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

"INTEREST RATE OPTIMIZATION FOR CONSUMER CREDITS: EMPIRICAL EVIDENCE FROM AN ONLINE CHANNEL"

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL

MARTÍN CARLOS LAVANDERO IVELIC

PROFESOR GUÍA:
MARCELO OLIVARES ACUÑA

MIEMBROS DE LA COMISIÓN:
ANDRÉS MUSALEM SAID
LUIS ABURTO LAFOURCADE

SANTIAGO DE CHILE
2019

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL
POR: MARTÍN CARLOS LAVANDERO IVELIC
FECHA: 2019
PROF. GUÍA: MARCELO OLIVARES ACUÑA

"INTEREST RATE OPTIMIZATION FOR CONSUMER CREDITS: EMPIRICAL EVIDENCE FROM AN ONLINE CHANNEL"

Mientras que industrias como las aerolíneas, hoteles o cadenas de supermercado continúan desarrollando innovadoras estrategias de pricing, los bancos rara vez discuten nuevas estrategias de precios como una herramienta para diferenciarse del resto y mejorar su rentabilidad.

En los últimos 40 años, la gran mayoría de los bancos se han basado en los modelos estadísticos de credit scoring para fijar las tasas de interés, sin embargo, sólo pocos han adoptado nuevos enfoques. Un credit score representa la probabilidad de que un cliente caiga en default crediticio. Por lo tanto, esta estrategia basada en el riesgo cobra una mayor tasa de interés a clientes más riesgosos. Esto permite a los bancos compensar por las pérdidas dadas por la incertidumbre en la ganancia esperada y sus costos asociados. Si bien este método ha sido vital en el importante crecimiento de los créditos de consumo, no considera un problema fundamental: la sensibilidad al precio y la selección adversa. Es decir, el fenómeno en que clientes de bajo riesgo son más sensibles a cambios en precios comparados con clientes más riesgosos.

Esta tesis investiga créditos de consumo de uno de los bancos más grandes en Chile utilizando datos del canal online por un periodo de dos años. Mediante el uso de un modelo logit, la sensibilidad al precio y selección adversa son integradas para revelar correctamente el comportamiento del consumidor. El modelo calibrado incorpora la probabilidad de aceptar como la variable dependiente y la tasa de interés, riesgo y características personales/crédito como controles. El problema de optimización en este trabajo toma el modelo de probabilidad de compra y genera una tasa de interés óptima para cada cliente. Los resultados basados en data histórica muestran que la rentabilidad puede mejorar en un 6.63% y mejorar la probabilidad de aceptar el crédito de un 13.50% a un 17.72% si se aplican los precios recomendados.

Un AB testing se implementa en el canal online del banco para verificar que las elasticidades al precio se estimaron de manera correcta y ver si hubo mejoras de rentabilidad contra un grupo de control. Los resultados muestran que las elasticidades tuvieron valores similares; sin embargo, el número de observaciones fueron insuficientes para probar estadísticamente ganancias o pérdidas. Pese a los problemas relacionados con el tamaño muestral del experimento, el análisis de los datos históricos entregó resultados convincentes. Se encuentra que, ante una simulación de crédito de consumo, los segmentos de ingreso más altos son más elásticos al precio, a mayores montos hay una mayor elasticidad y que clientes más riesgosos tienden a ser menos sensibles al precio.

Este estudio utiliza técnicas de data analytics y revenue management para identificar qué segmentos de clientes son más o menos elásticos. Presenta una política de precios que mejora las ganancias esperadas del banco y contribuye a la literatura mediante el desarrollo de un método que cuantifica los valores de la optimización de precios.

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL
POR: MARTÍN CARLOS LAVANDERO IVELIC
FECHA: 2019
PROF. GUÍA: MARCELO OLIVARES ACUÑA

"INTEREST RATE OPTIMIZATION FOR CONSUMER CREDITS: EMPIRICAL EVIDENCE FROM AN ONLINE CHANNEL"

Whereas many industries such as airlines, hotels or supermarket chains continue to experiment with innovative pricing strategies, retail banks rarely discuss pricing innovation as a tool to differentiate from the rest and improve profitability.

Over the last 40 years, the vast majority of banks have relied on statistical credit scoring models to adjust loan interest rates and only a few have adopted new means. A credit score represents the probability of the borrower defaulting on a loan. Thus, this ubiquitous risk-based approach charges higher interest rates to riskier customers, allowing banks to adjust for higher expected losses resulting from uncertainty in the expected revenue and associated costs. While this method has been vital in allowing an important growth in consumer credits, it fails to account for a major problem: price sensitivity, which leads to the adverse selection phenomenon in retail credit, where low risk customers are more sensitive to a change in prices compared to high risk customers.

This thesis investigates consumer loans from one of the largest banks in Chile using data from the online channel for a period of two years. Through the use of a logit model, adverse selection and price sensitivity are integrated to correctly reveal consumer behavior. The calibrated model incorporates the take-up probability as the dependent variable and the interest rate, risk and personal/loan characteristics as controls.

The price optimization problem addressed in this work takes the propensity model from above and generates an optimal interest rate for each individual client. The results based on historical data showed profitability can improve by 6.63% and increase take-up rates from 13.50% to 17.72% when the recommended prices are applied.

A field experiment was implemented utilizing an online channel to check the estimated price elasticities and see if there were any revenue improvements against a control group that received the actual bank interest rates. The results showed the estimated elasticities were similar; however, the number of observations were insufficient to statistically prove any gains or losses with acceptable confidence.

Despite the aforementioned circumstances, analysis of the historical data yielded feasible results. It was found that when faced to a consumer credit simulation, higher income segments are more price elastic, larger amounts lead to a higher elasticity and riskier clients tend to be less price sensitive.

This study used data analytics and revenue management techniques to identify which segments of customers were more/less price elastic. It presented a pricing policy that improved the expected profit for the bank and contributed to literature by developing a method of quantifying the values of pricing optimization.

To my parents and family

Acknowledgments

I would like to express my deepest appreciation to my academic advisors, Professors Marcelo Olivares, Andrés Musalem and Luis Aburto, for their incredible amount of support throughout this thesis. I'm very fortunate to work with them in this project and honored to have them as my advisors. I appreciate the countless meetings, their time to answer questions and their commitment, constituting exemplary figures as scholars and individuals which I will try to imitate during my professional career.

I'm fortunate to the pricing team at the bank, in particular, I would like to thank Heber Matamoro and Rodrigo Garcés for their suggestions, tolerance and time invested in this project. I'm grateful for the meetings, advices and comments. I also, wanted to thank the bank for giving us the opportunity to test our model and believe in our research. They showed their commitment towards a continuous learning and looking forward to new challenges.

I would like to express my gratitude to ABIF *(Asociación de Bancos e Instituciones Financieras)* for the scholarship that helped me with financial aid to cover the research expenses.

I'm thankful to my friends from Universidad de Chile, for the good moments we shared and the hurdles we faced. I'm sure friendship will continue for the rest of our lives.

Finally, I would like to thank my family for their unconditional love and support. I'm particularly grateful to my parents and their endless effort that made me who I am. This thesis is dedicated to them.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Over the last decade, customer quality information has improved to a great extent. Examples of such information include customer demographics (age, date of birth, income), purchase behavior and social media activities. Using sales data on these customer characteristics and optimization tools, it is now possible to capture an individual customers' willingness-to-pay and charge an optimal price, increasing profits and market penetration. Benefits of customized pricing are straightforward, the firm sets price at the highest acceptable point for each person and the customer value is maximized. This method has the potential to increase revenue over systems in which customers are homogeneous or firms that lack the resources to target individualized consumers, i.e., use a fixed pricing policy.

Many lenders consider risk-based pricing to be the ultimate in pricing sophistication. There is no question that identifying higher-risk customers and charging them higher rates to compensate for higher losses is a rational and profitable idea. However, risk-based pricing does not incorporate one of the key elements required to maximize financial return: customer price sensitivity.

Numerous firms must solve the following problem on a daily basis: first, whether to charge high prices and earn more money per transaction, which leads to a decrease in the number of people taking credits, or, second, charge low prices reaching a larger market, but decreasing spreads and making less profit per transaction deal.

The key to maximizing profitability through pricing is to trade off between price and take-up probability. A trade off that needs to be made for each combination of customer segment, product and channel and updated as cost of funds or market conditions change. In the airline industry for example, it is known that late booking customers (just a couple days before the flight) are likely to be business people who are price insensitive while travellers booking with anticipation are looking for good deals and are more price sensitive. By setting and adjusting rates using these insights, airline companies can make greater profits. While this approach is used in many industries, banking has not adopted it yet. Nonetheless, banks have a gold opportunity compared to the rest. They have extensive information of the customer like age, type of job, wage, debts, credits taken or failed sales and therefore, it is time for banks to pay attention to pricing in order to benefit from this important opportunities and increase profitability.

With the emergence of statistical pricing tools, institutions are moving to a more centralized fixed pricing system because they provide better price discrimination opportunities compared to the sales agents at branch offices. The best way to follow that direction is using technology, in particular, the web. Through the web, consumer credits can be offered at lower costs (an inexpensive channel that does not have to pay sales forces or maintain the branch offices) and also, avoid suboptimal interest rates as a product of a negotiation between the sales agent and the customer. Hence, it is objective, reproducible and accessible to people where information is transferred much better.

This thesis uses data from one of the largest banks in Chile and identify which segments of clients are more price elastic/inelastic to interest rates in order to implement a pricing policy that accounts for trade-off between the probability of taking the loan versus the financial margin.

Many factors influence sensitivity, some of them are related to the characteristics of the credit: amount requested, duration or interest rate. Personal/financial characteristics: age of the consumer, profession, years affiliated with the bank, numbers of credits taken, debts and finally, market conditions: maximum interest rate allowed, inflation, cost of funds, market competition, monetary policies or global economy.

The new method recommended to set prices in the credit environment is much more complete than the actual risk-based approach and it has already been used in other industries like auto-mobile lending in the United States. Our reference is the work methodology adopted by (Phillips et al., 2015).

The web channel will be the subject of study and presents the following characteristics that are helpful for our research:

- It represents the 56.1% of credits taken in a period of two years and the sum of amounts per loan accepted represents a 31.6% of total sales.
- Clients present specific characteristics that allow to have larger spreads compared to other channels.
- Data sets available for clients that simulate and those who take credits. Personal information is also available (wage, debts, risk level and more)

The proposed methodology of personalized pricing is structured in the following steps:

1. From online simulations (some of them failed and others taken), the price-response is estimated on take-up using rate as one of the explanatory variables.
2. Then, the coefficients for rate are estimated to model how customers would respond to different offered prices. Thus, the objective function is optimized, that is, financial margin incorporating willingness to pay. To the extent that the model accurately captured price-response it is found that by setting optimal prices produced by a centralized

data-driven profit maximization procedure, prices could have been set that would have increased profitability by 6.63% and the take-up probability from 13.50% to 17.72%.

3. A field experiment is implemented to check the estimated elasticities and see if there were any revenue improvements. This randomized controlled trial was applied in the high and middle/young professional income segments. Clients were randomly assigned from both segments and the control group received the actual bank interest rates, whereas the treatment group received the optimized interest rates.

The results make a diagnosis of the actual pricing system by determining how elastic are the segments to the interest rates and identify the areas in the pricing matrices where the bank is less competitive. Modeling the take-up probability and incorporating it in the optimization procedure allows us to recommend an optimal interest rate for each segment of client according to market conditions.

1.2 Problem description

A brief summary presents the problems faced throughout the elaboration of this thesis:

How to separate the risk effect and the willingness to pay for each client: According to the *price-dependent risk* phenomenon (Phillips et al., 2011), riskier customers tend to be less price-sensitive and have higher overall take-up rates. Thus, if the risk variable is not considered in the model, the estimates will come from different demand curves resulting in biased coefficients.

Figure 1.1 shows an example of the endogeneity problem in the price-response model. Imagine a bank having a customer segmentation with three levels of risk: low, medium and high. The omission of the risk variable would lead to only one flat and underestimated demand curve (in red), which is not representative of reality, instead of having three different demand curves for each risk level.

Determine how the price affects the take-up probability controlling for credit, personal and market conditions: The challenge consists in identifying different customers segments and estimate their price elasticities offering the correct interest rates. Therefore, the idea is to improve the existing customer's targeting and pricing mechanisms to focus credit discounts in those segments that are more price elastic. By contrast, in inelastic customer segments the strategy is to set higher rates in order to improve profitability.

Optimize and recommend changes to the existing price matrices: The aim is to offer a personalized interest rate for each client, where the take-up probability and financial margin are maximized.

Evaluate and compare the proposed model against the existing pricing policy and measure any significant effects: Each model's performance is evaluated through the use of historical data and a field experiment.

Design and maximize an objective function considering cost of funds, take-up probability, risk and financial margin: Many objective functions can be modeled depending upon the bank's goal. For instance, a bank might want to increase the number of credits, decrease credit defaults or maximize the financial margin. Thus, three different approaches would be needed. The complexity consists in identifying the bank's intention and adapt it to an objective function.

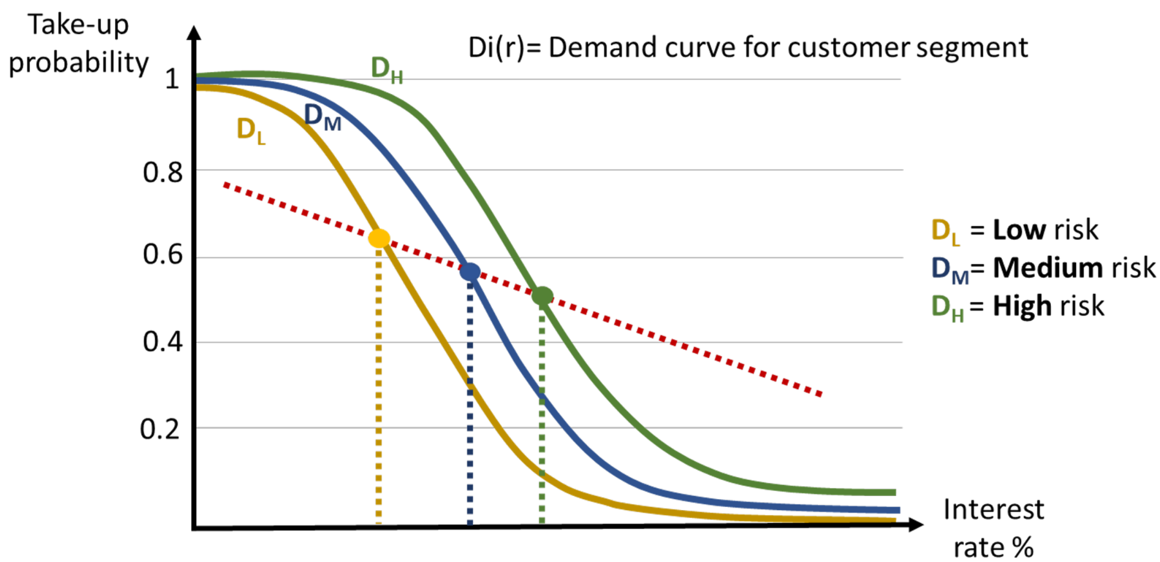


Figure 1.1: Representation of the endogeneity problem in the price-response model as result of variable omission.

1.3 Overall methodology

Figure 1.2 summarizes the methodology of the study:

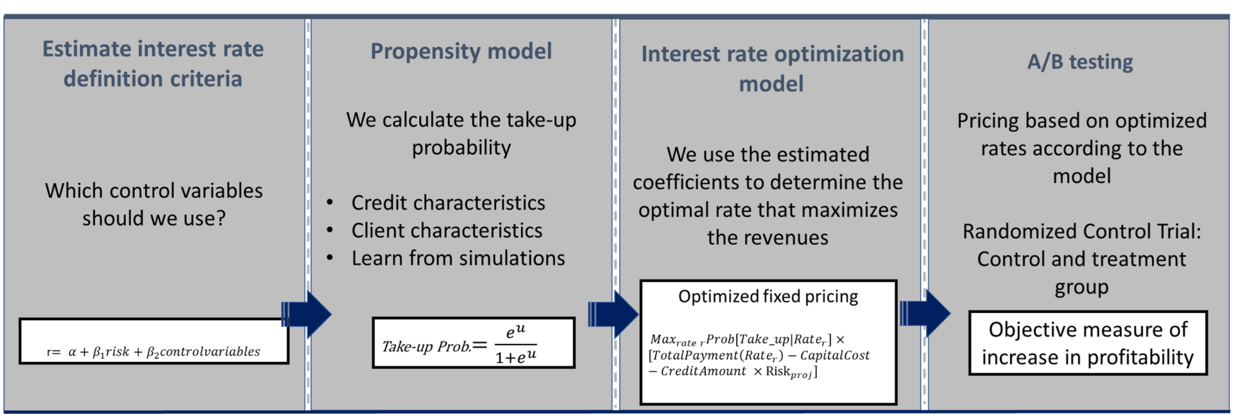


Figure 1.2: Big picture of the study

The steps of the analysis are as follows:

1. Estimate the interest rates the bank is charging based on historical data. The objective is to understand the bank's interest rate definition criteria and for which variables to control. This step is helpful in order to identify the most important variables and check if there are any endogeneity problems when estimating the model.
2. A logit model is used to estimate the take-up probability for each client. This input will be utilized in the objective function to maximize profits by setting a trade-off in between interest rate and take-up probability.
3. The estimated coefficients are used to set the optimal prices produced by a centralized, data-driven maximization procedure which utilizes the historical data.
4. The pricing model is tested based on the optimized rates using a randomized control trial with a control and treatment group. The objective is to validate the estimated price elasticities and evaluate whether the profits are larger than the control group.

1.4 Thesis structure

This thesis proceeds in section 2 where literature on credit pricing is reviewed.

In section 3 the context of the application is described showing some summary statistics and the data.

Section 4 examines the econometric model of take-up probability. In particular, customer choices are identified and simulations are connected to purchase occasions and purchase decisions. This is addressed by grouping simulations into clusters. Then, the endogeneity of price is reviewed and addressed by adding controls. The econometric model description ends with the analysis of heterogeneity in price sensitivity and captured by segmentation of the customer group and risk. Section 4 continues with preliminary results of the choice model and ends with a re-estimation of the initial model because it was found that a simulation step was omitted in the methodology. Hence, a variation of the model is presented.

Section 5 shows the optimization results of the initial and final model optimized interest rates for each historical transaction and the average profit earned with the proposed method. It finalizes with a recommendation of new pricing matrices that are later used in the field experiment.

Section 6 introduces the experimental design applied in the bank. Then, the field experiment data were used to estimate the elasticities and validate the historical model estimates. This chapter ends showing AB testing results: Take-up and revenue improvements for the treatment and control groups.

This thesis concludes in section 7 and discusses future work.

Chapter 2

Related literature

For many years, risk-based pricing has been the state-of-the-art in credit pricing. Edelberg (2006) shows that lenders increasingly used risk-based pricing of interest rates in consumer loan markets. The loan offered to a risky borrower should increase as the risk of default increases in order to cover the expected increase in a model in which low-risk customers are more price sensitive than high risk customers.

Having said that, knowing more about the customers would likely improve profits. Cross & Dixit (2005) state that with a customer centric pricing setting prices that accurately reflect the perceived value of products per customer segment benefit producers and consumers.

Attanasio et al. (2008) studied the credit constraints in the market for consumer durables taking evidence from micro data on car loans. They estimated the elasticities of loan demand with respect to the interest rate and the maturity of the loan using a three period model and showed that credit constraints are binding for young and low-income households.

Other strategies to measure elasticity have also been adopted, Karlan & Zinman (2005) used a logit model to identify demand curves that are downward-sloping with respect to price and used empirical evidence from a micro-finance lender in south Africa. Phillips (2005), discusses the use of customized-pricing bid response models in industry and later, in Agrawal & Ferguson (2007) bid-response models (logit and power functions) are used along with segmentation to increase profits. They additionally tested them using an industry dataset to compare their performance and estimate the percent improvement in expected profits.

Most of the literature on the topic of pricing, focuses on determining the factors that influence price setting and their relationships between consumer behavior and pricing. Only recently the retail credit price optimization has been formalized and lenders have begun to adopt pricing optimization approaches that consider consumer willingness-to-pay as well as risk in setting prices for credit. Phillips (2013), discusses the non-linearity of incremental contribution in interest rates, the uncertainty of the return on a loan, the role of information, and the interaction between pricing and risk. This work also describes how these aspects can be incorporated into the determination of optimal rates for consumer credits.

The majority of the papers presented above are primarily theoretical. Empirical studies of pricing optimization applied to car loans are found in Phillips et al. (2015). They proved that using a unique data set from an auto lender that used field negotiation to determine prices the local sales forces adjusted prices in a way that improved profits by approximately 11% on average. They also showed that a centralized fixed pricing system may better segment potential customers based on willingness to pay and risk improving profits up to 20% over those realized from the local sales forces adjustments. This result suggests that even though the dealers were improving profits relative to the nominal rates set by the lender, there was an additional opportunity to improve profitability further through optimized prices. It also implies that banks must look forward to an optimized pricing model for each customer to better capture price sensitivity and profit more.

Chen et al. (2015) and Javanmard & Nazerzadeh (2016) studied pricing problems with demand learning. In the presence of existing sales data on customer features, Chen et al. (2015) modeled decision problems in which actions can be personalized by taking into account the information available to the decision maker using a logit model. They tested their method on real transaction data for airline seating reservations and showed that effective customized pricing can increase revenue by at least 7% over the best single-price. Javanmard & Nazerzadeh (2016) studied a pricing problem faced by a firm that sales products to customers that arrive over time. They considered a dynamic pricing problem with a binary choice model and learned as the sales data accrued over time. With this information they proposed a pricing policy in their setting.

Ban & Keskin (2017) used machine learning techniques to dynamically adjust the price of a product at the individual customer level, by utilizing information about customers' characteristics. Their work is different from Chen et al. (2015) and Javanmard & Nazerzadeh (2016) since price sensitivity was captured and depended on individual customer features. They tested the theory on simulated data and on a data set from an online auto loan company in the United States and the proposed pricing policy improved by 32% the annual expected revenue.

Studies involving sensitivity and optimization are found in Vulcano et al. (2010), where choice behavior and customer preferences are estimated for price in a major U.S airline using an expectation-maximization algorithm. Their estimates were then used in a simulation study to assess the revenue performance of the current controls used by the airline relative to controls optimized showing a 1%-5% revenue improvement. Then, Terblanche & De la Rey (2014) determined optimal prices to quote prospected customers in credit retail, maximizing interest rates while taking price sensitivity and adverse selection into account.

The consumer credit optimization has little research and has been mainly addressed by Robert Phillips. This thesis is based on Phillips et al. (2015) and applied to the banking industry through an online channel where no negotiation is allowed.

The contributions of our work are as follows:

- Separate the risk effect and the willingness to pay for each client.
- Determine how the price affects the take-up probability using a logit model with data from a major bank in Chile.
- Optimize the interest rates according to the estimated elasticities and recommend changes to the existing price matrices.
- Design and maximize an objective function according to the bank's need.
- Evaluate and compare the proposed model against the existing pricing policies to measure their performance.
- Develop a method of quantifying the value of optimization in a specific organization.

Chapter 3

Context of the application and data description

3.1 Data

This section describes the data used throughout the thesis. Mainly, information from credits, clients and the market was gathered.

The bank provided the data related to online simulations performed by clients, the purchased transactions in every sales channel (branch office, web, phone call, proservice and ATM), client information: demographic data, internal debt, debt with other banks also known as "SBIF" debt), income, segment, trade agreements and projected risk rate.

The data related with the market belongs to the regulator of banks and financial institutions in Chile (hereafter SBIF[1]). Specifically, we worked with different indexes, some of them were the maximum rate of interest (TMC)[2] or the average banking interest rate.

Different types of data were used:

- ***Customer identification data:*** The customer identification field, client's ID (party_id), is uniquely assigned to each customer of the bank. This ID is extremely important because it will be used as a key to join different data sets.

- ***Demographic data:*** Information particularly useful to segment individuals. For example, one could use a predictive model to identify the demographic characteristics a high-value client has and improve the pricing system.

  The database contained 346.744 observations, where each line had the birth date, city, region, neighborhood, gender, income and the income segment to which the client

[1] SBIF is an acronym that refers to "Superintendencia de Bancos e Instituciones Financieras" that traslates as Superintendence of Banks and Financial Institutions.

[2] Usury laws specify a maximum interest rate to charge in consumer credits called TMC. TMC is an acronym that refers to "Tasa Máxima Convencional" that translates as Maximum Conventional Rate

belongs according to the bank. A second database contained *income information:* 7.552.120 observations, with each client's ID, and a monthly estimated income for a period of two years (2016 and 2017).

- ***Transactional data:*** According to researchers, the best predictors for customers' future purchase behaviors are their historical transactions. A transactional database was utilized and provided valuable information from credit simulations in order to identify different simulation characteristics: if the credits were taken/not taken, the requested amounts, terms, dates and time, ID, discounts, taxes or interest rates.

  Credit simulations: Individual-level data of all new credits extended through the online channel between 2016 and 2017 were used. Total number of observations were approximately 714.727. The data contained credit characteristics and consumer characteristics (client's ID, date and exact time of the simulation, amount, interest rate offered, term, taxes, discounts, amount and type of insurance, type of credit and the step the client reached in the simulation process).

  Purchased credits: This database contained 192.978 purchased operations from every channel in the bank and all the types of credit. These are the simulations that ended up in a purchase and generated the revenue of the business for the period 2016 and 2017. Among the total number of operations, a segmentation was performed based on the type of product (consumer credits CON300) remaining 94.928 left from which 92.068 belonged to the online channel[3]. The most important columns show the characteristics of the credit: interest rate, date, cost of funds, term, spread or amount. Also, it contains client's information: projected risk rate, income segment, if the client is new in the bank and finally the segmentation the bank generates for each variable mentioned above. The latter will be useful to guide the thesis.

  Cost of funds: From the purchased operations database a table was constructed with daily cost of funds depending upon the credit term/horizon for the period 2016-2017. This information was fundamental to compute the objective function.

  Credit incentives/benefits: 3.075.995 observations, with client's ID and a variable that points out whether the credit has or not a discount. Data covers the period from 09/2016 to 12/2017.

- ***Financial information:*** *Projected risk rate:* Approximately 10.000.330 observations with the client's ID and a monthly projected risk rate for each client from the years 2016 to 2017. The projected risk is defined as the probability of default within a year, it ranges from 0 to 100%.

[3] Our sample to estimate the logit choice model became smaller after performing a segmentation to target the interest group, see section 3.1.2 for more information.

Bank and SBIF debt: For a period of two years (2016 & 2017), there were 7.091.156 and 7.839.862 observations respectively with the client's ID, internal debt in the bank and banking system debt measured monthly. For both cases we have the total debt and debt by sector: consumption, commercial and mortgage.

- ***Market Data:*** A database was built from the SBIF's publications where the maximum rate of interest is set depending on the amount and date of the credit. The average banking interest rate was also collected[4] for a specific date and amount. Data period starts from 2016 to the end of 2017.

  Once different databases were uploaded, they were joined together with the PostgreSQL program through *querys*. As a result, only one consolidated database contained all the information and it was used to run different studies.

Table 3.1: Transactional database example.

Party id	Date	Amount	Interest Rate	Term	Discount
42	2017-11-08 09:50:51	$6,000,000	0.96%	12	NA
42	2017-11-24 16:41:23	$10,000,000	0.99%	24	NA
42	2017-11-27 12:01:43	$10,000,000	0.99%	24	NA
166	2017-05-16 12:40:14	$3,000,000	1.37%	24	15%
166	2017-05-16 12:43:30	$1,500,000	1.56%	12	15%
421	2017-09-24 23:26:42	$218,000	2.59%	6	NA

3.1.1 Summary statistics

Through five different channels, the clients of the bank can take consumer credits: at the branch offices where the client might negotiate the price with a sales agent, via web, phone calls, Proservice (a subsidiary of the bank) and at ATMs.

Considering the purchased credits in a two-year period (2016/01-2017/12), the web channel had the largest number of loans (56%) followed by the branch offices as shown in Figure 3.1. Although, if the amounts of the credits taken at each channel are added, the branch office represents about 59% of the total sales. Hence, the average loan is bigger compared to the rest of the channels, see Figure 3.2 for details.

[4]We use a document that includes a monthly series of average interest rates in the banking system, from December 1999 to date (Only 2016-2017 data were considered). Not indexed currency.

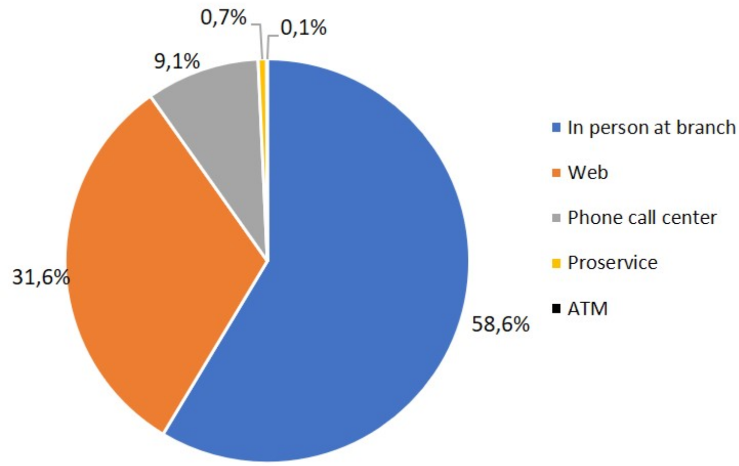


Figure 3.2: Sum of loan amounts per channel as percentage of total sales.

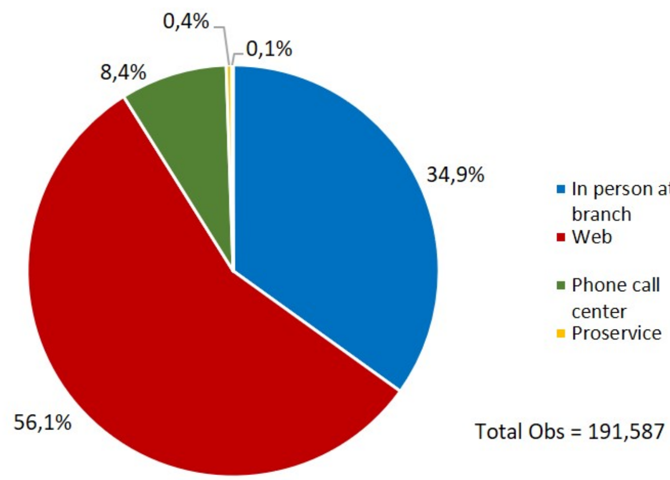


Figure 3.1: Total credits taken by channel in a 2-year period.

Table 3.2 shows summary statistics per channel, we can see the branch office presents the highest average loan amount, which is approximately 3 times the web channel's average, and that might also explain a lower APR[5]. Column 4 shows the average projected risk rate for clients by channel. Proservice concentrates the riskiest people (probability of default in a one-year period is about 9%), but this channel accumulates 840 observations during a two year period and represents a 0.4% from total observations, which is not significant.

Usually longer credit terms are correlated with larger loan amounts. Thus, there is a clear trend when comparing the average loan amounts and the terms of the different channels.

[5]APR is an acronym that refers to the Annual Percentage Rate charged for credit loans

Higher terms for the branch office loans and low terms for the web channel are expected.

Web customers are considered good clients for the bank, they present the lowest average projected risk rate and high spreads. Plus, no negotiation is allowed in the process of acquiring a credit, so it perfectly qualifies for a customer-centric-pricing strategy. In terms of estimating the parameters of the model, it is the best channel because no negotiation through sales agents is allowed, a problem that according to Phillips et al. (2015) might lead to endogeneity.

Table 3.2: Summary statistics per channel

Purchased operations - Channels						
Channel	Loan amount	APR	Projected risk rate	Term	Spread	Observations
Branch office	$ 9,020,625	13.28%	2.83%	37	9.09%	66,927
Web	$ 3,024,590	17.16%	1.78%	23	13.28%	107,562
Phone calls	$ 5,821,865	16.63%	3.07%	31	12.32%	16,068
Proservice	$ 8,320,313	10.89%	9.10%	41	6.52%	840
ATM	$ 3,034,467	22.36%	2.19%	16	18.15%	190

This table contains all type of credits

The web/online channel presents four income segments: massive/emerging/university, middle/young. Prof, high and preferential, ordered from lower to higher income respectively.

From Table 3.3, as the income of the customer is increased, higher average credit amount is requested. The group that outstands over the rest is the preferential income segment, having a much longer average loan amount $7.700.000 CLP aprox, although it represents a smaller portion of total purchased operations.

Following the logic, from Table 3.2, higher loans lead to longer terms. The highest APR's are charged in the massive/emerging/university segment and the largest spreads are earned (around 19%), but presents the riskiest clients. Meanwhile the lowest APR's are charged in the preferential segment along with the lowest spreads, the pros are that it has the least risky clients (0.9% probability of default in a one-year period).

Table 3.3: Purchased operations on the Web channel

Purchased operations - Web channel						
Income segment	Loan amount	APR	Projected risk rate	Term	Spread	Observations
Massive/Emerging/University	$ 1,260,340	22.65%	2.40%	19	18.85%	13,886
Middle/Young.Prof	$ 1,710,176	19.54%	1.90%	20	15.67%	35,652
High	$ 3,184,322	16.10%	1.33%	24	12.15%	37,445
Preferential	$ 7,724,488	11.54%	0.90%	28	7.64%	5,085
Total	$ 2,574,056	18.17%	1.70%	22	14.27%	92,068

Note monetary values were expressed in Chilean pesos CLP and the APR corresponds to the annualized interest rate (12-month period). Total observations are less compared to those reported on the web channel in Table 3.2 because we only considered a specific type of consumer credit in our study.

Most of the purchased operations were grouped in the middle/young. Prof and high-income segments, see Figure 3.3. Hence, this thesis was focused in both groups to estimate the model and then test it with control and treatment groups in the online channel.

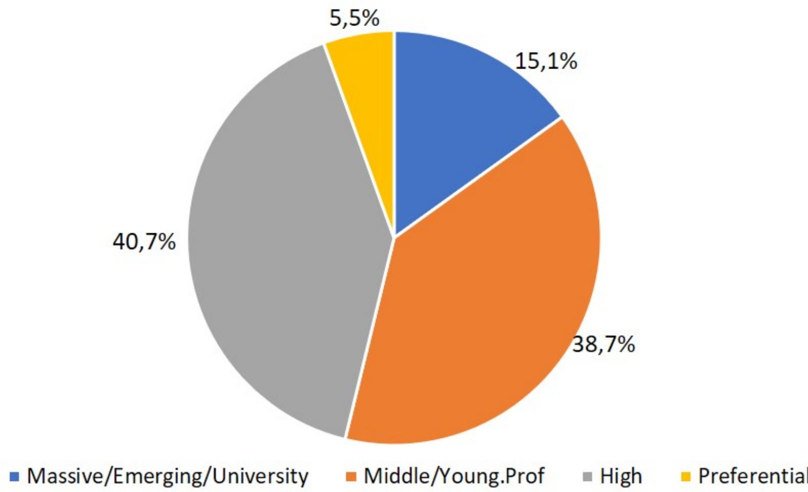


Figure 3.3: Percentage of credits purchased per income segment.

Table 3.4 shows the correlation matrix of the most important variables from the customer online simulations data set.

Table 3.4: Correlation matrix

	Amount	APR	Term	Proj.Risk	CF
Amount	1	-0.383	0.456	-0.089	0.237
APR	-0.383	1	-0.265	0.366	-0.067
Term	0.456	-0.265	1	0.003	0.456
Proj.Risk	-0.089	0.366	0.003	1	0.051
CF	0.237	-0.067	0.456	0.051	1

3.1.2 Data Cleaning

In this section the simulation concept is defined, a fundamental element of this thesis used to model the client's behavior and estimate the consumers' price response function in order to find the optimal interest rate to charge.

Simulation: Attempt of purchasing an online credit. The client gets an interest rate depending on the amount and term chosen for a specific date and time.

In this study, only unique simulations were considered. Hence, duplicated rows or information were deleted from the database, leaving just a single combination of client ID, amount, interest rate, term and date excluding the insurance options from the analysis. Also, the simulations were filtered, so they can satisfy the following conditions:

1. Only coming from the online channel
2. It must be a new sales operation: To avoid having an interest rate as a result from a negotiation in between the client - sales agent, the simulations involving refinanced prepayment, refinanced prepayment with arrears, extension or renegotiated were deleted.
3. Only consider the last step the client went through in the simulation process with at least one of these elements different: amount, interest rate, term or date.
4. Consumer credits with a specifict code (CON300) were studied because they meet the requirements wanted for a client. In other words, different online simulations that can end up in a purchase without a sales agent interaction (the whole process is managed via web)

3.1.3 Pilot test data selection:

A test pilot was designed to estimate the coefficients and later evaluate the model's performance. A more restrictive data selection was applied:

- Neither credit benefits/incentives nor initiatives
- Clients with a projected risk rate lower than 25%
- Only high and middle/young professional income segments
- Simulations from the second step of the simulation matrix on.

3.2 Bank's pricing method

Before working on the model, it is important to understand the bank's regulations and pricing policies. In this section, the actual scenery is described in order to have a better understanding of this thesis and identify potential problems, delivering suggestions or an ultimate improvement proposal.

Segmentation is one of the key elements to set prices that accurately reflect the perceived value of consumers. Since the customers are not homogeneous, the bank performs segmentation techniques that include income, risk, the loan amount and its term. Through a decision tree, the interest rates are calculated taking into account client and credit characteristics. Then, a base pricing matrix is defined for each income segment and depending on the risk tiers the interest rate is weighted with a penalization rate.

The vast majority of the online consumers get the rates from the base pricing matrix i.e., present low risk levels, therefore no penalization is applied. The bank, from time to time, also performs discounts or initiatives for specific groups of clients. Some examples are a discount for being a new affiliated and a special rate for the bank employees or selected institutions.

The risk factor is very important at the moment of defining an interest rate. Riskier clients usually have a lower probability of payment which may lead to bank losses. To prevent this, the pricing team established a penalty that affects the interest rate depending exclusively on the client's risk.

As we mentioned before, the objective of segmenting client data in tranches or tiers is useful for pricing rules. In this particular case, the bank distinguishes 5 groups with different levels of risk and each one of them is adjusted with a different rate. The actual system works as follows:

Step 1. An initial interest rate is computed considering the income segment of the client and the credit characteristics: amount, term and date. With this information the bank creates a base matrix with a combination of terms, amounts and income segments.

$$InterestRate_{\text{initial}} = \alpha + \beta_1 amount + \beta_2 term + \beta_3 incomeSegment + \beta_4 X\ldots$$

Step 2. Client's risk evaluation, initial interest rate is taken and adjusted as follows:

$$Final\,Interest\,Rate = \begin{cases} No\,adjustment\,over\,the\,initial\,interest\,rate & \text{if } Proj_{risk} < 0.8\% \\ 10\%\,of\,penalty\,over\,the\,initial\,interest\,rate & \text{if } 0,8\% \leq Proj_{risk} < 1.8\% \\ 20\%\,of\,penalty\,over\,the\,initial\,interest\,rate & \text{if } 1,8\% \leq Proj_{risk} < 3.8\% \\ 25\%\,of\,penalty\,over\,the\,initial\,interest\,rate & \text{if } 3,8\% \leq Proj_{risk} < 8\% \\ 100\%\,of\,penalty\,over\,the\,initial\,interest\,rate & \text{if } Proj_{risk} \geq 8\% \end{cases}$$

$Proj_{risk}$: Denotes the projected risk rate for the client. It measures client's probability of default within a year.

Finally, each income segment has a base matrix and other 5 matrices with the penalty applied over the interest rate depending on the tier defined.

3.3 Scope of the study

The objective of this thesis is to develop a base model that recommends interest rates for a specific segment, and it is going to be tested through the design and implementation of an experiment.

Given the variety of the bank offer, the study focused on specific types of credits, sales channels, clients or initiatives that will outline the project. The selection criteria considered the best configuration of elements to perform an analysis with the cleanest and most exact information, avoiding model bias.

Having said that, the scope of work is clarified:

- The study of only one sales channel: online.

- The insurance offered in the credit's sale is not considered to analyse the take-up probability.

- The model is estimated under certain conditions, thus it can only be applied to clients belonging from a specific segment. The credit and clients' characteristics are detailed in the A/B testing section.

Chapter 4

Econometric model of take-up probability

4.1 Logit Model

In the logit model, the outcome either occurs or does not. Therefore, a yes or no decision is modeled: to take (Y=1) or not to take the credit (Y=0). The outcome represents a probability and it is modeled as the net utility of taking up the credit at least once, which depends upon observables and unobservables for each customer. The standard form of the logit is:

$$Prob(Y = 1|x) = \frac{\exp^{x'\beta}}{1 + x'\beta}$$

A determined set of factors, such as age, marital status, education, and work history for example, are gathered in a vector $\boldsymbol{x}$ and explain the decision, so that

$$\begin{aligned} Prob(Y = 1|x) &= F(x, \beta) \\ Prob(Y = 0|x) &= 1 - F(x, \beta) \end{aligned}$$

The set of parameters β reflects the impact of changes in $\mathbf{x}$ on the probability.

A short summary of the main challenges in the econometric model is included below:

- To identify the customer choices and how to connect simulations to purchase occasions and purchase decisions. This is addressed by grouping simulations into clusters.
- The endogeneity of price. This problem is addressed by adding controls.
- The heterogeneity in price sensitivity. Captured by segmentation of customer group and risk.

4.1.1 Purchase Occasion clustering

The purchase occasion to which the customer decides whether to take or not the credit is defined as the group of banking simulations[1] within a period of time t that forms part of the decision-making process of a client in the purchase/non-purchase of a credit.

To determine whether simulations i and j were part of the same purchase occasion, i.e., set a time window, the client's activity was studied. To do so, the purchased operations for the consumer credits (between years 2016-2017) were reviewed and a retrospective analysis was performed to evaluate the number of daily simulations before the purchase.

The results from this study were helpful to better understand the customer purchasing behavior. Identifying the temporary windows that gathered most of the activity provided the reference point to cut.

Figure 4.1 shows a histogram with the number of daily simulations before the credit purchase. t is defined as the credit take-up day, thus *t-5* represents five days before the purchase.

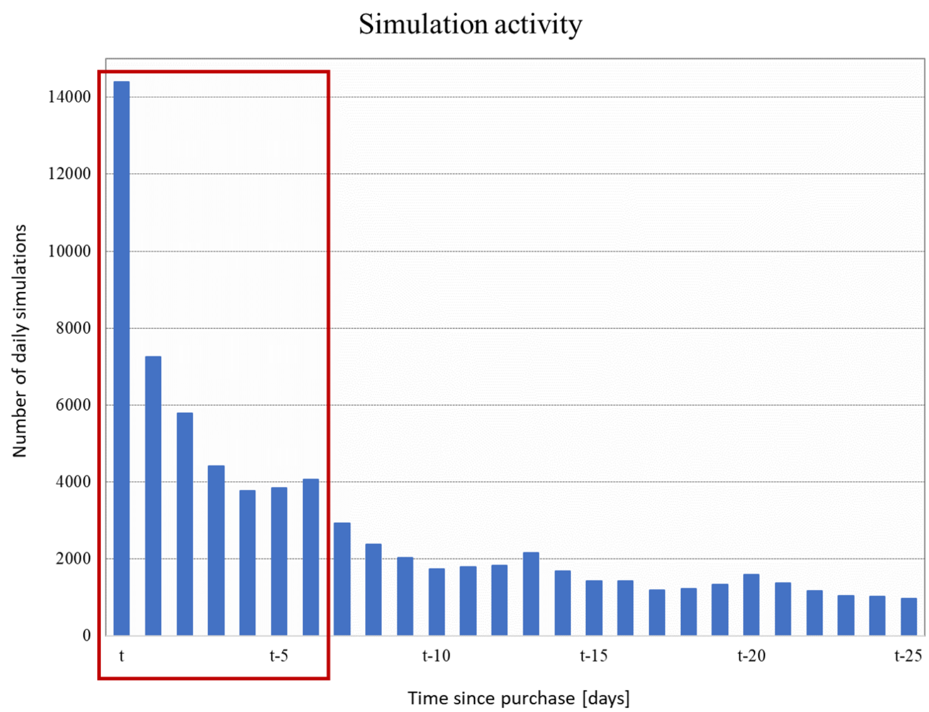


Figure 4.1: Daily simulations' distribution for clients acquiring online consumer credits.

As long as the purchase horizon increased daily simulations decreased. From the simulations' distribution it was concluded the customer behavior presented an intense activity before the week of the purchasing decision. This finding helped to set a temporary window of *t-6* days to perform the cut.

[1]A banking simulation is an attempt of purchasing an online credit at a specific date. The client has to fill the characteristics of the loan required like the amount, term or insurance options.

Each simulation associated to a same purchase occasion was grouped in a cluster. Clusters represent the bank's offer to clients according to the characteristics of the credit required and upon which they will have to make a binary decision: whether to buy or not.

As discussed above, the following criteria is used to define the simulations corresponding to the same purchase occasion:

1. Grouped simulations must belong to the same client and keep a temporary order.
2. There can be only one purchase within the group of chosen simulations. Otherwise, it is considered as another search process and independent from the one before.

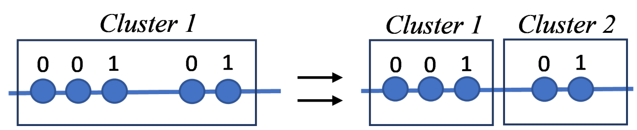


Figure 4.2: Example of corrected cluster. If the simulation ends up in a purchase it takes value 1, otherwise, it is represented with a 0.

3. In between a simulation and the preceding one there cannot be a time window bigger than 7 days.

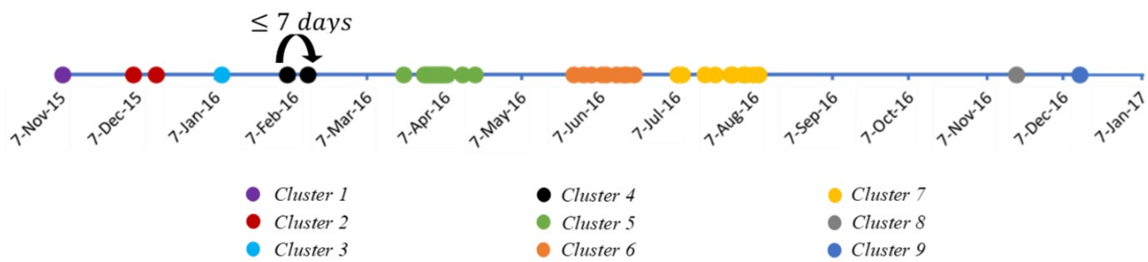


Figure 4.3: Representation of the client's simulations grouped by clusters for a period of 15 months.

Once the rules and the time window cut were established, the quality of the clusters formed were validated through two measures:

- *Average of the variance between amounts within a cluster:* The procedure consists in taking the variance between amounts in each cluster and then calculate this statistic's average considering all the clients.
- *Cluster overlap:* If the last simulation of a cluster has an amount that differs in less than a 10% with the preceding simulation, and the latter has less than x+1 days (where x is the time window to set a purchase occasion)

To illustrate an example of the "overlap", seven clusters are formed with a time window $\leq$ 7 days between simulations. As seen in Figure 4.4, the difference within the last simulation of cluster 1 and the first from cluster 2 is eight days (x+1). If the absolute value of the difference is less than a 10% of the loan amount requested in the last simulation of cluster 1, instead of considering them as separated clusters, we should have treated them as a whole. Same occurs in case of cluster 6 and 7, where we should have grouped those 3 simulations in one purchase occasion rather than two.

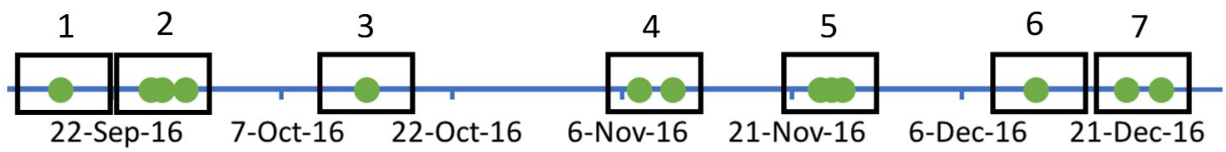


Figure 4.4: Cluster overlap example.

Figure 4.5 shows the results obtained with a graph that contains the average coefficient of variation for each time window defined (y-axis) and the percentage of clusters overlap (x-axis).

For the coefficient of variation: $C_v = \sigma/|\bar{x}|$, where $\bar{x}$ is the average of the amounts requested in the purchase occasion and σ its standard deviation.

Then, the average was computed C_v for all the clients and for each time window:

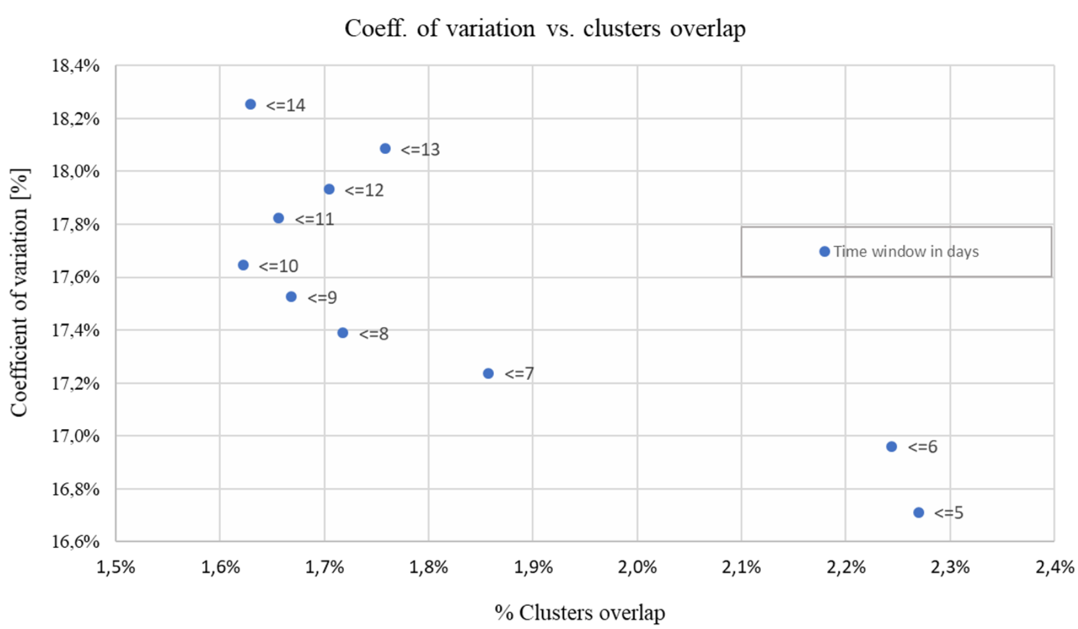


Figure 4.5: Relationship between the coefficient of variation and percentage of clusters overlap as the time window increases.

The graph suggests that with an increase in the time window (difference between consecutive simulations) greater is the coefficient of variation within a cluster, thus the amount's dispersion is larger. Furthermore, the percentage of "clusters overlap" or that were not formed well are at most 2.4% which is insignificant compared to the total. However, the change between a period of 6 to 7 days in between simulations decreases the error from 2.2% to 1.9% respectively and increasing to 8 days rises the coefficient of variation. From this analysis it was proposed to set 7 days as the time window.

With the histogram and the quality measures, the requisites that form a purchase occasion were defined. In order to cover this problem, other ideas were also explored to group in clusters like k-means: an iterative clustering algorithm that uses the distance to the centroids to gather similar simulations in k clusters. The validation method silhouette (see Rousseeuw, 1987) was also used to set the optimal number of clusters through an evaluation of the cohesion within the cluster and the separation with others. It was concluded this method did not meet the requirements because even though it grouped based upon amounts, terms, interest rates and dates, did not keep the temporary order of the simulations as seen in Figure 4.6.

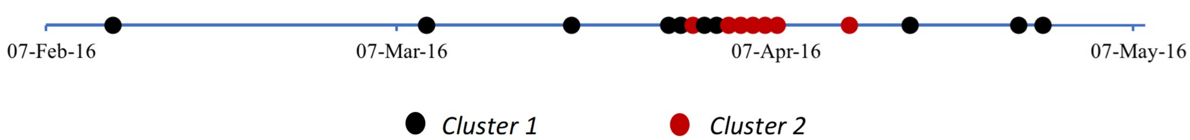


Figure 4.6: Cluster with k-means algorithm.

Our data shows that originally we had a sample of 45,940 simulations to run the propensity model. After applying the algorithm, they were grouped in 22.990 clusters and changed the purchase/simulation ratio from 11.32% to 22.63%. This clustering method was repeated in many different samples to test our model and clusters represented approximately half of the simulations.

4.1.2 Matching Problem

To estimate the customers' elasticity, it was necessary to identify the simulations purchased/non-purchased. The client's decision and credit characteristics were fundamental to understand the purchasing behavior .

The problem called "match problem", consisted in being able to identify which simulation corresponds to the purchased credit. Figure 4.7 represents the situation. If the match was found, the simulation was marked with a 1, otherwise 0.

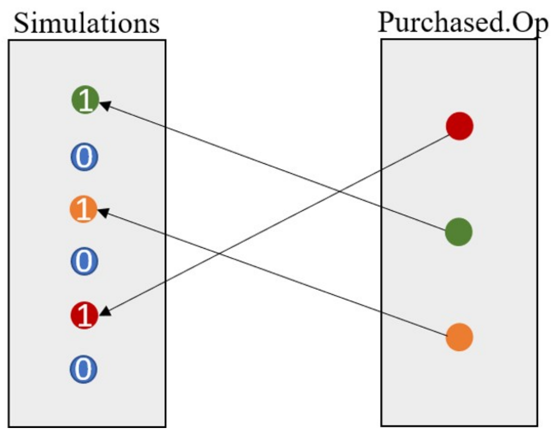


Figure 4.7: Simulations are in the left table whereas its counterparty, with the same characteristics, is in the purchased operation's table. In case both sides are matched, the simulation is marked with a tick stating it ended in purchase.

The data did not have a serial code linking a simulation with a purchased operation, but presented the following patterns:

- If the client simulated and took the credit, it came out in the purchased operations table the same day as long as it takes place before 8pm, otherwise, it was informed the next working day.
- If the purchase took place on a saturday, sunday or an official holiday it was informed the following working day, no matter the time.

  Example: Imagine a client that simulates on a friday night at 9pm and chooses to keep/buy the credit. The simulations' table will show the characteristics of the credit, time and hour at that exact moment. Although, the purchased operations' table will report the same credit characteristics with monday as the take-up date.

Thus, a couple of quick points were considered:

- Only the last simulation of the day and the most advanced stage the client reached were considered. The assumption was that the customer will not continue simulating the same day after taking up a credit.
- Date and schedule change were considered to perform the matching.

After these adjustments both tables were joined with common variables: date and client's ID.

There were some situations where official holidays or simulator errors caused bigger gaps on reported purchased operations. To solve this, a 14 day[2] range of difference in between the simulation and the purchased operations date was set (example shown in table 4.1). If the client simulated x days before the take-up, the same number of matchs will be generated. To prevent this, the pair with the minimum difference was matched.

[2]Determined from client' activity study, see chapter 4.1.1

The difference represented by the greek letter α for the client j was defined as:

$$\alpha = Purchased Op.Date_j - Simulation Date_j \leq 14$$

If $\alpha \geq 0$, thus must satisfy:

$$Paired\, match = \min \alpha$$

Table 4.1: Matching example

Table A, client j

Simulations	Simulation date	α
s.1	28-06-2017	null
s.2	11-06-2017	4
s.3	**15-07-2017**	**0**
s.4	20-12-2017	null

Table B, client j

Purchased operations	Date of purchase
p.1	**15-07-2017**

$$Paired\, match = \{s.3, p.1\}$$

First column of Table A shows each simulation's ID, second column the simulation date and the third column difference between the purchased operation and the simulation date. The value α in the first simulation is null because difference between both dates is larger than 14 days. The minimum α in table A is zero which corresponds to the third simulation, so we match s.3 and p.1.

Once the match was performed, results showed that there were purchased operations that could not be matched because of a gap produced by a web simulator error. To avoid problems and leave data as clean as possible, those client Id's with an error were excluded deleting all their simulations and keeping away a potential bias on the model estimates.

Let's suppose in Figure 4.8 we have a client with ID=15, the database shows three purchased operations, although, there's one that could not be matched with a simulation. As a result, the client with ID=15 would be deleted from the data.

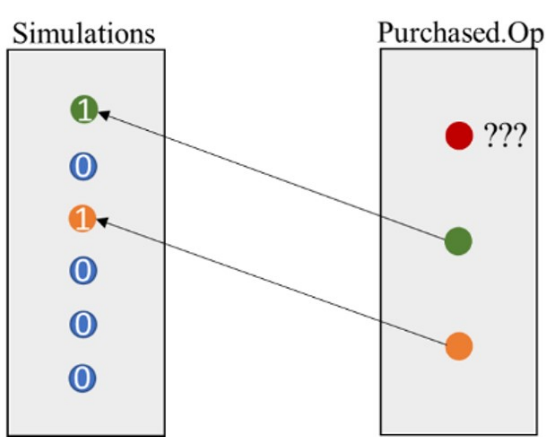


Figure 4.8: Matching error representation.

Table 4.2: Summary statistics for continuous variables

Statistic	N	Mean	St. Dev.	Min	Max
Amount	131,907	$5,185,900	$6,298,833	$100,000	$80,000,000
Interest rate	131,907	1.450%	0.497%	0.500%	3.100%
Term	131,907	28.369	16.871	4	80
Projected risk rate	129,188	1.4%	2.9%	0.2%	46.7%
Estimated Income	122,309	$5,224	$107,621	$333	$6,317,133
Cost of Funds	127,711	3.75%	0.49%	2.73%	5.09%

From a total of 92.068 online purchased operations, 59.074 remained after applying a matching filter.

4.2 Endogeneity of price

Once the cluster was defined as our unit of analysis, we tested whether we had enough control variables in the logit model. A linear regression was performed with the interest rate as the dependent variable and controls were added. If a large percentage of the dependent variable's variation is explained with our controls, means we are on the right path. If not, more explanatory variables will be needed since not accounting for endogeneity could lead to substantial mispricing. This test helps to identify the importance of each variable in the interest rate's definition and provides insights on the bank's pricing.

4.2.1 Linear regression model

Table 4.2 lists the continuous variables used to estimate the interest rate. The bank categorized online applications into tiers based on default risk, term, amount and income. Riskiness increases with tier i.e., tier 1 online applications were considered the least risky.

A regression was run on the simulations' table applying the bank's segmentation by tiers. Then, the coefficients of the model were estimated based on historical data.

The data set used was the same in the pilot model[3], but this time all the income segments and levels of risk were considered.

Model specification:

$$R_j = \alpha + \sum_w \beta_w Term_{\text{tier}_w} + \sum_x \beta_x Amount_{\text{tier}_x} + \sum_y \beta_y Income_{\text{tier}_y} + \sum_z \beta_z \text{Risk}_{\text{tier}_z} \quad (4.1)$$

where R is the interest rate for the transaction j and w, x, y, z ≥ 1 represent the number of segments for each variable. The model was estimated using standard OLS regression.

[3]See section 3.1.2 for details

The explanatory variables included in this model are dummies indicating the tiers from each specific category.

Most of the credit simulations fell into a relatively small number of terms (specified in months). As stated earlier, there were many 24, 36 or 60-month loans, but no 43-month or 61-month loans. Following the bank's applied segmentation, a categorical variable was created with possible values (0-12], (12,24], (24,36], (36,48], (48,60] or $>$60.

The categorical variable amount was used to group the loans in 9 tiers; (0-0.349MM], (0.349-0.499MM], (0.499-0.799MM], (0.799-1.499MM], (1.499,2.999MM], (2.999-6.999MM], (6.999-9.999MM], (9.999-19.999MM] and $>$19.999MM.

The variable *income* indicates the income segment of each client. Our data set shows there are four groups: Massive/Emerging/University, Middle/Young Professional, High and Preferential, ordered from lowest to the largest income. According to the bank, each segment has a unique pricing matrix in order to capture the different elasticities given by the customer characteristics. The 5 risk tiers mentioned in section 3.2 were included in the model.

Regression results from Table 4.3 are found to be intuitive. Larger amounts and terms lead to a lower interest rate, whereas higher levels of risk imply a significantly higher interest rate. These results suggest that riskier clients are a problem for the bank because of their probability of default.

The adjusted R^2 represents the percentage of variation explained by the independent variables that affect the dependent variable. As shown in Table 4.3, $R^2 = 0.73$, which means the terms selected fit well the data.

Overall, the coefficient signs make sense and the income variables suggest that a bigger income leads to lower interest rates. Hence, the preferential segment gets a better deal with smaller interest rates compared to the base level (massive/emerging/university) or lower income segments.

A priori, this finding might indicate that more liquidity constraint segments. Thus, younger and lower income households are less price sensitive compared to the consumer groups that do not seem likely to be candidates for liquidity constraints (e.g., high income, middle age consumers). So far, it is just a hypothesis, but it will be tested later with the logit model measuring the elasticity of demand for each income segment.

Table 4.3: Linear regression results

Explanatory variable	Estimate	Conf. Interval
Term:		
Tier 2 (13-24)	-0.160^{***}	$(-0.163, -0.156)$
Tier 3 (25-36)	-0.143^{***}	$(-0.148, -0.138)$
Tier 4 (37-48)	-0.129^{***}	$(-0.135, -0.123)$
Tier 5 (49-60)	-0.085^{***}	$(-0.091, -0.079)$
Tier 6 (>60)	-0.074^{***}	$(-0.099, -0.050)$
Amount:		
Tier 1 (<0.349 MM)	0.527^{***}	(0.516, 0.538)
Tier 2 (0.349-0.499.MM)	0.496^{***}	(0.482, 0.509)
Tier 3 (0.499-0.799.MM)	0.466^{***}	(0.456, 0.477)
Tier 4 (0.799-1.499.MM)	0.385^{***}	(0.375, 0.394)
Tier 5 (1.499-2.999.MM)	0.292^{***}	(0.283, 0.301)
Tier 6 (2.999-6.999.MM)	0.111^{***}	(0.102, 0.119)
Tier 7 (6.999-9.999.MM)	0.049^{***}	(0.040, 0.058)
Tier 8 (9.999-19.999.MM)	0.040^{***}	(0.032, 0.048)
Income:		
Tier 2 (Middle\Young.Prof)	-0.158^{***}	$(-0.163, -0.154)$
Tier 3 (High)	-0.368^{***}	$(-0.373, -0.364)$
Tier 4 (Preferential)	-0.655^{***}	$(-0.661, -0.648)$
Risk:		
Tier 2 (0.8%-1.8%)	0.105^{***}	(0.102, 0.109)
Tier 3 (1.8%-3.8%)	0.456^{***}	(0.451, 0.461)
Tier 4 (3.8%-8%)	0.838^{***}	(0.831, 0.844)
Tier 5 (>8%)	1.029^{***}	(1.019, 1.040)
Intercept	1.500^{***}	(1.490, 1.510)
Observations		129,188
Adjusted R^2		0.732
F Statistic		$17{,}674.420^{***}$ (df = 20; 129167)

Standard errors in parentheses
Note: Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Term tier base level: (<12), Amount tier base level: (>19.999 MM), Income tier base level: (Massive-Emerging-University), Risk tier base level: (<0.8%)

4.3 Price-Response model

Our analysis is based on a data set of online consumer credit simulations from a Chilean bank. Two income segments were used: Middle/Young Professional and High, with terms ranging from 13-months to 60-months and amounts from 3 MM above. The data were clustered in

22.297 purchase occasions of which 5.017 (22.5%) were taken up. The main interest is to predict how the probability of take-up changes as a function of the interest rate. Therefore, a logit regression is used on the data with take-up as the target variable.

4.3.1 Ways of modelling the offer

The idea of associating a purchase occasion to a cluster is being able to determine which was the offer the client saw in the group of simulations upon whether chooses to buy or not. Having said that, different ways were proposed to find the right combination of interest rate, amount and term that best fits the search process:

1. {Purchase, Average}: In case there is a purchase in the cluster, the combination of interest rate, amount and term of the credit chosen represents the offer. If there are no purchases the average of the interest rate, amount and term for all the simulations within the cluster are taken as the representative offer.
   In figure 4.9 a temporary line with credit simulations for a specific client is shown. Given the cluster formation rules, we identify two cluster occasions: in the first case, a simulation ends up in a purchase, thus it represents the offer seen by the client. For the second case, there are no purchases, so the offer is going to be represented as an average of the characteristics from all the simulations within the cluster.

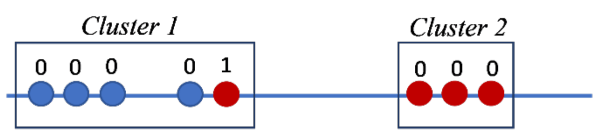


Figure 4.9: Offer in case of purchase (cluster1) and non-purchase (cluster2). For both situations the offer is represented in color red.

2. {Average, Average}: In case there's a purchase, the average of the interest rate, amount and term from all the simulations in the cluster represent the offer. In case of non-purchase the average of the interest rate, amount and term from all the simulations within the cluster represent the offer.
3. {Purchase, Last simulation}: In case of purchase, the combination of interest rate, amount and term of the credit chosen represent the offer. In case of non-purchase, the combination of interest rate, amount and term of the last credit simulated in the cluster represent the offer.

The model was tested with each offer option presented above, different variable transformations and segmentations were performed as well as crossing the rate variable with other variables (e.g., term, amount, risk, income segments). A combination of forward inclusion and backward elimination techniques[4] were used according to the AIC or BIC criteria and

[4]Stepwise selection R software

purchase occasions were re-clustered depending on amount segments and terms. The quality prediction was measured with indicators like the pseudo r-squared, sensitivity, accuracy, ROC and AUC (concordance). The model elected had the highest predictive performance and significance levels. Later, it was used in the bank to compare it with the existing model.

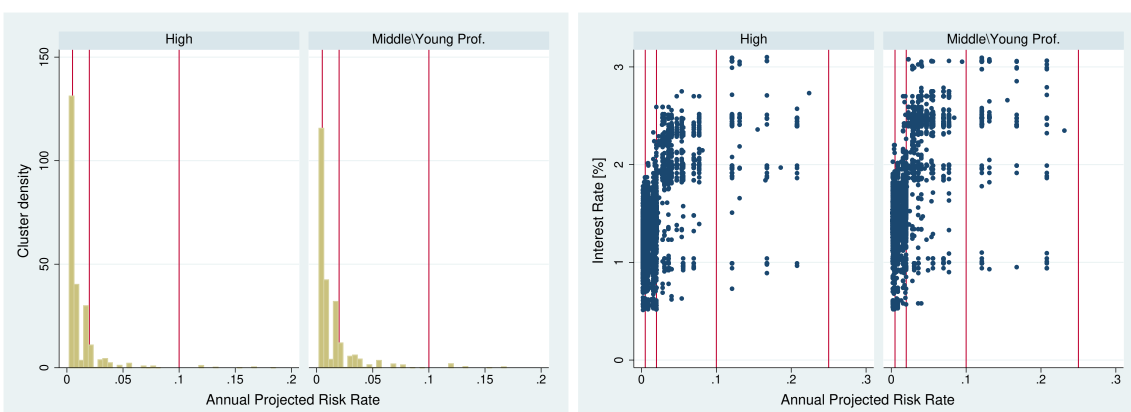
According to literature, segmentation is necessary to better capture price elasticities. Phillips (2013) states that sensitivity is different for different segments: borrowers applying for larger loans are more price-sensitive than those applying for smaller loans and higher risk customers tend to be less price sensitive than lower risk customers. Therefore, to estimate the Price-Response model we performed a segmentation considering credit and client characteristics like the amount, term, risk and income by tiers.

The variable TMC was included in the model and indicates the maximum interest rate to charge in consumer credits depending on the size of the loan. Also, the *Industry Rate* indicates the average interest rate for all the credits in the country depending on the size of the loan[5]. Once a month the SBIF publishes both values and prevail for a 30-day period.

Six dummies *AmountClass* were used to categorize cluster amounts in order to capture different elasticities among them. Equal observations and previous loan size bank segmentation helped defining the following tiers: (3-4.5MM], (4.5-5.5MM], (5.5-7MM], (7-10MM], (10-15MM] and >15MM.

Based upon the same argument, a categorical variable called *TermClass* was created with possible values [13-24], [25-36], [37-48] and [49-60].

Overall, clients taking online consumer credits are characterized for having a low-risk profile (See Figure 4.10). Therefore, the bank's risk segmentation was re-defined to better capture the population's risk distribution. Three risk tiers were used to categorize clients: [0-0.05%], (0.05%-0.2%], (0.2%-1%], Tier 1 applications were considered least risky and Tier 3 the riskiest ($RiskTier_i$).



(a) Risk histogram grouped by clusters.

(b) Risk scatter plot.

Figure 4.10: Population's risk distribution for the High and Middle/Young Professional-income segments

[5]The SBIF generates a segmentation based on the size of the loan (division in three tiers) and then sets for each one of them a different TMC and Industry Rate value.

To capture the relationship between riskiness and interest rate, we crossed both for each level of risk ($RiskTier_i \times InterestRate$).

$Amount_{>8MM} \times InterestRate$ indicates whether higher loan sizes (above 8MM), lead to a different price elasticity.

To differentiate between the two income segments, two categorical variables were created: *Middle/Young Professional* and *High*. Also, an interaction of income and interest rate is included Middle/Young Professional × InterestRate. Note the dependent variable TakeUp indicates whether a customer took up a loan (Y=1) or not (Y=0).

The probability of a customer taking up a loan is expressed as the following response function obtained from fitting the logistic regression model:

$$TakeUp\_Probability = \frac{e^u}{1 + e^u} \tag{4.2}$$

$$\begin{aligned} u = {} & \alpha + \beta_1 AmountClass_i + \beta_2 Amount_{8MM} \times InterestRate + \beta_3 RiskTier_i \times InterestRate \\ & + \beta_4 RiskTier_i + \beta_5 TermClass_i + \beta_6 TMC + \beta_7 IndustryRate + \beta_8 Middle/YoungProf. \\ & + \beta_9 Middle/YoungProf. \times InterestRate \end{aligned}$$

Figure 4.11 provides the results expected from the response graph of price vs. take-up probability. An increase in price leads to lower take-up rates showing that price elasticity exists.

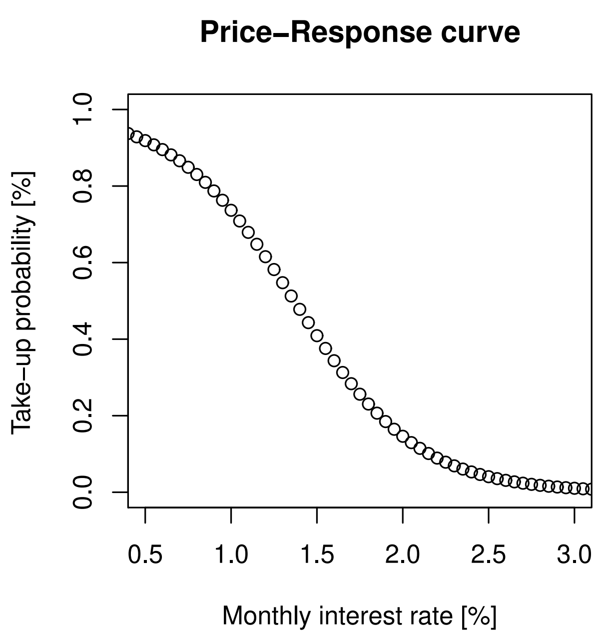


Figure 4.11: Response graph of price vs take-up.

4.3.2 Propensity model results

As shown in Table 4.4, $AmountClass$ is positive and decreasing with the size of the loan (note the $AmountClass$>15MM is the base level), consistent with the expectation that, all else being equal, take-up rates are smaller for larger loans. Plus, the coefficient of the dummy variable $Amount_{8MM} \times InterestRate$ indicates that people taking credits over 8 MM are more price elastic than people taking small credits.

On the other hand, the interaction between risk and interest rate is negative (i.e., higher interest rates lead to lower take-up), which makes sense. Although, the most interesting fact is that clients with higher levels of risk are less sensitive to the interest rate, compare marginal effects $Risk_{\text{Tier1}} \times InterestRate$= -0.27 vs. $Risk_{\text{Tier2}} \times InterestRate$ = -0.17. According the bank's interpretation, lower-income clients present more borrowing constraints and overall, banks offer them worse credit conditions because of their risk. Plus, they are less bancarized and therefore have lower bargaining power.

Our results are also consistent with the hypothesis that lower income segments are less price elastic compared to higher income segments. Attanasio et al. (2008), hereinafter ("AGK") state that while the demand for loans is sensitive to the interest rate, the interest rate sensitivity is largest for older rather than younger consumers, and for consumers with relatively large current income. For consumers who are not liquidity constrained, loan demand is solely a function of price of the loan (the interest rate), while liquidity constrained consumers also respond to maturity changes.

Karlan & Zinman (2005) find that price sensitivity increases with income, which is consistent with the presence of liquidity constraints that decrease with income. Their results show that maturity elasticity is significant for low-income, but not high-income borrowers. Although, higher income borrowers present the strongest interest rates effects. They also state that an explanation to liquidity constraints is that income and age proxy for financial sophistication i.e., the poor and inexperienced use a decision rule that lead them to focus on monthly payments rather than interest rates.

The marginal effects at means were also calculated:

$$\frac{1}{n}\sum_{\text{i}=1}^{n} p(\text{i}) * \big(1 - p(\text{i})\big) * \beta \tag{4.3}$$

where n is the number of observations $p(\text{i})$ the take-up probability and β is the coefficient.

Table 4.5 shows the marginal effects for each risk segment. As expected, riskier clients tend to be less price sensitive than low-risk clients. The marginal effect of increasing the interest rate in 10 basis points for the risk Tier 1, leads to a take-up decrease of 2.68% (E.g., if the bank increases the interest rate from 1.2% to 1.3%, the take-up probability should decrease from 40% to 37.2%).

Table 4.4: Choice model (logit) estimates for the online bank data

Variable	Estimate	Conf.Interval
$AmountClass_{3-4.5MM}$	1.877***	(1.549, 2.205)
$AmountClass_{4.5-5.5MM}$	1.296***	(0.982, 1.609)
$AmountClass_{5.5-7MM}$	1.042***	(0.778, 1.306)
$AmountClass_{7-10MM}$	0.691***	(0.467, 0.914)
$AmountClass_{10-15MM}$	0.516***	(0.287, 0.745)
$Amount_{>8MM}$ × InterestRate	−0.179*	(−0.318, −0.040)
$Risk_{\text{Tier1}}$ × InterestRate	−1.684***	(−1.974, −1.394)
$Risk_{\text{Tier2}}$ × InterestRate	−1.083***	(−1.374, −0.792)
$Risk_{\text{Tier3}}$ × InterestRate	−0.762***	(−1.117, −0.407)
$TermClass_{25-36}$	0.033	(−0.047, 0.113)
$TermClass_{37-48}$	−0.058	(−0.156, 0.039)
$TermClass_{49-60}$	−0.756***	(−0.868, −0.645)
$Risk_{\text{Tier2}}$	−0.417 ·	(−0.850, 0.016)
$Risk_{\text{Tier3}}$	−0.655 ·	(−1.356, 0.046)
TMC	0.085 ·	(−0.008, 0.178)
Middle\Young.Prof	−0.281	(−0.653, 0.091)
Middle\Young.Prof × InterestRate	0.228 ·	(−0.040, 0.497)
Industry rate	−0.082*	(−0.147, −0.017)
Intercept	−0.923	(−2.200, 0.354)
Observations		22,297
Log Likelihood		−11,013.910
Akaike Inf. Crit.		22,065.820

Standard errors in parentheses
Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Term tier base level: (13-24), Amount tier base level: (>15 MM), Income tier base level: (High), Risk tier base level: (0-0.05%)

Table 4.5: Marginal effects for the logit model

Marginal Effects					
Variable Name	df/dx	Std.Err	Z	P > |z|	
$AmountClass_{3-4.5MM}$	0.35	0.03	10.39	<2.2e-16	***
$AmountClass_{4.5-5.5MM}$	0.26	0.04	7.07	1.511e-12	***
$AmountClass_{5.5-7MM}$	0.20	0.03	6.78	1.189e-11	***
$AmountClass_{7-10MM}$	0.12	0.02	5.50	3.759e-08	***
$AmountClass_{10-15MM}$	0.09	0.02	4.02	5.792e-05	***
$Amount_{>8MM}$ × InterestRate	-0.03	0.01	-2.53	0.01151	*
$Risk_{\text{Tier1}}$ × InterestRate	-0.27	0.02	-11.43	<2.2e-16	***
$Risk_{\text{Tier2}}$ × InterestRate	-0.17	0.02	-7.30	2.831e-13	***
$Risk_{\text{Tier3}}$ × InterestRate	-0.12	0.03	-4.21	2.533e-05	***
$TermClass_{25-36}$	-0.01	0.01	0.80	0.42365	
$TermClass_{37-48}$	-0.01	0.01	-1.19	0.23548	
$TermClass_{49-60}$	-0.11	0.01	-15.19	<2.2e-16	***
$Risk_{\text{Tier2}}$	-0.06	0.03	-1.97	0.04886	*
$Risk_{\text{Tier3}}$	-0.09	0.04	-2.24	0.02526	*
TMC	0.01	0.01	1.79	0.07309	
Middle\Young.Prof	-0.04	0.03	-1.54	0.12402	
Middle\Young.Prof × InterestRate	0.04	0.02	1.66	0.09572	
Industry rate	-0.01	0.01	-2.46	0.01356	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Two price-response curves for different risk tiers are compared to represent graphically the effect of risk on price elasticity (see Figure 4.12). In particular, two clients are characterized from the high-income segment with risk tiers 1 and 3, where 1 is low risk and 3 is high risk.

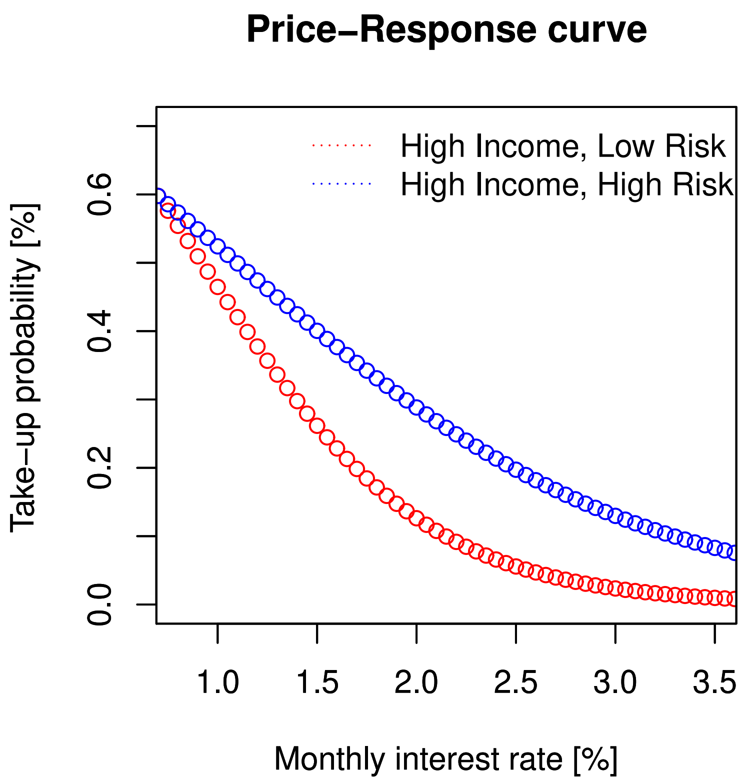


Figure 4.12: Response graph of price vs. take-up for different risk categories

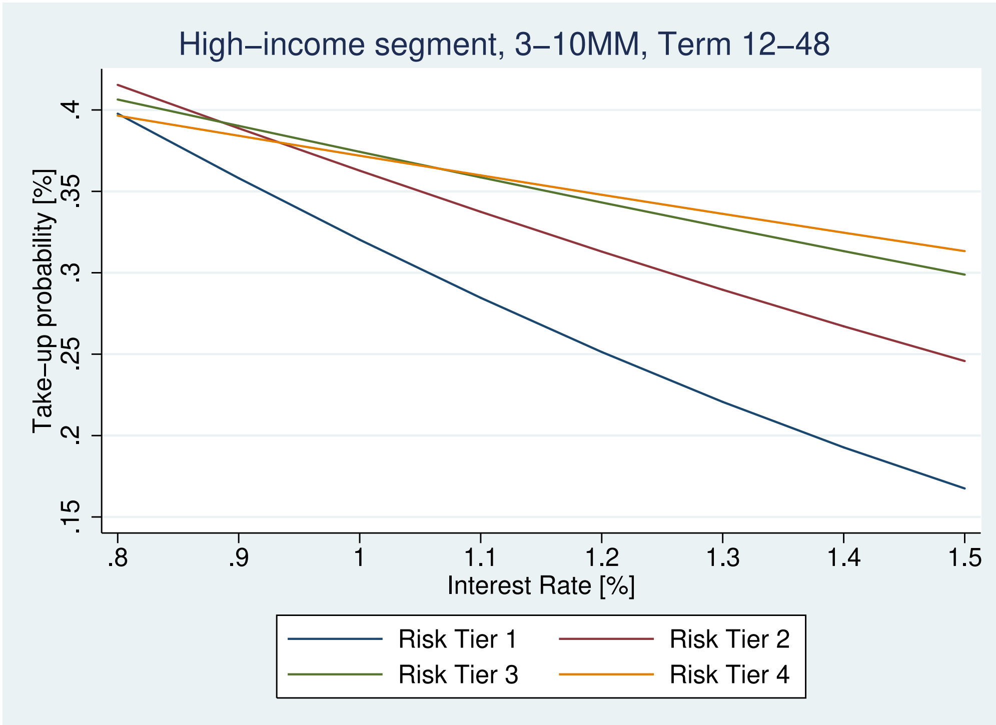
Note that this thesis specifically focused on the Middle/Young professional and the High-income segment in order to test them with a pilot, but we also performed several propensity models for the remaining segments: Massive-Emerging-University & Preferential. These two price-response curves also support the hypothesis that higher income segments are more price responsive than lower income segments. The elasticity ranking is as follows (from 1 to 4, where 4 is the most elastic):

Table 4.6: Customers from the massive/emerging/university income-segment are the least price-sensitive, whereas customers from the preferential-income segment are the most elastic to changes in the interest rate

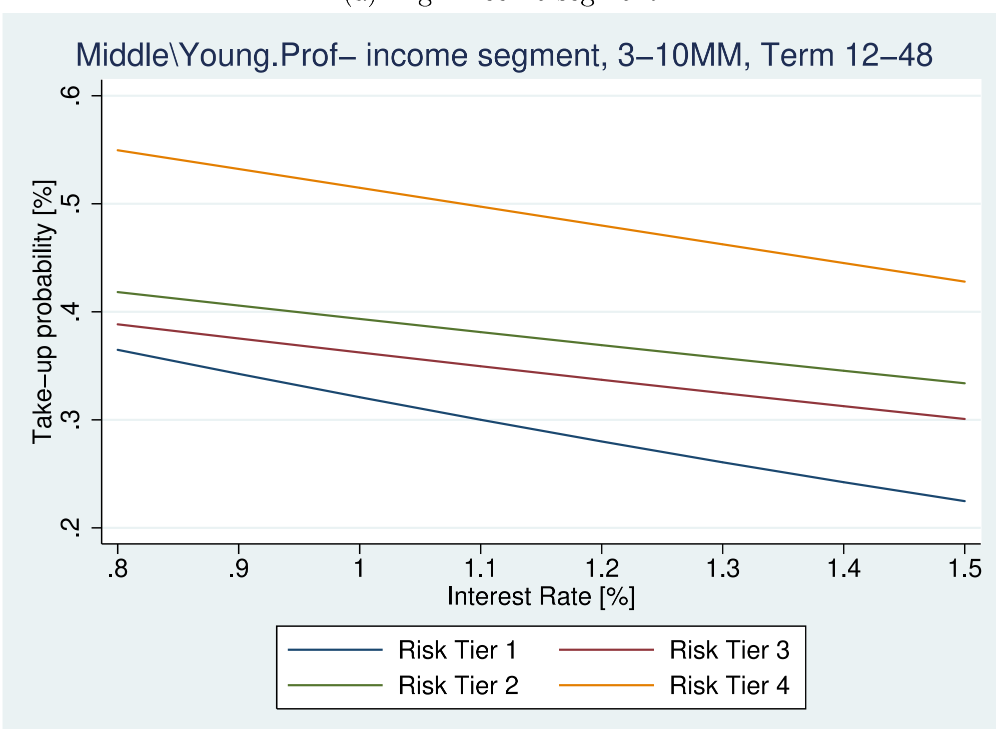
Elasticity ranking by income segment	
No.1	Massive/Emerging/University
No.2	Middle/Young Professional
No.3	High
No.4	Preferential

To illustrate with examples from the data the relationship between interest rate and take-up probability, we took a sample of the simulations with term between 12-48, amount between

3 to 10 MM and the income segments high and middle/young professional, see Figure 4.13. Pooled results are shown in Table 4.14



(a) High-income segment.



(b) MYP-income segment.

Figure 4.13: Take-up probability under different risk categories for the High-income & Middle/Young Professional-income segment. Riskier customers tend to be less price-sensitive and have higher overall take-up rates

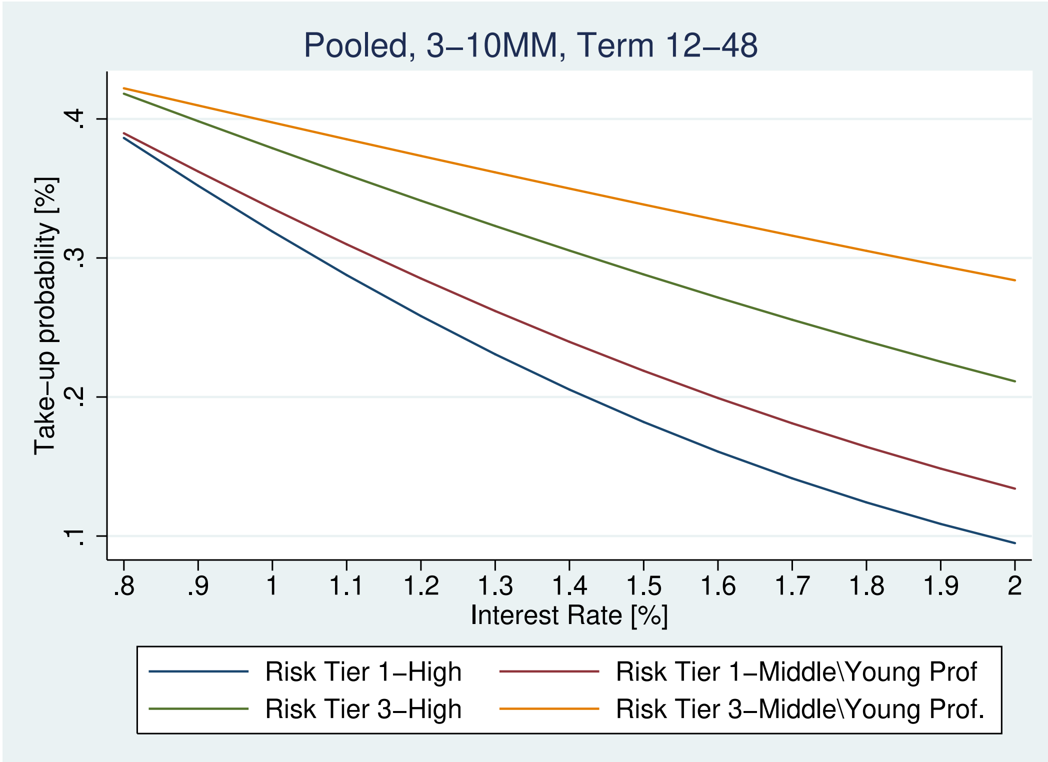


Figure 4.14: Pooled results

4.3.3 Re-estimation of the propensity model

The main finding after the analysis of the propensity model results is that a simulation step was excluded from the data which contained only zeros, i.e., clients entered to the web platform, simulated and did not buy the credit. To correct it, the simulation step was added, and the historical coefficients were re-calculated. As expected, observations almost duplicated due to the simulations included in the sample, the historical observations increased from 22,297 to 39,072 Table 4.7 shows the estimated coefficients for the historical model and Table 4.8 the marginal effects.

To re-estimate this model, two income segments were used: Middle/Young Professional and High, with terms ranging from 13-months to 60-months and amounts from 3MM above. The data set was clustered in 39.072 purchase occasions of which 5.275 (13.5%) were taken up. The variables used were the same than the previous model.

Table 4.7: Re-estimated coefficients using historical data and adding simulation step

Variable Name	Estimates	Conf.Interval
$AmountClass_{3-4.5MM}$	1.356^{***}	(1.056, 1.657)
$AmountClass_{4.5-5.5MM}$	0.709^{***}	(0.420, 0.998)
$AmountClass_{5.5-7MM}$	0.837^{***}	(0.583, 1.091)
$AmountClass_{7-10MM}$	0.551^{***}	(0.332, 0.770)
$AmountClass_{10-15MM}$	0.556^{***}	(0.332, 0.781)
$Amount_{>8MM}$ × InterestRate	0.248^{***}	(0.118, 0.379)
$Risk_{\text{Tier1}}$ × InterestRate	-1.847^{***}	(−2.113, −1.581)
$Risk_{\text{Tier2}}$ × InterestRate	-1.372^{***}	(−1.639, −1.104)
$Risk_{\text{Tier3}}$ × InterestRate	-0.995^{***}	(−1.308, −0.681)
$TermClass_{25-36}$	0.317^{***}	(0.245, 0.388)
$TermClass_{37-48}$	0.404^{***}	(0.315, 0.493)
$TermClass_{49-60}$	0.091	(−0.020, 0.201)
$Risk_{\text{Tier2}}$	−0.235	(−0.631, 0.162)
$Risk_{\text{Tier3}}$	$-0.525^{\cdot}$	(−1.141, 0.091)
TMC	0.101^{*}	(0.020, 0.182)
Middle\Young.Prof	−0.171	(−0.504, 0.162)
Middle\Young.Prof × InterestRate	0.112	(−0.129, 0.354)
Industry Rate	−0.040	(−0.097, 0.017)
Intercept	-2.729^{***}	(−3.856, −1.602)
Observations		39,072
Log Likelihood		−14,961.370
Akaike Inf. Crit.		29,960.750

Standard errors in parentheses
Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Term tier base level: (13-24), Amount tier base level: (>15 MM), Income tier base level: (High), Risk tier base level: (0-0.05%)

Table 4.8: Marginal effects for the logit model using historical data and step1

Marginal Effects					
Variable Name	df/dx	Std.Err	Z	P > |z|	
$AmountClass_{3-4.5MM}$	0.17	0.02	7.63	2.339e-14	***
$AmountClass_{4.5-5.5MM}$	0.09	0.02	4.16	3.084e-05	***
$AmountClass_{5.5-7MM}$	0.11	0.02	5.44	5.043e-08	***
$AmountClass_{7-10MM}$	0.06	0.01	4.35	1.327e-05	***
$AmountClass_{10-15MM}$	0.07	0.01	4.16	3.055e-05	***
$Amount_{>8MM}$ × InterestRate	0.02	0.01	3.73	0.0001848	***
$Risk_{\text{Tier1}}$ × InterestRate	-0.20	0.01	-13.74	< 2.2e-16	***
$Risk_{\text{Tier2}}$ × InterestRate	-0.15	0.01	-10.07	< 2.2e-16	***
$Risk_{\text{Tier3}}$ × InterestRate	-0.10	0.01	-6.22	4.836e-10	***
$TermClass_{25-36}$	0.03	0.01	8.27	< 2.2e-16	***
$TermClass_{37-48}$	0.04	0.01	8.09	5.652e-16	***
$TermClass_{49-60}$	0.01	0.01	1.56	0.1169668	
$Risk_{\text{Tier2}}$	-0.02	0.02	-1.19	0.2320413	
$Risk_{\text{Tier3}}$	-0.04	0.02	-1.99	0.0461136	*
TMC	0.01	0.01	2.43	0.0149767	*
Middle\Young.Prof	-0.01	0.01	-1.02	0.3044544	
Middle\Young.Prof × InterestRate	0.01	0.01	0.91	0.3615603	
Industry rate	-0.01	0.01	-1.38	0.1664157	

Once the elasticities were calculated for the historic data, the optimized interest rates were computed and a new pricing matrix was obtained. See tables 4.9 and 4.10 for the high and MYP income segments. Even though interest rate variation between the different risk tiers decreased, some inconsistencies were found. For instance, higher interest rates were charged while the credit amount increased, which is not logical. Other pitfalls are that pricing matrices for both income segments look almost the same. It was already proved that the MYP income segment was less elastic and the pilot test's results confirm it. Thus, the MYP matrix should charge on average, higher interest rates, but the model is not consistent with this finding.

Table 4.9: High-income segment pricing matrix

High-Income segment, optimal interest rates												
	Risk Tier 1 (<0.005)				Risk Tier 2 (0.005-0.02)				Risk Tier 3 (0.02-0.05)			
	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60
[3 , 4.5 MM]	1.078	1.131	1.152	1.127	1.382	1.446	1.472	1.442	1.763	1.819	1.842	1.819
(4.5 , 5.5 MM]	1.029	1.069	1.086	1.072	1.323	1.368	1.384	1.374	1.692	1.740	1.771	1.721
(5.5 , 7 MM]	1.020	1.058	1.074	1.064	1.311	1.350	1.374	1.361	1.682	1.729	1.739	1.722
(7 , 10 MM]	1.070	1.100	1.119	1.120	1.415	1.445	1.469	1.466	1.834	1.872	1.923	1.902
(10 , 20 MM]	1.100	1.136	1.151	1.137	1.473	1.514	1.532	1.511	1.953	1.994	2.004	1.979

Table 4.10: MYP-Income segment pricing matrix

MYP-Income segment, optimal interest rates												
	Risk Tier 1 (<0.005)				Risk Tier 2 (0.005-0.02)				Risk Tier 3 (0.02-0.05)			
	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60
[3 , 4.5 MM]	1.096	1.147	1.167	1.144	1.421	1.484	1.508	1.474	1.823	1.886	1.912	1.876
(4.5 , 5.5 MM]	1.050	1.088	1.105	1.094	1.363	1.411	1.433	1.406	1.740	1.803	1.819	1.786
(5.5 , 7 MM]	1.042	1.079	1.093	1.085	1.348	1.396	1.416	1.396	1.720	1.780	1.805	1.788
(7 , 10 MM]	1.076	1.107	1.123	1.133	1.427	1.439	1.472	1.498	1.943	1.822	2.022	2.042
(10 , 20 MM]	1.133	1.168	1.182	1.172	1.524	1.555	1.597	1.586	2.096	2.121	2.114	2.115

These findings led us to a new model that included more interactions between variables and will be explained in subsection 4.4.

4.4 Final model

Based upon the last model 4.7, six dummies (AmountClass) were used to categorize cluster amounts into the following tiers: (3-4.5MM], (4.5-5.5MM], (5.5-7MM], (7-10MM], (10-15MM] and >15MM.

A categorical variable called TermClass was created with values [13-24], [25-36], [37-48] and [49-60].

Given the big interest rate variation within risk tiers it was decided to shorten the risk intervals for the estimation. Increasing the number of risk tiers will lead to more accurate elasticities and homogeneous groups. Therefore, the mean interest rate statistic will improve. Six risk tiers were created; [0-0.05%], (0.05%-0.1%], (0.1%-1.8%], (1.8%-3.8%], (>3.8%).

Where lower values are considered least risky. This time the industry rate was omitted and the TMC remained due to high correlation between variables.

To capture the relationship between riskiness and interest rate, we crossed both for each level of risk $Risk_{\text{Tieri}}$ × InterestRate. Two categorical variables were created: Middle/Young Professional and High. Also, an interaction of income and interest rate was included Middle\Young.Prof × InterestRate to differentiate elasticities between income segments.

Last variation in the model is an interaction between the income segment and level of risk Middle\Young.Prof × $Risk_{\text{Tier}_\text{i}}$. According to the AB testing results, elasticities among low levels of risk were similar when comparing the High and MYP-income segments, but different at higher levels of risk. Including this variable will capture those changes.

4.4.1 Results

As seen in table 4.11 coefficient signs are intuitive. All the coefficients for $Risk_{\text{Tieri}}$ × InterestRate are negative, indicating that higher interest rates lead to lower take-up, as expected. Also, the same variable shows that with an increase in the level of risk, price-elasticity should decrease. Table 4.12 shows that risk tier 1 is more price-elastic than risk tier 2 and 3, but from tiers 3 through 6 we obtain the same marginal effects, meaning that from a certain risk profile price-elasticities remain stable.

TermClass coefficients tell us that is less probable to take-up long credits (49-60 months), while 37-48-month credits are the most preferred among clients.

Finally, the variable Middle\Young.Prof × InterestRate is positive indicating a less price sensitive segment compared to the high-income clients. As a result, higher interest rates in the MYP-income price matrix should be seen.

Table 4.11: Final model coefficients

Variable Name	Estimates	Conf.Interval
$AmountClass_{3-4.5MM}$	1.462^{***}	(1.206, 1.718)
$AmountClass_{4.5-5.5MM}$	0.458^{***}	(0.213, 0.703)
$AmountClass_{5.5-7MM}$	0.591^{***}	(0.379, 0.804)
$AmountClass_{7-10MM}$	0.444^{***}	(0.232, 0.656)
$AmountClass_{10-15MM}$	0.569^{***}	(0.345, 0.794)
$Risk_{\text{Tier1}}$ × InterestRate	-2.015^{***}	(−2.325, −1.705)
$Risk_{\text{Tier2}}$ × InterestRate	-1.675^{***}	(−2.105, −1.246)
$Risk_{\text{Tier3}}$ × InterestRate	-1.184^{***}	(−1.650, −0.718)
$Risk_{\text{Tier4}}$ × InterestRate	-1.182^{***}	(−1.485, −0.880)
$Risk_{\text{Tier5}}$ × InterestRate	-1.157^{**}	(−1.915, −0.400)
$Risk_{\text{Tier6}}$ × InterestRate	-1.180^{***}	(−1.585, −0.774)
$TermClass_{25-36}$	0.332^{***}	(0.261, 0.404)
$TermClass_{37-48}$	0.434^{***}	(0.344, 0.523)
$TermClass_{49-60}$	0.137^{*}	(0.026, 0.248)
$Risk_{\text{Tier2}}$	−0.065	(−0.636, 0.507)
$Risk_{\text{Tier3}}$	$-0.595^{\cdot}$	(−1.214, 0.025)
$Risk_{\text{Tier4}}$	-0.585^{*}	(−1.104, −0.067)
$Risk_{\text{Tier5}}$	−0.151	(−1.637, 1.336)
$Risk_{\text{Tier6}}$	0.056	(−0.756, 0.868)
TMC	$0.022^{\cdot}$	(−0.001, 0.046)
Middle\Young.Prof	-0.731^{*}	(−1.408, −0.055)
Middle\Young.Prof × InterestRate	$0.250^{\cdot}$	(−0.047, 0.546)
Middle\Young.Prof × $Risk_{\text{Tier1}}$	0.436^{*}	(0.004, 0.868)
Middle\Young.Prof × $Risk_{\text{Tier2}}$	0.315	(−0.130, 0.759)
Middle\Young.Prof × $Risk_{\text{Tier3}}$	0.327	(−0.123, 0.778)
Middle\Young.Prof × $Risk_{\text{Tier4}}$	0.345	(−0.095, 0.784)
Middle\Young.Prof × $Risk_{\text{Tier5}}$	−0.066	(−0.797, 0.665)
Constant	-1.120^{***}	(−1.781, −0.459)
Observations		39,072
Log Likelihood		−14,753.610
Akaike Inf. Crit.		29,563.220

Standard errors in parentheses
Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Term tier base level: (13-24), Amount tier base level: (>15 MM), Income tier base level: (High), Risk tier base level: (0-0.05%)

Table 4.12: Marginal effects for the final model

Marginal Effects					
Variable Name	df/dx	Std.Err	Z	P > |z|	
$AmountClass_{3-4.5MM}$	0.20	0.02	9.04	< 2.2e-16	***
$AmountClass_{4.5-5.5MM}$	0.05	0.01	3.42	0.0006235	***
$AmountClass_{5.5-7MM}$	0.07	0.01	4.80	1.583e-06	***
$AmountClass_{7-10MM}$	0.05	0.01	3.70	0.0002090	***
$AmountClass_{10-15MM}$	0.07	0.01	4.23	2.264e-05	***
$Risk_{\text{Tier1}}$ × InterestRate	-0.21	0.01	-12.82	< 2.2e-16	***
$Risk_{\text{Tier2}}$ × InterestRate	-0.18	0.02	-7.65	1.971e-14	***
$Risk_{\text{Tier3}}$ × InterestRate	-0.12	0.02	-4.98	6.275e-07	***
$Risk_{\text{Tier4}}$ × InterestRate	-0.12	0.01	-7.68	1.534e-14	***
$Risk_{\text{Tier5}}$ × InterestRate	-0.12	0.04	-2.99	0.0027442	**
$Risk_{\text{Tier6}}$ × InterestRate	-0.12	0.02	-5.70	1.162e-08	***
$TermClass_{25-36}$	0.03	0.01	8.62	< 2.2e-16	***
$TermClass_{37-48}$	0.05	0.01	8.58	< 2.2e-16	***
$TermClass_{49-60}$	0.01	0.01	2.34	0.0190975	*
$Risk_{\text{Tier2}}$	-0.01	0.03	-0.22	0.8212560	
$Risk_{\text{Tier3}}$	-0.05	0.02	-2.25	0.0242058	*
$Risk_{\text{Tier4}}$	-0.05	0.02	-2.69	0.0069982	**
$Risk_{\text{Tier5}}$	-0.01	0.07	-0.21	0.8334509	
$Risk_{\text{Tier6}}$	0.01	0.04	0.13	0.8947866	
TMC	0.00	0.00	1.85	0.0636887	.
Middle\Young.Prof	-0.07	0.03	-2.26	0.0233243	*
Middle\Young.Prof × InterestRate	0.02	0.01	1.65	0.0984565	.
Middle\Young.Prof × $Risk_{\text{Tier1}}$	0.05	0.02	1.79	0.0725097	.
Middle\Young.Prof × $Risk_{\text{Tier2}}$	0.03	0.03	1.26	0.2075780	
Middle\Young.Prof × $Risk_{\text{Tier3}}$	0.04	0.03	1.28	0.1994342	
Middle\Young.Prof × $Risk_{\text{Tier4}}$	0.04	0.03	1.37	0.1687441	
Middle\Young.Prof × $Risk_{\text{Tier5}}$	0.00	0.03	-0.18	0.8561965	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Chapter 5

Personalized pricing optimization

The optimization results presented in this section correspond to the model 4.4. It was found that a simulation step was omitted and changed the estimated coefficients (See 4.11). Thus, new optimization results were presented and the pricing matrix took new values. Please note the experiment was implemented with the prices from the model that omitted the simulation step

5.1 Description of the model

The price optimization problem facing the lender is determine which price to charge to each pricing segment in order to maximize expected total profitability. Specifically, for each customer j, we will find the rates R_j that solve:

$$\max_{R_j} \quad TakeUp\_Probability_j(R_j) * \Big(Amount_j * Term_j * (R_j - CostOfFunds_j - Risk_j)\Big)$$

Here, the take-up probability is calculated using the choice model estimated in equation (4.2), which includes the credit and client's characteristics. The *amount* variable indicates the size of the loan j and the *Term* represents the number of monthly payments. The *CostOfFunds* was also added for each credit (expenses the bank has to incur to lend money) along with the *Client's Risk*.

The objective function of this problem calculates the expected profit at the rate R_j and it is unimodal. For every historical transaction, we solved this problem in the R software.

5.2 Negotiated deals

So far, only situations without the presence of negotiation have been considered, i.e., the seller sets a price and the client either takes it or leaves it. There are many cases in which the final price is the product of one or more rounds of negotiation between both sides (e.g., branch office). For those situations in which negotiation is likely, an effective approach should provide guidance regarding an acceptable range of prices.

The price is not the only element of concern when it comes to business, a bank might be willing to offer a lower interest to increase the bank deposits for a specific period of time for example. Same applies for the online pricing channel.

This situation motivated us to set bounds for a guided price negotiation and show them along with the results of this thesis, the basic idea is illustrated in Figure 5.1. The interest rate that maximizes expected profit is 0.949, at which the expected profitability from the credit under consideration is $363.992. A lower interest rate and an upper interest rate have been set so the expected profitability is $360.000 at the lower bound and the upper bound. At any price between both, the expected profit will be within 1% of the optimum.

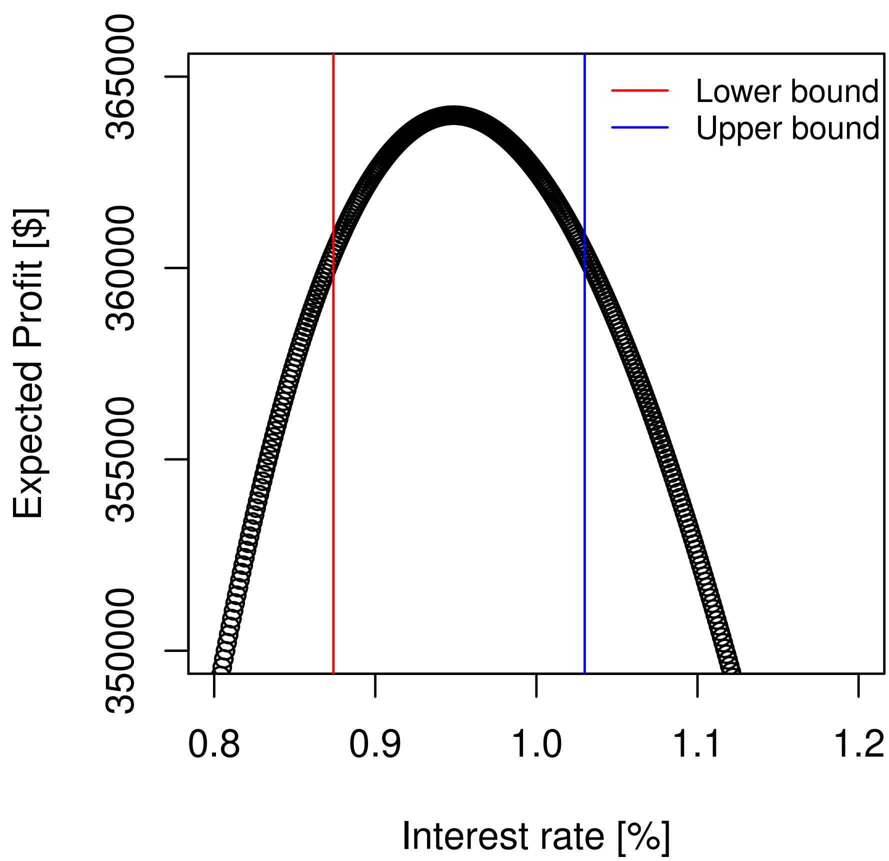


Figure 5.1: Setting bounds for a guided price negotiation.

5.3 Optimization Results

Table 5.1: Summary statistics for the optimized and actual models

	Mean Probability	Mean interest rate	Mean profit	Diff%
Optimized model	19.69%	1.403%	\$406.388	6.67%
Bank's actual model	22.48%	1.255%	\$379.179	-

The profits of the model proposed were compared with the actual profits. For each transaction j in the data set[1], the profitability was estimated as:

$$ActualProfit = \sum profit_j$$

$$= \begin{cases} Amount_j * Term_j * (R_j - CC_j - Risk_j) & \text{if } purchase_j = 1 \\ 0 & \text{if } purchase_j = 0 \end{cases}$$

and the proposed model is defined as:

$$ModelProfit_j = TakeUpProbability_j(R_j) * (Amount_j * Term_j * (R_j - CapitalCost_j - Risk_j))$$

where *ActualProfit* is the profit that was achieved for the period of the historical data and *ModelProfit* is our counterfactual estimate of the expected profit that would have been achieved if the bank had used the rates proposed from the price-elasticity model.

The second column of table 1 compares the mean interest rates between the optimized and actual model which are 1.40% and 1.26% respectively. The optimal rates in the optimized model are higher than the actual rates the bank offers, on average. Thus, we would expect a lower take-up probability (19.69% vs. 22.48%).

Table 5.1 also shows the average profit per loan and model. The mean profit in the optimized model is \$406.388 CLP. This suggests that adjusting prices would improve profits by \$27.209 CLP per transaction on average. Last column shows the difference in profit for both pricing methods during a 15-month period using our test set.

Figure 5.2. shows the optimal rate offered for a client that has a low-risk profile and belongs to the high-income segment. The *y-axis* represents the revenue or financial margin discounted from risk measured in Chilean pesos (CLP) and the *x-axis*, the monthly interest rate for the loan. The function shown is concave, thus, an optimal can be reached.

The transaction was settled on June 29th, 2017 and the bank offered a 1.12% monthly interest rate represented with a black continuous line. Our model proposes, under the same conditions, a monthly interest rate of 1.20% which is represented with a red dotted line. In this example the bank charged a slightly lower interest rate compared with the model's recommendation, which means the client was less price-sensitive than the bank's prediction.

[1]Historical dataset of transactions from the period 09/2016-12/2017.

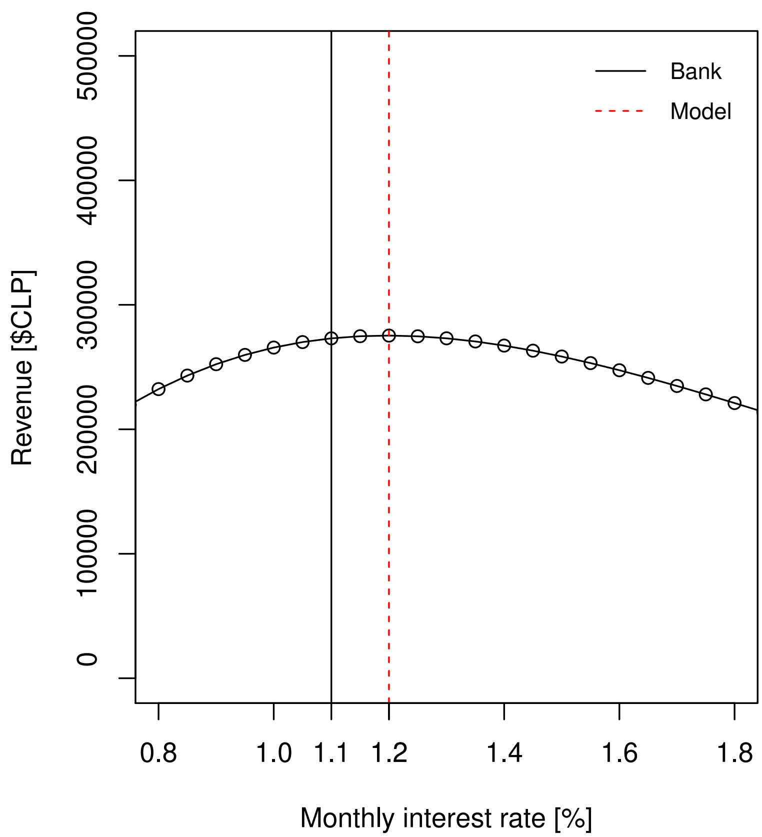


Figure 5.2: Optimal interest rates for a high-income client with a low-risk profile. In this case, the interest rate charged by the bank was below the optimal.

Our optimization results by solving the maximization problem are shown in tables 5.2 and 5.3 for the high and middle/young professional-income segments respectively. Both tables give an average of optimal prices obtained from historical data grouped by risk levels, term and amount. Consistent with the estimated model 4.4, three major observations can be made: higher credit amounts lead to lower interest rates, riskier clients are penalized with higher rates and the high-income segment is more price elastic compared to the MYP-income segment, which implies inferior interest rates.

In order to contrast our results, tables 5.4 and 5.5 show the bank's historical interest rates. The risk adjustment between the optimal and actual interest rates are considerably big and can be appreciated in tables 5.6 and 5.7 where the percentage variation is compared. Tables 5.8 and 5.9 show the revenue improvement from applying these price matrices. Note that lower risk segments present a lower revenue improvement and also gather most of the observations. Weighting the number of observations in each cell by the percentage change in revenue, reports gains of (6.67%).

Table 5.2: Optimal historical interest rate for the high-income segment, higher levels of risk were omitted due to too few observations

High-Income segment, historical optimal interest rates												
	Risk Tier 1 (<0.005)				Risk Tier 2 (0.005-0.02)				Risk Tier 3 (0.02-0.05)			
	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60
[3 , 4.5 MM]	1.223	1.239	1.234	1.138	1.707	1.731	1.722	1.596	2.366	2.379	2.367	2.205
(4.5 , 5.5 MM]	1.129	1.144	1.145	1.088	1.582	1.603	1.600	1.520	2.197	2.228	2.264	2.104
(5.5 , 7 MM]	1.125	1.142	1.141	1.079	1.579	1.598	1.592	1.513	2.196	2.231	2.194	2.108
(7 , 10 MM]	1.028	1.047	1.051	1.006	1.416	1.429	1.435	1.372	1.883	1.962	1.981	1.852
(10 , 20 MM]	0.971	0.991	0.999	0.970	1.315	1.337	1.343	1.307	1.772	1.785	1.788	1.759

Table 5.3: Historical optimal interest rates for the MYP-income segment

Middle/Young Professional-Income segment, historical optimal interest rates												
	Risk Tier 1 (<0.005)				Risk Tier 2 (0.005-0.02)				Risk Tier 3 (0.02-0.05)			
	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60
[3 , 4.5 MM]	1.316	1.336	1.332	1.227	1.990	2.019	2.007	1.853	3.036	3.055	3.034	2.826
(4.5 , 5.5 MM]	1.217	1.237	1.235	1.177	1.850	1.885	1.875	1.775	2.814	2.865	2.894	2.699
(5.5 , 7 MM]	1.217	1.237	1.231	1.172	1.842	1.871	1.861	1.771	2.808	2.867	2.835	2.697
(7 , 10 MM]	1.127	1.152	1.151	1.097	1.683	1.664	1.701	1.596	2.413	2.566	2.337	2.282
(10 , 20 MM]	1.063	1.065	1.070	1.049	1.506	1.513	1.501	1.505	2.234		2.171	2.134

Table 5.4: Bank's historical interest rates for the high-income segment

High-Income segment, bank's historical interest rates												
	Risk Tier 1 (<0.005)				Risk Tier 2 (0.005-0.02)				Risk Tier 3 (0.02-0.05)			
	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60
[3 , 4.5 MM]	1.097	1.119	1.146	1.184	1.183	1.208	1.253	1.320	1.855	1.898	1.957	2.093
(4.5 , 5.5 MM]	1.095	1.093	1.117	1.194	1.180	1.195	1.219	1.288	1.897	1.668	1.954	2.014
(5.5 , 7 MM]	1.068	1.080	1.114	1.153	1.156	1.167	1.202	1.261	1.886	1.768	1.832	1.788
(7 , 10 MM]	1.058	1.057	1.094	1.155	1.161	1.149	1.180	1.245	1.857	1.816	1.742	1.826
(10 , 20 MM]	1.031	1.050	1.061	1.133	1.143	1.138	1.147	1.203	1.865	1.799	1.772	1.851

Table 5.5: Bank's historical interest rates for the MYP-income segment

Middle/Young Professional-Income bank's historical interest rates												
	Risk Tier 1 (<0.005)				Risk Tier 2 (0.005-0.02)				Risk Tier 3 (0.02-0.05)			
	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60
[3 , 4.5 MM]	1.348	1.392	1.418	1.431	1.471	1.509	1.516	1.559	2.210	2.282	2.329	2.366
(4.5 , 5.5 MM]	1.340	1.355	1.363	1.381	1.346	1.376	1.393	1.374	1.975	2.126	1.961	1.897
(5.5 , 7 MM]	1.256	1.265	1.254	1.304	1.290	1.301	1.292	1.328	1.861	1.838	1.761	1.915
(7 , 10 MM]	1.215	1.244	1.293	1.292	1.278	1.245	1.276	1.300	1.711	1.529	1.753	1.915
(10 , 20 MM]	1.101	1.159	1.128	1.211	1.228	1.138	1.231	1.297	1.944		1.963	1.958

Table 5.6: Variation between the bank's (base) and the optimal interest rates for the high-income segment

High-Income segment, interest rate variation												
	Risk Tier 1 (<0.005)				Risk Tier 2 (0.005-0.02)				Risk Tier 3 (0.02-0.05)			
	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60
[3 , 4.5 MM]	11.41%	10.80%	7.60%	-3.89%	44.30%	43.31%	37.43%	20.90%	27.55%	25.35%	20.93%	5.38%
(4.5 , 5.5 MM]	3.10%	4.64%	2.53%	-8.84%	34.07%	34.15%	31.26%	18.04%	15.80%	33.61%	15.86%	4.47%
(5.5 , 7 MM]	5.37%	5.67%	2.43%	-6.40%	36.58%	36.97%	32.40%	19.94%	16.47%	26.18%	19.78%	17.89%
(7 , 10 MM]	-2.85%	-0.94%	-3.99%	-12.90%	22.03%	24.35%	21.58%	10.21%	1.44%	8.07%	13.71%	1.39%
(10 , 20 MM]	-5.83%	-5.62%	-5.87%	-14.44%	15.09%	17.44%	17.07%	8.64%	-4.97%	-0.77%	0.91%	-4.97%

Table 5.7: Variation between the bank's and optimized historic interest rates for the MYP-income segment

Middle/Young Professional-Income segment, interest rate variation												
	Risk Tier 1 (<0.005)				Risk Tier 2 (0.005-0.02)				Risk Tier 3 (0.02-0.05)			
	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60
[3 , 4.5 MM]	-2.36%	-4.00%	-6.08%	-14.24%	35.26%	33.82%	32.44%	18.89%	37.40%	33.86%	30.24%	19.45%
(4.5 , 5.5 MM]	-9.18%	-8.66%	-9.42%	-14.78%	37.44%	36.99%	34.62%	29.15%	42.50%	34.75%	47.56%	42.27%
(5.5 , 7 MM]	-3.12%	-2.24%	-1.86%	-10.16%	42.76%	43.75%	44.13%	33.29%	50.85%	55.97%	61.05%	40.89%
(7 , 10 MM]	-7.30%	-7.38%	-10.99%	-15.13%	31.69%	33.70%	33.35%	22.73%	41.09%	67.83%	33.32%	19.19%
(10 , 20 MM]	-3.52%	-8.13%	-5.07%	-13.40%	22.63%	33.01%	21.88%	16.07%	14.91%		10.60%	8.97%

Table 5.8: Expected revenue improvement applying the proposed interest rates for the high-income segment

High-Income segment, revenue improvement												
	Risk Tier 1 (<0.005)				Risk Tier 2 (0.005-0.02)				Risk Tier 3 (0.02-0.05)			
	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60
[3 , 4.5 MM]	3.46%	3.43%	2.77%	2.89%	18.61%	18.31%	14.90%	7.17%	11.65%	7.65%	9.02%	2.33%
(4.5 , 5.5 MM]	1.84%	1.66%	2.36%	2.94%	12.58%	13.28%	11.14%	6.01%	3.21%	21.78%	2.68%	0.81%
(5.5 , 7 MM]	2.73%	2.84%	1.49%	3.50%	14.83%	13.83%	12.02%	6.90%	3.10%	11.06%	6.61%	11.91%
(7 , 10 MM]	2.70%	2.49%	2.09%	5.04%	7.89%	9.46%	7.80%	3.83%	0.92%	5.07%	9.79%	6.03%
(10 , 20 MM]	3.34%	2.96%	3.37%	5.85%	6.57%	6.98%	7.29%	4.25%	0.57%	4.33%	5.47%	4.47%

Table 5.9: Expected revenue improvement applying the proposed interest rates for the MYP-income segment

Middle/Young Professional-Income segment, revenue improvement												
	Risk Tier 1 (<0.005)				Risk Tier 2 (0.005-0.02)				Risk Tier 3 (0.02-0.05)			
	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60
[3 , 4.5 MM]	1.36%	1.91%	2.04%	5.69%	11.33%	10.91%	11.14%	4.52%	13.88%	11.94%	8.43%	4.08%
(4.5 , 5.5 MM]	2.97%	2.59%	2.06%	4.26%	12.16%	11.91%	10.48%	8.40%	13.65%	9.39%	15.74%	13.68%
(5.5 , 7 MM]	2.98%	2.72%	3.08%	2.86%	14.15%	14.13%	14.95%	9.46%	16.75%	22.78%	25.92%	11.41%
(7 , 10 MM]	3.03%	2.98%	3.14%	4.85%	9.38%	10.31%	10.26%	5.80%	15.54%	41.12%	15.81%	4.61%
(10 , 20 MM]	5.79%	2.74%	5.15%	4.79%	6.51%	17.79%	4.68%	5.01%	2.21%		1.21%	1.02%

5.3.1 Final model

Final interest rates (with the re-estimated model that includes the simulation step omitted) are shown in tables 5.10 and 5.11 for the high and MYP-income segments.

Comparison with initial pricing matrices (pilot model) from the model 4.4:

- Lower recommended interest rates compared to those suggested in the first pilot.
- Differences within interest rates for different risk tiers decreased.
- Interest rates are similar among different income segments at low levels of risk but differ while riskiness increases.

Table 5.10: Optimal interest rates for the high-income segment, final model results.

High-Income segment historical optimal interest rates.																
	Risk Tier 1 (<0.005)				Risk Tier 2 (0.005-0.01)				Risk Tier 3 (0.01-0.018)				Risk Tier 4 (0.018-0.038)			
	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60
[3 , 4.5 MM]	1.018	1.065	1.091	1.056	1.193	1.242	1.260	1.229	1.506	1.560	1.591	1.552	1.572	1.631	1.657	1.622
(4.5 , 5.5 MM]	0.945	0.983	1.000	0.986	1.102	1.144	1.164	1.145	1.411	1.454	1.472	1.456	1.490	1.509	1.528	1.529
(5.5 , 7 MM]	0.944	0.982	0.998	0.986	1.102	1.144	1.163	1.143	1.409	1.454	1.474	1.456	1.480	1.531	1.537	1.532
(7 , 10 MM]	0.935	0.971	0.988	0.976	1.092	1.131	1.148	1.333	1.398	1.440	1.460	1.442	1.463	1.510	1.528	1.531
(>10 MM]	0.934	0.970	0.987	0.969	1.092	1.129	1.152	1.129	1.397	1.442	1.461	1.438	1.460	1.502	1.523	1.523

Table 5.11: Optimal interest rates for the MYP-income segment, final model results.

MYP-Income segment historical optimal interest rates.																
	Risk Tier 1 (<0.005)				Risk Tier 2 (0.005-0.01)				Risk Tier 3 (0.01-0.018)				Risk Tier 4 (0.018-0.038)			
	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60
[3 , 4.5 MM]	1.089	1.137	1.152	1.122	1.283	1.338	1.362	1.322	1.736	1.798	1.826	1.780	1.811	1.873	1.905	1.873
(4.5 , 5.5 MM]	1.013	1.051	1.068	1.055	1.200	1.239	1.258	1.241	1.635	1.680	1.699	1.680	1.723	1.733	1.781	1.769
(5.5 , 7 MM]	1.012	1.051	1.066	1.054	1.197	1.239	1.257	1.240	1.632	1.677	1.698	1.676	1.719	1.752	1.779	1.771
(7 , 10 MM]	1.004	1.039	1.056	1.044	1.188	1.226	1.244	1.230	1.626	1.665	1.687	1.669	1.736	1.744	1.777	1.760
(>10 MM]	1.006	1.043	1.056	1.042	1.192	1.223	1.231	1.229	1.628	1.676	1.674	1.663	1.795	1.740	1.751	1.757

Applying the prices described in the matrices above we would have the following results:

Table 5.12: Summary statistics for the optimized and actual models

	Mean Probability	Mean interest rate	Profit Diff.%
Optimized model	17.72%	1.199%	6.63%
Bank's actual model	13.50%	1.255%	-

Lowering the interest rates compared to the bank's actual model increases the take-up probability to 17.72%, which brings a total profit gain of 6.63%. The pricing matrices from above will become the new interest rates charged for a second experiment, and this time, using the final model structure. The implementation and results for the second experiment are out of the scope of this study, so it is left for further research.

Chapter 6

AB Testing

6.1 Description of the experimental design

Data for this study was provided by a large bank that operates in Chile. This institution offers a large number of services, but the focus is on online consumer credits for existing customers, i.e., customers can only purchase if they possess an account.

The bank operates an online web page where clients can look out for different consumer credit options and choose either to purchase or not. Also, the lender charges different interest rates depending upon the amount, term, market conditions and client characteristics, like the risk or income segment. In this thesis, the actual pricing policy was studied and a new model was recommended. To test the new model, a field experiment was implemented comparing both policies. In the treatment group the proposed interest rates were charged, whereas in the control group the existing interest rates were offered.

The first objective of this field experiment is to validate the new pricing method. Therefore, it is necessary to check whether the price elasticities calculated were correct. And second, the model's performance needs to be measured (if it is profitable compared to the actual pricing system).

Design Overview:

The experiment included 30.000 existing clients, they were randomly assigned to the treatment and control groups so each one of them has 15.000 clients. For a period of 99 days an interest rate will be charged from a pre-processed matrix that depends upon the credit conditions, market and client's characteristics. The clients can simulate and purchase as many times as they want but will remain in the same group until the end of the test period (i.e., will be offered interest rates from the same pricing matrix).

This field experiment aims to measure the profit in each experimental condition, validate the take-up probability and our estimates for price elasticity within different income, term, risk and amount segments.

Sample frame:

Two factors could influence the performance of our test. The first challenge is the risk of covariate shift. If the distribution of the targeting variables in the training data is different that the validation data, then the trained policies may not extrapolate well to the validation data.

The second challenge according to the machine learning literature is the "concept shift". This problem arises when the response function is not stationary due to seasonality, macroeconomic conditions, change in consumer behavior or exposure to intervening promotions. According to Simester et al. (2018), if the response function changes from the training data to the validation data, this may contribute to deterioration in the performance of the optimized policies. The pilot test is oriented to the clients that satisfy the rules we used to train the model in order to decrease the covariate and concept shift factor, consequently we defined the following conditions:

- The credit must be a new sale
- Simulated through the online platform
- Credit code CON300
- Without any promotions nor pricing initiatives
- Clients with an annual projected risk $\leq$ 10%
- Only High and Middle/Young.Prof income segments

The interest rate matrix for both, the treatment and control groups was segmented by term, amount, income and risk rate in the tiers stated below. For each cell the average optimized interest rate was calculated since the bank's system cannot offer it individualized.

Term:

- 13-24
- 25-36
- 37-48
- 49-60

Requested amount:

- $\geq$ \$3.000.000 and $\leq$ \$4.500.000
- $>$ \$4.500.000 and $\leq$ \$5.500.000
- $>$ \$5.500.000 and $\leq$ \$7.000.000
- $>$ \$7.000.000 and $\leq$ \$10.000.000
- $>$ \$10.000.000 and $\leq$ \$20.000.000

Income:

- High

- Middle/Young Professional

Annual projected risk:

- Risk Tier 1 between 0 and ≤0.005
- Risk Tier 2 >0.005 and ≤0.02
- Risk Tier 3 >0.02 and ≤0.05
- Risk Tier 4 >0.05 and ≤0.1

Summarizing, field experiment simulations satisfied the following requirements: term tiers between 13 and 60-months, amount tiers from between 3 and 20MM and risk tiers from 0 to 10%. Using a combination of different tiers, an interest rate matrix for the High and the Middle/Young Professional-income segment was generated.

Once the model was adjusted with historical data, a forecast for July was performed. The updated variables were the Cost of Funds for a credit, the Maximum Interest Rate and the average interest rate (Industry Rate).

For the Maximum Interest Rate and Industry Rate, we consulted the SBIF web site where the interest rates are published and valid from 15-06-2018 to 15-07-2018 period.

Table 6.1: SBIF rates

<50UF	32.35%	35.4%
Between 50 and 200 UF	24.65%	28.4%
Between 200 and 5.000 UF	14.4%	21.6%

Experimental limitations:

Our model provided an individual interest rate according to each client's profile, but the bank did not have the computational tools to offer a personalized deal. Thus, the interest rate offered was averaged within a range of parameters (income, risk, amount & term), losing some accuracy. Furthermore, as the pilot runs, some parameters that should be updated daily cannot be changed, like the cost of funds. Same happens with the industry rate and TMC that are updated once a month (usually the 15^{th}) and induce a lag. Luckily the last two variables do not vary much and represent a low fraction on the interest rate definition. Therefore, their change is not significative for the optimization results.

The SBIF published the TMC and industry rates for July on June 15th,2018 and both values are expressed in the U.F. currency. To transform it to Chilean pesos we used the U.F./Chilean peso conversion rate evaluated in $27.118. The information can be found in the following link: *SBIF information*

To calculate the Cost of Funds, from 18-05 to 18-06, we took the average rate for each term measured annually obtaining the following values:

Table 6.2: Averaged annual cost of funds

12-24 months	3.465%
24-36 months	3.801%
37 to 48 months	4.019%
49 to 60 months	4.176%

Once the historical database was updated with this information the optimization procedure was implemented. The historic optimal vs. the forecast optimal interest rates for July changed from 1.40% to 1.39% on average. This result indicates that variables can be forecasted and the model refreshed less frequently without losing accuracy assuming the cost of funds does not change much. The problem shows up when the cost of funds variates. For example, in October the Chilean central bank decided to change the monetary policy increasing the interest rates to maintain the inflation at 3%. This decision led to lower consumption levels and cost of funds increased gradually. If the model is not updated frequently, interest rates will not represent the correct consumer elasticities. Also, if the proposed interest rates exceed the TMC for a specific combination of amount, term, risk and income segment, the bank will charge the maximum interest rate allowed.

Tables 6.3 and 6.4 present the optimal-monthly interest rates offered for the High-income and Middle/Young Professional-income segments. Each cell shows an average of the interest rates the model predicted for a range of parameters. In both tables, as the risk of the customers is increase, interest rates also increase. Also, larger loans lead to lower interest rates and the high-income segment has lower interest rates compared to the Middle/Young Professional-income segment. Please note the pricing matrix used in this experiment corresponds to the model 4.4 and not the final model (specified in 4.11), which was improved later.

Table 6.3: High-income optimal interest rates

High-Income optimal interest rates												
	Risk Tier 1 (<0.005)				Risk Tier 2 (0.005-0.02)				Risk Tier 3 (0.02-0.05)			
	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60
[3 , 4.5 MM]	1.208	1.233	1.228	1.130	1.692	1.722	1.713	1.583	2.343	2.368	2.349	2.192
(4.5 , 5.5 MM]	1.111	1.138	1.138	1.074	1.567	1.593	1.591	1.506	2.174	2.216	2.267	2.093
(5.5 , 7 MM]	1.109	1.135	1.135	1.070	1.561	1.587	1.588	1.504	2.178	2.212	2.185	2.093
(7 , 10 MM]	1.011	1.038	1.045	0.995	1.402	1.420	1.429	1.362	1.870	1.949	1.978	1.842
(10 , 20 MM]	0.953	0.982	0.989	0.955	1.300	1.328	1.336	1.295	1.765	1.783	1.780	1.750

Table 6.4: Midlle/Young Professional-income optimal interest rates

Middle/Young Professional-Income optimal interest rates												
	Risk Tier 1 (<0.005)				Risk Tier 2 (0.005-0.02)				Risk Tier 3 (0.02-0.05)			
	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60	13-24	25-36	37-48	49-60
[3 , 4.5 MM]	1.303	1.329	1.324	1.222	1.977	2.011	1.999	1.841	3.018	3.053	3.024	2.807
(4.5 , 5.5 MM]	1.202	1.229	1.229	1.162	1.831	1.869	1.862	1.762	2.805	2.845	2.862	2.680
(5.5 , 7 MM]	1.200	1.227	1.226	1.161	1.823	1.859	1.853	1.755	2.783	2.830	2.818	2.685
(7 , 10 MM]	1.107	1.138	1.140	1.085	1.658	1.656	1.690	1.579	2.392	2.560	2.328	2.271
(10 , 20 MM]	1.030	1.058	1.060	1.032	1.468	1.484	1.497	1.481	2.161	2.140	2.138	2.118

6.1.1 Power analysis for the experiment:

Definitions:

- X_{i}= Revenue from purchase occasion i for a client in the treatment group (where we apply the optimized interest rates). It takes value zero if the client doesn't buy, and it is equal to the financial margin in case the client buys.
- Y_{i}= Revenue from purchase occasion i for a client in the control group. Analogous to the definition above.
- n is defined as the number of observations (purchase occasions) from each group, $2n$ observations in total.
- μ_x and μ_y are the expected revenues for X_{i} and Y_{i} respectively (equal to the take-up probability times the financial margin). μ_y can be directly computed from historical data.
- σ_x and σ_y are the standard deviations of X_{i} and Y_{i}. σ_y can be calculated from historical data. It is assumed that $\sigma_x = \sigma_y$ to simplify the analysis.

The following hypothesis test is formulated:

$$H_0 : \mu_x = \mu_y = \mu_0$$
$$H_a : \mu_x = (1 + \lambda)\mu_y$$

Where $\lambda > 0$ is the increase in revenue from the intervention (the size effect).

To perform the hypothesis test, the following statistic is used:

$$z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{2\sigma^2}{n}}}$$

with a rejection area $z > k$, where k is defined to achieve a given level of significance. Under H_0, z follows a normal standard distribution (mean zero and variance one). To achieve a level of significance (error type I) α is set k=z_α where Pr(N(0,1)>z_α)= α (one tail test). In this case we take α = 5% and z_α = 1.65. Note this is independent from the size of the sample. The size of the sample is chosen to meet the statistical power desired and the power is defined as the probability to detect H_a when H_a is correct:

$$Power = Pr(z > z_\alpha | \mu_x = (1 + \lambda)\mu_y)$$

Under H_a the statistic z has mean $\lambda \cdot \frac{\mu_0}{\sqrt{\frac{2\sigma^2}{n}}}$. Then the power is given by:

$$Power = \left(N(0, 1) > z_\alpha - \frac{\lambda\sqrt{n}}{\sqrt{2}} \times \frac{\mu_0}{\sigma} = -z_\beta \right)$$

where z_β is fixed to meet the statistical power desired. For a power = 1-β, z_β is chosen as the inferior tale of a Normal(0,1) in the β percentile. Solving:

$$-z_\beta = z_\alpha - \frac{\lambda\sqrt{n}}{\sqrt{2}} \times \frac{\mu_0}{\sigma}$$

we get:

$$n = 2 \times \left[\frac{z_\alpha + z_\beta}{\lambda} \cdot \frac{\sigma}{\mu_0} \right]^2$$

which gives us the size of the sample given. Note the size of the sample depends on:

- The size effect of the intervention, λ. Larger expected effects require smaller sample sizes.
- The coefficient of variation (CV), σ/μ_0, from revenues. Bigger CV, requires larger sample sizes. From the data, it is obtained a coefficient of variation around 2.

The following Table 6.5 shows the size of the sample for different levels of power and size effects (λ). The rest of the parameters are fixed at $\alpha = 0.05$, CV=2.0. Based upon this analysis, it is defined a minimum sample size for each group (control, treatment) of 15 thousand clients (30 thousand total), assuming an average effect of 5% increase in revenue to achieve a power of 70%. Being more conservative and assuming an effect of 2.5%, would require 60,000 clients per group (120 thousand total).

Table 6.5: Sample size for each power and size effect (λ)

	Size effect (λ)		
Power	0.01	0.025	0.05
70%	376,453	60,232	15,058
80%	494,605	79,137	19,784
90%	685,108	109,617	27,404
95%	865,774	138,524	34,631

6.2 AB Testing Results

Results were divided in two sections: first, the pilot test data were used to test whether the estimated elasticities of the final model were correct. The pricing matrices in the experiment did not represent the interest rates recommended in the final model, but the customer elasticities can be estimated. Second, the AB testing results are presented. Due to the lack of observations it is found that it is not statistically possible to prove an improvement/decrease of profits. Even though there was bias in the estimated coefficients from the pilot model, results showed characteristics of the customer behavior and helped to re-estimate the model. The recommended pricing matrices are shown in this study, but not the second experiment results.

The AB Testing data were collected in a period of 99 days (July 10th-October 17th) and come from the online consumer credits for the treatment and control groups specified in section *Description of the experimental design*.

In order to estimate the pilot coefficients, we followed the same methodology applied for the historical interest rates. The data was cleaned and segmented by risk, income, term and amount. The purchased simulations were identified, the clients filtered and the same interactions/model variables were created.

Table 6.6 shows the pilot model coefficients and table 6.7 the marginal effects.

Table 6.6: Estimated coefficients for the pilot model

Variable Name	Estimates	Conf.Interval
$AmountClass_{3-4.5MM}$	1.676^{*}	(0.264, 3.088)
$AmountClass_{4.5-5.5MM}$	0.956	(−0.375, 2.287)
$AmountClass_{5.5-7MM}$	0.785	(−0.303, 1.874)
$AmountClass_{7-10MM}$	0.455	(−0.420, 1.331)
$AmountClass_{10-15MM}$	0.418	(−0.427, 1.264)
$Amount_{>8MM}$ × InterestRate	0.293	(−0.352, 0.938)
$Risk_{\text{Tier1}}$ × InterestRate	-1.817^{**}	(−2.933, −0.701)
$Risk_{\text{Tier2}}$ × InterestRate	-1.420^{*}	(−2.633, −0.208)
$Risk_{\text{Tier3}}$ × InterestRate	-	-
$TermClass_{25-36}$	0.492^{*}	(0.107, 0.877)
$TermClass_{37-48}$	0.614^{*}	(0.116, 1.113)
$TermClass_{49-60}$	−0.102	(−0.705, 0.501)
$Risk_{\text{Tier2}}$	−0.099	(−1.877, 1.679)
$Risk_{\text{Tier3}}$	2.357	(−2.519, 7.233)
TMC	−0.008	(−0.139, 0.124)
Middle\Young.Prof	−0.705	(−2.370, 0.960)
Middle\Young.Prof × InterestRate	0.378	(−0.827, 1.582)
Constant	−0.678	(−3.935, 2.580)
Observations		1,553
Log Likelihood		−551.895
Akaike Inf. Crit.		1,139.790

Standard errors in parentheses
Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Term tier base level: (13-24), Amount tier base level: (>15 MM), Income tier base level: (High), Risk tier base level: (0-0.05%)

Table 6.7: Marginal effects for the pilot model

Marginal Effects					
Variable Name	df/dx	Std.Err	Z	P > |z|	
$AmountClass_{3-4.5MM}$	0.22	0.12	1.81	0.069067	.
$AmountClass_{4.5-5.5MM}$	0.11	0.10	1.15	0.246400	
$AmountClass_{5.5-7MM}$	0.09	0.07	1.20	0.229754	
$AmountClass_{7-10MM}$	0.05	0.05	0.91	0.360446	
$AmountClass_{10-15MM}$	0.04	0.05	0.85	0.392800	
$Amount_{>8MM}$ × InterestRate	0.02	0.03	0.89	0.372766	
$Risk_{\text{Tier1}}$ × InterestRate	-0.17	0.05	-3.23	0.001203	**
$Risk_{\text{Tier2}}$ × InterestRate	-0.13	0.06	-2.30	0.021122	*
$Risk_{\text{Tier3}}$ × InterestRate	-	-	-	-	-
$TermClass_{25-36}$	0.05	0.02	2.27	0.022689	*
$TermClass_{37-48}$	0.07	0.03	2.05	0.040135	*
$TermClass_{49-60}$	-0.01	0.02	-0.33	0.734347	
$Risk_{\text{Tier2}}$	-0.01	0.08	-0.11	0.911456	
$Risk_{\text{Tier3}}$	0.44	0.60	0.74	0.458649	
TMC	-0.00	0.01	-0.11	0.909919	
Middle\Young.Prof	-0.06	0.08	-0.84	0.398682	
Middle\Young.Prof × InterestRate	0.03	0.06	0.61	0.538697	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Our pilot test elasticities look similar to those estimated with historical data, this means it was achieved one of the objectives: verifying that the estimated elasticities were correct.

Overall, there were 1.646 purchase occasions and a total conversion rate of 15.6% summarized in Table 6.8. Note conversion rates changed because extra observations were added to the data set.

Table 6.8: Conversion rates

	Purchase Occasions	Conversion rate [%]
High-Income segment	891	16.8%
MYP-Income segment	755	13.5%
Total	**1,646**	**15.6%**

The first column in table 6.9 shows the number of purchase occasions for the treatment and control groups divided by risk tiers. Third column shows the conversion percentages and the fifth column the average objective function conditional to a purchase. In column 4 the interest rate was weighted by the size of the loan for the credits simulated. Column 2 shows the number of credits purchased and the last column the objective function over the number of purchased occasions. Rows represent the segmentation we performed in risk tiers, risk tier 3 was omitted due to lack of observations. From table 6.9, at low levels of risk, the control group presented more purchase occasions compared to the treatment group, but both look similar in the number of credits taken. Thus, conversion rates were higher in the treatment group, explained by the decrease in the average interest rate (1.08% vs. 1.22%). If the objective function is compared against the number of purchase occasions, also the treatment group performs better. For the second risk tier, the treatment group charged a higher monthly average interest rate (1.43%) compared to the control group (1.28%). Purchase occasions were almost the same and conversion rates for the control group were almost twice the treatment group which indicates the price-elasticity was misestimated.

Table 6.9: Pilot test results high-income segment

	# P.Occasions		**P.Credits**		**Conversion rate**		**Interest rate**		**Avg.Obj.Func | P**		**Obj.Func/P.Ocassions**	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
$RiskTier_1$	267	360	49	52	18.35%	14.44%	1.08%	1.22%	\$2.223.405	\$2.282.181	\$408.041	\$329.648
$RiskTier_2$	116	120	15	29	12.93%	24.17%	1.43%	1.28%	\$2.087.698	\$2.448.106	\$269.961	\$591.626

Table 6.10: Pilot test results MYP-income segment

	# P.Occasions		**P.Credits**		**Conversion rate**		**Interest rate**		**Avg.Obj.Func | P**		**Obj.Func/P.Ocassions**	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
$RiskTier_1$	218	242	33	24	15.14%	9.92%	1.22%	1.47%	\$1.541.252	\$1.818.921	\$233.309	\$180.389
$RiskTier_2$	150	126	20	21	13.33%	16.67%	1.75%	1.48%	\$2.477.669	\$1.492.777	\$330.356	\$248.796

The MYP-income segment (see table 6.10) showed promising results, with lower purchase occasions for the risk tier 1, the conversion rate was 15.14% for the treatment group vs. 9.92% in the control group, which came as a consequence of lowering the interest rates. An interesting analysis is that in the second risk tier, with relatively similar purchase occasions for both groups, the average interest rate was much higher in the treatment group (1.75%

vs 1.48%) and presented practically the same purchased credits (20 vs. 21 respectively). Comparing both tables, the high-income segment showed a bad response when the interest rate for risk tier two was increased, but the MYP behaved very well. Even though there might be a misestimation in the elasticities for risk tier 2, there was a significant difference within groups and the MYP-income segment was less elastic. The field experiment did not have enough observations. Thus, it was not possible to statistically prove profit gains/losses or the correct elasticities. In the power analysis section it was calculated that 30.000 purchased credits were needed assuming an average effect of 5% in revenue to achieve a power of 70%.

Our intuition suggests that prices increased to much from one risk tier to the other. Specifically, the propensity model results showed risk tiers 2 and 3 were less price-elastic, but there was a misestimation problem. This led us to consider other ways of modeling or potential errors in our methodology. Hence, the omitted simulation step was incorporated and new recommended interest rates were presented for a second experiment.

Chapter 7

Recommendations for implementation and conclusions

A pricing policy was empirically studied for a major bank in Chile, which solves a maximization problem that incorporates risk and willingness-to-pay to charge the optimal interest rates to people applying for consumer credits. To empirically estimate the pricing policies, a data set was used from the online purchases and simulations for consumer credits. First, the effect of price response on take-up for customers were estimated. Then the interest rates using the estimated elasticities were optimized and the value added was quantified using a counterfactual analysis in which the status quo profits were compared to the proposed method profits. Finally, a field experiment was implemented to check the estimated elasticities and see if there were any revenue improvements.

Our main empirical findings can be stated as follows: 1.- After realizing there was an omitted simulation step, the historical and pilot model were estimated again and confirmed that our estimated elasticities were very similar. In particular, it was found that while the amount increased, clients were more price-elastic. As risk increased, clients were less price sensitive and that lower income segments were less price elastic than higher income segments. An interesting finding though, is that at low levels of risk, interest rates looked almost the same. This indicates that both groups behave very much alike, but as we increase the risk level, differences between elasticities also increase. Which leads to charging a considerably higher interest rate for the MYP-income segment clients.

Having said this, the bank's intuition that lower income segments are less price sensitive than high income segments is correct and they reflect it in the pricing matrices.

2.- Interest rates used by the bank were suboptimal. Adjusting the interest rates generated by a maximization procedure that considers willingness-to-pay and risk can significantly increase profitability about 6.63% approximately.

The take-up probability with the proposed pricing method increased from 13.50% to 17.72% as a response to lowering the interest rates in some areas.

3.- The field experiment gave us some insights for our model and estimated coefficients. Even though there were favorable profits compared with the control group under some income segment or risk level conditions, there was not sufficient confidence to statistically prove revenue gains due to a lack of observations. The pilot test lasted about 3 months and reported 252 purchased credits out of the 30.000 needed. We learned that if we wanted more observations it is necessary to extend the sample to a wider spectrum of amounts, terms and income-segments.

Branch office sales represented a 58,6% of total sales for the bank. Hence, a solution to this channel that allows negotiation was also presented. The approach set and upper and lower bound for a guided negotiation where the middle point represents the optimal interest rate. It is thought that this application will help to take centralized decisions.

Throughout this thesis the following problems were solved: The risk effect and the willingness to pay were separated for each client because it was observed that all being equal riskier clients tend to be less price sensitive. Using a logit model, it was determined and measured how the price affects the take-up probability. Then, the interest rates were optimized according to the estimated elasticities and pricing matrices were recommended. An objective function was designed and maximized according to the bank's need and finally, the performance of the proposed model was evaluated and compared against the existing pricing policies.

As is the case of most empirical studies, our work had limitations that stem from the data. Besides not achieving the sample we required in order to statistically prove our pricing method, there were database collection problems. For example, it was found that online operations were purchased in one table, but on the other table did not have any simulations creating a matching issue. This situation might cause problems with estimating the elasticity because it changes the take-up rates. Some of this limitations suggest future research directions.

7.1 Future work

Multichannel effects: It is not studied if customers after simulating online go to the branch offices or get a phone call. What would happen if after getting an online offer the client goes to the branch offices and negotiates a better deal? There is a cannibalization effect between channels that is subject of further research.

Uncertain elasticities for credits below 3MM and $\leq$ 12 months for the high-and middle young professional income-segments: The elasticities the model delivered were not reliable, to correct this two solutions are presented:

- Design a field experiment in order to correctly measure the elasticities. The idea is to modify the interest rate with a disturbance to generate six scenarios $\pm$ 10%, $\pm$ 20% and $\pm$ 30%, taking as the baseline the actual interest rates the bank is offering.

- Add the simulation step omitted and included in subsection 4.4 for credits below 3MM and $\leq$ 12 months. Omission might have affected the estimation of the model.

Extend model to other income-segments: The model was implemented in two income segments out of four. To cover all the online applications, the model must be extended to the massive-emergent-university and the preferential-income segments.

Study the change of incentives for the sales agents: Headquarters establishes a price list based on objective and observable factors related to each deal. Then, the headquarters sets the limits on the adjustments that field sales staff can apply to the list of prices for the individual deals and finally the sales staff can then negotiate a price within the discretion limits applied to the list price for that deal. Many questions arise in this setting, did the local sales staff increase or reduce profits through the adjustments they made to the list prices? Could additional profit have been achieved if the list prices had been optimized? To control this situation the bank changed the sales agent incentives and told them to achieve sales and spread goals rather than just sales. Even though they have already changed this policy they have not done any evaluations, so it might be subject of future research.

Improve the existing model: Add insurance options, extend to all types of clients. Model the take up probability using the monthly payment (includes interest rate, insurance and taxes) instead of just the interest rate.

Extend model to other channels: There is a possibility to extend the model to other channels: branch offices, phone calls, proservice and ATMs.

Bibliography

Agrawal, V. & Ferguson, M. (2007). Bid-response models for customised pricing. *Journal of Revenue and Pricing Management*, 6(3), 212–228.

Annim, S. K. (2010). Sensitivity of loan size to lending rates evidence from ghana's microfinance sector. *African Review of Money Finance and Banking*, (pp. 85–107).

Attanasio, O. P., Koujianou Goldberg, P., & Kyriazidou, E. (2008). Credit constraints in the market for consumer durables: Evidence from micro data on car loans. *International Economic Review*, 49(2), 401–436.

Ausubel, L. M. (1999). *Adverse selection in the credit card market.* Technical report, working paper, University of Maryland.

Ban, G.-Y. & Keskin, N. B. (2017). Personalized dynamic pricing with machine learning.

Berman, R. & Feit, E. M. (2018). Enhancing power of marketing experiments using observational data.

Chen, X., Owen, Z., Pixton, C., & Simchi-Levi, D. (2015). A statistical learning approach to personalization in revenue management.

Cross, R. G. & Dixit, A. (2005). Customer-centric pricing: The surprising secret for profitability. *Business Horizons*, 48(6), 483–491.

Edelberg, W. (2006). Risk-based pricing of interest rates for consumer loans. *Journal of monetary Economics*, 53(8), 2283–2298.

Javanmard, A. & Nazerzadeh, H. (2016). Dynamic pricing in high-dimensions. *arXiv preprint arXiv:1609.07574*.

Juster, F. T. & Shay, R. P. (1964). Consumer sensitivity to finance rates. In *Consumer Sensitivity to Finance Rates: An Empirical and Analytical Investigation* (pp. 6–46). NBER.

Karlan, D. S. & Zinman, J. (2005). Elasticities of demand for consumer credit.

Oliver, B. & Oliver, R. (2014). Optimal roe loan pricing with or without adverse selection. *Journal of the Operational Research Society*, 65(3), 435–442.

Özer, Ö., Ozer, O., & Phillips, R. (2012). *The Oxford handbook of pricing management*. Oxford University Press.

Phillips, R. (2013). Optimizing prices for consumer credit. *Journal of Revenue and Pricing Management*, 12(4), 360–377.

Phillips, R., Raffard, R., & Solutions, N. (2011). *Price-driven adverse selection in consumer lending*. Technical report, Working Paper. Center for Pricing and Revenue Management, Columbia University.

Phillips, R., Şimşek, A. S., & Van Ryzin, G. (2015). The effectiveness of field price discretion: Empirical evidence from auto lending. *Management Science*, 61(8), 1741–1759.

Phillips, R., Simsek, A. S., & VanRyzin, G. (2012). Endogeneity and price sensitivity in customized pricing. *Columbia University Center for Pricing and Revenue Management Working Paper*, 4.

Phillips, R. & Solutions, N. (2012). Customized pricing. *The Oxford Handbook of Pricing Management, Oxford University Press, Oxford*, (pp. 465–490).

Phillips, R. L. (2005). *Pricing and revenue optimization*. Stanford University Press.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20, 53–65.

Schuermann, T. (2004). What do we know about loss given default?

Simester, D., Timoshenko, A., & Zoumpoulis, S. I. (2017). Efficiently evaluating targeting policies using field experiments.

Simester, D., Timoshenko, A., & Zoumpoulis, S. I. (2018). *Targeting Prospective Customers: Robustness of Machine Learning Methods to Typical Data Challenges*. Technical report, working paper.

Simsek, A. S. (2013). *Pricing Decentralization in Customized Pricing Systems and Network Models*. Columbia University.

Sundararajan, R., Bhaskar, T., Sarkar, A., Dasaratha, S., Bal, D., Marasanapalle, J. K., Zmudzka, B., & Bak, K. (2011). Marketing optimization in retail banking. *Interfaces*, 41(5), 485–505.

Terblanche, S. & De la Rey, T. (2014). Credit price optimisation within retail banking. *ORiON*, 30(2), 85–102.

Thomas, L., Oliver, R., & Hand, D. (2005). A survey of the issues in consumer credit modelling research. *Journal of the Operational Research Society*, 56(9), 1006–1015.

Vulcano, G., Van Ryzin, G., & Chaar, W. (2010). Om practice—choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing & Service*

Operations Management, 12(3), 371–392.

Appendices

Appendix A

To check whether our treatment effect is significant we run a linear regression with the revenue *(R)* as the dependent variable. We define i as a purchase occasion

$$R_{\mathrm{i}} = \begin{cases} 0 & No\,Purchase \\ Margin & Purchase \end{cases}$$

$$R_{\mathrm{i}} = \alpha + \beta * Income + \beta * Risk + \beta * Amount + \beta * Term + \beta * Treatment + \beta * TMC$$

Table 7.1: Pilot test linear regression

Variable Name	Estimates	Conf.Interval
Middle/Young.Prof	−0.077	(−0.183, 0.029)
$Risk_{\mathrm{tier2}}$	0.092 ·	(−0.015, 0.199)
Treatment	−0.025	(−0.123, 0.074)
Term	0.008^{***}	(0.004, 0.012)
$AmountClass_{3-4.5MM}$	0.112	(−0.235, 0.458)
$AmountClass_{4.5-5.5MM}$	0.032	(−0.282, 0.347)
$AmountClass_{5.5-7MM}$	−0.026	(−0.243, 0.191)
$AmountClass_{7-10MM}$	−0.037	(−0.252, 0.178)
$AmountClass_{10-15MM}$	0.092	(−0.147, 0.330)
TMC	−0.022	(−0.062, 0.019)
Constant	0.506	(−0.429, 1.440)
Observations		1,553
R^2		0.026
Adjusted R^2		0.020
Residual Std. Error		0.976 (df = 1542)
F Statistic		4.138^{***} (df = 10; 1542)

Standard errors in parentheses
Note: Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Amount tier base level: (>15 MM)
Income tier base level: (High), Risk tier base level: (0-0.05%)

The regression results show the treatment is not significant, thus we cannot report with confidence that our model performed better in some segments.

Appendix B

Pre-Processing the database: We order the simulations from the lowest to the highest client's **id** and from the oldest to the most recent **date**.

In the algorithm we always compare two consecutive simulations, to form a new cluster at least one of the following conditions must be satisfied:

- Condition (1) If the cluster already has a simulation that ended being bought.
- Condition (2) If the difference is bigger than 7 days.
- Condition (3) If they belong to different clients.

Purchase[i]: Vector that takes value 1 if the simulation i ends up being purchased, 0 otherwise.

ID[i]: Client's ID for a simulation i.

Date[i]: Vector with date of simulation i.

Cluster[i]: Vector initially $\emptyset$. Elements $\in \mathbb{N}_0$ are being added pointing the cluster's number to which the simulation i belongs to.

Note: To the group of simulations within the same purchase occasion, we generate a unique combination {Id Client,Cluster number}

Algorithm 1 Cluster formation within time window

```
procedure CLUSTER
    S : # Simulations data set
    n = 0
    counter_purchase = 0
    x = Date[1]                                   ▷ We start with an inicial date
    for i = 1 to S do
        if ID[i] = ID[i + 1] then
            if Date[i + 1] − Date[i] ≤ 7 then
                if purchase[i] = 1 then
                    if counter_purchase = 0 then
                        cluster[i] = n
                        counter_purchase = counter_purchase+1
                    else if counter_purchase > 0 then     ▷ condition (1)
                        cluster[i] = n
                        n = n + 1
                        counter_purchase = 0
                    end if
                else if purchase[i]= 0 then
                    cluster[i] = n
                end if
            else                                  ▷ condition(2)
                cluster[i] = n
                n = n + 1
                x = Date[i+1]
            end if
            if ID[i] ≠ ID[i + 1] then             ▷ condition (3)
                cluster[i] = n
                x = Date[i+1]
                n = 0
                cluster[i+1] = 0
            end if
        end if
    end for
end procedure
```