene Hipoteca						

3.6.3 Indicadores de capacidad de predicción

GRÁFICO 14: Indicadores de capacidad de predicción del modelo de regresión de Cox en muestras de desarrollo y validación

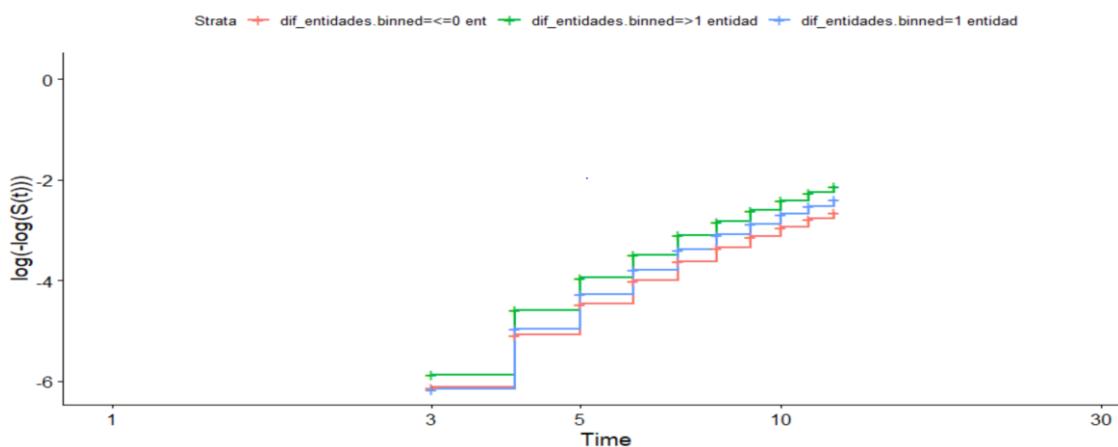


3.6.4 Análisis de los supuestos de riesgos proporcionales

3.6.4.1 Análisis gráfico del supuesto de riesgos proporcionales

Se muestra el gráfico log-log de la función de supervivencia de las tres categorías de la variable “Diferencia de Entidades acreedoras”, con el fin de evaluar si su evolución respecto al tiempo muestra un comportamiento paralelo que nos dé un indicativo de que el supuesto de riesgos proporcionales explicado en el capítulo 2 se cumple para este estudio.

GRÁFICO 15: Gráfico Log-Log de la variable “Diferencia Entidades”



Las funciones de supervivencia de las categorías de la variable “Diferencia Entidades” que describe el incremento del número de entidades acreedoras en el sistema financiero al momento de la solicitud del crédito consumo analizado, muestran un comportamiento paralelo a lo largo del tiempo, por lo que gráficamente se podría inferir que para esta variable el supuesto de riesgos proporcionales es cumplido.

De la misma manera en anexo 2 se presentan los gráficos Log-Log de las demás covariables del modelo de Cox ajustado

3.6.4.2 Análisis de los residuos de Schoenfeld

Se estimaron los residuos de Schoenfeld para cada una de las variables que cumplen con el criterio de riesgos proporcionales, buscando no rechazar la hipótesis nula de riesgos proporcionales

TABLA 6: Análisis de residuos de Shoenfeld – Modelo de Cox

Variable	Descripción	chisq	df	p
rng_ing	Rango Ingresos	55.22	7	0.000000001
max_lin_no_us.binnes	Línea no utilizada	28.33	5	0.000031424
ut_linea.binned	Utilización de líneas	35.77	3	0.000000084
dif_entidades.binned	Diferencia entidades	0.89	2	0.639281179
var_saldo.binned	Rangos de Variación de saldos Sistema financiero	25.04	3	0.000015150
flg_hipot.binned	Flag Hipoteca	6.80	1	0.009097394

La tabla 6 muestra que el supuesto de riesgos proporcionales de Schoenfeld se cumple para la variable “Diferencia de Entidades”.

Estos resultados pueden estar influenciados por la discretización de las variables predictoras aplicada bajo el método de *Weight of evidence* que busca maximizar su capacidad de discriminación individual.

Hay que mencionar así mismo, que, al graficar los residuos de Schoenfeld contra el tiempo para cada una de las variables predictoras, no muestran una tendencia clara respecto al tiempo, por lo cual no sería posible descartar la hipótesis nula de riesgos proporcionales.

GRÁFICO 16: Gráfico Residuos de Schoenfeld de variable Ingresos versus el tiempo

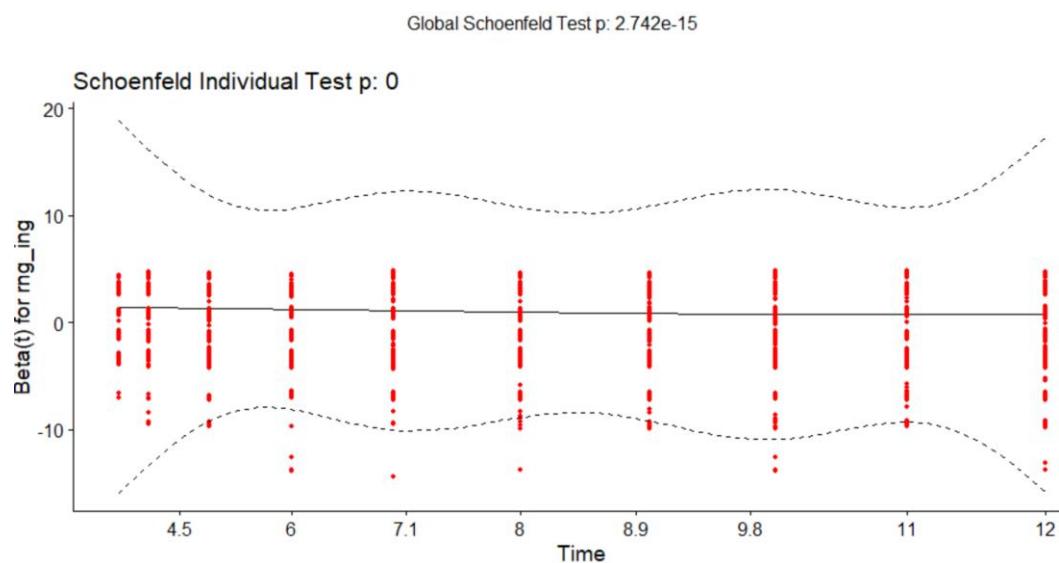
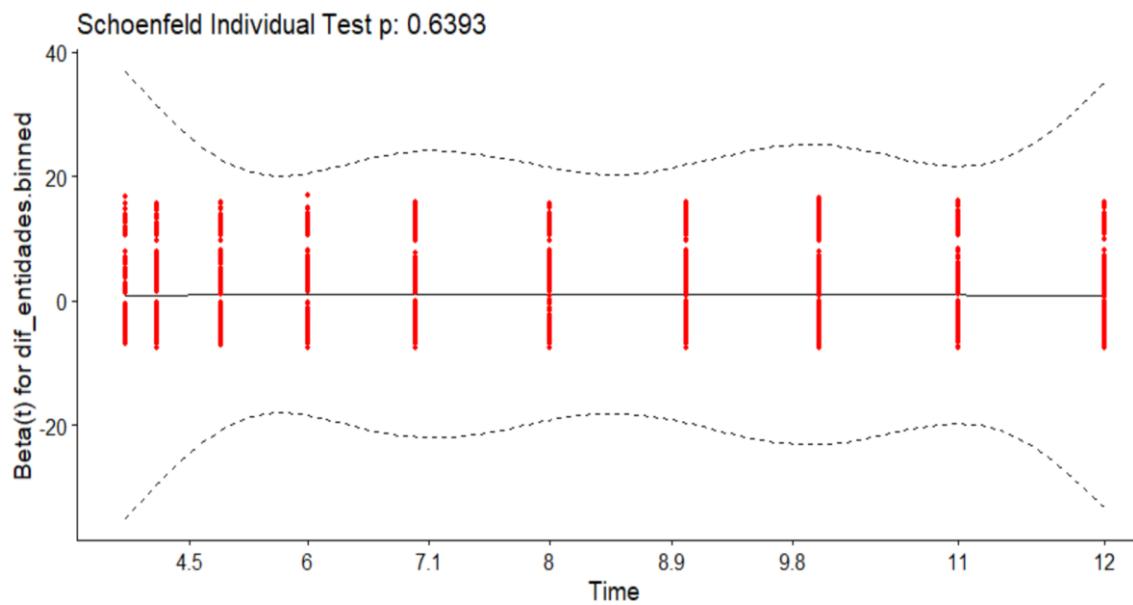


GRÁFICO 17: Gráfico de los residuos de Schoenfeld para la variable “Diferencia de Entidades” respecto al tiempo



Los gráficos de los residuos de Schoenfeld respecto al tiempo para cada una de las variables del modelo se muestran en el Anexo 3.

3.6.5 Análisis del supuesto de linealidad

Dado que en el modelo de Cox todas las variables fueron discretizadas tratando de maximizar el indicador de *Weight of evidence*, la no lineariadad de la forma funcional de los coeficientes que se generan en función de estas variables no es un supuesto a probar, tal como se detalló en la revisión teórica del modelo de Cox en el capítulo II.

3.7 Análisis comparativo de pérdidas esperadas

Se realizó la estimación de la pérdida esperada tanto bajo el método de regresión logística bajo el enfoque de IRB de Basilea II, como por el método de IFRS9 bajo modelos de supervivencia – Regresión de Cox.

Los resultados de las pérdidas esperadas estimadas bajo ambas metodologías serán comparadas con las pérdidas reales materializadas al finalizar el período de observación de los sujetos de estudio (12 meses).

3.7.1 Resultado de pérdidas esperadas estimadas bajo modelo de regresión logística

PARÁMETROS

PD: Probabilidad de Incumplimiento calibrada con base test

LGD: 100%

EAD: Monto Desembolsado

PERDIDA REAL: Saldo de operaciones en Incumplimiento (> 60 días) al mes 12 de maduración (se suman castigos realizados)

TABLA 7: Calibración de la probabilidad de Incumplimiento

	PD Calibrada	N° observaciones	% de Observaciones
Bucket 1	2.2%	8 607	20.0%
Bucket 2	5.2%	17 028	39.6%
Bucket 3	8.5%	2 071	4.8%
Bucket 4	9.9%	10 760	25.1%
Bucket 5	15.0%	4 487	10.4%
Total		42 953	100.0%

TABLA 8: Resumen agregado de Pérdida esperada y observada para el período analizado bajo método de regresiones logísticas

	Exposición	Pérdida Real 12 Meses(a)	Pérdida Esperada(b)	MAE	Error % (*)
TOTALES	569,253,084	20,900,555	35,056,384	2,359,305	67.7%
2018_09	92,983,643	3,359,954	5,596,658	2,236,703	66.6%
2018_10	97,862,505	3,673,752	5,945,440	2,271,688	61.8%
2018_11	89,436,208	3,537,825	5,587,753	2,049,928	57.9%
2018_12	87,865,376	3,304,570	5,553,481	2,248,911	68.1%
2019_01	102,842,749	3,538,583	6,323,426	2,784,844	78.7%
2019_02	98,262,603	3,485,871	6,049,626	2,563,754	73.5%

(*) Se define el error como el valor Absoluto de: (a-b)/a

Se puede observar que, existe una sobre estimación de pérdidas esperadas respecto de las pérdidas efectivamente materializadas por alrededor de S/. 14.1 MM para todo el período analizado, para una exposición inicial de S/. 569 MM. El desvío en términos relativos es de aproximadamente 67.7%.

3.7.2 Resultado de pérdidas esperadas estimadas bajo modelo de regresión de Cox

PARÁMETROS:

PD: PD Marginal *Lifetime*

LGD: 100%

EAD: MONTO DESEMBOLSADO

CURVA DE AMORTIZACION: Amortización Francesa a 18 Meses

TASA PARA ACTUALIZACION DE FLUJOS A VALOR PRESENTE:
18% TEA

PERDIDA REAL: Saldo de operaciones en Default (> 60 días) al mes 12 de maduración (se suman castigos realizados)

El cálculo de las pérdidas esperadas bajo el método de Cox con los parámetros definidos, se realiza siguiendo lo definido en el numeral 2.4.3.2 del capítulo II. Dado que se trata de la estimación del modelo al momento del desembolso de las operaciones de crédito, se aplica para todos los casos el *Stage 1* de la metodología IFRS9 al ser operaciones en estado de no incumplimiento ni de alto riesgo identificado.

GRÁFICO 18: Pérdidas observadas y esperadas según período de desembolso bajo método de regresiones de Cox

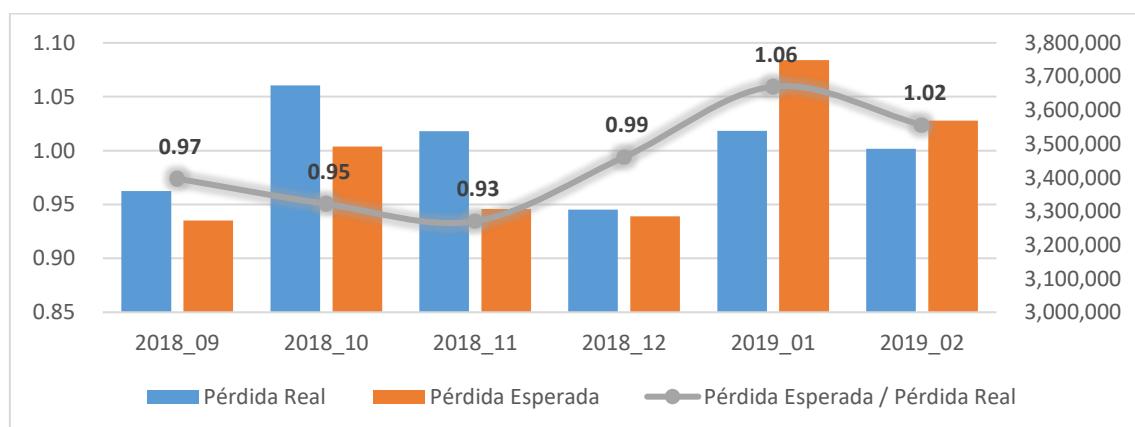


TABLA 9: Resumen agregado de Pérdida esperada y observada para el período analizado bajo método de regresiones de Cox

AGREGADO	Exposición	Pérdida Real 12 Meses(a)	Pérdida Esperada(b)	MAE	Error % (*)
AGREGADO	569,253,084	20,900,555	20,673,076	135,618	1.1%
2018_09	92,983,643	3,359,954	3,272,414	87,540	2.6%
2018_10	97,862,505	3,673,752	3,492,133	181,619	4.9%
2018_11	89,436,208	3,537,825	3,306,306	231,519	6.5%
2018_12	87,865,376	3,304,570	3,284,655	19,915	0.6%
2019_01	102,842,749	3,538,583	3,748,774	210,191	5.9%
2019_02	98,262,603	3,485,871	3,568,794	82,922	2.4%

(*) Se define el error como el valor Absoluto de: (a-b)/a

Se puede observar que las diferencias entre la pérdida esperada bajo el modelo de Cox y la pérdida efectivamente materializada se encuentran en un valor tan bajo como S/. 227 Mil para todo el portfolio analizado, para una exposición inicial de S/. 569 millones y una pérdida real observada de S/. 20. 9 Millones. El desvío en términos relativos es de aproximadamente 1.1%.

Al realizar el comparativo entre ambas metodologías se observa claramente un mejor acercamiento de las pérdidas estimadas bajo el modelo de regresión de Cox, respecto del modelo de regresión logística convencional, entre los motivos que sustentan esta mejor aproximación se encuentran:

- a) La metodología basada en métodos de supervivencia toma en cuenta el perfil de amortización de las operaciones de crédito
- b) La metodología basada en métodos de supervivencia, en combinación con el punto anterior, realiza estimaciones basadas en probabilidades de incumplimiento durante el período de vida de la operación

GRÁFICO 19: Perdida real y brechas versus modelo de cox y regresión logística

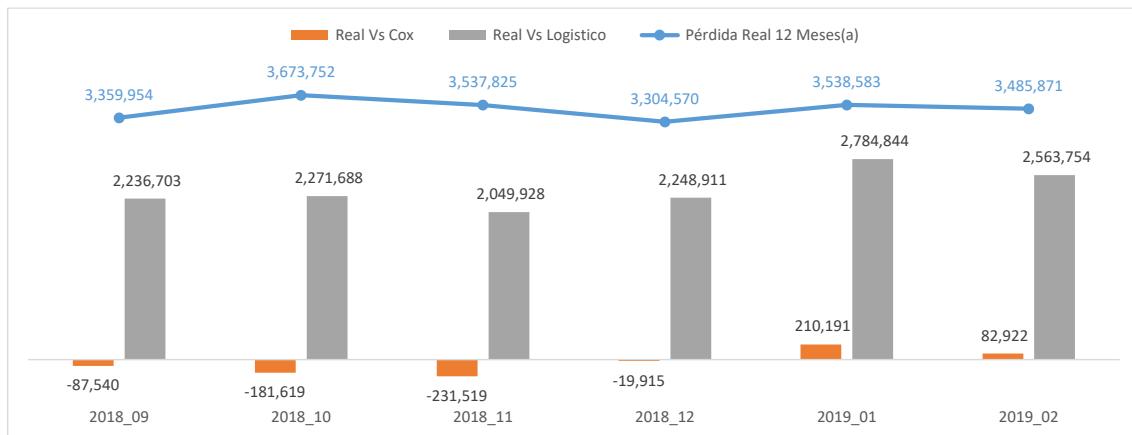


TABLA 10: Comparativo Pérdida Esperada bajo modelos de Regresión Logística y de Cox comparados con la pérdida efectivamente materializada

AGREGADO	Pérdida Real 12 Meses(a)	Pérdida Esperada Cox	Pérdida Esperada Logístico	Real Vs Cox	Real Vs Logistico
2018_09	20,900,555	20,673,076	35,056,384	87,540	2,236,703
2018_10	3,359,954	3,272,414	5,596,658	181,619	2,271,688
2018_11	3,673,752	3,492,133	5,945,440	231,519	2,049,928
2018_12	3,537,825	3,306,306	5,587,753	19,915	2,248,911
2019_01	3,304,570	3,284,655	5,553,481	210,191	2,784,844
2019_02	3,538,583	3,748,774	6,323,426	82,922	2,563,754
	3,485,871	3,568,794	6,049,626		

3.8 Estimación del número de incumplimientos durante la vida de las operaciones de crédito

El modelo de Cox, a diferencia del modelo de regresión logística, permite generar estimaciones del número de operaciones que caerían en incumplimiento en cada período posterior, de esta forma se pueden comparar el número de incumplimientos pronosticados versus el número de incumplimientos efectivamente materializados. El incumplimiento se define de la forma cómo está detallado en el numeral 3.2 del presente capítulo

El número de incumplimientos estimados a través del modelo de Cox se estima de la siguiente forma:

$$N_T = \sum_{t=1}^{12} \rho_{it} - \rho_{it-1}$$

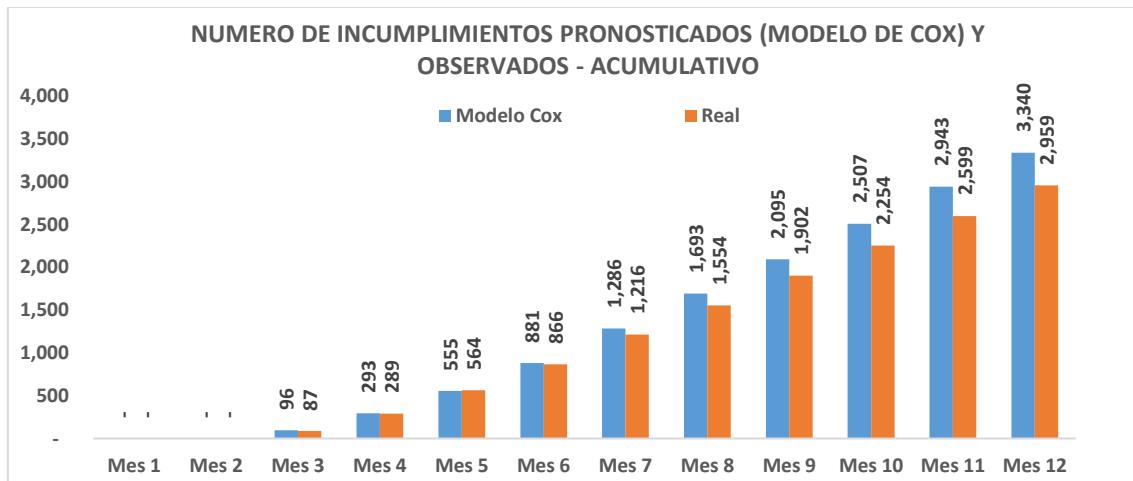
Donde:

N_T = Número de incumplimientos observados en el portafolio evaluado durante todo el período de evaluación T

ρ_{it} = Probabilidad de incumplimiento del sujeto i en el tiempo t

ρ_{it-1} = Probabilidad de incumplimiento del sujeto i en el tiempo t-1

GRÁFICO 20: Número de incumplimientos pronosticados y estimados bajo el modelo de regresión de Cox para cada punto temporal de las operaciones de crédito



Se observa que la estimación del número de incumplimientos bajo el modelo de regresión de Cox mantiene una alta precisión hasta el mes 6 posterior a su estimación, empezando a mostrar desvíos a partir de dicho período, los cuales se van ampliando.

Este comportamiento es esperable en cualquier modelo analítico de pronóstico de probabilidad de incumplimiento y sugiere la necesidad de actualizar las estimaciones de probabilidad de incumplimiento a partir del mes 6 posterior a su cálculo inicial.

Esta propiedad del modelo de Cox podría ser de alta utilidad para la gestión del portafolio de créditos, dado que podría ayudar a: (1) Planificar mejor los esfuerzos de recuperación de cartera, (2) Generar proyecciones financieras más precisas en el tiempo respecto las pérdidas esperadas por riesgo de crédito y por tanto mejorar las decisiones y planificación de la gestión financiera en las entidades financieras.

CAPITULO IV: PROPUESTAS

1. Debido a la mejora en la precisión de la estimación de la pérdida real a través de la utilización de regresiones basadas en métodos de supervivencia, se considera que es factible y recomendable la adopción de éste tipo de métodos por parte de las entidades financieras con el fin de realizar mejores estimaciones de pérdidas esperadas por riesgo de crédito.
2. Se propone se puedan establecer incentivos por parte del regulador para que las entidades financieras adopten metodologías de estimación de probabilidades de incumplimiento y en específico de los modelos de regresión basados en métodos de supervivencia, en vista que éstos pueden mejorar las estimaciones de pérdidas por riesgo de crédito en los portafolios de las entidades supervisadas. Fortaleciendo aspectos estratégicos como la asignación de tasas de interés en función del riesgo, así como la planificación financiera y de capital.
3. Es recomendable la adopción de los modelos de Cox en los procesos de planificación de la gestión de cobranza de las entidades financieras, podría ser beneficiosa, toda vez que permite estimar el número de incumplimientos para cada período posterior a su estimación, cuidando la actualización de sus estimaciones en un intervalo no menor a 6 meses

CONCLUSIONES Y RECOMENDACIONES

1. A través de los resultados de la investigación, que concluyen la superioridad del modelo de regresión de Cox sobre el modelo de regresión logística en su eficiencia para la estimación de perdidas esperadas sobre el portafolio de créditos analizado, se ha cumplido con el objetivo general y los objetivos específicos de la misma planteados al inicio de la investigación
2. De la misma manera, se da respuesta al problema general planteado, concluyendo que el modelo de regresión de Cox es superior en su eficiencia para la estimación de las pérdidas materializadas por riesgo de crédito para el portafolio de créditos analizado.
3. La adopción de métodos avanzados para la estimación de pérdidas esperadas se hace de necesidad imperiosa en el sistema financiero peruano considerando los siguientes aspectos regulatorios y competitivos:
 - Los límites de tasa de interés impuestos por el gobierno peruano a las operaciones de crédito consumo y MYPES en el 2021, obligan a las entidades financieras a ser más eficientes en sus esquemas de asignación de precios en base a riesgo con el fin de mitigar los impactos que ésta nueva regulación pueda tener en sus ingresos financieros.
 - El ambiente cada vez más competitivo del mercado financiero peruano, genera que un esquema de asignación de precios eficiente y preciso sea una ventaja competitiva fundamental entre entidades financieras.
4. Es importante que el regulador peruano, genere los incentivos y marco regulatorio necesarios para que las entidades financieras puedan en su mayoría adoptar este tipo de métodos para la estimación de pérdidas por riesgo de crédito.
5. En específico, los métodos basados en análisis de supervivencia como el método de regresión de Cox, en combinación con la metodología de pérdidas esperadas definidas en la nueva norma contable IFRS9 muestran una clara superioridad frente a los métodos de regresión logística usualmente utilizados en el sistema financiero peruano, y basados en los resultados del presente estudio, su utilización es altamente recomendable.

REFERENCIAS BIBLIOGRÁFICAS

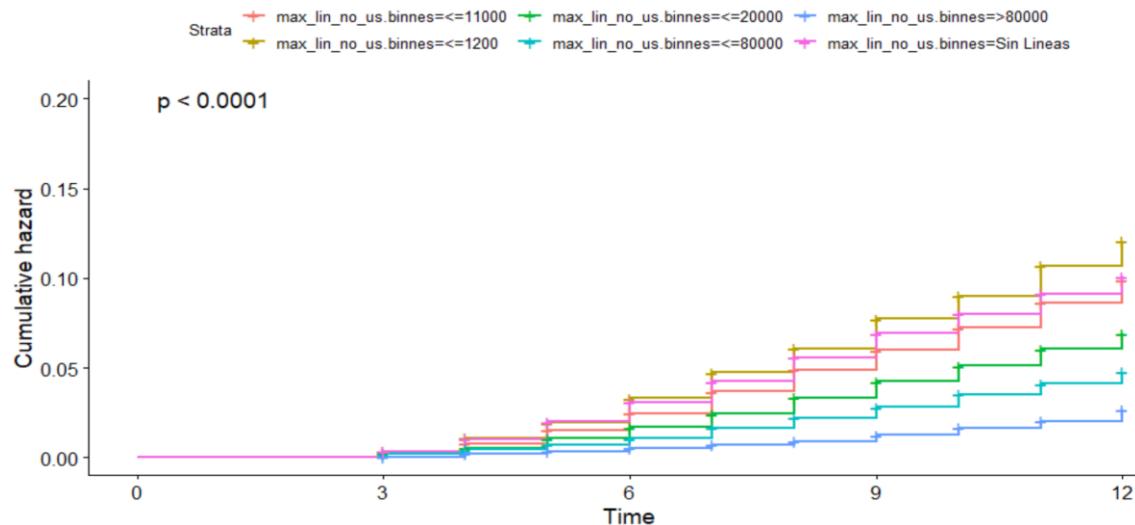
- Stepanova M., Thomas L(2002). *Survival analysis methods for personal loan data*
- Kleinbaum D., Klein M.(2012). *Survival Analysis, a Self Learning Text, Third Edition*
- Diricl L., Claeskens G., Baesens B.(2015). *Time to default in credit scoring using survival analysis: a benchmark study*
- Bellotti T, Crook J (2009), 'Credit scoring with macroeconomic variables using survival analysis', *Journal of the Operational Research Society*
- Eva Boj del Val(2017). *El modelo de Regresión de Cox de Riesgos Proporcionales*
- Uriarte F.,Agüero Y.(2017). *Análisis de supervivencia como alternativa metodológica para estimar probabilidades de incumplimiento de los deudores de créditos corporativos y a grandes empresas en el Perú*
- Man R. (2014). *Survival Analysis in credit Scoring, A framework for PD Estimation*
- Cao R, Devia A, Vilar J.(2009) *Modelling consumer credit risk via survival analysis*
- Fenek, Kai, Shapik(2016) *Economic Letters: Modelling the recovery outcomes for defaulted loans: A survival analysis approach*

ANEXOS

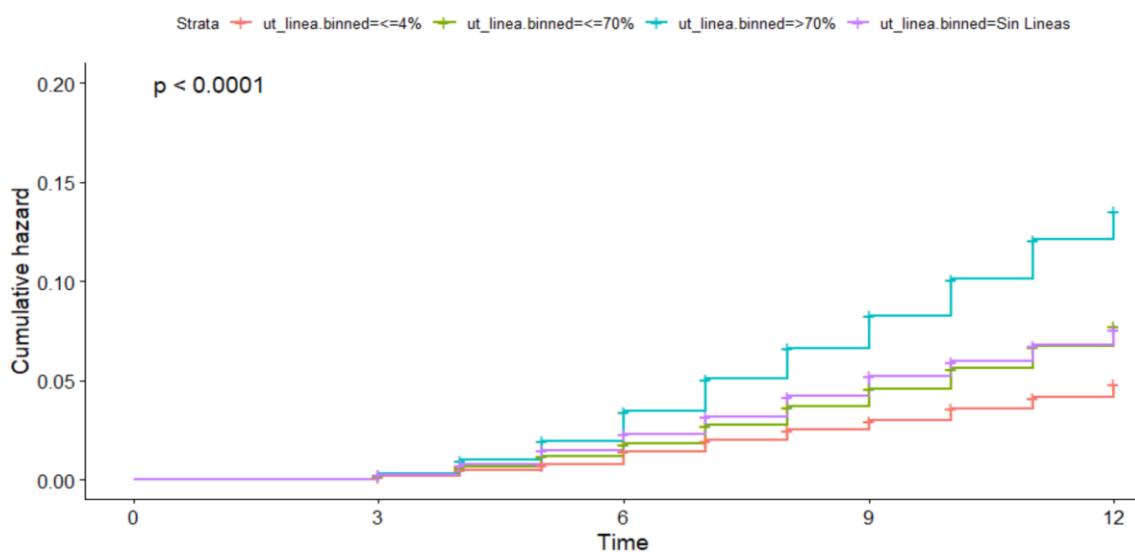
ANEXO 1: Gráficos de riesgo acumulativo basados en el estimador kaplan meir para las variables incluidas en el modelo de regresión de COX	1
ANEXO 2: Gráficos Log - Log versus el tiempo de las variables incluidas en el modelo	4
ANEXO 3: Gráficos de residuos de Shoenfeld para las variables endógenas del modelo de regresión de Cox.....	7

ANEXO 1: Gráficos de riesgo acumulativo basados en el estimador kaplan meir para las variables incluidas en el modelo de regresión de COX

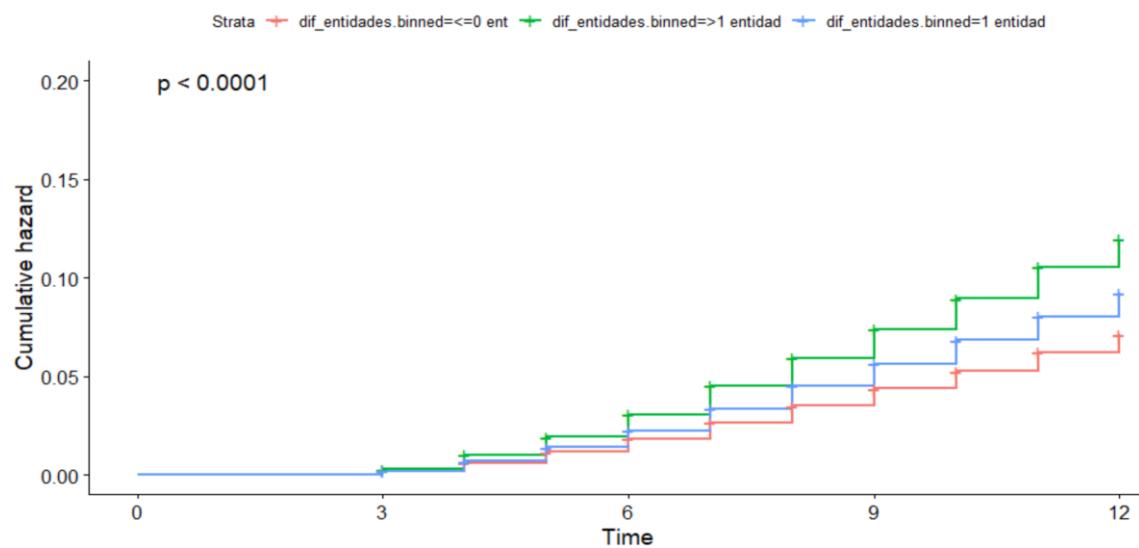
Monto de líneas no utilizadas en el sistema financiero



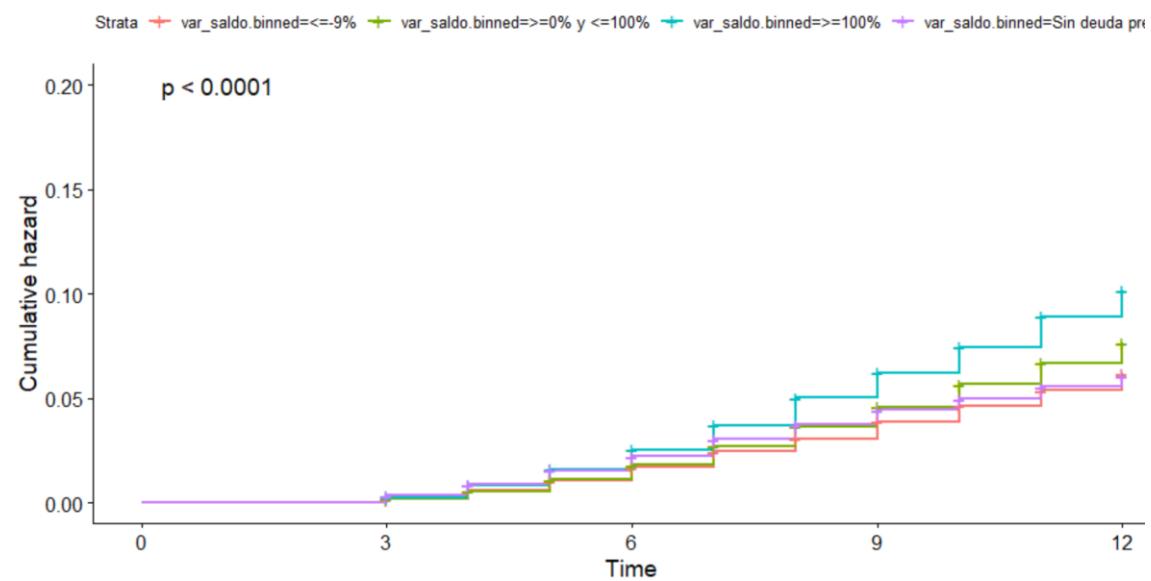
Porcentaje de utilización de líneas de crédito en el sistema financiero



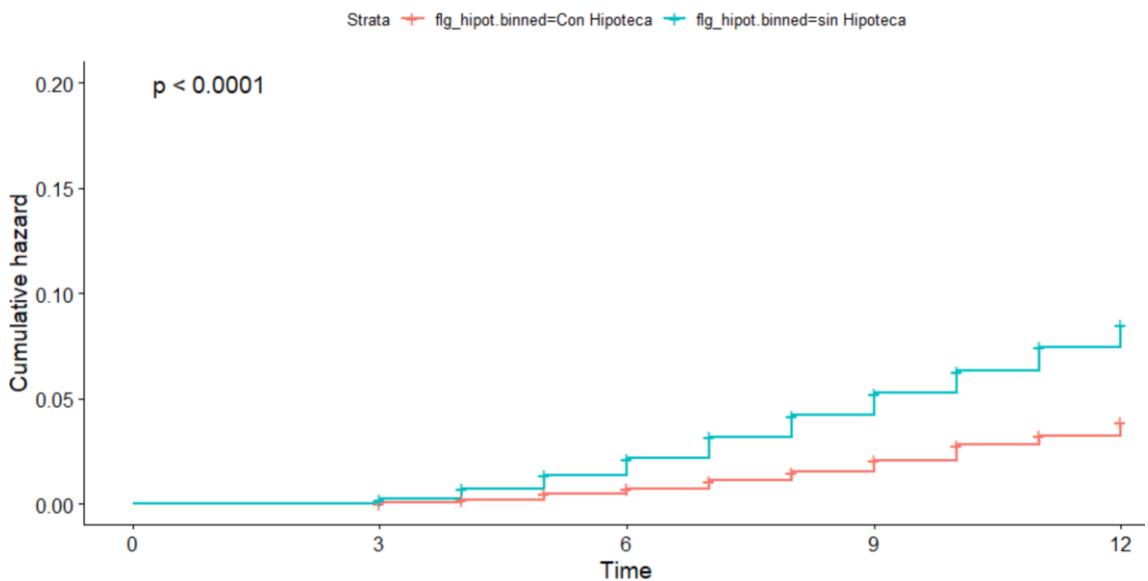
Variación de entidades acreedoras en el sistema financiero en los últimos 6 meses



Variación porcentual del saldo deudore en el sistema financiero en los últimos 12 meses



Flag de tenencia de crédito hipotecario en el sistema financiero



ANEXO 2: Gráficos Log - Log versus el tiempo de las variables incluidas en el modelo

Gráfico Log - Log de la variable ingresos

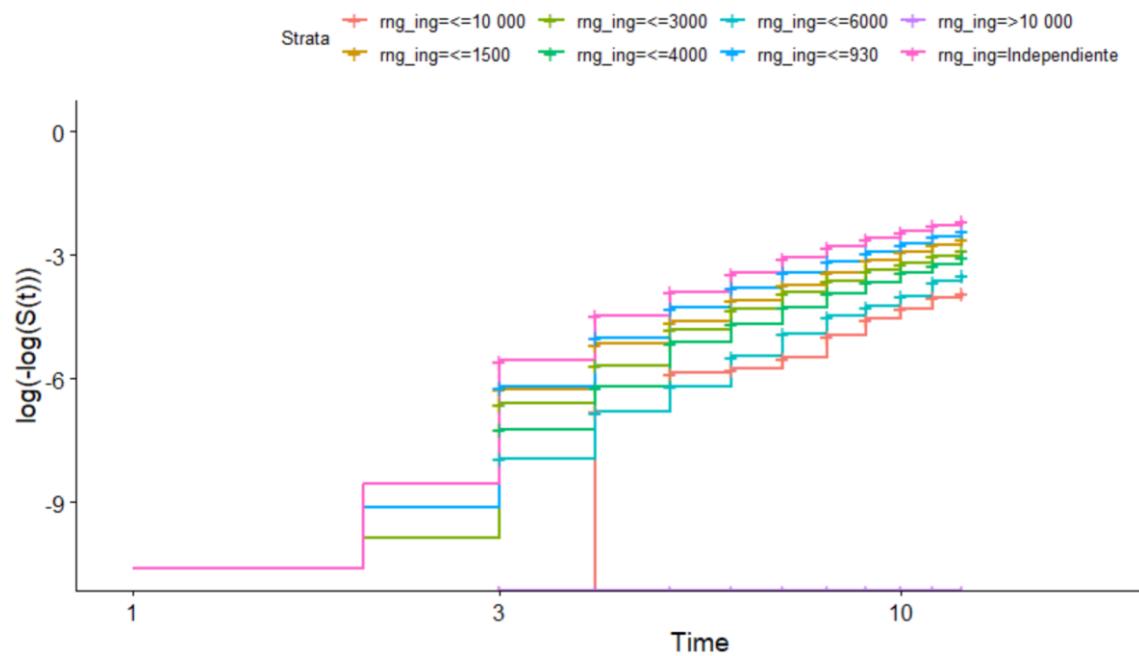


Gráfico Log – Log de la variable monto de líneas no utilizadas en el sistema financiero

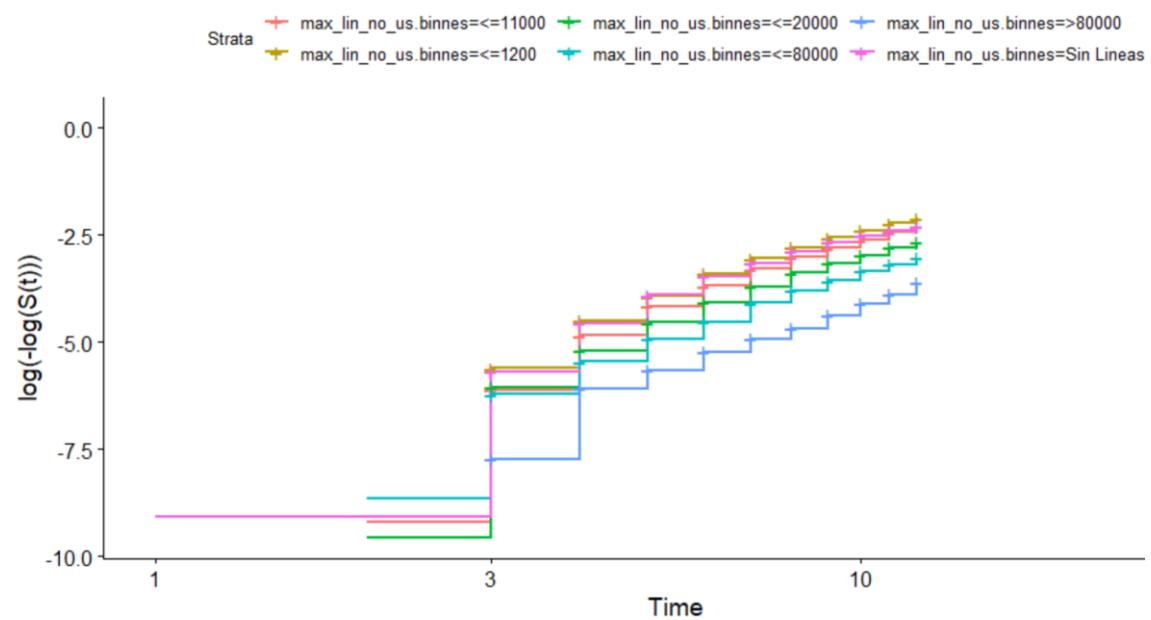


Gráfico Log -Log de la variable porcentaje de utilización de líneas

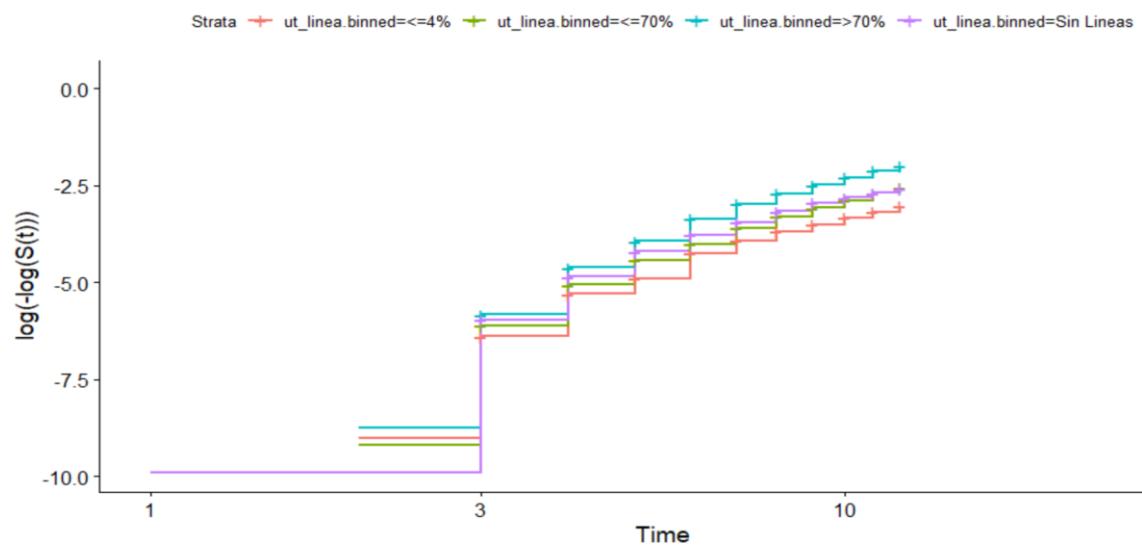


Gráfico Log-Log de la variable variación % de saldos deudores en el sistema financiero

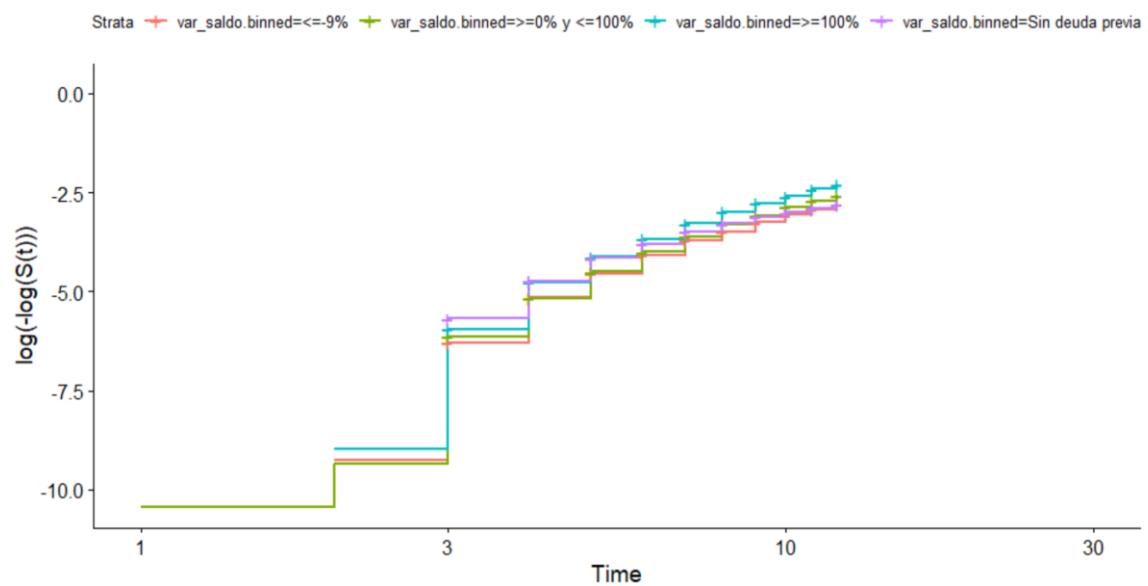
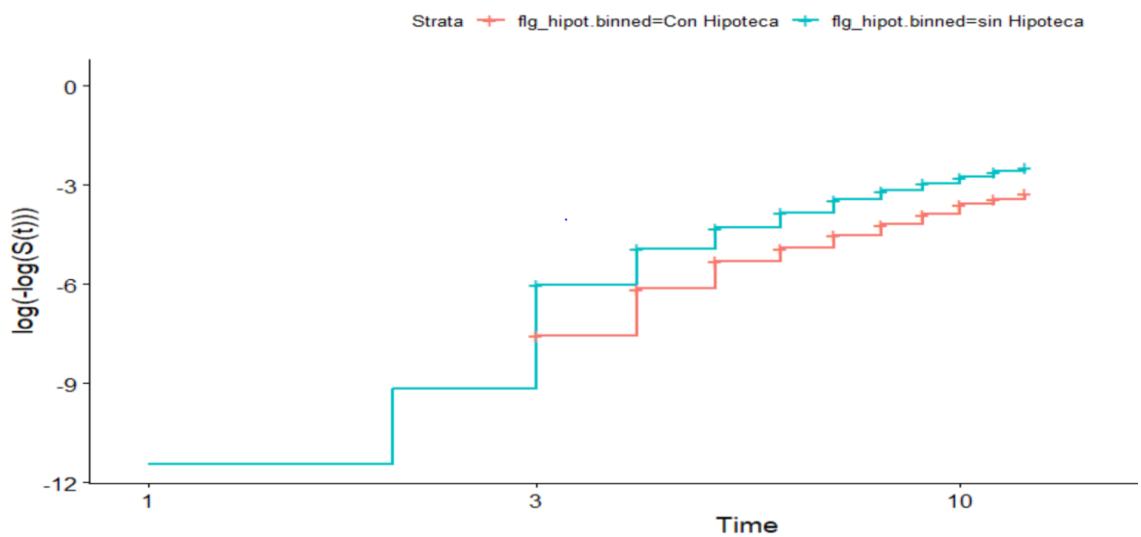


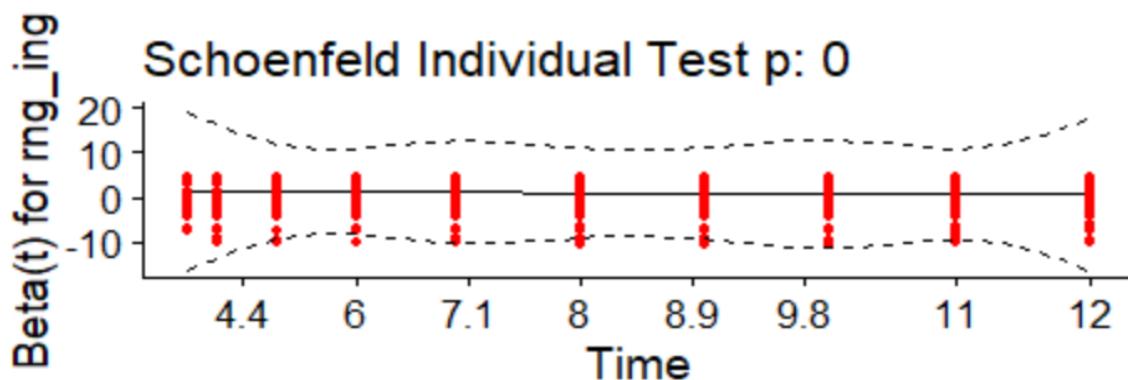
Gráfico Log-Log de la variable variación % de saldos deudores en el sistema financiero



ANEXO 3: Gráficos de residuos de Schoenfeld para las variables endógenas del modelo de regresión de Cox

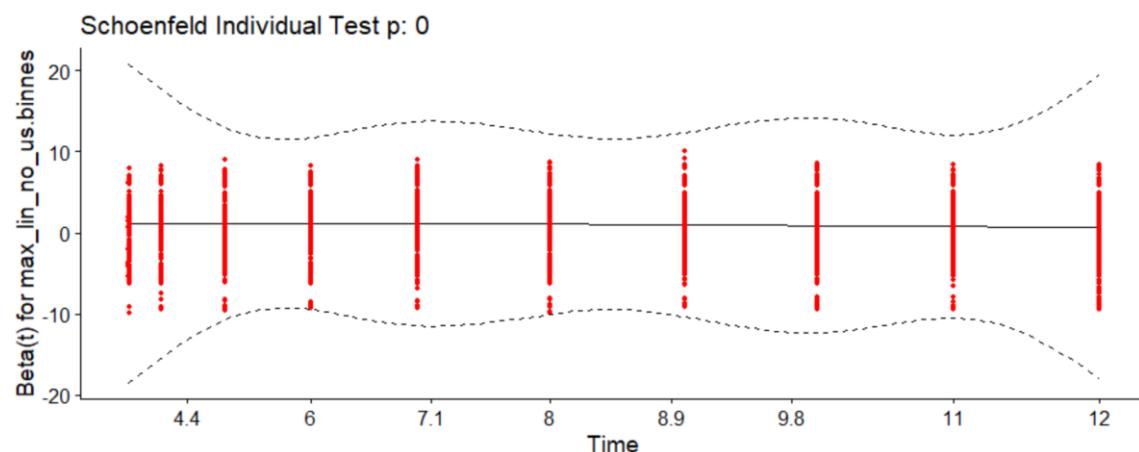
Ingresos

Global Schoenfeld Test p: 1.869e-15



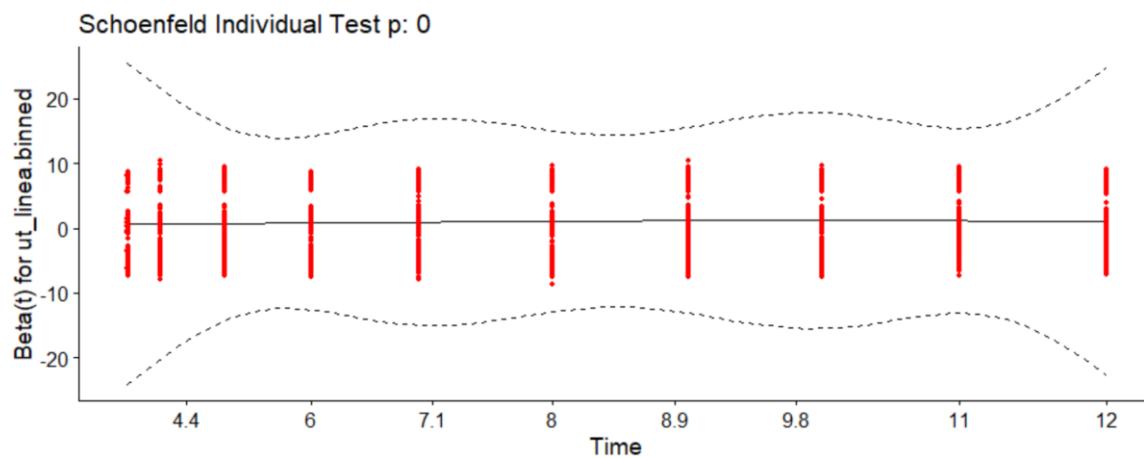
Máxima línea no utilizada

Global Schoenfeld Test p: 1.869e-15

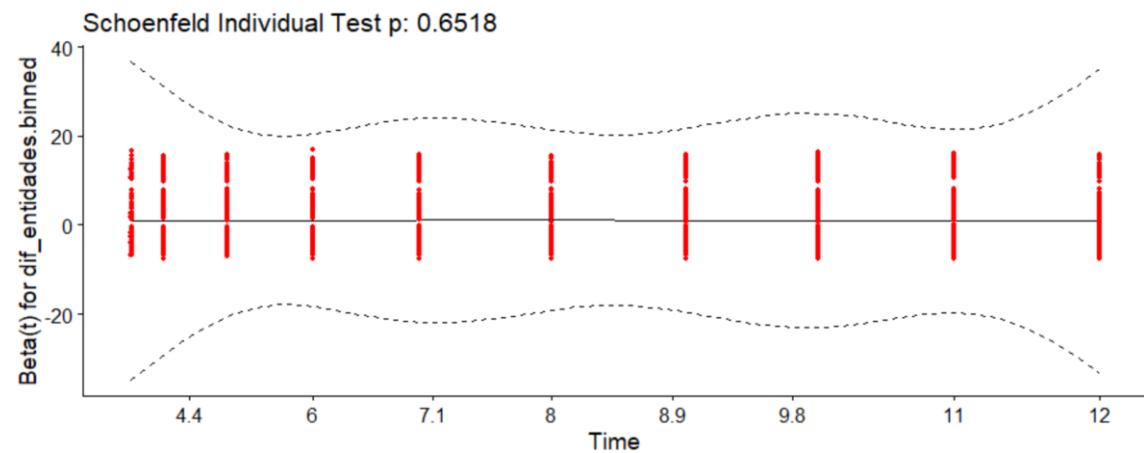


Porcentaje de Utilización de líneas

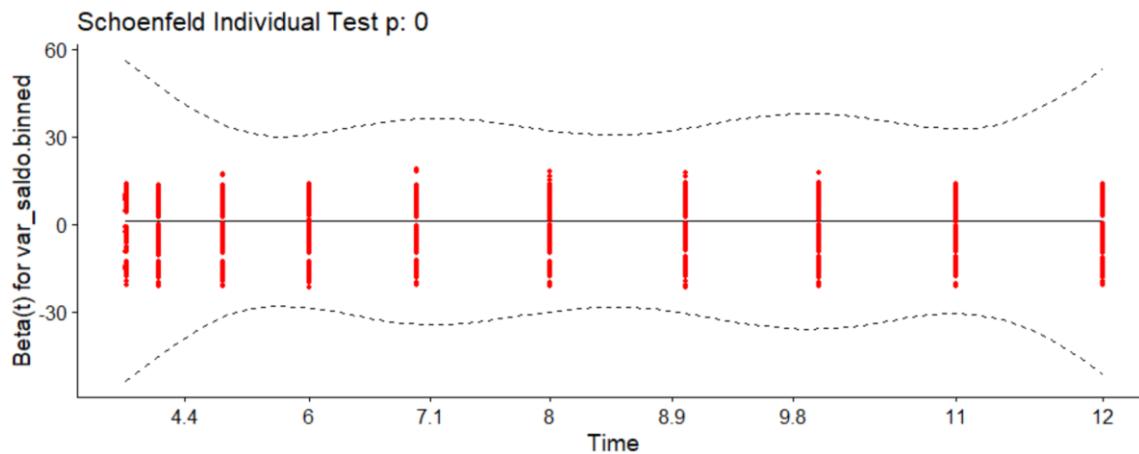
Global Schoenfeld Test p: 1.869e-15



Diferencia de entidades



Variación de saldos



Flag Crédito Hipotecario en el sistema financiero

