



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

ANÁLISIS DE COMPORTAMIENTO DE PREPAGO EN CRÉDITOS HIPOTECARIOS
DE CLIENTES BANCARIOS.

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL

DIEGO ANDRÉS RIVAS GIRALDO

PROFESOR GUÍA:

JOSE MIGUEL CRUZ GONZALEZ

MIEMBROS DE LA COMISIÓN:

CARLOS PULGAR ARATA

JUAN PABLO RISCO

SANTIAGO DE CHILE

2018

RESUMEN DE LA MEMORIA
PARA OPTAR AL TÍTULO DE:
INGENIERO CIVIL INDUSTRIAL
POR: Diego Andrés Rivas Giraldo
FECHA: 5/03/2018
PROFESOR GUÍA: SR. José Miguel
Cruz González

ANÁLISIS DE COMPORTAMIENTO DE PREPAGO EN CRÉDITOS HIPOTECARIOS DE CLIENTES BANCARIOS.

En el trabajo realizado, se buscó encontrar el mejor modelo econométrico para la estimación de la tasa de prepago para los clientes de un banco en particular.

Es importante poder calcular el prepago total debido a que actualmente de los créditos otorgados en el año 2009 un 42,9% ha sido prepagado, siendo esta cifra muy elevada con tendencia al alza debido al tiempo que le restan a los créditos vigentes.

Existen 4 principales causas de ejercer la opción de prepago, éstos son: rotación habitacional, tasa de interés de mercado, nivel de actividad económica y características de los créditos. Dado estas causas se trató de canalizar la mayor cantidad de efectos según las variables que teníamos a disposición, por lo cual las principales características de los créditos con los cuales se trabajó fueron: monto original, valor de la cuota, si el crédito presentó mora en algún período, el tipo de la casa a la cual se contrajo el crédito y su plazo residual.

Por el otro lado para canalizar el efecto de la actividad económica e interés de mercado se creó la variable nombrada refinanciamiento que compara, la tasa de mercado con la del crédito, de esta manera se puede conseguir la idea de cómo afecta la tasa de interés al crédito y al tener esta una alta correlación con la actividad económica se puede observar ese efecto también.

La tasa que se buscó explicar con los modelos logísticos estudiados fue de un 0,51% promedio mensual la que tiene asociada un monto de prepago de: 57.106.070 UF.

El modelo obtenido como el óptimo presenta significancia estadística del 95% para todas las variables que se encuentran en él y clasifica al 97,2% de los casos de forma correcta tomando como referencia el análisis fuera de la muestra, por lo cual se cree que se logró un buen resultado para la estimación del comportamiento de prepago.

Agradecimientos

Se agradece principalmente a mi familia por el apoyo constante durante este largo proceso. A mi profe guía, coguía y los profesores de la sección del trabajo de título por su orientación y consejos, además de presionar cuando los resultados exigían.

También se agradece a los grupos de amigos que he formado durante todo el proceso, los del colegio, los de la u, los de la vida que con sus invitaciones a salir complicaron más de una vez los estudios y el trabajo de título.

Agradezco al grupo de difusión por cumplir un rol en mi crecimiento dentro de la universidad, en especial a la Mane por su carisma, por siempre estar ahí y ahora último su muestra de esfuerzo en todo sentido.

Tabla de Contenido

Agradecimientos	iii
Índice de Ilustraciones	vi
Índice de Tablas.....	vii
1. Introducción	1
1.1. Justificación del tema.....	1
1.2. Antecedentes.....	3
1.3. Objetivos	6
1.3.1 Objetivo General.....	6
1.3.2 Objetivos Específicos	6
1.4. Alcances	6
2. Marco Conceptual.....	7
2.1. Créditos Hipotecarios.....	7
2.1.1. Introducción	7
2.1.2 Tipos de Créditos Hipotecarios.....	9
2.2. Prepago de Créditos Hipotecarios.....	11
2.2.1. Prepago Óptimo	12
2.2.2. Prepago Sub-óptimo.....	12
2.2.3. Opción de prepago	13
2.3. Data Mining.....	13
2.4. Análisis Univariado.....	14
2.4.1. Segmentación	14
2.4.2. Test de Medias	16
2.4.3. Test X ² de Pearson.....	16
2.5. Análisis multivariante.....	17
2.5.1. Regresión Logística	17
2.7.2. Backtesting	17
3. Metodología	19
3.1. Procesamiento de datos.....	19
3.2. Variable Prepago	20
3.3. Análisis Exploratorio.....	21
3.4. Tasa de prepago por período	25
3.5. Segmentación.....	28
3.5.1. Monto Original	28
3.5.2. Valor Cuota.....	29
3.5.3. Spread Tasa y sus desfases.....	30
3.5.4. Plazo Residual(PR).....	32
3.5.5. Tipo de Tasa	33
3.5.6. Mora.....	34
3.5.7. Correlación Variables	35
3.6. Modelos.....	36
4. Interpretación de Resultados	40
5. Conclusiones.....	42

5.1. Acápites I: Recomendaciones.....	43
6. Bibliografía	44
7. Anexos.....	45

Índice de Ilustraciones

<i>Ilustración 1: Créditos prepagados sobre totales</i>	2
<i>Ilustración 2: Porcentaje Créditos Prepagados</i>	3
<i>Ilustración 3: Colocaciones en vivienda Julio 2017</i>	9
<i>Ilustración 4: Esquema segmentos créditos hipotecarios</i>	10
<i>Ilustración 5: Evolución temporada del financiamiento habitacional bancario segmentado por tipo de producto</i>	11
<i>Ilustración 6: Ejemplificación Clusters</i>	15
<i>Ilustración 7: Número de Créditos Totales otorgados por Año</i>	22
<i>Ilustración 8: Información Créditos en UF</i>	22
<i>Ilustración 9: Comparación Nivel de Colocaciones(MMM)</i>	23
<i>Ilustración 10: Tasa Cartera Vivienda</i>	23
<i>Ilustración 11: Comparación Tasas de Prepago</i>	26
<i>Ilustración 12: TPT Promedio Cartera Vivienda</i>	27
<i>Ilustración 13: Segmentación por Monto Otorgado</i>	29
<i>Ilustración 14: Segmentación Valor Cuota</i>	30
<i>Ilustración 15: Segmentación Plazo Residual</i>	33
<i>Ilustración 16: Segmentación Tipo Tasa</i>	34
<i>Ilustración 17: Predicción Modelo</i>	41
<i>Ilustración 18: Test de Medias Monto Original</i>	46
<i>Ilustración 19: Test de Medias Valor Cuota</i>	47
<i>Ilustración 20: Test de Medias Spread Tasa</i>	47
<i>Ilustración 21: Test de Medias Spread Tasa Desfase 1</i>	48
<i>Ilustración 22: Test de Medias Spread Tasa Desfase 2</i>	48
<i>Ilustración 23: Test de Medias Spread Tasa Desfase 3</i>	48
<i>Ilustración 24: Test de Medias Spread Tasa Desfase 4</i>	49
<i>Ilustración 25: Test de Medias Spread Tasa Desfase 5</i>	49
<i>Ilustración 26: Segmentación Spread sin Desfase</i>	50
<i>Ilustración 27: Segmentación Spread Tasa 1 Desfase</i>	50
<i>Ilustración 28: Segmentación Spread 2 Desfases</i>	51
<i>Ilustración 29: Segmentación Spread Tasa 3 Desfases</i>	51
<i>Ilustración 30: Segmentación Spread Tasa 4 Desfases</i>	52
<i>Ilustración 31: Segmentación Spread Tasa 5 Desfases</i>	52
<i>Ilustración 32: Test de Medias Plazo Residual</i>	53
<i>Ilustración 33 Test de Medias Tipo Tasa</i>	53
<i>Ilustración 34: Test de Medias Mora</i>	54
<i>Ilustración 35: Segmentación Mora</i>	54
<i>Ilustración 36: Historia Mora por Segmentos</i>	55

Índice de Tablas

<i>Tabla 1: Comparación número de operaciones HIP300 y PLC300 a septiembre 2017</i>	20
<i>Tabla 2: Estadísticas Descriptivas de la Base de Datos</i>	24
<i>Tabla 3: Datos Grupos</i>	25
<i>Tabla 4: Estadísticas para la tasa de prepago para TPr1</i>	26
<i>Tabla 5: Estadísticas para la tasa de prepago TPr2</i>	26
<i>Tabla 6: Montos asociados al Prepago</i>	27
<i>Tabla 7: Estadísticas Segmentación Monto Original</i>	29
<i>Tabla 8: Estadísticas Segmentación Valor Cuota</i>	30
<i>Tabla 9: TPT Segmentación sin Desfase</i>	31
<i>Tabla 10: TPT Segmentación Spread Tasa 1 Desfase</i>	31
<i>Tabla 11: TPT Segmentación Spread Tasa 2 Desfase</i>	31
<i>Tabla 12: TPT Segmentación Spread Tasa 3 Desfases</i>	31
<i>Tabla 13: TPT Segmentación Spread Tasa 4 Desfases</i>	32
<i>Tabla 14: TPT Segmentación Spread Tasa 5 Desfases</i>	32
<i>Tabla 15: Estadísticas Segmentación Plazo Residual</i>	33
<i>Tabla 16: Estadística Segmentación Tipo de Tasa</i>	34
<i>Tabla 17: Estadística Segmentación Mora</i>	35
<i>Tabla 18: Matriz de Correlaciones</i>	35
<i>Tabla 19: Estadísticos Modelos</i>	37
<i>Tabla 20: Resumen Modelos Método de Wald</i>	38
<i>Tabla 21: Modelo Óptimo</i>	39
<i>Tabla 22: Participación de Mercado del Financiamiento de Viviendas Segmentado por Institución Financiera a junio del 2017</i>	45
<i>Tabla 23: Montos en \$MM de la evolución temporal de créditos hipotecarios en stock del sistema</i>	46
<i>Tabla 24: Porcentajes de la evolución temporal del stock de los créditos hipotecarios</i>	46

Capítulo 1

1. Introducción

El sistema financiero y la banca en particular tienen un rol clave en la economía, no sólo ayudándola a expandirse, sino también permitiendo que los shocks reales se puedan absorber y/o diluir, y no se agraven o exacerben en el tiempo.

Los créditos son una forma de que un cliente acceda a una cierta suma de dinero que debe pagar a un plazo y a una tasa de interés determinada en la concepción del contrato del crédito. En particular los créditos hipotecarios son un instrumento que permite al cliente, acceder al financiamiento de una vivienda, bienes de consumo, del mercado automotriz, entre otras cosas.

Factores como la disminución de la tasa de interés, aumento en los ingresos de las personas y los incentivos tributarios, han generado que las colocaciones para el sector habitacional, represente a junio de 2017 a un 27.81% del total de colocaciones del sistema bancario. Esta participación, a enero de 2009, ascendía a un 23.5%, registrando un crecimiento del 4,31% de su participación en las colocaciones totales en los últimos ocho años. Por otro lado, la tasa de interés ha caído desde un 6,87% a 3,61% para créditos a más de 20 años tomando las mismas fechas de referencia. Dado que en el presente trabajo se trabajó con una base de datos que comienza con una ventana del tiempo desde enero del 2009 es importante ver cómo ha evolucionado la tasa y el nivel de colocaciones hasta el presente.

Entre las flexibilidades que estos instrumentos de largo plazo poseen, se encuentra la posibilidad de prepago parcial o total del crédito hipotecario adquirido, siendo el último caso el que se estudiará en el presente estudio, buscando las principales variables intrínsecas del crédito que pueden afectar el comportamiento de prepago.

1.1. Justificación del tema

Los créditos hipotecarios tienen la opción de recibir un prepago parcial o total del crédito hipotecario adquirido. De esta manera le genera a la institución que provisionó el crédito, incertidumbre, debido a que no se sabe cuál es la probabilidad que algún crédito se prepague, siendo importante saber la duración real de este debido a que el crédito genera beneficios según su duración.

El prepago genera un descalce entre la duración entre pasivos y activos, esto le genera un riesgo al banco debido que ellos se proyectan con los flujos futuros a recibir, y dado estos flujos futuros buscan igualar la duración entre sus activos y pasivos.

A continuación, se profundiza en el principal problema que genera el prepago, este es la incertidumbre que se genera en los flujos futuros; el cual lo veremos tomando un activo (el crédito) y un pasivo (el financiamiento de ese crédito) calzados. Cada uno se contrae a una determinada tasa, en el momento se realiza un prepago se rompe ese calce, y se genera el descalce anteriormente mencionado. Las opciones que tiene el banco cuando se genera esta ruptura, es re-colocar o re-invertir, pero en estos casos las condiciones de tasa ya cambiaron, por lo tanto, en esa re-inversión se podría colocar a una mayor o menor tasa, de esta manera se podría mejorar o deteriorar la

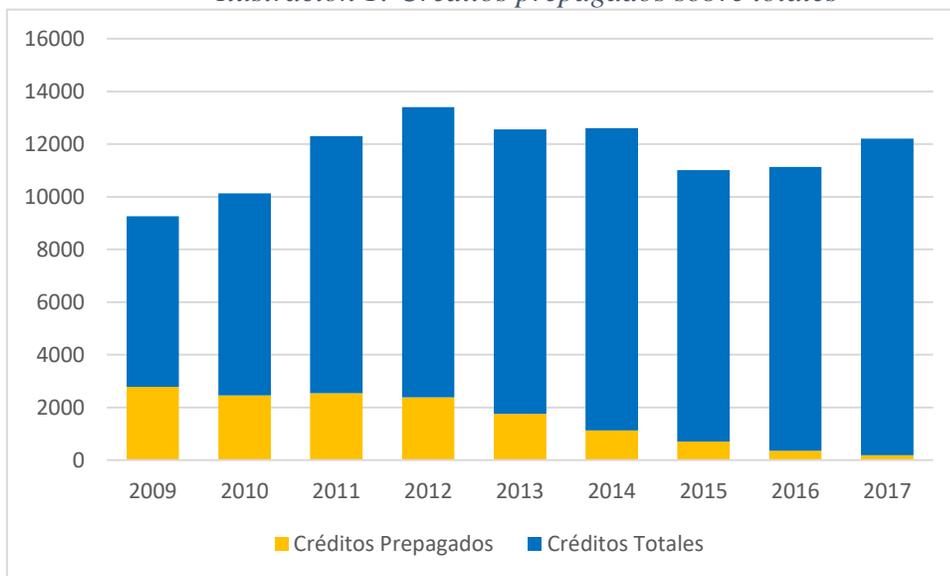
posición. La incertidumbre existente se traduce en la práctica en riesgo, y este riesgo generado por el movimiento de las tasas a las cuales se otorgaron los créditos, puede ser muy grave para el banco, y en consecuencia para el sistema financiero. Mientras mayor la magnitud del descalce mayor es el problema para la institución financiera, siendo el principal dolor para ella, re-colocar a una peor tasa de la cual se contrajo el contrato.

Es importante para los bancos poder identificar el comportamiento de prepago, para así estimar el vencimiento efectivo del producto. Esto se puede estimar con modelos basados en metodologías que sean conceptualmente sólidas, además de matemática y estadísticamente robustas. De esta forma, podrán ajustar, la duración de activos y pasivos con el fin de minimizar el riesgo antes cambios en las tasas.

Por lo anteriormente mencionado, resulta interesante desarrollar un modelo que identifique el comportamiento de prepago.

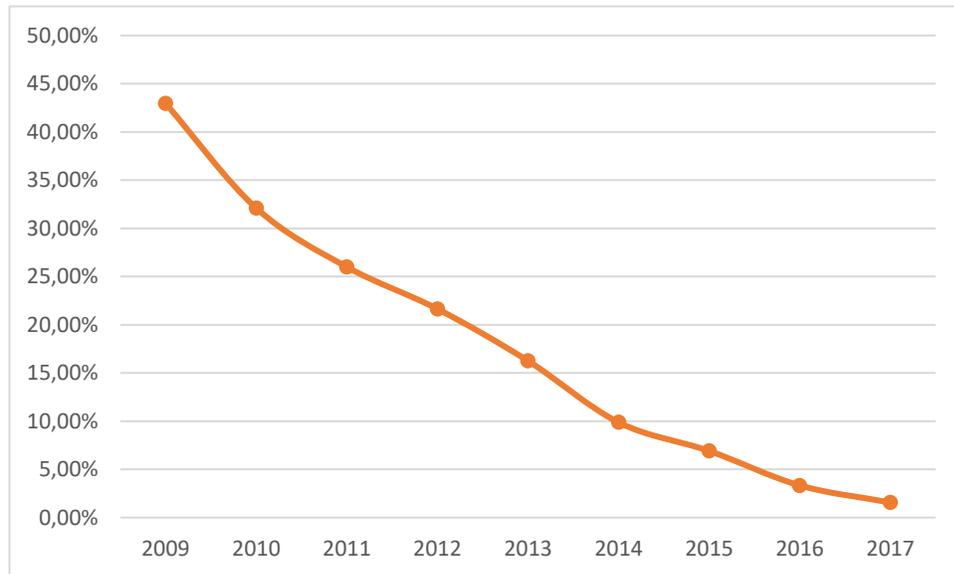
Para complementar la justificación a continuación se presenta el porcentaje de los créditos prepagados a partir desde el año 2009 dada su fecha de apertura, debido a que la base entregada contiene la información a partir de esa fecha. Se puede observar la cantidad de créditos otorgados por año y ver qué porcentaje de ellos se han prepagado hasta septiembre del 2017 que es el último período que se tiene registro en la base de datos. Para simplificar, cada vez que se mencione hasta la fecha se hará referencia a septiembre del 2017.

Ilustración 1: Créditos prepagados sobre totales



Fuente: Elaboración propia a partir de los datos de la base.

Ilustración 2: Porcentaje Créditos Prepagados



Fuente: Elaboración propia a partir de los datos de la base.

En las Ilustraciones 1 y 2 se puede observar la cantidad de créditos prepagados dada su fecha de apertura con la información tenida hasta la fecha. Es importante notar que dados los créditos originados en el año 2009 actualmente un 42,95% de estos han sido prepagados. La cifra mencionada va en aumento debido a que existen muchos créditos vigentes que tendrán un comportamiento de prepago asociado en el futuro. Por lo cual estudiar el comportamiento de prepago resulta de vital importancia debido a la gran magnitud de prepagos que se generan.

El decaimiento observado en la Ilustración 2 se debe a que, dado que los créditos han tenido menor tiempo de vigencia en la cartera, la cantidad de prepagos observada es menor. Cabe mencionar que los créditos con los cuales se trabajó tienen una media de duración de 241,33 meses lo que equivale a 20,11 años aproximadamente, siendo lo más otorgados los créditos a 20 años. Dada la información anterior, al grueso de los créditos de la base le quedan más de 11 años para recibir el comportamiento de prepago.

Para complementar esta sección, cabe resaltar que en promedio los créditos prepagados se realizan 5,19 años antes de su término con una desviación estándar de 3,34 años.

1.2. Antecedentes

En esta sección se busca detectar variables relevantes que podrían afectar el comportamiento del prepago, dado que el trabajo va en la línea de poder caracterizar individualmente a los clientes que prepagan. Se buscó características intrínsecas de los créditos para poder explicar la decisión de prepago, pero también era importante analizar cómo factores macroeconómicos podían incidir en este comportamiento. Los factores que se consideren relevantes serán incluidos en el modelo econométrico planteado en una sección posterior.

Para realizar esta sección se apoyó en el trabajo realizado por Cristian Majluf en el año 2007, donde su tema de memoria fue el de “Desarrollo de un modelo de prepago de letras hipotecarias para el banco BICE”.

Majluf desarrollo un análisis, para tener una base teórica sobre variables endógenas (propias del crédito) y exógenas (en este caso, macroeconómicas) que puedan tener un impacto en la decisión del prepago hipotecario, la cual se complementa y se presenta a continuación.

Basándonos en el trabajo antes mencionado, bibliografía estudiada y en particular el trabajo de Mariano Selvaggi, que distingue cuatros tipos de efectos (Selvaggi, Agosto de 2000) para el prepago.

A continuación, se muestra en detalle cada uno de estos efectos:

a. Rotación habitacional

Estudios evidencian (De Ovando, 2005) en las tasas anuales de mudanzas o rotación, que las familias en intervalos de tiempo más o menos regulares, cambian de vivienda. Esta movilidad habitacional propia de las sociedades actuales se da por diversas razones, entre las que se destacan las puramente laborales, el deseo de habitar una vivienda más lujosa o una búsqueda por ajustar las necesidades habitacionales familiares (tamaño familiar principalmente) a la realidad de dicho hogar. Entre otros aspectos importantes a destacar, que afectan este tipo de variable, está el gusto por determinadas zonas geográficas de un país o ciudad, así como también, factores socioculturales. Entre estos últimos, los divorcios o la edad de independencia de los hijos apresuran o retrasan el endeudamiento de nuevas generaciones a través de este tipo de créditos.

La relación existente entra la rotación habitacional y los prepagos es que cuando el titular de una hipoteca cambia de casa, por lo general, vende la que habita y con eso paga el saldo remanente del crédito vigente, para tomar otro para la nueva vivienda, de esta manera termina la hipoteca de forma anticipada.

b. Tasas de interés en el mercado

El costo financiero total de las hipotecas es otro de los principales factores que explican los prepagos. En efecto, cuando el costo de obtener un nuevo préstamo se ubica por debajo de la tasa pactada, existirán incentivos para que el deudor cancele su deuda para tomar una nueva a menor costo, ahorrándose así, el diferencial de tasas por el plazo remanente. Este comportamiento, que en mercados más desarrollados toma vital importancia, se denomina “refinanciamiento”. Matemáticamente este efecto se puede contextualizar de la siguiente forma:

$$VP + Costos < \sum_T \frac{C_t}{(1 + r_t)^t}$$

Fórmula 1

Donde VP, corresponde al valor presente de la deuda remanente; Costos, representa a los costos relativos al prepago y eventualmente, de emitir una nueva deuda por el monto remanente; C_t ,

representa el monto a pagar en el período t ; r_t , la tasa de descuento en t ; y T , la cantidad de períodos restantes al vencimiento del crédito.

La fórmula 1 nos muestra que existe el incentivo de prepago cuando al pagar el saldo remanente del crédito hoy cuando el valor presente del crédito más los costos transaccionales de realizar esta operación es menor que pagar las cuotas restantes en los plazos establecidos.

Tanto la evidencia empírica como teórica muestran que el incentivo al refinanciamiento sólo se hace presente a partir de un cierto spread mínimo sobre la caída en la tasa de interés. Esto, a causa de que tanto el prepago de un crédito como una nueva emisión, conlleva una serie de gastos fijos y variables, siendo estos traspasados directamente al consumidor. Investigaciones realizadas en EE. UU. (Vucina Ljubetic, 2004) demuestran que los prepagos se incrementan sustancialmente sólo a partir de un diferencial mínimo de 200 puntos base. Para captar estos comportamientos con mayor precisión, se suele trabajar con tasas rezagadas entre dos y cinco períodos, entendiéndose que el proceso de gestión de la nueva hipoteca demora cierto tiempo, y, por lo tanto, la tasa de prepago observada hoy, en realidad responde a incentivos existentes en períodos anteriores.

Asimismo, a menudo los prepagos se desencadenan luego de que las tasas de mercado, tras tocar un piso, comienzan su tendencia alcista. Este hecho, en principio contradictorio, se debe a que ciertos deudores demoran demasiado sus prepagos, intentando sacar la mayor ventaja posible del spread existente. Así pues, cuando estiman que el piso del mercado ha sido finalmente alcanzado, se lanzan en forma retardada a la refinanciación de sus créditos.

c. Nivel de actividad económica

Otra variable macroeconómica frecuentemente señalada en la literatura (LaCour-Little, 2002), es el efecto positivo de esta variable (actividad económica) sobre los prepagos. Sin embargo, este efecto es indirecto debido a su alta correlación con las tasas de interés, de rotación, y con el volumen de transacciones inmobiliarias. Se asume, además, que la combinación de tasas bajas y crecimiento económico, facilitan la creación de empleos y el aumento de los ingresos personales. Esto a su vez dinamiza la rotación laboral y habitacional. El aumento de los ingresos personales se denomina efecto ingreso, pero en el presente trabajo no se evaluará esta variable por falta de información. Cabe resaltar que es una variable considerada continuamente en la bibliografía estudiada.

d. Características de los créditos

Finalmente, uno de los aspectos más significativos en el comportamiento futuro de prepago de los créditos hipotecarios, son las características propias de él. En ese sentido, de las identificadas como las más importantes se destacan: el tipo de moneda y tasa de interés pactada; el destino (compra, remodelación, construcción, etc.); el monto y plazo convenido, el tiempo(edad) que lleva el préstamo, la localización de la propiedad hipotecada; el nivel de ingresos del deudor, la relación entre el monto del crédito y su valor de mercado, y la relación cuota-ingresos. En lo particular de este trabajo el tipo de moneda no se toma en cuenta, debido a que todos los créditos con los cuales se trabajó están en pesos chilenos, también se centró en el destino de compra de primera vivienda por lo cual otros tipos de créditos hipotecarios no fueron tomados en cuenta. Por otro lado, la tasa de interés pactada es de suma importancia ya que es una de las primeras señales al compararla con la de mercado, y que puede generar incentivos al refinanciamiento.

Estas características serán las principales agregadas al modelo, donde se buscará ver que variables tienen una mayor influencia en la decisión de anticipar las cuotas de un crédito.

1.3. Objetivos

1.3.1 Objetivo General

El objetivo de este trabajo es estimar el mejor modelo econométrico de comportamiento de prepago de los clientes del banco, para poder estimar la probabilidad de prepago de ellos.

1.3.2 Objetivos Específicos

Los objetivos específicos del presente trabajo son identificar comportamiento de prepago de los clientes de la institución bancaria desde el año 2009 hasta septiembre del 2017, estimando la tasa de prepago por el periodo definido

También se analizarán las variables relevantes en la decisión de prepago, trabajando con las variables proporcionadas por la institución bancaria y otras creadas a partir de esas.

Por otro lado, se quiere cuantificar el efecto de prepago, es decir, ver los montos asociados al prepago generado por los clientes.

1.4. Alcances

Dentro de los alcances de la presente memoria se encuentra que la ventana de tiempo con la cual se trabajó fue desde enero del 2009 a septiembre del 2017, también se trabajó con características intrínsecas de los créditos y no demográficas de los clientes. En este trabajo se estudio el prepago total del crédito, que se define como un pago total de las cuotas remanentes del crédito.

Se asumió amortización francesa para todos los créditos estudiados. Los créditos considerados son habitacionales, dejando de lado los créditos para fines generales. Estos son para construcción, remodelación, compra de terrenos. Por ende, el alcance del presente trabajo es únicamente habitacional.

La masa crítica de los créditos de tipo mutuos endosables es marginal, por lo cual se trabajó de forma genérica para los tres tipos de créditos que se detallan en el capítulo 2.

Capítulo 2

2. Marco Conceptual

Este capítulo permite al lector entender mejor el tema que se abordó y comprender los conceptos que se relacionan a la investigación. Sirve para crear la noción de como se comportan los créditos hipotecarios y facilitar al lector la contextualización de los conceptos utilizados durante el trabajo. También se muestra las técnicas utilizadas para realizar la investigación.

En un comienzo se enuncian características propias de los créditos para comprender como funcionan. Se detallan conceptos relacionados al prepago para lograr un mejor entendimiento de este comportamiento. Luego se presenta el mecanismo con el cual se desarrollo la investigación, en primer lugar, se comenzó con un análisis univariable. Este análisis fue realizado para poder determinar en detalle el comportamiento individual de cada variable estudiada y crear una noción más clara de cómo se podría comportar cada una. Para esto se realizaron cluster de los clientes para entender sus comportamientos de prepago.

A continuación, se comenzó con el análisis multivariado, donde se estudiaba el comportamiento de prepago relacionando las variables estudiadas en el análisis univariado en conjunto. Este análisis permitió comprender con mayor robustez el problema. Para lograr el análisis univariado se utilizó la técnica de la regresión logística que nos permitió separar los créditos en los cuáles no han recibido el comportamiento con los que sí, esto a partir de una variable “dummy”.

Para concluir con el marco conceptual se detalla el proceso cual el se comprueba el nivel de predicción del modelo obtenido a partir del análisis las diversas regresiones logísticas.

En la literatura se puede apreciar que existen diversas formas de abarcar el problema estudiado. Una de ellas es el análisis de supervivencia donde se tiene como objetivo modelar el tiempo en que se tarda que ocurra un determinado suceso (en este caso sería el prepago). Se prefirió utilizar la regresión logística dado que era el método más utilizado para estudiar el comportamiento de prepago según la literatura revisada.

2.1. Créditos Hipotecarios

2.1.1. Introducción

Los créditos hipotecarios son préstamos de mediano y largo plazo, adquiridos por personas naturales y empresas, y emitidos por distintas instituciones financieras. Estos pueden ser solicitados por diversos motivos; compra, ampliación o construcción de viviendas, compra de sitios, oficinas o locales comerciales, o para libre disponibilidad.

La hipoteca es un derecho de garantía, que se constituye para asegurar el cumplimiento de una obligación (normalmente de pago de un crédito o préstamo), que confiere a su titular un derecho de realización de valor de un bien (generalmente inmueble), por lo tanto, la hipoteca se concibe como un contrato entre dos partes.

Existen diferentes formas para realizar la devolución de capital del préstamo. La más conocidas son:

- El sistema francés
- El sistema americano
- El sistema alemán

Para el presente trabajo se asumió el sistema francés como referencia debido a que el grueso de los créditos con los cuales se trabajó posee este sistema de devolución de capital y será el que se detallará a continuación.

Es el modelo más común, en este sistema son las cuotas son fijas, salvo que el tipo de interés sea variable, en cuyo caso, será necesario repetir los cálculos cada vez que se fije el valor del interés.

La mayor parte de los intereses se pagan durante las primeras cuotas del préstamo, ya que el devengo de intereses a cada vencimiento se calcula únicamente en función de dos variables: el tipo de interés aplicado y capital pendiente de amortizar. Por lo tanto, al adeudar una mayor cantidad, se pagan más intereses en cada vencimiento de cuota, y conforme se vaya amortizando capital, el devengo mensual de los intereses será menor.

De esta manera, para el sistema francés existe la cuota de amortización de la hipoteca que se divide en dos partes, una son los intereses mientras que la otra es la amortización del capital.

Tenemos que la cuota hipotecaria se calcula de la siguiente manera:

$$Cuota = \frac{\text{capital} \cdot \text{interés}}{100 \cdot \left(1 - \left(1 + \frac{\text{interés}}{100}\right)^{-\text{plazo}}\right)}$$

Siendo el interés el valor en porcentaje para cada periodo, es decir, mensual.

Dado que la tasa a la que se fija el contrato es nominal y anual, esta hay que transformarla a mensual de la siguiente manera:

$$i_{\text{mensual}} = \frac{i_{\text{anual}}}{12}$$

Entonces dada que la cuota de amortización de la hipoteca que se divide en dos partes, caracterizamos cada una de ellas como:

$$\text{Intereses} = \frac{\text{capital pendiente} \cdot \text{tipo de interés}}{12}$$

Y la amortización de la cuota como:

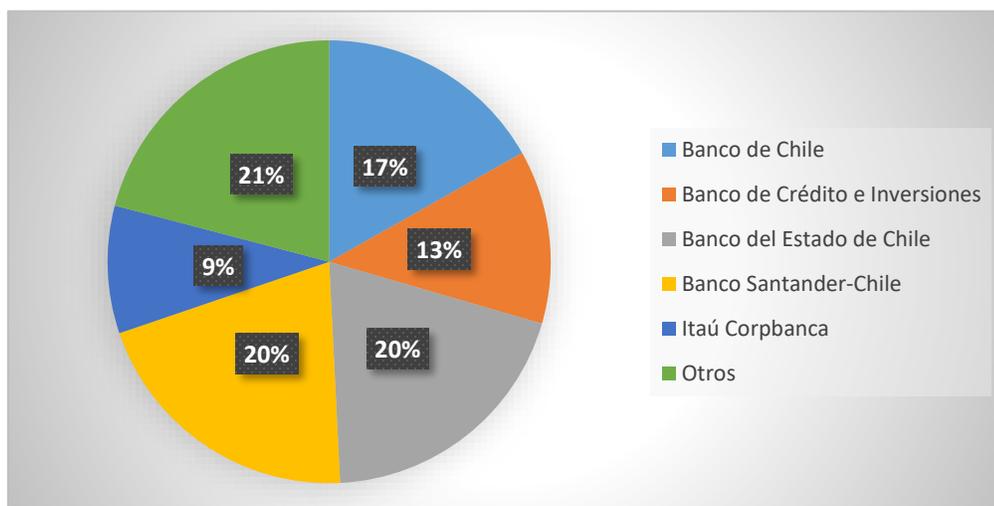
$$\text{Amortización por cuota} = \text{Cuota} - \text{Intereses}$$

Por lo tanto, en cada periodo se comienza a pagar menos intereses y aumentar la amortización por cuota.

Se presenta una breve explicación de cuáles son los otros mecanismos de la devolución de capital. Se tiene el sistema americano en el cual se tiene un pago periódico de intereses y sólo se amortiza el capital a la finalización del préstamo. Por otro lado, se tiene el sistema alemán de amortización, el cual se caracteriza por pagar los tipos de interés a de manera anticipada en cada cuota. Las cuotas para este tipo de amortización van disminuyendo a lo largo del tiempo, y los intereses se calculan según el saldo pendiente.

A continuación, se detalla en la Ilustración 3 las principales instituciones bancarias participantes en el mercado de vivienda. Este gráfico fue construido a partir de la información proporcionada por la SBIF (superintendencia de bancos e instituciones financieras) en su reporte de información financiera mensual.

Ilustración 3: Colocaciones en vivienda Julio 2017



Fuente: Elaboración propia a partir de datos de la SBIF

En el anexo A, se puede apreciar el detalle para los 20 bancos e instituciones financieras del sistema, los datos presentados son a junio del 2017. En el anexo A también se puede observar que cerca del 70% de las colocaciones para vivienda, corresponde a los bancos Santander, del Estado, de Chile y BCI. Esto denota una alta concentración de mercado en este producto financiero.

2.1.2 Tipos de Créditos Hipotecarios

Para caracterizar los créditos hipotecarios, se muestra en Ilustración 4, los tres tipos de créditos disponibles en el mercado chileno. La principal diferencia entre ellos es el modo en que una institución financiera costea un crédito solicitado por una persona natural o una empresa.

Ilustración 4: Esquema segmentos créditos hipotecarios



La clasificación mostrada en la Ilustración 4, posee una segmentación adicional en cuanto al giro del crédito otorgado, es decir, cada uno de estos puede subdividirse en dos categorías; vivienda; y fines generales, siendo la primera (vivienda), un crédito diseñado para personas naturales que buscan financiar sus requerimientos habitacionales de vivienda, y el segundo (fines generales), créditos otorgados para actividades de tipo profesional. En el presente trabajo se centrará en las personas naturales buscando financiar su vivienda, es decir, solo créditos habitacionales.

En el caso de las Letras de Crédito Hipotecario(LCH), para financiar el préstamo se genera un bono llamado letra hipotecaria, este es generado por la institución financiera que esta realizado el préstamo del dinero solicitado. El bono puede ser transado en la Bolsa de Valores o ser adquirida por el propio banco o un tercero relacionado. El monto otorgado al crédito quedará definido por el precio al cual se transó la letra hipotecaria. Dado que el precio de transacción puede ser distinto al valor par de la letra, se genera una diferencia, debiendo quedar explícito en el contrato, cuál de las partes se hará cargo de ella, y de qué forma, por lo tanto, el valor final definido para el papel en cuestión es al cual se transo el papel en la Bolsa de Valores.

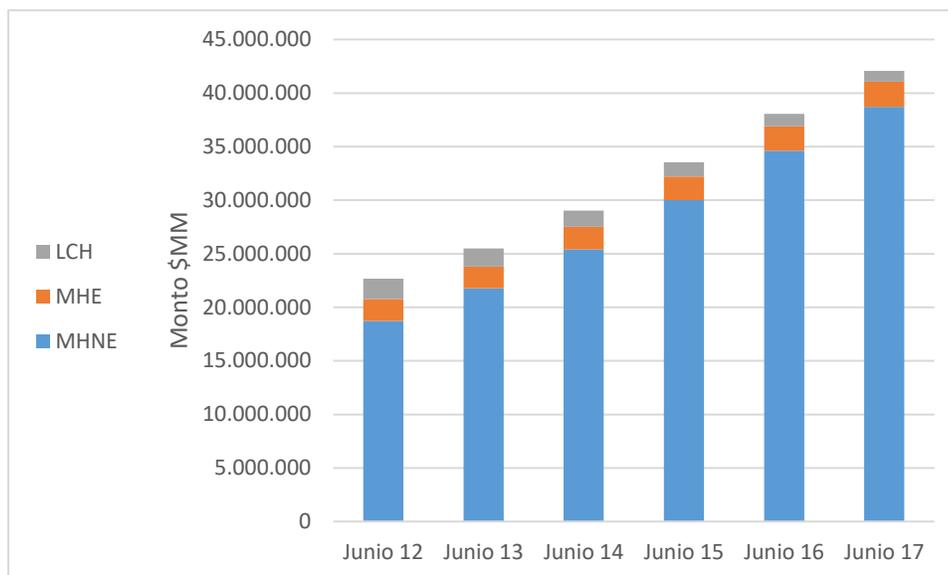
Los Mutuos Hipotecarios Endosables (MHE), a diferencia de las LCH, consiste en que se elimina esa diferencia de precios (estando endosada, puede ser transada al valor de emisión por un tercero), asumiendo el cliente, una tasa mayor a la que se podría obtener con una letra hipotecaria, de esta manera, el papel en este caso no cambia de valor al cual fue transado por las partes.

Finalmente, los Mutuos Hipotecarios No Endosables (MHNE), siendo los con mayor participación en el mercado, se diferencian de los otros en cuanto impide su transacción en el mercado, generando que los fondos para costear este préstamo deban ser otorgados completamente por el banco emisor del instrumento. En la presente memoria se modelara todos los créditos como MHNE dado que no tienen diferencia para el problema desarrollado el tipo de crédito.

Estos tres tipos de crédito han tenido una composición en las colocaciones totales del sector habitacional a lo largo de los últimos tres años, con clara tendencia al aumento por parte del financiamiento habitacional vía los MHNE.

Se presenta en la Ilustración 5 la evolución de los tres tipos de créditos hipotecarios a través del tiempo. Se contempla desde junio del 2012 hasta junio del 2017, para las colocaciones totales del sector habitacional. En el anexo B se puede contemplar con mayor detalle los montos y porcentajes de cada tipo de crédito hipotecario.

Ilustración 5: Evolución temporal del financiamiento habitacional bancario segmentado por tipo de producto



Fuente: Elaboración propia a partir de los datos de la SBIF

Se puede desprender de la información anterior que existe un aumento en los MHNE frente a MHE y LCH los cuales van en descenso. Esto se debe, a que los MHNE poseen un menor riesgo dado que los créditos de este tipo se fijan solo entre la institución financiera y la persona natural que contrajo el crédito, por lo cual no hay involucramiento de un tercero y menos una transacción en la bolsa que haga variar el valor del papel.

La particularidad que poseen estos instrumentos de renta fija (IRF), es que a pesar de que sus pagos futuros son conocidos, estos pueden ser total o parcialmente pre pagados. Es decir, que el crédito hipotecario puede ser liquidado antes de su fecha nominal de vencimiento. Esto genera que el plazo de vencimiento establecido en el contrato sea incierto, generando una disminución en la duración de estos instrumentos, lo que genera el problema de descalce entre pasivos y activos. Esto implica un riesgo para la institución, la cual fue mencionada en la justificación del problema en el capítulo 1 del presente informe.

2.2. Prepago de Créditos Hipotecarios

Existen muchas causas por las cuales se puede generar el prepago de un crédito hipotecario; venta de vivienda por mudanza u otras cuestiones socioeconómicas, prepagos parciales, ejercicio de seguros por siniestralidad u ocurrencia de hechos con cobertura (incendio, inundación, fallecimiento del deudor, etc.), refinanciamientos, entre otras. En teoría, toda acción llevada a cabo por un individuo es para maximizar su utilidad total, siendo muchas de estas, sub-óptimas económicamente hablando, es decir, se paga el crédito en su totalidad, pero no el óptimo para el individuo, este aspecto se detallará más adelante.

Con ello, se genera una serie de distorsiones en la información, que los modelos teóricos de prepago deberían capturar al ser desarrollados. Es por ello, que, dada ciertas condiciones, existe una probabilidad desconocida, asociada al prepago de un crédito en particular.

En términos generales, se catalogan a los prepagos en dos categorías; prepago óptimo y prepago sub-óptimo. A continuación, se hace un análisis de ambos casos con el fin de entender el trasfondo de un prepago y así, determinar cómo y qué variables podrían influir en él.

2.2.1. Prepago Óptimo

Como se mencionó, un prepago es desencadenado por variados motivos. Teóricamente, éste se debiera observar cuando el valor presente de ejercer la opción de prepago es estrictamente menor al valor presente de la suma de los flujos futuros que involucran al crédito. A continuación, se plasma matemáticamente lo mencionado en el paper de LaCour-Little, Marschoun & Maxam, 2001; esto se puede observar en la fórmula 1.

De manera equivalente, se puede representar la opción de prepago cuando el precio del instrumento es menor a su valor par más los costos asociados al prepago, es decir, cuando la opción está *in the money*. Luego,

$$P < \text{PrecioPar} + \text{Costos}$$

Fórmula 2

Donde P es el precio de mercado del instrumento, PrecioPar es el que mismo que ofrece el mercado por él y los Costos, representa a los costos relativos al prepago y eventualmente, de emitir una nueva deuda por el monto remanente.

El artículo concluye, que la tasa de interés presente en el mercado es posiblemente una variable relevante en la decisión y descripción de un prepago. Considerarla será de vital importancia en la construcción del modelo.

2.2.2. Prepago Sub-óptimo

Se puede definir al prepago sub-óptimo, cuando un individuo que posee un crédito hipotecario decide ejercer la opción aun cuando parezca, a simple vista, que económicamente no sea racional. Sin embargo, se vive en un mundo dinámico que crece y se desarrolla constantemente y junto a él, sus habitantes, las familias y la economía, por lo cual las decisiones se toman de forma veloz, lo que puede traer consigo el comportamiento descrito en este punto.

Es por lo anterior, que en ocasiones se observa en el mercado a individuos que ejercen la opción de prepago aun cuando esta está *out of the money*. Uno se encuentra en la posición *out of the money* cuando realizar el prepago es más costoso que esperar el pago regular de las cuotas del crédito.

Esto último, tiene racionalidad económica en el sentido que la utilidad que genera un inmueble más acorde a las reales necesidades que poseen los individuos, ya sea para uso doméstico o profesional, sí está finalmente maximizando la utilidad total de ellos.

Luego, el prepago se puede ejercer por diferentes motivos sub-óptimos. Estos pueden darse por el crecimiento de una familia, la cual necesite más espacio y con ello, adquirir una nueva vivienda. O que esta familia simplemente evolucione a una vivienda distinta, quedando implícita la rotación habitacional existente en el mercado.

La evolución de la economía, puede ser entonces una variable importante en cuanto a que el desarrollo y progreso de los negocios y familias, trae a las sociedades nuevas necesidades y nuevas instalaciones, lugares de trabajo y zonas para habitar.

De esa manera, al analizar las causas de un prepago sub-óptimo, se pueden inferir y tomar en cuenta diferentes variables para incluir en el desarrollo del modelo. Algunas de ellas serían por ejemplo la rotación habitacional y el crecimiento económico.

Cabe mencionar que el prepago óptimo o sub-óptimo es de punto del tomador del crédito, dado que para la institución financiera que otorga el préstamo, el problema sigue siendo el mismo. Este es que se le genera un descalce en sus flujos, dado el término anticipado del contrato. Este es el problema desarrollado en este trabajo.

2.2.3. Opción de prepago

Como se mencionó en los dos puntos anteriores, se puede notar el que individuo puede ejercer la opción de prepago o no. Esto se debe a diversos motivos, los cuales fueron mencionados anteriormente, pero en este punto se profundizará acerca de porqué el cliente no ejerce su opción de prepago cuando le es conveniente económicamente.

Un individuo bien informado y con educación le es más fácil notar cuando le conviene refinanciar su crédito para obtener una mejor tasa y por lo tanto maximizar sus utilidades en ese instante. También existen sujetos que no tienen la información y no poseen las herramientas necesarias para obtener un buen mecanismo de refinanciamiento por lo cual no saben cuándo les es conveniente realizar el prepago. Esto genera pérdida de beneficios para el individuo, y la institución financiera obtiene ganancias a partir de esto debido a que el individuo mantiene una tasa más alta de la cual puede optar. Es decir, se les está cobrando más caro, además de evitar el descalce de sus flujos, debido a que el cliente mantiene el crédito actual, vigente.

2.3. Data Mining

La minería de datos es el conjunto de técnicas y tecnologías que permiten explotar grandes bases de datos. Con el objetivo de encontrar patrones repetitivos, tendencias, reglas que expliquen el comportamiento de los datos en un determinado contexto, para la toma de decisiones de manera inteligente pero automatizada. Se centra en llenar la necesidad de descubrir el porqué, para luego predecir y pronosticar las posibles acciones con cierto factor de confianza para cada predicción.

La minería de datos puede ayudar a encontrar cuál es el mejor paquete de productos que le puede ofrecer a sus clientes existentes para que su relación sea más rentable. Por ejemplo, se pueden encontrar afinidades entre los productos que se consumen para realizar promociones y ofertas focalizadas (Marketing), detecta patrones de uso fraudulento de tarjetas de crédito (Banca),

detección de fraude telefónico (telecomunicaciones), identificación de terapias medicas satisfactorias para diferentes enfermedades (medicina), entre otros.

Los pasos a seguir para la realización de un proyecto de minería de datos son siempre los mismos, independientemente de la técnica especificada de extracción de conocimiento usada.

El proceso de minería de datos pasa por las siguientes fases:

1. **Selección de datos:** En esta etapa se determinan las fuentes de datos y el tipo de información a utilizar.
2. **Preprocesamiento:** Esta etapa consiste en la preparación y limpieza de datos extraídos desde las fuentes de datos de una forma manejable, necesaria para las fases posteriores en esta etapa se utilizan estrategias para manejar datos faltantes o en blanco, datos inconsistentes o que están fuera de rango, obteniéndose al final una estructura de datos adecuada para su posterior transformación.
3. **Transformación:** Consiste en el tratamiento preliminar de los datos, transformación y generación de nuevas variables a partir de las existentes con una estructura de datos apropiada. Aquí se realizan operaciones de agregación o normalización, consolidando los datos de una forma necesaria para la fase siguiente.
4. **Data Mining:** Es la fase de modelamiento propiamente tal, en donde los métodos inteligentes son aplicados con el objetivo de extraer patrones previamente desconocidos, válidos, nuevos, potencialmente útiles y comprensibles y que están contenidos u “ocultos” en los datos.
5. **Interpretación y Evaluación:** Se identifican los patrones obtenidos y que son realmente interesantes, basándose en algunas medidas y se realiza una evaluación de los resultados obtenidos

Data mining utiliza los datos existentes para:

- **Predecir:** Predice la pertenencia a una categoría.
- **Agrupar:** Descubre grupos de clientes homogéneos basados en sus características.
- **Identificar:** Identifica casos que no siguen un comportamiento esperado.
- **Asociar:** Encuentra eventos que ocurren simultáneamente o en una secuencia

2.4. Análisis Univariado

Es el análisis estadístico más sencillo, dado que trabaja todas las variables de un problema por separado, tomando estadísticas descriptivas de cada una para ver su comportamiento y ver cuáles son relevantes para describir el problema estudiado. En este caso el análisis univariado se utilizó para ayudar en el proceso de segmentación y poder distinguir cuál eran las variables más relevantes en el problema a partir de test estadísticos.

2.4.1. Segmentación

El tipo de información de cliente usado en la segmentación ha evolucionado en paralelo al desarrollo de los sistemas de información.

Desde las segmentaciones demográficas generales, más propias de la segmentación de mercados, pasando por el análisis comportamental basado en el valor se llega a los modelos de valor – necesidad, dominantes en la actualidad. Actitudes, prescripción, vinculación, y análisis

comportamental online constituyen las nuevas dimensiones que deben enriquecer los modelos de valor – necesidad.

Algunas de las variables demográficas más comunes son edad, sexo, educación, ingreso, ocupación, estado civil, tenencia de bienes.

2.4.1.1. K-Means

En un método de agrupamiento, que tiene como objetivo la partición de un conjunto de n observaciones en k grupos en el que cada observación pertenece al grupo cuyo valor medio es más cercano.

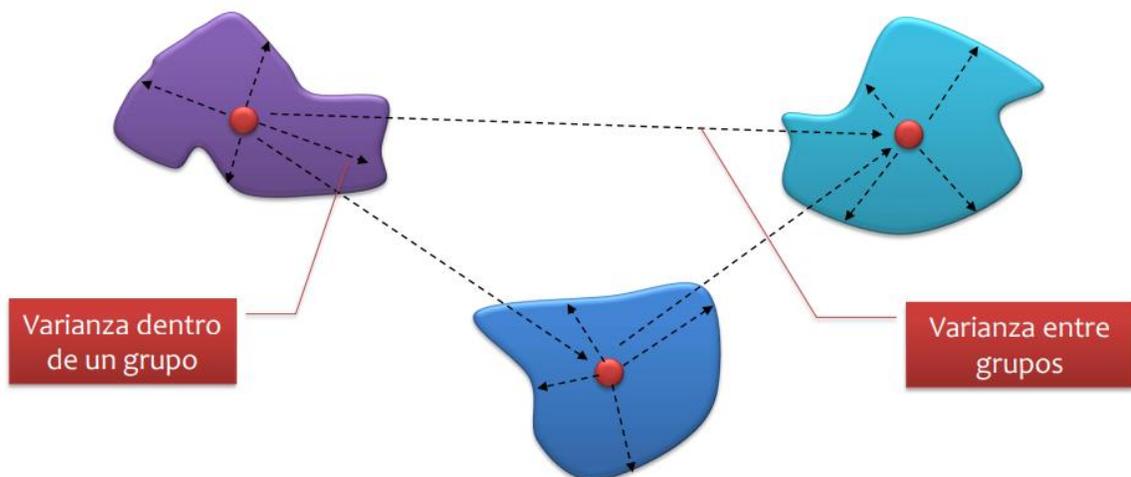
Para realizar la clusterización de los clientes se utiliza una técnica de análisis de conglomerado, que tiene las siguientes características:

- Existen diferentes métodos por los cuales cada objeto es asignado a un segmento
- Se busca que los objetos dentro de cada grupo sean similares entre sí y diferentes a los objetos de los otros grupos
- Se utiliza un principio de maximización de la varianza entre cluster mientras se minimiza la varianza dentro de un cluster.

$$\max_{c \in C} \left\{ \frac{\text{varianza entre grupos}}{\text{varianza en los grupos}} \right\}$$

$C =$ conjunto de clusters posibles

Ilustración 6: Ejemplificación Clusters



2.4.1.2. Objetivos de la segmentación

La segmentación se utiliza para determinar conjuntos de clientes que poseen un comportamiento similar y de esta manera poder estudiarlos de mejor forma. De esta manera al identificar grupos

similares se pueden aplicar estrategias diferenciadas para cada segmento, consiguiendo una mayor rentabilidad de las acciones de la institución.

2.4.2. Test de Medias

Los test para comparación de medias tienen como principal objetivo evaluar el efecto de un tratamiento comparando entre un grupo de clientes al que se le aplica un tratamiento con el grupo de control, que no ha recibido tratamiento.

Para $X_1, X_2, X_3, \dots, X_n$ con distribución $N(\mu_1, \sigma_1^2)$ y $Y_1, Y_2, Y_3, \dots, Y_n$ con distribución $N(\mu_2, \sigma_2^2)$ independientes entre sí. Se quiere contrastar la hipótesis sobre $\mu_1 - \mu_2$, donde $H_0: \mu_1 - \mu_2 = \Delta_0$, por lo cual, se calcula el estadístico:

$$Z = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1); \text{ Cuando } H_0 \text{ es verdadera}$$

Por lo que se rechaza la hipótesis nula y se acepta la alternativa en los siguientes casos:

Hipótesis Alternativa H_A

$$\begin{aligned} H_A: \mu_1 - \mu_2 &> \Delta_0 \\ H_A: \mu_1 - \mu_2 &< \Delta_0 \\ H_A: \mu_1 - \mu_2 &\neq \Delta_0 \end{aligned}$$

Región de rechazo para un nivel de significancia α

$$\begin{aligned} z &> z_\alpha \\ z &< -z_\alpha \\ z &< \frac{z_\alpha}{2} \text{ o } z > -\frac{z_\alpha}{2} \end{aligned}$$

En muchos casos solo se requiere saber si las medias de las poblaciones son diferentes, por lo que $\Delta_0 = 0$. Este es el caso que se aplicó en este trabajo.

2.4.3. Test X^2 de Pearson

La prueba de X^2 de Pearson se considera una prueba no paramétrica, que se utiliza para probar la independencia de dos variables entre sí, mediante la presentación de los datos.

La independencia de dos variables consiste en que la distribución de una de las variables es similar sea cual sea el nivel que examinemos de la otra.

La prueba de independencia chi-cuadrado contrasta la hipótesis de que las variables son independientes, frente a la hipótesis alternativa de una variable se distribuye de modo diferente para diversos niveles de la otra. Por lo cual a un nivel de significancia determinado se puede observar si dos grupos distribuyen de la misma forma.

Esta prueba tiene muy pocas limitaciones, aunque se deben considerar ciertas observaciones, como que, si las muestras son muy grandes, la prueba de independencia dará resultados significativos incluso donde, posiblemente, no los haya. No estudia la asociación entre variables, por ende, sirve únicamente para el análisis univariado.

2.5. Análisis multivariante

Es un método estadístico utilizado para determinar la contribución de varios factores en un evento o resultado. La idea de este análisis es estudiar el comportamiento de las diversas variables que se determinaron relevantes a partir del análisis univariado, para ver cómo se comportan en conjunto y cuál es su influencia en el problema estudiado.

2.5.1. Regresión Logística

Los métodos de regresión se han convertido en un componente integral de cualquier análisis de datos, que consiste en la descripción de la relación entre una variable respuesta y una o más variables explicativas. A menudo es el caso de que la variable respuesta es dicotómica teniendo dos valores posibles. En la última década el modelo de Regresión Logística se ha convertido, en muchos campos en el método estándar de análisis en esta situación.

Según Hosmer y Lemeshow (2000), “lo que distingue a un modelo de Regresión Logística a partir del modelo de regresión lineal es que la variable respuesta en la Regresión Logística es binaria o dicotómica”.

En general la Regresión Logística es adecuada cuando la variable respuesta es politómica (admite varias categorías de respuesta, tales como mejora mucho, empeora, se mantiene, mejora); pero es especialmente útil en particular cuando solo hay dos posibles respuestas (dicotómica), que es el caso común. En esta situación el investigador está interesado en la predicción y explicación de las relaciones que influyen en la categoría en que un objeto está situado.

La regresión logística puede ayudar a obtener una solución para el problema, dado que el comportamiento de prepago se puede modelar como una elección discreta de si prepago en un determinado instante, esto se ve de la siguiente manera:

$$Y_t \begin{cases} 1, & \text{si prepagó en período } t \\ 0, & \text{si no prepagó en el período } t \end{cases}$$

De esta manera se podrá buscar P_{nt} que es la probabilidad que el cliente n , prepague en el período t .

La regresión logística tiene los siguientes supuestos:

1. Las variables explicativas no tienen correlación.
2. Distribución Binomial, describe la distribución de los errores.
3. No linealidad de la variable respuesta.

2.7.2. Backtesting

El backtesting es un método para hacer predicciones de forma “forecast”, es decir, con la información actual, poder realizar pronósticos futuros. En el presente trabajo se utilizó el backtesting para estimar fuera de la muestra, se eligen un porcentaje de casos aleatorios para

pronosticar el comportamiento del otro porcentaje que no fue seleccionado y estudiar su comportamiento, este caso, cuales tendrán una probabilidad más alta de ser prepagados.

Este es un método ampliamente usado en el ámbito de las inversiones, donde se toman los datos históricos, para ver que estrategias son más convenientes y donde se debe poner nuestro foco en posiciones futuras de una empresa o agente de inversión.

Capítulo 3

3. Metodología

La metodología del presente trabajo se divide en 4 etapas: pre procesamiento de datos, análisis descriptivo y segmentación de variables, creación del modelo e interpretación de los resultados.

3.1. Procesamiento de datos

Para contextualizar se detalla el contenido de la base que fue entregada por el banco, la cual contenía 8.681.3166 observaciones. A la fecha se tiene una suma aproximada de \$6.000 MMM de pesos en colocaciones. Las observaciones tenían asociada las siguientes variables con los cuales se trabajó, en el anexo K se pueden ver el nombre de las llaves en la base:

- Id del cliente: indica el código asociado a un cliente
- Código de la operación: indica el código asociado a la operación para un respectivo crédito.
- Tipo de operación: podía ser HIP300 o PLC300 dependiendo el tipo de crédito.
- Fecha proceso: indica la fecha en que se encuentra el crédito en ese momento.
- Fecha apertura: indica la fecha en que se concibió el crédito.
- Código de moneda: indica la moneda a la cual se contrajo el crédito, 998 UF, 999 pesos chilenos.
- Valor original: valor del monto del crédito, varía según la UF de determinada fecha.
- Saldo de mora 1: es el monto que se encuentra en mora entre 0 y 30 días.
- Saldo de mora 2: es el monto que se encuentra en mora entre 31 y 89 días.
- Saldo vencido: es el monto que se encuentra en mora más de 90 días.
- Monto pagado en el mes: es el monto que el cliente pago en esa fecha.
- Vencimiento del mes: es el monto que debe pagar en ese mes el cliente.
- Número de cuotas por pagar: es la cantidad de cuotas faltantes del crédito.
- Saldo capitalizado: es el saldo que le va quedando al cliente.
- Interés devengado: es el monto que se genera por los intereses
- Reajuste devengado: es el monto total, del crédito más lo generado por los intereses.
- Tasa: es la tasa a la cual se contrajo el crédito.
- Cuotas originales: es la cantidad de cuotas originales del crédito.

En esta etapa se depuró la base de datos disponible para eliminar todas aquellas observaciones que no sean útiles para la estimación de los modelos propuestos. Por lo cual dado la base que contenía dos tipos de operaciones, las denominadas HIP300 y las PLC300 siendo las primeras el objeto de estudio que contiene a todos los créditos asociados a la primera vivienda. Los créditos tipo HIP300 son el grueso de los créditos de la base que el banco facilitó, el segundo caso corresponde a los créditos otorgados para fines generales que contiene segundas viviendas, ampliaciones, terrenos, oficinas entre otros casos. A continuación, se presenta el desglose por cada operación:

Tabla 1: Comparación número de operaciones HIP300 y PLC300 a septiembre 2017

Operación	Nº de observaciones	Porcentaje
HIP300	7.078.815	81,5%
PLC300	1.602.501	18,5%

Fuente: Elaboración propia a partir de los datos de la base

En esta etapa se eliminaron los datos con campos vacíos o nulos y también los correspondientes a PLC300 dado que se sugirió por parte del banco trabajar solo con los HIP300 dado la mal confección de los datos para la operación asociada a fines generales.

Luego de la limpieza de los datos, la base quedo confeccionada por 7.708.815 observaciones las cuales hacen referencia a 115.757 clientes, con una cantidad de 135.999 créditos distintos.

Posterior a tener la data limpia, se armó la variable “PREPAGO_PARCIAL” y “PREPAGO_TOTAL” que identifica a los clientes que poseen ese tipo de comportamiento, siendo el prepago parcial una disminución de las cuotas por pagar de los clientes de un periodo a otro y el prepago total siendo un cese del crédito antes de terminar las cuotas pendientes. Para proceder se trabajó solo con la variable “PREPAGO_TOTAL”, la cual de aquí en adelante se denominará como “Variable Prepago”. Debido a que los datos asociados al prepago parcial eran muy pocos para obtener resultados significativos.

3.2. Variable Prepago

En esta sección se detallará como se obtuvo la variable prepago, mencionada en la sección anterior. Esta variable se creó como una variable binaria donde el 1 indica que la operación recibió un prepago total en algún periodo del tiempo, mientras que el 0 indica que la operación fue finalizada o sigue vigente. De esta manera, se obtienen dos tipos de créditos, los tipo 1 que son los que recibieron el comportamiento y por ende finalizan antes de su fecha de término, mientras que los de tipo 0, que son los que no han recibido el prepago en algún periodo del tiempo o los que terminaron en la fecha estipulada de cuotas del contrato original.

En un comienzo, para entender el procedimiento realizado, se debe explicitar como se encontraban los datos en la base de datos. Estos eran datos de panel, los cuales estaban ordenados por el código de operación, es decir, se tenía un registro y el siguiente correspondía a la misma operación si es que el crédito continuaba vigente. Por lo cual, cuando la operación cambiaba, era porque el crédito cesaba o existía el comportamiento de prepago o el último caso corresponde a que el registro era la fecha actual, por lo tanto, en el último caso el crédito seguía vigente. Cabe destacar que cuando el crédito se renegocia, cambia el código de operación y al cancelarlo anticipadamente si es considerado como prepago.

Explicado lo anterior, se realizó el siguiente procedimiento, la cual se realizó en el software R para poder ir iterando registro a registro el comportamiento de cada crédito en particular. Se tomó como variable clave el número de cuotas por pagar, de esta forma, se podía registrar cuando el crédito cesaba antes de su finalización. Dada la variable clave había que tener varias consideraciones para poder determinar la variable de prepago de forma óptima. En primer lugar, había que condicionar que la operación de un registro con el siguiente fuera distinta. Luego había que ver que la fecha de

la operación fuera distinta a la actual de la base que es 30 de septiembre del 2017. Por último, había que condicionar que el crédito no poseyera ninguna deuda de ningún tipo en el periodo anterior, debido a que si tenía deuda y el cliente las cancelaba no correspondía a un comportamiento de prepago.

Por lo tanto, la variable binaria creada se puede caracterizar de la siguiente forma:

$$X_{ijt} \begin{cases} 1, & \text{si prepagó el crédito } i \text{ el cliente } j \text{ en el tiempo } t \\ 0, & \text{si no prepagó el crédito } i \text{ el cliente } j \text{ en el tiempo } t \end{cases}$$

Sujeto a las siguientes restricciones:

$$\begin{aligned} N_{it} &\neq 1, \\ Op_{i,p} &\neq Op_{i,p+1} \\ Fecha_{ip} &\neq "2017 - 09 - 30" \\ Mora_{1_{i,t-1}} &= 0 \\ Mora_{2_{i,t-1}} &= 0 \\ Saldo_Vencido_{i,t-1} &= 0 \end{aligned}$$

Donde:

x_{it} : variable binaria que indica si la operación i fue prepagada en el período t
 N_{it} : variable que indica el número de cuotas remante para la operación i en el período t
 Op_{ip} : Id de la operación i en el registro p
 $Fecha_{ip}$: Fecha en el que se encuentra la operación i para el registro p
 $Mora_{1_{it}}$: Es la mora de entre 0 a 30 días del crédito i en el periodo $t-1$
 $Mora_{2_{it}}$: Es la mora de entre 31 a 89 días del crédito i en el periodo $t-1$
 $Saldo_Vencido_{it}$: Es la mora de más de 90 días del crédito i en el periodo $t-1$

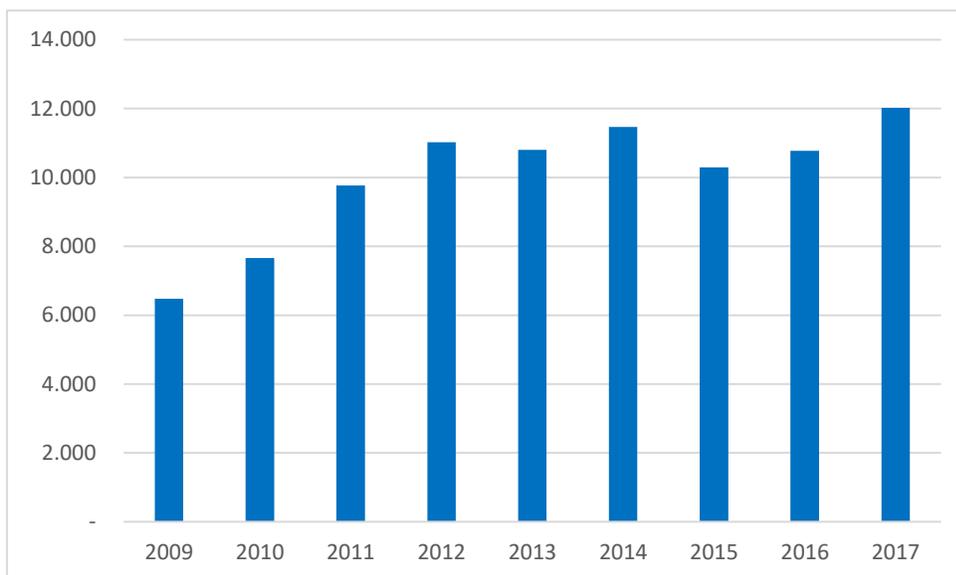
3.3. Análisis Exploratorio

Una vez terminada la depuración de los datos y creada la variable prepago, se realizó un análisis exploratorio de los datos, donde el objetivo del análisis es tener una intuición del comportamiento de los créditos.

Para comenzar con el análisis se obtuvo la estadística de cuantos créditos fueron generados desde el año 2009 en adelante y cuáles eran los montos asociados, lo comentado se detalla en la ilustración 7 y 8.

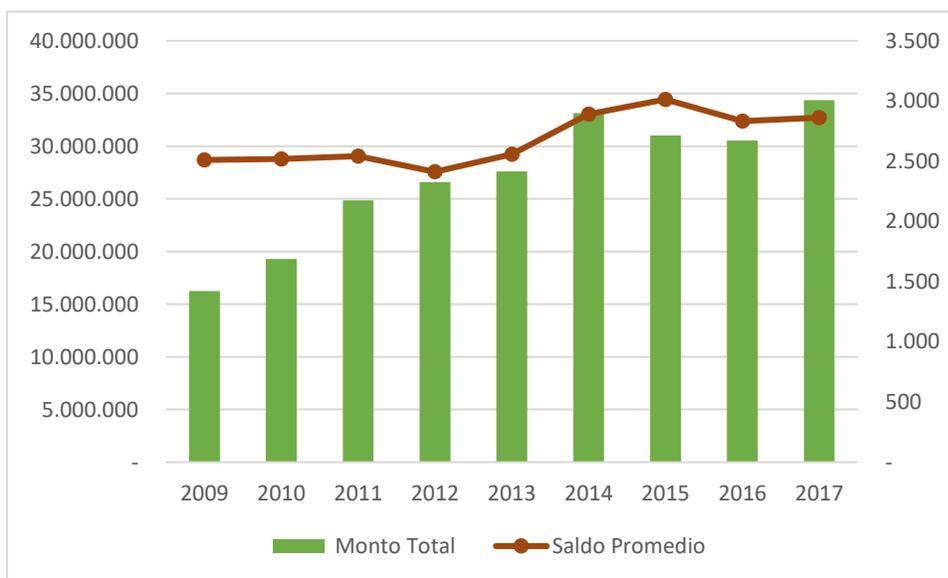
Se muestra la cantidad de créditos de vivienda otorgados desde el año 2009 por el banco, luego el monto total otorgado por año y su respectivo saldo promedio. Cabe destacar que la información presentada es la que está vigente en la base de datos, por ende, para la cantidad de créditos y montos previos al año 2009 no se encuentra la información en su totalidad dado que los créditos previos a esa fecha pudieron ser prepagados o finalizados.

Ilustración 7: Número de Créditos Totales otorgados por Año



Fuente: Elaboración propia a partir de los datos de la base

Ilustración 8: Información Créditos en UF

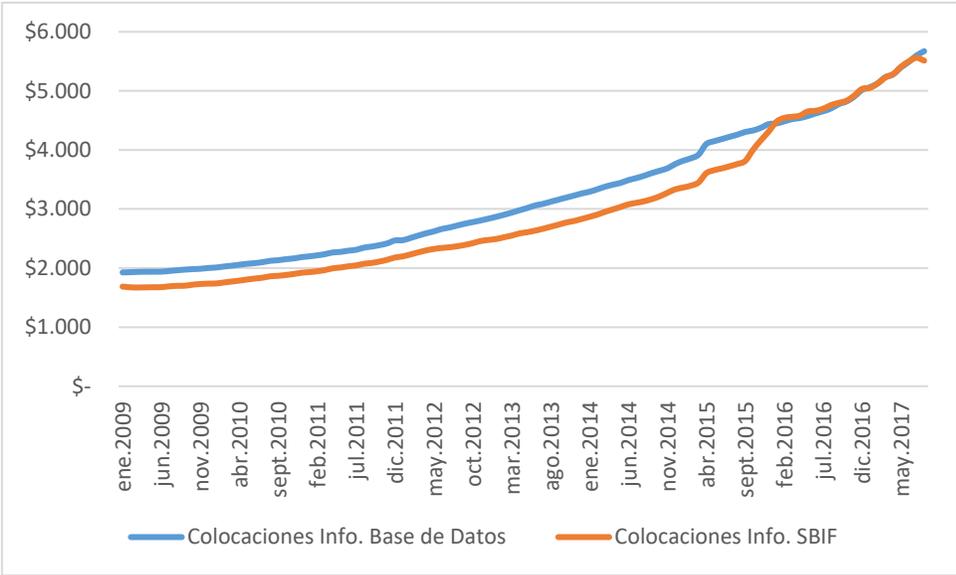


Fuente: Elaboración propia a partir de los datos de la de la base

A continuación, se compara el nivel de colocaciones y su evolución para la cartera de vivienda respecto al banco con el cual se trabajó, para poder determinar la veracidad de los datos que fueron entregados y si éstos se encuentran en el mismo rango de magnitud.

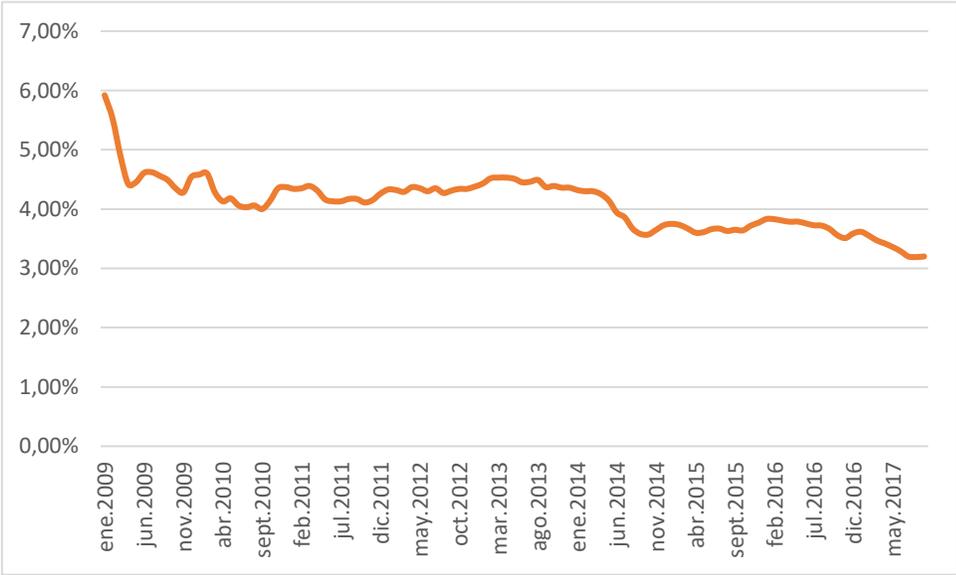
En el Ilustración 9 se presenta la comparación de las colocaciones de la cartera respectivo al banco con el cual se trabajó. Los montos presentados se encuentran en miles de millones de pesos. Para complementar en la Ilustración 10 se muestra la variación de la tasa de la cartera de vivienda durante el mismo período, los datos fueron extraídos desde la página del banco central.

Ilustración 9: Comparación Nivel de Colocaciones(MMM)



Fuente: Elaboración propia a partir de los datos de la de la base y la SBIF

Ilustración 10: Tasa Cartera Vivienda



Fuente: Elaboración propia a partir de los datos del Banco Central

Se puede observar que los valores son muy similares, por lo cual con la información con la cual se está trabajando es coincidente con la que maneja el sistema. En este caso la SBIF tiene pequeñas

diferencias observadas se pueden atribuir a que los datos para los créditos PLC300 no están muy bien construidos. También se observa que en el primer semestre del año 2015 existe una leve alza en los datos de la SBIF, esto se puede atribuir a una mala entrega de la información desde el banco a la SBIF, pero esto se ajusta a partir del enero del 2016.

Se continua con el análisis exploratorio extrayendo las estadísticas de las principales variables de las de la base de datos. Las cuales son el valor original de los créditos, las cuotas originales, la tasa de emisión y el monto de la cuota mensual. Otras variables que fueron trabajadas son el tipo de tasa a la cual se contrajo el crédito y la mora, pero para esas variables se detallara la información cuando se hable de la segmentación realizada. Las estadísticas presentadas para cada variable son la media, el mínimo, máximo y desviación estándar. Para no sobre estimar a los créditos que tenían un mayor número de observaciones, se decidió extraer la última observación de cada crédito, la cual contiene toda la información genérica de la operación respectiva y de esta manera contabilizar una única vez el crédito estudiado.

Tabla 2: Estadísticas Descriptivas de la Base de Datos

	Valor Original (\$)	Cuotas Originales	Monto Cuota (\$)	Tasa Emisión
Media	59.769.549	241,33	375.617	4,36%
Mínimo	559.793	22	-	0,34%
Máximo	2.262.080.000	364	42.857.368	12,00%
Desviación Estándar	48.866.779	64,743	381.508	1,13%

Fuente: Elaboración propia a partir de los datos de la base

La tabla 2 sirve para generar una comprensión mayor de los datos con los cuales se trabajó, en particular mencionar que el monto de cuota 0, considera el último periodo del crédito donde el cliente no debe nada, es decir, el crédito concluye. Por otro lado, el 12% de tasa de emisión corresponde a los créditos más antiguos, alrededor del año 1990 cuando la tasa para la cartera de vivienda era muy alta.

Antes de comenzar a hablar de las tasas de prepago y posterior segmentación, se enuncian los datos para ambos grupos, los cuales son, el grupo de créditos que si recibió un prepago frente al que no recibió el comportamiento.

En total existen 135.999 operaciones únicas en el sistema hasta septiembre del 2017, estas se dividen en 34.895 operaciones con el comportamiento de pre pago y 101.104 sin ese comportamiento, es decir, un 25,66% de las operaciones de la data entregada fueron prepagadas. Cabe mencionar que desde el año previo al 2009 no se tiene la totalidad de información de los créditos por lo cual el 25,66% es solo un dato de referencia, pero coincide con lo estudiado por el paper de Lacour-Little, Marschoun y Maxan en el año 2002.

Para el grupo de no prepago tenemos que 101.104 operaciones no recibieron un pago anticipado antes de la última cuota faltante para completar el crédito. En tanto para el grupo de prepago fueron 34.895 operaciones las que sufrieron el comportamiento Se ejemplifica esto en la tabla 3.

Tabla 3: Datos Grupos

	Cantidad de Créditos	Porcentaje
Prepagados	34.895	25,66%
No Prepagados	101.104	74,34%
Total	135.999	100,00%

Fuente: Elaboración propia a partir de los datos de la base

3.4. Tasa de prepago por período

Para continuar con el análisis, se creó una tabla con las tasas de prepago para cada período, esta se realizó de dos formas. En primer lugar, se calculó la tasa mediante la cantidad de operaciones prepagadas por período sobre la cantidad de operaciones totales en un período, para luego comparar esa tasa, con la calculada como el saldo prepagado en el periodo sobre el saldo total del periodo. La tasa con la que se trabajó es una tasa instantánea mensual de cómo es el prepago en un determinado momento. Se usó como referencia la variable “PREPAGO_TOTAL” para separar ambos grupos, en primer lugar, se calculó la tasa como:

$$TPr_1 = \frac{\sum_J \sum_I X_{ijt}}{\text{Total Operaciones Período } t}$$

Con:

TPr_1 : Tasa de prepago por unidad de operaciones para un período

$$\sum_J \sum_I X_{ijt} : \text{N}^\circ \text{ de operaciones prepagadas en período } t$$

En segundo lugar, la tasa se calculó de la siguiente forma:

$$TPr_2 = \frac{\sum_J \sum_I S_{ijt}}{S_T}$$

Con:

TPr_2 : Tasa de prepago por unidad de monto

$$\sum_J \sum_I S_{ijt} : \text{Saldo prepagado en período } y$$

S_T : Saldo total en stock período y

A continuación, se presentan los datos más relevantes para ambos tipos de tasa.

Tabla 4: Estadísticas para la tasa de prepago para TPr1

	Tasa Prepago	Período
Media	0,49%	-
Máximo	0,76%	2012-05
Mínimo	0,20%	2009-02
Desviación estándar	0,106%	-

Fuente: Elaboración propia a partir de los datos de la base

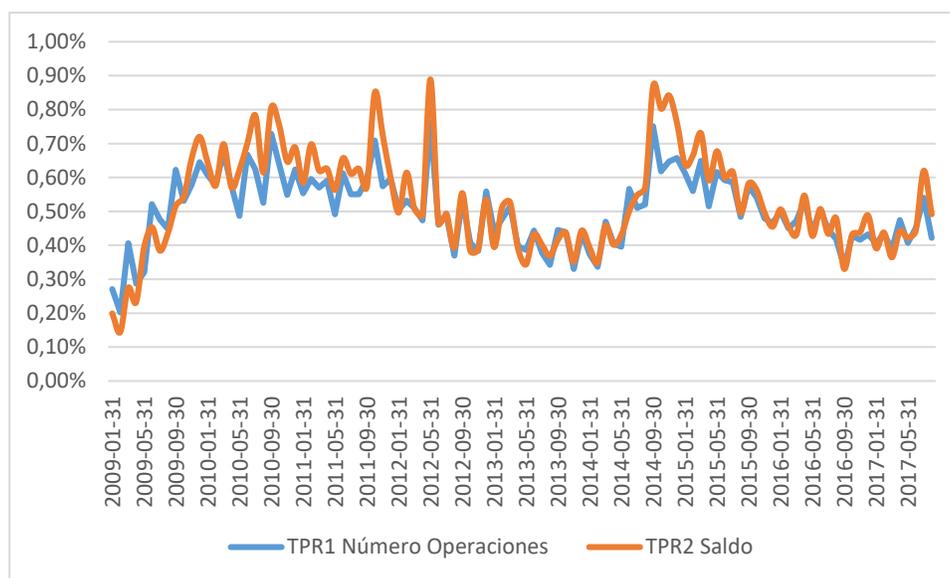
Tabla 5: Estadísticas para la tasa de prepago TPr2

	Tasa Prepago	Período
Media	0,51%	-
Máximo	0,89%	2012-05
Mínimo	0,14%	2009-02
Desviación estándar	0,14%	-

Fuente: Elaboración propia a partir de los datos de la base

Para complementar las estadísticas anteriores, se presenta los gráficos con la historia de prepago para cada tipo de tasa. Los cuales muestran la variación de la tasa en el tiempo donde se puede ver una alta volatilidad que es lo que se quiere explicar con variables endógenas y exógenas de los créditos.

Ilustración 11: Comparación Tasas de Prepago



Fuente: Elaboración propia a partir de los datos de la de la base

Se puede apreciar que la Ilustración 11 para la TPr_2 es más suavizada que la TPr_1 , por lo cual se considera trabajar con esta para los pasos siguientes de la investigación. Dado que las curvas que se observan en el gráfico presentan una alta volatilidad, se construyó una tasa agrupada por

trimestre, la que corresponde la tasa promedio trimestral(TPT). Se pensó crear una tasa trimestral de prepago de la forma:

$$TTP = \frac{Prepagos_i}{Stock\ Créditos_i}$$

En la expresión anterior el numerador denota la cantidad de prepagos realizados durante el trimestre, mientras que el denominador muestra el stock inicial de créditos activos al principio del trimestre.

La TTP tiene una magnitud mayor a la TPT, pero en la investigación realizada se busca explicar el 0,51% de prepago promedio mensual, por lo cual se opta por trabajar con la TPT en la segmentación que se presenta en la sección a continuación que sirve para crear una noción de cómo afectan las variables al modelo y dar una intuición del comportamiento de prepago. En la ilustración N°12 se presenta la TPT para el saldo prepagado. Dado que esta tasa es un promedio, no presenta un valor del orden de 3 veces la tasa.

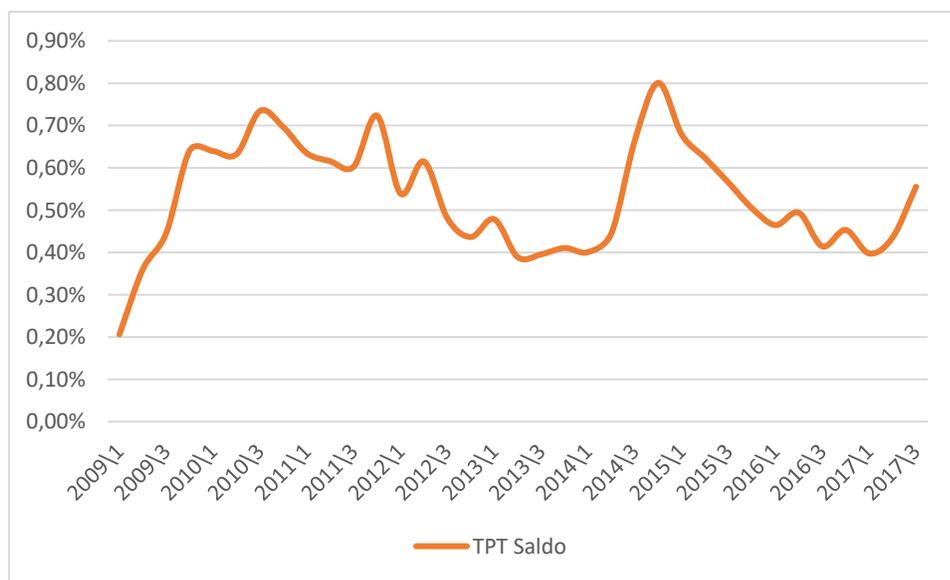
El monto total prepagado desde el año 2009 hasta la fecha queda explicitado en la tabla 6.

Tabla 6: Montos asociados al Prepago

	Montos (UF)
Saldo Prepagado	57.106.070
Saldo en Stock	11.109.325.178
Tasa Prepago	0,51%

Fuente: Elaboración propia a partir de los datos de la base

Ilustración 12: TPT Promedio Cartera Vivienda



Fuente: Elaboración propia a partir de los datos de la de la base

3.5. Segmentación

En este punto se comienza con el análisis univariado, donde se trabajó cada una de las variables por separado para ver su incidencia en el comportamiento de prepago. Se quiere ver como una segmentación de las variables seleccionadas puede dar una mejor idea de que características de los créditos y de esta manera observar si existe una mayor incidencia sobre el comportamiento estudiado.

El proceso de cómo se realiza la segmentación es poder determinar las variables relevantes a partir del análisis exploratorio realizado y la bibliografía estudiada. Luego de varias segmentaciones por variables, se decide generar dos grupos por variable para poder mostrar y observar de manera más clara el comportamiento por grupo. En la mayoría de las variables trabajadas se utilizó el algoritmo K-Means (detallado en el marco teórico) para determinar los centros de cada grupo y poder realizar el corte y para pasar a la posterior segmentación. Se eligió dos grupos debido a que la intención de esta sección es entender si existe evidencia de diferencias entre grupos, por lo cual 2 núcleos nos permite generar la intuición adecuada del comportamiento.

Las variables trabajadas se enuncian según su importancia dado el test X^2 de Pearson que se utilizó para categorizar las variables seleccionadas, las variables con las cuales se trabajó fueron las siguientes:

- Monto Original
- Valor de la Cuota
- Spread Tasa y sus desfases
- Plazo Residual
- Tipo de Tasa
- Si el cliente tuvo mora según su segmentación

Estas variables fueron seleccionados luego del levantamiento de información a partir de la bibliografía y del análisis exploratorio. Luego de tener los grupos segmentados se utilizó el test de medias para ver si existía evidencia estadística de diferencia entre los grupos creados, en caso de que no se presentara significancia estadística igualmente se agregaban esas variables a los modelos, dado que en un análisis multivariado podrían tener influencia sobre el comportamiento estudiado.

3.5.1. Monto Original

La variable monto original hace referencia a la cantidad de dinero prestado por la institución financiera al cliente. En esta variable se utilizó el algoritmo mencionado para determinar los centros de los clusters y se denominó a los grupos como créditos altos(CA) y créditos bajos(CB). En la Ilustración 13 se puede observar la historia de prepago para ambos grupos y su comparación con el promedio.

Para esta variable el punto de corte fue de \$46.276.187 pesos y presenta evidencia significativa de diferencia de medias, el resultado del test se puede ver en anexo C. En la tabla 7 se puede observar los datos principales de esta segmentación.

Ilustración 13: Segmentación por Monto Otorgado

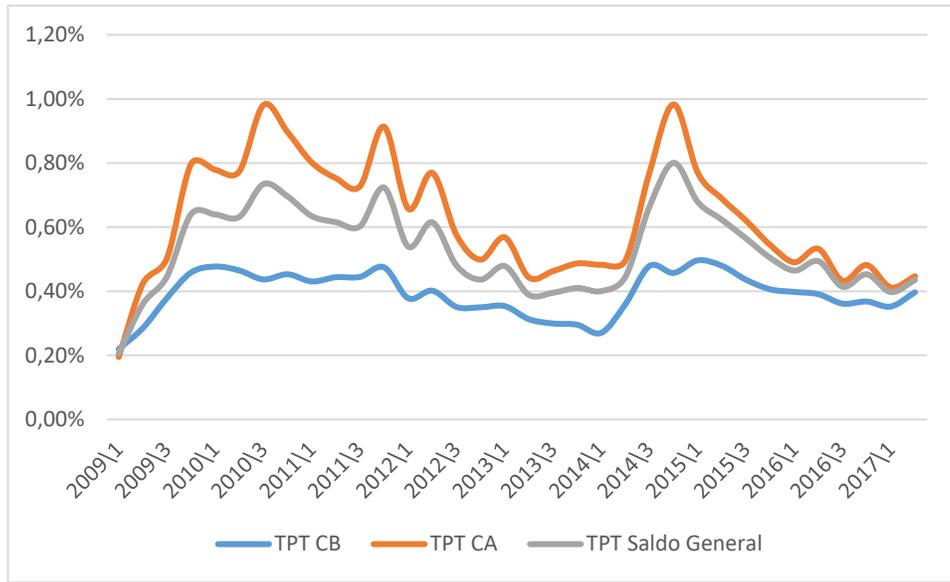


Tabla 7: Estadísticas Segmentación Monto Original

	N	Porcentaje	Monto Prepagado	Tasa Promedio	Corte
CA	67.983	49,99%	41.891.387 UF	0,58%	>\$46.276.187
CB	68.016	50,01%	15.214.682 UF	0,39%	<=\$46.276.187

Fuente: Elaboración propia a partir de los datos de la de la base

3.5.2. Valor Cuota

El valor de la cuota hace referencia al vencimiento del mes por parte del cliente en un crédito, es decir, el monto que debe pagar el contratista del préstamo en el mes. Esta es una variable que al principio no se tenía contemplada, pero luego de leer la bibliografía en particular los artículos de Lacour-Little se decidió añadir. Cabe mencionar con respecto a esta variable que existe una alta correlación con el monto original, la tabla de correlaciones de las variables será presentada al final de esta sección.

La variable se segmento en cuotas grandes(CG) y cuotas pequeñas(CP), en la Ilustración 14 se puede ver la historia de prepago de la segmentación comparada con el promedio.

En cuanto a esta variable también se presentó una evidencia estadística significativa, el test se puede ver en anexo D, donde las cuotas grandes presentan un mayor comportamiento de prepago que las cuotas pequeñas. El corte obtenido por el algoritmo fue \$375.000 pesos. En la tabla 8 se presentan las estadísticas para esta segmentación.

Ilustración 14: Segmentación Valor Cuota

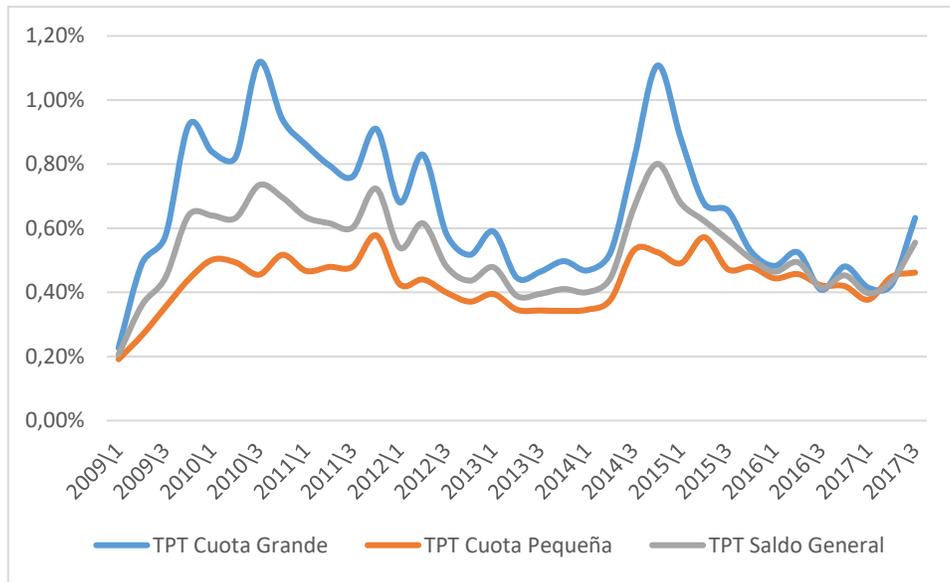


Tabla 8: Estadísticas Segmentación Valor Cuota

	N	Porcentaje	Monto Prepagado	Tasa Promedio	Corte
CG	46.072	33,88%	32.341.484 UF	0,61%	>\$375.000
CP	89.927	66,12%	24.764.586 UF	0,43%	<=\$375.000

Fuente: Elaboración propia a partir de los datos de la de la base

3.5.3. Spread Tasa y sus desfases

El spread tasa se calculó de la siguiente manera:

$$\text{Spread Tasa} = \text{Tasa Otorgada} - \text{Tasa Mercado}_i$$

Donde la tasa otorgada es a la cual se contrajo el crédito para tasa fija y si en el contrato se fijó tasa mixta, es la tasa, a la cual se encuentra el crédito en ese momento. La tasa de mercado se tomó dada la información que entrega el banco central para la cartera de vivienda, se puede observar en la ilustración N°10. Esta variable se trabajó con desfases porque la tasa de mercado varía dependiendo el desfase de forma mensual. Se trabajó con 5 desfases dado que la literatura menciona que con un tiempo de tres meses el cliente demora en tomar la decisión de prepago, por lo cual se agregó dos instancias más para ver su efecto y tener modelos más robustos posteriormente.

En esta ocasión no se utilizó el algoritmo debido a que se decidió tomar el spread como positivo o negativo, teniendo la intuición de que cuando el spread es positivo el comportamiento de prepago debería aumentar debido a que la tasa actual es menor frente a la que se estipuló en el contrato original del crédito. Esta variable tiene una fuerte relación con el termino refinanciamiento que se explicó en el capítulo 1.

Para esta sección dado que los gráficos no muestran una clara evidencia del comportamiento de prepago para los diversos spreads presentados, se muestran únicamente las tablas con la tasa de prepago para cada segmentación.

Tabla 9: TPT Segmentación sin Desfase

Spread Tasa sin Desfase	TPT
Negativo	0,45%
Positivo	0,51%

Fuente: Elaboración propia a partir de los datos de la de la base

Tabla 10: TPT Segmentación Spread Tasa 1 Desfase

Spread Tasa 1 Desfase	TPT
Negativo	0,43%
Positivo	0,52%

Fuente: Elaboración propia a partir de los datos de la de la base

Tabla 11: TPT Segmentación Spread Tasa 2 Desfase

Spread Tasa 2 Desfases	TPT
Negativo	0,43%
Positivo	0,52%

Fuente: Elaboración propia a partir de los datos de la de la base

Tabla 12: TPT Segmentación Spread Tasa 3 Desfases

Spread Tasa 3 Desfases	TPT
Negativo	0,43%
Positivo	0,52%

Fuente: Elaboración propia a partir de los datos de la de la base

Tabla 13: TPT Segmentación Spread Tasa 4 Desfases

Spread Tasa 4 Desfases	TPT
Negativo	0,43%
Positivo	0,52%

Tabla 14: TPT Segmentación Spread Tasa 5 Desfases

Spread Tasa 5 Desfases	TPT
Negativo	0,44%
Positivo	0,52%

Fuente: Elaboración propia a partir de los datos de la de la base

Se puede notar en las tablas anteriormente presentadas que efectivamente existencia una tendencia mayor al prepago en los grupos que tienen un spread positivo. Por ende, cuando la tasa otorgada es mayor a la de mercado el cliente tendera a realizar el prepago en una mayor cantidad.

En esta ocasión, para la variable spread tasa, tenemos significancia estadística para el spread con 1,2,3,4 y 5 desfases para el test de media de una cola, en el caso sin desfase no se presenta evidencia robusta para concluir la diferencia. Se comentará más acerca de esto, en la interpretación de los resultados apoyado por el modelo obtenido. En el anexo E se pueden ver los resultados de los test de hipótesis para esta variable y en el anexo F los gráficos con la historia de prepago para cada segmentación

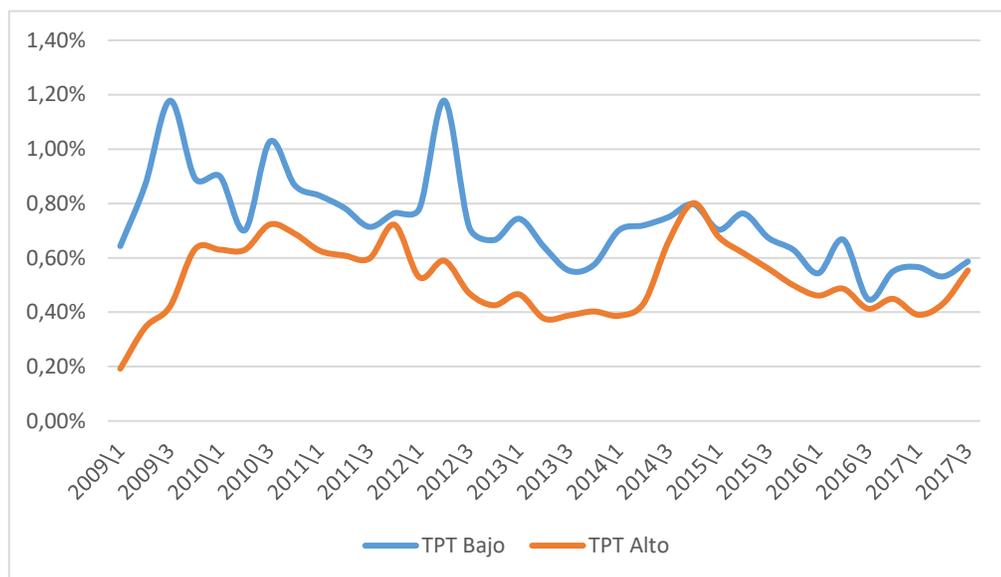
3.5.4. Plazo Residual(PR)

El plazo residual nos indica cuanto tiempo le queda al crédito de vida, se construyó como una variable porcentual que indica cuan cerca está el crédito de terminar, esta variable se puede expresar de la siguiente forma:

$$PR = \frac{\text{Cuotas por Pagar}}{\text{Cuotas Originales}}$$

La segmentación que se decidió para esta variable fue el de PR alto y PR bajo con un corte del 0,5, es decir, si el crédito se encuentra en su primera mitad el PR es alto, mientras que si se encuentra bajo el 0,5 el PR es bajo. En la Ilustración 15 se puede ver la comparación de ambos grupos.

Ilustración 15: Segmentación Plazo Residual



Esta variable también evidencia significancia estadística, ver anexo G, donde el plazo residual bajo presenta mayor comportamiento de prepago, es decir, cuando el crédito está más cerca del término, las personas lo terminan antes. En la tabla 15 se presentan las principales estadísticas para esta segmentación.

Tabla 15: Estadísticas Segmentación Plazo Residual

	Monto Prepagado	Tasa Promedio	Corte
PR Alto	54.002.197 UF	0,51%	>0.5
PR Bajo	3.103.873 UF	0,68%	<=0.5

Fuente: Elaboración propia a partir de los datos de la de la base

En la tabla 15 se puede observar que el monto prepago para el PR bajo es menor al monto prepago por los créditos con PR alto, esto se debe a que esta variable tiene una alta correlación el saldo residual, es decir, cuando menor plazo le queda al crédito menor cantidad de dinero el cliente debe cancelar para terminar el crédito de manera anticipada.

3.5.5. Tipo de Tasa

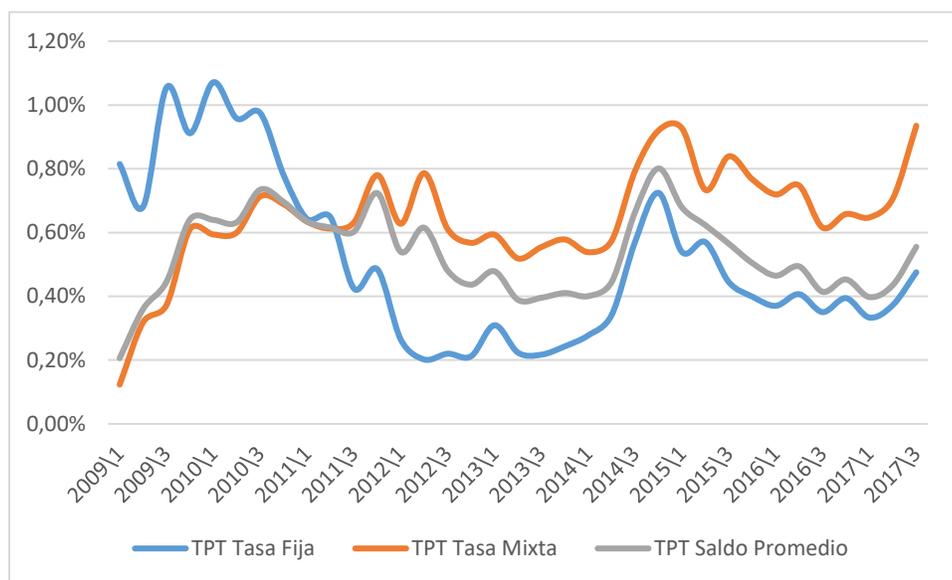
La quinta variable trabajada fue el tipo de tasa. Los créditos en su concepción se pueden fijar a tasa fija(TF), lo que indica que el crédito se paga a la misma tasa durante toda su duración. Por otro lado, tenemos los créditos que se fijan a tasa mixta(TM). Si el crédito se concibe con TM, la tasa tendrá un componente de tasa variable y otro de tasa fija, es decir, el crédito comienza con una tasa para luego de ciertos periodos de tiempo ir ajustando la tasa con respecto a las expectativas de mercado.

En lo particular esta variable se trabajó segmentado por el tipo de tasa. En primer lugar, se observó la historia de prepago para la tasa fija y la tasa mixta. Luego la tasa mixta se segmento en que

componente de la tasa mixta se prepago, es decir, si el prepago se efectuaba cuando la tasa era fija o está en su componente variable.

En la Ilustración 16 se presenta la comparación entre la historia de prepago para ambos tipos de tasa señalados.

Ilustración 16: Segmentación Tipo Tasa



Para la variable de esta sección también se presenta evidencia estadística (anexo H) en la diferencia de las medias de ambos grupos, en donde la tasa mixta presenta un porcentaje mayor de prepago que la tasa fija, la intuición a priori nos llevaba a una conclusión distinta, pero existe una hipótesis para este comportamiento que se detallará en la sección de análisis de resultados. Los datos de ambos grupos se encuentran en la tabla 16.

Tabla 16: Estadística Segmentación Tipo de Tasa

	N	Porcentaje	Monto Prepagado	Tasa Promedio	Corte
TF	82.459	60,63%	22.783.908 UF	0,40%	Tasa Fija
TM	53.540	39,37%	34.322.161 UF	0,63%	Tasa Mixta

Si se analiza únicamente la tasa mixta, se extrae que el 100% de los créditos que fueron prepagados con este tipo de tasa se realizaron durante el período de su componente fija.

3.5.6. Mora

La última variable de la segmentación corresponde a la mora, donde se trabajó como una variable binaria donde se indicaba si el crédito presento mora en algún periodo o no tuvo atraso en los pagos para ningún período. En anexo I se encuentra la historia de prepago para ambos grupos.

Para este caso, no se presenta una significancia estadística para la diferencia de medias entre ambos grupos, ver test en anexo I. Igualmente en la tabla 17 se presentan las estadísticas para la segmentación de la mora.

En el anexo J se puede observar la historia de prepago para la segmentación de la mora que se detalló al comienzo de este capítulo. Aunque esta variable no presentó una evidencia estadística en la diferencia de grupos, igualmente estará los modelos que se presentan en el capítulo a continuación, para ver cómo es su comportamiento en el análisis multivariado.

Tabla 17: Estadística Segmentación Mora

	N	Porcentaje	Monto Prepagado	Tasa Promedio
Mora	31.360	23,06%	16.095.642 UF	0,51%
No tuvo Mora	104.639	76,94%	41.010.428 UF	0,54%

3.5.7. Correlación Variables

En la última sección del análisis univariado, se presenta la tabla de correlaciones de las variables anteriormente presentadas, para poder complementar la interpretación de los resultados y analizar de mejor forma el comportamiento de las variables seleccionadas. En la matriz que se presenta a continuación se presenta la variable spread tasa como refinanciamiento que se detalla en la sección 3.5.

Tabla 18: Matriz de Correlaciones

	Constante	MONTO ORIGINAL	MONTO CUOTA	Refinanciamiento	Refinanciamiento Desfase1	Refinanciamiento Desfase2	Refinanciamiento Desfase3	Refinanciamiento Desfase4	Refinanciamiento Desfase5	Plazo Residual	Indicador Tasa	Indicador Mora
Constante	1,000	-,717	,309	-,015	,004	-,003	-,001	-,001	-,006	,112	-,113	-,054
MONTO ORIGINAL	-,717	1,000	-,882	,001	-,001	,003	,003	-,003	,010	-,302	,004	,078
MONTO CUOTA	,309	-,882	1,000	,003	,001	-,002	-,002	,003	-,006	,268	,032	-,080
Refinanciamiento	-,015	,001	,003	1,000	-,839	,387	-,132	,080	-,129	,011	,027	,007
Refinanciamiento Desfase1	,004	-,001	,001	-,839	1,000	-,745	,312	-,107	,068	,000	-,006	-,002
Refinanciamiento Desfase2	-,003	,003	-,002	,387	-,745	1,000	-,740	,327	-,131	,000	,001	-,004
Refinanciamiento Desfase3	-,001	,003	-,002	-,132	,312	-,740	1,000	-,758	,416	-,001	-,004	,002
Refinanciamiento Desfase4	-,001	-,003	,003	,080	-,107	,327	-,758	1,000	-,847	,000	,003	,006
Refinanciamiento Desfase5	-,006	,010	-,006	-,129	,068	-,131	,416	-,847	1,000	,006	-,001	-,012
Plazo Residual	,112	-,302	,268	,011	,000	,000	-,001	,000	,006	1,000	,135	,060
Indicador Tasa	-,113	,004	,032	,027	-,006	,001	-,004	,003	-,001	,135	1,000	-,010
Indicador Mora	-,054	,078	-,080	,007	-,002	-,004	,002	,006	-,012	,060	-,010	1,000

La matriz de correlaciones es importante para mostrar que variables tienen alta correlación, por lo cual estas deben ser eliminadas del modelo para evitar problemas de colinealidad de las variables. En la tabla 18, se consideran como correlación alta que las variables presentan más de 0,3 en el coeficiente de correlación.

3.6. Modelos

Aquí se comienza con el análisis multivariado y podremos observar cómo se comportan las variables antes mencionadas en forma de conjunto. Las variables como el monto original y el tamaño de la cuota se trabajaron agregando el logaritmo natural dando que son variables de cantidad, las variables binarias y porcentuales se agregaron a los modelos sin variación. En el caso de la variable spread tasa, se creó la variable denominada refinanciamiento de la siguiente forma:

$$\text{Refinanciamiento} = \ln\left(\frac{r_t}{i_t}\right)$$

Con:

r_t : Tasa del crédito en el período t

i_t : Tasa de mercado en el período t

La idea de crear esta variable surge a partir del estudio realizado por Lacour-Little, Marschoun y Maxam del año 2002.

De esta manera se podía capturar de mejor manera la diferencia de tasas entre la que tiene el crédito con la de mercado. La variable refinanciamiento fue confeccionada para los 5 desfases presentados en la segmentación. Esta variable se trabajó de manera distinta en el análisis univariado dado que la importancia de ese análisis era solo crear la intuición del comportamiento de las variables.

Se decidió trabajar la mora como si el cliente tuvo mora menor a 30 días o mayor a los 30 días para generar una diferencia en los tipos de clientes que incurren en el incumplimiento de algún pago, por lo tanto, ambas variables son binarias en caso de que el cliente tuvo aquel comportamiento durante la vigencia del crédito. Por último, la variable indicadora de la tasa mixta es una variable binaria que es 1 si el cliente tiene el crédito con tasa mixta y 0 si el cliente tiene tasa fija.

La variable por explicar es el prepago total que es una variable binaria, por lo cual, para estos modelos será la variable dependiente. Los modelos presentados corresponden a regresiones logísticas para explicar el comportamiento de prepago, donde se buscó conocer cuál era la incidencia de cada variable al comportamiento estudiado.

Por simplicidad en la tabla de resumen de las regresiones logísticas se agregó la magnitud de los betas, el error cuadrado y la significancia por variable para cada modelo. La regresión logística explica de la siguiente manera la probabilidad de realizar un prepago.

$$P(Y = 1) = \frac{1}{1 + \exp(-\alpha - \beta_1 X_1 - \beta_2 X_2 - \beta_3 X_3 - \dots - \beta_k X_k)}$$

Donde $\alpha, \beta_1, \beta_2, \beta_3, \dots, \beta_k$ son los parámetros del modelo, y exp denota la función exponencial. $X_1, X_2, X_3, \dots, X_k$ son las variables del modelo.

Para obtener los modelos se utilizó el método de selección hacia delante de Wald. El método utilizado va introduciendo las variables independientes en el análisis, contrastando la entrada de las variables basándose en la significación estadística de puntuación y contrasta la eliminación basándose en la probabilidad del estadístico de Wald.

El modelo elegido es el modelo número 11, de la tabla 20, esto usando el criterio de información de Akaike(AIC), complementando con la capacidad de clasificar los datos de manera correcta. En la tabla 19 se presentan los datos y características relevantes para los 11 modelos presentados anteriormente. El menor AIC es el modelo óptimo que en esta ocasión es último modelo de la tabla 19.

Tabla 19: Estadísticos Modelos

Modelo	(-2)Logaritmo de la verosimilitud	R cuadrado de Nagelkerke	Porcentaje de Clasificación(%)
1	360449,153	0,007	99,3
2	358755,629	0,011	98,1
3	357118,612	0,016	97,4
4	356680,649	0,017	97,4
5	356493,709	0,018	97,3
6	356365,77	0,018	97,2
7	356324,73	0,018	97,2
8	356285,727	0,018	97,2
9	356280,47	0,018	97,2
10	356280,584	0,018	97,2
11	356276,197	0,018	97,2

El AIC se calcula de la siguiente manera: $-2LL + 2N$ con la función LL la que aparece en la segunda columna de la tabla 19 y N el número de observaciones consideradas en la regresión. El R cuadrado muestra la cantidad de varianza explicada, mientras que la última columna indica la cantidad de estimación de casos teóricos bien clasificados, vemos que el descenso que se muestra en esa columna corresponde a un alza por parte de la clasificación de los créditos que obtuvieron el comportamiento de prepago mientras que desciende la clasificación correcta por parte de los no prepagados.

Otro aspecto que considerar en nuestro modelo es el porcentaje de clasificación de los casos, es decir, cuantos casos de la muestra son clasificados correctamente según el comportamiento. Para el modelo óptimo se puede notar que el 97,2% de los casos tuvo una clasificación correcta.

Tabla 20: Resumen Modelos Método de Wald

Modelos	Datos	Constante	LN Monto Original	LN Tamaño Cuota	Refinanciamiento	Refinanciamiento D1	Refinanciamiento D2	Refinanciamiento D3	Refinanciamiento D4	Refinanciamiento D5	Plazo Residual	Mora 30 días	Mora más de 30	Indicador TM	
Modelo 1	β	-4,568									-1,192				
	Error Estándar	0,017									0,024				
	Sig.	0,000									0,000				
Modelo 2	β	-11,715	0,416								-1,419				
	Error Estándar	0,172	0,010								0,024				
	Sig.	0,000	0,000								0,000				
Modelo 3	β	-11,244	0,395								-1,411	-0,671			
	Error Estándar	0,172	0,010								0,024	0,018			
	Sig.	0,000	0,000								0,000	0,000			
Modelo 4	β	-11,476	0,397								-1,358	-0,702		0,259	
	Error Estándar	0,173	0,010								0,025	0,018		0,013	
	Sig.	0,000	0,000								0,000	0,000		0,000	
Modelo 5	β	-12,150	0,430				0,352				-1,291	-0,723		0,294	
	Error Estándar	0,180	0,010				0,026				0,025	0,018		0,013	
	Sig.	0,000	0,000				0,000				0,000	0,000		0,000	
Modelo 6	β	-12,108	0,428				0,359				-1,292	-0,498	-0,277	0,305	
	Error Estándar	0,180	0,010				0,026				0,025	0,027	0,025	0,013	
	Sig.	0,000	0,000				0,000				0,000	0,000	0,000	0,000	
Modelo 7	β	-12,077	0,426				1,059			-0,719	-1,295	-0,498	-0,277	0,305	
	Error Estándar	0,180	0,010				0,112			0,111	0,025	0,027	0,025	0,013	
	Sig.	0,000	0,000				0,000			0,000	0,000	0,000	0,000	0,000	
Modelo 8	β	-12,066	0,425				-0,486	2,147			-1,326	-1,297	-0,497	-0,277	0,304
	Error Estándar	0,181	0,010				0,273	0,343			0,147	0,025	0,027	0,025	0,013
	Sig.	0,000	0,000				0,075	0,000			0,000	0,000	0,000	0,000	0,000
Modelo 9	β	-12,057	0,425		-0,403		0,130	1,833			-1,230	-1,298	-0,497	-0,277	0,302
	Error Estándar	0,181	0,010		0,176		0,383	0,370			0,153	0,025	0,027	0,025	0,013
	Sig.	0,000	0,000		0,022		0,735	0,000			0,000	0,000	0,000	0,000	0,000
Modelo 10	β	-12,057	0,425		-0,361			1,941			-1,25	-1,298	-0,497	-0,277	0,302
	Error Estándar	0,181	0,010		0,126			0,187			0,14	0,025	0,027	0,025	0,013
	Sig.	0,000	0,000		0,004			0,000			0,000	0,000	0,000	0,000	0,000
Modelo 11	β	-12,056	0,425		-0,451			2,551	-1,041		-0,729	-1,298	-0,497	-0,277	0,302
	Error Estándar	0,181	0,010		0,133			0,347	0,498		0,286	0,025	0,027	0,025	0,013
	Sig.	0,000	0,000		0,001			0,000	0,000		0,000	0,000	0,000	0,000	0,000

Por lo cual a continuación se presenta el modelo óptimo:

Tabla 21: Modelo Óptimo

	B	Error estándar	Wald	gl	Sig.	Exp(B)	
Modelo 11	LN MONTO ORIGINAL	0,425	0,01	1726,005	1	0	1,53
	Refinanciamiento	-0,451	0,133	11,492	1	0,001	0,637
	Refinanciamiento Desfase3	2,551	0,347	54,165	1	0	12,821
	Refinanciamiento Desfase4	-1,041	0,498	4,372	1	0,037	0,353
	Refinanciamiento Desfase5	-0,729	0,286	6,493	1	0,011	0,482
	Plazo Residual	-1,298	0,025	2600,403	1	0	0,273
	Mora 0-30 días	-0,497	0,027	337,021	1	0	0,608
	Mora más 30 días	-0,277	0,025	122,108	1	0	0,758
	Indicador TM	0,302	0,013	557,621	1	0	1,353
	Constante	-12,056	0,181	4458,19	1	0	0

El modelo presentado en la tabla N°21 es el mejor según los criterios utilizados, que fue el AIC y el porcentaje de clasificación. Se puede comentar que tal como el análisis univariado evidenciaba, los créditos con mayor monto aportan positivamente al prepago, cabe mencionar que dado la alta correlación con el monto de la cuota la segunda variable fue eliminada en el proceso realizado. Observando el refinanciamiento, se puede notar que el refinanciamiento con 3 desfases es el que absorbe toda la diferencia de la tasa del crédito con la de mercado, siendo esta la variable con mayor influencia sobre el prepago.

Para el plazo residual se puede decir que mientras menos tiempo le quede al crédito de vida mayor será la tendencia a prepagarlos. Si analizamos la mora se infiere que si el cliente ha incurrido en ella no tendera a prepagar los créditos. Por último, observamos que si el crédito fue concebido a tasa mixta existirá una mayor tendencia a prepagar que frente a los otorgados a tasa fija.

En general se extrae que el análisis multivariado complementa numéricamente la intuición generada por el análisis univariado, siendo un complemento para entender de mejor manera el comportamiento de prepago.

Capítulo 4

4. Interpretación de Resultados

Comenzaremos comentando acerca de los resultados que trajo la segmentación presentada para luego complementar los resultados con el modelo óptimo que se obtuvo.

El principal resultado obtenido del trabajo realizado fue que existen variables que inciden en cierto grado el comportamiento de prepago. Luego de realizar un análisis univariado para posteriormente realizar el multivariado se pueden extraer los siguientes resultados.

Tanto la variable monto original como el tamaño de la cuota inciden positivamente al prepago, esto quiere decir que mientras mayores sean estos valores, la tendencia a prepagar los créditos aumenta. La intuición de esto es que los créditos con montos mayores tienen un mayor poder de negociación con las diferentes instituciones bancarias por lo cual pueden conseguir mejores tasas y elegir la mejor opción para ellos. También se encuentra la idea de que las personas que piden créditos con mayores montos tienen mayores ingresos y esta variable según estudios previos tienen una mayor cantidad de años de estudios por lo cual tienen una mejor formación y son más cultos financieramente. Cabe destacar que en el modelo obtenido la variable monto original absorbe el efecto del tamaño de la cuota por la alta correlación existente entre ambas.

La variable spread tasa es la variable donde canalizamos el efecto macroeconómico de la economía dado que tomamos la tasa hipotecaria de mercado para compararla con la tasa del crédito en una fecha determinada. La tasa hipotecaria varía según las expectativas del mercado, la interpretación que tenemos es que cuando la tasa de mercado es menor a la tasa del crédito, es decir, el spread es positivo, el prepago debiese ser mayor. La lógica de esta variable es que las personas refinancian sus créditos para contraer otro a menor tasa, es decir, que el banco le preste dinero a un menor precio, una de las principales causas del prepago según la bibliografía y el levantamiento de información es este, el saldo remanente se paga con otro crédito. Se obtuvo que el spread con mayor incidencia sobre el prepago es el spread con 3 desfases, se explica de que el cliente se demora tres meses luego de observar las tasas de mercado para decidir prepagar, también se relaciona con los costos y tiempo transaccionales de parte de la institución bancaria.

A continuación, vemos que el plazo residual presenta un mayor prepago cuando al crédito le queda menos tiempo para ser terminado, la intuición en este caso va en la línea que el cliente tiene incentivos mayores a pagar el crédito si le queda un plazo menor, por un lado, esta terminar el compromiso de pago con algún acreedor, lo que da holgura financiera al cliente, por otro lado, es más fácil pagar un crédito que le queda menor cantidad de dinero para ser pagado en su totalidad. Al pagar un crédito por completo la persona puede obtener otro crédito con mayor facilidad para endeudarse en otro aspecto que él encuentre necesario.

Evaluaremos en este caso el tipo de tasa a la cual se contrajo el crédito, donde se presente evidencia que la tasa mixta se prepagaba más que la fija, a priori, la intuición nos decía que la tasa fija se prepagaba en una mayor cantidad pero el análisis dio el resultado contrario, pero

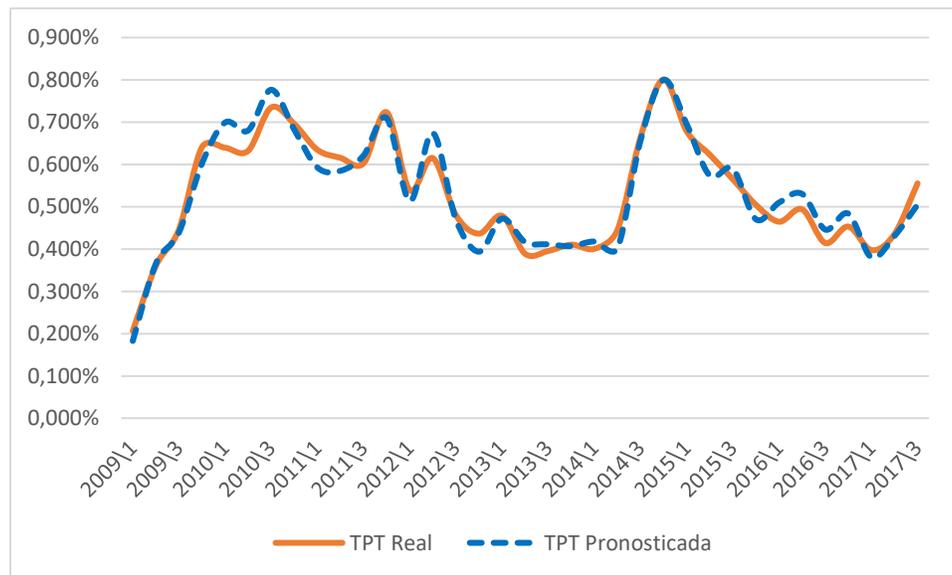
cabe mencionar que la tasa mixta tiene dos aristas las cuales son un segmento con tasa variable y otro con tasa fija, se vio en que periodos el crédito se pagaba y dio como resultado que la tasa mixta se prepaga cuando esta con tasa fija. La intuición de lo anterior es que los clientes que conciben créditos con tasa mixta están más preocupados de cómo se mueve el mercado entonces son personas más informadas y están revisando constantemente el crédito para refinanciarlo y poder abrir uno nuevo a menor costo.

Por último esta la variable de la mora, donde se vio si los créditos que habían incurrido en ese comportamiento se prepagaban más o menos, la segmentación no entrego evidencia significativa pero al segmentar por mora de 0 hasta 30 días y mora de 31 a más vemos que la mora de 0 hasta 30 día tiene un menor efecto al prepago que la mora de 31 o más, esto se cree que ocurre dado que en la mora de 31 o más se encuentran créditos que tienen mucho tiempo de incumplimiento reiterativo por lo cual el banco elimina de la base esos créditos, entonces se tiene que el crédito cesa antes de su fecha de término.

Observando los resultados del modelo óptimo el spread tasa con 3 desfases en la variable que mayor incidencia tiene sobre el prepago, esto coincide con la literatura estudiada, y como se mencionó anteriormente el refinanciamiento termina siendo el principal motivo del prepago.

Para finalizar, en la ilustración 17 se observa el poder predictivo del modelo, donde clasifica correctamente al 97% aproximado de los casos, el modelo falla principalmente en los picos de prepago donde la clasificación peor. Para realizar la predicción de la TPT, se tomaron aleatoriamente 70% de los casos y se evaluó como predice el modelo en la misma ventana de tiempo estudiada.

Ilustración 17: Predicción Modelo



Capítulo 5

5. Conclusiones

En este informe se determinó el mejor modelo para explicar la probabilidad de prepago con las variables que se tenían. La idea era ver buscar y cuantificar que variables podían explicar de la mejor manera tasa de 0,51% de prepago mensual. Luego del trabajo realizado se puede decir que el objetivo general del trabajo fue cumplido satisfactoriamente.

Se trabajó con la información a partir del año 2009 hasta septiembre del 2017. El levantamiento de información bibliográfica fue crucial para lograr un entendimiento global del problema y poder identificar las principales variables relevantes que influyen en el comportamiento de prepago. De esta manera se logró cumplir el primer objetivo específico que era crear la historia de prepago mensual para la cartera.

También se determinó que existen principalmente cuatro efectos que influyen en la decisión de ejercer la opción de prepago, éstos son: rotación habitacional, tasa de interés de mercado, nivel de actividad económica y características de los créditos. Dado estos efectos se buscaron variables que pudieran explicar los cuatro efectos mencionados, por lo cual las variables seleccionadas fueron el monto original, el valor de la cuota, la tasa y sus spreads; esta variable es la que canalizaba el efecto macroeconómico al incluir la tasa de mercado en sus análisis; el tipo de tasa a la cual se contrajo el crédito, el plazo residual del crédito y por último se trabajó la mora.

Se comenzó haciendo un análisis exploratorio de la información contenida en la base de datos, para proceder con la segmentación que ayudó a identificar en primer lugar cuál era el efecto de cada variable al prepago. De este análisis se creó la intuición que la mayoría de las variables expuestas efectivamente tenían un comportamiento de prepago distinto para cada grupo, pudiendo identificar las variables que influían en la decisión de prepago el cual era nuestro segundo objetivo específico.

Finalmente, para cumplir nuestro tercer objetivo específico y de esta manera lograr el general con las variables trabajadas se estimaron diversos modelos logísticos para poder determinar el óptimo. El criterio de AIC fue el que se utilizó para elegir el modelo, dado que es un modelo logístico se puede estimar la probabilidad de prepago dadas ciertas características del crédito.

Tomando el mejor modelo, se extrae que la variable que tiene una mayor influencia positiva en la decisión de prepago es el refinanciamiento con tres desfases. Por lo cual se comprueba que los acreedores tienen un tiempo de retardo en tomar la decisión de refinanciar un crédito, en este caso para los clientes del banco son 3 meses los que se demoran en tomar de decisión luego de observar las tasas de mercado.

Del modelo también se observa que los créditos con mayores montos tienen una tasa más alta de prepago al igual que los créditos con tasa mixta, la intuición para ambos casos fue

explicada en el capítulo 5. En cuanto para el refinanciamiento, refinanciamiento con desfase 4 y 5, plazo residual y la mora aportan de manera negativa al comportamiento de prepago.

Por otro lado, se concluye dado que los créditos que son prepagados son clientes “buenos”, es decir, no tendrán problemas de incumplimiento en el futuro, la cartera de vivienda se ve deteriorada al prepagarse los créditos.

Finalmente se concluye que dado a la investigación realizada el banco puede anticiparse a que créditos tienen una probabilidad mayor de prepagarse. En el trabajo se descubrió cuales eran algunas de las variables intrínsecas de los créditos que favorecían el prepago. Dado que el prepago le genera al banco descalce en sus flujos, poder anticiparse y canalizar cual de los créditos de su cartera tiene una tendencia mayor a prepagarse podría ayudar a disminuir la incertidumbre, es decir, el riesgo que el prepago genera.

También dado que uno de los principales motivos del prepago es el refinanciamiento el cual puede que el cliente se fugue a otra institución bancaria, poder saber cual de los créditos tiene mayor probabilidad de generar un prepago aumentaría la retención de los clientes.

5.1. Acápites I: Recomendaciones

Luego de finalizado el trabajo se tienen un par de sugerencias para complementar la estimación del prepago. En primer lugar, para poder estimar de manera más robusta la decisión de prepago sería interesante agregar variables demográficas de los clientes, las cuales pueden ser el ingreso, nivel de estudios, el valor de la vivienda, la zona de la vivienda entre otras. Estas variables se utilizan constantemente en la bibliografía estudiada y pueden complementar de buena forma la estimación de la probabilidad de prepago.

La segunda recomendación que se realiza para continuar profundizando el trabajo realizado es que en el artículo de Lacour-Little 2002 para el concepto de refinanciamiento utilizan una variable llamada “burnout”.

$$Burnout_{it} = \sum_{\tau=\tau_{orig}}^t \max\{\log(c_i/r_\tau), 0\}$$

donde c es la tasa de la cuota del crédito i , r es la tasa de mercado para el período τ y el τ_{orig} es el tiempo de originación del crédito. Este concepto hace relación con el agotamiento del crédito y la idea es acumular la liquidez de la opción de prepago a través del tiempo. Además, durante los períodos de alta volatilidad de la tasa de interés esta medida asigna valores grandes a la variable.

Capítulo 6

6. Bibliografía

[1] MATTEO CONSALVI, GIOVANNI SCOTTO DI FRECA; Measuring prepayment risk: an application to UniCredit Family Financing

[2] IVAN BRATKO, RYSZARD S. MICHALSKI, MIROSLAV KUBAT; Machine learning and data mining methods and applications

[3] ALLISON, PAUL. 1995. Survival Analysis Using SAS®: A Practical Guide. 7ª Edición. SAS Institute, Inc.

[4] ALLISON, PAUL. 1999. Logistic Regression Using The SAS® System. 4ª Edición. SAS Institute, Inc.

[5] BERRY, MICHAEL J.A., LINOFF, GORDON S. 2004. Data Mining Techniques for Marketing, Sales and Customer Relationship Management. 2ª Edición. Wiley

[6] ANDREAS BOHN, MARIJE ELKENBRACHT-HUIZING, ALM in banking: interest rates, liquidity and the balance sheet

[7] BARROWS H. 1995. Rational Prepayment and the Valuation of Mortgage- Backed Securities. Richard Stanton. Haas School of Business. U.C. Berkeley.

[8] DE OVANDO, N. 2005. Evidencia de Prepago en Créditos Hipotecarios a la Vivienda: El Caso de México. México. Instituto Tecnológico Autónomo de México.

[9] SELVAGGI, M.. 2000. Precancelaciones hipotecarias en Argentina: primeras evidencias basadas en modelos de duración. 2000. Buenos Aires, Argentina. Gerencia de Estudios Económicos. Banco Hipotecario S.A.

[10] VUCINA LJUBETIC, E. A. 2004. Valorización, opciones y spread de letras de crédito hipotecario en Chile. Tesis para optar al grado de Magíster en Ciencias de la Ingeniería. Santiago, Chile: Pontificia Universidad Católica de Chile, 2004.

[11] LACOUR-LITTLE, Michael. MARSCHOUN, Michael A. MAXAM, Clark L. Improving Parametric Mortgage Prepayment Models with Non-Parametric Kernel Regression. Journal of Real Estate Research, Vol. 24, No. 3, 2002.

[12] LACOUR-LITTLE, Michael,. MAXAM, Clark L. Applied Nonparametric Regression Techniques Estimating Prepayments on Fixed-Rate Mortgage-Backed Securities.

[13] MAJLUF ADAUY, Cristian. Desarrollo de un Modelo de Prepago de Leras Hipotecarias para el banco BICE.

Capítulo 7

7. Anexos

ANEXO A

Tabla 22: Participación de Mercado del Financiamiento de Viviendas Segmentado por Institución Financiera a junio del 2017

Banco	Monto \$MM	Participación de Mercado	Participación de Mercado Acumulada
Banco Santander	8.861.371	20,57%	20,57%
Banco del Estado	8.488.328	19,70%	40,27%
Banco de Chile	7.249.122	16,83%	57,1%
Banco de Crédito e Inversiones	5.464.171	12,68%	69,78%
Itaú Corpbanca	4.010.253	9,31%	79,09%
Banco Bilbao Vizcaya Argentaria, Chile	3.646.717	8,46%	87,55%
Scotiabank	3.416.490	7,93%	95,48%
Banco Bice	783.476	1,82%	97,3%
Banco Security	574.532	1,33%	98,63%
Banco Falabella	439.112	1,02%	99,65%
Banco Consorcio	93.154	0,22%	99,87%
Banco Ripley	34.980	0,08%	99,95%
Banco Internacional	23.651	0,05%	100%
Banco BTG Pactual	0	0%	100%
Banco de la Nación Argentina	0	0%	100%
Banco do Brasil S.A.	0	0%	100%
China Construction	0	0%	100%
HSBC Bank (Chile)	0	0%	100%
JP Morgan Chase Bank, N.A.	0	0%	100%
The Bank of Tokyo-Mitsubishi UFJ, Ltd.	0	0%	100%
Sistema Bancario	43.085.357	100%	100%

Fuente: Elaboración propia a partir de datos de la SBIF.

ANEXO B

Tabla 23: Montos en \$MM de la evolución temporal de créditos hipotecarios en stock del sistema

Fecha	MHNE	MHE	LCH	Total
Junio 12	18.720.978	2.045.680	1.903.966	22.670.624
Junio 13	21.791.708	1.999.862	1.687.777	25.479.347
Junio 14	25.416.428	2.110.381	1.510.864	29.037.674
Junio 15	30.015.158	2.178.975	1.334.989	33.529.122
Junio 16	34.604.062	2.280.716	1.175.807	38.060.585
Junio 17	38.722.456	2.331.389	1.019.309	42.073.154

Tabla 24: Porcentajes de la evolución temporal del stock de los créditos hipotecarios

Fecha	MHNE	MHE	LCH
Junio 12	82,58%	9,02%	8,40%
Junio 13	85,53%	7,85%	6,62%
Junio 14	87,53%	7,27%	5,20%
Junio 15	89,52%	6,50%	3,98%
Junio 16	90,92%	5,99%	3,09%
Junio 17	92,04%	5,54%	2,42%

ANEXO C

Ilustración 18: Test de Medias Monto Original

Prueba t para dos muestras suponiendo		
	Variable 1	Variable 2
Media	0,00616318	0,0038913
Varianza	3,5384E-06	5,2155E-07
Observacion	35	35
Varianza agr	2,03E-06	
Diferencia hi	0	
Grados de lib	68	
Estadístico t	6,67053835	
P(T<=t) una c	2,7674E-09	
Valor crítico	1,66757228	
P(T<=t) dos c	5,5348E-09	
Valor crítico	1,99546893	

ANEXO D

Ilustración 19: Test de Medias Valor Cuota

Prueba t para dos muestras suponiendo		
	Variable 1	Variable 2
Media	0,004317	0,0066652
Varianza	6,6403E-07	4,1184E-06
Observacion	35	34
Varianza agrupada	2,3655E-06	
Diferencia hipotética	0	
Grados de libertad	67	
Estadístico t	-6,34054286	
P(T<=t) una cola	1,1244E-08	
Valor crítico t	1,66791611	
P(T<=t) dos colas	2,2489E-08	
Valor crítico t	1,99600835	

ANEXO E

Ilustración 20: Test de Medias Spread Tasa

Prueba t para dos muestras suponiendo		
	Variable 1	Variable 2
Media	0,00483768	0,00521
Varianza	1,3599E-06	8,1877E-07
Observacion	35	35
Diferencia hipotética	0	
Grados de libertad	64	
Estadístico t	-1,49225845	
P(T<=t) una cola	0,07027235	
Valor crítico t	1,66901303	
P(T<=t) dos colas	0,1405447	
Valor crítico t	1,99772965	

Ilustración 21: Test de Medias Spread Tasa Desfase 1

Prueba t para dos muestras suponiendo		
	Variable 1	Variable 2
Media	0,00477661	0,00523065
Varianza	1,2889E-06	7,6241E-07
Observacion	35	35
Diferencia hip	0	
Grados de lib	64	
Estadístico t	-1,87549532	
P(T<=t) una c	0,03264378	
Valor crítico	1,66901303	
P(T<=t) dos c	0,06528756	
Valor crítico	1,99772965	

Ilustración 22: Test de Medias Spread Tasa Desfase2

Prueba t para dos muestras suponiendo		
	Variable 1	Variable 2
Media	0,00476533	0,00524065
Varianza	1,4256E-06	7,1182E-07
Observacion	35	35
Diferencia hip	0	
Grados de lib	61	
Estadístico t	-1,9234416	
P(T<=t) una c	0,02954752	
Valor crítico	1,67021948	
P(T<=t) dos c	0,05909505	
Valor crítico	1,99962358	

Ilustración 23: Test de Medias Spread Tasa Desfase 3

Prueba t para dos muestras suponiendo		
	Variable 1	Variable 2
Media	0,0047617	0,00524975
Varianza	1,4862E-06	6,8855E-07
Observacion	35	35
Diferencia hip	0	
Grados de lib	60	
Estadístico t	-1,95792129	
P(T<=t) una c	0,02744711	
Valor crítico	1,67064886	
P(T<=t) dos c	0,05489423	
Valor crítico	2,00029782	

Ilustración 24: Test de Medias Spread Tasa Desfase 4

Prueba t para dos muestras suponiendo		
	Variable 1	Variable 2
Media	0,00474058	0,00525274
Varianza	1,4069E-06	7,4396E-07
Observacion	35	35
Diferencia hip	0	
Grados de lib	62	
Estadístico t	-2,06604357	
P(T<=t) una c	0,02150459	
Valor crítico	1,66980416	
P(T<=t) dos c	0,04300919	
Valor crítico	1,99897152	

Ilustración 25: Test de Medias Spread Tasa Desfase 5

Prueba t para dos muestras suponiendo		
	Variable 1	Variable 2
Media	0,0047471	0,0052371
Varianza	1,3409E-06	7,726E-07
Observacion	35	35
Varianza agr	1,0568E-06	
Diferencia hip	0	
Grados de lib	68	
Estadístico t	-1,99402102	
P(T<=t) una c	0,02508091	
Valor crítico	1,66757228	
P(T<=t) dos c	0,05016182	
Valor crítico	1,99546893	

ANEXO F

Ilustración 26: Segmentación Spread sin Desfase

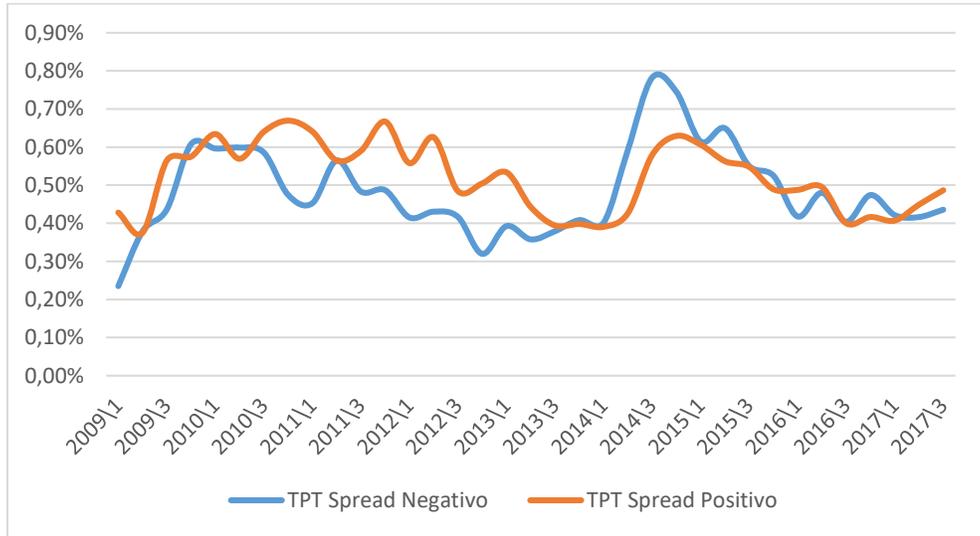


Ilustración 27: Segmentación Spread Tasa 1 Desfase

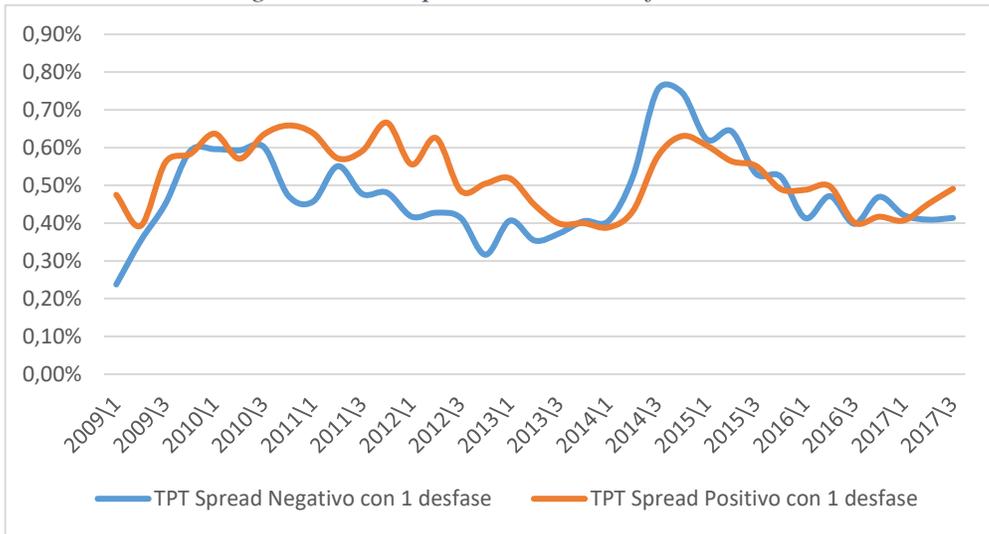


Ilustración 28: Segmentación Spread 2 Desfases

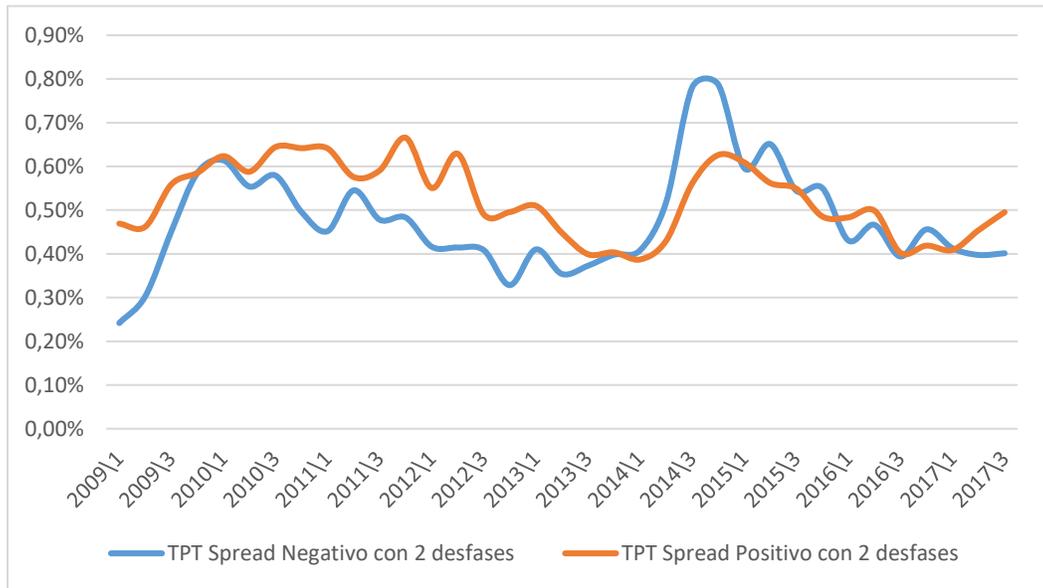


Ilustración 29: Segmentación Spread Tasa 3 Desfases

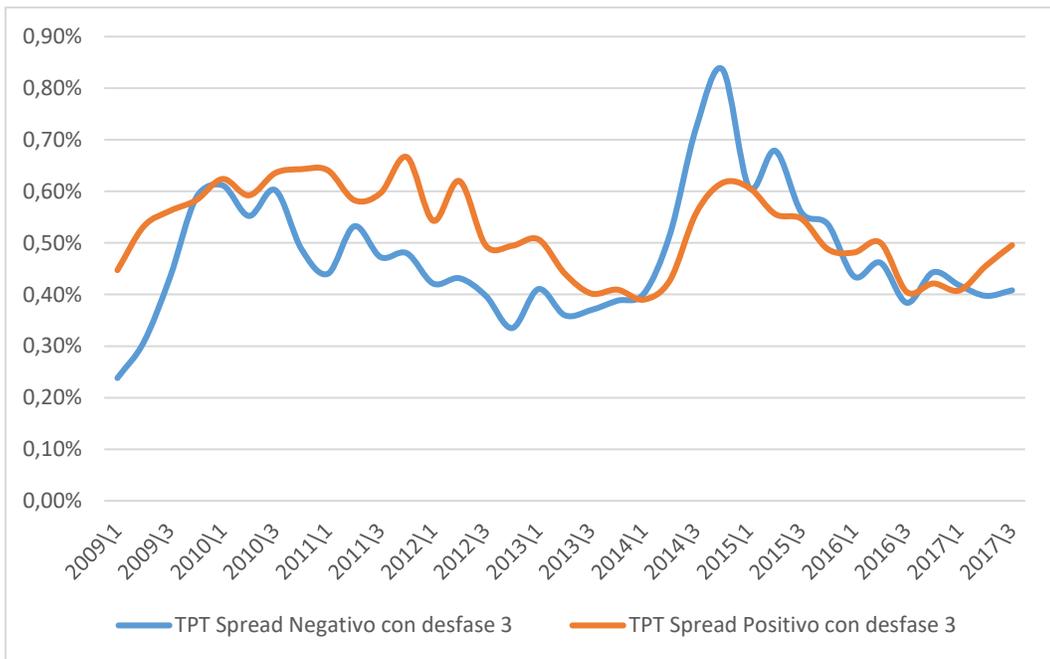


Ilustración 30: Segmentación Spread Tasa 4 Desfases

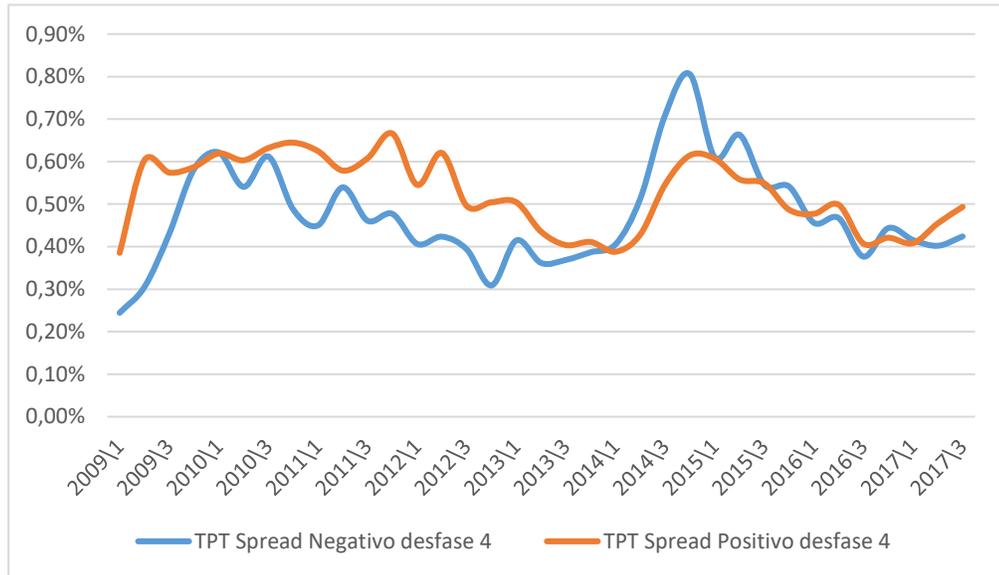
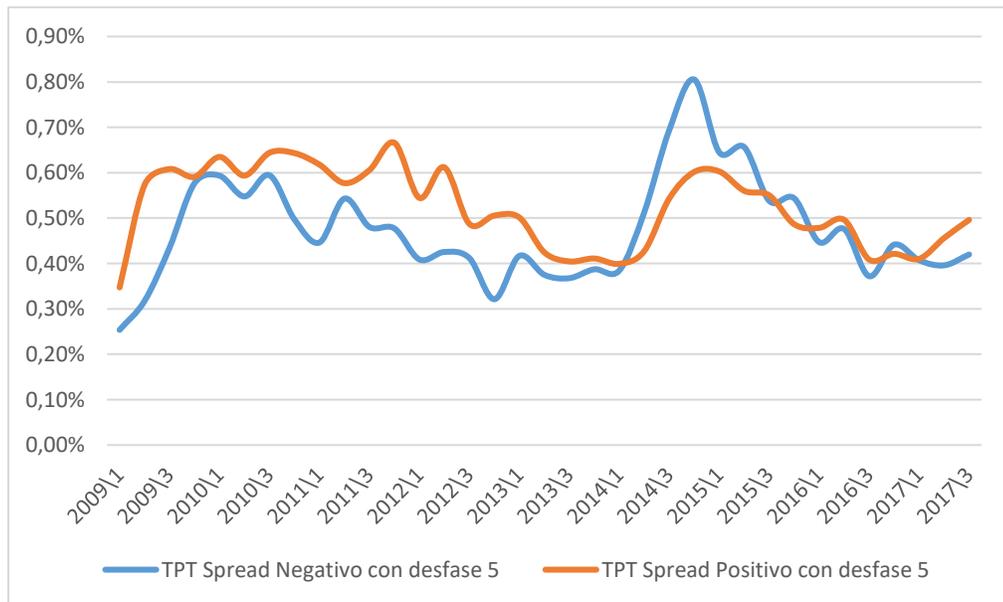


Ilustración 31: Segmentación Spread Tasa 5 Desfases



ANEXO G

Ilustración 32: Test de Medias Plazo Residual

Prueba t para dos muestras suponiendo		
	Variable 1	Variable 2
Media	0,00732732	0,00522368
Varianza	2,7493E-06	1,7583E-06
Observacion	35	35
Diferencia hipotética	0	
Grados de libertad	65	
Estadístico t	5,86179696	
P(T<=t) una cola	8,3571E-08	
Valor crítico (una cola)	1,66863598	
P(T<=t) dos colas	1,6714E-07	
Valor crítico (dos colas)	1,99713791	

ANEXO H

Ilustración 33 Test de Medias Tipo Tasa

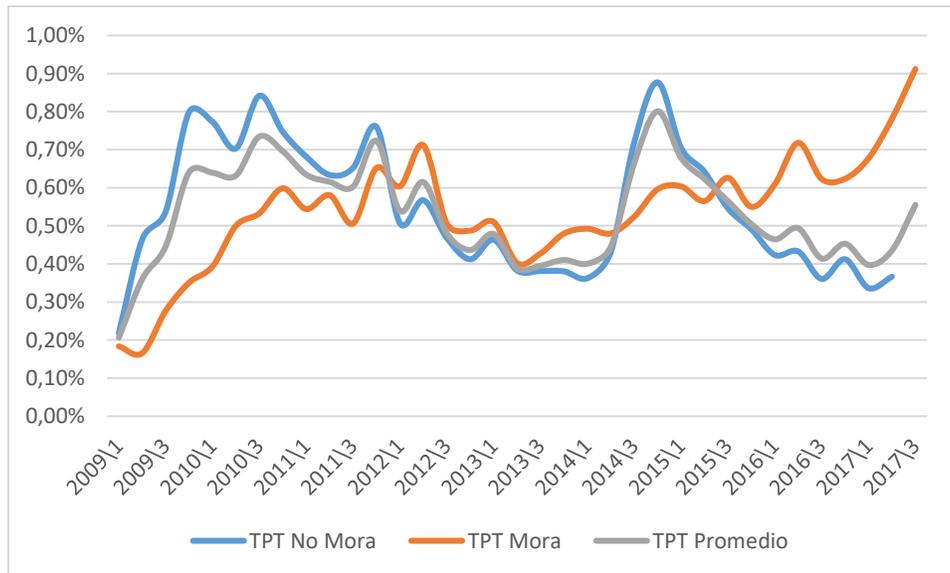
Prueba t para dos muestras suponiendo		
	Variable 1	Variable 2
Media	0,00511541	0,00646791
Varianza	6,7934E-06	2,6536E-06
Observacion	35	35
Varianza agrupada	4,7235E-06	
Diferencia hipotética	0	
Grados de libertad	68	
Estadístico t	-2,60330863	
P(T<=t) una cola	0,00566304	
Valor crítico (una cola)	1,66757228	
P(T<=t) dos colas	0,01132607	
Valor crítico (dos colas)	1,99546893	

ANEXO I

Ilustración 34: Test de Medias Mora

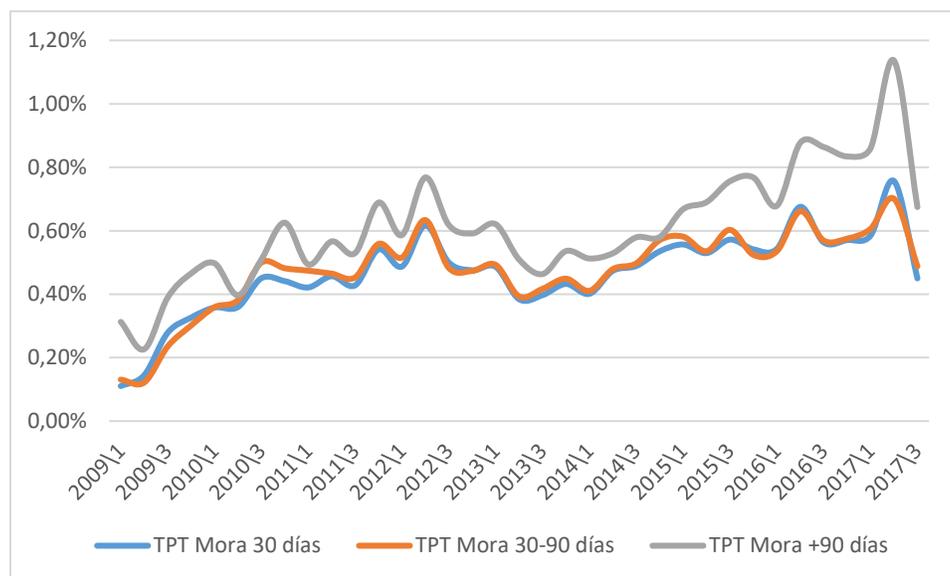
Prueba t para dos muestras suponiend		
	Variable 1	Variable 2
Media	0,0054196	0,00530617
Varianza	2,8293E-06	1,704E-06
Observacion	35	35
Diferencia hi	0	
Grados de lib	64	
Estadístico t	0,3151566	
P(T<=t) una c	0,37683403	
Valor crítico	1,66901303	
P(T<=t) dos c	0,75366806	
Valor crítico	1,99772965	

Ilustración 35: Segmentación Mora



ANEXO J

Ilustración 36: Historia Mora por Segmentos



ANEXO K: Claves de la base

- **ID_CLIENTE:** indica el código asociado a un cliente
- **OPERACIÓN:** indica el código asociado a la operación para un respectivo crédito.
- **TIPO_OPERACIÓN:** podía ser HIP300 o PL300 dependiendo el tipo de crédito.
- **FECHA_PROCESO:** indica la fecha en que se encuentra el crédito en ese momento.
- **FECHA_APERTURA:** indica la fecha en que se concibió el crédito.
- **CODIGO_MONEDA:** indica la moneda a la cual se contrajo el crédito, 998 UF, 999 pesos chilenos.
- **VALOR_ORIGINAL:** valor del monto del crédito, varía según la UF de determinada fecha.
- **SALDO_MORA_1:** es el monto que se encuentra en mora entre 0 y 30 días.
- **SALDO_MORA_2:** es el monto que se encuentra en mora entre 31 y 89 días.
- **SALDO_VENCIDO:** es el monto que se encuentra en mora más de 90 días.
- **MONTO_PAGADO_EN_EL_MES:** es el monto que el cliente pago en esa fecha.
- **VENCIMIENTOS_DEL_MES:** es el monto que debe pagar en ese mes el cliente.
- **NRO_CUO_POR_PAGAR:** es la cantidad de cuotas faltantes del crédito.
- **SALDO_CAP_IFRS:** es el saldo que le va quedando al cliente.
- **INT_DEVENG_IFRS:** es el monto que se genera por los intereses
- **REAJ_DEVENG_IFRS:** es el monto total, del crédito más lo generado por los intereses.
- **OPE_TAS_INT:** es la tasa a la cual se contrajo el crédito.
- **CUOTAS_ORIGINALES:** es la cantidad de cuotas originales del crédito.

ANEXO L

Tecnología

Este anexo se indica cuáles fueron los softwares utilizados para el desarrollo de esta memoria y porque se utilizaron además del sistema operativo con el cual se trabajó.

SPSS Statistics

SPSS es un software de la compañía IBM, es un software estadístico que ofrece técnicas de recolección de datos y analítica predictiva para solucionar múltiples problemas empresariales y de investigación. Brinda varias técnicas, que incluyen pruebas de hipótesis lo que facilita la gestión de los datos, la selección y la ejecución de análisis y el intercambio de resultados.

Para la presente memoria se usó este software para manejar las bases de datos de una manera sencilla y poder crear nuevas variables necesarias para el análisis. También se utilizó el programa para sacar diversas estadísticas descriptivas de las variables mencionadas en el capítulo 3 y poder comparar los grupos de estudio.

Se selecciono este software debido a su extenso uso, la facilidad para manejar bases de datos y también dado que el banco que facilito la información se utiliza.

R

R es un entorno y lenguaje de programación con un enfoque al análisis estadístico, R es una implementación de software libre del lenguaje S pero con soporte de alcance estático. Se trata de uno de los lenguajes más utilizados en investigación por la comunidad estadísticas, siendo además muy popular en el campo de la minería de datos, la investigación biomédica, la bioinformática y las matemáticas financieras. A esto contribuye la posibilidad de cargar diferentes bibliotecas o paquetes con funcionalidades de cálculo y gráficas.

En esta memoria se trabajó con este software para realizar la variable “PREPAGO_TOTAL” con una función recursiva donde se iba comparando un período con el siguiente para ver si el crédito había cesado según el número de cuotas remanentes, esta variable es de vital importancia para la presente memoria dado que sirve para segmentar por grupos y luego usarla de variable dependiente para los diversos modelos planteados.

Se eligió este software para trabajar en esta memoria debido a que es gratuito y tiene mucho material de donde sacar información en la web en caso de tener alguna duda. Por otro lado el lenguaje de programación es bastante intuitivo y su enfoque estadístico es de vital importancia dado el contexto de la memoria.

Microsoft Excel

Microsoft Excel es una aplicación de hojas de cálculo que forma parte de la suite de oficina Microsoft Office, es una aplicación utilizada en tareas financieras y contables, con fórmulas, gráficos y un lenguaje de programación.

Este programa se utilizó para ordenar los resultados obtenidos por el software SPSS y poder presentarlos de buena manera. En Excel se realizaron la mayoría de las tablas y gráficos presentes en esta memoria, fue un complemento para los programas mencionados en los dos puntos anteriores.

Se utiliza este software por el amplio dominio que tiene el investigador con él y por su facilidad de manejo y estar asociado al sistema operativo con el cual se trabajó que se describe en el siguiente punto.

Sistema Operativo

El grueso de esta memoria se trabajó con un notebook Lenovo ideapad 310 con las siguientes características:

- Edición de Windows: Windows 10 Home Single Languaje
- Procesador: Intel(R) Core™ i7-6500U CPU 2,5GHz 2.59GHz
- Memoria Instalada(RAM): 4,00 GB
- Tipo de sistema: Sistema operativo de 64 bits, procesador x64