



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

DISEÑO DE MECANISMOS PARA LA ASIGNACIÓN ÓPTIMA DE PRODUCTOS DE
UN BANCO EN MICROFINANZAS

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN GESTIÓN DE OPERACIONES

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

ANDRÉS GARRIDO CASTRO

PROFESOR GUÍA:

SR. JOSÉ MIGUEL CRUZ GONZÁLEZ

MIEMBROS DE LA COMISIÓN:

SR. GIORGI GIULIO PARRA DE BLASI

SR. RICHARD WEBER HAAS

SANTIAGO DE CHILE
2018

RESUMEN DE LA MEMORIA PARA OPTAR AL
TÍTULO DE: Ingeniero Civil Industrial y grado de
Magíster en Gestión de Operaciones
POR: Andrés Garrido Castro
FECHA: 08/06/2018
PROF. GUÍA: José Miguél Cruz

DISEÑO DE MECANISMO PARA LA ASIGNACIÓN ÓPTIMA DE PRODUCTOS DE UN BANCO EN MICROFINANZAS

Muchos países de la OECD, poseen programas de apoyo y desarrollo de pequeñas y medianas empresas, debido a que representan aproximadamente el 95 % del total de empresas, entre el 60 % y 70 % del empleo en la mayoría de los países de la OECD y un 33 % del PIB en economías emergentes. En Chile, la microempresa se distribuye de forma más homogénea a lo largo del país, que la mediana y gran empresa, en particular en las zonas más extremas y, si bien son pequeñas, tienen opciones de crecimiento, con tasas de rentabilidad muy superiores al resto de la industria.

Este trabajo surge como un desafío de entender mejor a los clientes del BancoEstado Microempresas (BEME) y tiene como objetivo, formular un modelo de optimización que permita diseñar un mecanismo de asignación de productos para un segmento de clientes de BEME con decisiones óptimas del “éxito” de los microempresarios en un horizonte de tiempo, teniendo como variable de control la entrega de un producto de crédito, capturando la dinámica de pago como el mínimo entre lo que puede y lo que debe pagar un cliente con variables de estado del balance, considerando adicionalmente restricciones de pérdida esperada. Para su resolución, se emplea la metodología *Sample Average Approximation* (SAA), garantizando una significancia estadística del 99 % en la mayoría de los experimentos.

Se realizaron diversos experimentos computacionales con diferentes distribuciones de rentabilidad de clientes, evaluando convergencia y tiempos de resolución. Se concluye que el modelo es satisfactorio al lograr modelar los comportamientos de pago como una opción e interpretar sus restricciones lineales como un portafolio de opciones. Al incorporar las restricciones de riesgo en provisiones, se logra capturar la dinámica de pérdida esperada y de pago mínimo de agente, actuando como un conjunto de restricciones degeneradas. Finalmente se logra generar una metodología de asignación de productos en base a rentabilidad de clientes, extendible al incorporar otras aproximaciones al comportamiento de pago o medidas de riesgo en función objetivo. No obstante, el Banco debe incorporar otro tipo de variables a la estimación de rentabilidad, que reflejen los resultados poblacionales de pérdida e incumplimiento observados, para que pueda ser implementada.

Agradecimientos

Al finalizar este proceso quiero agradecer a mi familia, mis padres Samuel y Anita, que me han enseñado y dado todo lo que han podido (y más) para que yo esté donde estoy, soy lo que soy gracias a ustedes. A mis abuelos, tíos, primos por el apoyo y la preocupación. A mi Tata y Yaya, por darme esos abuelos extra, que a pesar de no tener su sangre siempre seré su nieto.

A mis industriamiguís, Jesu, Ale y Pablo, compañeros de tantas historias. A la Dafne y la Jesu (te ganaste doble mención por hinchar) por ser como las hermanas que quiero, odio y aguanto, pero apañan en todas. A mis amigos de la sección 6 y plan común, Rodrigo, Seba, Juampi, Super, Otu, Ricardo, por tantas historias en esos primeros pasos en la Universidad que nos marcaron. Al negro, Emilio, Vale, Coni, Lore, Mati, Nachita, Nico, Pipes Cortines y Lorca, Maldonado (casi se me olvida), Cami, Rita y Cristian, Caro y tantos otros por ser parte de esa familia elegida que son los amigos. Al CEIN 2013, la Feria Empresarial, la semana de inducción, Tutoría.

A mis profesores, José Miguel, Richard y en particular a Giorgio por acompañarme y enseñarme en este proceso, fue largo y sufrido, pero sin duda aprendí mucho, no solo de mi tema de tesis y de ustedes, si no de mí y de lo que puedo lograr. A los profesores y funcionarios que pude conocer en mi paso por la Universidad a Karen, Gaby y Mónica de secretaría docente, a Linda y Fernanda de MGO por estar siempre dispuestas para ayudar en todo.

Me gustaría agradecer a tanta gente que probablemente se me quedaron varios nombres en el tintero y si los empiezo a nombrar a todos, me quedaré sin páginas. Los años que pasé en la Universidad fueron sin duda una etapa maravillosa, donde aprendí, crecí y conocí a muchísima gente que me marcó. Para no perder la costumbre estoy escribiendo esto apurado, pero ya buscaré la forma de agradecerles a cada uno.

Gracias totales.

Andrés Garrido Castro

Tabla de contenido

Agradecimientos	II
1. Introducción	1
1.1. Antecedentes y Motivación	1
1.1.1. Caracterización y efecto de la microempresa en la economía	1
1.1.2. Descripción del Problema	5
1.2. Sobre esta Tesis	7
2. Marco Teórico	10
2.1. Knowledge Discovery in Databases y Data Mining	11
2.1.1. Knowledge Discovery in Databases, (KDD)	11
2.1.2. Data Mining	12
2.2. Modelación basada en agentes	14
2.3. Optimización MIP estocástica	15
2.4. Metodología de resolución: Sample Average Approximation	16
2.4.1. Validación de una solución	18
2.5. Modelo general de acumulación de capital	20
2.5.1. Descripción del modelo	20
2.5.2. Conjuntos	21
2.5.3. Variables	22
2.5.4. Variable aleatoria	22
2.5.5. Parámetros	23
2.5.6. Función objetivo	23
2.5.7. Restricciones	23
2.5.8. Modificaciones al modelo para resolución	26
2.6. Analisis de la función objetivo	29
2.7. Supuestos importantes	30

3. Tratamiento de Datos	31
3.1. Caracterización de los datos y estimación de parámetros	31
3.1.1. Estimación de parámetros	31
3.1.2. Política actual del Banco	42
4. Experimentos Computacionales y Resultados	44
4.1. Consideraciones iniciales	44
4.2. Experimentos en distribución	44
4.3. Análisis de resultados	47
4.3.1. Estrategias de productos	53
4.3.2. Dinámicas capturadas	55
4.3.3. Tiempos de ejecución y convergencia estocástica	56
5. Extensiones Futuras	67
5.1. Modelo generalizado	67
5.1.1. Función objetivo	67
5.2. Mayor refinamiento en modelación de flujos de agentes	68
5.3. Variables exógenas en la generación de escenarios	68
5.4. Algoritmo de descomposición para la resolución del problema	69
5.5. Modelo de optimización dinámica estocástica para la resolución del problema	69
5.6. Modelo de valorización en dos etapas con opciones reales	69
5.6.1. Descripción de variables	70
5.6.2. Parámetros	70
5.6.3. Parámetros de estado	71
6. Conclusiones y recomendaciones	74
6.1. Conclusiones	74
6.2. Recomendaciones	76
Bibliografía	77
Apéndices	80
A . Primeras iteraciones de la formulación del modelo con creencia de tasa del cliente . .	80
A .1. Diagrama de negociación, Modelo con aceptación del producto por parte del cliente	80
A .2. Formulación del modelo	81
B . Tabla de grupos generados por productos y plazos	84

C . Gráfico de grupos generados por productos y plazos	85
D . Variables seleccionadas de la base de datos para la descripción de productos	85

Índice de tablas

1.1. Caracterización de empresas segun Ventas y Empleo promedio	2
1.2. Distribución de Microempresas por Región en Chile año 2016.	3
1.3. Tabla de caracterización subsegmentos.	6
1.4. Caracterización subsegmentos.	6
3.1. Productos considerados para el modelo	37
3.2. Resultados Ajuste distribución	42
4.1. Resumen de parámetros por experimento	45
4.2. Cantidad de productos entregados por periodo y por tamaño in-sample	64
6.1. Agrupación de 143 tipos de productos utilizada	84
6.2. Variables seleccionadas para la descripción de productos	86

Índice de figuras

1.1. Distribución respuestas pregunta EME 2, 2013	4
2.1. Diagrama rolling & forward del modelo periodo a periodo	15
2.2. Diagrama de la estructura de un Balance	21
3.1. Diagrama proceso de agrupación en productos característicos	32
3.2. Histograma con Montos y varlores representativos considerados	34
3.3. Histograma con plazos en meses y meses respresentativos considerados	34
3.4. Punto de corte que explica el 70 % y 80 % de los datos	35
3.5. Distribución de tasas	36
3.6. Distribución multinormal de retornos del negocio descompuesta con un Modelo de mixtura	41
3.7. Distribución de retornos por productos	43
4.1. Distribucion \log_{10} funcion objetivo exp1 (izq.) y exp2 (der.) por tamaño in-sample .	48
4.2. Distribucion \log_{10} funcion objetivo exp3 (izq.) y exp4 (der.) por tamaño in-sample .	48
4.3. Distribucion \log_{10} funcion objetivo exp5 (izq.) y exp6 (der.) por tamaño in-sample .	49
4.4. BoxPlots distribuciones activo (izq) y mora (der) en último periodo experimento 4 .	50
4.5. BoxPlots distribución de activo por periodo para repetición n°10 de experimento 4 .	50
4.6. BoxPlots distribuciones activo (izq.) y mora (der.) en último periodo experimento 5	51
4.7. BoxPlots distribución de activo por periodo para repetición n°10 de experimento 5 .	52
4.8. Distribuciones activo en repetición 10 (izq) y 8 (der) por periodo	52
4.9. BoxPlots distribuciones activo (izq) y mora (der) en último periodo experimento 6 .	53
4.10. BoxPlots distribuciones activo (izq.) y mora (der.) en último periodo experimento 4*	55
4.11. BoxPlots distribuciones deuda contable (izq) y mora (der) por periodo para repetición 10 en exp 6	56
4.12. Promedio geométrico de tiempos de resolución por tamaño in-sample	57
4.13. Error relativo (izq.) y tiempo promedio de resolución (der.) por número de repeticiones	59

4.14. Error relativo de las diferentes configuraciones versus tamaño out-sample para los experimentos 4,5 y 6	60
4.15. Evolución tiempo promedio de cálculo de cota inferior para experimentos 4,5 y 6, versus tamaño out-sample	61
4.16. Histograma de crédito $j=1$ entregados por periodo	62
4.17. LB (izq.) y error relativo (der.) por periodo del candidato a solución para el crédito 1	63
4.18. LB (izq.) y GAP relativo (der.) por periodo del candidato a solución para experimento 6	64
4.19. Gap relativo experimento y por tamaño in-sample	65
5.1. Representación del árbol de simulación	72
5.2. Representación del árbol de simulación con valores	73
6.1. Flujo primera iteración modelo con aceptación de producto	80
6.2. Distribución de créditos vendidos por monto versus plazo	85

Capítulo 1

Introducción

1.1. Antecedentes y Motivación

1.1.1. Caracterización y efecto de la microempresa en la economía

Muchos países de la *Organisation for Economic Cooperation and Development* (OECD), poseen programas de apoyo y desarrollo de pequeñas y medianas empresas. Por ejemplo Alemania, Islandia, Japón y Nueva Zelanda dedican más del 50% de sus programas de apoyo y desarrollo a éstas [OECD, 2004]. Esto se debe, entre otras razones, a que en la mayoría de estos países estas empresas representan aproximadamente el 95% del total, entre el 60% y 70% del empleo y un 33% del producto interno bruto (PIB) en economías emergentes [OECD, 2017]. En países de la Unión Europea se puede observar que existen 25 millones de pequeñas empresas, que representan el 99% del total de empresas y emplean 95 millones de personas, quienes conforman el 55% del total de empleos en el sector privado [Keskin et al., 2010]. Por otro lado, se estima que contribuyen entre un 20% y un 35% en promedio a las exportaciones manufactureras directas en el mundo. En síntesis, generan mucho empleo, pero poco PIB, exportaciones y montos, no obstante le dan inercia y resistencia a la economía.

El concepto de micro, pequeña y mediana empresa posee varias definiciones, dependiendo del país; ventas anuales, empleo generado, entre otros. Por ejemplo, en países de la Unión Europea se entiende como mediana aquellas que poseen menos de 250 empleados, pequeña empresa menos de 50 y micro menos de 10 empleados [OECD, 2004] y [Keskin et al., 2010].

En Chile la Ley 20.416 y decreto 76 de 2010 de Ministerio de economía, establecen definiciones de tamaños de empresa basada en ventas anuales para efectos tributarios y para temas laborales según cantidad de trabajadores. Esta caracterización se puede observar en la tabla 1.1.

Tabla 1.1: Caracterización de empresas segun Ventas y Empleo promedio

Definición general de tamaño	Ventas anuales (UF)	Empleo promedio
Microempresa	De 0,1 a 2.400	De 1 a 9
Pequeña Empresa	De 2.401 a 25.000	De 10 a 49
Mediana Empresa	De 25.001 a 100.000	De 50 a 199
Gran Empresa	Superior a 100.001	De 200 y más

Según la cuarta encuesta de microemprendimiento en Chile del Ministerio de economía, en 2015 existían 1.865.860 emprendedores, ya sean empleadores o trabajadores por cuenta propia (TCP), de los cuales 1.814.938 son microemprendedores, es decir, aproximadamente 1 por cada 10 habitantes [Ministerio de economía, 2016]. Las microempresas se caracterizan por su gran heterogeneidad productiva, ya que abarcan una gran variedad de rubros, siendo mayoritaria su presencia en el sector terciario en comercios, restaurantes y hotelería y en menor medida en enseñanza, minería, electricidad gas y agua según datos de [SERCOTEC, 2013].

La importancia del sector microempresario radica en la generación de empleo, mayor igualdad en la distribución de ingresos, la ocupación del territorio, incremento de la competencia en los mercados, aporte en las exportaciones, el crecimiento económico. Además, debido a su gran diversificación, muchas microempresas son anticíclicas en periodos de crisis. No obstante también sufren desventajas respecto a empresas grandes en cuanto a acceso de mercados, financiamiento, tecnologías y otras dimensiones importantes para su funcionamiento [Solimano et al., 2007].

En cuanto a la generación de empleo, similar a la evidencia internacional, este sector tiene una importante presencia en el país. Al año 2011, según la encuesta CASEN, la microempresa abarca el 36,1 % de población ocupada. Si a esto se le suma la mediana empresa, ambas alcanzan casi la mitad con más de 2.8 millones de personas, lo que incluye trabajadores por cuenta propia, empleadores y empleados.

Con respecto a la ubicación territorial, las micro y pequeñas empresas (MIPE), están distribuidas de forma más homogénea a lo largo del territorio nacional que las medianas y grandes. Estas están mucho más concentradas en la Región Metropolitana, por lo que tiene gran relevancia en la presencia regional, y puede ser importante en la cobertura de políticas públicas que busquen llegar a todos los puntos del país.

Se destaca la presencia de las microempresas en las regiones extremas del país, como lo son la

región de Aysen, Magallanes y Arica donde superan el 45 % del total de las empresas de la región, siendo que el promedio país es de un 39,7 % [SERCOTEC, 2013].

Tabla 1.2: Distribución de Microempresas por Región en Chile año 2016.

Región	Número de MIPE	% del Total	% Regional MIPE	Número de Mediana y grande	% del Total	% Regional Mediana y grande	Total General
RM	328853	38.8 %	93.2 %	23919	2.8 %	6.8 %	352772
I	14972	1.8 %	93.2 %	1091	0.1 %	6.8 %	16063
II	24072	2.8 %	95.7 %	1079	0.1 %	4.3 %	25151
III	12805	1.5 %	96.7 %	439	0.1 %	3.3 %	13244
IV	31657	3.7 %	97.3 %	890	0.1 %	2.7 %	32547
V	81460	9.6 %	96.9 %	2617	0.3 %	3.1 %	84077
VI	43930	5.2 %	97.4 %	1172	0.1 %	2.6 %	45102
VII	55721	6.6 %	97.7 %	1313	0.2 %	2.3 %	57034
VIII	86158	10.2 %	97.1 %	2597	0.3 %	2.9 %	88755
IX	39898	4.7 %	97.6 %	979	0.1 %	2.4 %	40877
X	42197	5.0 %	97.0 %	1311	0.2 %	3.0 %	43508
XI	6364	0.8 %	98.1 %	124	0.01 %	1.9 %	6488
XII	9462	1.1 %	96.8 %	317	0.04 %	3.2 %	9779
XIV	17709	2.1 %	97.8 %	403	0.05 %	2.2 %	18112
XV	12554	1.5 %	98.5 %	190	0.02 %	1.5 %	12744
Sin información	536	0.1 %	96.4 %	20	0.0 %	3.6 %	556
Total	808348	95.5 %	95.5 %	38461	4.5 %	4.5 %	846809

Fuente: Servicio de Impuestos Internos (SII)

Si bien existe una alta concentración de las ventas en las medianas y grandes empresas, las MIPes formales alcanzan un 8,37 % del monto total de ventas para el año 2015, según datos del servicio de impuestos internos (SII). Este sector desde el año 2005 al 2015 ha tenido un crecimiento en sus ventas de un 55,8 %, con una tasa promedio de crecimiento de 4,56 % anual. Esto reafirma que si bien son pequeñas, tienen opciones de crecimiento y son un aporte al desarrollo del país, con tasas de rentabilidad muy superiores al resto de las industrias.

Las MIPes, son un sector que genera muchas implicancias positivas en el desarrollo económico del país, como beneficios sociales, donde se han enfocado esfuerzos en hacerlo crecer y fomentar su inversión y desarrollo. Por un lado acercándolo a servicios financieros tradicionales, que si bien han ido en aumento, aún poseen poca penetración. El 70,3 % de los microempresarios declara no haber solicitado jamás un préstamo bancario y entre las razones declaradas para esto: un 32,2 % dice no necesitarlo, un 20 % que “no le gusta pedir préstamos”, un 21 % piensa que no cumpliría con los requisitos, un 18 % declara no poder cumplir frente a los pagos y el resto, que representa un 8 %, da razones de desconfianza o desconocimiento en el procedimiento para solicitarlo, entre otros [Ministerio de economía, 2013]. Estos números muestran que aún existe una brecha importante relacionada a la desconfianza y desconocimiento del sistema financiero y de los posibles beneficios que esto podría generar con mayor educación financiera a este sector.

Dentro de los problemas para las MIPes se tiene el acceso a financiamiento, un 30,1 % de los microempresarios piensa que sus principales limitaciones para crecer es la falta de financia-

miento [Ministerio de economía, 2012]. Según un estudio de la OECD, algunas diferencias entre el financiamiento de pequeñas y grandes empresas son las mayores tasas y exigencias en los contratos. Esto, debido a que instituciones financieras asocian un mayor riesgo inherente a estos segmentos. Este riesgo, en parte, se puede explicar a la varianza en rentabilidad, sobrevivencia y crecimiento de este sector, al existir asimetrías de información entre las instituciones financieras y las microempresas. Estas asimetrías hacen que las evaluaciones sean peores para su *rating* crediticio, limitando su acceso a créditos, ya que las instituciones financieras no saben si efectivamente se le está dando un crédito a un buen o mal pagador (asimetría de información y riesgo crediticio) o a alguien que utilizará ese financiamiento para mejorar su negocio o para otra cosa (riesgo moral).

Se puede diagnosticar a partir de los datos de la segunda encuesta de microemprendimiento [Ministerio de economía, 2012], una deficiente educación financiera, sumado al desconocimiento y desconfianza del sistema financiero, al momento de solicitar financiamiento en la banca tradicional, lo que dificulta aún más conocer las verdaderas necesidades de los microempresarios y las reales posibilidades de pago. En esta encuesta se hace el ejercicio de preguntar “¿Si un banco paga una tasa de 1 % al mes, la tasa de interés anual es?”, teniendo como resultados la distribución de respuestas de la figura 1.1, donde se observa que sólo al rededor de un 5 % responde correctamente la pregunta.

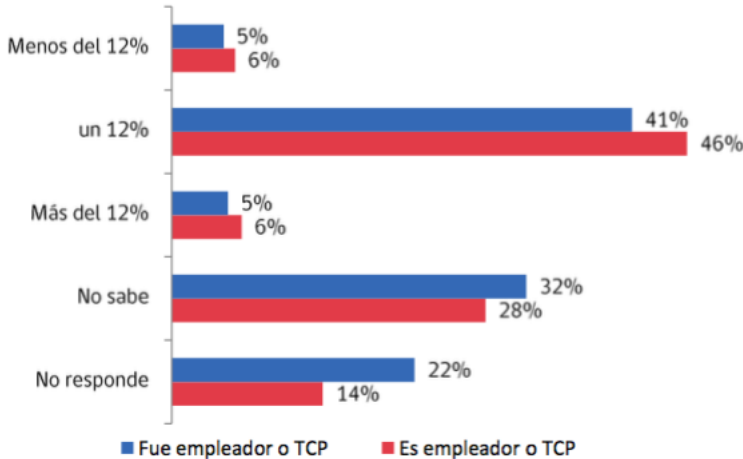


Figura 1.1: Distribución respuestas pregunta EME 2, 2013

Esta deficiente educación financiera sumada a la necesidad de financiamiento genera un problema al no entender qué producto necesita el microempresario y para el Banco se refleja en asignaciones que no son óptimas, debido al costo de conocer exactamente qué necesita cada cliente. Es por esto que se hace necesario encontrar mecanismos de asignación óptimos de productos, que mejoren la

relación entre lo que necesita el microempresario y lo que oferta el Banco.

Un área de las finanzas especializada en el sector de las micro y pequeñas empresas es la microfinanza, que surge a partir de un movimiento global dedicado a aumentar el acceso a créditos de menor escala (microcréditos), mecanismos de ahorro, inversión, seguros y servicios financieros en general, a comunidades o personas con bajos niveles de ingreso, que generalmente tienen un acceso limitado a la banca tradicional. Esta área de las finanzas presenta entonces un campo de estudio para el desarrollo de posibilidades de extensión de mercados, reducción de pobreza y generación de beneficios sociales, y también plantea preguntas a responder. En el libro “*The economics of micro-finance*” de los autores Armendáriz y Morduch del año 2010, se discuten algunas hipótesis sobre por qué se hacen tan necesarias las microfinanzas, y se plantean los problemas de selección adversa y riesgo moral al momento de financiar pequeñas y medianas empresas, donde las asimetrías de información y el caracterizar mejor a un individuo terminan siendo desafíos fundamentales para el sistema financiero [Armendáriz and Morduch, 2010], .

Cuando se habla de MIPes, se suele asociar a sectores de escasos recursos y de alta vulnerabilidad social. Para el año 2011 el 49,4% de los ocupados en MIPes registraba ingresos menores a dos salarios mínimos (364.000 pesos) y el 17,3% reconocía ingresos bajo un salario mínimo [SER-COTEC, 2013]. Es por esto la importancia y el interés que existe en el crecimiento y desarrollo de sus negocios, para mejorar la situación socioeconómica de los microempresarios y ser un motor de movilidad social.

Como es sabido las empresas no quiebran por falta de rentabilidad, si no que por falta de liquidez [Merton, 1974], en especial en el sector microempresario con menores rentas, como lo mencionan [Alfaro-Castillo and Castillo-Vergara, 2013]. Por esto es también de especial importancia el análisis de las variables contables de las microempresas.

1.1.2. Descripción del Problema

BancoEstado Microempresas (BEME) surge en el año 1995 como una filial del BancoEstado, con el objetivo de “generar igualdad de oportunidades en el acceso a servicios financieros”. Propiciando la participación en mercados financieros imperfectos con alto impacto social, como es la industria de las microfinanzas. [Larraín et al., 2007]

Este impacto social se reflejó al dar acceso de forma pionera al sector microempresarial a los servicios financieros, en particular a los créditos, segmentos que habían estado desatendidos y con nulo acceso a la banca tradicional. BEME ha tenido un impacto significativo y un crecimiento que se observa en que, al día de hoy, cuenta con más de 500.000 clientes de distintos rubros y distintas necesidades.

Como se ha indicado en la sección anterior, el sector microempresario en Chile abarca la mayor parte del territorio nacional y una gran variedad de rubros, lo que provoca que su cartera de clientes sea bastante compleja de manejar.

BEME segmenta a sus clientes en urbanos y rurales, los que a su vez se subsegmentan en otras categorías. Este trabajo se enfocará en el segmento urbano, cuya subsegmentación y composición se puede observar en la tabla 1.3 y, en particular, en el subsegmento comercio. El sub segmento comercio, se puede observar en la tabla 1.4, con un breve resumen de la caracterización de este.

Tabla 1.3: Tabla de caracterización subsegmentos.

Segmento	Subsegmentos	Porcentaje	Ejemplos
Comercio	Negocios de barrio	40 %	Bar, distribuidora, ferretería, fuente de soda, librería, minimarket
	Comercio detallista	39 %	Almacén, bazar, amasandería, botillera, verdulería
	Informal y semiformal (ferias libres)	21 %	Ambulantes, persas, semanero, por catálogo, suplementero
Transporte	Taxis colectivos	53 %	-
	Taxis básicos	25 %	-
	Carga	12 %	-
	Escolares	10 %	-
Oficios, manufactura y servicios	Oficios	26 %	Electricista, gasfiter, jardinero, contratistas, reparaciones
	Manufactura	13 %	Artesanos, fabricación de cuero y calzado, artículos varios, imprenta
	Servicios	23 %	Diseñadores, belleza, profesionales y técnicos, aseo, gastronómicos, educación
	Otros	38 %	-

Fuente: BancoEstado Microempresas

Tabla 1.4: Caracterización subsegmentos.

Segmento	Edad promedio	Formalidad	Penetración de mercado	Monto promedio
Comercio	53 años	79 %	32 %	\$3,9 MM
Transporte	52 años	S.I.	61 %	\$5,7MM
Oficios, manufactura y servicios	52 años	S.I.	29 %	\$4,1MM

Fuente: BancoEstado Microempresas

El rol de BEME es asegurarse que selecciona adecuadamente a los clientes, porque logra con sus productos financieros sacar adelante a los microempresarios, desde la perspectiva de que tenga mayores probabilidades de ser exitoso en su negocio. Por ejemplo:

Consideremos el problema que enfrenta un microempresario dueño de una distribuidora de pro-

ductos, que busca financiar la compra de un vehículo propio para el despacho de estos, y así disminuir los gastos que actualmente significan el arriendo de una camioneta para este fin. Para esto acude al banco buscando financiamiento. El banco, frente a la solicitud, posee un conjunto grande de productos, entendiendo estos últimos como créditos con distintos montos, plazos y tasas. ¿Cómo poder identificar cuál es el producto que realmente se ajusta a su perfil?, para un mismo monto ¿Qué plazo le conviene?, ¿Alguno de los productos sobredeudará al cliente? Actualmente, estas preguntas se resuelven en base a la capacidad de pago estimada en una evaluación de clientes en terreno, observando principalmente la cuota que deberá pagar el cliente por el producto.

Para responder alguna de estas preguntas BEME requiere el análisis del segmento comercio, de forma de identificar y poder estimar parámetros relevantes, distribuciones de rentabilidad de los negocios, conjunto de productos ofrecidos, entre otros. Se necesitan mejores mecanismos de asignación de productos, por lo que en este trabajo se propone, como una posible solución a este problema, la formulación de un modelo de programación matemática estocástico, que permita maximizar el éxito de los microempresarios, definiendo políticas de colocación que mejor se adapten a las necesidades de los clientes, considerando las restricciones que debe cumplir el banco, de rentabilidad y de liquidez.

1.2. Sobre esta Tesis

Este trabajo de tesis, surge como un desafío de poder entender mejor a los clientes del BancoEstado Microempresas y tiene como objetivo principal, diseñar un mecanismo óptimo de asignación de productos, que se implementa a través de un modelo de optimización para un segmento de clientes de BEME con decisiones óptimas en valor esperado del “éxito” de los microempresarios en un horizonte de tiempo. Donde actualmente se entiende como “éxito” o bienestar del cliente, su acumulación de capital, lo que se considerará como la primera aproximación para el planteamiento del modelo.

Para poder cumplir este objetivo, se deben cumplir hitos intermedios como: caracterizar clientes del segmento comercio, a partir de datos de evaluaciones de clientes; en base a rentabilidad del negocio, utilizando variables financieras contables. Además, realizar una caracterización de los productos de crédito que reflejen la oferta del Banco a sus clientes de forma discreta.

Adicionalmente, como aporte metodológico, se necesita formular un modelo de optimización que

replique la interacción de los agentes Banco y Cliente en el proceso de asignación de un producto, y en el desarrollo del negocio de un microempresario, incorporando las dinámicas de pago y restricciones de pérdidas esperadas del Banco. Reconociendo la aleatoriedad del problema, enfrentándola como un problema de optimización *Mixed-Integer Programming* (MIP) estocástico e incorporándola en la rentabilidad de los negocios del cliente.

Finalmente, se requiere implementar diseños experimentales, que busquen resolver el problema en base a distintas distribuciones de rentabilidad para grupos de clientes y así establecer un mecanismo de asignación de productos.

Se espera con el análisis de datos, poder entender cómo actualmente se realiza la asignación de productos y poder encontrar características que permitan diferenciar y mejorar esta última gracias a nociones de optimalidad entregadas por el modelo de optimización estocástico planteado.

Al encontrar estas características diferenciadoras de los clientes, se espera poder analizar si es conveniente o no realizar esta sub segmentación, en particular bajo criterios de rentabilidad de negocios, lo cual, en una primera instancia, se prevee que sea positivo, debido a que se podría entregar mejores productos en base a las necesidades y características de los distintos negocios.

Finalmente plantear un modelo que permita encontrar políticas óptimas para el banco, mediante la maximización del éxito de los clientes sujeto a restricciones de rentabilidad o riesgo, permitiendo al banco generar un mecanismo de asignación basado en riesgo-rentabilidad, con parámetros escogidos óptimamente para sus políticas de asignación de productos a sus distintos tipos de clientes y, adecuarlas mejor a sus necesidades.

El trabajo se divide en 6 capítulos. El primer capítulo busca introducir el tema y el problema a tratar, explicando el contexto de la microempresa, la descripción del problema, los objetivos y resultados esperados de este trabajo.

El segundo capítulo consiste en el marco teórico para el desarrollo del trabajo de tesis, en los temas relacionados al tratamiento de datos, optimización estocástica y los métodos de resolución del problema, concluyendo con el modelo de optimización planteado.

El tercer capítulo explica el trabajo realizado con los datos para la estimación de parámetros relevantes para su incorporación al modelo. Considerando esto, la limpieza y selección de atributos, metodologías, supuestos utilizados y finalmente los resultados de las estimaciones.

El cuarto capítulo explica los experimentos computacionales, su diseño experimental, medidas de tiempo de resolución, convergencia para garantizar la significancia estadística y finalmente el análisis de resultados, considerando candidatos a solución, distribuciones de la función objetivo y variables relevantes. Finalmente, los mecanismos de asignación propuestos y las dinámicas capturadas por el modelo.

El capítulo 5 explica las extensiones futuras que se pueden desarrollar a partir de este trabajo.

Finalmente el capítulo 6, entrega las conclusiones del trabajo realizado junto a recomendaciones para el Banco.

Capítulo 2

Marco Teórico

La literatura relacionada a modelos de riesgo crediticio, en particular, a modelos estructurales, se remontan a lo planteado por Merton [Merton, 1974] y su modelo de riesgo de crédito, que en pocas palabras, plantea que, considerando el balance de una compañía, el valor total del activo de una firma A_t es un proceso estocástico, en particular, sigue un *Movimiento Browniano Geométrico* de la siguiente forma:

$$dA_t = \mu A_t dt + \sigma A_t dW_t, A_0 > 0 \quad (2.1)$$

Donde μ es la media del retorno del activo, σ su volatilidad y considerando la deuda D , se puede interpretar el patrimonio (*equity*) de la firma como una opción sobre el activo de una firma bajo la condición de identidad contable $A_t = E_t + D_t$. Donde E_t corresponde al *equity*.

El evento de incumplimiento o *default*, ocurre cuando el valor del activo en algún t sea inferior al de la deuda ($t : A_t < D_t$), lo que comúnmente se le denomina tiempo de parada para un proceso estocástico, es decir, cuando un movimiento browniano geométrico baja de cierto umbral y puede evaluarse a través de la teoría de opciones financieras.

Este trabajo, se basa en las variables de estado del balance contable, al igual que el modelo de Merton, sin embargo, el aporte metodológico radica en incorporar decisiones óptimas sobre la colocación y la deuda a partir de la información de las variables financieras contables de las microempresas. Modelando con dichas variables las decisiones de agentes, en un modelo de optimización debiendo para ello modelar e incorporar restricciones que establecen la lógica de un *default*.

Como se presentó en los objetivos de este trabajo, en una primera parte se busca hacer un análisis exploratorio de los datos obtenidos de parte del BancoEstado Microempresas para la caracterización

de clientes del segmento comercio, siendo por lo tanto, necesario explicar algunos conceptos para este análisis.

2.1. Knowledge Discovery in Databases y Data Mining

2.1.1. Knowledge Discovery in Databases, (KDD)

La metodología de “*Knowledge Discovery in Databases*” (KDD) se ha utilizado mucho en el último tiempo para el manejo de bases de datos y se define como el proceso no trivial de identificación de patrones en datos que sean válidos, novedosos, potencialmente útiles y por último comprensibles [Fayyad, 1996]. La predicción y/o entendimiento de patrones a partir de los datos, resulta fundamental en la toma de decisiones de distintas organizaciones para el entendimiento de negocios o gestión de organizaciones.

El proceso KDD consiste en varios procesos, que si bien se presentan de forma secuencial, se realizan ciclos entre éstos de forma de ir refinando los resultados de una etapa previa en base a resultados de etapas posteriores.

- **Selección de datos:** Consiste en establecer cuáles serán las fuentes de información y la selección del subconjunto de datos necesarios para la aplicación de los siguientes pasos, estudiarlos para entender el significado de las variables o atributos.
- **Preprocesamiento de datos:** Consiste en detectar errores en los datos seleccionados, ya sea por ejemplo: datos faltantes, repetidos o fuera de rango. Existen varias estrategias para la limpieza de datos que se pueden observar en la literatura y dependerá mucho del análisis y del conocimiento que se tenga de las fuentes de información de donde se obtienen estos datos.
- **Transformación de los datos:** Se refiere al proceso de decidir cuáles son las variables relevantes y representativas de los datos, en base al objetivo del análisis, se busca también reducir la dimensionalidad de la base de datos de estudio, para que sea más fácil trabajar con ella y encontrar soluciones. A su vez, dado un modelo de minería de datos, se debe transformar las variables de forma que se ajusten al formato de entrada de los modelos y algoritmos de la siguiente etapa. Generalmente entre esta etapa y la siguiente se realizan ciclos para poder ajustar de mejor forma los parámetros de algoritmos o modelos.
- **Data Mining:** Corresponde a la etapa principal en donde se aplican métodos para la búsqueda de patrones, y cuya parte importante es determinar a partir del objetivo del proceso, si se

busca clasificar, regresionar, segmentar, entre otros. Definir los algoritmos de resolución e interpretación de parámetros para poder tomar decisiones.

- **Implementación y evaluación de resultados:** A partir de los resultados de los pasos anteriores se busca interpretar los resultados y se determinan cómo podrán ser utilizados e implementados en la toma de decisiones a partir de los resultados obtenidos.

Para este trabajo se implementó una lógica similar a esta metodología para la estimación de parámetros que se detalla en el capítulo 3 del tratamiento de datos.

2.1.2. Data Mining

Existen varias definiciones de minería de datos en la literatura, pero se entenderá para este trabajo como una serie de técnicas posibles de aplicar sobre un conjunto de datos que, permiten extraer conocimiento de un determinado sistema en estudio.

El uso de estas técnicas, dependiendo de su objetivo puede ser:

- No supervisados: Se utilizan cuando no se tiene una variable a explicar (\hat{y}), por lo que se busca encontrar posibles relaciones entre los datos y describir dichas relaciones, un ejemplo de estos métodos son los algoritmos de segmentación o clustering.
- Supervisados: Se utilizan cuando se tiene una variable a explicar a partir del resto de los datos ($\hat{y} = F(\vec{x})$). Son utilizados extraer patrones de comportamiento que pueden ser aplicados a otros conjuntos, generalmente se asocian a métodos predictivos o de clasificación.

Para este trabajo, se busca realizar un análisis exploratorio de los datos, para buscar relaciones en los datos que se tienen por parte del banco, como se detalla en el capítulo 3, existen dos objetivos claros e indispensables para el modelo, que corresponden a realizar una segmentación de los productos que actualmente tiene el banco, por lo que se detallará la metodología de segmentación para posteriormente estimar una distribución de rentabilidad, usando un modelo finito de mixtura, del cual también se detallará la teoría relacionada.

Segmentación

La Segmentación o *clustering*, tiene referencia a la agrupación de un conjunto de objetos en subconjuntos que sean más similares entre sí que a los otros subconjuntos. Corresponde a uno de los principales campos de *data mining*, como una técnica no supervisada de análisis estadístico de datos, utilizado en muchas áreas.

Permite encontrar relaciones entre datos que no son claras en un principio y disminuir la dimensionalidad de los datos a trabajar. Existen diversos tipos de algoritmos, dependiendo de lo que se quiera considerar como *cluster*, y de las formas más eficientes de encontrarlos. Un ejemplo bastante conocido es el algoritmo *K-medias*, que agrupa los objetos en torno a K núcleos o centroides e iterativamente va reacomodandolos como “centros de masa” hasta que se llegan a grupos estables.

En este trabajo se utilizará una metodología para segmentar los productos similar a *K-medias*, al mantener una cantidad de centroides fijos, se agruparán en torno a una grilla uniforme y se seleccionarán los grupos de forma que representen un conjunto discreto de opciones que sea representativo a la oferta actual del Banco.

Modelos Finitos de mixtura

Los modelos finitos de mixtura permiten una representación natural de heterogeneidad en una distribución de probabilidad, en un número finito de clases latentes, es decir, encontrar relaciones entre los datos que permita segmentar la información en grupos heterogéneos. Para este trabajo se utilizará esta metodología para estimar la distribución de rentabilidad como una combinación de dos o más tipos de clientes.

Mucha de la teoría sobre estos modelos, involucra el supuesto que existen subgrupos que están distribuidos de acuerdo a una forma paramétrica, la cual, generalmente se asume son distribuciones normales univariada o multivariada. Como se verá en el capítulo siguiente del tratamiento de datos, la distribución de rentabilidad se puede aproximar como la mezcla de dos distribuciones normales de grupos distintos.

De forma más concreta, se asume un vector de realizaciones de variables aleatorias X_1, \dots, X_n , de una mezcla de $m > 1$ distribuciones arbitrarias, la distribución de cada X_i puede escribirse como

:

$$g_{\theta}(x_i) = \sum_{j=1}^m \lambda_j \phi_j(x_i), \quad x_i \in \mathbf{R}^r \quad (2.2)$$

Donde $\theta = (\lambda, \phi) = (\lambda_1, \dots, \lambda_m, \phi_1, \dots, \phi_m)$ correspondiente a la combinación de los parámetros de proporción o “peso” λ_i , cuya $\sum_{j=1}^m \lambda_j = 1$ y las densidades ϕ_i que tiene cada una de las distribuciones $i \in 1, \dots, m$ que son parte de la mezcla.

A partir de esto, para realizar la estimación de los parámetros de estos modelos, se utilizan algoritmos *Expectation-maximization* (EM), explicados por [Dempster et al., 1977], maximizando la verosimilitud. En pocas palabras consisten de dos tipos de pasos: “*Expectation*” (E), que plantea la función esperada de *log-verosimilitud* evaluada en el punto actual de la estimación de parámetros y, el paso de “*Maximization*” (M), que computa nuevamente los parámetros maximizando la función encontrada en el paso E, los cuales se actualizarán en el siguiente paso E.

Esta metodología y algoritmos se introduce en funciones del lenguaje estadístico R, en el paquete “*mixtools*” los cuales se detallan en [Benaglia et al., 2009], y serán utilizados en el siguiente capítulo de tratamiento de datos.

2.2. Modelación basada en agentes

La modelación basada en agentes o *agent-based models*, es un tipo de modelamiento computacional para simular acciones e interacciones de agentes, quienes al observar su estado toman decisiones basadas en un conjunto de reglas evaluando un sistema global, es decir, consiste de un sistema de agentes y de las relaciones entre ellos. Esta metodología provee información valiosa sobre patrones de comportamiento de los agentes y las dinámicas que se pueden observar en el mundo real del sistema que se modela [Bonabeau, 2002].

Para este trabajo, se considera el estudio de las dinámicas e interacción de dos agentes: el Banco y un cliente (microempresario), en particular la entrega de productos y las dinámicas de pago involucradas. Se espera además incorporar la dinámica del *default*. Estas dinámicas se observan periodo a periodo y todas las decisiones involucradas se reflejan en las variables de estado intertemporales del microempresario.

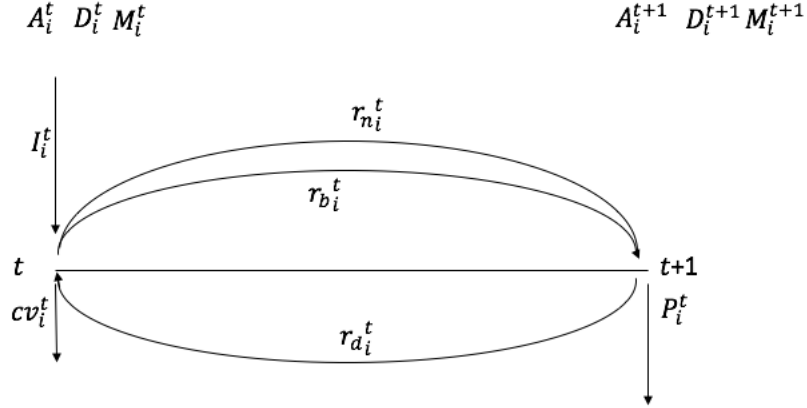


Figura 2.1: Diagrama rolling & forward del modelo periodo a periodo

La figura 2.1 describe lo que ocurre en cada momento en el sistema considerado. Sea un periodo t cualquiera, las variables de estado activo (A^t), deuda (D^t) y mora (M^t) representan la situación en el momento t del sistema. Al comienzo del periodo se reciben los ingresos por producto (I^t), en caso de que corresponda y se descuenta el costo de vida (CV^t). Para pasar al final del periodo, se realiza el negocio del cliente, que tiene una rentabilidad asociada (r_n^t) y se debe realizar el pago de la cuota del producto (P^t), este pago se definirá en base al mínimo entre lo que puede pagar y lo que debe pagar el cliente. Posteriormente se actualizan las variables de estado y se avanza al siguiente periodo. Este diagrama presenta la intuición de lo que se produce periodo a periodo entre los agentes involucrados y como base del modelo que se presenta más adelante.

2.3. Optimización MIP estocástica

Cuando se habla de optimización o programación matemática, se remonta a su desarrollo en la segunda guerra mundial con los trabajos iniciales de [Kantorovich, 1939] y [Hitchcock, 1941] en la introducción de problemas lineales y redes de transporte. Posteriormente con la propuesta del método *Simplex* para la resolución de problemas lineales realizada por [Dantzig, 1948]. Finalmente pudo demostrar, a través del método de punto interior, que los problemas lineales pueden ser resueltos en tiempo polinomial. Resultado que es muy potente al permitir explicar que a pesar de que un problema tenga un número exponencial de restricciones, si este puede ser descompuesto, es posible resolver en tiempo polinomial [Khachiyan, 1980].

Existen muchos subcampos en el ámbito de la optimización además de la programación lineal

(LP), tales como: programación convexa; no lineal; combinatorial; infinitesimal, entre otros. En particular, existe la programación entera, que se caracteriza por que las variables del problema de optimización deben ser valores enteros y como extensión se tienen los denominados *Mixed Integer Programming* (MIP), que hace referencia a problemas que tienen algunas variables enteras, pero las demás corresponden a variables lineales. Las diferencias con los problemas lineales se ven al momento de la resolución de los problemas, ya que se utilizan otro tipo de algoritmos como *Branch and Bound* y variantes de este como *Branch and cut* u otras heurísticas [Schrijver, 1998].

Además de esto, para este trabajo se utilizará el marco de la optimización bajo incertidumbre u optimización estocástica, que hace referencia a modelos de optimización donde existen parámetros que con incertidumbre, es decir, que no son fijos, actúan de forma aleatoria bajo una cierta ley de probabilidad conjunta.

En optimización bajo incertidumbre existen dos enfoques metodológicos y de modelamiento: uno denominado optimización robusta, que busca entregar soluciones, que como su nombre lo dice, son “robustas” frente a la aleatoriedad de sus parámetros, es decir, que la solución pueda hacerse cargo de “peores casos”, o casos extremos, y a su vez tener soluciones factibles para cualquier realización de la variable aleatoria, el segundo enfoque, comúnmente llamado optimización estocástica a secas, donde la incertidumbre se introduce tanto en el poliedro como en la función objetivo mediante probabilidad, esperanza o momentos superiores bajo cierta ley de probabilidad o pueden ser generados.

A partir de esto, en este trabajo se desarrollará un modelo de optimización estocástica mixta (MIP), es decir con algunas de sus variables enteras y otras lineales, con restricciones y función objetivo en esperanza y cuyos parámetros siguen una distribución particular de distribución de probabilidad. En la siguiente sección se presenta el método para resolver dichos problemas.

2.4. Metodología de resolución: Sample Average Approximation

Es necesario entonces, encontrar un método para resolver dichos problemas de optimización bajo incertidumbre. Donde en el último tiempo el enfoque *Sample Average Approximation* (SAA) [Shapiro and Ruszczyński, 2008], ha tomado fuerza al permitir actuar como un estimador del problema original, mediante realizaciones de los parámetros aleatorios. Bajo ciertas condiciones generales, se obtienen garantías de convergencia tanto en el valor de la función objetivo como en la solución

misma.

Esta metodología, se presenta como una herramienta versátil para la solución de problemas de optimización estocástica que reúnen las condiciones de:

- No es viable una resolución mediante un método cerrado.
- Es sencilla la evaluación de la función objetivo, para un escenario fijo de los parámetros aleatorios.
- El conjunto de posibles soluciones es lo suficientemente grande.

En pocas palabras, la técnica busca resolver muchas repeticiones de problemas equivalentes ciertos, a partir de realizaciones o escenarios del conjunto de parámetros incierto. Para luego hacer inferencia estadística a partir de las soluciones de las repeticiones del problema, permitiendo estimar el gap estocástico, es decir, un intervalo de confianza para el candidato a solución, con garantías estadísticas sobre el mismo.

Usualmente se describe esta metodología resolviendo un problema de minimización de la forma:

$$v^* := \min_{x \in \mathcal{P}} \mathbb{E}_\xi [\mathbf{F}(x, \xi)] \quad (2.3)$$

No obstante, para el desarrollo de este trabajo, se requiere plantear un problema que maximice un valor esperado, por lo que, análogamente, se trabajarán las definiciones cambiando el sentido de la optimización para el problema análogo:

$$v^* := \max_{x \in \mathcal{P}} \mathbb{E}_\xi [\mathbf{F}(x, \xi)] \quad (2.4)$$

Se definirá $f(x) := \mathbb{E}_\xi [\mathbf{F}(x, \xi)]$ donde ξ es el vector aleatorio con ley de probabilidad \mathbb{P} , asumiendo que $f(x)$ está bien definida y es finita para todo $x \in \mathcal{P}$.

Sea $\{\xi_i\}_{i=1, \dots, N}$ una muestra independiente e idénticamente distribuido (iid), de realizaciones del vector ξ . Para cualquier valor de $x \in \mathcal{P}$ se puede aproximar el valor de $f(x)$ por el promedio de los valores de $\mathbf{F}(x, \xi_i)$ con $i \in 1, \dots, N$, de la siguiente forma:

$$\hat{v}_N^* := \min_{x \in \mathcal{P}} [\hat{f}_N(x)] := \min_{x \in \mathcal{P}} \frac{1}{N} \sum_{i=1}^N [\mathbf{F}(x, \xi_i)] \quad (2.5)$$

Se tiene que $\hat{f}_N(x)$ es un estimador insesgado, matemáticamente, $\mathbb{E}[\hat{f}_N(x)] = f(x)$ y además considerando una muestra *iid*, bajo ciertas condiciones de regularidad descritas en [Shapiro and Ruszczyński, 2008], $\hat{f}_N(x)$ converge puntualmente a $f(x)$ cuando $N \rightarrow \infty$. Por lo tanto, se espera que tanto la solución como el valor óptimo de la función 2.5 converjan a los valores del problema real 2.3 cuando N tiende a infinito.

2.4.1. Validación de una solución

Como se trata de un problema de carácter aleatorio, se desconoce el valor de \hat{v}^* como solución real al problema. Se tiene un candidato a solución $\bar{x} \in \mathcal{P}$ y por lo tanto se debe estimar vía cotas estadísticamente válidas el gap de optimalidad, por lo que se presentará la metodología para calcular la cota inferior, superior y gap.

Cota superior

Se tiene que $\hat{v}^* \geq \mathbb{E}[\hat{v}_N^*]$, es decir \hat{v}_N^* genera una cota superior del valor de \hat{v}^* utilizando el promedio del resultado de M repeticiones independientes, con N muestras *iid* del vector parámetros aleatorios en cada una, del problema. Esto genera $\{\hat{v}_{N,j}^*\}_{j=1,\dots,M}$ valores óptimos en cada repetición $j \in M$ y se tendrá como estimador insesgado de $\mathbb{E}[\hat{v}_N^*]$:

$$\bar{v}_{N,M} = \frac{1}{M} \sum_{p=1}^M \hat{v}_{M,r}^* \quad (2.6)$$

A su vez, al ser muestras *iid*, lo son los valores de $\{\hat{v}_{N,j}^*\}_{j=1,\dots,M}$ y se puede estimar su varianza por la siguiente relación:

$$\hat{\sigma}_{N,M}^2 := \frac{1}{M} \left[\frac{1}{M-1} \sum_{q=1}^M (\hat{v}_{M,q}^* - \bar{v}_{N,M})^2 \right] \quad (2.7)$$

Por el teorema central del límite (TCL) la distribución de probabilidad de $\bar{v}_{N,M}$ es aproximadamente normal para un M suficientemente grande y para esto se presenta como cota inferior (LB) del problema para un nivel de confianza α :

$$LB_{N,M}^\alpha := \bar{v}_{N,M} - z_\alpha \hat{\sigma}_{N,M} \quad (2.8)$$

Cota inferior

Considerando a \bar{x} como el candidato a solución y dado que se resuelve un problema de maximización, se puede utilizar una estimación de $f(\bar{x})$ con un muestreo $N' \gg N$, es decir un muestreo mucho mayor que al generado para obtener el candidato a solución y la cota superior $UB_{N,M}$, completamente iid. De esta forma se buscará aproximar la integral de $f(x)$ considerando a $\hat{f}_{N'}(\bar{x}) := \frac{1}{N'} \sum_{i=1}^{N'} \mathbf{F}(\bar{x}, \xi_i)$ como un estimador muestral del promedio y a su vez poder calcular su varianza muestral $\hat{\sigma}_{N'}^2$:

$$\hat{\sigma}_{N'}^2 := \frac{1}{N' - 1} \sum_{j=1}^{N'} \left[\mathbf{F}(\bar{x}, \xi_j) - \hat{f}_{N'}(\bar{x}) \right]^2 \quad (2.9)$$

De igual forma que para la cota superior, dado que las muestras son iid y por TCL, se considera como cota superior (UB):

$$UB_{N'}^\alpha(\bar{x}) := \hat{f}_{N'}(\bar{x}) + z_\alpha \hat{\sigma}_{N'} \quad (2.10)$$

Gap de optimalidad

Se entiende por gap el posible error del candidato óptimo a solución del problema, como a su vez, se está resolviendo un problema bajo incertidumbre de algunos parámetros, se busca un intervalo donde pueda encontrarse v^* , la solución real al problema con un α nivel de confianza. En particular con lo presentado anteriormente se establece que para tamaños muestrales N, N' y número de repeticiones M , se cumple que $UB_{N,M}^\alpha \geq LB_{N'}^\alpha$ y el intervalo es $\left[LB_{N'}^\alpha, UB_{N,M}^\alpha \right]$.

Se define como *gap absoluto* (gap_A 2.11) al tamaño de este intervalo, que dependerá de las unidades en las que se presente la función objetivo y a su vez un *gap relativo* (gap_R 2.12) relativizando esta magnitud al valor absoluto de la cota inferior.

$$gap_A(\bar{x}) := UB_{N,M}^\alpha - LB_{N'}^\alpha \quad (2.11)$$

$$gap_R(\bar{x}) := \frac{gap_A(\bar{x})}{|LB_{N'}^\alpha|} \quad (2.12)$$

Teniendo en consideración esta herramienta de resolución de problemas de optimización MIP estocástica, se presenta ahora la propuesta metodológica para resolver el problema de asignación de productos, bajo criterios óptimos, para la acumulación de capital de clientes del BancoEstado Microempresas.

2.5. Modelo general de acumulación de capital

Como base del Modelo General de acumulacion de capital, se considera como éxito de un Micro-empresario, para BancoEstado Microempresas, la acumulación de capital del cliente, por lo tanto, en su rol social, se plantea este modelo buscando maximizar el éxito o bienestar del cliente, cuya variable de control será la entrega de productos y en el momento que se entregan.

2.5.1. Descripción del modelo

En cuanto a la revisión bibliográfica de modelos similares, existen trabajos sobre el uso de optimización multiobjetivo para la administración de carteras de créditos, por ejemplo [Soyibo et al., 1991] en un banco en Nigeria, que permiten gestionar las carteras cumpliendo las restricciones del negocio buscando maximizar las utilidades del Banco. Otros modelos utilizan como base el modelo de valorización presentado por [Merton, 1974] en base a opciones para calcular riesgo de crédito en los retornos de capital como [Maria and Yuhang, 2004]. Por otro lado en otras lineas de investigación asociados a la relación de clientes (*customer relationships*) presentado por [Sharpe, 1990] con diseños de contratos implícitos asociados a productos de créditos.

A continuación se detallará una breve descripción y explicación del modelo de optimización estocástica que se plantea para resolver este problema, el cual surge de un proceso iterativo de modelamiento para representar la relación entre un cliente y el Banco de forma de entregar productos, que en primera instancia fue mucho más ambicioso y se puede observar en el anexo A , donde se incorporaba como modelo de equilibrio la aceptación de productos a partir de la creencia del cliente de la tasa ofrecida por el banco, no obstante no se contaban con datos necesarios y por tanto requeriría de muchos supuestos para su resolución.

La variable de control del modelo consiste en qué producto, y en qué periodo del horizonte de tiempo entregar al cliente, de forma de maximizar el bienestar del microempresario, entendido esto último, para los criterios del banco y por este trabajo como la acumulación esperada de capital de los clientes, y en otras palabras, como el *Equity* neto del cliente al final del periodo.

Esta modelación se basa en el balance contable de un negocio como el de la figura 2.2, donde se tiene el activo (A) , la deuda (D) y el patrimonio o *equity* (E), se entenderá la tasa de rentabilidad como la tasa de crecimiento del activo, donde estas componentes del balance pasan a ser variables

de estado en cada periodo teniendo en consideración tanto el saldo de deuda contable como la mora por parte del cliente.

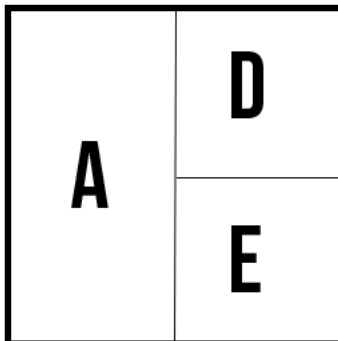


Figura 2.2: Diagrama de la estructura de un Balance

También existirán restricciones de pérdida esperada por parte del banco dada por las provisiones que se esperan tener.

A su vez, para este trabajo se considera un supuesto importante en la racionalidad del cliente al momento del pago y se modelará la variable, como una opción entre el mínimo de lo que le corresponde pagar y lo que puede pagar, considerado esto último como parte importante de su activo contable.

Esto último continúa la idea de [Merton, 1974], que si el activo menos la deuda baja de un cierto punto, se considera un *default*, en este caso un atraso.

Teniendo esto en consideración se presenta ahora el detalle del Modelo de Optimización con sus conjuntos, variables, parámetros, restricciones y función objetivo.

2.5.2. Conjuntos

$$J : \text{Conjunto de Créditos} \tag{2.13}$$

$$T : \text{Conjunto de periodos} \tag{2.14}$$

$$S : \text{Conjunto de escenarios aleatorios} \tag{2.15}$$

2.5.3. Variables

$$X_j^t \in \{0, 1\} : \text{Se entrega crédito } j \text{ al cliente } i \text{ en el periodo } t \quad (2.16)$$

$$Y_j^t \in \{0, 1\} : \text{Corresponde pagar cuota de crédito } j \text{ en periodo } t \quad (2.17)$$

$$A_s^t \geq 0 : \text{Activo del cliente en el periodo } t \text{ en escenario } s \quad (2.18)$$

$$D_s^t \geq 0 : \text{Saldo de deuda del cliente en el periodo } t \text{ en escenario } s \quad (2.19)$$

$$M_i^{t,s} \geq 0 : \text{Mora del cliente en el periodo } t \text{ en el escenario } s \quad (2.20)$$

$$P_s^t \geq 0 : \text{Pago del cliente en el periodo } t \text{ en escenario } s \quad (2.21)$$

$$\delta_s^t \in \{0, 1\} : \text{variable auxiliar binaria para definir pago en caso de que pueda pagar lo que debe} \quad (2.22)$$

La variable 2.16 (X_j^t), principal variable de control para el banco, establece si se le asigna el producto j en el periodo t al cliente. Esta variable, a su vez, es una decisión en primera etapa para el problema estocástico y le da robustez a la solución, dado que se toma la decisión considerando todas las posibles trayectorias o escenarios s que presente la variable aleatoria.

La variable 2.17 (Y_j^t) en función de la variable de control 2.16, para que indique en qué periodos deben pagarse cuotas del crédito j . Establece que si se le entregó el producto al cliente en a lo más el plazo máximo de duración del producto, la variable debe ser 1.

En síntesis, existen variables de control de parte del banco, que asignan o no el producto, variables de estado para el cliente sobre su negocio y la variable de pago, que representa la opción de pago del cliente como su propia variable de decisión sobre esta interacción Banco - Cliente. Que para este problema se modeló de la forma descrita, entre el mínimo de lo que puede y lo que debe pagar.

2.5.4. Variable aleatoria

$$\tilde{r}_s^t : \text{Rentabilidad del negocio del cliente en periodo } t, \text{ en escenario } s \quad (2.23)$$

2.5.5. Parámetros

$$rb_j : \text{Rentabilidad cobrada por el banco del crédito } j \quad (2.24)$$

$$TJ_j : \text{Plazo del crédito } j \quad (2.25)$$

$$S_j : \text{Monto del crédito } j \quad (2.26)$$

$$c_j : \text{Cuota producto } j \quad (2.27)$$

$$A_0 : \text{Activo inicial} \quad (2.28)$$

$$PE : \text{Pérdida Esperada, provisiones del banco} \quad (2.29)$$

$$\beta : \text{Costo financiero de la deuda} \quad (2.30)$$

$$G_1 : \text{Gran } M \text{ para la asignación de producto sin deuda} \quad (2.31)$$

$$G_2 : \text{Gran } M \text{ para definición de pago} \quad (2.32)$$

2.5.6. Función objetivo

$$\text{máx } \mathbb{E}_s [E_s^{T+1}] = \mathbb{E}_s [A_s^{T+1} - M_s^{T+1}] \quad (2.33)$$

Esta función, busca maximizar el *equity* esperado del cliente al final del horizonte de tiempo estudiado, entendiendo como se dijo anteriormente el *éxito* de los microempresarios como acumulación de capital.

2.5.7. Restricciones

Definición variable de estado activo

$$A_s^{t+1} = (A_s^t + \sum_j S_j \cdot X_j^t)(1 + \tilde{r}_s^t) - P_s^t \quad \forall t > 0 \in T, s \in S \quad (2.34)$$

Definición variable de estado deuda

$$D_s^{t+1} = D_s^t + \sum_j S_j \cdot X_j^t(1 + rb_j^t) - P_s^t \quad \forall t > 0 \in T, s \in S \quad (2.35)$$

Definición variable de estado mora

$$M_s^{t+1} = \sum_j c_j \cdot Y_j^{t+1} + b \cdot M_s^t - P_s^t \quad \forall t > 0 \in T, s \in S \quad (2.36)$$

Las restricciones 2.34, 2.35 y 2.36, definen la no anticipatividad y el estado del balance del negocio periodo a periodo, para las variables del activo 2.18, deuda 2.19 y mora 2.20, denominadas de estado. Donde las variables se definen en base a su valor anterior, más la actualización de lo que ocurre en el periodo actual en base a la variable aleatoria de rentabilidad 2.23, pago 2.21 y asignación de productos 2.16.

Definición de opción de pago

$$P_s^{t+1} = \min \left\{ \sum_j c_j \cdot Y_j^t + b \cdot M_s^t ; \alpha(A_s^t + \sum_j S_j \cdot X_j^t)(1 + \tilde{r}_s^t) \right\} \quad \forall t > 0 \in T, s \in S \quad (2.37)$$

La restricción 2.37, define la variable de pago 2.21 como la opción de pagar lo mínimo entre lo que corresponde según las cuotas del producto más mora, o una fracción del activo del cliente en el periodo, que representa a su vez, lo que puede pagar para poder subsistir. Esta restricción es no lineal, por lo que se deben realizar algunas modificaciones para poder resolver el problema, que se presentarán más adelante.

Definición Default

$$G_1(1 - x_j^t) \geq M_s^t \quad \forall j \in J, t > 0 \in T, s \in S \quad (2.38)$$

La restricción 2.38, corresponde a la definición de incumplimiento, es decir, si un cliente se encuentra con cuotas atrasadas. Si la variable M_s^t es positiva en algún t, no se le podrá asignar productos mientras no se haya regularizado su situación. Esto genera un incentivo a que el cliente prefiera estar en una buena relación con el banco, para seguir accediendo al sistema financiero. Mientras sus condiciones permitan hacerlo.

Restricción de riesgo máximo

$$\mathbb{E}_s [D_s^{T+1}] \leq PE \quad \forall s \in S \quad (2.39)$$

La restricción 2.39, representa el riesgo máximo que el Banco está dispuesto a asumir al entregar productos.

Definición de cuotas por producto entregado

$$Y_j^t = \sum_{k=t-TJ_j}^t X_j^k \quad \forall j \in J, t > 0 \in T \quad (2.40)$$

Se asigna al menos un producto

$$\sum_{j,t} X_j^t \geq 1 \quad (2.41)$$

Condiciones iniciales

Para este trabajo se consideran las siguientes condiciones iniciales para las variables del activo, deuda, mora y asignación inicial.

$$A_s^1 = A_0 \quad \forall s \in S \quad (2.42)$$

$$D_s^1 = M_s^1 = X_j^1 = 0 \quad \forall s \in S, j \in J \quad (2.43)$$

No obstante estas pueden ser modificadas dependiendo del caso de estudio y del problema a analizar, por ejemplo se podría considerar una deuda inicial a ser pagada o clientes que empiezan el periodo a estudiar con mora.

Naturaleza de variables

$$X_j^t, Y_j^t, \delta_s^t \in \{0, 1\} \quad (2.44)$$

$$A_s^t, D_s^t, M_s^t, P_s^t \geq 0 \quad (2.45)$$

2.5.8. Modificaciones al modelo para resolución

A este modelo se le deben hacer algunas modificaciones para poder resolverlo. Es necesario linealizar la restricción que establece el pago 2.37, aproximar la restricción en valor esperado de pérdida esperada y la función objetivo.

Linealización de pago

Para esto se agregan las siguientes restricciones, para reemplazar la ecuación 2.37.

$$G_2(\delta_s^{t+1} - 1) \geq -P_s^{t+1} + \sum_j c_j \cdot Y_j^t + b \cdot M_s^t \quad \forall t > 0 \in T, s \in S \quad (2.46)$$

$$G_2(\delta_s^{t+1}) \geq -P_s^{t+1} + \alpha(A_s^t + \sum_j S_j \cdot X_j^t)(1 + \tilde{r}_s^t) \quad \forall t > 0 \in T, s \in S \quad (2.47)$$

$$P_s^{t+1} \leq \sum_j c_j \cdot Y_j^t + b \cdot M_s^t \quad \forall t > 0 \in T, s \in S \quad (2.48)$$

$$P_s^{t+1} \leq \alpha(A_s^t + \sum_j S_j \cdot X_j^t)(1 + \tilde{r}_s^t) \quad \forall t > 0 \in T, s \in S \quad (2.49)$$

Estas restricciones, se pueden interpretar desde un punto de vista financiero como un portafolio replica de opciones lineales, que hacen análoga la opción de pago presentada en el modelo general 2.37.

Aproximación restricción de riesgo

Para la ecuación 2.39, se aproxima el valor esperado de la mora en el último periodo como el promedio de ésta en los N escenarios generados, acorde al tamaño in-sample de la metodología SAA.

$$\frac{1}{N} \sum_s D_s^{T+1} \leq PE \quad (2.50)$$

Sin esta restricción, con este modelo el Banco estaría actuando como un agente neutro al riesgo. Ésta permite un análisis de riesgo para el resguardo de provisiones del Banco, con su pérdida esperada. No obstante como se detalla en el capítulo de Extensiones futuras 5, es posible modificar el valor esperado por otra medida de riesgo ρ_θ para algún valor θ de aversión al riesgo y considerando una función $\mu(\theta, PE)$ como medida para las provisiones del Banco la restricción 2.50 se puede escribir de forma más general como: $\rho_\theta(D_s^{T+1}) \leq \mu(\theta, PE)$.

Al incorporar esta restricción, además de incorporar la noción de riesgo para el banco, sucede algo interesante que se pudo observar posteriormente en el capítulo de experimentos computacionales 4, pero dada sus implicancias teóricas se detallará en este capítulo.

Al resolver el problema con las restricciones de linealización de pago y de pérdida esperada, se observó que se produce degenerancia, por lo que se detallará el análisis.

En primer lugar se define la pérdida esperada como:

$$PE = \sum_j \sum_t PI \cdot S_j X_j^t R_b \quad (2.51)$$

Donde PI corresponde la probabilidad de incumplimiento del segmento. Por construcción también se sabe que $\sum_j \sum_t S_j X_j^t R_b = \sum_j \sum_t c_j Y_j^t$ y por lo tanto la ecuación 2.51 se puede escribir como:

$$PE = \sum_j \sum_t PI \cdot c_j Y_j^t \quad (2.52)$$

Se puede ver que la variables auxiliares binarias δ_s^t , modela el incumplimiento, tal que las veces que para un periodo t y escenario s sería 1 si el cliente no puede pagar, por lo que se tiene que el promedio de todos los casos con incumplimiento es una estimación de la probabilidad de incumplimiento:

$$\frac{1}{N \cdot T} \sum_s \sum_t \delta_s^t = PI \quad (2.53)$$

Con estas definiciones, en la ecuación de riesgo 2.50, descomponiendo recursivamente $D_s^{T+1} = \sum_t \sum_j RB_j - \sum_t P_s^t$, y reemplazando la equivalencia por construcción entre los montos de créditos y suma de cuotas queda:

$$\frac{1}{N} \sum_s (\sum_t \sum_j c_j Y_j^t - \sum_t P_s^t) \leq \sum_j \sum_t PI \cdot c_j Y_j^t \quad \forall j \in J, t \in T, s \in S$$

Reordenando términos con el objetivo de encontrar una cota inferior a los pago se tiene:

$$\frac{1}{N} \sum_s \sum_t \sum_j c_j Y_j^t - \sum_j \sum_t PI \cdot c_j Y_j^t \leq \frac{1}{N} \sum_s \sum_t P_s^t$$

Reemplazando a su vez la relación entre la PI y la variable auxiliar δ_s^t :

$$\frac{1}{N} \sum_s \sum_t \sum_j c_j Y_j^t - \left(\frac{1}{N \cdot T} \sum_s \sum_t \delta_s^t \right) \sum_j \sum_t c_j Y_j^t \leq \frac{1}{N} \sum_s \sum_t P_s^t$$

Esta restricción establece que la suma de la totalidad de los pagos sea mayor igual que lo que le corresponde pagar menos una parte asociada a la pérdida esperada, es decir, plantean una cota inferior y considerando las restricciones de linealización de pago 2.54 y 2.55 que representan una cuota superior a los pagos por periodo t y escenario s :

$$P_s^{t+1} \leq \sum_j c_j \cdot Y_j^t + b \cdot M_s^t \quad \forall t \in T, s \in S \quad (2.54)$$

$$P_s^{t+1} \leq \alpha (A_s^t + \sum_j S_j \cdot X_j^t) (1 + \tilde{r}_s^t) \quad \forall t \in T, s \in S \quad (2.55)$$

Con esto, se tiene que a partir de la restricción de riesgo y pérdida esperada se pueden implicar una parte de las restricciones de linealización de pago, y se pueden omitir en el modelo las restricciones 2.46 y 2.47 que producían degenerancia.

Sin embargo, la importancia de este resultado, al entenderlo como un problema general del regulador se logra capturar que la dinámica de pérdida esperada y el comportamiento de la opción de pago mínimo de agente actúan como un conjunto de restricciones degeneradas, más aún que a partir de una restricción de riesgo de parte del Banco, se puede internalizar el comportamiento de pago mínimo esperado del agente, conectando implícitamente una regulación de riesgo, con la dinámica de pago esperado.

Aproximación función objetivo

Finalmente, la aproximación de la función objetivo la ecuación 2.33, reemplazará el valor esperando, por el promedio de la liquidación neta del Activo por la Mora en el último periodo:

$$\max \frac{1}{N} \sum_s [E_s^{T+1}] = \frac{1}{N} \sum_s [A_s^{T+1} - M_s^{T+1}] \quad (2.56)$$

Ya definido el modelo y su aproximación, en la siguiente parte se busca realizar un análisis de las distintas componentes de éste y de sus implicancias en la resolución, partiendo por un análisis de la función objetivo.

2.6. Análisis de la función objetivo

$$\max \frac{1}{N} \sum_s [A_s^{T+1} - M_s^{T+1}] \quad (2.57)$$

Para este análisis se desarrollará la función 2.57 recursivamente. De esto, se pueden agrupar los términos de dos formas distintas, con las interpretaciones financieras, uno de ellos desde el punto de vista del Banco 2.58 y el segundo del Microempresario 2.59 las cuales se describirán a continuación.

Forma A: Desde el punto de vista del cliente

$$\max \frac{1}{N} \sum_s \left[A_s^1 \prod_{t=1}^T Rn_i^{T-t} - b^T M_s^1 + \sum_j \sum_{t=1}^T ((S_j X_j^{T-t} \prod_{t=1}^T Rn_i^{T-t}) - c_j b^{T-t} Y_j^{T-t}) + \sum_{t=1}^T P_s^{T-t} (b^{T-t} - \prod_{t=1}^T Rn_i^{T-t}) \right] \quad (2.58)$$

De esta ecuación 2.58, se pueden agrupar algunos términos, en primer lugar $(A_s^1 \prod_{t=1}^T Rn_i^{T-t} - b^T M_s^1)$, corresponden a las condiciones iniciales del cliente. En particular, para el problema planteado se entiende que la Mora inicial es 0 (de todas formas no es excluyente), así que el término relevante es el activo inicial por la productoria de las rentabilidades, es decir el valor futuro del activo inicial, ajustado por rentabilidad del negocio.

Un segundo término que se puede interpretar corresponde a: $\sum_j \sum_{t=1}^T ((S_j X_j^{T-t} \prod_{t=1}^T Rn_i^{T-t}) - c_j b^{T-t} Y_j^{T-t})$, que representa a los ingresos recibidos por los productos menos los pagos que se deben efectuar, ajustados por rentabilidad. Se interpreta financieramente como el retorno extra que generan los productos de financiamiento al negocio.

Finalmente, un tercer término $\sum_{t=1}^T P_s^{T-t} (b^{T-t} - \prod_{t=1}^T Rn_i^{T-t})$ permite observar las ‘fuerzas’ en el sentido al que se somete la variable del pago, en particular, si las rentabilidades son mayores al costo financiero por no pagar las cuotas correspondientes, el cliente preferiría, de no ser por las restricciones que establecen la mecánica de pagos, dejar de pagar (o ponerse al día después). En cambio si las rentabilidades son más bajas que el costo financiero, el cliente prefiere estar ‘en buena relación’ con el banco, pagando sus productos.

Reagrupando los términos en base a los otros parámetros, se puede ver la siguiente ecuación 2.59, e interpretarla según el otro agente del modelo, el Banco.

Forma B: Desde el punto de vista del Banco

$$\max \frac{1}{N} \sum_s \left[A_s^1 \prod_{t=1}^T Rn_i^{T-t} - b^T M_s^1 - \sum_{t=1}^T \prod_{t=1}^T Rn_i^{T-t} (P_s^{T-t} - (\sum_j S_j X_j^{T-t})) - \sum_{t=1}^T b^{T-t} (\sum_j c_j Y_j^{T-t} - P_s^{T-t}) \right] \quad (2.59)$$

La primera parte de la ecuación 2.59, al igual que la anterior 2.58, corresponde a las condiciones iniciales del cliente.

Sin embargo al reagrupar los términos, el segundo $(\sum_{t=1}^T \prod_{t=1}^T Rn_i^{T-t} (P_s^{T-t} - (\sum_j S_j X_j^{T-t})))$ se puede interpretar como el retorno que recibe el banco, dado los pagos del cliente menos las colocaciones que entrega, ajustado por la rentabilidad del cliente a valor futuro.

Finalmente, el tercer término $(\sum_{t=1}^T b^{T-t} (\sum_j c_j Y_j^{T-t} - P_s^{T-t}))$ representa los pagos correspondientes menos los pagos efectivos que realiza el cliente, ajustado por el costo financiero.

Dado que se está maximizando el equity neto esperado del cliente al final del periodo, es más interesante el análisis del punto de vista del microempresario, debido a que se pueden observar las “intenciones racionales de pago” .

2.7. Supuestos importantes

Algunos de los supuestos importantes que se consideran en este modelo tienen relación a la eficiencia de los clientes. Pudiéndose esto representar en la siguiente ecuación:

$$\frac{\partial r}{\partial A} = 0 \quad (2.60)$$

Esto significa que la distribución de rentabilidad no depende del nivel de activo que tenga el cliente, es decir, en el periodo de tiempo que se está observando en el modelo, no se consideran saltos entre grupos o en “niveles” de la distribución de rentabilidad de los clientes. Lo cual podría no ser cierto si se analiza al largo plazo a los clientes. Por ejemplo un cliente con una distribución de baja rentabilidad, puede irle bien en su negocio, acumular capital suficiente para “saltar” de grupo y tener mejores rentabilidades. El análisis de saltos entre rentabilidades también quedará propuesto en extensiones futuras y en cómo modelar la eficiencia de los productos que se le entregan a los clientes.

Capítulo 3

Tratamiento de Datos

3.1. Caracterización de los datos y estimación de parámetros

En el marco de este trabajo, se utilizaron datos del BancoEstado Microempresas para la calibración del modelo de optimización. En este capítulo, se describe el tratamiento, análisis y estimación de parámetros y como se mencionó en los alcances de este trabajo, se considera sólo el segmento comercio.

3.1.1. Estimación de parámetros

Productos ofrecidos

La idea, es caracterizar acorde a las variables de tasas, montos y plazos, entre otras, los productos "tipos" ofrecidos por el Banco a sus clientes, por lo que a continuación se detalla el proceso de análisis exploratorio de los datos y su posterior agrupación en productos característicos, utilizando las bases de datos de registros de ventas de productos a clientes, generando así un "set de oferta" que se incorpora al modelo.

El proceso parte con un análisis exploratorio de los datos de ventas de productos, limpieza, selección de atributos, transformación de los datos. A continuación, como se observa en la figura 3.1 se seleccionan conjuntos de Plazos (T) y Montos (M) significativos a partir de la data, para conformar una grilla *lattice* con grupos con centros en el conjunto $C = \{(T \times M)\}$, en los cuales se agrupan los productos cercanos y, a partir de ahí, se seleccionan los más representativos de la oferta actual del banco.

Por alcances del problema se trabajará con los productos de "créditos en cuotas" para la caracterización. No obstante como se detalla en el capítulo 5 de extensiones futuras, es posible incorporar

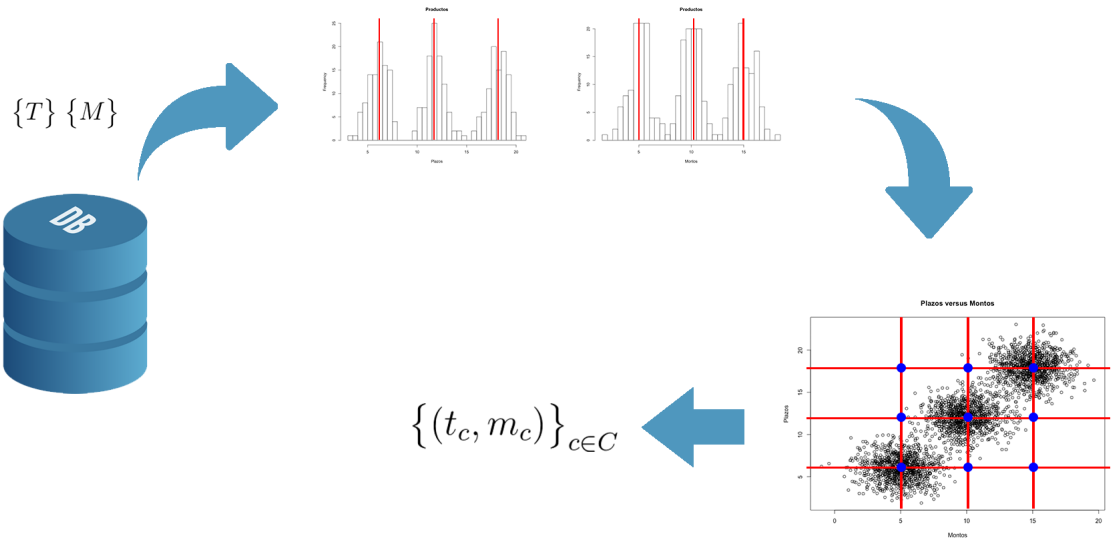


Figura 3.1: Diagrama proceso de agrupación en productos característicos

otro tipo de productos como seguros y garantías haciendo modificaciones e incorporando otras variables.

A partir de estos créditos, se obtendrán los plazos, montos y tasas que permiten obtener la cuota a pagar, parámetros necesarios para incorporarlos al modelo de optimización estocástica. Se realizará una segmentación, para la identificar un conjunto más acotado de créditos, debido a la gran cantidad de combinaciones posibles de estos mismos que actualmente se observan en los datos. En primer lugar se observarán montos y plazos para realizar la segmentación por grupos y finalmente observar la distribución de tasa a partir de las agrupaciones creadas para los segmentos más relevantes.

La base completa del detalle de operaciones tiene 461.959 registros y 102 variables. A ésta, se le incorporarán algunas categorías a partir de las tablas de ventas y detalle de productos. Para obtener mayor información en los registros, se filtrará por el segmento y por tipo de producto para obtener los códigos operacionales detallados anteriormente y los clientes pertenecientes al segmento comercio.

En la base de datos, no todas las variables sirven para el propósito de caracterizar los créditos, debido a que muchas de estas tienen como finalidad otros modelos e indicadores internos del funcionamiento del banco, por lo que, luego del análisis exploratorio, se seleccionaron las variables relevantes a plazos, montos, tasas y tipos de producto. El detalle de variables se encuentra en la

tabla 6.2 del anexo.

Se agregarán a su vez, otras variables construidas a partir de las originalmente seleccionadas entre ellas: la tasa anual equivalente al producto por el plazo que tenga; la cuota (C) asociada al producto a partir de su monto (M); plazo (T) y, tasa con composición mensual (r) según la siguiente relación presentada en la ecuación 3.1. El cálculo de esta cuota, incorpora tanto amortización como intereses y es la convención que se utilizará para los productos en el modelo.

$$C = M \frac{(1+r)^T \cdot r}{(1+r)^T - 1} \quad (3.1)$$

A su vez, a partir de la fecha de venta y la fecha del primer vencimiento, se obtiene la cantidad de periodos de gracia, hasta el primer pago del producto y se agrega también una variable para saber el año del producto.

Luego de este tratamiento de datos, considerando los productos de “*créditos en cuotas*”, la base finalmente queda con 126.857 observaciones y 27 variables. Si bien, se reduce en una gran cantidad los datos, esto se debe a que muchos de los registros de productos consideran “*líneas de crédito*”, créditos hipotecarios, cuentas vista (rut) u otro tipo de productos más específicos, que no aportan al objetivo del modelo, no obstante podrían ser utilizados para tener una idea más clara de la dinámica del activo (A^t) y el equity (E^t) de los clientes.

Se analiza el histograma de montos de productos, para obtener los valores significativos. Por esto, se decide filtrar los datos a montos menores a 6MM\$, considerando que montos mayores representan muy pocos en relación al resto de los registros y serán considerados *outliers*. Así, se tiene una base de 105.985 registros y se pueden observar algunos montos que, en promedio, destacan por sobre los demás como se observa en el gráfico 3.2.

Del gráfico 3.2 se observa que existen montos significativos y estos se observan en los cortes 0, 25MM\$; 0, 5MM\$; 1MM\$; 1, 5MM\$; 2MM\$; 2, 5MM\$; 3MM\$; 3, 5MM\$; 4MM\$; 4, 5MM\$; 5MM\$; 5, 5MM\$ y 6MM\$, por lo que en lo que respecta a los montos, tomará estos valores como una variable discreta. Estos valores son coherentes con lo esperado, concentrándose en cifras “exactas” cada \$500.000 y además se observan altas concentraciones de montos pequeños dada las bajas capacidades de pago de los clientes y al sector y los rubros a los que pertenecen.

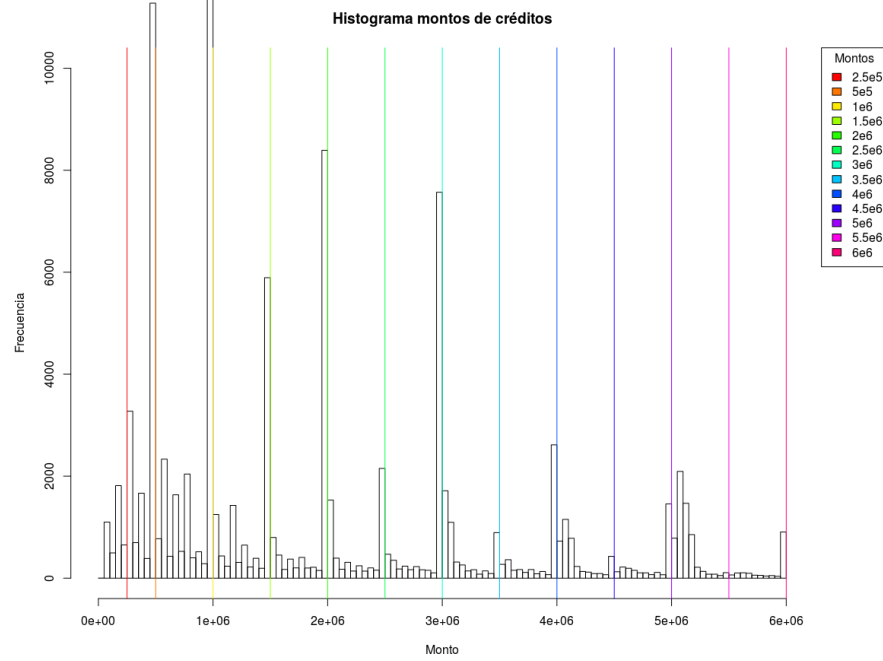


Figura 3.2: Histograma con Montos y varlores representativos considerados

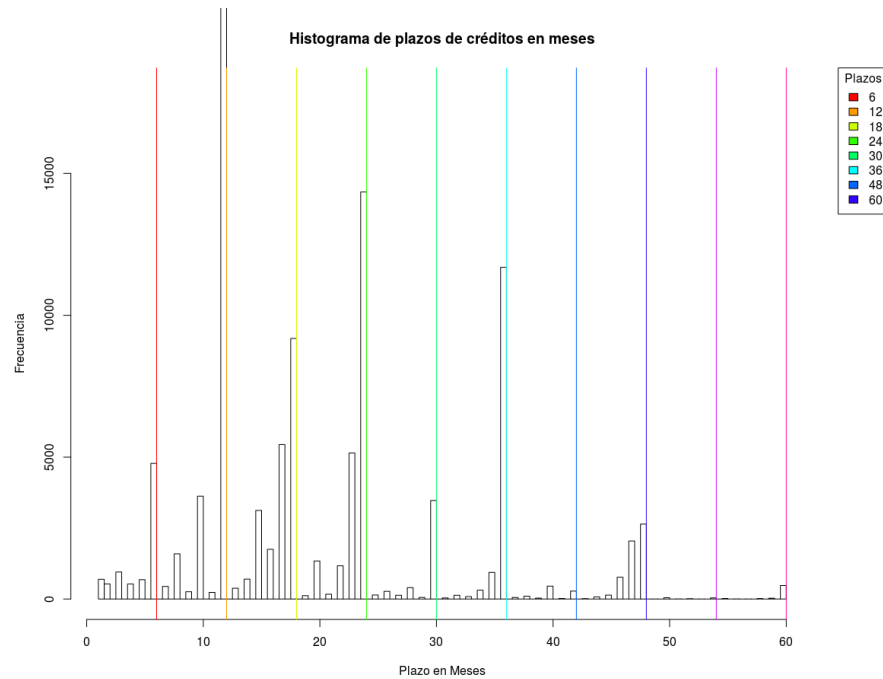


Figura 3.3: Histograma con plazos en meses y meses respresentativos considerados

Ahora se realizará un análisis similar para los plazos, considerando plazos mayores a 60 meses como *outliers*, generando los cortes en los plazos cada 6 meses, considerando entonces 6, 12, 18, 24, 30, 36, 42, 48, 54 y 60 meses como se observa en el gráfico 3.3.

Teniendo esta discretización de plazos y montos, se pueden generar productos representativos de su clase a partir de sus combinaciones, como por ejemplo un crédito de \$500.000 a 6 meses o uno de \$6.000.000 a 48 meses, etc. Todas las combinaciones posibles entre estos cortes da un total de 143 grupos, cantidad que resulta excesiva para poder agregar al modelo de optimización debido a que incrementa en demasía la dimensionalidad del problema. Por lo que se busca disminuir la cantidad de segmentos a las combinaciones relevantes, considerando las combinaciones que expliquen gran parte de las observaciones y dejando de lado aquellas menos recurrentes, como por ejemplo un crédito de \$250.000 a 36 meses.

Para la agrupación de montos, se establecieron los cortes en los montos (ΔM) y plazos (ΔT) según lo observado de los datos en los gráficos 3.2 y 3.3, cuyo detalle se puede observar en la tabla 6.1 del anexo. Finalmente, esta agrupación se utiliza para observar cuántos productos se transan de ese estilo, poder caracterizarlos y, encontrar el conjunto característico.

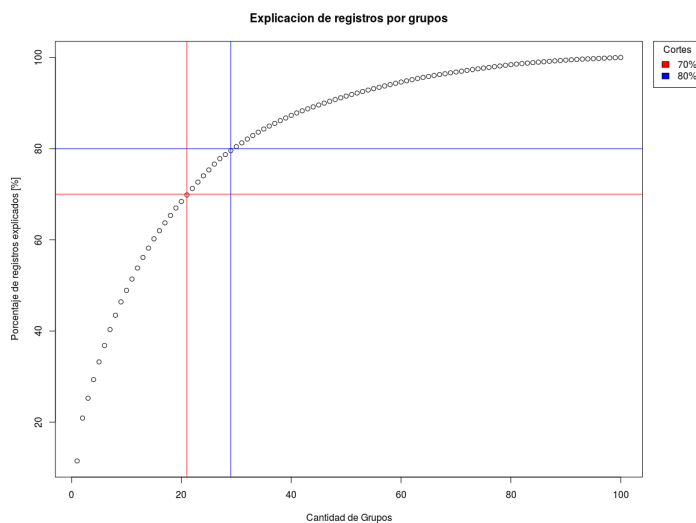


Figura 3.4: Punto de corte que explica el 70 % y 80 % de los datos

Se busca disminuir la cantidad de créditos a utilizar, de modo de seleccionar los más representativos, por lo tanto, al ordenar los grupos por cantidad de registros, existe un crecimiento marginal decreciente al aumentar la cantidad de grupos incorporados, como se observa en la figura 3.4, se consideraron dos puntos al 70 % y al 80 % de la muestra explicada con 21 y 29 productos respectivamente. Se decide finalmente utilizar el corte al 70 %, debido a que la incorporación de más segmentos, al 80 %, no genera mucha más información para el análisis, debido a que al agregar 9 grupos nuevos sólo se explica un 13 % más, del total de datos.

A partir de esto, se considerarán un total de 21 tipos de créditos que se muestran en la tabla 3.1 y cuya clasificación por número de grupo se observa en la tabla 6.1 del anexo. Con estos se identificará la distribución de tasas para cada uno, y se estimará la tasa a utilizar en el modelo, además de calcular la cuota para el producto respectivo a partir de los demás parámetros, con la relación 3.1 explicada anteriormente.

Se filtró la base para estos grupos obteniendo un total de 74.022 registros que corresponden como se mencionó al 70 % de los registros totales. A su vez, para cada uno de estos, se observó gráficamente el histograma y densidad de la distribución de las tasas de interés para estos productos. Se esperaba una distribución normal (o bastante similar) para cada uno de los productos, en particular, se esperaba trabajar con la media de esa distribución, para posteriormente en trabajos futuros poder hacer variaciones y análisis de sensibilidad frente a cambios en la tasa. No obstante, lo que se observó es que la distribución de las tasas es multimodal en los grupos escogidos, donde se pueden observar 3 grupos. Un análisis más en profundidad de estos grupos podría describir clientes “buenos”, “promedio” y “malos”.

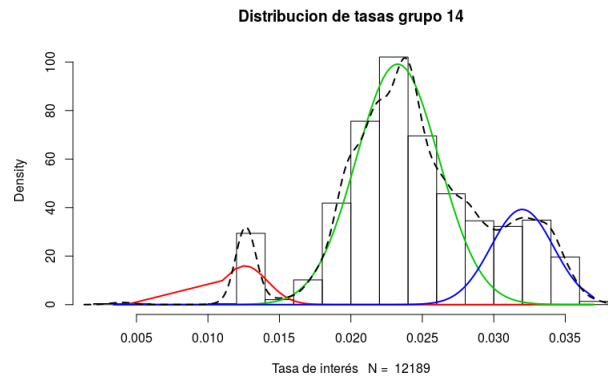


Figura 3.5: Distribución de tasas

Para comprender mejor este análisis se presenta el gráfico 3.5 de una distribución de tasas, ajustada con la misma metodología utilizada para encontrar la distribución de la rentabilidad de los clientes, explicada en el marco teórico, de un modelo de mixtura de normales, pero esta vez con tres grupos, donde se observan las 3 posibles distribuciones de clientes, donde en particular en este caso se observa un pequeño grupo de clientes “buenos” , con tasas menores, uno mayor representando a los “promedio” y otro un poco menor que correspondería al tercer grupo un poco más riesgoso de “malos” .

Finalmente, se trabajará con la media de estas distribuciones de tasas para la creación de los productos y quedará propuesto hacer un análisis más fino de los productos diferenciando estos sub-grupos de clientes por tasa, debido a que se incrementa exponencialmente el número de productos y en consecuencia la dimensionalidad del problema de optimización y, para los objetivos y alcances de este trabajo no aporta mayor información. Por lo tanto, los productos que se seleccionaron son los presentados en la tabla 3.1.

Tabla 3.1: Productos considerados para el modelo

Producto	Grupo	N	Monto	Tasa Mensual	Plazo	Cuota
1	2	1.510	\$250.000	2,505 %	6	\$45.396
2	3	1.808	\$250.000	2,638 %	12	\$24.576
3	13	4.127	\$500.000	2,422 %	6	\$90.539
4	14	12.189	\$500.000	2,447 %	12	\$48.587
5	15	3.719	\$500.000	2,438 %	18	\$34.651
6	16	1.899	\$500.000	2,397 %	24	\$27.640
7	24	1.517	\$1.000.000	2,130 %	6	\$179.308
8	25	9.967	\$1.000.000	2,130 %	12	\$95.315
9	26	4.624	\$1.000.000	2,161 %	18	\$67.652
10	27	4.315	\$1.000.000	2,163 %	24	\$ 53.854
11	36	2.559	\$1.500.000	2,045 %	12	\$142.233
12	37	2.635	\$1.500.000	2,111 %	18	\$101.033
13	38	2.172	\$1.500.000	2,124 %	24	\$80.424
14	47	2.459	\$2.000.000	1,950 %	12	\$188.542
15	48	2.669	\$2.000.000	1,988 %	18	\$133.261
16	49	3.801	\$2.000.000	2,007 %	24	\$105.830
17	70	1.726	\$3.000.000	1,675 %	18	\$194.426
18	71	3.327	\$3.000.000	1,715 %	24	\$153.547
19	73	3.104	\$3.000.000	1,739 %	36	\$112.654
20	95	1.730	\$4.000.000	1,492 %	36	\$144.414
21	117	2.165	\$5.000.000	1,423 %	36	\$178.449

Fuente: Elaboración propia.

Rentabilidad del negocio

Teniendo los créditos representativos de la oferta del Banco, se procede a estimar la distribución de la rentabilidad del negocio de los clientes a partir de los datos entregados, que será uno de los principales parámetros del modelo de optimización estocástica, debido a que presenta aleatoriedad del problema.

Una de las características de BancoEstado Microempresas, en su manejo comercial, es la evaluación en terreno de sus clientes por parte de los ejecutivos del banco. Por lo tanto, para poder estimar la rentabilidad de los negocios, se utilizó la base datos de evaluaciones de clientes. Esto, además, porque no se tenían otras fuentes de información con datos históricos suficientes como para generar una trayectoria de “rentabilidad” del negocio estadísticamente significativa.

Cada registro en esta base de datos consiste de un atributo de una evaluación para un cliente en tres escenarios, por lo que debió realizarse un tratamiento para poder normalizar su estructura, reordenando y agrupando los registros para que cada observación indicara una evaluación por cliente. Esta base quedó finalmente con 1.835.634 registros y 17 variables, entre las que se destacan: *idClientes*, *idEvaluación*, *Deuda*, *Capacidad de pago*, *Costos Fijos*, *Costos Variables*, *Costos Totales*, *Venta Formal*, *Venta Informal*, *Ventas totales*, *Resultado del ejercicio*, *Otros ingresos*, *Gastos no operacionales* y *Gastos familiares*.

La metodología y justificación financiera para el cálculo de la rentabilidad se tendrá a partir de que las variables mencionadas permiten hacer un análisis del estado financiero contable de las micro empresas, en particular obtener la capacidad de pago del cliente.

Teniendo en consideración, como base, las normativas internacionales de información financiera (*IFRS: International Standards Financial Reporting*), al Resultado del ejercicio obtenido por la ecuación 3.2, se le incorporará un ajuste dado por otros ingresos, gastos familiares y otras deudas para obtener finalmente la “*capacidad de pago*”, como se observa en la ecuación 3.3. Debido a que en el segmento microempresario, es complejo separar los negocios, de otros ingresos o gastos familiares por lo que son significativos en los resultados de los negocios, siendo esta una característica muy común de este segmento..

$$\text{Resultado del Ejercicio} = \text{Ventas formales} + \text{Ventas Informales} - \text{Costos Fijos} - \text{Costos Variables} \quad (3.2)$$

$$\text{CapacidadPago} = \text{Resultado del Ejercicio} + \text{Otros Ingresos} - \text{Gastos No Operacionales} - \text{Gastos Familiares} - \text{Deuda} \quad (3.3)$$

Por otro lado, debido a que no se tienen datos tan exactos de venta formal e informal, no se pudo realizar un ajuste por resultados antes y después de impuestos. Esto para el segmento microempresario, en la literatura, suele ser importante, por lo que se tendrá en consideración para extensiones

futuras cómo incorporar este efecto.

Finalmente, se considerará como rentabilidad del negocio lo obtenido por la ecuación 3.4, como el cociente entre la capacidad de pago y las ventas totales como una aproximación del margen después del ejercicio.

$$r_n := \frac{CapacidadPago}{VentasTotales} \quad (3.4)$$

Considerando la metodología anterior se realizará un tratamiento y principalmente limpieza de datos para calcular la distribución de rentabilidad mencionada.

Se realizó imputación de datos en aquellos que podrían ser obtenidos de forma sencilla a partir de otras variables, como los costos totales, ventas totales y resultados operacionales. A su vez, luego del tratamiento de *outliers* y la imputación, se analizó la consistencia de los datos, reconstruyendo la variable capacidad de pago a partir de las demás, y contrastarla con la registrada en la base para el posterior cálculo de la rentabilidad.

Se eliminaron los registros cuyos “*gastos familiares*” fuesen nulos. Considerando los gastos familiares superiores a 5 millones y menores a 10.000 considerandolos *outliers* (Dejando un total de 1.140.735 registros).

Es importante señalar que la variable “*Resultado del ejercicio*” se obtiene a partir de los *costos totales* y *ventas totales*, que a su vez cada una resulta de *costos fijos* y *variables*, y de *ventas formales* e *informales respectivamente*. Al hacer un análisis de los registros de la base de datos, se tiene que algunos de éstos poseen resultados operacionales sin tener necesariamente el detalle completo de sus variables de *costos totales* y *ventas totales*, por lo que para el análisis de datos faltantes, se comienza con esta última variable para imputar datos.

Se revisó la variable “*Resultado del ejercicio*”, y se trató de imputar los datos faltantes o nulos a partir de las ventas y costos declarados en las demás variables. Sin embargo, al analizar estos registros, se observó que la mayoría de los datos de las variables que podrían recrear el resultado del ejercicio también tenían gran cantidad de datos faltantes en sus registros. En un total de 128.525 registros, las variables: *venta formal*, *venta informal*, *total ventas*, *costos variables*, *costos fijos* y *total costos* tenían un 99,96 %; 99,97 %; 98,1 %; 98,06 %; 99,79 % y 97,92 % respectivamente, de datos

faltantes por sobre el total de registros incompletos de la variable “*resultado del ejercicio*”.

Eliminando estos datos, se tienen un total de 1.012.210 registros. Se decide eliminar la variable “*margen operacional*” por tener la totalidad de registros como datos faltantes o con valor 0.

Se realiza un análisis de los datos faltantes y nulos de los costos totales, debido a su importancia para la obtención de la rentabilidad, sin embargo la relación entre los datos faltantes de costos fijos y costos variables. En 52.152 registros faltantes de la variable se tiene que un 99,89 % y 99,94 % los costos fijos y variables respectivamente son datos faltantes, por lo que se decide eliminar estos registros dado a la imposibilidad de imputarlos. A partir de esto la base tiene 960.058 registros.

Se podría utilizar el mismo tipo de análisis para tratar de imputar datos faltantes de ventas totales, sin embargo, ya se tienen registros de sus resultados operacionales y costos totales, por lo que no es necesario realizarlo.

Tampoco se utilizará la variable *gastos no operacionales*, puesto que no se tienen datos para poder ser utilizados, sin embargo esto debería incorporarse en un análisis posterior, con una base de datos limpia y registros más confiables.

La variable *otros ingresos* no tienen sus datos completos, sin embargo sus datos faltantes se asumirán como 0, puesto que no necesariamente todos los microempresarios tienen otros ingresos, y asumir otros ingresos nulos, no daña el modelo, lo hace más robusto.

Para el tratamiento de *outliers*, se comenzó analizando la variable *resultado del ejercicio*, donde se observan algunos datos que están excesivamente fuera de rango, por lo que se consideró un intervalo al 95 % con los registros, de forma que los valores fuesen más coherentes. Luego de esto la base de datos ahora tiene 912.054 registros.

De igual forma, se realizó ahora con los *costos totales*, *ventas totales*, *otros ingresos*, *deudas y capacidad de pago*, considerando el intervalo al 95 % de los datos y eliminando los datos nulos no imputados se obtiene una base con 496.993 registros.

A partir de estos datos, se realiza la comparación entre la capacidad de pago registrada en la

base de datos cuya distribución se concentra en cero, salvo algunos *outliers*, por lo que se puede confiar en la consistencia de los datos.

Finalmente, se utilizarán sólo una evaluación por cliente del total de registros resultante de este análisis quedando un total de 232.477 registros, en particular, se utilizarán los atributos: *capacidad de pago* y *ventas totales*, para calcular la rentabilidad por ventas con la aproximación 3.4 explicada anteriormente.

La distribución de rentabilidad obtenida, permite realizar un análisis interesante, ya que como se explicó anteriormente por rentabilidad de negocio se puede segmentar en dos grupos de clientes. Para estimar las distribuciones de ambos grupos se utilizará el paquete de R *mixtools*, con un modelo de mixtura de distribuciones normales. Esta herramienta utiliza algoritmos de *Expectation-Maximization*, para encontrar estimadores de máxima verosimilitud en modelos probabilísticos.

Se puede observar las distribuciones resultantes en la figura 3.6. Donde se observan dos distribuciones normales, a los que se les pueden asociar tipos de clientes representativos. Un primer grupo muy concentrado con baja rentabilidad y baja varianza y otro grupo de menor proporción con rentabilidades más altas y a su vez, mayor varianza, en otras palabras un grupo de bajo y otro de alto riesgo-retorno.

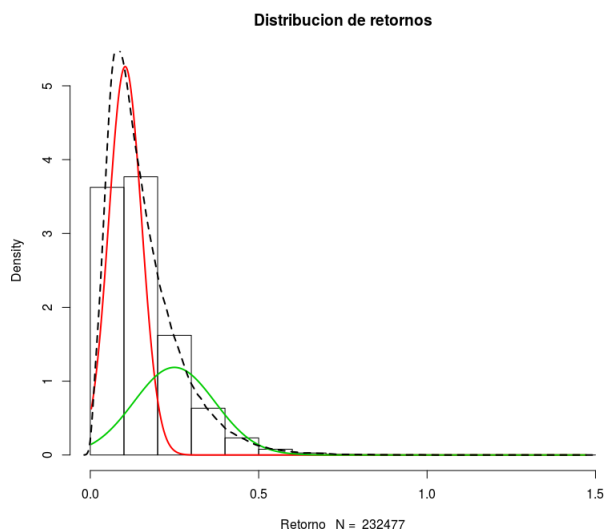


Figura 3.6: Distribución multinormal de retornos del negocio descompuesta con un Modelo de mixtura

Tabla 3.2: Resultados Ajuste distribución

Grupo	λ	μ	σ	N
1	0,682	11,03 %	5,50 %	158.550
2	0,317	28,13 %	14,72 %	73.927

Fuente: Elaboración propia.

Los parámetros obtenidos de ambos grupos se pueden observar en la tabla 3.2. Un importante análisis sobre estas distribuciones es que tienen un sesgo de selección y entrada de clientes, debido a que están calculadas en base a las evaluaciones de clientes del banco, por lo que muchos clientes potenciales, o que a partir de su evaluación fueron rechazados por la institución que, en el papel, les va peor en sus negocios o tienen menos capacidad de pago, pueden modificar la distribución poblacional.

3.1.2. Política actual del Banco

Luego de observar estos dos potenciales subgrupos de clientes, resulta interesante realizar una análisis de cómo se incorpora esta información en la actual asignación de productos del banco. Para esto, se realizó una mezcla entre las observaciones del análisis de productos y de las rentabilidades de clientes.

En la figura 3.7 se puede observar que la distribución de rentabilidad es consistente para los distintos los productos, es decir, actualmente no se le entregan productos distintos a los subgrupos encontrados de rentabilidad alta y rentabilidad baja, si no, se esperaría que algún producto presentara una distribución de rentabilidad distinta, pero se observa que no se discrimina por esta variable de rentabilidad.

En resumen, se logró caracterizar la oferta del banco en un conjunto discreto de 21 tipos de créditos presentados en la tabla 3.1. Además se caracterizaron los clientes en dos tipos en base a su rentabilidad, presentados en la tabla 3.2 uno con alta y otro de baja rentabilidad.

La pregunta ahora a desarrollar y que busca resolver este trabajo es si ¿Será o no conveniente realizar una segmentación por estos dos grupos de clientes? Para lo cual se detallará en el siguiente capítulo el diseño de los experimentos computacionales que permitan abordar una respuesta. Esto permitiría pasar de una asignación basada en capacidad de pago y “ojo experto” a una asignación basada en rentabilidad de clientes, caracterizando y permitiendo mejorar la asignación con el objetivo de tener carteras más “sanas”.

Distribucion de rentabilidades por productos

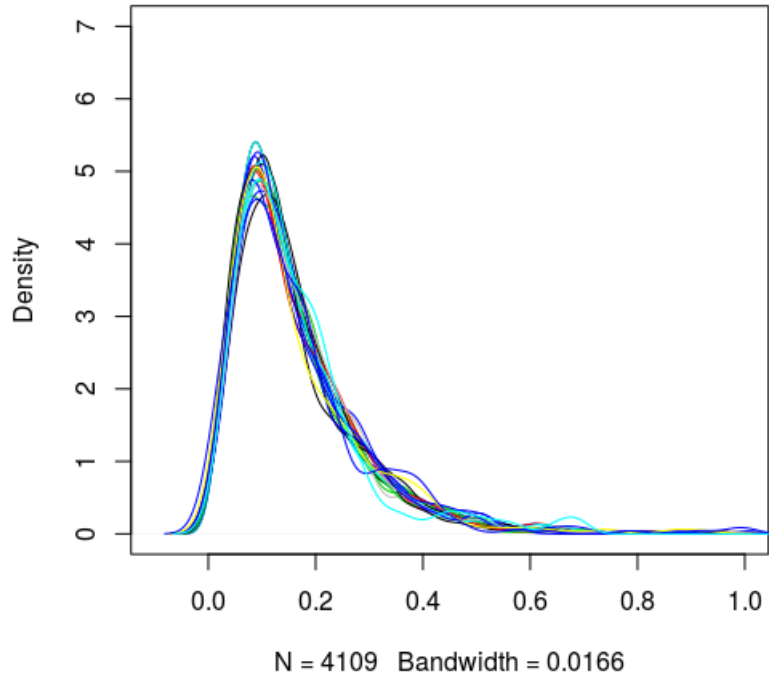


Figura 3.7: Distribución de retornos por productos

Capítulo 4

Experimentos Computacionales y Resultados

4.1. Consideraciones iniciales

La implementación de los modelos, se realizó en el lenguaje de modelamiento AMPL. Adicionalmente se desarrollaron aplicaciones en R para la generación de escenarios y su ejecución con CPLEX 12.0.

Las ejecuciones de estos experimentos se realizaron en el cluster científico de cómputo *Vulcan* del Instituto de Sistemas Complejos de Ingeniería (ISCI) en el Departamento de Ingeniería Civil Industrial de la Universidad de Chile. El cluster se encuentra compuesto por diez nodos de cómputo. Se usaron ocho máquinas con dos quad-core Intel Xeon E5620 de 2.4GHz, 48 GB de RAM, arquitectura x84_64, sistema operativo Linux 2.6.32.

Para el análisis de datos se utilizaron aplicaciones desarrolladas en R para la visualización y lectura de datos de CPLEX y AMPL.

4.2. Experimentos en distribución

En este capítulo se presentan los experimentos computacionales, donde se resuelve el modelo con los parámetros estimados anteriormente. Como se mencionó en el capítulo 3 de Análisis de datos, se identificaron dos grupos de clientes en base a la estimación de la distribución de la variable de rentabilidad del negocio, buscando responder la pregunta de si es conveniente o no diferenciar a estos clientes para la asignación de productos. Por lo tanto, en el diseño de los experimentos computacionales se plantean tres experimentos, en base a las distribuciones de rentabilidad.

El primer experimento se conoce que el cliente pertenece al primer grupo de clientes, caracterizado por bajas rentabilidades y baja volatilidad; un segundo experimento, donde se conoce que el cliente pertenece al segundo grupo, con mayores rentabilidades y mayor volatilidad y finalmente, un experimento donde el Banco no puede distinguir al cliente, por lo tanto su distribución de probabilidad es una mezcla de distribuciones normales de ambos grupos.

A partir de los resultados de estos experimentos se observa que los datos obtenidos de parte del Banco, si bien reflejan la evaluación del cliente, poseen un sesgo de selección y no se registran las colas de distribuciones de pérdida que se observan en los datos reales del Banco. Esto, debido a que las rentabilidades obtenidas presentan valores muy altos.

Para hacerse cargo de este problema, se plantea un diseño experimental donde se ajustan los escenarios a los resultados poblacionales que posee el Banco en sus análisis de las pérdidas e incumplimiento. Esto además, para presentar la metodología y las dinámicas de default que incorpora el modelo planteado en el capítulo 2, dado que los datos reales y actuales del Banco no permiten observar esta dinámica.

A modo de síntesis de los experimentos se presenta la tabla 4.1 junto a su descripción, distribución y activo inicial que se explica a continuación con los demás parámetros iniciales.

Tabla 4.1: Resumen de parámetros por experimento

Experimento	Descripción	A_0	Distribución r_n
1	Grupo 1 de baja rentabilidad con distribución original	\$ 300.000	$N(11, 03\%; 5, 5\%)$
2	Grupo 2 de alta rentabilidad con distribución original	\$ 500.000	$N(28, 13\%; 14, 72\%)$
3	Mezcla de grupos con distribución original	\$ 400.000	$MN\left(\begin{matrix} 0,682; N(11,03\%;5,5\%) \\ 0,317; N(28,13\%;14,72\%) \end{matrix}\right)$
4	Grupo 1 de baja rentabilidad con distribución desplazada	\$ 300.000	$N(0\%; 5, 5\%)$
5	Grupo 2 de alta rentabilidad con distribución desplazada	\$ 500.000	$N(17, 10\%; 14, 72\%)$
6	Mezcla de grupos con distribución desplazada	\$ 400.000	$MN\left(\begin{matrix} 0,682; N(0\%;5,5\%) \\ 0,317; N(17,10\%;14,72\%) \end{matrix}\right)$

Parámetros iniciales

Uno de los parámetros necesarios para la resolución del modelo consiste en el activo inicial del cliente (A_0). Para su estimación se consideraron dos enfoques, el primero como el cociente de la capacidad de pago promedio por grupo y su rentabilidad media (analogando a una perpetuidad de su capacidad de pago) y, como segunda opción simplemente la capacidad de pago promedio por grupo de cliente. Se opta por el segundo enfoque debido a que al considerar microempresarios, los activos físicos en los negocios no poseen mucha liquidez y, al momento de observar el estado del cliente en cada periodo la capacidad de pago obtenida es una mejor aproximación del “capital líquido real” que poseen los clientes.

En base a esto los montos como activos iniciales para los distintos experimentos son los presentados en la tabla 4.1. Por periodo se entenderán meses, y se considerarán un total de 42 meses (3,5 años). Como probabilidad de incumplimiento (PI) se considera un 8%, acorde a la cola en la distribución de pérdidas presentada por el banco en sus datos históricos de incumplimiento y castigo.

Se presentan los tres experimentos base antes mencionados, considerando cada uno una distribución de rentabilidad del negocio del cliente distinta. El primer y segundo experimento se establecen bajo la idea que se conoce al grupo de rentabilidad al que pertenece el microempresario y, un tercer experimento en el que no se tiene noción de a qué subsegmento es parte, por lo que su rentabilidad pertenece a la mezcla de las distribuciones entre grupos.

Para el primer y segundo experimento se considera una distribución normal $\tilde{r} \sim N(\mu_i, \sigma_i)$ con $i \in 1, 2$, es decir, la rentabilidad del negocio distribuye como una normal con media y desviación respectiva al grupo al que pertenece según la tabla 4.1.

Para el experimento 3, utilizando la notación mencionada en el capítulo 2 se tiene que $\tilde{r} \sim g_\theta$, que en este caso, $\theta = (\lambda_1, \lambda_2, (\mu_1, \sigma_1^2), (\mu_2, \sigma_2^2)) = (0.682, 0.317, (11, 03 \%, 5, 50 \%), (28, 13 \%, 14, 72 \%))$. En otras palabras, para la obtención de un escenario aleatorio se tiene que si se lanza una moneda al aire con probabilidad λ sería del grupo 1, al igual que su distribución de rentabilidad, y con probabilidad $1 - \lambda$ del grupo 2, resultado que converge en probabilidad al modelo de mixtura de normales.

Como se mencionó, en la estimación de parámetros en el capítulo 3, la variable de rentabilidad posee un sesgo de selección que se observó en los primeros experimentos. Este sesgo, tampoco logra

capturar los resultados poblacionales observados en los datos históricos, referentes al incumplimiento actual del banco. La “rentabilidad” que plantean estos datos, es coherente en cuanto a su distribución, no obstante, plantean escenarios demasiado favorables, que en la realidad no se observan.

Mantener los resultados con este nivel de rentabilidades, no tiene mucho sentido, debido a que no se observan casos donde los clientes no pueden pagar o caen en mora. Lo que significa que las variables utilizadas para la obtención de las capacidades de pago, y en su defecto rentabilidad, no capturan completamente otros efectos que causan incumplimiento.

Para hacerse cargo de ese problema, se plantea este nuevo experimento donde se asume que existe algún efecto externo, por ejemplo, una crisis como un proceso de poisson que genera “*shocks*” aleatorios sobre la distribución de rentabilidad con cierta tasa λ , que en valor esperado y para efectos prácticos del problema significa desplazar la distribución de rentabilidad del cliente en un $x\%$, lo que permite obtener resultados coherentes con los datos poblacionales observados.

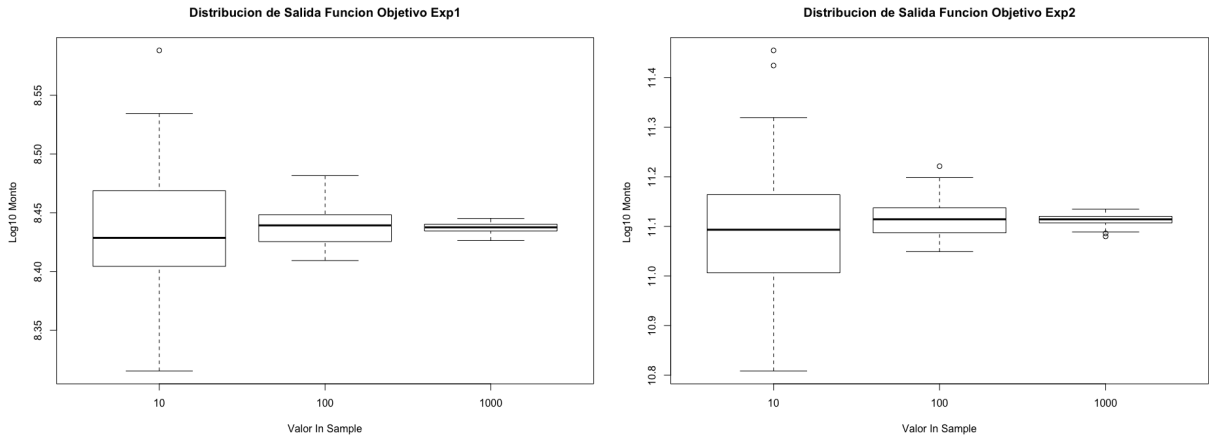
No existieron los datos necesarios para estimar las distribuciones de estos efectos exógenos, por ello se diseñaron tres experimentos adicionales considerando un desplazamiento lineal en las distribuciones. El ajuste se consideró fijando el promedio del primer grupo en torno a 0. Por lo tanto, la media del segundo grupo queda en $\mu_2 = 17,1\%$, manteniendo los parámetros de mezcla y dispersión constantes. Cabe mencionar que agregar *shocks* aleatorios a la rentabilidad a través de un proceso de Poisson, converge en probabilidad a un desplazamiento en el promedio.

4.3. Análisis de resultados

Los resultados obtenidos de los distintos experimentos se presentan a continuación. Estos fueron obtenidos posterior a un análisis de convergencia, tiempos de resolución y candidatos a solución óptima, que se detalla más adelante como justificación.

Los gráficos presentados a continuación, muestran la acumulación de capital esperada de los clientes y las variables de estado del modelo que son gatilladas por la variable de control, de entregar un producto en algún periodo, cumpliendo las restricciones de riesgo. Se observa las diferencias entre los grupos en los últimos experimentos diferenciando los productos a entregar lo que permitirá definir estrategias de asignación de productos para diferenciar los clientes en base a rentabilidad

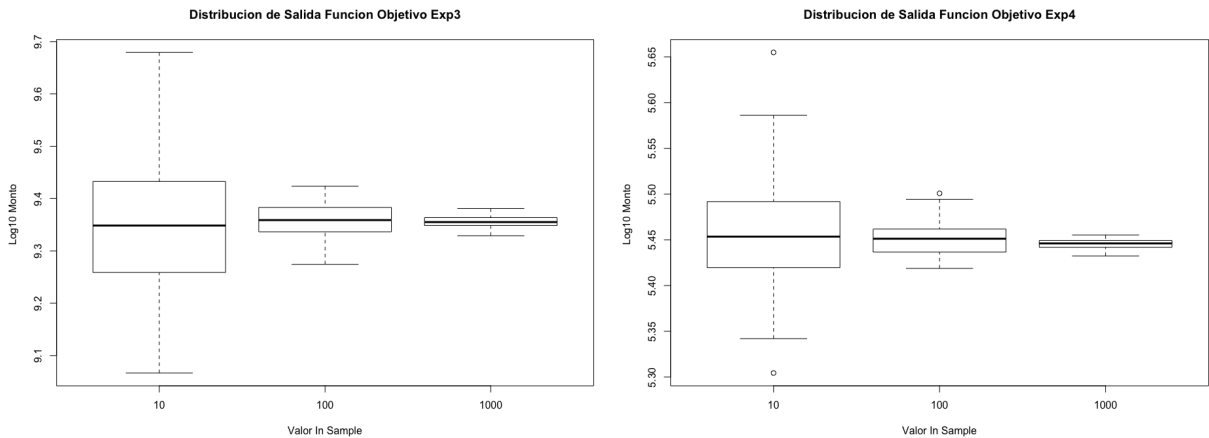
esperada.



(a) Distribucion \log_{10} Funcion objetivo exp1

(b) Distribucion \log_{10} Funcion objetivo exp2

Figura 4.1: Distribucion \log_{10} funcion objetivo exp1 (izq.) y exp2 (der.) por tamaño in-sample



(a) Distribucion \log_{10} Funcion objetivo exp3

(b) Distribucion \log_{10} Funcion objetivo exp4

Figura 4.2: Distribucion \log_{10} funcion objetivo exp3 (izq.) y exp4 (der.) por tamaño in-sample

En primer lugar se puede observar en las siguientes figuras 4.1, 4.2 y 4.3 las distribuciones del resultado de la función objetivo por experimento. Los experimentos 1, 2 y 3 presentan montos excesivamente altos, (más de dos mil veces el monto inicial), que como ya se mencionó fueron el gatillante para el diseño de los siguientes experimentos (4,5 y 6) con desplazamiento en su distribución.

Los resultados de los experimentos 1, 2 y 3 permiten la verificación de la metodología, cálculos de convergencia estocástica y funcionamiento del modelo. No obstante, no son de tanta utilidad para el objetivo final de este trabajo que es generar estrategias de producto en base a la caracterización

de clientes de un segmento del BancoEstado Microempresas. Por lo tanto, para la generación de estrategias se detallarán más en profundidad los experimentos 4, 5 y 6

El experimento 4 presenta resultados bastante similares al activo inicial de este grupo (\$300.000), lo cual tiene mucho sentido ya que sus rentabilidades fueron fijadas en torno a cero. Para el experimento 5, se observan montos más altos, producto de las mayores rentabilidades e influenciado además por el producto que se le entrega en los candidatos a solución.

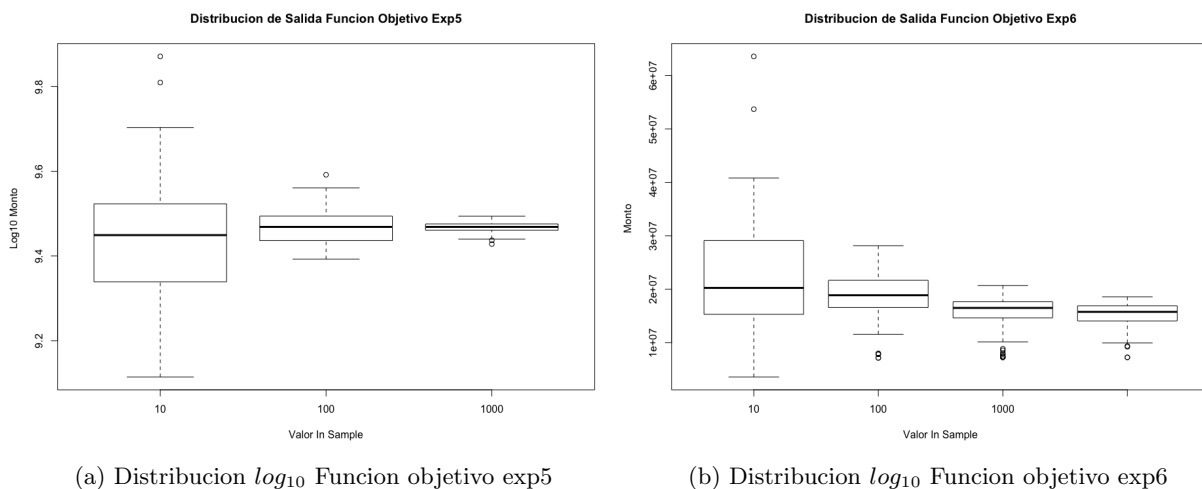


Figura 4.3: Distribucion \log_{10} funcion objetivo exp5 (izq.) y exp6 (der.) por tamaño in-sample

El experimento 6 se observa que a medida que aumenta el tamaño *in-sample* comienza a converger, no obstante se siguen viendo algunos *outliers* producto de la mezcla de candidatos a solución.

Para el caso del experimento 4, la entrega del producto 1 (de menor plazo y monto) para las 100 repeticiones se produce casi homogéneamente, con una leve concentración al principio o al final del horizonte de tiempo. Se puede concluir de esto, que dependerá de la trayectoria de rentabilidades, si tiene trayectorias más altas, puede preferir entregar antes, si tiene más bajas se entrega hacia el final del horizonte de tiempo, si son mixtas, en períodos intermedios, es decir, hay una relación temporal con la rentabilidad que es coherente a lo esperado. En los gráficos de la figura 4.4, se observa el activo (izq.) y mora (der.) en el último periodo del horizonte de tiempo, por repetición, donde en el caso de la mora se observan algunos casos *outliers* que escapan de la concentración en cero, que corresponderían a casos de *default*. Por otro lado el activo, dada la distribución normal de los retornos, se observa también una distribución similar a una normal pero con varios *outliers*

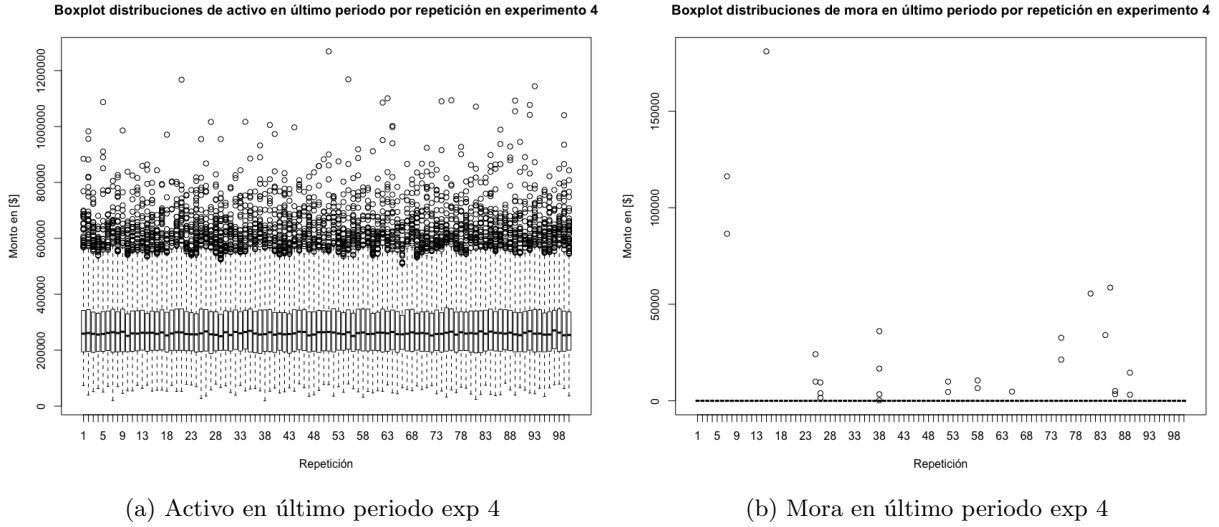


Figura 4.4: BoxPlots distribuciones activo (izq) y mora (der) en último periodo experimento 4

de trayectorias que les fue mucho mejor que al promedio. Es cuestionable, que los valores de la cola del activo llegue a montos menores de 50.000 o que estén más cercanos a cero, debido a que esto no les dejaría margen de acción para su negocio.

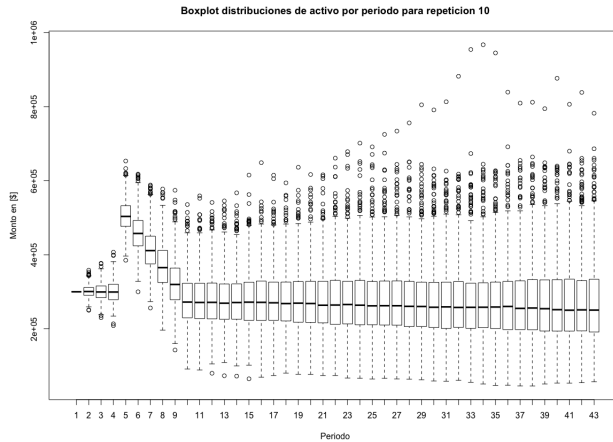


Figura 4.5: BoxPlots distribución de activo por periodo para repetición n°10 de experimento 4

Es interesante analizar las distribuciones del activo por periodo, por ejemplo para la repetición 10 en la figura 4.5, donde se observa que se le entrega el producto en el periodo 5 (por el salto en el activo), y los siguientes meses se van realizando los pagos. Por un lado, se observa que posterior a la realización de los pagos, existen trayectorias que se escapan donde a los negocios les va “muy bien” en comparación a los demás. No obstante, la media se tiende a mantener en torno al activo inicial, y como se explicaba anteriormente existen trayectorias que alcanzan niveles bastante bajos

y cercanos a cero. Se podría hablar de escenarios donde si bien no hay *default* por no pagar, no les alcanza para seguir manteniendo o gestionando el negocio. Si bien esta es una repetición particular, se observan comportamientos similares en diferentes escenarios, variando el periodo de entrega del producto y la cantidad de *outliers*. Otro efecto que se observa, es que al considerar la distribución normal en torno a cero, hace que estos negocios no “crezcan” si no que vayan “subsistiendo” en el tiempo.

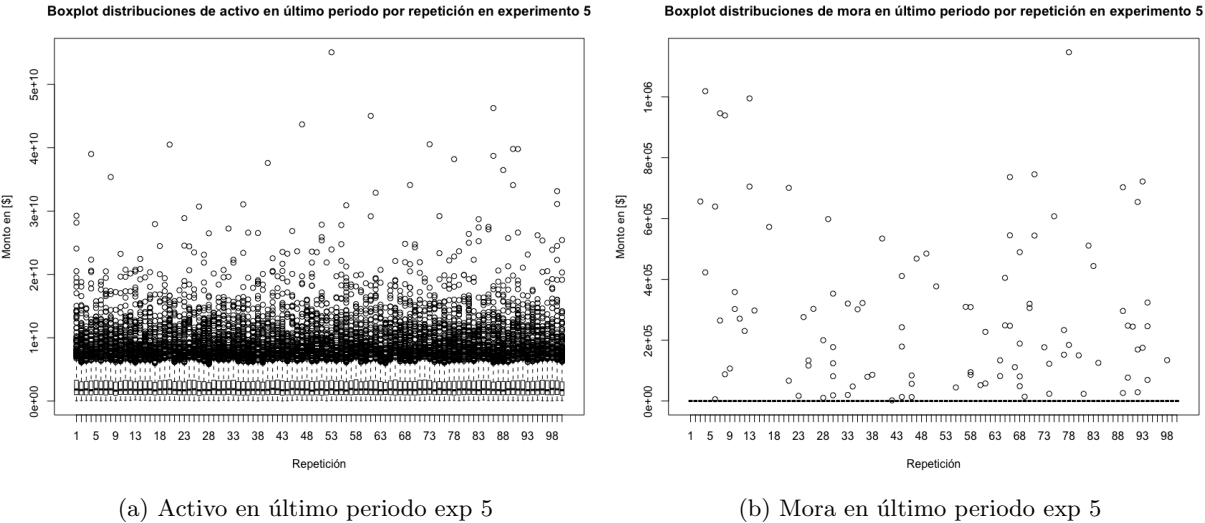


Figura 4.6: BoxPlots distribuciones activo (izq.) y mora (der.) en último periodo experimento 5

Haciendo un análisis análogo al experimento 5, se puede observar en la figura 4.6 que por parte del activo, los montos que se manejan son mucho mayores, gatillados por las altas rentabilidades de este grupo y además por el producto que se entrega que es de mayor monto (5MM), se observan muchas más trayectorias *outliers* con activos mucho mayores a los del promedio, lo que al tener como función objetivo el valor esperado, permite que existan mayores montos de mora para algunas trayectorias como se ve en el gráfico de la derecha, es decir se observan casos de default con montos mucho mayores lo que conlleva mayor riesgo.

Al observar la distribución del activo en una repetición en el gráfico 4.7 después de entregar el producto las trayectorias se mantienen concentradas en torno al promedio y, cuando ya termina de pagar el producto, crecen considerablemente ya que se les permitió acumular mucho más capital y no se les va descontando mes a mes los pagos. La media en las distribuciones también tiene una tendencia al alza en el tramo final del horizonte de tiempo, explicado como se señaló, por las altas rentabilidades de este grupo.

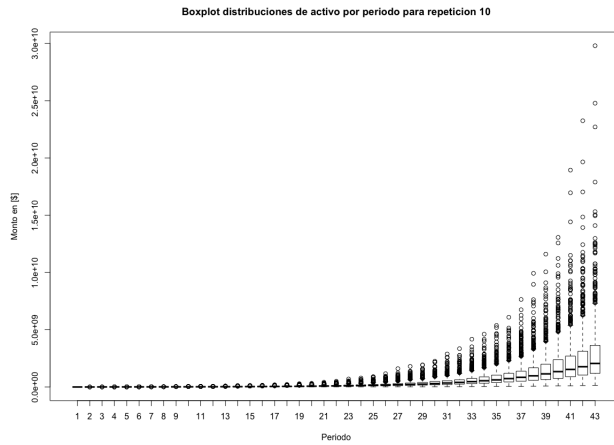
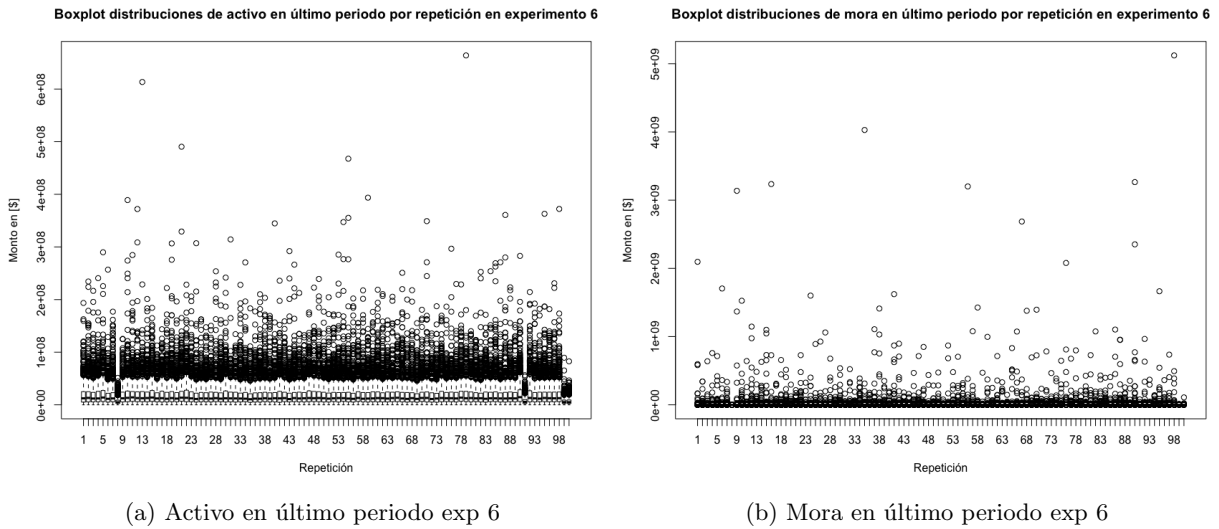


Figura 4.7: BoxPlots distribución de activo por periodo para repetición n°10 de experimento 5



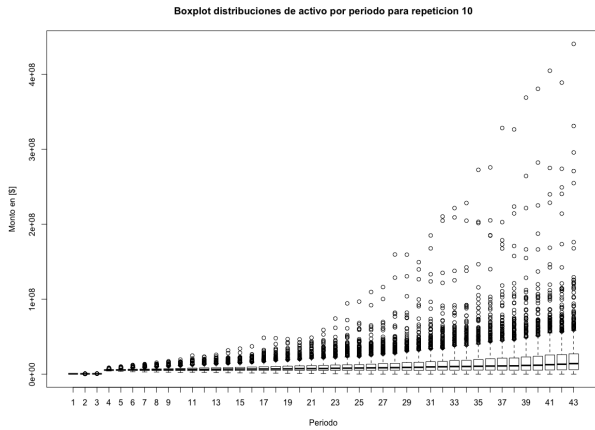
(a) Activo en último periodo exp 6

(b) Mora en último periodo exp 6

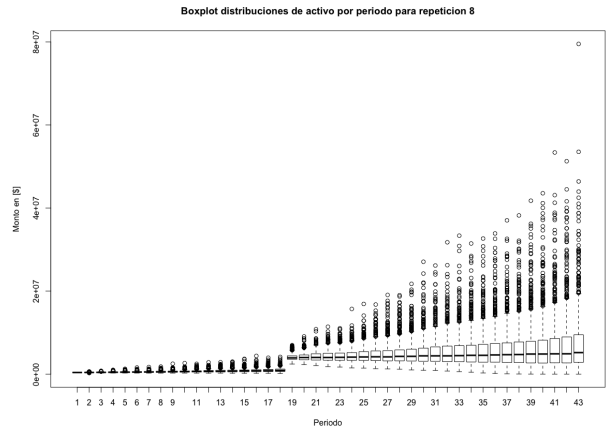
Figura 4.8: Distribuciones activo en repetición 10 (izq) y 8 (der) por periodo

Finalmente el experimento 6, con la mezcla de normales en los gráficos de activo 4.8 se observan montos menores que en el experimento 5, y además se observan 4 repeticiones considerablemente distintas a las demás, debido a que la solución de esas repeticiones entrega productos distintos al 21 y como se vió en la tabla 4.2, se entrega el producto 18 en los periodos 15, 16, 17 y 18, lo que en terminos de acumulación de capital, posee menos periodos para hacerlo y se manejan montos menores. También se observan significativamente más trayectorias con mora, producto de las diferencias y saltos en rentabilidad que ocurren al ser una mezcla de las distribuciones del grupo 1 y 2.

En las figuras 4.9 se observan los activos por periodo para las repeticiones 10 (izq) y 8(der),



(a) Activo por periodo repeticion 10



(b) Activo por periodo repticion 8

Figura 4.9: BoxPlots distribuciones activo (izq) y mora (der) en último periodo experimento 6

donde se entrega el producto 21 y 18 respectivamente y se observan las diferencias con respecto al experimento 5, al no tener una explosión en la distribución en los periodos finales del horizonte de tiempo, a su vez la tendencia de la media es plana, pero con “*colas gordas*” asimétricas en montos mayores.

A modo de síntesis, se observa que el modelo refleja bien las interacciones del Banco y cliente, capturando dinámicas de pago y que acorde a lo esperado se observan diferencias entre los productos entregados para ambos grupos de clientes que dan pie a las estrategias presentadas a continuación.

4.3.1. Estrategias de productos

Como se ha mencionado a lo largo de los capítulos anteriores, se encontraron dos grupos de clientes y el objetivo final es encontrar estrategias de productos distintas para cada uno, considerando el caso donde un cliente se acerca a solicitar algún producto y el banco tiene como diferenciar a cuál grupo pertenece; utilizar dichas estrategias y, por otro lado, si no se tuviese la certeza a qué grupo pertenece, plantear una solución robusta considerando la posibilidad de que un cliente pertenezca a cualquiera de los dos grupos.

En los primeros experimentos se observó que no existe diferenciación en las estrategias de productos entregadas, debido a los sesgos de entrada presentados en los datos de las rentabilidades y dado que a todos los clientes siempre les va bien y tienen buenas rentabilidades debido a que no existen registros de los casos “malos”. La mejor estrategia es entregarles el producto de mayor

monto, ya que de todas formas lo podrán pagar y así invertir en su negocio.

Lo anterior en la realidad no ocurre, porque no todos los clientes tienen las rentabilidades suficientes para poder costear los pagos de los productos, por lo que se realizó un desplazamiento $-L$ en la distribución de rentabilidad de forma de centrar el primer grupo en 0. El modelo general planteado presentado en el capítulo 2 del marco teórico es convexo y es sensible a las distribuciones de rentabilidad, por lo que con este desplazamiento se generó un escenario “extremo” al primero. Donde ambos grupos son notoriamente distintos a los del grupo 1 y les va “muy mal”, es decir su negocio no crece y subsiste y al grupo 2 les sigue yendo “bien”, pero tienen casos donde pueden caer en mora.

Grupo1

Estos resultados llevan a una conclusión algo simple, donde las estrategias para los clientes del grupo 1 serán entregar el producto de menor monto dentro de lo posible, de forma que el cliente pueda pagar y no sobreendeudarlo. Con respecto a los periodos, los mejores resultados se dan en el comienzo o final del horizonte de tiempo, pero en periodos intermedios los resultados son inciertos.

Grupo2

Con respecto al grupo 2 entregarles el producto de mayor monto y plazo de forma de que le siga yendo bien, sin embargo esto de todas formas implicará ciertas pérdidas esperadas, dadas por la probabilidad de incumplimiento. La importancia de este grupo es que son menores en cantidad a los del grupo 1 (30 % del total observado). La dificultad a futuro recará en lograr identificarlos.

Será interesante encontrar el punto de desplazamiento L “ideal”, que permita encontrar resultados que reflejen de mejor forma la realidad de los microempresarios, porque esto permitiría encontrar estrategias distintas para el grupo 1 en particular, ya que no sería una respuesta obvia entregarles el producto más bajo, si no, que existirá algún set de productos que permita maximizar su capital acumulado esperado.

A modo de prueba, se resolvió el experimento 4 con un desplazamiento en la distribución de rentabilidad de $\frac{3L}{4}$ a modo de ilustrar lo señalado anteriormente.

En este caso, se entrega mayoritariamente el producto 18 (\$3MM en 24 meses) en los periodos 15,16, 17 y 18 y se entrega el producto 21(\$5MM en 36 meses), 1 vez en el periodo 6 y finalmente el

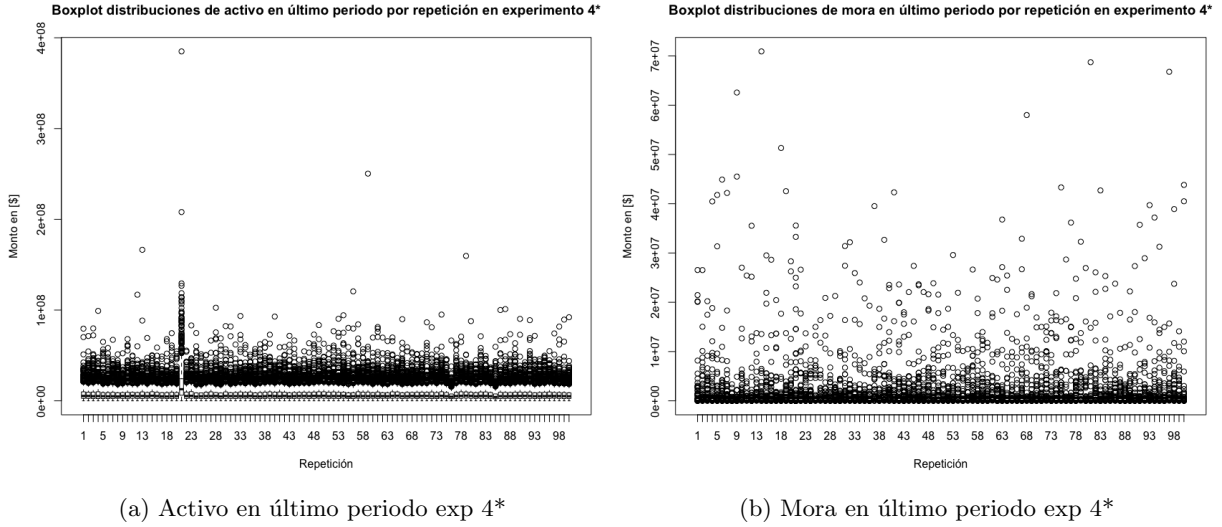


Figura 4.10: BoxPlots distribuciones activo (izq.) y mora (der.) en último periodo experimento 4*

producto 17 (\$3MM en 18 meses) en el periodo 22 y 21, 1 vez en cada uno. Dado esto, ya que existe un crecimiento en la rentabilidad, se entregan productos que permiten acumular más capital a los microempresarios como se observa en la figura 4.10. No obstante, en contraste con los resultados con la rentabilidad en torno a 0, en este caso se observan mayores montos de mora en las distintas repeticiones.

El desafío a futuro será encontrar los factores que hacen que la distribución de rentabilidad no sea la calculada en este trabajo. Tal vez identificar variables exógenas y poder calcular mejores estrategias. Dado que el modelo es lo suficientemente robusto para cambios en la distribución.

4.3.2. Dinámicas capturadas

Una de las principales dinámicas que se buscaba capturar con este trabajo, corresponde al pago, entendido como se explicó en el capítulo 2, como la opción de pago mínimo entre lo que puede y lo que le corresponde. Si bien en los primeros experimentos se tenían escenarios donde no se producen casos de *default*, al realizar los experimentos siguientes si se pueden apreciar las dinámicas de pago. Como se observa en la figura 4.11, que corresponde a la deuda contable por periodo para una repetición (izq.) del experimento 6, se observa como escalonadamente se van realizando los pagos en los periodos hasta que llega un punto en algunas trayectorias donde no se puede seguir pagando y se “mantiene” la deuda contable. A su vez, cuando deja de pagar, la mora “aparece” y empieza a crecer, acumulando a su vez el interés que se observa en la figura 4.11 (der.).

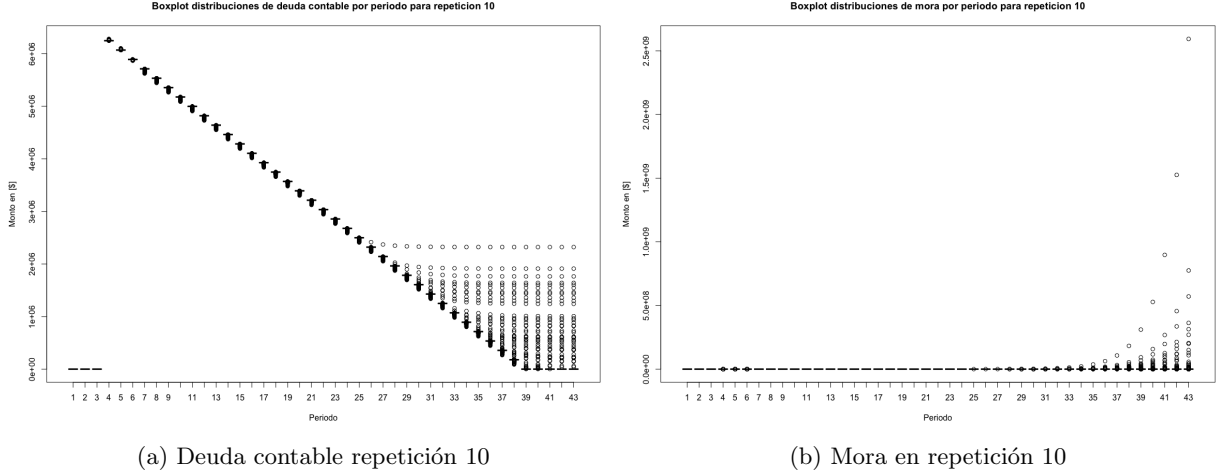


Figura 4.11: BoxPlots distribuciones deuda contable (izq) y mora (der) por periodo para repetición 10 en exp 6

4.3.3. Tiempos de ejecución y convergencia estocástica

Transversal a los distintos experimentos en distribución, es necesario que los resultados realizar un análisis de convergencia estocástica. Como se discutió anteriormente en el marco teórico 2, la convergencia estocástica hace referencia al tamaño relativo del intervalo de confianza de SAA. Se presenta entonces el tamaño relativo del intervalo de confianza de la función objetivo generado por SAA.

Para la validación de estos datos, se realizó una primera corrida de experimentos considerando un conjunto de tamaños *in-sample* $N = \{10, 100, 1000\}$ para el vector de escenarios ($s \in S$) o trayectorias de rentabilidad y donde se realizaron hasta $M = 100$ repeticiones del problema de optimización equivalente cierto, con muestras independientes idénticamente distribuidas (*iid*) de tamaño $n \in N$, con sus respectivas soluciones $\{v_r, X_r\}$. El objetivo de esto es determinar de forma robusta el número de repeticiones y tamaños muestrales *in-sample* (N) y *out-sample* (N') a utilizar considerando el gap estocástico como medida de convergencia y los tiempo de resolución.

En cuanto a los tiempos de solución, se midió el tiempo de resolución de las 100 distintas repeticiones para los distintos tamaños *in-sample* $n \in N$. Calculando el \log_{10} del promedio geométrico de los tiempos de las 100 repeticiones y se presenta en el gráfico 4.12. De este mismo gráfico se observa que a medida que aumenta el tamaño in sample en potencias de 10, los tiempos de resolución en la

escala logarítmica aumentan linealmente, para los experimentos 1, 2, 3 y 6, es decir, el tiempo de resolución promedio del problema aumenta exponencialmente.

Sin embargo, para los demás experimentos 4 y 7 se observa un quiebre en el tiempo de resolución, lo que se puede explicar porque tienen que lidiar con muchos más escenarios de bajas rentabilidades, donde la estructura del problema de optimización se hace más compleja (con más restricciones activas). En particular el experimento 4 presenta un aumento considerable en el tiempo de la resolución, a medida que se aumenta el tamaño *in-sample*.

Otro aspecto interesante de observar en el gráfico 4.12, es que los experimentos 1 y 6 parten dominados por los demás en tiempo de resolución al tener un *in-sample* de 10, efecto que se revierte al ir aumentando el tamaño del conjunto de escenarios. Esto se puede explicar ya que al tener más escenarios además de aumentar su complejidad, en estos dos experimentos, presenta más casos desfavorables de rentabilidad y aumenta la cantidad de iteraciones del algoritmo de resolución del problema. Por otro lado, el quiebre que se ve en la linealidad para el experimentos 6 se puede asociar a que a partir de 100 escenarios se comienza a observar la distribución real de rentabilidad.

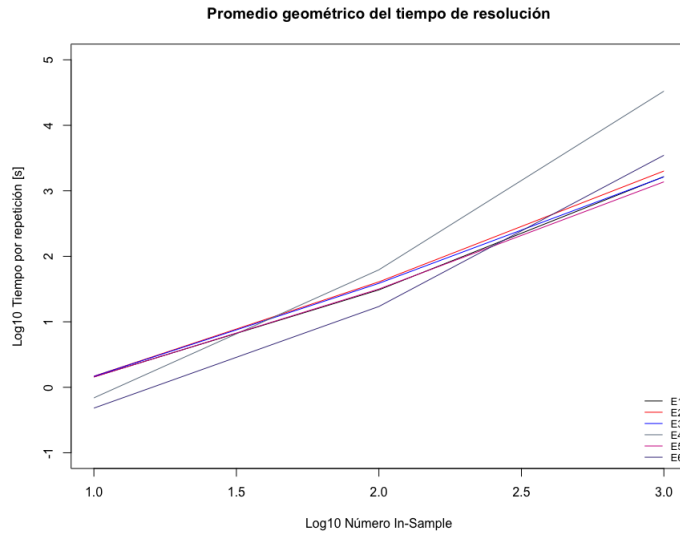


Figura 4.12: Promedio geométrico de tiempos de resolución por tamaño in-sample

Algo interesante que se puede observar que el tiempo del experimento 7, es decir, la distribución de mixtura tiene tiempos menores para los tamaños *in-sample* 10 y 100, pero al aumentar a 1000 se produce el quiebre y supera a los demás experimentos, salvo el de la normal del grupo 1 desplazada, que debido a sus niveles de rentabilidad su tiempo de resolución es mucho mayor.

Esto tiene implicancias en la resolución del problema, si se quiere agregar escenarios, los tiempos de resolución crecen de manera exponencial, por lo que se hace complejo aumentar la dimensionalidad del problema y en el futuro debiese ser considerado algún algoritmo de descomposición, por ejemplo al aumentar la cantidad de productos que se quieran incorporar al modelo.

GAP estocástico

El análisis del GAP estocástico, se utiliza para el ajuste de los parámetros de SAA (n, n', m) de tamaño *in-sample*, *out-sample* y número de repeticiones para garantizar validez estadística, por lo que en lo siguientes cálculos se muestra el cálculo del GAP para los distintos experimentos para obtener los parámetros de SAA. En particular, se utilizará el gap relativo que se presentó en el marco teórico 2, debido a que el gap absoluto depende de las magnitudes de la función objetivo y a su vez de las unidades en que se encuentra esta última.

Primero se analizará el número de repeticiones y cómo varía el promedio acumulado a medida que se agrega una repetición, para luego incorporar la diferencia mediante el tamaño *in-sample* y, finalmente, el tamaño *out-sample* para el resultado del gap estocástico.

El valor de la cota superior depende directamente del tamaño *in-sample* y el número de repeticiones y el valor de la cota inferior del tamaño muestral del *out-sample*.

Número de repeticiones (m)

El número de repeticiones, hace referencia al número de instancias a resolver del problema equivalente cierto, su aumento permite reducir la varianza de la cota inferior para un tamaño *in-sample* representativo y también genera más candidatos a solución, lo que puede reducir el GAP absoluto al permitir refinar el cálculo de la cota superior.

Para este análisis se consideró un rango entre 1 y 100 repeticiones, que se presentan en el siguiente gráfico 4.13 donde se muestra para cada experimento el cálculo del error relativo (izq.) y el tiempo promedio de resolución (der.) por número de repeticiones para los distintos tamaños *in-sample* $N = \{10, 100, 1000\}$. De este gráfico se observa que, a partir de 60 repeticiones para el error relativo tiene un comportamiento estable en la mayoría de los experimentos, salvo el experimento 4, el que a medida que aumentan las repeticiones se comienza a estabilizar al llegar a 100. En relación

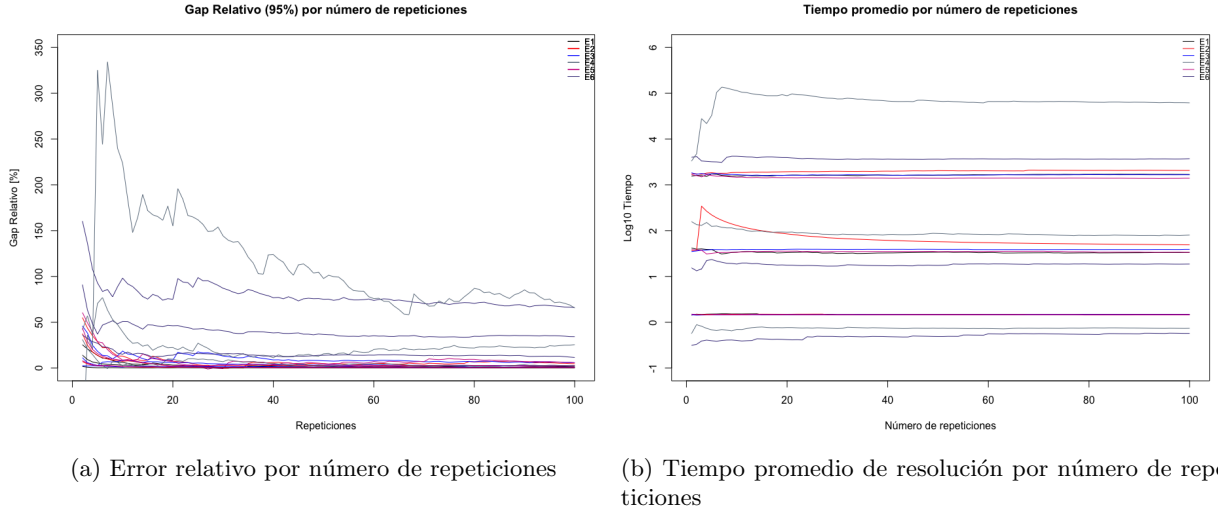


Figura 4.13: Error relativo (izq.) y tiempo promedio de resolución (der.) por número de repeticiones

al tiempo de resolución, se observa que, a medida que aumenta el número de repeticiones, se tiende a estabilizar mucho más rápido y entorno a 40 y se observan las 3 bandas para cada uno de los tamaños *in-sample*. También se observa que el experimento 4 tiene un salto mayor en el tiempo de resolución como se pudo concluir en el gráfico 4.12 de tiempos de resolución por tamaño *in-sample* para $N = 1000$.

En el error relativo del experimento 4, se puede observar saltos en el valor a medida que aumenta el número de repeticiones, lo que se atribuye a incorporar nuevas soluciones y nuevos candidatos a solución, y que justamente es este experimento el que más variadas soluciones plantea, las que se detallarán más adelante en el análisis de las soluciones y en el cálculo del tamaño *out-sample*.

Tamaño *out-sample* (n') y candidatos a solución

Este parámetro (n') se asocia a la evaluación de las soluciones candidatas de las repeticiones, escogiendo el candidato con menor error (absoluto o relativo) para generar la cota superior. La elección del tamaño *out-sample* y el cálculo de la cota inferior implícitamente busca aproximar numéricamente la integral asociada al problema subyacente con el candidato a solución, eligiendo el que al ser evaluado tenga menor error.

Este parámetro afecta directamente el valor de la cota inferior se reduce la varianza y, a su vez, por los distintos valores que puede tomar dependiendo del candidato a solución que sea electo. En

base a esto y para poder tomar una mejor decisión sobre este parámetros, en el gráfico 4.14 log-log, se observa el error relativo de las soluciones a medida que aumenta el tamaño *out-sample*.

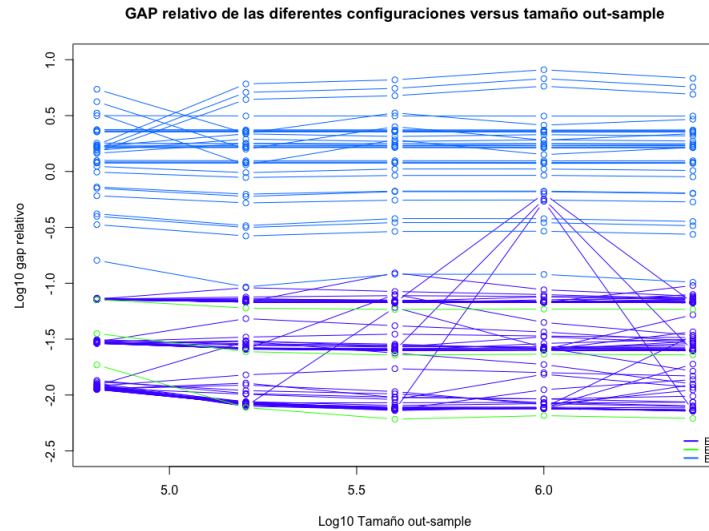


Figura 4.14: Error relativo de las diferentes configuraciones versus tamaño out-sample para los experimentos 4,5 y 6

Se puede apreciar de la figura 4.14, que es bastante estable a partir de un tamaño *out-sample* de $1 \cdot 10^6$. No obstante, se observan un par de puntos que se escapan del rango para el experimento 4, por lo que se revisará más en detalle los candidatos a solución, debido a que se pueden presentar errores numéricos. Adicionalmente, se hace necesario incorporar la información sobre los tiempos de de cálculo de la cota inferior versus el tamaño *out-sample* que se muestra en el gráfico 4.15. De este gráfico se puede ver que siguen un comportamiento lineal en el gráfico log-log, y que el experimento 6 tiene dominado a los otros dos. Esto se debe, en parte, a que se tiene en consideración en el tiempo la generación de números aleatorios que crece exponencialmente en tiempo a medida que aumenta el tamaño *out-sample*, al igual que el cálculo de la cota inferior.

De los resultados anteriores se propone que un valor de $1 \cdot 10^6$, para el tamaño *out-sample*, es razonable y suficiente, teniendo en consideración que el tiempo de cómputo aumenta y no se obtiene una mayor reducción en la convergencia estocástica, la que ya con este valor se encuentra dentro de valores aceptables.

Por otro lado los candidatos a solución permitirán dar indicios de los mecanismos o estrategias de asignación de productos, lo que se detallará a continuación. Para los experimentos 1,2 y 3 se

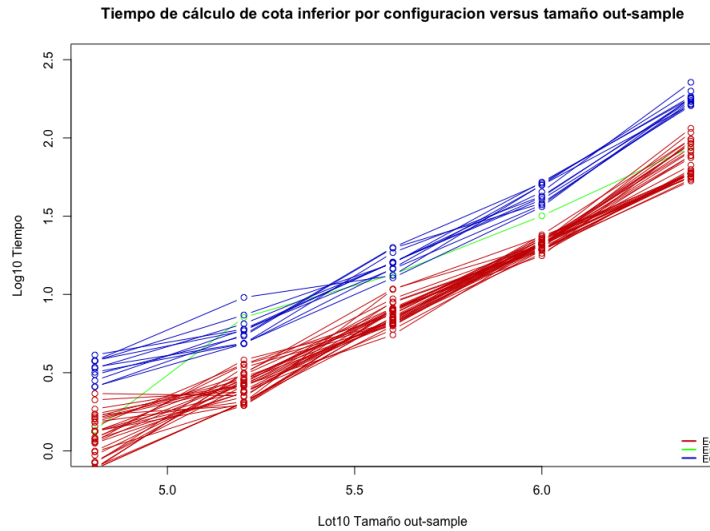


Figura 4.15: Evolución tiempo promedio de cálculo de cota inferior para experimentos 4,5 y 6, versus tamaño out-sample

tuvo un sólo tipo de crédito para todas las repeticiones y para todos los tamaños *in-sample*, que consiste en el entregar el crédito 21 en el periodo 2 (primer periodo posible de entrega de productos). Este crédito es el de mayor Monto (\$ 5 MM) a un plazo de 36 meses. Estos resultados fueron los gatillantes a realizar los experimentos siguientes, debido a que las rentabilidades al ser siempre positivas y excesivamente altas, deja en evidencia que la solución sería entregar el mayor crédito dado que el cliente siempre podrá pagar. Es por esto que se detallarán los candidatos a solución y el análisis del tamaño *out-sample* para los experimentos 4, 5 y 6.

Los productos se detallaron en el capítulo 3 en la tabla 3.1, por lo que se les nombrará bajo su índice $j \in J = 1, \dots, 21$ según esta misma tabla.

Candidatos a solución experimento 4: Normal Grupo 1 con baja rentabilidad desplazada

Para este grupo, al revisar los candidatos a solución se observó que sólo se le entrega el producto $j = 1$ en distintos periodos, lo que se puede observar en el gráfico 4.16.

De este gráfico no se puede establecer algún periodo predominante, no obstante cuando el tamaño *in-sample* es de 1000, hay un *peak* de entrega de producto entre el periodo 12 y 13. Cuando el tamaño *in-sample* es menor, hay una leve tendencia a entregar los productos al comienzo o al final. Esto se puede explicar en que se puede separar clientes entre los que les fue “bien” y los que les fue “mal” respectivamente, en particular, frente a valores atípicos se les entrega en los extremos: a clien-

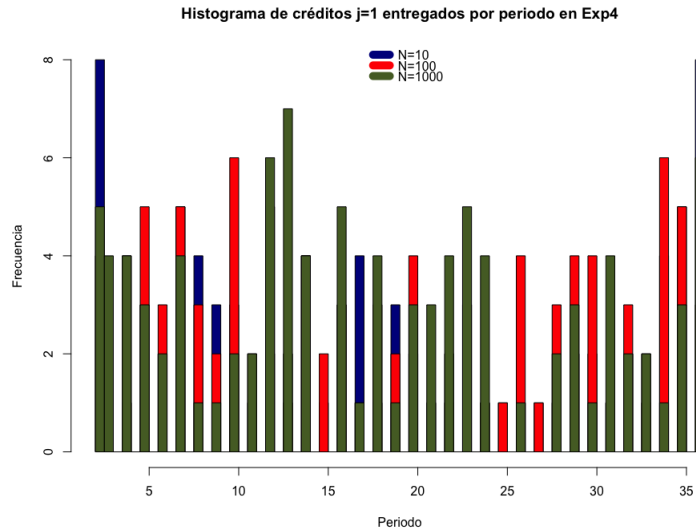


Figura 4.16: Histograma de crédito $j=1$ entregados por periodo

tes que les va “muy bien” se le entrega al inicio y a clientes que les va “muy mal” se les entrega al final.

Para escoger el mejor candidato, se suele buscar la solución que maximiza la cota inferior, es decir, el que minimiza el valor del error estocástico, en este caso se tiene un resultado interesante de analizar, debido a que el valor de la cota y el error es bastante similar en casi todos los periodos, salvo en un rango entre el periodo 10 y 20. En caso de ser tomar la primera regla de decisión, se debería escoger el periodo 2 que genera un error relativo de 0.72%, no obstante, al ser tan similares lo que puede estar ocurriendo es que las diferencias son producidas por errores numéricos que no permiten tener una distinción clara de cuál podría ser el mejor candidato y en particular esto se puede asociar a los mecanismos de decisión, en que a este grupo, es indiferente en qué periodo se le entrega el producto 1, al principio de la ventana de tiempo o al final. No obstante en los periodos intermedios se producen algunos fenómenos extraños.

Candidatos a solución experimento 5: Normal Grupo 2 con alta rentabilidad desplazada

De este grupo, al revisar los candidatos a solución se observó que sólo se le entrega el producto $j = 21$ a diferencia del experimento 4, si existe un periodo dominante, independiente del tamaño *in-sample* se le entrega el producto $j = 21$ en el periodo 2, al igual que a los experimentos 1,2 y 3, lo cual se puede explicar en que a este grupo le sigue “yendo bien” con sus rentabilidades posterior a su desplazamiento, por lo tanto al entregarles este producto con mayor monto permite potenciar el negocio de igual forma y al entregar el mayor monto, a mayor plazo al inicio, maximiza

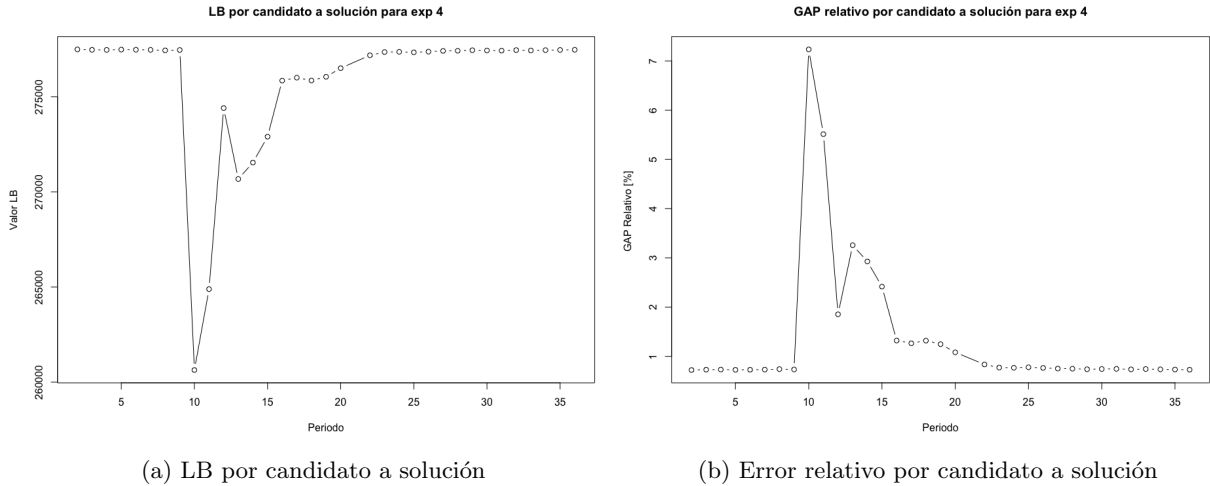


Figura 4.17: LB (izq.) y error relativo (der.) por periodo del candidato a solución para el crédito 1 la acumulación.

Para el cálculo de la cota inferior, se considera el candidato a solución del producto 21 en el periodo 2, al ser el mejor candidato de las dos opciones posibles.

Candidatos a solución experimento 6: Mixtura de Normales desplazada

Este experimento resulta más interesante al generar más productos como candidatos a solución y en mayor diversidad de periodos, lo cual se debe a la mezcla que existe en la distribución resultante. Al revisar los candidatos a solución se observó que se les entrega el producto $j = \{17, 18, 20, 21\}$ en distintos periodos.

La distribución de los resultados se puede ver en la tabla 4.2, donde si bien los candidatos se concentran en el producto 21 en los primeros periodos, a diferencia de los experimentos relacionados al Grupo 2 de clientes, de mayores rentabilidades, estos no sólo se entregan en el primer periodo, es más a medida que aumenta el tamaño in sample se desplaza a periodos mayores hasta llegar a 6, si no que también se muestran como candidatos a solución algunos productos como el 17, 18 y 20 en periodos mucho mayores. Esto se explica debido a que a mayor número de escenarios, se comienzan a observar el grupo menos representado en la mezcla de distribuciones, es decir, el grupo de mayores rentabilidades.

Al igual que en los experimentos anteriores, para escoger el mejor candidato a solución se utiliza

Tabla 4.2: Cantidad de productos entregados por periodo y por tamaño in-sample

Periodo	In-Sample	Producto			
		17	18	20	21
2	N=10	-	-	-	42
	N=100	-	-	-	20
	N=1000	-	-	-	1
3	N=10	-	-	-	14
	N=100	-	-	-	16
	N=1000	-	-	-	4
4	N=10	-	-	-	16
	N=100	-	-	-	8
	N=1000	-	-	-	6
5	N=10	-	-	-	9
	N=100	-	-	-	18
	N=1000	-	-	-	18
6	N=10	-	-	-	14
	N=100	-	-	1	33
	N=1000	-	-	1	66
15	N=10	-	-	-	-
	N=100	-	1	-	-
	N=1000	-	1	-	-
16	N=10	-	1	-	-
	N=100	-	2	-	-
	N=1000	-	1	-	-
17	N=10	-	2	-	-
	N=100	-	1	-	-
	N=1000	-	1	-	-
18	N=10	-	1	-	-
	N=100	-	-	-	-
	N=1000	-	1	-	-
24	N=10	1	-	-	-
	N=100	-	-	-	-
	N=1000	-	-	-	-

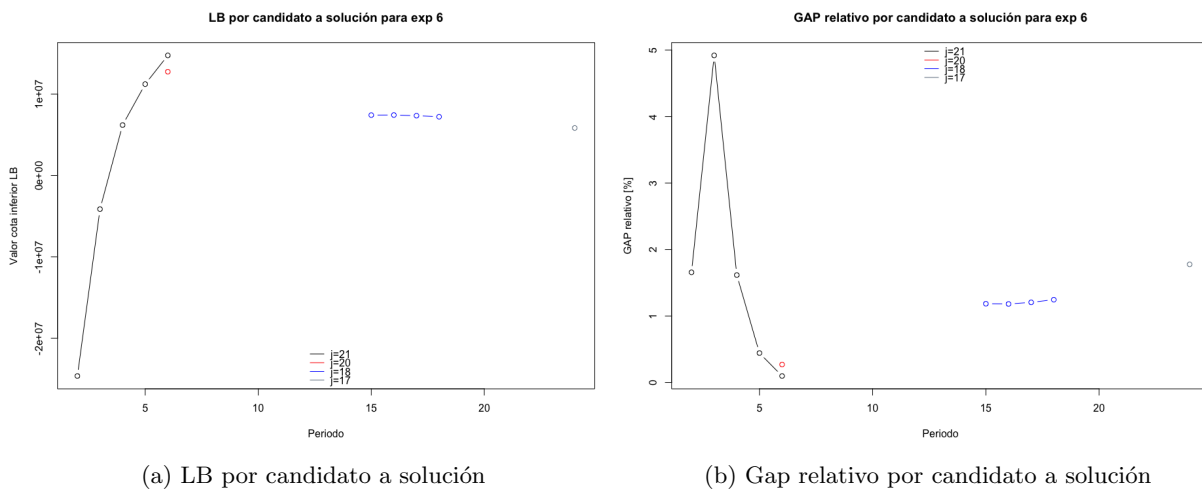


Figura 4.18: LB (izq.) y GAP relativo (der.) por periodo del candidato a solución para experimento 6

el cálculo de la cota inferior. En los gráficos 4.18 se presentan los cálculos de la cota inferior (izq.) y el error relativo (der.) para los distintos candidatos a solución por productos, donde se observa que el producto 21 a medida que se avanza en los periodos disminuye el error relativo y es en el mes 6 donde alcanza su valor mínimo de 9.86 %.

Tamaño in-Sample (n)

El tamaño *in-sample* es fundamental para la obtención de soluciones representativas de la incertidumbre del problema, para este trabajo, el tamaño *in-sample* se puede asociar a la cantidad de trayectorias consideradas en el problema de optimización que se busca modelar, por lo tanto mientras mayor sea, se pueden encontrar soluciones más estables y con mayor convergencia, y por ello representativas.

Al hablar de convergencia estocástica, para poder definir un tamaño razonable, se debe considerar además el tiempo de resolución, debido a que una instancia con un mayor número de escenarios aumenta significativamente el tiempo de resolución por repetición como se ve en el gráfico 4.12.

En la figura 4.19 se puede observar el error relativo por tamaño *in-sample*, donde se observa un comportamiento similar en la tendencia de todos los experimentos, llegando a una convergencia estocástica en torno a un 99% de confianza (error menor a 1%), salvo el experimento 6 que se encuentra en torno a un 90%.

Exp	$N = 10$	$N = 100$	$N = 1000$
1	1.21 %	0.68 %	0.053 %
2	4.67 %	2.07 %	0.59 %
3	5.80 %	2.42 %	0.80 %
4	6.8 %	2.68 %	0.91 %
5	5.79 %	2.25 %	0.58 %
6	65.94 %	34.21 %	11.47 %

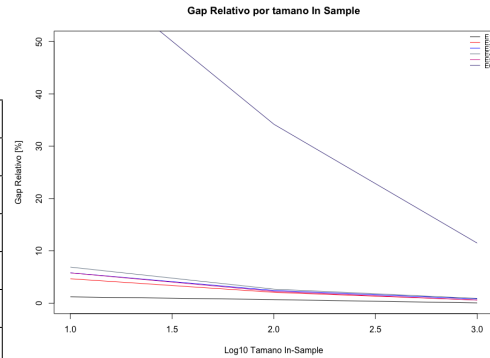


Figura 4.19: Gap relativo experimento y por tamaño in-sample

En la figura 4.19, se tienen la tabla (izq.) con los valores del error relativo por experimento y tamaño in sample en porcentajes y su gráfico (der.). En todos los experimentos se observa un decaimiento importante en el valor del error y representan resultados estadísticamente significativos a un 99% de confianza en la convergencia estocástica para todos los experimentos, con un tamaño de 1.000 escenarios, salvo para el experimento 6 que disminuye el nivel de confianza.

Que el experimento 6, presente un mayor valor de su error estocástico merece un mayor análisis,

lo que significó realizar otra serie de repeticiones con un tamaño *in-sample* de 2000, para observar su comportamiento. Los resultados de este análisis significaron una disminución del error relativo a 7,9%. Que este, presente un mayor error se explica en que al ser una mezcla de dos grupos de distribución, existan saltos en las trayectorias de grupos los cuales presentan grandes diferencias, como a su vez, el desplazamiento de las distribuciones normales es significativo y presenta una mayor varianza en la solución final. En el análisis de resultados, a continuación, se detallarán más estas diferencias y los resultados de los posibles candidatos a solución óptima.

Capítulo 5

Extensiones Futuras

5.1. Modelo generalizado

Como se explicó en el capítulo 2, el modelo planteado se puede extender y generalizar a partir de tres de sus principales componentes. Por un lado la función objetivo, que en este trabajo se consideró como el *Equity* neto del cliente al final del periodo, considerando así el rol del BancoEstado de maximizar el bienestar de sus clientes, podría modificarse de forma de aumentar el *retorno sobre patrimonio* o “*return on equity*” (ROE) de la institución bancaria, manteniendo la misma estructura general.

Incorporar medidas de riesgo distintas para la restricción de riesgo o provisiones que considera el banco, en caso de mantener la misma función objetivo 2.33.

Si se considera otro forma de modelar la interacción del pago del cliente, también se puede modificar esto, cambiando la restricción 2.37.

5.1.1. Función objetivo

Por ejemplo, se puede hacer un análisis de la rentabilidad del banco, considerando la esperanza del valor presente de los pagos recibidos del cliente menos la colocación entregada.

$$\text{máx } \mathbb{E}_s \left[\sum_t \sum_j \delta^t (P_s^t - M_j X_j^t) \right] \quad (5.1)$$

Esta función es interesante, dado que es posible separarla en una parte estocástica y otra determinista, donde queda sujeto a la esperanza la variable pago del cliente, que efectivamente dependerá de

las trayectorias de la variable aleatoria. A su vez, se podría modificar e incorporar otra medida de riesgo, como lo sería el *Value at Risk* (*VaR*) o *Conditional Value at Risk* (*CVaR*), para hacer un mejor análisis del riesgo implicado y considerar los peores casos posibles.

$$\text{máx} \left[\mathbb{E}_s \left[\sum_t \sum_j \gamma^t P_s^t \right] - \sum_t \sum_j \gamma^t M_j X_j^t \right] \quad (5.2)$$

Se puede observar que si este fuese un problema determinista, en el cuál el cliente siempre pueda cumplir con su obligación de pago, se reduce el problema en la selección del producto con mayor rentabilidad para el banco, es decir mayor Rb_j .

5.2. Mayor refinamiento en modelación de flujos de agentes

Con el objetivo de capturar otros efectos en la interacción de los agentes, se propone como futura extensión un mayor refinamiento en el modelamiento de los flujos de los agentes, como por ejemplo efectos de liquidez, impositivos o pago de proveedores en el fenómeno de *default*, por ejemplo con el desfase de pago que hacen los supermercados.

Por un lado, en los alcances de este trabajo se dejó fuera del análisis los efectos impositivos, debido a la falta de datos para su estimación, como se explica en la literatura [Alfaro-Castillo and Castillo-Vergara, 2013], si tienen un efecto al igual que los distintos pagos de proveedores en la liquidez de la operación de una microempresa.

Es precisamente por liquidez, que las microempresas, terminan incumpliendo pagos y entender mejor este efecto podría permitir refinar el mecanismo de default que se plantea metodológicamente con este trabajo.

5.3. Variables exógenas en la generación de escenarios

La generación de escenarios en este trabajo se basó en la estimación de la rentabilidad de los clientes, aproximada como el cuociente entre la capacidad de pago y las ventas totales, como se explica en el capítulo 3, en la ecuación 3.4, obtenidos a partir de las evaluaciones de los clientes.

Sería entonces interesante analizar, posibles efectos exógenos a la generación de estos números aleatorios. Por ejemplo, incorporar el efecto del ciclo económico como un shock aleatorio en la distribución de rentabilidad, o el efecto de una crisis y cómo afectan los retornos de las microempresas.

A modo de ejemplo, se podría definir una variable exógena como un proceso de *Poisson* con una cierta tasa de ocurrencia λ , que genera shocks en la rentabilidad de los clientes, por ejemplo una crisis o un efecto macroeconómico. Un trabajo a futuro entonces será, definir una variable y como ésta afecta la rentabilidad con shocks aleatorios.

5.4. Algoritmo de descomposición para la resolución del problema

Como se observó en el capítulo 4, el problema de optimización estocástica presentado en cuanto a sus tiempos de resolución es de carácter exponencial a medida que se aumentan los escenarios. Si bien presenta una dificultad la dimensionalidad del problema, se plantea abordar su resolución mediante algún algoritmo de descomposición estilo Benders [Benders, 1962].

5.5. Modelo de optimización dinámica estocástica para la resolución del problema

Otra forma de abordar la resolución del problema, es a través de un modelo de optimización dinámica estocástica para la resolución del problema.

5.6. Modelo de valorización en dos etapas con opciones reales

Para la justificación conceptual del modelo general de acumulación de capital presentado en este trabajo, se planteó un modelo de valorización en dos etapas con opciones reales, con el objetivo de contrastar teóricamente los resultados. Se describirá este último modelo a continuación, los supuestos y justificaciones metodológicas que tiene detrás, dejando propuesto su posible implementación y desarrollo futuro.

Se consideran dos etapas, la primera t_0 , en la cual se decide si entregar o no el producto y la segunda etapa T , en la que se realiza el pago y a su vez la “liquidación” de la microempresa. Como se detalló anteriormente, siguiendo el balance de la empresa considerado en la figura 2.2 del capítulo

2, se tiene que al realizar esta liquidación del negocio, lo que se está modelando es el patrimonio, cuyo valor se busca maximizar en valor esperado. Para este modelo se realizarán simulaciones de la variable aleatoria, rentabilidad del negocio, que permitirán ir armando árboles de trayectorias posibles del patrimonio, en 2 etapas. Definiendo el default cuando el patrimonio baja del valor 0, es decir, cuando la deuda supera a los activos, según el modelo de [Merton, 1974].

Algunas consideraciones que se tienen es que la rentabilidad del negocio, se puede asociar a la veocidad de crecimiento del activo y la deuda, al igual que el modelo global, se considera como la suma de todos los pagos o cuotas que deberá pagar el microempresario al momento de acceder a un producto, estas cuotas consideran tanto interés como amortización. Sin embargo, el pago no es relevante para esta forma de modelar, dado que se cancelan al momento de la liquidación, por lo tanto nos interesa sólo la deuda total que contrae el cliente.

5.6.1. Descripción de variables

$$X_j : \text{Se asigna producto } j \text{ al cliente} \quad (5.3)$$

$$A : \text{Activo del cliente} \quad (5.4)$$

$$D : \text{Deuda del cliente} \quad (5.5)$$

$$E_t : \text{Equity del cliente } (A_t - D_t) \quad (5.6)$$

$$M : \text{Mora del cliente} \quad (5.7)$$

$$I : \text{Variable auxiliar para definir Ingreso por los productos} \quad (5.8)$$

5.6.2. Parámetros

$$\tilde{r} : \text{Rentabilidad del negocio del cliente} \quad (5.9)$$

$$rb_j : \text{Rentabilidad del producto } j \quad (5.10)$$

$$\beta_j : \text{Costo financiero de la deuda del producto } j \quad (5.11)$$

$$S_j : \text{Monto del producto } j \quad (5.12)$$

$$c_j : \text{cuota producto } j \quad (5.13)$$

5.6.3. Parámetros de estado

$$A_0 : \text{Caja inicial del cliente} \quad (5.14)$$

$$D_0 : \text{Deuda inicial del cliente} \quad (5.15)$$

$$E_0 : \text{Patrimonio inicial del cliente} \quad (5.16)$$

Dado esto, se tiene que las variables se definen de la siguiente forma:

$$A_t = (A_{t-1} + \sum_j X_{j,t-1} S_j) \tilde{R} \quad (5.17)$$

$$D_t = (D_{t-1} + \sum_j S_j R b_j X_j) \quad (5.18)$$

$$E_t = (A_{t-1} + \sum_j X_{j,t-1} S_j) \tilde{R} - (D_{t-1} + \sum_j S_j R b_j X_j) \quad (5.19)$$

Una consideración importante es que al realizar estas trayectorias del patrimonio de la micro-empresa, como un método de árbol binomial, cuando se quiera obtener el valor esperado de esta última, se tendrá que considerar una tasa de descuento para obtener el valor presente esperado del patrimonio.

Como el valor de esta tasas de descuento son difícil de calcular, se puede utilizar una metodología de valorización neutra al riesgo, cambiando el espacio de probabilidades del árbol de simulación periodo a periodo, calculando la martingala equivalente. En otras palabras, como se ve en la ecuación 5.20, se iguala el valor esperado del patrimonio calculado en base a las probabilidades neutras al riesgo, traído a valor presente con la tasa libre de riesgo, con el valor esperado del valor presente del patrimonio. Para ayudar a entender este método de forma gráfica este árbol de simulación de escenarios se puede observar en la figura 5.1. Cada nodo del árbol es condicional a la variable $X_{j,t}$, en donde se le pudo o no, haber dado el producto j .

$$\frac{\mathbb{E}_{\pi^*}(E_t)}{1 + r_f} = \mathbb{E}_R(VP(E_t)) \quad (5.20)$$

Para ilustrar esto último, se presenta un ejemplo simple de un árbol binomial de un periodo, en donde se considerará π^* como la probabilidad neutra al riesgo de un escenario positivo de rentabilidad (subir en el árbol) y $(1 - \pi^*)$ como la probabilidad de bajar. Se tendrá una caja inicial $A_0 = 1$, deuda inicial $D_0 = 0$, dos realizaciones de la variable aleatoria $\tilde{r}^1 = 0.5$ y $\tilde{r}^1 = 0.1$, se considera un

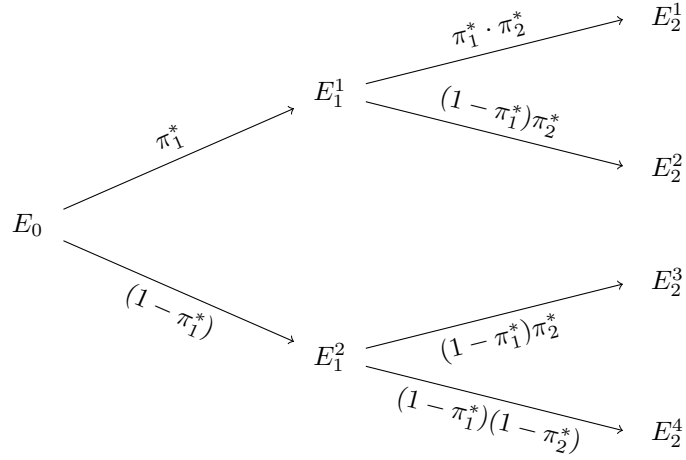


Figura 5.1: Representación del árbol de simulación

conjunto de producto unitario $J = \{1\}$, es decir se tiene un solo producto cuya tasa es de $rb_1 = 0.14$ y cuyo monto es de $S_1 = 1$ y finalmente una tasa libre de riesgo de $r_f = 0.1$.

Considerando el árbol binomial en un periodo cuando el valor de la variable $X_{1,1} = 1$ se pueden observar en la figura 5.2 los cálculos de ambos nodos para luego así, poder obtener las probabilidades neutras al riesgo (π^*), y de los factores de “alza” (u) y “descenso” (d) de patrimonio dados por las fórmulas 5.21, 5.22 y 5.23 respectivamente.

$$\frac{(1 + r_f) - d}{u - d} \tag{5.21}$$

$$u = \frac{E_1^1}{E_0} \tag{5.22}$$

$$d = \frac{E_1^2}{E_0} \tag{5.23}$$

Finalmente, reemplazando los valores de este ejemplo, se tiene el árbol generado en la figura 5.2, donde se observan los valores de la probabilidad y los patrimonios resultantes, lo que en definitiva permitirá calcular el valor esperado mediante la valorización neutra al riesgo.

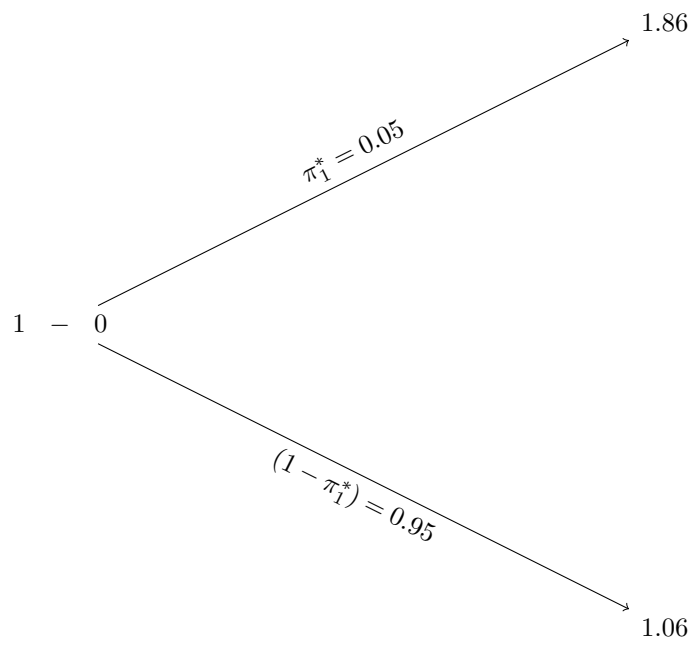


Figura 5.2: Representación del árbol de simulación con valores

Capítulo 6

Conclusiones y recomendaciones

6.1. Conclusiones

Con respecto al Modelo en sí, se considera satisfactorio e interesante el hecho de poder modelar un comportamiento no lineal como la opción de pago de un cliente en un problema de optimización MIP estocástico; el poder interpretar las restricciones de linealización como un portafolio réplica de opciones lineales y, finalmente, uno de los resultados más importantes es al incorporar la restricción de riesgo de parte del banco, y entenderlo como un problema general del regulador, se logra capturar la dinámica de pérdida esperada y de pago mínimo de agente, las cuales actúan como un conjunto de restricciones degeneradas.

Deja muchas posibilidades de extensiones al ser posible la generalización e incorporación de otro tipo de aproximaciones al comportamiento de pago, medidas de riesgo tanto para la función objetivo o restricciones presupuestarias y otro tipo de función objetivo, como por ejemplo maximizar el ROE del banco.

Se tienen perfiles de estrategia básicos para ejemplos de rentabilidad, pero esto no puede ser aplicado directamente, debido a que los desplazamientos en distribución no poseen un respaldo mayor debido a la falta de información.

En lo relativo al trabajo con los datos de BancoEstado Microempresas se logra caracterizar de manera exitosa los productos y las rentabilidades; se logra incorporar una buena aproximación de rentabilidad a partir de variables contables de las microempresas. No obstante esto requirió ajustes sobre la marcha debido a la ausencia de datos, estructura de las bases de datos y sesgos que no se habían considerado en el comienzo del trabajo, de los cuales se pudo hacer cargo e incorporar, pero deja planteada algunas preguntas a futuro para la recopilación de los datos y quedan como

recomendaciones para el Banco.

La metodología utilizada para la segmentación y caracterización de los productos permitió discretizar la oferta de créditos para ser incorporada al modelo, logrando un buen acercamiento al reflejar lo que actualmente entrega el Banco. Por otro lado, la metodología de modelos de mixtura finitos para la estimación de los parámetros de las distribuciones de los grupos encontrados también posibilitó incorporar subsegmentación a un segmento de clientes bastante grande, lo que permitiría tomar decisiones de gestión al banco con mayor información.

El proceso iterativo de modelación de la interacción de agentes, si bien parte como un problema muy ambicioso y complejo, que no era posible de resolver con los datos entregados, permitió un entendimiento del problema y un refinamiento hasta llegar al modelo general planteado, el cual presenta una metodología robusta para incorporar comportamiento de pago. Este proceso también permitió plantear problemas similares que se dejan propuestos como el modelo de valorización en dos etapas con opciones reales descrito en parte en el capítulo 5.

Gracias a la metodología SAA, utilizada para la resolución del problema estocástico planteado, se logró tener certeza estadística significativa para las soluciones presentadas. No obstante, también se pudo corroborar una de las críticas que se le hace a este método, como es el tener errores numéricos que hacen no se tenga mucha diferencia entre algunos candidatos a solución. Esto se pudo observar en el experimento 4, donde no se podía asegurar con certeza cuál era la solución que producía el menor “error” relativo, debido a que las diferencias eran muy pequeñas y se pudo deber a errores numéricos o aproximaciones. No obstante, se puede concluir a grandes rasgos para ese experimento que hay intervalos donde no es conveniente o es incierto lo que puede pasar, y otros donde es indiferente en cuál entregar el producto.

A partir de los resultados obtenidos, se permite entregar recomendaciones para la recopilación de información por parte del Banco, pues las evaluaciones de clientes no están capturando todas las variables relevantes en la rentabilidad del cliente o tienen un sesgo de entrada muy alto. Por otro lado se logra plantear una metodología para encontrar estrategias de entrega de productos a partir de la segmentación en grupos de clientes a partir de rentabilidad esperada y no de la capacidad de pago, como se hace actualmente.

En resumen se concluye que es un trabajo satisfactorio al poder plantear un problema de gran envergadura con un análisis sistémico que requirió de muchas iteraciones, pero generó aprendizaje y oportunidades de seguir desarrollando la teoría y a su vez se logra simplificar y abordar de una forma más concreta y aplicada, resolviendo el problema original planteado.

Respecto al segmento de clientes que se enmarca este trabajo, la microempresa en Chile, este trabajo permitió desarrollar un modelo que incorpora la noción de bienestar del microempresario en la función objetivo a maximizar, no sólo maximizar el beneficio del Banco.

6.2. Recomendaciones

Se recomienda a BancoEstado Microempresas mejorar la estructura de sus datos, en particular, la base de datos utilizada para las evaluaciones de clientes, debido a que esta debió ser pre-procesada para que fuese posible su análisis. Debiese estar normalizada y estructurada, de tal manera que permita un manejo sencillo y confiable de la información.

Por otro lado, se debe caracterizar mejor a los clientes en base a su rentabilidad, para que la estimación sea concordante con los resultados observados posteriormente a la entrega de sus productos. Esto requerirá identificar algunas variables que afecten directamente el *default*. Se debe considerar de todas formas, que la estimación de cola de pérdidas debe estar ajustada a la realidad de pérdida de sus clientes y no a una distribución poblacional, debido a que incorporar, por ejemplo, los clientes rechazados buscando eliminar el sesgo de selección, generaría un desbalance en las distribuciones de ganancia lo que sesgaría el modelo de pérdida.

Realizar un mayor seguimiento de clientes que sean representativos para poder entender mejor las variables contables de las evaluaciones, permitiendo tener la dinámica de rentabilidad por periodo y así poder incorporar esa información a las distribuciones de rentabilidad del modelo.

Bibliografía

- [Alfaro-Castillo and Castillo-Vergara, 2013] Alfaro-Castillo, M. and Castillo-Vergara, M. (2013). Financial-accounting variables analysis affecting the closure of smes in early stagesness. *Revista Universitaria Ruta*, 15(1):126–138.
- [Armendáriz and Morduch, 2010] Armendáriz, B. and Morduch, J. (2010). *The economics of microfinance*. MIT press.
- [Benaglia et al., 2009] Benaglia, T., Chauveau, D., Hunter, D. R., and Young, D. S. (2009). mixtools: An R Package for Analyzing Finite Mixture Models. *Journal of Statistical Software*, 32(6):1–29.
- [Benders, 1962] Benders, J. F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4(1):238–252.
- [Bonabeau, 2002] Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(suppl 3):7280–7287.
- [Dantzig, 1948] Dantzig, G. B. (1948). Programming in a linear structure. Comptroller, USAF Washington D.C.
- [Dempster et al., 1977] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.
- [Fayyad, 1996] Fayyad, U. M. (1996). Data mining and knowledge discovery: making sense out of data. *IEEE Expert*, 11(5):20–25.
- [Hitchcock, 1941] Hitchcock, F. L. (1941). The distribution of a product from several sources to numerous localities. *J. Math. phys*, 20(2):224–230.
- [Kantorovich, 1939] Kantorovich, L. V. (1939). The mathematical method of production planning and organization. *Management Science*, 6:363–422.

- [Keskin et al., 2010] Keskin, H., Şentürk, C., Sungur, O., and Kiriş, H. M. (2010). The importance of smes in developing economies.
- [Khachiyan, 1980] Khachiyan, L. G. (1980). Polynomial algorithms in linear programming. *USSR Computational Mathematics and Mathematical Physics*, 20(1):53–72.
- [Larraín et al., 2007] Larraín, P. et al. (2007). *BancoEstado Microcréditos: lecciones de un modelo exitoso*. CEPAL.
- [Maria and Yuhang, 2004] Maria, V. and Yuhang, X. (2004). Default risk in equity returns. *The Journal of Finance*, 59(2):831–868.
- [Merton, 1974] Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of finance*, 29(2):449–470.
- [Ministerio de economía, 2012] Ministerio de economía, C. (2012). Segunda encuesta de microemprendimiento.
- [Ministerio de economía, 2013] Ministerio de economía, C. (2013). Tercera encuesta de microemprendimiento.
- [Ministerio de economía, 2016] Ministerio de economía, C. (2016). Cuarta encuesta de microemprendimiento.
- [OECD, 2004] OECD (2004). Small and medium-sized enterprises in turkey: issues and policies.
- [OECD, 2017] OECD (2017). Enhancing the contributions of smes in a global and digitalised economy.
- [Schrijver, 1998] Schrijver, A. (1998). *Theory of linear and integer programming*. John Wiley & Sons.
- [SERCOTEC, 2013] SERCOTEC (2013). La situación de la micro y pequeña empresa en Chile.
- [Shapiro and Ruszczyński, 2008] Shapiro, A. and Ruszczyński, A. (2008). Lectures on stochastic programming. *Preprint, Georgia Tech*.
- [Sharpe, 1990] Sharpe, S. A. (1990). Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships. *The journal of finance*, 45(4):1069–1087.

- [Solimano et al., 2007] Solimano, A., Pollack, M., Wainer, U., and Wurgaft, J. (2007). Micro empresas, pyme y desarrollo económico. chile y la experiencia internacional. *Documento de Trabajo Working Paper*, 3.
- [Soyibo et al., 1991] Soyibo, A., Sobodu, O. O., Osayameh, R. O., and Soyibo, A. (1991). A multi-objective lending and credit administration model for commercial banks in a developing country. *Savings and Development*, 15(2):167–186.

Apéndices

A . Primeras iteraciones de la formulación del modelo con creencia de tasa del cliente

A .1. Diagrama de negociación, Modelo con aceptación del producto por parte del cliente

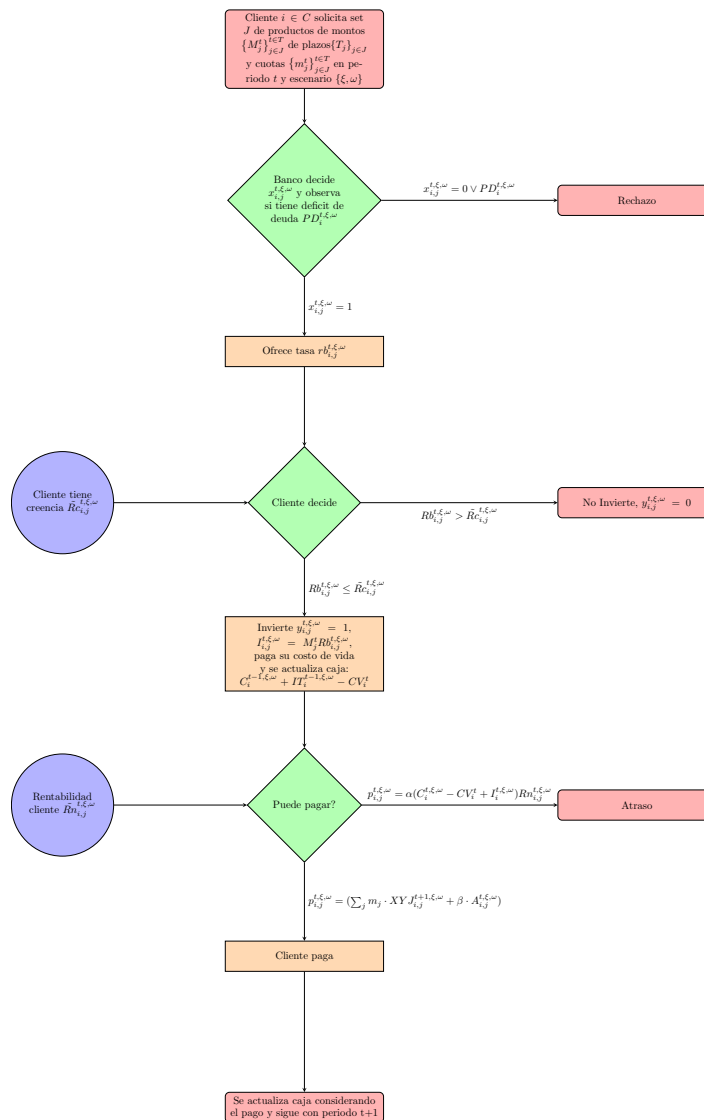


Figura 6.1: Flujo primera iteración modelo con aceptación de producto

A .2. Formulación del modelo

Conjuntos

$$C = \{i \in 1 \dots Clientes\} \quad \text{Conjunto de Clientes} \quad (6.1)$$

$$J = \{j \in 1 \dots Productos\} \quad \text{Conjunto de Productos ofrecidos} \quad (6.2)$$

$$T = \{t \in 1 \dots Periodos\} \quad \text{Conjunto de Periodos considerados} \quad (6.3)$$

$$E = \{e \in 1 \dots Estados\} \quad \text{Conjunto de Estados de cliente considerados} \quad (6.4)$$

Escenarios Aleatorios

$$S = \{\xi \in 1 \dots N\} \quad \text{Escenarios de realizacion de creencia de tasa ofrecida por banco} \quad (6.5)$$

$$\Omega = \{\omega \in 1 \dots M\} \quad \text{Escenarios de realización de rentabilidad del negocio del cliente} \quad (6.6)$$

Variables

$$x_{i,j}^t: \text{Banco ofrece producto } j \text{ al Cliente } i, \text{ en el periodo } t \quad (6.7)$$

$$Rb_{i,j}^t = 1 + r_{bi,j}^t: \text{Tasa ofrecida por el Banco para producto } j \text{ al Cliente } i, \text{ en el periodo } t \quad (6.8)$$

$$y_{i,j}^{t,\xi}: \text{Cliente } i \text{ invierte en el producto } j, \text{ en el periodo } t, \text{ en escenario } \xi \quad (6.9)$$

$$Y_i^{t,\xi}: \text{Cliente } i \text{ invierte, en el periodo } t, \text{ en escenario } \xi \quad (6.10)$$

$$I_{i,j}^{t,\xi,\omega}: \text{Inversión del cliente } i, \text{ del producto } j, \text{ en el periodo } t, \text{ en escenario } \xi, \omega \quad (6.11)$$

$$IT_i^{t,\xi}: \text{Inversión total del cliente } i, \text{ en el periodo } t, \text{ en escenario } \xi, \omega \quad (6.12)$$

$$P_{i,j}^{t,\xi,\omega}: \text{Pago del cliente } i, \text{ por producto } j, \text{ en el periodo } t, \text{ en escenario } \xi, \omega \quad (6.13)$$

$$PT_i^{t,\xi,\omega}: \text{Pago total del cliente } i, \text{ en el periodo } t, \text{ en escenario } \xi, \omega \quad (6.14)$$

$$D_i^{t,\xi,\omega}: \text{Cliente } i \text{ hace default, en el periodo } t, \text{ en escenario } \xi, \omega \quad (6.15)$$

$$C_i^{1,\xi,\omega}: \text{Caja acumulada por el cliente } i, \text{ en el periodo } t, \text{ en escenario } \xi, \omega \quad (6.16)$$

$$Z_{i,e}: \text{Variable de estado } e \text{ del cliente } i \quad (6.17)$$

$$O_{i,j}^{t,\xi,\omega}: \text{Pérdida por cliente } i, \text{ del producto } j, \text{ en el periodo } t, \text{ en escenario } \xi, \omega \quad (6.18)$$

$$u_{i,j}^{t,\xi,\omega}: \text{Costo oportunidad por cliente } i, \text{ del producto } j, \text{ en el periodo } t, \text{ en escenario } \xi, \omega \quad (6.19)$$

Función Objetivo

$$\text{Max} \quad \frac{1}{N \cdot M} \sum_{i,\xi,\omega} C_i^{T+1,\xi,\omega} \quad (6.20)$$

$$\text{Max} \quad \mathbf{E}\left(\sum_i \delta_t^t \cdot C_i^{T+1}\right) \quad (6.21)$$

$$(6.22)$$

Restricciones

s.t:

$$C_i^{1,\xi,\omega} = C0_i^{\xi,\omega} \cdot \delta_1 + IT_i^{1,\xi,\omega} \cdot \delta_1 \cdot Rn_i^{1,\xi,\omega} - PT_{i,j}^{1,\xi,\omega} - \sum_j \alpha_{i,j} O_{i,j}^{1,\xi,\omega} \quad \forall_{i,j,\xi} \quad (6.23)$$

$$D_i^{T,\xi,\omega} = 0 \quad \forall_{i,j,\xi,\omega} \quad (6.24)$$

$$IT_i^{t,\xi,\omega} = \sum_j I_{i,j}^{t,\xi,\omega} \quad \forall_{i,t,\xi,\omega} \quad (6.25)$$

$$PT_i^{t,\xi,\omega} = \sum_j P_{i,j}^{t,\xi,\omega} \quad \forall_{i,t,\xi,\omega} \quad (6.26)$$

$$y_{i,j}^{t,\xi,\omega} \leq Y_i^{t,\xi} \quad \forall_{i,j,t,\xi} \quad (6.27)$$

$$\epsilon_{i,j}^{t,\xi} \cdot y_{i,j}^{t,\xi} \leq 0, \quad \forall_{i,j,t,\xi} \quad (6.28)$$

$$y_{i,j}^{t,\xi} \leq x_{i,j,t} \quad \forall_{i,j,t,\xi} \quad (6.29)$$

$$I_{i,j}^{t,\xi} \leq M_j^t \cdot y_{i,j}^{t,\xi} \quad \forall_{i,j,t,\xi} \quad (6.30)$$

$$O_{i,j}^{t,\xi,\omega} \leq I_{i,j}^{t,\xi} + (Rb_{i,j}^t - 1) \cdot M_j^t \quad \forall_{i,j,t,\xi,\omega} \quad (6.31)$$

$$O_{i,j}^{t,\xi,\omega} \leq M_1 \cdot y_{i,j}^{t,\xi,\omega} \quad (6.32)$$

$$u_{i,j}^{t,\xi,\omega} \leq M_2 \cdot (1 - y_{i,j}^{t,\xi}) \quad \forall_{i,j,t,\xi,\omega} \quad (6.33)$$

$$u_{i,j}^{t,\xi,\omega} \leq M_3 \cdot (1 - D_{i,j}^{t,\xi,\omega}) \quad \forall_{i,j,t,\xi,\omega} \quad (6.34)$$

$$O_{i,j}^{t,\xi,\omega} \leq M_j^t \cdot Rb_{i,j}^t \quad (6.35)$$

$$u_{i,j}^{t,\xi,\omega} \leq M_j^t \cdot Rb_{i,j}^t \quad \forall_{i,j,t,\xi,\omega} \quad (6.36)$$

$$\sum_{i \in C} x_{i,j}^t \cdot M_j^t \leq F_j^t \quad \forall_{j,t} \quad (6.37)$$

$$\sum_{j \in J} x_{i,j}^t \leq K \quad \forall_{i,t} \quad (6.38)$$

$$R_f + \sum_e \beta_e \cdot Z_{i,e} \leq Rb_{i,j}^t \quad \forall_{i,j,t} \quad (6.39)$$

$$M_j^t \cdot Rb_{i,j}^t - O_{i,j}^{t,\xi,\omega} - u_{i,j}^{t,\xi,\omega} \leq P_{i,j}^{t,\xi,\omega} \quad \forall_{i,j,t,\xi,\omega} \quad (6.40)$$

$$O_{i,j}^{t,\xi,\omega} \leq M_4 \cdot (1 + D_i^{t,\xi,\omega} - Y_i^{t,\xi,\omega}) \quad \forall_{i,j,t,\xi,\omega} \quad (6.41)$$

$$P_{i,j}^{t,\xi,\omega} \leq M_5 \cdot (1 - D_i^{t,\xi,\omega}) \quad \forall_{i,t,\xi,\omega} \quad (6.42)$$

$$P_{i,j}^{t,\xi,\omega} \leq I_{i,j}^{t,\xi} + (Rb_{i,j}^t - 1) \cdot M_j^t \quad \forall_{i,j,t,\xi,\omega} \quad (6.43)$$

$$P_{i,j}^{t,\xi,\omega} \leq y_{i,j}^{t,\xi} \cdot M_6 \quad \forall_{i,j,t,\xi,\omega} \quad (6.44)$$

$$P_{i,j}^{t,\xi,\omega} \leq M_j^t \cdot Rb_{i,j}^t \quad \forall_{i,j,t,\xi,\omega} \quad (6.45)$$

$$\frac{1}{N \cdot M} \sum_{i,\xi,\omega} D_i^{t,\xi,\omega} \leq PI \quad \forall_t \quad (6.46)$$

$$\mathbf{P}(\sum_i D_i^t) \leq PI \quad \forall_t \quad (6.47)$$

$$D_i^{t,\xi,\omega} \leq y_{i,j}^{t,\xi} \quad \forall_{i,j,t,\xi,\omega} \quad (6.48)$$

$$(C_i^{t,\xi,\omega} + IT_i^{t,\xi,\omega}) \cdot \delta_t^t \cdot Rn_i^{t,\xi,\omega} - \sum_j M_j^t \cdot Rb_{i,j}^t + M_7 \cdot (1 - Y_i^{t,\xi} + D_i^{t,\xi,\omega}) \geq 0 \quad \forall_{i,t,\xi,\omega} \quad (6.49)$$

$$(1 - D_i^{t,\xi,\omega}) \geq x_{i,j}^{t+k} \quad \forall_{i,t,\xi,k=1..T-t} \quad (6.50)$$

$$C_i^{t,\xi,\omega} = (C_i^{t-1,\xi,\omega} + IT_i^{t-1,\xi,\omega}) \cdot \delta_t^{t-1} \cdot Rn_i^{t-1,\xi,\omega} - PT_{i,j}^{t-1,\xi,\omega} - \sum_j \alpha_{i,j} O_{i,j}^{t,\xi,\omega} \quad \forall_{i,t>1,\xi,\omega} \quad (6.51)$$

$$\frac{1}{N \cdot M \cdot F_t} \sum_{i,\xi,\omega} P_{i,j}^{t,\xi,\omega} + \alpha_{i,j} \gamma_{i,j} O_{i,j}^{t,\xi,\omega} \geq R_t \quad \forall_{i,t,\xi,\omega} \quad (6.52)$$

$$\mathbf{E}(\frac{\sum_i P_{i,j}^t + \alpha_{i,j} \gamma_{i,j} O_{i,j}^t}{F_t}) \geq R_t \quad \forall_{i,t} \quad (6.53)$$

$$\frac{1}{N \cdot M} \sum_{i,\xi,\omega} O_{i,j}^{t,\xi,\omega} \leq P_t \quad \forall_{i,t,\xi,\omega} \quad (6.54)$$

$$\mathbf{E}(\sum_i O_{i,j}^t) \leq P_t \quad \forall_{i,t} \quad (6.55)$$

Variables aleatorias

$Rc_{i,j}^{t,\xi} = Rb_{i,j}^t + \tilde{\epsilon}_{i,j}^{t,\xi}$: Distorsión creencia de tasa para producto j , del cliente i , en periodo t , en escenario ξ

$$(6.56)$$

$Rn_i^{t,\xi,\omega} = 1 + r_{ni}^{t,\xi,\omega}$: Rentabilidad del negocio del cliente i , en el periodo t , en el escenario ξ, ω

$$(6.57)$$

Parámetros

N : Cantidad de realizaciones de escenarios ξ

$$(6.58)$$

$$M: \text{Cantidad de realizaciones de escenarios } \omega \quad (6.59)$$

$$C0_i: \text{Capital inicial del cliente } i \quad (6.60)$$

$$Rb_{i,j}^t = 1 + r_{bi,j}^t: \text{Tasa ofrecida por el Banco para producto } j \text{ al Cliente } i, \text{ en el periodo } t \quad (6.61)$$

$$M_{j,t}: \text{Monto solicitado por producto } j \text{ en periodo } t \quad (6.62)$$

$$M_k: \text{Gran M} \quad \forall_{k=1..10} \quad (6.63)$$

$$F_j^t: \text{Fondo disponible para ofrecer por producto } j \text{ en periodo } t \quad (6.64)$$

$$K: \text{Cantidad máxima de productos que puede tener cada cliente} \quad (6.65)$$

$$R_{fj} = 1 + r_{fj}: \text{Tasa mínima a la que puede prestar el banco por producto } j \quad (6.66)$$

$$\beta_e: \text{Spread asociado a variables de estado } e \text{ de los clientes} \quad (6.67)$$

$$PI: \text{Probabilidad de incumplimiento permitida} \quad (6.68)$$

$$P_t: \text{Provisiones de pérdida permitida} \quad (6.69)$$

$$\alpha_{i,j}: \text{Recuperación de la pérdida} \quad (6.70)$$

$$\gamma_{i,j}: \text{Costo de recuperación} \quad (6.71)$$

$$R_t = 1 + r_t: \text{Rentabilidad exigida} \quad (6.72)$$

$$\delta_t: \text{Factor de descuento asociada al periodo } t \quad (6.73)$$

B . Tabla de grupos generados por productos y plazos

Tabla 6.1: Agrupación de 143 tipos de productos utilizada

Montos \ Plazos	[0; 3]	(3; 9]	(9; 15]	(15; 21]	(21; 27]	(27; 33]	(33; 39]	(39; 45]	(45; 51]	(51; 57]	(57; ∞^+)
[0; 250.000]	1	2	3	4	5	6	7	8	9	10	11
(250.000; 750.000]	12	13	14	15	16	17	18	19	20	21	22
(750.000; 1.250.000]	23	24	25	26	27	28	29	30	31	32	33
(1.250.000; 1.750.000]	34	35	36	37	38	39	40	41	42	43	44
(1.750.000; 2.250.000]	45	46	47	48	49	50	51	52	53	54	55
(2.250.000; 2.750.000]	56	57	58	59	60	61	62	63	64	65	66
(2.750.000; 3.250.000]	67	68	69	70	71	72	73	74	75	76	77
(3.250.000; 3.750.000]	78	79	80	81	82	83	84	85	86	87	88
(3.750.000; 4.250.000]	89	90	91	92	93	94	95	96	97	98	99
(4.250.000; 4.750.000]	100	101	102	103	104	105	106	107	108	109	110
(4.750.000; 5.250.000]	111	112	113	114	115	116	117	118	119	120	121
(5.250.000; 5.750.000]	122	123	124	125	126	127	128	129	130	131	132
(5.750.000; ∞^+)	133	134	135	136	137	138	139	140	141	142	143

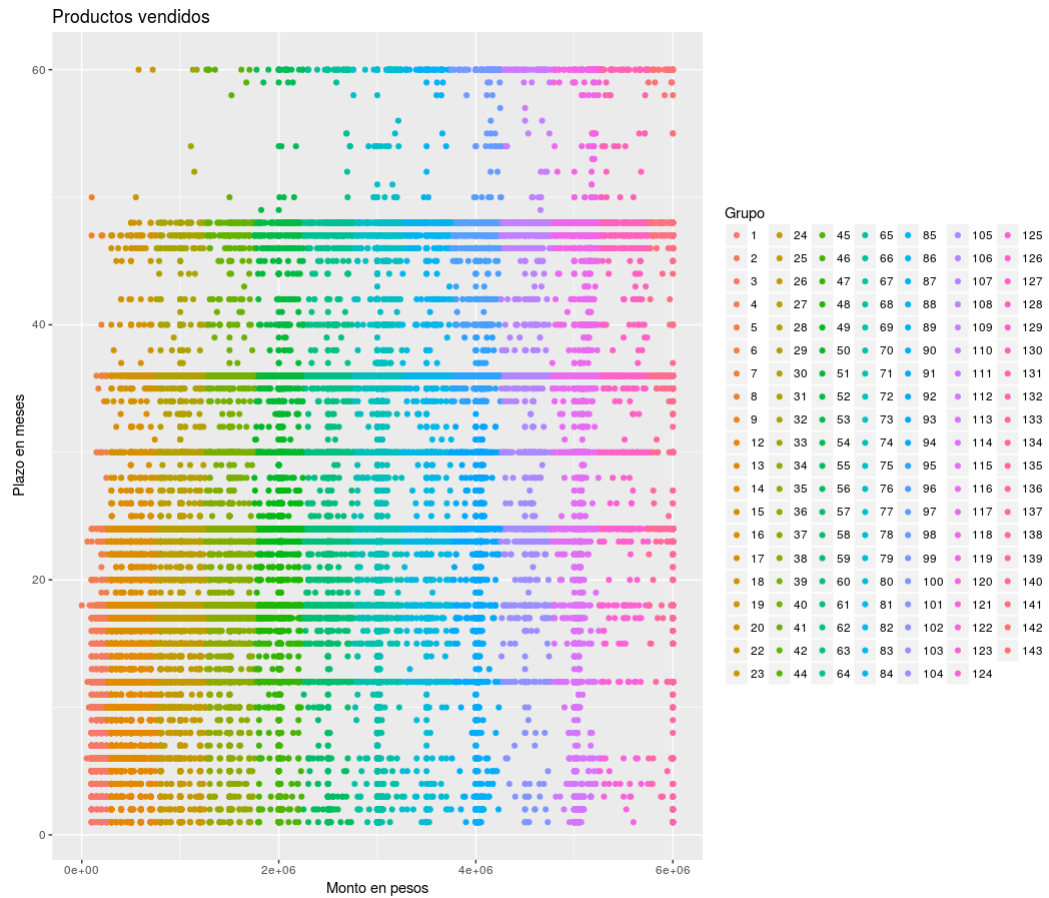


Figura 6.2: Distribución de créditos vendidos por monto versus plazo

C . Gráfico de grupos generados por productos y plazos

D . Variables seleccionadas de la base de datos para la descripción de productos

Tabla 6.2: Variables seleccionadas para la descripción de productos

Variable en base de datos	Detalle
ID	Id operación
OPE_FECHA_CURSE	Fecha operación
OPE_PLAZO	Plazo operación
OPE_TIPO_TASA_PENAL	Tipo tasa penal
cpd.descripcion	Descripción producto
OPE_MONTO_LIQ	Monto líquido
OPE_MONTO_PROX_VCMTO	Monto próximo vencimiento
ID_Cliente	ID cliente
OPE_FECHA_1ER_VENC	Fecha primer vencimiento
OPE_TIPO_TASA	Tipo de tasa
OPE_VALOR_TASA_PENAL	Valor tasa penal
cpd_producto_esencial	Descripción producto detalle
OPE_VAL_SEG_DES	-
OPE_CUOTAS_PAGADAS	Cuotas pagadas
OPE_OPERACION	Código operación
OPE_MONTO_ORIGEN_PES	Monto origen
OPE_TASA	Tasa mensual
OPE_TIPO_OPERACION	Tipo de operación
cpd.dop_ctr	Detalle operación
OPE_CANT_CUO	Cantidad de cuotas
tasa_porc	Tasa en porcentaje
meses_prox_venc	Meses para el próximo vencimiento
monto_grupo	Corte al que corresponde en monto
tasa_anual	Tasa anualizada
dias_prox_venc	Días para el proximo vencimiento
segmento	Segmento de clientes del banco
grupo	Grupo asignado según monto y plazo
cuota	Valor de la cuota según parámetros
years	Año de la operación
plazo_grupo	Corte al que corresponde en plazo