



## Determinantes del riesgo comportamental en préstamos de consumo y microcrédito: Un estudio de caso en Centro América

Adriana Uquillas  
Escuela Politécnica Nacional  
adriana.uquillas@epn.edu.ec

Recibido el 13 de junio de 2017; aceptado el 12 de octubre de 2017; Disponible en Internet el 14 de diciembre de 2017.

### Resumen:

Aunque en la industria financiera existen modelos de riesgo de crédito en el momento de la originación del préstamo, los bancos en general, todavía no formalizan la utilización de modelos de riesgo de crédito para la cartera de los créditos concedidos. Estos modelos son los llamados modelos comportamentales y son uno de los requerimientos del Comité de Supervisión Bancaria de Basilea. Este trabajo aborda la modelización del riesgo comportamental para las carteras concedidas de consumo y microcrédito. Para este propósito, se estimó un modelo de elección binaria con especificación logit. Los datos de corte transversal, utilizados en el análisis empírico, provienen de Guatemala. El estudio incluye el análisis de 62,000 individuos y el tratamiento de 157 variables explicativas recogidas en un período histórico entre 2015/01 y 2015/11. Se evaluó el riesgo de manera integral (mercado) y no solamente en silos de producto, lo que representa un aporte técnico importante.

Los resultados de la investigación sugieren que la probabilidad de mora disminuye si la edad del prestatario aumenta. Existe una relación positiva entre la probabilidad de mora y las variables comportamentales relativas a atraso en pagos (en el producto, en otros productos o en otras instituciones financieras) y entre el compromiso de la renta (deuda sobre salario) y la probabilidad de mora. La probabilidad de mora aumenta con la disminución de la renta y cuando el prestatario tiene relación crediticia con un mayor número de instituciones financieras (proxy de endeudamiento). Además, se alerta una fragilidad en el segmento de microcrédito, el mismo requiere de una tecnología y análisis crediticio más especializado.

**Palabras Claves:** Riesgo de Crédito; Scoring de Comportamiento; Modelo Logístico; Mercados Emergentes; Modelos econométricos

**Códigos JEL:** G20; G21; C4; C5; C21; C80

## Factors Affecting Behavioral Credit Risk in Personal Lending and Microcredit: A case study in Central America

### Abstract:

Although credit risk models exist in the financial industry at the time of loan origination, banks in general have not yet formalized the use of credit risk models for the loan portfolio. These models are the so-called behavioral models and are one of the requirements of the Basel Committee on Banking Supervision. This work deals with the behavioral risk modeling of personal lending and microcredit.

For this purpose, a binary choice model with logit specification was estimated. The cross-sectional data used in the empirical analysis come from Guatemala. The study includes the analysis of 62,000 individuals and the treatment of 157 explanatory variables collected in a historical period between January and November 2015. The risk was evaluated in an integral way (market view) and not only in product silos, which represents an important technical contribution.

The research results suggest that the probability of default decreases if the borrower's age increases. There is a positive relation between the probability of default and the behavioral variables related to delays (in the product, in other products or in other financial institutions) and between the commitment of the rent (debt over salary) and the probability of default. The probability of default increases with the decrease in income and when the borrower is related to a greater number of



financial institutions (need for debt). In addition, a fragility in the microcredit segment is alerted, it requires a credit technology and a more specialized analysis.

**Keywords:** Credit Risk; Behavioral Scoring.; Logistic Model; Emerging markets; Econometric models

**JEL Codes:** G20; G21; C4; C5; C21; C80

## 1. Introducción

La concesión de créditos es una de las actividades más significativas en la banca. Esta actividad se encuentra expuesta al riesgo de crédito que se define como el riesgo de pérdida debido al incumplimiento del prestatario o de la contraparte en la devolución del crédito, lo que conlleva el no pago, el pago parcial o la falta de oportunidad en el pago de las obligaciones pactadas. Las instituciones financieras deben establecer procesos adecuados de administración, gestión, medición, control, mitigación y monitoreo del riesgo de crédito al que se encuentran expuestas en el desarrollo del negocio, con el fin de mantener una adecuada cobertura de provisiones.

Para esto es indispensable considerar la existencia de diferentes perfiles de riesgo y las características de tanto los mercados en los que opera la institución financiera como de los productos que ofrece.

Por su parte, el Comité de Supervisión Bancaria de Basilea en su artículo "Convergencia internacional de medidas y normas de capital", en la tercera sección "Segundo Pilar - El proceso del examen supervisor", establece que: El proceso de examen supervisor establecido en este Marco no tiene por objetivo únicamente garantizar que los bancos posean el capital necesario para cubrir los riesgos de sus actividades, sino que también insta a los bancos a que desarrollen y utilicen mejores técnicas de gestión de riesgos en el seguimiento y control de los mismos.

Los avances tecnológicos y los requerimientos de Basilea II han aumentado el uso de modelos econométricos de riesgo de crédito en las últimas décadas, principalmente en los países desarrollados.

Por otro lado, aunque la mayoría de los países de América Latina han ido fortaleciendo sus regulaciones prudenciales y sus supervisiones en sintonía con las recomendaciones y los lineamientos



propuestos por el Comité de Supervisión Bancaria de Basilea, esto lo han hecho a través del refuerzo de sus normas de suficiencia de capital (Guerrero. R et al; 2011) pero han descuidado en diferentes medidas la modelización cuantitativa del riesgo crediticio.

Basilea II establece un mínimo del 8% para el indicador de Solvencia. En lo que se refiere a los países centroamericanos, según el Working Paper 136 del Banco de Guatemala, Panamá adoptó este nivel, mientras que Costa Rica, Guatemala, Nicaragua y Honduras coincidieron con un 10% y con una mayor exigencia está El Salvador con 12%. Sin embargo, al evaluar los riesgos considerados en los activos ponderados por riesgo, se tiene que Costa Rica considera dentro de los activos ponderados por riesgo: crédito, precio, tipo de cambio y operativo. Nicaragua, en cambio, considera los activos ponderados por riesgo de crédito y por tipo de cambio, mientras que en el resto de países únicamente se considera el riesgo de crédito, es decir, se estaría considerando únicamente lo establecido en Basilea I.

En el mismo Working Paper se señala que los reguladores tienen aún camino por recorrer en cuanto a la evaluación, retroalimentación y monitoreo de acuerdo al esquema de Supervisión Basada en Riesgos. En todos los países centroamericanos existen normas generales para la gestión de los riesgos y diferentes normativas para regular riesgos específicos. Esta regulación sigue las reformas que se han dado en Basilea, donde en primera instancia se consideró el riesgo de crédito, para luego, en Basilea II, incorporar los riesgos de mercado y el riesgo operativo, impactando todos los anteriores en el capital, mientras que recientemente con la reforma de Basilea III, se impulsó el marco regulatorio para el riesgo de liquidez.

Además, uno de los aspectos que el acuerdo de Basilea menciona (Basilea, C; 2005) es que aunque en la industria financiera existen modelos de probabilidad de mora para concesión de crédito, los bancos en general todavía no formalizan la utilización de modelos de riesgo de crédito para la cartera de los



créditos concedidos. Estos modelos son los llamados modelos de riesgo comportamental, y justamente son el objeto de estudio en el presente artículo. Este tipo de modelo no sólo es importante para la gestión de cartera de crédito sino también para entender comportamientos de recuperación de deudas y para mejorar la gestión de cobranzas (Chamboko, Bravo, 2016).

Para evaluar cuidadosamente el riesgo asociado con el préstamo de dinero, las instituciones financieras recopilan información sobre los prestatarios (por ejemplo: ingresos mensuales, deuda pendiente, datos geográficos, etc.) y, utilizando un modelo econométrico o estadístico, clasifican a los deudores como buenos o malos pagadores. La puntuación de crédito es un proceso de varias etapas. En general, implica los siguientes pasos: i) entendimiento del negocio, del mercado y del producto que se desea modelizar; ii) recopilación de información idiosincrática, sistémica y de políticas crediticias; iii) desarrollo de un modelo estadístico a partir de datos históricos; iv) aplicación del modelo estimado para calcular las puntuaciones de riesgo de los prestatarios potenciales; v) medición del desempeño del modelo; vi) desarrollo e implantación de políticas de concesión y gestión; y, finalmente, vii) monitoreo de la cosecha y de la cartera de crédito.

Uno de los principales desafíos en los mercados de crédito en los países en desarrollo es la escasez de información y por lo tanto las teorías basadas en estas limitaciones de información. El trabajo empírico inspirado por estas teorías genera tanto el apoyo a las propias teorías, como la posibilidad de producir cuestionamientos y mejorar la comprensión de estos mercados. La relación que se da entre la teoría y los datos sobre los mercados de crédito en las economías en desarrollo representa una oportunidad prometedora de investigación. Este artículo propone la estimación de un modelo comportamental de riesgo crediticio. Se desea modelizar la capacidad de pago de los prestatarios una vez que el crédito ya fue concedido. Más explícitamente, se modeliza el comportamiento crediticio de la cartera concedida de consumo y de microcrédito en una institución financiera de Guatemala. Los



modelos comportamentales incluyen un análisis más profundo que los modelos de concesión de crédito pues, en el estudio es posible incorporar factores relacionados al comportamiento de pago del cliente respecto al producto de crédito, así como las características del contrato de crédito, la utilización del crédito en el tiempo e incluso las cantidades y tipos de productos adquiridos.

Para este propósito, un modelo de elección binaria con especificación logit es construido. El estudio incluye el análisis de 62000 individuos y el tratamiento de 157 variables explicativas recogidas entre 2015/01 y 2015/11. Se espera que el modelo de crédito propuesto pueda contribuir para una gestión efectiva de riesgo de crédito a través del entendimiento de los factores que determinan el comportamiento de riesgo de los prestatarios.

Como fue mencionado anteriormente, se estudiará el comportamiento de los prestatarios de crédito de consumo y de microcrédito. De acuerdo a la Codificación de Regulaciones del Banco Central del Guatemala, se define el crédito de consumo como: “créditos otorgados a una sola persona individual destinados a financiar la adquisición de bienes de consumo o atender el pago de servicios o de gastos no relacionados con una actividad empresarial”. Por otro lado, el microcrédito se define como: “los activos crediticios otorgados a una sola persona individual o jurídica destinados al financiamiento de la producción o comercialización de bienes y servicios”. Se espera que el tipo de producto adquirido sea uno de los determinantes que afectan el riesgo comportamental.

El resto del artículo está estructurado de la siguiente manera: La Sección 2 presenta una revisión de literatura. La Sección 3 describe la metodología, el marco teórico y los datos disponibles. La Sección 4 presenta los resultados empíricos. Finalmente, la Sección 5 concluye el trabajo.

## **2. Revisión de Literatura**

Sobre el desempeño y uso de los modelos de riesgo de crédito (Martín, 2015) presenta una revisión de literatura académica relacionada a evaluación de desempeño de los modelos de riesgo de crédito. A



partir del análisis de 62 trabajos el autor concluye la falta de rigidez en la consistencia del desempeño de los modelos econométricos. El autor manifiesta que la mayoría de los modelos no se ajustan a los requerimientos prácticos aunque los mismos posean toda la rigidez estadística necesaria. (Leonard Onyiriuba, 2016) debate sobre si los modelos de crédito deberían suplantar el análisis cualitativo en la gestión de riesgo de crédito y muestra cómo esto afecta a los bancos en las economías emergentes.

Sobre los determinantes de riesgo de crédito, Clifford W. y Smith Jr. (1980), al asumir expectativas racionales en su trabajo, mencionan que los diferentes incentivos tanto para los prestamistas cuanto para los prestatarios determinan la complejidad de los contratos de créditos personales. Además, manifiesta que la venta de contratos ligada a una garantía es un mecanismo eficiente para controlar el conflicto de incentivos pues, en un mercado competitivo los beneficios de esta provisión se manifiestan a través del prestatario que puede tener acceso a crédito más barato.

(Linda Allen et al., 2004) realizan un resumen de los procedimientos establecidos por Basilea II en lo que se refiere a regulación de capital para bancos minoristas o de mercadeo. Además mencionan la relación existente entre el crédito y los préstamos transaccionales, lo que permite analizar una base de datos más completa para medir el comportamiento del riesgo de crédito. Malik, M., Thomas, L.C (2010) exploran el paralelismo entre los modelos comportamentales y la clasificación de riesgo de los bonos corporativos. Los autores incorporan en el modelo aspectos idiosincráticos de los clientes y características macroeconómicas a través del modelo de riesgo proporcional de Cox. Los resultados muestran que la mora es significativamente influenciada por factores macroeconómicos.

(Giray Gozgor, 2014) analiza los determinantes de crédito en 24 economías emergentes, estima un modelo de datos de panel y concluye que los balances externos y las percepciones de riesgo están negativamente asociados con el crédito doméstico. (Norlida Abdul Manab et al., 2014) examina el impacto de la gestión en la predicción del comportamiento crediticio. Los resultados muestran que



rentabilidad, liquidez y productividad son factores importantes. Por otro lado, G. Fernandes y R. Artes (2016) discuten el impacto de la economía local en el riesgo de crédito de pequeñas empresas. Incorporan en su análisis la dependencia espacial entre firmas proponiendo el método de kriging y lo comparan con las técnicas tradicionales basadas en la agrupación por el código postal. Los autores argumentan sobre la robustez de la metodología propuesta. (Anne-Sophie Bergerès et al., 2015) estudia la relación entre los diferentes productos de crédito. Los autores modelan la utilización de los créditos y la tasa de mora en un sistema de ecuaciones simultáneas y encuentran evidencia fuerte entre los dos instrumentos financieros. En el trabajo se verifica que los prestatarios usan la liquidez de crédito como fuente de pago del préstamo personal, lo que sugiere que los bancos deberían gestionar ambos instrumentos de manera simultánea.

(J. Trejo et al. 2015) realizan un estudio enfocado en la selección de variables. El modelo econométrico de comportamiento crediticio propuesto considera los cambios consolidados en las características de los prestatarios. Los autores concluyen que el modelo propuesto podría minimizar la pérdida esperada a través de una correcta gestión de riesgo. (Manju Puri et al., 2017) analiza información de 296 bancos alemanes con un total de un millón de préstamos y concluye que el hecho de que los bancos puedan usar información histórica sobre sus clientes disminuye la asimetría de información y por lo tanto mejora la previsibilidad del modelo. En este trabajo se define la relación entre banco y prestatario a través de diferentes variables como el número de transacciones y tiempo de relacionamiento con el banco.

Sobre los puntos ciegos entre prestamistas y prestatarios en el mercado de crédito, Linda Dezsó y George Loewenstein (2012) realizan una investigación descriptiva de los préstamos personales. En su trabajo los autores encuestan a 971 individuos sobre sus experiencias en créditos personales y detectan la presencia de sesgo en los prestatarios. La mora es frecuentemente más reportada por los



prestamistas que por los prestatarios. Existen sentimientos negativos entre las partes debido a la mora y, principalmente, los prestatarios no demuestran aprecio a los prestamistas. (Shweta Arya et al., 2013) incorpora medidas de impulsividad, actitud frente al riesgo y puntos ciegos para determinar preferencias sobre comportamiento crediticio. El material del trabajo es producto de entrevistas de orden financiero y psicológico. Controlando por el tamaño de la deuda, los autores encuentran que la puntuación crediticia está correlacionada con los factores antes descritos.

Sobre las diferentes metodologías de modelización en la literatura se ha propuesto un gran número de algoritmos de clasificación alternativos. Todos estos métodos relacionan una variable dependiente discreta con predictores de riesgo usando alguna forma funcional. (Sun Ho Ha, 2010) desarrolla un modelo de predicción de comportamiento crediticio con el objetivo de gestionar las cuentas deudoras en el producto tarjeta de crédito. El autor presenta un modelo híbrido que combina la red de Kohonen y el modelo de riesgo de Cox. El modelo estima el tiempo esperado de pago de la deuda. Posteriormente, Sun Ho Ha y Ramayya Krishnan (2012) agrupan a los deudores de créditos rotativos, específicamente tarjetas de crédito, en un banco de mercadeo. Además, desarrollan un modelo de predicción de pago para cada segmento usando para esto el análisis de riesgo proporcional de Cox. B. Cardoso y J. Díaz (2015) introducen la modelización de riesgo de crédito a través de modelos mixtos de sobrevivencia donde es posible plasmar la heterogeneidad no observada de los prestatarios. Mediante la metodología propuesta se identifican grupos de clientes con diferentes comportamientos de riesgo. Además, los parámetros del modelo pueden ser interpretados a través de las variables independientes en la regresión. El trabajo plantea que el tiempo entre el primer atraso y la mora de hecho puede ser modelado a través de una combinación de distribuciones log normal. Por otro lado, (Gang Dong et al. 2010) propone el uso de una regresión logística con coeficientes aleatorios. El



modelo propuesto demostró mejoras en el ajuste de predicción cuando fue comparado con la regresión logística habitual.

(Hussein A. et al. 2016) se enfoca en el sector bancario de Camerún y desarrollan modelos de riesgo de crédito usando diferentes técnicas econométricas y estadísticas como: regresión logística, árboles de clasificación y redes neuronales. Estos autores comparan las diferentes metodologías mediante estadísticos de desempeño habitualmente usados en este tipo de problemas, como son la curva ROC, el coeficiente de Gini y el estadístico de Kolmogorov-Smirnov. Los resultados muestran que las redes neuronales tienen mejor poder predictivo. Además, un análisis de sensibilidad identifica que variables como: ocupación, cuenta bancaria activa, presencia de garantías, otros tipos de créditos y gastos mensuales, son variables claves que determinan el comportamiento crediticio y también son importantes en el proceso de gestión de políticas de crédito.

Muchas técnicas se han desarrollado para modelizar riesgo de crédito. Técnicas como redes neuronales y el aprendizaje de máquinas han demostrado mayor capacidad de predicción que la técnica más comúnmente usada, la regresión logística, sin embargo, características deseables como la robustez e interpretabilidad que ofrece la regresión logística han hecho que estas técnicas no sean aplicadas en la práctica en la industria bancaria.

### **3. Metodología**

La herramienta estadística utilizada con frecuencia para estudiar los fenómenos y su relación con variables exógenas es la regresión. La técnica de regresión logística es utilizada para predecir una respuesta de dos niveles en función de un conocimiento previo expresado en un conjunto de variables. Si se dispone de los datos en el nivel individual o micro, se requerirá la estimación de los coeficientes asociados a los regresores mediante el método de máxima verosimilitud (MV).

De acuerdo a (Lemeshow S. & Hosmer D. W, 1989), se dice que un proceso es binomial cuando sólo tiene dos posibles resultados: "éxito" o "fracaso", siendo la probabilidad de cada uno de ellos constante en una serie de repeticiones. A la variable número de éxitos en n repeticiones se le denomina variable binomial. A la variable resultado de un sólo ensayo y, por tanto, con sólo dos valores: 0 para fracaso y 1 para éxito, se le denomina binomial puntual o Bernoulli.

En este trabajo interesa estimar la probabilidad de que un individuo sea un mal pagador, dado un conjunto de variables X. Esto es:  $P(Y = 1|X)$ . Por sencillez, si consideramos una sola variable explicativa, esta probabilidad se expresa mediante la función logística:

$$P(Y = 1|X) = \pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (1)$$

donde  $\pi(x)$  es la probabilidad de ocurrencia del evento de interés (el éxito), y por lo tanto recibe valores entre 0 y 1,  $\beta_0$  es una constante general y  $\beta_1$  es una constante que, de forma exponencial, influencia la probabilidad de éxito de acuerdo con el valor de X.

Para lograr propiedades deseables de un modelo de regresión lineal, la función de regresión sufre una transformación que matemáticamente no cambia la relación entre la variable independiente (predictor) y variable dependiente (a ser predicha), la función tendrá entonces la siguiente forma:

$$L_x = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x \quad (2)$$

En este caso, la ecuación de regresión pasa a ser lineal en sus parámetros y su interpretación se hace más simple.

Con relación a la estimación de los parámetros, de acuerdo con (Gujarati, 2010), al tener una muestra aleatoria de n observaciones, se considera la función  $f_i(Y_i)$ , ( $i = 1, 2, \dots, n$ ) que denota la probabilidad de que  $Y_i = 1$  o 0; la probabilidad conjunta de observar los n valores Y, es decir,  $f(Y_1, Y_2, \dots, Y_n)$ , entonces se expresa como:

$$f(Y_1, Y_2, \dots, Y_n) = \prod_{i=1}^n f_i(Y_i) = \prod_{i=1}^n \pi(x_i)^{Y_i} (1 - \pi(x_i))^{1-Y_i} \quad (3)$$

Se supone entonces que cada  $Y_i$  se obtiene de manera independiente y cada  $Y_i$  tiene la misma función de densidad (logística). La probabilidad conjunta dada en la ecuación (3) se conoce como función de verosimilitud (FV). Al tomar el logaritmo natural, se obtiene la función log de verosimilitud (FLV):

$$\ln f(Y_1, Y_2, \dots, Y_n) = \sum_{i=1}^n \left[ Y_i \ln \left( \frac{\pi(x_i)}{1 - \pi(x_i)} \right) \right] + \sum_{i=1}^n \ln(1 - \pi(x_i)) \quad (4)$$

A partir de (2) es posible expresar (4) como:

$$\ln f(Y_1, Y_2, \dots, Y_n) = \sum_{i=1}^n Y_i [\beta_0 + \beta_1 x_i] - \sum_{i=1}^n \ln(1 + e^{(\beta_0 + \beta_1 x_i)}) \quad (5)$$

Como se observa en (5), la función log de verosimilitud es una función de los parámetros  $\beta_0$  y  $\beta_1$ , pues los valores de  $X_i$  son conocidos. La estimación de MV consiste en obtener los valores de los parámetros desconocidos de forma que la probabilidad de representar los valores de  $Y$  observados sea tan grande (máximo) como sea posible, es decir maximizar la FLV. En el proceso de maximización no es posible obtener soluciones explícitas y por lo tanto se utilizan métodos de estimación no lineal para obtener soluciones numéricas. Una vez que se obtienen los valores numéricos de  $\beta_1$  y  $\beta_2$ , la ecuación (1) se estima con facilidad.

### Razón de probabilidades (Odds ratio)

A partir de la ecuación (2) se obtiene

$$\frac{\pi(x)}{1-\pi(x)} = e^{\beta_0 + \beta_1 x} \quad (6)$$

donde (6) representa la razón de probabilidades (denominada odds en inglés) en favor del éxito. Si por ejemplo, entonces la razón de probabilidades es igual a 2. Es decir, cuando  $x = 50$ , un éxito es dos veces tan probable como una falla (2 éxitos por 1 fracaso) o simplemente la odds es 2 a 1.

### Interpretación de los Coeficientes Estimados

Un coeficiente positivo aumenta la odds ratio, es decir la probabilidad de suceso del evento aumenta. Un coeficiente negativo disminuye la odds ratio. De manera formal, según (Gujarati, 2010), la interpretación del modelo dado en (2) es la siguiente:  $\beta_1$ , la pendiente, mide el cambio en  $L$  ocasionado por un cambio unitario en  $X$ , es decir, dice cómo cambia el logaritmo de las probabilidades en favor de ser un mal pagador a medida que  $X$  cambia en una unidad. El intercepto  $\beta_0$  es el valor del logaritmo de las probabilidades en favor de ser un mal pagador si  $X$  es cero. Al igual que la mayoría de las interpretaciones de los interceptos, esta interpretación puede no tener significado físico alguno.

### Análisis de Multicolinealidad

Para asegurar que el modelo estimado sea el correcto, es necesario asegurarse que no hay multicolinealidad entre los predictores en el modelo final. Las consecuencias de multicolinealidad en una regresión son los altos errores estándar e incluso la imposibilidad de cualquier estimación.

La multicolinealidad puede ser medida a partir de la estructura de correlación conjunta entre las variables de predicción del modelo. En última instancia, podemos representar estas variables como combinaciones lineales de las otras.

Para evaluar la multicolinealidad del modelo, se calcula los valores propios de la matriz de correlación lineal y se obtiene el índice de condición (IC) de la siguiente forma:

$$IC = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \quad (7)$$

Algunas referencias internacionales consideran inaceptables valores del IC mayores que 10 (Milone, 2009).

### **Método Chaid de Árbol de Decisión:**

El árbol de decisión es una metodología de modelación supervisada (con variable respuesta), así como la regresión logística. Esta metodología no se basa en parámetros y supuestos de linealidad, pero sí en la segmentación de la base de datos de las variables analizadas, siempre buscando maximizar la diferencia entre grupos con respecto a la variable de interés (en este caso la mora o no del cliente).

Dado que este método segmenta la base de datos maximizando la explicación de la variable respuesta, este trabajo propone utilizar este método para categorizar las variables explicativas antes de la aplicación de la regresión logística y después de haber definido la variable dependiente (Sección 3.1.1). Otros métodos predictivos de segmentación pueden encontrarse en (Jiawei Han et al., 2012).

El algoritmo más utilizado es el CHAID (Chi-cuadrado Detección automática) desarrollado por Kaas, en 1980. En cada paso del algoritmo se evalúa si una reducción en el número de categorías de la variable independiente es significativo o no, teniendo en cuenta la variable dependiente. El objetivo final es construir un árbol de clasificación que contiene sólo las variables más importantes para la clasificación, junto con sus categorías más importantes para la variable respuesta. Los pasos del algoritmo CHAID se pueden encontrar en (Santos, Oliveira; 2007).



### 3.1 Fuentes de datos y variables

Teniendo en cuenta la información disponible en la base de datos, apoyándose en investigaciones anteriores y en la literatura revisada, se incluye en la especificación del modelo variables explicativas de acuerdo a los siguientes seis grandes grupos de información: características del contrato y comportamiento de pago en el mismo, comportamiento de pago en otros productos, histórico crediticio en otros bancos, informaciones socio demográficas, tipo de producto concedido. El período considerado para el modelado fue entre 2015/07 y 2015/11 debido a la escasez de información e histórico disponible.

Para componer la información histórica de las variables explicativas del modelo, se observaron 6 meses anteriores a cada una de las fechas de referencia. A partir de las informaciones brutas de los seis grupos de información antes indicados se construyeron variables transformadas con el fin de incorporar nociones de comportamiento temporal y dar dinamismo al modelo. Esto permitió pasar de un conjunto de aproximadamente 70 variables a la disponibilidad de 157 variables explicativas.

Una vez que la metodología y la base de datos fueron definidas, se procede a definir las condiciones bajo las cuales un prestatario puede ser considerado un mal pagador.

#### 3.1.1 Definición de la variable respuesta

(Basilea, C; 2005) propone considerar como malos pagadores a aquellos individuos que presentan mora de más de 90 días de atraso en los siguientes 12 meses a partir del mes de referencia. Este enfoque permite mirar a un cliente en un plazo de tiempo grande, por lo que el modelo podría resultar más fiable. En este caso, sin embargo, no se poseía información con más de 6 meses de seguimiento debido a las restricciones en las bases de datos por lo que fue necesario realizar un análisis para definir la variable de buen pagador o mal pagador. Para esto, considerando la disponibilidad de información, notando que varias versiones de la marcación de mora están

altamente correlacionadas y que el modelo desarrollado debe estar alineado con la gestión de políticas y de negocio, fueron analizadas 4 posibilidades de marcación de mora: más de 30 días de atraso en 3 meses; más de 30 días de atraso en 6 meses; más de 60 días de atraso en 3 meses y más de 60 días de atraso en 6 meses.

Con relación a la marcación de 60 días de atraso en 3 meses, es necesario notar que quien atrasa más de 60 días en 3 meses será parte del conjunto de prestatarios que atrasan más de 60 días en los próximos 6 meses, pero no lo contrario. Por lo tanto, la marcación de 60 en 6 meses es más conservadora.

De acuerdo a la Tabla 1 de frecuencia cruzada observamos que 77% de los clientes que atrasos de 30 días en 3 meses se convirtieron en atrasos de 60 días en 6 meses.

**Tabla 1. Frecuencia cruzada**

30 en 3	Estadística	60 en 6	
		0	1
0	N	126333	5246
	% línea	96	3.99
1	N	5306	17991
	% línea	22.78	77.2

**Fuente:** *Elaboración Propia.*

Por otro lado, en la Tabla 2. Considerando el atraso de más de 30 días y asociándolo a las ventanas de seguimiento de 3 meses y 6 meses, observamos que 64.49% de los individuos que se atrasan en los próximos 3 meses, 35.51% se atrasan en los meses 4, 5 y 6. Es decir, en 3 meses se deja de identificar 35% de los atrasos con relación a los 6 meses. Por lo tanto, asumiendo una posición conservadora, la variable dependiente escogida es de 30 días de atraso en 6 meses. Es decir, un prestatario será considerado mal pagador si tiene más de 30 días de atraso en los próximos 6 meses a partir de la fecha de referencia.

**Tabla 2. Frecuencia cruzada seguimiento de 3 a 6 meses**

30 en 3	Estadística	30 en 6	
		0	1
0	N	118753	12826
	% columna	100	35.51
1	N	0	23297
	% columna	0	64.49

**Fuente:** *Elaboración Propia.*

Desde el punto de vista empresarial, la variable escogida asegura una ventana más grande (6 meses de seguimiento) y por lo tanto tiende a dar lugar a estimaciones más robustas, y de su ordenamiento predictivo es posible hacer políticas de crédito incluso en el corto plazo, pues se espera que garantice un buen ordenamiento de los clientes.

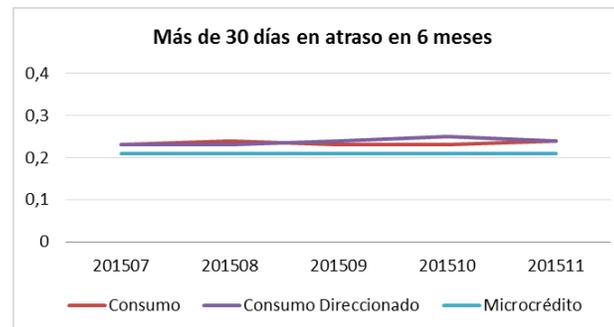
La Figura 1 muestra la variable dependiente en el período de modelado. Se observa que el porcentaje de mora de 30 días de retraso en 6 meses muestra estabilidad en el tiempo observado, y esto revela un aspecto positivo en el estudio.

**Figura 1. Tasa de mora observada a lo largo del tiempo**



**Fuente:** *Elaboración Propia.*

Figura 2. Tasa de mora observada a lo largo del tiempo por segmento.



Fuente: *Elaboración Propia.*

En la figura 2 se muestra la tasa de mora por segmento. Se observa estabilidad a lo largo del tiempo y comportamientos muy similares entre segmentos en términos de nivel de mora. Sin embargo, verificar la homogeneidad de las variables exógenas a través de los segmentos en términos de la variable respuesta sería un trabajo de investigación interesante pues, propone un estudio de segmentación con el objetivo de dividir la población total en subpoblaciones (por ejemplo microcrédito, consumo, consumo direccionado) que podrían tener un perfil de comportamiento más homogéneo en términos de mora. Modelos estimados para segmentos por separado deberían ser más robustos y con más poder de discriminación siempre que sea verificada la heterogeneidad de comportamiento a través de los segmentos.

Cifras de la Superintendencia de Bancos (SIB) de Guatemala señalan que al 30 diciembre 2015 los índices de morosidad se encontraban en 2.2% para Consumo y 3.4% para Microcrédito. Sin embargo, ese indicador de morosidad es una función del saldo en atraso y no es un indicador de mora o no de un individuo, conforme requerimiento de Basilea (Basilea, C; 2005). Al comparar el indicador de tasa de mora propuesto con los indicadores de gestión de la institución financiera de la

cual se obtuvieron los datos, se observaron índices similares de morosidad en cada segmento y en toda la cartera.

Finalmente, se realiza un estudio de asociación entre la marcación de incumplimiento y las variables explicativas. Para el análisis de correlación entre la variable dependiente y las variables cuantitativas se utiliza el análisis de correlación de Spearman. Se puede observar en la Tabla 3, que la variable ‘más de 30 días de atraso en los próximos 6 meses’ presenta el mayor número de variables significativas tanto en el 15% y el 25% de confianza. La definición de este nivel se debe al hecho de que se espera que el nivel de significancia aumente cuando sean consideradas las asociaciones de forma multivariante, es decir, el modelo de regresión con otras variables.

**Tabla 3. Correlación de Spearman entre la variable dependiente y las variables cuantitativas**

Marcación de atraso (x días de atraso en y meses)	Número de variables significativas al 15% y al 25% de confianza	
	15% de significancia	25% de significancia
30 en 3	54	60
60 en 3	53	60
30 en 6	62	65
60 en 6	53	68

**Fuente:** *Elaboración Propia.*

Para el caso de variables categóricas se utilizó la prueba de Kruskal-Wallis, que es una versión no paramétrica de ANOVA. En la tabla 4, se observa que las variables ‘30 días en 3 meses’ y ‘30 días en 6 meses’ muestran la mayor cantidad de variables correlacionadas con el nivel de 15% y 25% de confianza.

**Tabla 4. Prueba Kruskal-Wallis entre la variable dependiente y las variables categóricas**

Marcación de atraso (x días de atraso en y meses)	Número de variables significativas al 15% y al 25% de confianza	
	15% de significancia	25% de significancia
30 en 3	8	10
60 en 3	3	6
30 en 6	9	11
60 en 6	2	4

**Fuente:** *Elaboración Propia.*

El análisis bivariado muestra la capacidad de explicación de las variables dispuestas en la base de datos, la idea es tener un modelo de regresión que pueda explicar la mayor parte de la probabilidad de incumplimiento, de esta manera se espera que la variable dependiente tenga una alta correlación con las informaciones disponibles en la base de datos.

#### **4. Resultados y Discusión**

Para la modelización se consideraron las fechas de referencia entre 2015/07 y 2015/09. Las fechas restantes fueron separadas para realizar la validación out of sample del modelo. La información considerada fue la información del contrato de crédito actual en su origen, como el número de cuotas, tipo de producto y el monto bruto; información sobre el comportamiento de este producto de crédito, como atrasos y pagos; información de otros productos de crédito tales como la cantidad de productos de crédito y retraso en estos productos; productos de crédito de otros bancos; y la ubicación e información demográfica de los clientes, tales como el ingreso, tipo de vivienda, género y la edad.

Se descartaron de la muestra individuos sin información histórica, operaciones con más de 30 días de atraso en la fecha de referencia (malos pagadores en la fecha de referencia) y operaciones



próximas a finalizar el contrato. Además variables con más de 25% de datos faltantes fueron eliminadas automáticamente del análisis. No se eliminaron.

#### 4.1 Tratamiento y Selección de variables

Las variables originalmente numéricas fueron categorizadas, a excepción de la variable de edad. La categorización es una manera simple y parsimoniosa de lidiar con relaciones no lineales entre las variables determinísticas y la variable dependiente.

La variable edad fue elegida como excepcionalmente numérica por tener sólo 3 datos faltantes en toda la base (que fueron excluidos) y por presentar una relación lineal con la tasa de morosidad.

Para todas las demás variables numéricas (y categóricas) se realizó el proceso de categorización (y re categorización) siguiendo el método CHAID del árbol de decisión descrito en la sección anterior.

El proceso de re-categorización consiste en agrupar de manera óptima las variables categóricas evaluando si una reducción en el número de categorías de la variable independiente es significativo o no, teniendo en cuenta la variable dependiente. Considerando que se cuenta con más de 100 variables se usa el método de selección automática por pasos Stepwise para realizar una pre selección de variables, ver la referencia (Robert B. Bendel, A.A. Afifi, 2012) para detalles sobre este método. Posteriormente se realiza el análisis de parámetros y se procede a la fase de refinamiento de las categorizaciones. Este es un proceso interactivo de modelado y re-categorización de las variables hasta obtener un modelo final.

El modelo final obtenido en la estimación logística se muestra en la Tabla 5. Debido a la confidencialidad de información requerida por la institución financiera, fuente de los datos usados para el estudio, se han eliminado algunos valores de los coeficientes estimados y se ha procedido solamente a explicitar los respectivos impactos (-): negativo, (+): positivo y (++) positivo y de mayor impacto que (+).

Para el modelo propuesto se observa que todas las variables son estadísticamente significativas al 95% de confianza. Además, en la evaluación de multicolinealidad, no se observó ninguna variable con IC alto, el valor máximo fue apenas de 3.6.

**Tabla 5. Resultados del Modelo.**

Variable	Categoría	Parámetro	p-valor	IC
Intercepto		2	<.0001	0
Edad		-0,0154	<.0001	1,0472
¿Saldo en atraso actualmente?	no	-0,6344	<.0001	1,7706
	si			
Atraso corto (histórico reciente)	si	-0,3721	<.0001	3,568
	no			
Atraso corto (histórico medio plazo)	si	(-)	<.0001	3,4008
	no			
¿Más de 1 mes en atraso (histórico reciente)?	no	(-)	<.0001	1,0833
	si			
Dos o más meses consecutivos sin pagar? (histórico reciente)	no	(-)	<.0001	1,0141
	si			
saldo en atraso (histórico medio plazo)/ monto bruto	hasta 0.02	-0,3714	<.0001	1,8126
	> 0.02			
¿Posee saldo en atraso (histórico medio plazo)?	no	-0,5085	<.0001	1,6906
	si			
¿Posee saldo en atraso en los demás productos de crédito actualmente?	no	(-)	<.0001	1,0515
	si			
Comprometimiento renta interna	hasta 0.15	-0,181	<.0001	1,0662
	> 0,15 y missing			
¿Posee saldo en atraso en créditos de otros bancos?	no	(-)	<.0001	1,0833
	si, missing			
Número de instituciones financieras con crédito concedido sin considerar crédito hipotecario	hasta 2	-0,0804	<.0001	1,0639
	> 2, missing			
Saldo contrato/ monto bruto	hasta 0,7	(-)	<.0001	1,1173
	> 0,7			
Renta	baja	0,0751	0,003	1,0718
	media	0,0657	0,001	1,1163
	alta			
Tipo de Morada	Arrendada/propia hipotecada	-0,3964	<.0001	1,2674
	prestada	0,4802	<.0001	1,0341
	Vive com familiares			
Canal de Venta	Casas comerciales	(-)	<.0001	1,2592
	Direccionado a renegociados	(++)	<.0001	1,0539
	Fuerza de ventas	(+)	0,001	1,1514
	otros			
Familia Producto	Electrónica	(+)	<.0001	1,0934
	otros			
Segmento	Segmento en extinción	0,1964	<.0001	1,5183
	Consumo	-0,091	<.0001	1,433
	Microcrédito			

Fuente: *Elaboración Propia.*

#### 4.2 Poder de discriminación del modelo

La medida convencional de la bondad de ajuste, R cuadrado, no es particularmente significativa para los modelos con regresión binaria. Existen otras medidas similares a R cuadrado, llamadas pseudo R. En los modelos con regresión binaria, la bondad del ajuste tiene una importancia secundaria. Lo

que interesa son los signos esperados de los coeficientes de la regresión y su importancia práctica y/o estadística (Gujarati, 2010).

(Basilea, C; 2005) propone el uso de varias metodologías estadísticas para la evaluación del poder discriminatorio basándose en literatura previa o por ser populares en la industria financiera. El estadístico más popular es el estadístico KS (Kolmogorov-Smirnov) donde se evalúa la capacidad de diferenciar los buenos de los malos deudores a través del modelo de puntuación.

El estadístico utilizado es:

$$KS = \sup_x |F_x - G_x| \quad (8)$$

donde  $F_x$  representa la función de distribución acumulada empírica para la población 1 (malos pagadores) y  $G_x$  representa la función de distribución acumulada empírica para la población 0 (buenos pagadores). El KS corresponde a la distancia vertical máxima entre los gráficos de  $F_x$  y  $G_x$  sobre la amplitud de los posibles valores de  $x$  (score estimado por el modelo).

En la industria financiera, se esperan valores superiores a 50% como punto de referencia para los modelos de comportamiento. En la Tabla 6 se observa que todos los KS fueron altos, mayores que 0,5, destacando los segmentos de Microcrédito y Consumo. Además, podemos decir que los modelos presentan robustez en la muestra de validación pues sus valores no presentan grandes alteraciones.

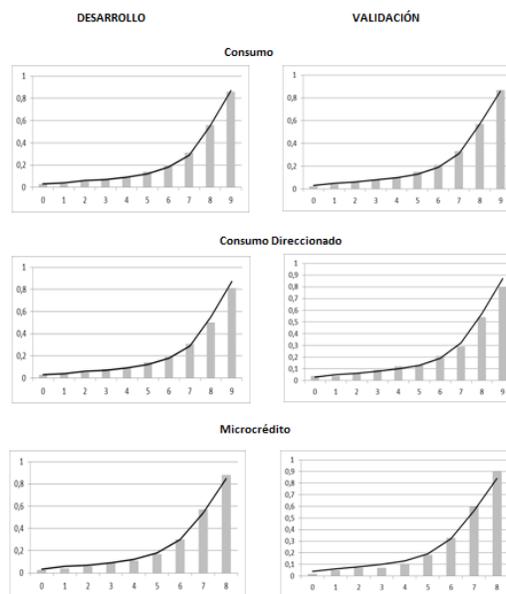
**Tabla 6. Poder de discriminación: KS del Modelo**

Segmento	Desarrollo	Validación
Consumo	0.59	0.58
Crédito Direccionado	0.51	0.51
Microcrédito	0.62	0.63
General	0.58	0.57

Fuente: *Elaboración Propia.*

Por otro lado, se espera que un modelo con buena capacidad de clasificación presente una tasa de morosidad creciente a medida que aumenta la probabilidad de mora estimada por el modelo. La Figura 2 muestra gráficos de ordenamiento: se observa el porcentaje promedio de morosidad (barras) y la puntuación media del modelo (líneas) para los segmentos de consumo, microcrédito y consumo direccionado, tanto para la muestra de desarrollo como para la muestra de validación.

**Figura 3. Tasa de mora observada a lo largo del tiempo**



**Fuente:** *Elaboración Propia.*

Estos resultados muestran una buena capacidad de clasificación, no hay inversiones de porcentaje de incumplimiento a lo largo de los deciles en el eje horizontal. La única excepción fue el microcrédito que mostró una pequeña inversión entre los grupos 2 y 3 en el período de validación, pero el hecho de que estas inversiones están en los grupos con probabilidad de mora inferior permite también afirmar que hay una buena capacidad para discriminar entre buenos y malos pagadores.



#### 4.4 Interpretación del modelo

Como se mencionó anteriormente, la variable modelada fue la mora. Al respecto, los parámetros del modelo muestran si una categoría del modelo final influye positivamente aumentando la probabilidad de incumplimiento (parámetro positivo) o negativamente disminuyendo la probabilidad de incumplimiento (parámetro negativo).

El coeficiente de EDAD igual a  $-0.0154$  indica que, si EDAD se incrementa en una unidad, en promedio el logit estimado disminuye aproximadamente 0.015 unidades, ceteris paribus. Lo cual indica una relación negativa entre ambos.

Existe una relación positiva entre las variables relativas a atraso (en el producto, en otros productos y en otras instituciones financieras) y la probabilidad de mora. Por ejemplo, al tomar el antilogaritmo del coeficiente de la variable atrasos cortos en el histórico reciente se obtiene aproximadamente 0.6892. Esto indica que al pasar de atrasos largos a atrasos cortos, es probable que, en promedio, sólo alrededor de uno de cada 1.5 clientes sea un mal pagador.

No poseer saldo en atraso en su comportamiento crediticio a mediano plazo, hace que el logit disminuya alrededor de 0.5085, si las demás variables permanecen constantes. Si tomamos el antilogaritmo obtenemos casi 0.6. Esto indica que si no se posee saldo en atraso en el comportamiento crediticio a mediano plazo, es probable que, en promedio, sólo alrededor de uno de cada 1.7 clientes sea un mal pagador.

A menor comprometimiento de renta del prestatario (deuda total sobre renta), menor probabilidad de mora. Un comprometimiento de renta bajo hace que el logit disminuya alrededor de 0.181, si las demás variables permanecen constantes.

Mientras más créditos concedidos posea el prestatario en diferentes instituciones financieras, mayor probabilidad de mora, pues esto da un indicio de mayor endeudamiento por parte del prestatario.



Menor número de instituciones financieras con crédito concedido hace que el logit disminuya alrededor de 0.0804, si las demás variables permanecen constantes.

Menor salario mensual por parte del prestatario aumenta la probabilidad de mora. Prestatarios con los salarios más bajos son aproximadamente 1.08 veces más propensos a ser malos pagadores que quienes tienen salarios altos. Prestatarios con salarios medios son aproximadamente 1.06 veces más propensos a ser malos pagadores que quienes tienen salarios altos.

‘Residencia prestada’ presenta mayor probabilidad de mora seguida de ‘vive con familia’ y ‘arrendada/ propia’. ‘Venta por medio de casas comerciales’ disminuye la probabilidad de mora. La ‘venta por el canal fuerza de ventas’ a través de agencias aumenta la probabilidad de mora y esta aumenta más todavía al usar el canal para renegociaciones de deuda.

Parece ser que créditos direccionados a compra de aparatos electrónicos tienen impacto positivo en la probabilidad de mora.

Finalmente, con relación al segmento crediticio, se observa que el crédito de consumo a través del crédito direccionado a la compra principalmente de línea blanca presenta menor probabilidad de mora seguida del microcrédito. Si el crédito concedido es de consumo y no es microcrédito, es probable que, en promedio, sólo alrededor de uno de cada 1.1 clientes sea un mal pagador, *ceteris paribus*.

## 5. Conclusiones

El riesgo de crédito es el riesgo más importante al que debe hacer frente una institución financiera, principalmente si hablamos de bancos sudamericanos y centroamericanos donde la principal actividad del sistema financiero es la concesión de créditos. La manifestación de este riesgo es el índice de morosidad, es decir, las obligaciones que no han sido pagadas a tiempo por los acreditados.



Uno de los principales desafíos en los mercados de crédito en los países en desarrollo es la escasez de información y por lo tanto las teorías basadas en estas limitaciones de información. El trabajo empírico inspirado por estas teorías, a su vez, ha generado tanto el apoyo a las teorías, como los enfoques para mejorar la teoría. Este enfoque bidireccional que tiene lugar entre la teoría y los datos sobre los mercados de crédito en las economías en desarrollo es una oportunidad prometedora de investigación para este mercado.

El modelo obtenido demuestra parsimonia, robustez y buen poder de discriminación. El mismo representa una herramienta importante para la gestión de crédito a través de políticas de renovación y es un punto de partida para el desarrollo de modelos de probabilidad de mora bajo los requerimientos de Basilea. Sin embargo, debido a la limitación de información se hizo uso de un indicador de mora de más corto plazo. Se contaron solamente con tres fechas de referencia para el desarrollo del modelo y dos fechas para la validación del mismo, pero al mismo tiempo se obtuvo más dinamismo y excelentes resultados al incorporar variables de comportamiento no sólo en las fechas de referencia sino variables que describen el histórico comportamental con hasta 6 meses de histórico a partir de la fecha de referencia. Son estas variables las que en su conjunto generan un aporte significativo en términos de discriminación. Mediante esta construcción de variables se pasa de un conjunto de aproximadamente 70 variables a la disponibilidad de 157 variables explicativas.

El entendimiento de los determinantes que afectan al riesgo de crédito conlleva a implicaciones de control, política crediticia y de gestión. La institución financiera tiene la posibilidad de, en algunos casos, manejar las variables de control para producir el nivel deseado de la variable objetivo, en este caso la morosidad de crédito y en otros casos la gestión de la cartera a través de políticas de crédito más alineadas al comportamiento crediticio de sus clientes, ofreciendo productos y condiciones más



adecuados a cada perfil crediticio. Por lo tanto, este trabajo constituye un aporte técnico importante y de referencia para la gestión y control de la cartera de créditos concedidos.

Evaluando el modelo propuesto podemos observar las siguientes características, el estudio apunta a concordancias en relación a la literatura existente en esta área, pero también contribuye con nuevos hallazgos: la probabilidad de mora disminuye con la edad del prestatario. Existe una relación positiva entre las variables comportamentales relativas a atraso (en el producto, en otros productos y en otras instituciones financieras) y la probabilidad de mora. El compromiso de la renta del prestatario (deuda total sobre renta) es un factor crucial para el comportamiento crediticio, en este estudio se verifica su importancia. El compromiso de renta tiene un impacto mayor de forma inversamente proporcional a la renta mensual, así que los salarios mayores pueden tener un compromiso mayor y los salarios más bajos merecen atención especial y un compromiso de renta menor.

Por ejemplo, cualquier variación positiva en la canasta básica de alimentos tiene un impacto mucho más fuerte para quien está en los cuartiles inferiores de ingresos mensuales. Los perfiles de consumo y el compromiso de los gastos esenciales son muy diferentes. En este sentido, la variable renta también es un factor determinante y el resultado está en línea con lo esperado, la probabilidad de mora aumenta con la disminución de renta, *ceteris paribus*.

El tipo de residencia parece ser un buen discriminador de comportamiento. ‘Residencia prestada’ presenta mayor probabilidad de mora seguida de ‘vive con familia’ y ‘arrendada/ propia’. El comportamiento crediticio en otras instituciones financieras y en otros productos también son determinantes del riesgo comportamental. Relaciones de crédito con menor número de instituciones financieras causan menor probabilidad de mora, entendiéndose este efecto como un descriptor de necesidad de endeudamiento del prestatario.



Poseer un crédito de consumo y no un microcrédito, hace que el logit disminuya alrededor de 0.091, si las demás variables permanecen constantes. Esto indica que para el segmento de consumo, es probable que, en promedio, sólo alrededor de uno de cada 1.1 clientes sea un mal pagador. El crédito de consumo es un producto diseñado para ofrecer facilidades para financiar estudios, comprar un vehículo para la familia, adquirir activos para el hogar, etcétera. El microcrédito está diseñado para los microempresarios informales cuyo destino específico es la compra de bienes o servicios que luego serán comercializados a los clientes finales, obteniendo así un margen de ganancia. Pero los mismos podrían estar destinados para pagar deudas más caras o financiar cuentas pendientes de cobro. Por la naturaleza y características propias de sus negocios, los microempresarios requieren de una tecnología crediticia y de un análisis más especializado y que no puede ser otorgado sin tener pleno conocimiento de la funcionalidad y beneficios para ambas partes, tanto de la entidad como del microempresario. El resultado del modelo alerta sobre la fragilidad de este segmento.

Finalmente, esta investigación abre las puertas para investigaciones futuras pues, un factor determinante en el comportamiento crediticio es la tasa de interés, factor que no ha sido incluido en este estudio por limitaciones en la información, pero que podrá ser considerado en próximos trabajos. La inclusión de este determinante de riesgo no es trivial ya que no puede ser considerado una variable determinística y es clara la relación de endogeneidad entre ambas variables de interés. Siendo así, técnicas estadísticas y econométricas más sofisticadas son necesarias para abordar este problema que está estrechamente ligado a cuestiones de asimetría de información y riesgo moral.

Además, avanzando en la modelización de riesgo bajo los requerimientos de Basilea, dos sistemas de clasificación de riesgo deben ser evaluados (en inglés, PIT: Point in time y TTC: Through the cycle). Los mismos difieren en el tipo de variable utilizada para la discriminación de riesgo siendo el primero pro cíclico y el segundo a cíclico (Basilea, 2005). ¿Las variables incluidas en este trabajo



podrían ser consideradas sistémicas? Si la respuesta es sí, ¿en qué grado? El modelo propuesto es PIT o TTC?, y ¿Cuáles son las implicaciones en gestión y control al tratar con cada sistema de clasificación? Estas son preguntas relevantes más aún en economías en desarrollo donde existe mucho dinamismo económico.

## 6. Referencias

- Abdou, Hussein, Tsafack, M. Dongmo (2016). Predicting creditworthiness in retail banking with limited scoring data. Knowledge-Based Systems. Forthcoming.
- A.Steenackers (1989). Credit scoring model for personal loans. Insurance: Mathematics and Economics. Volume 8, Issue 1, March 1989, Pages 31-34.
- Anne-Sophie Bergerès, Philippe d'Astous, Georges Dionne (2015). Consumer credit line utilization and default probability on a term loan? Evidence from bank-customer data. Journal of Empirical Finance. Volume 33, Pages 276–286.
- Artem Bequé, Stefan Lessmann (2017). Extreme Learning Machines for Credit Scoring: An Empirical Evaluation. Expert Systems with Applications. Volume 86, Pages 42–53
- Basilea, C. (2005). Studies on the Validation of Internal Rating Systems. BIS Working Paper 14.
- Bruno Cardoso Alves, José G. Díaz (2015). Survival mixture models in behavioral scoring. Expert Systems with Applications. Volume 42, Issue 8, Pages 3902-3910.
- Clifford W. Smith Jr. (1980). On the theory of financial contracting: The personal loan market. Journal of Monetary Economics. Volume 6, Issue 3, Pages 333-357.
- Gabriel Jimenez, Jesús Saurina (2004). Collateral, type of lender and relationship banking as determinants of credit risk. Journal of Banking & Finance. Volume 28, Issue 9, Pages 2191–2212.



- Gang Dong, Kin Keung Lai, Jerome Yen (2010). Credit scorecard based on logistic regression with random coefficients. *Procedia Computer Science*. Volume 1, Issue 1, Pages 2463-2468.
- Giray Gozgor (2014). Determinants of domestic credit levels in emerging markets: The role of external factors. *Emerging Markets Review*. Volume 18, Pages 1–18.
- Guilherme Barreto Fernandes, Rinaldo Artes (2016). Spatial dependence in credit risk and its improvement in credit scoring. *European Journal of Operational Research*. Volume 249, Issue 2, Pages 517–524.
- Gujarati, D. and Porter, D. (2010). *Econometría* (5th ed). Mexico, D.F, Mexico:McGRAW-HILL.
- Hakan Turan (2016). The Weighting of Factors Affecting Credit Risk in Banking. *Procedia Economics and Finance*. Volume 38, Pages 49-53.
- Hussein A. Abdou, Marc D. Dongmo Tsafack, Collins G. Ntim. (2016) Predicting creditworthiness in retail banking with limited scoring data. *Knowledge-Based Systems*. Volume 103, Pages 89–103.
- Jiawei Han, Midelin Kamber, Jian Pei. (2012). *Data Mining Concepts and Tehniques* (Third Edition). USA. Elsevier.
- José Carlos Trejo García, Miguel Ángel Martínez García, Francisco Venegas Martínez. (2015). Credit risk management at retail in Mexico: An econometric improvement in the selection of variables and changes in their characteristics. *Contaduría y Administración*. Volume 62, Issue 2, April–June 2017, Pages 399–418.
- Lemeshow S. & Hosmer D. W. (2005). *Applied Logistic Regression* , Segunda Edición. John Wiley & Sons, Inc.
- Leonard Onyiriuba (2016). *Emerging Market Bank Lending and Credit Risk Control*. Elsevier.



- Linda Allen, Gaylen DeLong, Anthony Saunders (2004) .Issues in the credit risk modeling of retail markets. *Journal of Banking & Finance*. Volume 28, Issue 4, April 2004, Pages 727–752.
- Linda Dezsó, George Loewenstein (2012). Lenders’ blind trust and borrowers’ blind spots: A descriptive investigation of personal loans. *Journal of Economic Psychology*. Volume 33, Issue 5, Pages 996–1011.
- Malik, M., Thomas, L.C (2010). Modelling credit risk of portfolio of consumer loans. *The Journal of the Operational Research Society*. Vol. 61, No. 3, Consumer Credit Risk Modelling; Transportation, Logistics and the Environment, pp. 411-420.
- Manju Puri, Jörg Rocholl,, Sascha Steffen (2017). What do a million observations have to say about loan defaults? Opening the black box of relationships. *Journal of Financial Intermediation*. Available online 11 February 2017.
- Martin, N (2015). Assessing scorecard performance: A literature review and classification. Institute of Marketing and Media, University of Hamburg, Welckerstrasse 8, D-20354 Hamburg, Germany.
- Mukesh Eswaran and Ashok Kotwal (1990). Implications of Credit Constraints for Risk Behaviour in Less Developed Economies. *Oxford Economic Papers New Series*, Vol. 42, No. 2, pp. 473-482.
- Norlida Abdul Manab, Ng Yen Theng, Rohani Md Rus (2014).The Determinants of Credit Risk in Malaysia. *Procedia - Social and Behavioral Sciences*, Volume 172, 27 January 2015, Pages 301-308.
- Richard Chamboko, Jorge M. Bravo (2016). On the modelling of prognosis from delinquency to normal performance on retail consumer loans. *Risk Management*. Volume 18, Issue 4, pp 264–287.



- 
- Robert B. Bendel, A.A. Afifi (2012). Comparison of Stopping Rules in Forward “Stepwise” Regression. *Journal of the American Statistical Association*. Volume 72. 1977. Issue 357.
- Santos, A. R.; Oliveira, L. A (2007). Comparação entre os algoritmos CHAID, CHAID-Exaustivo, CART e QUEST para dados com variável resposta categórica nominal via simulação bootstrap. no. 180. 40 f. Relatório Técnico do Departamento de Estatística – UFSCar.
- Sun Ho Ha (2010). Behavioral assessment of recoverable credit of retailer's customers. *Information Sciences*. Volume 180, Issue 19, 1. Pages 3703-3717.
- Sun Ho Ha, Ramayya Krishnan (2012). Predicting repayment of the credit card debt. *Computers and Operations Research*. Volume 39, Issue 4. Pages 765-773.
- Shweta Arya, Catherine Eckel, Colin Wichman (2013). Anatomy of the credit score. *Journal of Economic Behavior & Organization*. Volume 95, Pages 175–185.