



BCU

Modelos de Puntaje Crediticio: revisión metodológica y análisis a partir de datos de encuesta

Cecilia Dassatti

Nº 003 - 2023

Documento de trabajo
ISSN 1688-7565



Modelos de Puntaje Crediticio: revisión metodológica y análisis a partir de datos de encuesta[☆]

Cecilia Dassatti^{a*}

a Banco Central del Uruguay (Inveco), 777 Diagonal J.P. Fabini 11100 Montevideo, Uruguay

Documento de trabajo del Banco Central del Uruguay 003-2023

Autorizado por: Jorge Ponce
Disponible en línea: 18/5/2023

Resumen

Las instituciones financieras enfocadas en la concesión de préstamos al consumo se enfrentan a dos tipos de decisiones: por un lado, deben decidir si conceder un préstamo a un cliente nuevo y, por otro, deben decidir cómo gestionar los préstamos que ya concedieron. Los modelos de puntaje crediticio son técnicas muy utilizadas por las instituciones financieras a la hora de tomar este tipo de decisiones. El objetivo del presente trabajo es analizar los modelos de puntaje crediticio y las herramientas de validación de los mismos. Adicionalmente, se evalúa el grado de implementación de puntaje crediticios en Uruguay a partir de datos de una encuesta realizada en 2014.

JEL: C10, G21

Palabras clave: riesgo de crédito, puntaje crediticio, validación

[☆]Las opiniones vertidas en este documento son de exclusiva responsabilidad del autor y no representan la opinión del BCU.

Se agradecen los aportes de Alejandro Pena y Germán Ezquerra.

* Correo electrónico: cdassatti@bcu.gub.uy

1 Introducción

Los modelos de puntaje crediticio o credit scoring son técnicas muy utilizadas por las instituciones financieras para apoyarse en sus decisiones de concesión de créditos al consumo. El antecedente científico de estos modelos se ubica en el trabajo pionero de Fisher (1936), quien ideó la técnica estadística de análisis discriminante para diferenciar grupos dentro de una población a través de atributos medibles. Más adelante, en 1941, Durand (1941) observó que el mismo enfoque podría usarse para distinguir entre préstamos buenos y malos.

Las instituciones financieras enfocadas en la concesión de préstamos al consumo se enfrentan a dos tipos de decisiones: por un lado, deben decidir si conceder un préstamo a un cliente nuevo y, por otro, deben decidir cómo gestionar los préstamos que ya concedieron. Las herramientas estadísticas usualmente utilizadas para tomar el primer tipo de decisiones son las asociadas a los modelos de puntaje de admisión u originación, en tanto que las decisiones de gestión de límites de créditos o de estrategias de marketing para clientes ya existentes se apoyan en las técnicas de puntaje de comportamiento.

El objetivo del presente trabajo es analizar las metodologías más usualmente utilizadas a la hora de construir un puntaje crediticio, así como las herramientas aplicadas para validar este tipo de modelos. Adicionalmente, se evalúa el grado de implementación de puntajes crediticios entre las Empresas Administradoras de Crédito (EACs) en Uruguay a partir de datos de una encuesta realizada en 2014. El análisis se motiva por dos grandes factores: primero, el crecimiento registrado en la actividad de estas empresas dentro del sistema financiero uruguayo (los activos de las EACs crecieron aproximadamente un 60% en términos reales entre Diciembre de 2010 y Diciembre de 2018) y, en segundo lugar, la importancia relativa que presenta el riesgo de crédito entre los riesgos asociados al sistema financiero uruguayo.

El resto del documento se organiza de la siguiente manera. En la Sección 2, se realiza una breve introducción acerca de la metodología de puntaje crediticio. En la Sección 3 se describe la etapa inicial de tratamiento de los datos para construir modelos de puntaje crediticio. En la Sección 4 se describen los distintos modelos utilizados habitualmente, en tanto que en las Secciones 5 y 6 se describe la construcción final del tablero de puntaje y se analizan las herramientas para la validación de los modelos. Finalmente, en la Sección 7 se analizan los datos para Uruguay y en la Sección 8 se concluye.

2 Metodología de Puntaje Crediticio

Un modelo de puntaje crediticio consiste en una herramienta para evaluar el riesgo de crédito asociado tanto a solicitudes nuevas de préstamos (puntaje de admisión) como a préstamos ya concedidos anteriormente (puntaje de comportamiento). La idea detrás de estos modelos es identificar el efecto de ciertas características del solicitante del préstamo sobre su probabilidad de impago.

Este tipo de modelos surgieron como una primera respuesta formal al problema de evaluar el riesgo crediticio de un solo deudor de forma científica y automatizada, como consecuencia de la necesidad de procesar grandes volúmenes de solicitudes para préstamos relativamente pequeños. Durante los años treinta, algunas empresas de ventas por correo con una demanda creciente, diseñaron sistemas de puntuación con el fin de superar las inconsistencias detectadas entre las decisiones de otorgamiento de créditos de diferentes analistas. De esta manera, se identificaron los atributos observables de los deudores buenos o malos, para luego ser medidos y calificados numéricamente y llegar así a un puntaje final (simple suma de los componentes individuales). Durante la Segunda Guerra Mundial, al escasear el talento y la experiencia analítica crediticia, las compañías dedicadas al crédito al consumo a gran escala solicitaron a analistas experimentados que escribieran las reglas que los llevaron a decidir si otorgar o no un crédito. Como resultado, se llegó a una serie de diferentes tipos de sistemas de puntuación y de condiciones que la experiencia enseñó que se debían cumplir para que un préstamo tuviera una probabilidad razonable de ser repago.

Luego de la guerra, con la llegada de las calculadoras automáticas, los técnicas mencionadas pasaron por un proceso de automatización. En particular, el antecedente de las técnicas de calificación crediticia moderna surge del trabajo de Fisher (1936), quien ideó el *análisis discriminante* para diferenciar grupos dentro de una población a través de atributos medibles. Más adelante, en 1941, Durand (1941) observó que el mismo enfoque podría utilizarse para distinguir entre préstamos buenos y malos.

Inicialmente, tanto las variables seleccionadas como los puntajes asignados fueron principalmente basados en *juicios de experto*, pero la aplicación sistemática de los métodos contribuyó a la uniformidad de los criterios, proporcionando así consistencia y previsibilidad al proceso de originación del crédito. La introducción de las tarjetas de crédito en la década de 1960 proporcionó un impulso adicional para el uso extensivo de los sistemas de puntaje crediticio (Thomas (2000)).

Los elementos básicos del sistema de puntaje crediticio son los siguientes: (i) un conjunto de variables,

con sus correspondientes atributos, (ii) el puntaje asociado a cada atributo en particular, y (iii) un umbral o “valor de corte”. Las características se refieren a las preguntas planteadas al individuo, en tanto que los atributos son las respuestas a cada una de dichas preguntas, ya sea que se obtuvieron a partir del mismo solicitante del crédito o de fuentes externas. De esta manera, se asignan puntajes a cada característica y se suman para obtener un valor final. Dicho valor se compara con el umbral o valor de corte, tomándose la decisión de concesión del préstamo si el puntaje final supera el umbral.

De acuerdo al “Manual de Gestión del Riesgo para Actividades relativas a Tarjetas de Crédito” de la Federal Deposit Insurance Corporation (FDIC), existen distintos tipos de puntaje crediticio ¹. Por un lado, se encuentran los modelos utilizados para tomar la decisión de otorgar o no un préstamo por primera vez (*puntaje de admisión*). Estos modelos tienen como variables explicativas a características del solicitante, tales como su antigüedad laboral, el tipo de vivienda, la antigüedad en la residencia actual o el ingreso. Por otro lado, los *modelo scoring* son modelos utilizados para identificar a aquellos clientes que tienen una mayor tendencia a cancelar sus cuentas o a ser inactivos (deserción de clientes). Por último, los modelos de *puntaje de comportamiento*, se utilizan para evaluar el riesgo crediticio de individuos que ya son clientes de la institución (para evaluar cuáles están en peligro de no pago en el corto o mediano plazo). Estos últimos modelos son utilizados para ajustar los límites de crédito y decidir las políticas de comercialización a aplicar a cada tipo de cliente.

Si bien los modelos de puntaje crediticio tienen ventajas desde el punto de vista de ahorro de tiempo en el análisis, si la calibración de los mismos no es hecha con precisión, las consecuencias de su uso pueden ser muy negativas. De acuerdo a Mester et al. (1997), la muestra sobre la cual se desarrolla el modelo debe contar con una cantidad significativa de casos buenos y malos, los datos deben ser lo más actuales posibles, y el modelo debe ser revisado periódicamente para evaluar si las relaciones entre las variables son similares a las existentes en el momento inicial. Adicionalmente, la muestra debe incluir tanto a los casos de préstamos concedidos así como también a los rechazados, de modo de evitar sesgo en la misma. Posteriormente, para evaluar su desempeño, la muestra se divide en una muestra de desarrollo y una de testeo, con la que se evaluará el poder predictivo de la primera.

Las etapas básicas para la construcción de un modelo de puntaje crediticio son las siguientes (Leung et al. (2008)):

¹https://www.fdic.gov/regulations/examinations/credit_card/ch8.html

1. Conformación de la base de datos a partir de la información de las solicitudes de los clientes y del comportamiento de pago de los clientes incluidos en la misma.
2. Agrupación o segmentación de las variables de la base de datos en categorías o grupos (*atributos*).
3. Determinación de los clientes *buenos* (baja mora) y *malos* (morosos).
4. Determinación de las variables que entran en el modelo, seleccionando aquellas características con mayor poder predictivo. Con las variables seleccionadas se obtiene un modelo de predicción (aplicando métodos paramétricos o no paramétricos).
5. Validación del modelo, a través de técnicas como las curvas ROC y el Índice Gini, entre otras.
6. Definición del punto de corte, según el cual se separa a las solicitudes nuevas en aceptadas o rechazadas.

3 Diseño de los grupos y análisis del poder predictivo

Con el fin de lograr un modelo parsimonioso y de fácil interpretación, se realiza una agrupación inicial de las variables en categorías o grupos (atributos). Posteriormente, se calculan las medidas de predicción de cada grupo y se los ordena de acuerdo a su poder predictivo.

Una vez que se tienen las características agrupadas en categorías, el siguiente paso es analizar si los criterios de agrupación seleccionados tienen un buen poder predictivo.

3.1 Test de Chi-cuadrado

Esta prueba sirve para decidir si dos criterios de clasificación sobre el mismo conjunto de individuos son independientes o no. La misma se utiliza para seleccionar los puntos de corte o umbrales que mejor separen la muestra en base a la distribución de buenos y malos pagadores. Se basa en una tabla de contingencia según la cual se separa la muestra a partir de un punto de corte en una variable:

X/Y	Y=0	Y=1	n_y
X_iC	n_{11}	n_{12}	n_{1*}
X_lC	n_{21}	n_{22}	n_{2*}
n_x	n_{*1}	n_{*2}	n

El estadístico se calcula de la siguiente manera:

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

en donde: n_{ij} son las frecuencias observadas clasificadas en la fila i de la columna j y e_{ij} son las frecuencias esperadas o teóricas que se observarían en caso de que se cumpliera la hipótesis nula de independencia. Entonces, el estadístico mide cuánto se parecen las frecuencias observadas a las esperadas. Cuanto mayor el valor del estadístico, menos creíble es que ambas frecuencias se parecen. El test Chi-cuadrado da una idea de cuáles son los puntos que mejor diferencian a la población en función de la tasa de mora o la probabilidad de incumplimiento.

A modo de ejemplo, en la Figura 3.1, se supone que existe una característica que se puede agrupar en tres categorías diferentes:

Atributos	Valores Observados			Valores Esperados	
	Buenos	Malos	Total	Buenos	Malos
Grupo 1	50,000	2,500	52,500	49218.75	3,281
Grupo 2	6,500	400	6,900	6468.75	431
Grupo 3	17,000	2,000	19,000	17812.50	1,188
Total	73,500	4,900	78,400	73,500	4,900

Figure 3.1

El estadístico χ^2 se calcula de la siguiente manera:

$$\chi^2 = \frac{(50.000-49.218,75)^2}{49.218,75} + \frac{(2.500-.3281,25)^2}{3.281,25} + \frac{(6.500-6.468,75)^2}{6.468,75} + \frac{(400 - 431,50)^2}{431,50} + \frac{(17.000 - 17.812,50)^2}{17.812,50} + \frac{(2.000 - 1.187,50)^2}{1.187,50} = 793,84$$

Finalmente, si los datos se pudieran agrupar de acuerdo a otra clasificación, para evaluar cuál es la mejor, se selecciona aquella con un valor mayor del estadístico.

3.2 Indicador Weight of Evidence (WOE)

El indicador WOE refleja el poder predictivo a partir de la diferencia entre la proporción de buenos y malos dentro de cada uno de los rangos o categorías que puede tomar cada variable independiente. Es decir, el indicador mide la probabilidad de que un cliente con ese atributo (categoría o rango) pueda ser bueno o malo.

$$WOE_i = 100 \cdot \ln\left(\frac{\%Buenos_i}{\%Malos_i}\right)$$

$$\%Buenos_i = \frac{Buenos_i}{TotalBuenos}$$

$$\%Malos_i = \frac{Malos_i}{TotalMalos}$$

Algunas puntos a tener en cuenta a la hora de interpretar los valores del indicador son los siguientes:

- Un indicador con valores negativos implica una alta proporción de malos sobre buenos.
- Para que el WOE esté definido, ninguna de las clases puede estar conformada únicamente con buenos o malos.
- Dado que el indicador se puede interpretar como el nivel de riesgo asociado al rango de la variable independiente, se suele ordenar de forma creciente.
- Para que el análisis sea significativo, se aplica la regla de que al menos tiene que haber un mínimo del 5% de los datos en cada grupo.
- Tanto la tasa de malos como el valor del WOE deben ser lo suficientemente diferentes entre cada grupo; la agrupación o segmentación se debe realizar de forma tal que se maximice la diferencia entre buenos y malos. Cuanto mayor sea la diferencia entre el WOE de los grupos, mayor será la capacidad predictiva asociada a ese grupo o atributo.

De acuerdo a Zeng (2013), el indicador WOE contribuye a la medición de la fuerza de cada atributo de la variable independiente, de acuerdo a la clasificación de bueno y malo otorgada por la variable dependiente.

3.3 Indicador Information Value (IV)

Se trata de un indicador del poder predictivo o discriminante de cada una de las variables independientes. Se construye a partir de la suma de los WOE a lo largo de los rangos o categorías, ponderados por las diferencias proporcionales halladas en dichos rangos.

Para un número k de categorías o rangos, se define de acuerdo a la siguiente expresión:

$$IV = \sum_{i=1}^k (\%Buenos_i - \%Malos_i) * WOE_i$$

Las variables con un IV menor a 0,10 se consideran generalmente como débiles, en tanto que valores superiores a 0,30 suelen ser asociados con variables con fuerte poder predictivo. Por último, valores por encima de 0,5 suelen asociarse con variables que sobre-predicen a la variable independiente y se aconseja no considerar a dichas variables o considerarlas de forma controlada (Siddiqi (2012), Zeng (2013)).

Cabe aclarar que mientras que el WOE describe la relación entre una variable independiente y la variable dependiente, el IV mide la fuerza de esa relación. Es decir, con el IV se puede medir el poder de predicción de dos maneras diferentes de agrupar los atributos de una característica.

A continuación, en base a los datos de ejemplo detallados en la Figura 3.2, se interpretan los resultados de calcular ambos indicadores.

Edad	N° Clientes	Distribución	Buenos	Distri. Buenos	Malos	Distri. Malos	Tasa de Malos	WOE	Dif. Distrib.	IV parciales
18-23	4000	11.76%	2050	6.91%	950	28.61%	23.75%	-142.14	-21.71%	0.30854
24-30	8000	23.53%	6800	22.91%	1200	36.14%	15.00%	-45.59	-13.23%	0.06033
30-35	12000	35.29%	11100	37.40%	900	27.11%	7.50%	32.18	10.29%	0.03311
35-42	6000	17.65%	5800	19.54%	200	6.02%	3.33%	117.68	13.52%	0.15907
42+	4000	11.76%	3930	13.24%	70	2.11%	1.75%	183.74	11.13%	0.20455
Total	34000	100%	29680	100%	3320	100%				0.76562

Figure 3.2

El procedimiento para elaborar la tabla fue el siguiente:

1. Para una variable continua, se definen rangos para agrupar las características (en caso de que se trate de una variable categórica, este paso no es necesario).
2. Se calcula el número de buenos y malos dentro de cada grupo.
3. Se calcula el porcentaje de buenos y malos en cada grupo.

4. Se calculan el WOE y el IV.

Por ejemplo, dentro del grupo de edades entre 18 y 23 años, hay 4.000 clientes, los cuales representan el 11,76% del total de clientes considerados. De éstos, 2.050 son buenos y representan el 6,91% de los clientes buenos de la muestra y 950 son malos (28,61% del total de malos). El WOE para este grupo de edad es:

$$WOE_{18-23} = 100 \cdot \ln\left(\frac{6,91\%}{28,61\%}\right) = -142,14$$

El IV parcial para este grupo de edad es:

$$IV = (0,0691 - 0,2861) * (-142,14) = 0,30854$$

Por último, el IV total es de 0,76562. De acuerdo a los valores de referencia mencionados previamente, la categoría edad parece tener un poder predictivo fuerte.

4 Modelos de Puntaje Crediticio

Como ya se mencionó, los modelos de Puntaje Crediticio buscan obtener una regla general que permita determinar la probabilidad de incumplimiento de una determinada solicitud de crédito. Este objetivo se puede resumir mediante la siguiente expresión:

$$PD = f(x_1, x_2, \dots, x_k) + \epsilon$$

En donde: PD es la probabilidad de incumplimiento, x_i son los atributos del solicitante del crédito, $f(x)$ es la forma funcional que describe la relación entre ambas variables y ϵ es el error aleatorio. Estaremos ante modelos paramétricos o no paramétricos, según que la forma funcional sea conocida o desconocida.

4.1 Modelos Paramétricos

Entre los modelos paramétricos utilizados para desarrollar sistemas de puntaje crediticio se encuentran los modelos de probabilidad lineal, los modelos logit y probit y los modelos de análisis discriminante. Los tres primeros se basan en datos históricos acerca del desempeño de un préstamo y las características del

solicitante.

Como es conocido, el modelo de probabilidad lineal asume una relación lineal entre la probabilidad de incumplimiento y las variables explicativas. Los modelos logit y probit, en cambio, son no lineales en los parámetros. El modelo logit asume que la probabilidad de incumplimiento sigue una distribución logística, en tanto que el modelo probit asume una distribución normal.

El análisis discriminante, por otro lado, en lugar de devolver una probabilidad de incumplimiento, divide la muestra entre clientes de alto y bajo riesgo de incumplimiento. El objetivo de análisis discriminante es encontrar aquellos factores explicativos que permitan distinguir a un grupo de otro.

Se trata de un modelo de regresión lineal en el cual las variables independientes representan las características diferenciadoras de cada individuo (Thomas (2000), Altman (1968)), siendo éstas las que permiten realizar la clasificación. La idea es obtener la mejor combinación lineal de variables independientes que maximicen la diferencia entre grupos (función discriminante). Adicionalmente, se busca predecir la pertenencia de un individuo a los grupos preestablecidos (capacidad predictiva). Para ello, la muestra se divide aleatoriamente en dos subgrupos: un grupo de entrenamiento, que se utilizará para la obtención de las funciones discriminantes, y un grupo de testeo, que servirá para determinar la capacidad predictiva del modelo obtenido.

Una de las limitaciones del modelo de regresión lineal es que la variable PD puede tomar valores por fuera del intervalo $[0,1]$, un requerimiento básico para ser una medida de probabilidad. Una forma de corregir esto se basa en la aplicación de modelos de elección binaria, entre los cuales se destaca el modelo logit (Luce (2012), Marschak (1959), McFadden et al. (1973)).

Si se considera a $p_i = P(Y = 1/X_i)$, entonces se define $g(p_i) = \beta_0 + \beta' X_i$, siendo $\beta' = (\beta_1, \dots, \beta_p)$ un vector de parámetros de coeficientes de las variables explicativas del modelo. A esta función $g(\cdot)$ se le llama función de enlace. En el caso concreto del modelo logit, la función de enlace presenta la siguiente expresión:

$$g(p_i) = \log\left(\frac{p_i}{1 - p_i}\right)$$

Utilizando la función inversa de $g(\cdot)$, dada por $g^{-1}(t) = \frac{e^{(t)}}{1 + e^{(t)}}$, se tiene que:

$$p_i = \frac{e^{(\beta_0 + \beta' X_i)}}{1 + e^{(\beta_0 + \beta' X_i)}}$$

4.2 Modelos No Paramétricos

Métodos no paramétricos más recientemente utilizados para estimar las probabilidades de incumplimiento se refieren a los modelos de precios de opciones, los modelos de árboles de decisión, y las redes neuronales. Estos métodos tienen el potencial de ser más útiles en la aplicación para préstamos comerciales, los cuales tienden a ser más heterogéneos que los préstamos de consumo o hipotecarios.

El modelo de precios de opciones consiste en comparar los pasivos del solicitante del crédito con una opción de venta (put) sobre sus activos. En base al modelo de Merton (1974), se deduce la probabilidad de incumplimiento de una empresa a partir de la estimación de la volatilidad del precio de sus activos (la cual suele depender de la volatilidad del precio de sus acciones). A partir de la identidad contable $Activos = Pasivos + PatrimonioNeto$ y del supuesto según el cual los accionistas reciben el valor residual de la empresa, si los pasivos superan el valor de los activos ($PN = 0$), se esperará que la empresa ejerza la opción de declararse en incumplimiento.

Para estimar la probabilidad de incumplimiento se calcula el valor de los pasivos y el valor de los activos. El valor de los pasivos se considera como un bono cupón cero que vence en el corto plazo, por lo tanto, es un valor fijo que surge del balance de la empresa. Sin embargo, en el caso de los activos, lo que se puede observar a corto plazo es su valor en libros, el cual puede diferir de su valor de mercado. El modelo de Merton (1974) permite establecer un vínculo entre los valores inobservables (valor de activos de la empresa y su volatilidad) y las variables observables.

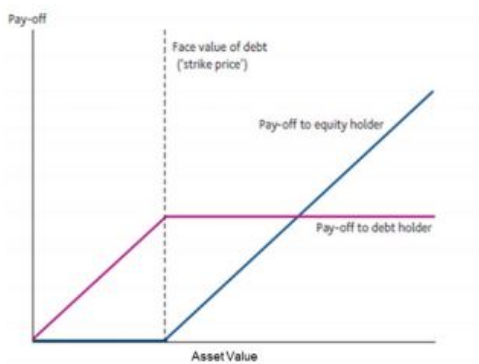


Figure 4.1: *Pagos de accionistas y acreedores*

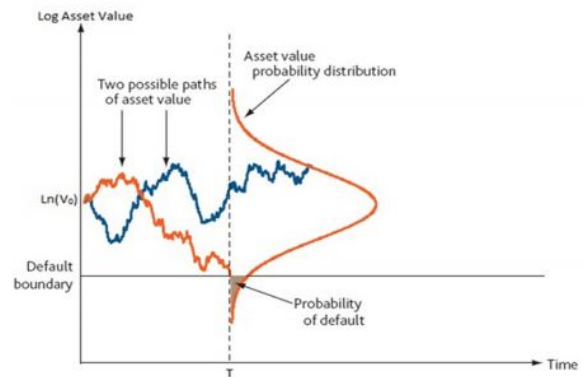


Figure 4.2: *Probabilidad de Incumplimiento*

Por otro lado, bajo el método de árboles de decisión se realizan particiones binarias de los datos de forma

recursiva (Laurent and Rivest (1976)). Las “ramas” del modelo representan conjuntos de decisiones, las cales generan sucesivas reglas para la clasificación de un conjunto de datos en subgrupos de datos disjuntos y exhaustivos. Para encontrar la mejor regla de división, el algoritmo estudia cada una de las variables explicativas, analizando los puntos de corte para así poder elegir la regla que otorgue la mayor homogeneidad posible a los nuevos subgrupos. El proceso culmina cuando ya no es posible realizar una nueva división que mejora la homogeneidad existente.

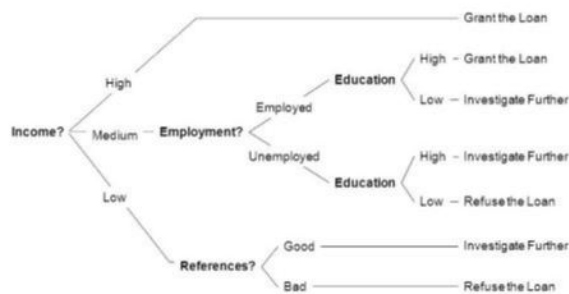
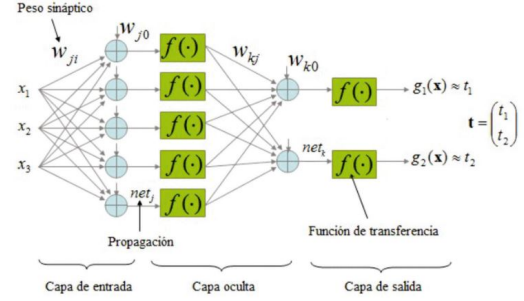
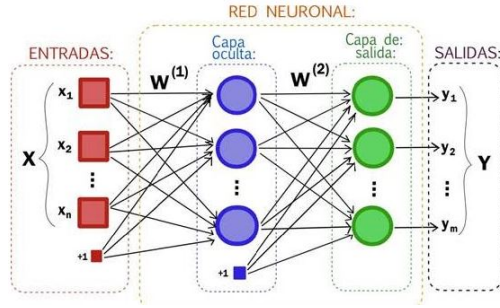


Figure 4.3

Finalmente, las redes neuronales son algoritmos de inteligencia artificial que permiten, a través de la experiencia, aprender automáticamente la relación existente entre las características del prestatario y la probabilidad de incumplimiento, así como la identificación de aquellas características que son las más importantes para predecir dicho incumplimiento. Se trata de un método más flexible, ya que no implica supuestos acerca de la forma funcional de la relación entre las características y la probabilidad de incumplimiento o acerca de las distribuciones de las variables o los errores del modelo.

Una red neuronal consiste en un modelo de nodos e interconexiones que recrea el diseño de las interconexiones neuronales del cerebro humano. La *neurona artificial* recibe información de diversas fuentes (X_i); dicha información no llega a la neurona en estado puro, sino que es ponderada (W_i). Al llegar a la neurona, todos los datos de las informaciones ponderados por sus pesos se suman (*regla de propagación*). Cuando dicha suma supera un cierto umbral, entonces la neurona se activa y el número resultante de la regla de propagación es el insumo de una *función de transferencia*. Generalmente, una red neuronal típica tiene tres capas: una de entrada de datos, otra oculta, donde se procesan los datos, y una tercera, de salida de los resultados. Todas las neuronas que contiene una capa se conectan con todas las neuronas de la siguiente capa.



5 Construcción del Puntaje

Una *scorecard* es una tabla que contiene los puntajes asignados a cada atributo de cada una de las variables utilizadas (a mayor puntaje mayor probabilidad de pago).

Una vez estimado el modelo, es posible re-escalar las probabilidades estimadas en puntuaciones, de acuerdo a la siguiente expresión:

$$Score = Offset + Factor \cdot \ln(Odds)$$

El *Offset* es lo que se denomina como término de traslación o compensación, en tanto que el *Factor* es un término de re-escalamiento. Se trata de valores impuestos por la institución financiera para estandarizar las puntuaciones y así comparar distintos modelos de puntaje. El *Offset* determina la puntuación equivalente a un *Odds* 1:1 (igual proporción de clientes buenos y malos), y el *Factor* determina el incremento en la puntuación necesario para doblar los *Odds*.

Los valores de la tabla de puntaje crediticio surgen de transformar los coeficientes del modelo estimado. La transformación busca obtener valores enteros para cada atributo j de la característica i dentro de un rango considerado adecuado por la institución financiera. Se suele calibrar la *scorecard* de forma de que, cada cierto aumento en el puntaje P_0 , se obtenga el doble de la relación *Odds*.

Entonces, para obtener los valores de *Offset* y de *Factor*, se resuelve el siguiente sistema de ecuaciones:

$$Score = Offset + Factor.ln(Odds)$$

$$Score + P_0 = Offset + Factor.ln(2.Odds)$$

Se obtiene:

$$Factor = \frac{P_0}{ln(2)}$$

$$Offset = Score - Factor.ln(Odds)$$

De esta manera, si se parte de un puntaje de 500 como el equivalente a un *odds* de 1:1 y que los *odds* se duplican cada $P_0 = 50$ puntos en la scorecard, entonces 550 puntos equivalen a un *odds* de 2:1, y 600 equivalen a una relación de 4:1 y así sucesivamente. Por lo tanto:

$$Factor = \frac{50}{ln(2)} = 72,135$$

$$Offset = 500 - 72,135.ln(1) = 500$$

Entonces, se obtiene la función de Puntaje o Score:

$$Score = 500 + 72,135.ln(Odds)$$

Finalmente, combinando esta expresión con el modelo de regresión logística:

$$Score = Offset + Factor.ln(Odds)$$

$$Score = Offset + Factor. \left(\hat{\beta}_0 + \sum_{i=1}^p \hat{\beta}_i.woe_{ij} \right)$$

$$Score = Offset + Factor\hat{\beta}_0 + Factor. \left(\sum_{i=1}^p \hat{\beta}_i.woe_{ij} \right)$$

De esta manera, los puntajes en una scorecard se pueden descomponer en un puntaje inicial ($Offset + Factor\hat{\beta}_0$) y un puntaje asociado al atributo j de la característica i ($Factor.\hat{\beta}_i.woe_{ij}$).

Cuando ingresa una nueva solicitud de crédito, se calcula el puntaje del cliente y se decide otorgar o no el crédito según que su puntaje supere lo que se conoce como *punto de corte*.

6 Técnicas de validación de los modelos de puntaje crediticio

Una vez que se construye el modelo, es necesario validarlo. Para ello, se aplican técnicas estadísticas que permiten medir y comparar la bondad del ajuste del modelo, la exactitud de sus predicciones y su poder discriminante.

6.1 Criterios globales

Se pueden utilizar una serie de criterios globales que nos ayudan a elegir un modelo de modo que se tengan en cuenta el ajuste y el número de parámetros del modelo. Algunos de los criterios globales más utilizados son:

- **Criterio de Información de Akaike (AIC):** cuantifica la distancia entre el modelo teórico y el estimado. El estadístico viene dado por: $AIC = -2L(\hat{\beta}) + 2p$, donde $l(\hat{\beta})$ es el logaritmo de la función de verosimilitud y el $2p$ es un término de penalización que depende del número de parámetros estimados en el modelo (p). Se basa en la idea de que un modelo es mejor cuanto mayor sea su función de verosimilitud y menor su número de parámetros. La **regla de decisión** es que, entre dos modelos, es preferible aquel con menor AIC.
- **Criterio de Información de Bayes (BIC) o Schwarz:** este criterio ha demostrado ser más adecuado que el anterior cuando el tamaño de la muestra incide negativamente en la complejidad del modelo. También es un criterio de verosimilitud penalizada, pero con una mayor penalización con respecto al número de variables del modelo. El estadístico viene dado por: $BIC = -2L(\hat{\beta}) + \log(n)p$, en donde el término $\log(n)p$ busca penalizar por el efecto del tamaño de la muestra (n) y por el número de parámetros estimados (p). **Regla de decisión:** se elige aquel modelo con menor BIC.

6.2 Evaluación del poder discriminante del modelo

Para evaluar la capacidad discriminante del modelo se pueden utilizar los siguientes métodos: el test de Kolmogorov-Smirnov, el Índice de Gini, y las Curvas ROC.

6.2.1 Test de Kolmogorov-Smirnov

El estadístico KS es un test de bondad de ajuste utilizado para probar que dos muestras independientes provienen de la misma distribución. A partir del cálculo de la máxima distancia absoluta entre dos distribuciones empíricas, el estadístico KS busca detectar las discrepancias entre las frecuencias relativas acumuladas de las muestras consideradas (en este caso, de buenos y malos). Estas diferencias están determinadas no sólo por las medias, sino también por la dispersión y, en general, la forma de las dos distribuciones. Si la distancia máxima es menor que un valor crítico, entonces se concluye que las distribuciones son iguales.

Se parte de las siguientes hipótesis:

H0: Las distribuciones de buenos y malos son iguales.

H1: Las distribuciones de buenos y malos son distintas.

Concretamente, a partir de las muestras de clientes buenos y malos (de tamaño n_1 y n_2 respectivamente), se agrupan los datos en k categorías o intervalos y se calculan las frecuencias relativas acumuladas $F_i(x)$ y $G_i(x)$, con $i = 1, 2, \dots, k$ (x es el puntaje del scorecard). A continuación, se calculan las diferencias de las frecuencias relativas acumuladas.

El estadístico de Kolmogorov-Smirnov es la diferencia máxima de las distribuciones de frecuencias relativas acumuladas:

$$KS = \max_{1 \leq i \leq k} |F_{n_1}(x) - G_{n_2}(x)|$$

Se concluye comparando el valor del estadístico con un valor crítico, el cual se calcula de acuerdo a la siguiente expresión:

$$Val_{crit} = 100 \cdot K \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

donde K es el valor obtenido de la tabla Kolmogorov-Smirnov con $n_1 + n_2 - 2$ grados de libertad, para un nivel de confianza dado.

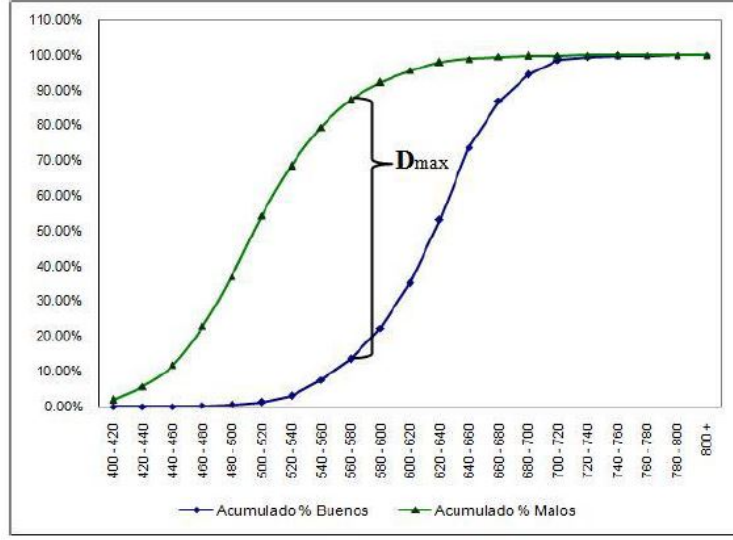


Figure 6.1: *Distribución Acumulada de buenos y malos*

Se rechaza H_0 si $KS > Valcrit$, lo que quiere decir que existe evidencia suficiente para rechazar que las distribuciones de buenos y malos son iguales (hay discriminación entre las dos poblaciones). Por ejemplo, en la Figura 6.1, se obtiene un KS de 87,3 a un nivel de puntaje de 570.

6.2.2 Curva de Lorenz

La Curva de Lorenz representa la relación entre el Score y las funciones de distribución acumuladas de los casos buenos y malos. Formalmente, la curva de Lorenz de las funciones de distribuciones de clientes buenos y malos, $F(x)$ y $G(x)$ es el subconjunto del producto cartesiano dado por:

$$L(F, G) = \{(u, v) / u = F(x), v = G(x); \text{con } x \in R\}$$

Si $F(x) = G(x)$ entonces $L(F, G)$ describe la recta $u = v$ con $u \in (0, 1)$ entre las distribuciones $F(x)$ y $G(x)$. Por lo tanto, cuanto más se separe $L(F, G)$ de la recta $u = v$, mayor será la diferencia entre $F(x)$ y $G(x)$.

Como se puede observar en la Figura 6.2, el área que se encuentra entre la recta de 45 y la curva de Lorenz es una medida de desigualdad entre las distribuciones $F(x)$ y $G(x)$. Se denota por T al triángulo delimitado por la recta $u = v$, el eje horizontal y la recta $u = 1$.

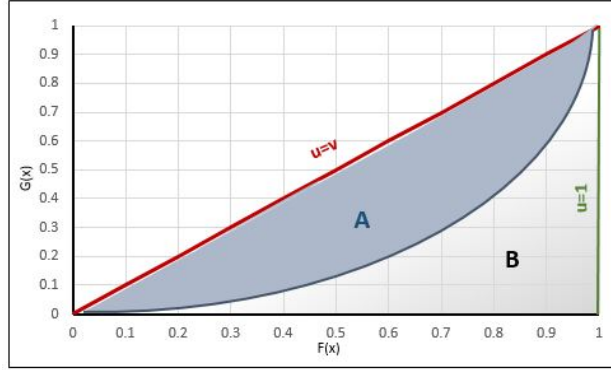


Figure 6.2

Algunas variantes de la Curva de Lorenz son la Curva ROC (receiver operating characteristic) y la Curva de Performance.

6.2.3 Índice Gini

Otra herramienta usualmente utilizada para evaluar el poder discriminante de los modelos de clasificación es el Índice de Gini. Se trata de un método gráfico que ilustra el grado de desigualdad en la distribución de la variable en cuestión (el mismo se deriva de la curva de Lorenz). Sin embargo, Gini también se puede utilizar para determinar la potencia de una variable individual. Las características de la variable se agrupan de la mejor a la peor situación de riesgo, mediante el establecimiento de un punto de corte, según el modelo que se construirá, y las variables se pueden elegir para su inclusión en el proceso de modelado en base a este punto de corte.

El área debajo de la Curva de Lorenz se puede aproximar a partir de la suma de las áreas de un conjunto de trapecios. El área del i -ésimo trapecio surge de la siguiente expresión:

$$B_i = \frac{(F_i - F_{i-1}) - (G_i - G_{i-1})}{2}$$

Por lo tanto, la superficie del área total que se encuentra por debajo de la Curva de Lorenz es:

$$B = \sum_{i=2}^k B_i$$

Finalmente, el Índice Gini es:

$$Gini = \frac{1/2 - B}{1/2}$$

Gráficamente (Figura 6.2), el índice de Gini se expresa como:

$$Gini = A/A + B$$

Es decir, es el cociente entre el área comprendida entre la línea de igualdad perfecta y la curva de Lorenz, y el área delimitada entre la línea de igualdad perfecta y el eje horizontal.

El rango de valores del coeficiente de Gini es $[0, 1]$. Cuanto más cerca esté el coeficiente a uno, mejor será la separación de los clientes morosos con respecto a los no morosos.

6.2.4 Curva ROC y AUC

Una alternativa gráfica a la Curva de Lorenz es la Curva ROC. La misma se construye trazando el ratio de verdaderos positivos (“sensibilidad”) con respecto al ratio de falsos positivos (“1-especificidad”). Una de las ventajas de la Curva ROC como medida de validación se refiere a la flexibilidad de su construcción, ya que no depende de supuestos sobre el método probabilístico detrás de la regla de discriminación entre buenos y malos.

Formalmente, supongamos que tenemos una cartera de n clientes, algunos provenientes de la población de morosos y otros de la población de no morosos.

Sea M una variable aleatoria que sigue una distribución Bernoulli con parámetro p (“prevalencia del evento”) y que toma el valor de 1 si el individuo pertenece a la población de morosos y 0 en caso contrario. De esta manera, $p = P(M = 1)$ y $1 - p = P(M = 0)$.

Por otro lado, sea Y una variable aleatoria que se denomina prueba o marcador y que toma dos resultados posibles: Positivo, P , si el individuo se clasifica como moroso o Negativo, N , en caso de que se clasifique como no moroso. La idea es evaluar el poder discriminante de Y .

Un estimador de p para una muestra de la población es:

$$p = \frac{\text{cantidad de morosos en la muestra}}{\text{total de individuos en la muestra}}$$

Para evaluar la validez o poder discriminatorio de la prueba con respecto al verdadero estado (moroso o no moroso) se establece una comparación entre el resultado Y y el estado M , dando lugar al siguiente cuadro de contingencias:

	$Moroso \equiv M = 1$	$NoMoroso \equiv M = 0$
$Prueba+ \equiv Y = 1$	Verdadero Positivo V_+	Falso Positivo F_+
$Prueba- \equiv Y = 0$	Falso Negativo F_-	Verdadero Negativo V_-

Si un individuo no moroso se clasifica y predice correctamente como no moroso, entonces estamos ante un “verdadero positivo”, V_+ ; por otro lado, si un individuo moroso se clasifica correctamente como moroso, entonces se trata de un “verdadero negativo”, V_- . Puede haber dos tipos de errores, un resultado falso positivo, F_+ (clasificar a un individuo no moroso como moroso, error de tipo I) y un resultado falso negativo, F_- (clasificar a un individuo moroso como no moroso, error de tipo II).

Para medir la eficacia de la prueba se utilizan los conceptos de especificidad y sensibilidad.

La sensibilidad es un parámetro que se mide en el grupo de individuos que verdaderamente son morosos y se define como el cociente entre verdaderos positivos sobre el total de individuos morosos. De esta manera, la sensibilidad es la probabilidad de obtener un resultado positivo cuando el cliente es moroso ($E = 1yY = 1$, proporción de verdaderos positivos o True Positive Fraction, TPF).

$$S = TPF = \frac{\text{Morosocorrectamenteclasificados}}{\text{Totaldemorosos}}$$

$$S = \frac{V_+}{F_- + V_+}$$

De esta manera, si $S = 1$, entonces $F_- = 0$.

Por otro lado, la especificidad es un parámetro que se mide dentro del grupo de individuos no morosos. Se define como el cociente entre verdaderos negativos y el total de no morosos. Por lo tanto, es la probabilidad de obtener un resultado negativo cuando el individuo no es moroso ($E = 0yY = 0$, proporción de verdaderos negativos o True Negative Fraction, TNF).

$$E = TNF = \frac{\text{NoMorososincorrectamenteclasificados}}{\text{Totaldenomorosos}}$$

$$E = \frac{V_-}{V_- + F_+}$$

Por lo que el complemento de E es:

$$1 - E = FPF = \frac{F_+}{V_- + F_+}$$

	$Moroso \equiv M = 1$	$NoMoroso \equiv M = 0$
$Prueba+ \equiv Y = 1$	$S = TPF = \frac{V_+}{F_- + V_+}$	$FPF = 1 - E = \frac{F_+}{F_+ + V_-}$
$Prueba- \equiv Y = 0$	$1 - S = FNF = \frac{F_-}{F_- + V_+}$	$E = TNF = \frac{V_-}{F_+ + V_-}$

La Curva ROC se obtiene a partir del cociente de morosos correctamente clasificados contra los no morosos incorrectamente clasificados, donde el eje de abscisas se corresponde con FPF y el eje de las ordenadas con TPF . La información aportada por la Curva ROC suele resumirse a través del área debajo de la curva (area under curve, AUC o AUROC, Hanley and McNeal (1982)), cuyos valores se ubican en el rango $[0,1]$.

Si el $AUROC$ toma el valor de 1, entonces la prueba es perfecta, lo que equivale a decir que el modelo clasifica correctamente al 100% de los morosos como morosos y al 100% de los no morosos como no morosos. Si el área debajo de la Curva ROC toma el valor de 0,5, entonces quiere decir que la probabilidad de clasificar incorrectamente a los individuos es de 50%. Por lo tanto, estaríamos ante una prueba o marcador que no proporciona información para discriminar entre clientes morosos y no morosos. La Curva ROC asociada a este caso presentaría valores de sensibilidad iguales a (1-especificidad) en todos los posibles puntos de corte. El área de un marcador no informativo se corresponde a la diagonal del cuadrado, por lo que se suele representar esta diagonal para visualizar en cuánto supera la prueba que se valora a lo que sería una *prueba inútil* (la diagonal representa una situación de clasificación al azar). Por último, si el área debajo de la Curva ROC toma un valor menor a 0,5, esto significaría que la clasificación comete un error mayor que la mera clasificación al azar. Por lo tanto, cuanto mayor sea el valor del $AUROC$, mayor será la capacidad discriminante del modelo evaluado.

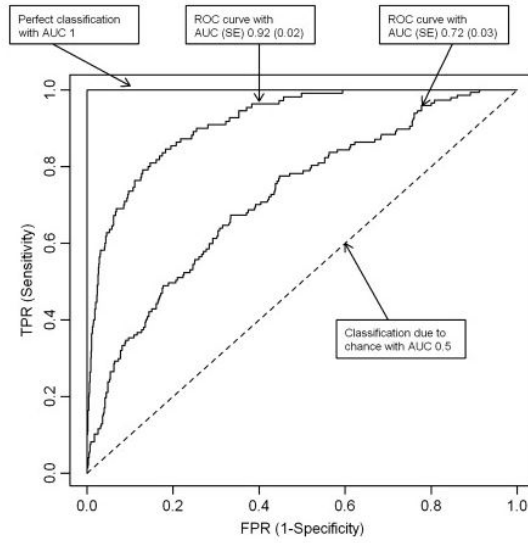


Figure 6.3

Como se puede observar en la Figura 6.4, si las distribuciones de morosos y no morosos fuesen iguales, las mismas estarían solapadas y la prueba sería inútil para detectar diferencias entre ambos.

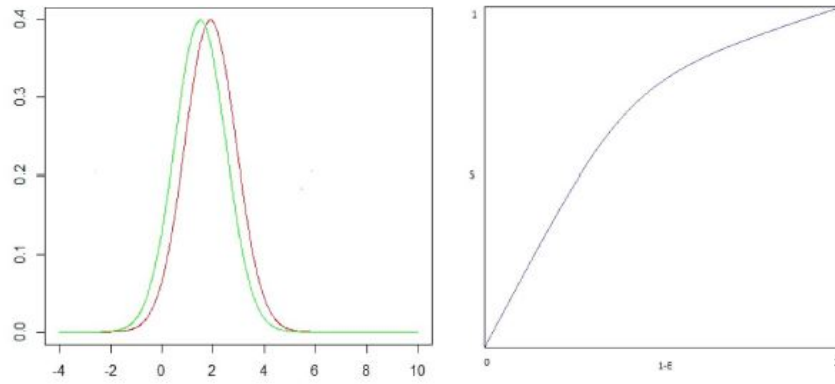


Figure 6.4

Cuánto más alejadas estén las distribuciones asociadas a las poblaciones de morosos y no morosos, menos solapamiento habrá y, por tanto, menos falsos negativos y falsos positivos (Figura 6.5).

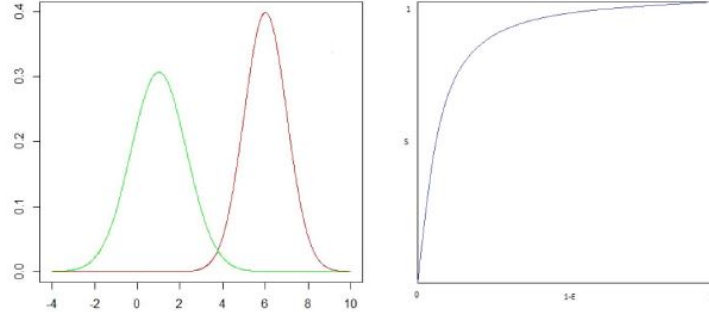


Figure 6.5

Por último, cabe mencionar que el Índice Gini es una función del área bajo la curva:

$$Gini = 2AUROC - 1$$

6.2.5 Population Stability Index (PSI)

Un estadístico muy utilizado para comprobar que la distribución actual de la población es similar a la utilizada para construir el modelo es el Population Stability Index (Yurdakul (2018), Taplin and Hunt (2019)).

El índice de estabilidad de la población (PSI) busca medir cuánto ha cambiado la distribución de una variable entre dos muestras o con el tiempo. Es ampliamente utilizado para monitorear posibles cambios en las características de una población y para diagnosticar cualquier problema en el desempeño del modelo. Si una o más variables han cambiado significativamente, generalmente sugiere la necesidad de actualizar o reconstruir un modelo.

El PSI se calcula de la siguiente manera:

$$PSI = \sum_{i=1}^K \ln \left(\frac{f_{Observada_i}}{f_{Esperada_i}} \right) \cdot (f_{Observada_i} - f_{Esperada_i})$$

En donde $f_{Observada}$ y $f_{Esperada}$ son las funciones de distribución de clientes en la muestra actual (*muestra de validación*) y la muestra utilizada en el modelo (*muestra de modelo*) respectivamente, y K son las categorías según las cuales se dividen los datos (rangos de edad, por ejemplo).

Generalmente, se establecen los siguientes criterios a la hora de valorar el PSI:

- PSI menor a 0,1: la población presenta cambios muy poco significativos.

- PSI entre 0,1 y 0,25: la población presenta pequeños cambios.
- PSI mayor de 0,25: la población presenta cambios significativos.

A modo de ejemplo, supongamos que contamos con el siguiente puntaje crediticio:

Grupo	Mediana del Score	Casos Esperada	Casos Observada	% Esperada	% Observada	PSI
1	42	3718	154	10.6%	6.0%	0.0255
2	69	3795	172	10.8%	6.8%	0.0191
3	90	3239	141	9.2%	5.5%	0.0188
4	111	3537	195	10.1%	7.7%	0.0066
5	138	3320	189	9.5%	7.4%	0.0049
6	168	3596	301	10.2%	11.8%	0.0023
7	201	3457	298	9.8%	11.7%	0.0032
8	249	3515	369	10.0%	14.5%	0.0165
9	321	3444	412	9.8%	16.2%	0.0318
10	444	3503	317	10.0%	12.4%	0.0055
Total		35124	2548	100%	100%	0.1342

De esta manera, el PSI asociado al Grupo 2 es:

$$PSI_2 = \ln \left(\frac{10.8\%}{6.8\%} \right) \cdot (10.8\% - 6.8\%) = 0.0191$$

En tanto que el PSI total es la suma de los índices asociados a cada grupo, es decir, 0.1342. De acuerdo a los criterios previamente mencionados, se observa un pequeño cambio en la distribución de la población en el score, por lo que podría ser necesario ajustar el modelo.

Como se observa en la Figura 6.6, en la muestra del modelo se registra una distribución bastante equitativa a lo largo de los distintos grupos, en tanto que en la muestra de validación se observan mayores porcentajes de casos en los tramos de puntaje más elevados, lo cual podría explicarse por un cambio en las políticas crediticias.

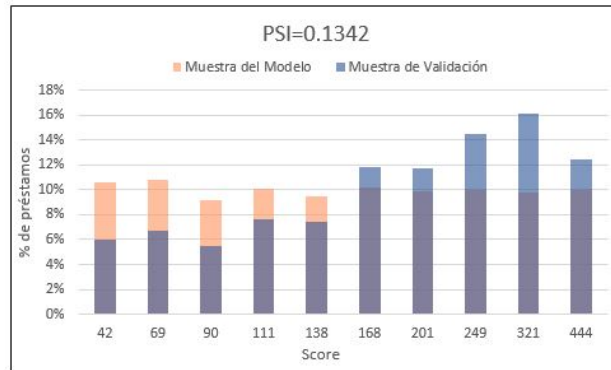


Figure 6.6

7 Aplicación de puntajes crediticios en Uruguay

En la presente Sección, se procede a analizar el grado de implementación de modelos de scoring o puntaje crediticio en las Empresas Administradoras de Crédito en Uruguay. La motivación de dicho análisis responde a dos factores: primero, el crecimiento registrado en la actividad de estas empresas dentro del sistema financiero uruguayo (los activos de las EACs crecieron aproximadamente un 60% en términos reales entre Diciembre de 2010 y Diciembre de 2018) y, en segundo lugar, la importancia relativa que presenta el riesgo de crédito entre los riesgos asociados al sistema financiero uruguayo.

De acuerdo a la normativa de la Superintendencia de Servicios Financieros (SSF) del Banco Central del Uruguay (BCU), las Empresas Administradoras de Crédito (EACs) “son aquellas personas físicas o jurídicas que en forma habitual y profesional intervienen en el financiamiento de la venta de bienes y servicios realizada por terceros, otorgando crédito mediante el uso de tarjetas, órdenes de compra u otras modalidades similares”. Adicionalmente, se “consideran Empresas Administradoras de Crédito las cooperativas de consumo, asociaciones civiles y otras personas jurídicas con giro no financiero que emitan en forma habitual y profesional órdenes de compra en tanto tal actividad sea significativa dentro del conjunto de actividades que conformen su giro”.

Las EACs se agrupan en dos grandes categorías: las empresas administradoras de crédito de mayores activos y las de menores activos. Para pertenecer al grupo de EACs de mayores activos se debe cumplir la condición de que la suma de los activos totales más contingencias de la institución al cierre del ejercicio

superen el equivalente a 100.000 Unidades Reajustables (RNRCSF²).

Para evaluar el grado de avance en materia de puntaje crediticio en las EACs, se trabajó con datos de una encuesta realizada en el año 2014, en la cual se incluyeron preguntas referidas a la implementación de un puntaje de admisión y de comportamiento³. Las respuestas analizadas corresponden a 14 EACs mayores y a 15 EACs menores (este grupo incluye una institución que dejó de otorgar créditos a partir de setiembre 2009). Actualmente, el sistema financiero uruguayo cuenta con 12 EACs de menores activos y 16 empresas de mayores activos.

Administradores de Créditos Mayores		
Cantidad de empresas entrevistadas	14	
Utilizan credit scoring?	Si	8
	No	6
Si respondió que no, tienen planado hacerlo?	Si	6
	No	0
Si respondió que sí, es vinculante?	Si	4
	No	2
	Vinculante para algunos productos	1
	Vinculante para otorgar, no para negar créditos	1
Desde qué fecha es vinculante?	menos 1 año	1
	entre 1 y 2 años	2
	mas de 2 años	3
El uso del Scoring ha mejorado la eficiencia en el sentido de	Mayor rapidez en la autorización de créditos	1
	Disminución de la morosidad en clientes nuevos	1
	Disminución de los costos de evaluación de créditos	0
	Mayor rapidez en la evaluación/autorización y Disminución de los costos de evaluación	2
	Mayor rapidez en autorización, Disminución morosidad clientes nuevos y Disminución costos evaluación créditos	3
	Otros (segmentación cartera, mejor asignación límites, estandarización criterios)	3
Cantidad de variables consideradas	cantidad de variables < 5	2
	5 ≤ cantidad de variables ≤ 10	4
	10 ≤ cantidad de variables ≤ 20	2
Tipo de variables consideradas	Sociodemográficas	7
	Relativas a la solicitud	0
	Todas las anteriores	1
Cuentan con validaciones del modelo de score?	Si	3
	No	5
Motivo de no uso	Los costos no compensan los beneficios	1
	La imposibilidad de que el modelo abarque todo lo que hay que tener en cuenta para conceder un crédito	1
	No cree en su utilidad	0
	Otros (empresa nueva, barreras de sistemas de gestión, se hará en el futuro, tenían y quedó obsoleto)	4
	No contesta	1
Cuentan con un score de comportamiento?	Si	2
	No	12
En caso de no tener, tienen planeado tenerlo?	Si	4
	No	6
	A evaluar	2

Como se puede observar en la Tabla, un 57% de las EACs mayores tenía implementado un modelo de credit scoring para las decisiones de admisión u originación de los créditos. En cambio, sólo un 14% declaraban contar con un score de comportamiento para evaluar el riesgo crediticio de los individuos que ya formaban parte de su clientela. Considerando que ha pasado un período de tiempo razonable desde que se realizó la encuesta, resultaría importante diseñar una nueva versión de la misma para evaluar si el 33% de las empresas que en su momento respondieron que planeaban implementar un puntaje de comportamiento efectivamente lo hicieron. Asimismo, como se puede observar en la Tabla, la totalidad de las EACs mayores que declaró no contar con un puntaje de originación respondió que tenía planeado implementarlo en el futuro,

²<https://www.bcu.gub.uy/Acerca-de-BCU/Normativa/Documents/Reordenamiento%20de%20la%20Recopilaci%C3%B3n/Sistema%20Financiero/RNRCSF.pdf>

³Ver detalle de cuestionario en el Anexo A.

con lo que también sería relevante verificar su cumplimiento al haber pasado cinco años.

Por otro lado, como se puede observar en la Figura 7.1, entre las empresas que en su momento contaban con un puntaje de admisión, el 50% lo definía como un elemento vinculante para la toma de decisión de concesión del crédito. Adicionalmente, el principal motivo de mejora de eficiencia derivada del scoring es la de mayor rapidez en la autorización de créditos, seguida por el motivo de disminución de costos en la evaluación de créditos.



Figure 7.1

Por último, en lo que refiere a las variables utilizadas en los modelos, la mayor parte de las empresas encuestadas incluían entre 5 y 10 variables, siendo la edad, el nivel de ingresos y los antecedentes crediticios las más significativas a la hora de conceder o no un crédito.

En lo que refiere a las EACs menores, tal como se observa en la Tabla siguiente, un 50% de las mismas contaban con modelos de puntaje crediticio en 2014, en tanto que aproximadamente un 43% de las restantes tenían planeado implementar uno o lo estaban evaluando. De las que contaban con scoring, más del 70% lo consideraban un elemento vinculante a la hora de conceder un crédito. Adicionalmente, sólo un 21% declaraban contar con un score de comportamiento, en tanto que un 50% de las que respondieron que no contaban con el mismo consideraban probable su incorporación en un futuro cercano. Nuevamente, habiendo pasado ya cinco años, en una nueva versión de la encuesta se podrá verificar el grado de cumplimiento de los planes declarados.

Administradores de Créditos Menores		
Cantidad de empresas entrevistadas	14	
Utilizan credit scoring?	Sí	7
	No	7
Si respondió que no, tienen planeado hacerlo?	Sí	2
	No	3
	Se va a evaluar	1
Si respondió que sí, es vinculante?	No contesta	1
	Sí	5
	No	1
Desde qué fecha es vinculante?	Es vinculante para seguir el proceso de otorgamiento	1
	menos 1 año	2
	entre 1 y 2 años	2
El uso del Scoring ha mejorado la eficiencia en el sentido de	más de 2 años	2
	Mayor rapidez en la evaluación/autorización de créditos	1
	Disminución de la morosidad en clientes nuevos	0
	Disminución de los costos de evaluación de créditos	0
	Mayor rapidez en la evaluación/autorización y Disminución de la morosidad clientes nuevos	2
	Mayor rapidez en la evaluación/autorización y Disminución de los costos de evaluación	1
	Disminución morosidad clientes nuevos y Disminución de costos de evaluación	2
	Mayor rapidez en autorización, Disminución morosidad clientes nuevos y Disminución costos evaluación créditos	0
Cantidad de variables consideradas	Otros (mayor dinámica y eficiencia)	1
	cantidad de variables < 5	3
	5 ≤ cantidad de variables ≤ 10	4
	10 ≤ cantidad de variables ≤ 20	0
Tipo de variables consideradas	Sociodemográficas	6
	Relativas a la solicitud	0
	Todas las anteriores	1
Cuentan con validaciones del modelo de score?	Sí	1
	No	7
Motivo de no uso	Los costos no compensan los beneficios	2
	La imposibilidad de que el modelo abarque todo lo que hay que tener en cuenta para conceder un crédito	1
	No cree en su utilidad	0
	Otros (la automatización lleva a un aumento de la morosidad, no conoce los costos, no se están otorgando créd.)	3
	No contesta	2
Cuentan con un score de comportamiento?	Sí	3
	No	10
	No contesta	1
En caso de no tener, tienen planeado tenerlo?	Sí	3
	No	5
	A evaluar	2

En tanto, los principales motivos de mejora derivados del score crediticio se refieren a la disminución de la morosidad en los clientes nuevos y la mayor rapidez en el proceso de autorización de los créditos. Si se compara con los motivos declarados por parte de las EACs mayores, se observa que mientras que el peso del motivo asociado a la mayor rapidez en el proceso es el mismo, el de disminución de la morosidad presenta una mayor participación dentro de las respuestas de las EACs menores.



Por otro lado, entre las variables que resultan más significativas en los modelos de scoring utilizados por las EACs menores, la más importante es la asociada a los antecedentes de pago del cliente, seguida luego

por la edad, el nivel de ingresos y las consultas en el Clearing de Informes. El resto de las variables son de carácter socio-demográfico, en tanto que sólo en un caso se incluyen variables relativas a la solicitud del crédito (características del crédito solicitado).

Finalmente, cabe destacar que, si bien existen diferencias entre las EACs mayores y menores, ambas presentan bajos niveles de utilización de herramientas de validación de sus modelos de puntaje crediticio (cerca de un 38% en el caso de las EACs mayores y sólo un 13% de las EACs menores). Las diferencias entre ambos grupos de instituciones tiene sentido, ya que el reducido tamaño de algunas de las EACs menores puede impactar en los costos de asignar recursos a la aplicación de dichas herramientas. No obstante, dado que el 71% de las EACs menores utiliza el scoring como un elemento vinculante en la decisión de concesión de crédito, podría ser relevante evaluar el poder predictivo de los modelos utilizados.

8 Conclusiones

Las instituciones financieras enfocadas en la concesión de préstamos al consumo se enfrentan a dos tipos de decisiones: por un lado, deben decidir si conceder un préstamo a un cliente nuevo y, por otro, deben decidir cómo gestionar los préstamos que ya concedieron. Las herramientas estadísticas usualmente utilizadas para tomar el primer tipo de decisiones son las asociadas a los modelos de puntaje de admisión u originación, en tanto que las decisiones de gestión de límites de créditos o de estrategias de marketing para clientes ya existentes se apoyan en las técnicas de puntaje de comportamiento.

En este trabajo se analizaron las metodologías más usualmente utilizadas a la hora de construir un puntaje crediticio, así como las herramientas aplicadas para evaluar la validación de los modelos. Adicionalmente, se evalúa el grado de implementación de puntajes crediticios entre las Empresas Administradoras de Crédito en Uruguay a partir de datos de una encuesta realizada en 2014.

A partir de los datos analizados, se observa que un 57% de las EACs de mayores activos tenía implementado un modelo de credit scoring para las decisiones de admisión de los créditos. En cambio, sólo un 14% declaraban contar con un score de comportamiento para evaluar el riesgo crediticio de los individuos que ya formaban parte de su clientela. Entre las empresas que contaban con un score de admisión, el 50% lo definía como un elemento vinculante para la toma de decisión de concesión del crédito. En lo que refiere a las EACs de menores activos, un 50% de las mismas contaban con modelos de scoring crediticio en 2014, en

tanto que sólo un 21% declaraban contar con un score de comportamiento. De las que contaban con scoring, más del 70% lo consideraban un elemento vinculante a la hora de conceder un crédito. Cabe destacar que, si bien existen diferencias entre las EACs de mayores y menores activos, ambas presentan bajos niveles de utilización de herramientas de validación de sus modelos de scoring. Por último, en lo que refiere a las variables utilizadas en los modelos, la mayor parte de las empresas encuestadas incluían entre 5 y 10 variables, siendo la edad, el nivel de ingresos y los antecedentes crediticios las más significativas a la hora de conceder o no un crédito.

Dado que ha pasado un período de tiempo razonable, sería interesante realizar una nueva versión de la mencionada encuesta con el fin de verificar el grado de avance en la implementación de modelos de scoring en las principales instituciones financieras dedicadas a la concesión de créditos al consumo en Uruguay. Adicionalmente, también sería interesante evaluar si han habido cambios en las variables consideradas, principalmente a la luz de los recientes cambios regulatorios que definieron un precio para los datos masivos de la central de riesgos crediticios del BCU.⁴

⁴Circular N° 2.317, BCU.

A Cuestionario sobre métodos de Scoring de Admisión y de Comportamiento

Las 9 primeras preguntas están relacionadas a los scoring de admisión y la última a los scoring de comportamiento.

1. ¿Utiliza un método de scoring como una herramienta para autorizar los créditos?
2. En caso de que la respuesta a la pregunta 1) fuera NO ¿Tiene planeado instrumentar un método de scoring en el futuro?
3. ¿ El scoring que utiliza es vinculante o un elemento más a tener en cuenta?
4. En el caso de que el scoring fuera vinculante, ¿en que fecha adquirió este carácter?
5. El uso del Scoring ha mejorado la eficiencia en el sentido de:
 - Mayor rapidez en la autorización de créditos
 - Disminución de la morosidad en clientes nuevos
 - Disminución de los costos de evaluación de créditos
 - Las dos anteriores
 - Otros. Especificar
6. Especifique las variables que incluye su scoring
7. ¿Cuales son las variables que explican en mayor medida la concesión o no del crédito en el modelo de scoring ?
8. ¿Se han realizado estudios acerca de la capacidad discriminante del método? En caso afirmativo, establezca la metodología utilizada para evaluar la capacidad discriminante y los resultados obtenidos.
9. En caso que la respuesta 1) fue NO. ¿Cuál es el motivo?
 - Los costos no compensan los beneficios del sistema

- La imposibilidad de que el modelo abarque todo lo que hay que tener en cuenta para conceder un crédito
- No los cree útiles
- Otros – explicitar

10. ¿Utiliza un scoring de comportamiento? En caso negativo, ¿piensa desarrollarlo en el futuro?

References

- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*.
- Durand, D. (1941). *Risk elements in consumer installment financing*. National Bureau of Economic Research, New York.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*.
- Hanley, J. and B. McNeal (1982). A simple generalization of the area under the roc curve to multiple class classification problems. *Radiology*.
- Laurent, H. and R. L. Rivest (1976). Constructing optimal binary decision trees is np-complete. *Information processing letters*.
- Leung, K., F. Cheong, C. Cheong, S. O'Farrell, and R. Tissington (2008). Building a scorecard in practice. In *Proceedings 7th International Conference on Computational Intelligence in Economics and Finance, Taiwan*.
- Luce, R. D. (2012). *Individual choice behavior: A theoretical analysis*. Courier Corporation.
- Marschak, J. (1959). Binary choice constraints and random utility indicators. Technical report, YALE UNIV NEW HAVEN CT COWLES FOUNDATION FOR RESEARCH IN ECONOMICS.
- McFadden, D. et al. (1973). Conditional logit analysis of qualitative choice behavior.
- Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of finance*.
- Mester, L. J. et al. (1997). What is the point of credit scoring? *Business review*.
- Siddiqi, N. (2012). *Credit risk scorecards: developing and implementing intelligent credit scoring*. John Wiley & Sons.
- Taplin, R. and C. Hunt (2019). The population accuracy index: A new measure of population stability for model monitoring. *Risks*.
- Thomas, L. C. (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International journal of forecasting*.
- Yurdakul, B. (2018). Statistical properties of population stability index.
- Zeng, G. (2013). Metric divergence measures and information value in credit scoring. *Journal of Mathematics*.