

Modelo para la medición del riesgo de insolvencia empresarial: PYME de Colombia, un caso de estudio¹

Model for measuring the risk of corporate failure: SMEs in Colombia, a case study

E. A. Cruz, J. Espinosa y S. Aristizábal

Recibido Agosto 01 de 2013 – Aceptado Noviembre 20 de 2014

Resumen— El documento muestra una descripción matemática y operativa de la técnica de regresión logística de acuerdo a disertaciones de autores que han trabajado profundamente su aplicación empírica. Se define el concepto de riesgo de crédito en un contexto estadístico. Sigue una descripción de las actividades económicas consideradas en el estudio. Luego, se sintetiza la construcción en SPSS 19 del modelo de regresión logística: definición del grupo de entrenamiento y la selección de variables relacionadas con la solvencia empresarial. Finalmente, los resultados se presentan en tablas, con la codificación de las variables, estadísticos de bondad de ajuste y de poder de pronóstico para determinar la probabilidad de que una empresa PYME caiga en estado de insolvencia y genere un riesgo a sus acreedores. Se encontró que las empresas que reportaron pérdidas acumuladas tienen ocho veces más probabilidades de caer en estado de insolvencia que las empresas con utilidades acumuladas positivas.

Palabras Clave— probabilidad de insolvencia, PYME colombianas, regresión logística, riesgo de crédito, riesgo de insolvencia.

Abstract- The paper shows a mathematical description of the operational and logistic regression technique based on lectures by authors who have profoundly worked its empirical application. The concept of credit risk is defined in a statistical context. A description of economic activities considered in the study. Then, construction was synthesized in 19 SPSS logistic regression model: definition of group training and selection of variables related to business solvency. Finally, results are presented in tables, with the coding of variables, statistical goodness of fit and predictive power to determine the probability that an SME company insolvent fall and create a risk to creditors. It was found that companies reported accumulated losses are eight times more likely to fall insolvent companies with positive retained earnings.

Key Words— *insolvency likely, colombian SME, logistic regression, credit risk, insolvency risk.*

I. INTRODUCCIÓN

El estudio de investigación llevado a cabo tuvo como objetivo determinar la relación estadística entre las variables que afectan la insolvencia de las empresas PYME, y determinar la probabilidad de que una de estas caiga en estado de insolvencia; construyendo para esto, un modelo estadístico de regresión logística binaria estructurado en SPSS 19 que calculará los coeficientes del modelo. Se planteó si las variables independientes que posiblemente explicaban el evento de insolvencia, estaban relacionadas. Esto dio entrada a determinar cuál era la probabilidad que alguna de las empresas PYME estudiadas cayera en estado de insolvencia de acuerdo a una relación de variables tanto categóricas como numéricas, implementando un modelo

¹ Producto derivado del proyecto de Investigación “Modelo para la Medición del Riesgo de Insolvencia Empresarial: Un Caso de Estudio en las Empresas PYME de Colombia”, apoyado por la Facultades de Ingenierías, de las Universidades Tecnológica de Pereira y Libre Seccional Pereira a través de los Grupos de Investigación Grupo en Administración Económica y Financiera e Ingeniería Financiera.

Eduardo Arturo Cruz Trejos imparte docencia en la Facultad de Ingeniería Industrial de la Universidad Tecnológica de Pereira, Vía La Julita Pereira (correos e.: ecruz@utp.edu.co).

Jaime Espinosa Peña es docente del Programa de Ingeniería Financiera de la Universidad Libre Seccional Pereira, Av. Las Américas Sede Belmonte Pereira (correo e.: jaimespino@unilibrepereira.edu.co).

Sergio Aristizábal es investigador del Programa de Ingeniería Financiera de la Universidad Libre Seccional Pereira, Av. Las Américas Sede Belmonte Pereira (correo e.: finanzasquantitativas@gmail.com).

de regresión. Para dar respuesta a la determinación de la probabilidad era necesario resolver el modelo estructurado hallando los parámetros o coeficientes de los elementos del modelo matemático; para encontrar estos parámetros se utilizó el método de máxima verosimilitud.

Revisando los antecedentes, y considerando la experiencia de los investigadores y de expertos en el tema de insolvencia empresarial, se encontró que las variables que se podrían considerar influyentes en el estado de insolvencia de una empresa eran las relacionadas con el flujo de efectivo, las utilidades y la estructura financiera de la empresa [1]. El criterio para la determinación de las variables a utilizar en el estudio de acuerdo a las categorías antes mencionadas, fue considerar a las cuentas de los estados financieros básicos (balance, general, flujo de efectivo y estado de resultados) como fuente de variables para la selección de las apropiadas para la estructuración del modelo.

La estructuración del modelo de regresión logística binaria en la determinación de la probabilidad de insolvencia, considerando variables tanto cuantitativas como categóricas, permite darles a los administradores de riesgo financiero una comprensión más intuitiva para la formulación de estrategias que permitan mitigar las posibles pérdidas que pudieran presentar las empresas en un momento futuro determinado, y facilitar el direccionamiento estratégico de la empresa. En síntesis, este modelo se convierte en un referente de indicadores que alertan sobre posibles efectos negativos reflejados implícitamente en las cuentas de los estados financieros básicos con su respectiva interrelación.

II. MARCO REFERENCIAL

A. Regresión logística

La referencia [2] expone que el objetivo de la regresión logística consiste en modelar cómo influye en la probabilidad de aparición de un suceso, habitualmente dicotómico, la presencia o no de diversos factores y el valor o nivel de los mismos. También puede ser usada para estimar la probabilidad de aparición de cada una de las posibilidades de un suceso con más de dos categorías (suceso politómico).

Si se utiliza como variable dependiente la probabilidad P de que ocurra el suceso, se construye la siguiente función:

$$Z_n = \frac{P}{1-P} \quad (1)$$

Ahora, si se tiene una variable que puede tomar cualquier valor, se plantea una ecuación de regresión tradicional:

$$Z_n = \frac{P}{1-P} = a + b[\text{independiente}] \quad (2)$$

Que se puede convertir con una pequeña manipulación algebraica en:

$$\begin{aligned} \text{Log} \left[\frac{P}{1-P} \right] &= a + bx \rightarrow \frac{P}{1-P} = e^{a+bx} \\ \frac{P}{1-P} * (1-P) &= e^{a+bx} * (1-P) \\ \frac{1}{(1+e^{a+bx})(e^{-a-bx})} &= \frac{1}{e^{-a-bx} + e^{a+bx} - 1 - bx} \\ P &= \frac{1}{1+e^{-a-bx}} \end{aligned} \quad (3)$$

1) Regresión logística múltiple

La referencia [3] explica que para el caso de una regresión logística con múltiples estados de respuesta, los coeficientes se estiman y los contrastes de hipótesis se realizan del mismo modo que en el modelo simple; aunque con el modelo múltiple (igual que en regresión lineal) se pueden hacer contrastes no solo sobre cada coeficiente sino también sobre el modelo completo o para la comparación de modelos.

La técnica de regresión logística es ampliamente utilizada, cuando se tiene por objetivo identificar el grupo al cual pertenece una unidad experimental, pero que, además, genera la probabilidad de tal clasificación y es menos estricta en el cumplimiento de supuestos para la utilización del modelo [4]. Por ejemplo, uno de los supuestos más difíciles de cumplir y que se requiere en la aplicación del Análisis Discriminante, es el supuesto de normalidad, en el caso de la regresión logística no se requiere cumplir el supuesto de normalidad multivariado del conjunto de variables regresivas; lo que se traduce en que se puede trabajar con las variables originales sin necesidad de transformarlas [5].

En cuanto a la caracterización de cada una de las variables explicadoras se busca encontrar, a través del método de Máxima Verosimilitud, los coeficientes que maximicen la función logística. Como explica [5]: "... Mediante la interpretación de los coeficientes del modelo estimado se busca hallar la características considerándolas simultáneamente a todas ellas... que son más discriminatorias... si alguna de las variables independientes es una variable discreta con k niveles, se debe incluir en el modelo como un conjunto de $k-1$ variables de diseño o "variables dummies". El cociente de las probabilidades correspondientes a los dos niveles de la variable respuesta se denomina ODDS (cociente de probabilidad) y su expresión es:

$$\begin{aligned} \frac{P(Y = 1/x)}{1-P(Y = 1/x)} &= e^{0 + 1^x 1 + \dots + p^x p} \quad (4) \\ \text{Log} \left[\frac{P(Y = 1/x)}{1-P(Y = 1/x)} \right] &= \text{Log} \left(e^{0 + 1^x 1 + \dots + p^x p} \right) \end{aligned}$$

Donde:

Log = Logaritmo de la razón de proporciones de los niveles de la variable respuesta.

Los β_j estimados representan la tasa de cambio de una función de la variable dependiente "Y" por unidad de cambio de la variable independiente "X".

El coeficiente β_j expresa el cambio resultante en la escala de medida de la variable "Y" y para un cambio unitario de la variable "X"; por ejemplo, para la variable X1, $\beta_1 = \ln(g(x_1+1)/g(x_1))$ representa el cambio en el Logit frente a un incremento de una unidad en la variable X1. La interpretación se hace en términos de la razón de OR (cociente de probabilidad condicional).

$$OR = \frac{\frac{P(Y = \frac{1}{x_j + 1})}{1 - P(Y = \frac{0}{x_j + 1})}}{\frac{P(Y = \frac{1}{x_j})}{P(Y = \frac{0}{x_j})}} = \frac{e^{0 + 1^{x_1} + \dots + P^x P}}{e^{0 + 1^{x_1} + \dots + P^x P}} \quad (5)$$

Un interrogante en este tipo de análisis, es determinar si todas las variables consideradas en la función discriminante contienen información útil y si solamente algunas de ellas son suficientes para diferenciar los grupos. Cuando las variables independientes tienen mucha relación entre sí, el modelo no puede distinguir que parte de la variable dependiente es explicada por una u otra variable, esto se conoce como multicolinealidad [6]. Por lo que es imperativo evaluar la correlación entre las variables explicadoras del modelo.

2) Función de distribución logística

La referencia [3] enuncia la explicación de la función de la distribución logística, planteando la siguiente cuestión: ¿Qué significado tienen los coeficientes del modelo β_0 y β_1 ?

Respondiendo al interrogante se dice que $\exp(\beta_0)$ representaría el valor del ODDS cuando la variable explicativa toma el valor cero, es decir, cuanto más probable es el éxito que el fracaso cuando la variable explicativa vale cero:

- $\exp(\beta_1)$ representa el OR por unidad de incremento de la variable explicativa X.
- OR = Medida de asociación
- $\exp(\beta_0)$ es el OR por el aumento de una unidad en la variable Xi manteniendo constantes las otras (controlado por ellas).

3) Las variables en el modelo logístico

Las variables cualitativas

Es incorrecto que en el modelo intervengan variables cualitativas, ya sean nominales u ordinales. La solución a este problema es crear tantas variables dicotómicas como

respuestas menos dos. Estas son las variables dummy (variables internas; indicadores de diseño).

El coeficiente de la ecuación para cada variable dummy se corresponde al Odds Ratio de esa categoría con respecto al nivel de referencia (la primera respuesta); cuantifica cómo cambia el riesgo respecto a primera respuesta. El hecho de que la constante en el modelo logístico no sea significativa, manifiesta que cuando las variables independientes toman el valor de cero, el logaritmo también toma un valor de cero.

Inconvenientes en la selección de variables

La selección de variables explicativas, como en cualquier otro modelo estadístico, requiere de un cuidadoso proceso de selección y subsecuente manipulación, con el fin de alcanzar la mayor precisión a la hora de describir o predecir un conjunto de características de un objeto de estudio. Para el caso de la regresión logística, se presentan, entre otros, los siguientes inconvenientes, los cuales pueden invalidar o sobreestimar el verdadero nivel de descripción o predicción del modelo:

- La falta de un sustento teórico que facilite la selección de variables: para la selección de las variables independientes se acude a los principios de diagnóstico y análisis de los estados financieros; y se combinan los rubros de inversión, financiación y aporte de capital con los rubros de resultado como utilidad operacional, utilidad neta y el flujo de efectivo. Todo lo anterior, sustentado en la revisión exhaustiva realizada por [15], en la que concluye parcialmente que los principales autores en la aplicación de la regresión logística para la medición del riesgo de crédito han seguido los criterios más convenientes para cada uno a la hora de realizar esta selección.

- Los problemas de multicolinealidad que se pueden presentar entre las variables independientes: este es un aspecto que puede generar modelos inconsistentes; lo anterior, debido que la redundancia generada por variables independientes muy correlacionadas, puede generar un alto nivel de predicción pero con incrementos exagerados en los errores estándar del modelo.

Para solucionar el problema de multicolinealidad, inicialmente se debe analizar la matriz de correlación de las variables independientes y eliminar la variable altamente correlacionada que tenga menor importancia para los autores en la construcción del modelo de predicción.

Así mismo, la solución a este problema se mitiga ampliando la muestra de observación para incrementar la información del modelo, tomando la información en un período de más de quince años.

- La oscilación en el tiempo: de acuerdo con [16], los modelos logísticos que solo consideran un periodo, son sesgados, ineficientes e inconsistentes. No obstante, la

mayoría de autores ha modelado estáticamente (en un solo periodo); es importante manifestar y/o reconocer las intenciones de cada una de las investigaciones de medición de riesgo de crédito para determinar si un modelo estático es más válido que un modelo dinámico.

B. Riesgo de crédito

“El riesgo de crédito se define como la pérdida potencial que se registra con motivo del incumplimiento de una contraparte en una transacción financiera (o en alguno de los términos y condiciones de la transacción). También se concibe como un deterioro en la calidad crediticia de la contraparte o en la garantía o colateral pactada originalmente”. [7].

“El riesgo crédito... surge cuando las contrapartes están indispuestas o son totalmente incapaces de cumplir sus obligaciones contractuales”. [8]

La referencia [9], comenta: “... el riesgo de crédito es la posibilidad, con mayor o menor incertidumbre, de incurrir en una pérdida cuando la contrapartida incumple total o parcialmente sus compromisos financieros según lo estipulado en el contrato que los riga” (p. 47).

Según estos autores y demás literatura revisada acerca del Riesgo de Crédito, las pérdidas y el incumplimiento son consecuencias que hacen del crédito, a través de los diferentes instrumentos financieros, un producto que requiere un detallado estudio por parte de las entidades que lo ofrecen para garantizar las ganancias futuras de los dueños de la empresa. El riesgo de crédito, en perspectiva, es definido en 5 tipos de riesgo [9].

- Riesgo de contraparte.
- Riesgo emisor.
- Riesgo país.
- Riesgo de liquidación.
- Riesgos relacionados, como el deterioro de la calidad de cartera.

Cada uno de estos grupos expresa que el estudio del riesgo crediticio es realizado de acuerdo al interés del gestor de riesgo. La perspectiva del presente estudio de investigación es determinar la probabilidad de que una entidad deudora sea incapaz de cumplir con los términos de pago y fecha pactados en un contrato de crédito con otra entidad. Esto es lo que se denomina riesgo de contraparte.

1) Cálculo del riesgo de crédito

En la actualidad hay desarrollados, entre otros, cuatro tipos de modelos para determinar el riesgo de crédito de las empresas [7]:

- Modelos tradicionales, entre los que se encuentra el de las 5 Ces.
- Modelos econométricos.
- Modelo KMV y Moody's.
- Redes neuronales artificiales.

Estos modelos tienen por objeto determinar la probabilidad de que una entidad incumpla en sus compromisos crediticios, exceptuando a los modelos tradicionales, los cuales son de carácter cualitativo. Quienes toman las decisiones deben interactuar armoniosamente tanto con los estudios cualitativos como con los cuantitativos, ya que ninguno debe ser percibido de forma mutuamente excluyente sino que, por el contrario, deben complementarse con el fin de generar resultados más fundamentados y certeros.

A continuación se presenta una descripción de los modelos econométricos utilizados en el presente estudio para la medición del riesgo crediticio.

2) El coeficiente de correlación múltiple ρ

Esta es una medida que indica el grado de relación existente entre dos o más variables. Los valores generados en un análisis de correlación van desde -1 (los valores negativos representan una relación inversa) hasta 1 (los valores positivos representan una relación directa).

A una correlación de -1 se le llama correlación inversa perfecta; un resultado de 1 se denomina correlación directa perfecta. Si el resultado llegase a ser muy próximo a cero, se dice que no existe relación entre las variables estudiadas.

Para el cálculo del coeficiente de correlación es necesario determinar cuál es el objetivo del estudio. Esto debido a que la regresión y la correlación son dos técnicas separadas: la regresión tiene como fin la predicción y la correlación la asociación. “Para situaciones en las que el interés principal es el análisis de regresión, el coeficiente de correlación de la muestra (r) se obtiene a partir del coeficiente de determinación (r^2) [10]:

$$r^2 = \frac{\text{sumadecuadrosregresion}}{\text{sumadecuadros total}} = \frac{SCR}{SCT} \quad (6)$$

$$r = \text{Coeficientedecorrelacion} = \sqrt{r^2} \quad (7)$$

3) Evaluación del modelo de regresión lineal múltiple

Cuando el modelo de regresión múltiple está construido, inicialmente debe ser evaluada la bondad de ajuste. Esta es medida con el error estándar y el coeficiente de determinación múltiple. Finalmente es llevada a cabo la prueba de significancia para el modelo, la cual consta del coeficiente de regresión estandarizado (prueba t) y el cociente F (prueba de Fisher).

La bondad de ajuste

Medir la bondad de ajuste significa determinar qué tan aproximados fueron los resultados obtenidos en la

regresión, a los datos originales. Si una regresión no genera un ajuste considerable, entonces el modelo será rechazado inmediatamente sin necesidad de hacer más pruebas [11]. El ajuste es medido con los siguientes indicadores.

- *Error estándar de estimación (Se)*: Este mide los grados de dispersión alrededor del plano de la regresión. Entre más pequeño sea “Se”, más ajustado y preciso será el modelo de regresión [12].

$$Se = \sqrt{\frac{\sum (Y_i - \bar{Y}_i)^2}{n - k - 1}} \quad (8)$$

Donde $n-k-1$ es el número de grados de libertad y k es el número de variables a la derecha; el numerador bajo el radical de la fórmula es la suma de los errores elevada al cuadrado y se minimizará de acuerdo con el concepto de mínimos cuadrados ordinarios.

- *Coefficiente de determinación múltiple (R^2)*: expresa la fuerza de relación entre la variable dependiente y las variables explicadoras. Entre más alto sea el valor de R^2 mayor poder explicativo tendrá el modelo. El valor de esta medida se encuentra siempre entre 0 y 1, es decir: $0 \leq R^2 \leq 1$.

- *El coeficiente de determinación ajustado (R^2)*: el coeficiente de determinación tiene la desventaja de que es vulnerable a manipular sus resultados incluyendo nuevas variables al modelo así no tengan significancia, esto debido a algún tipo de correlación casual que se pueda presentar, pero que en el fondo no tiene nada que ver con el objeto del estudio. El coeficiente de determinación ajustado se ajusta a la medida del poder explicativo para el número de grados de libertad [13].

$$R^2 = 1 - (1 - R^2) \frac{n - 1}{n - k - 1} \quad (9)$$

C. Empresas PYME en Colombia

Para la descripción de los respectivos subsectores objeto de trabajo, la investigación se basó en la referencia [14], que describe cada uno de los mismos así:

TABLA I
ACTIVIDADES ECONÓMICAS ESTUDIADAS

| Código CIU | Clasificación |
|------------|---|
| A0121 | Cría especializada de ganado vacuno |
| F4521 | Construcción de edificaciones para uso residencial |
| F4530 | Construcción de obras de ingeniería civil |
| G5030 | Comercio de partes, piezas (autopartes) y accesorios (lujos) para vehículos automotores |
| G5141 | construcción, vidrio, artículos de ferretería y equipo y materiales de fontanería y calefacción |
| G5190 | Comercio al por mayor de productos diversos npe |
| J6599 | Intermediación financiera, excepto los seguros y los fondos de pensiones y cesantías |
| K7010 | Actividades inmobiliarias realizadas con bienes propios o arrendados |

1) Cría especializada de ganado vacuno

Realiza las siguientes actividades: explotación de recursos naturales vegetales y animales. Comprende las actividades de cultivo, cría de animales, explotación (aprovechamiento) de madera, cosecha de plantas, cuidado de animales y captura de los mismos en su hábitat natural. En cuanto a la cría especializada:

- La cría y/o levante de ganado vacuno.
- La producción de leche cruda de vaca.
- La producción de semen bovino.
- Servicios de engorde del ganado en corrales.
- La producción de mantequilla, queso y otros productos lácteos, como actividad secundaria no modifica la clasificación de la unidad.

2) Construcción de edificaciones para uso residencial y construcción de obras de Ingeniería Civil

Actividades que realizan: preparación de terreno, construcción corriente y construcción especial; construcción de obras de ingeniería civil; el acondicionamiento y la terminación de edificios y obras; acondicionamiento de edificaciones y de obras civiles y terminación y acabado de edificaciones y de obras civiles. También se incluyen las obras nuevas, reparaciones, ampliaciones y reformas, la erección de edificios y estructuras prefabricadas in situ y también la construcción de obras de índole temporal.

3) Comercio de partes, piezas (autopartes) y accesorios (lujos) para vehículos a nivel nacional; comercio al por mayor de materiales de construcción ferretería y vidrio; y venta al por mayor de productos diversos.

Actividades que realiza: venta al por mayor y al por menor (venta sin transformación) de cualquier tipo de productos y la prestación de servicios relacionados con la venta de mercancía; reparación de vehículos automotores y la instalación y reparación de efectos personales y enseres domésticos.

4) Intermediación financiera, excepto los seguros y los fondos de pensiones y cesantías. Distribuir fondos por medios distintos del otorgamiento de préstamos

Actividades que realiza:

- Inversiones en valores mobiliarios, por ejemplo, acciones, obligaciones, títulos, lotes de valores mobiliarios de sociedades o fondos comunes de inversión, etc.
- Transacciones por cuenta propia de corredores de bolsa.
- Inversiones en bienes inmuebles efectuadas primordialmente por cuenta de otros intermediarios financieros (por ejemplo, las sociedades de inversión inmobiliaria).
- Suscripción de créditos recíprocos, opciones y otras operaciones financieras de cobertura.

5) *Actividades inmobiliarias realizadas con bienes propios o arrendados*

Actividades que realiza:

- La compra, venta, alquiler y explotación de bienes inmuebles propios o arrendados tales como edificios de apartamentos, viviendas y edificios no residenciales, incluso salas de exposiciones, terrenos; la urbanización y el fraccionamiento de terrenos en lotes, etc.
- El acondicionamiento y la venta de terrenos, la explotación de apartamentos amoblados y de zonas residenciales para viviendas móviles.

III. METODOLOGÍA

A. Tipo de investigación

El trabajo de investigación desarrollado en este proyecto alcanzó dos niveles: un nivel descriptivo y un nivel exploratorio. El nivel descriptivo en la investigación se debió a la caracterización detallada que se hizo de las cuentas de los estados financieros de las empresas PYME en Colombia; de las actividades económicas más representativas, en unidades de producción del sector real en Colombia; y del ajuste teórico de las cuentas de los estados financieros, elegidas para medir el riesgo de insolvencia.

El nivel exploratorio se dio en el diseño de un modelo de regresión logística aplicado a la ingeniería financiera, se plantea la hipótesis nula H_0 : una empresa con su mayoría de variables independientes con estado 1, está en camino a un estado de crisis, en el sentido de trabajar con variables categóricas como posibles explicadoras del suceso de que una empresa PYME de Colombia tenga más probabilidad, o no, de caer en estado de insolvencia, de acuerdo a su ubicación en los diferentes estados de las variables explicadoras. La hipótesis alternativa H_1 : si la mayoría de sus variables tienen una valoración de cero, es en ese momento solvente.

B. Definición preliminar de la muestra

Como población se eligieron las pequeñas y medianas empresas que presentaron sus estados financieros frente a la Superintendencia de Sociedades para el año 2011. Para tal año fueron 21.518 empresas las que se reportaron. Este tamaño de empresas según sus activos representó, aproximadamente, el 77% del total de empresas.

De todas las PYMES consideradas se encontraron 160 en estado de reestructuración, 105 en estado de reorganización, 27 en concordato y 9 empresas en liquidación. Para efectos del estudio, estos estados representan unas condiciones relacionadas en cuanto a la caracterización jurídica de estar en uno u otro estado, pero matemáticamente presentan una similitud. Por ejemplo, las empresas que están en reestructuración son empresas que se acogieron a la Ley 550 de 1999 para la reactivación empresarial, ley que fue

modificada por la Ley 1116 de 2006, que dio una nueva denominación a las empresas que cayeran en estado de insolvencia el cual es “empresas en reorganización o liquidación”. Con todo lo anterior se decidió considerar como parte de la muestra de estudio las empresas denominadas en camino a un estado de crisis como son: en reestructuración, reorganización, concordato y liquidación; en total se tomaron 301 empresas con estas características.

La otra parte de la muestra estuvo compuesta por empresas que estuvieran “aparentemente sanas”, es decir, por empresas que no tuvieran la denominación de estar en estado de reorganización o liquidación. Para tratar de tener homogeneidad en la comparación de empresas en reorganización y las “aparentemente sanas”, se eligió un mismo número de empresas: 301 unidades “aparentemente sanas”. También se tuvo cuidado de que la elección de estas empresas fuera proporcional en tamaño, igual que las empresas del primer grupo. Es decir, las 301 empresas en estado de insolvencia (reorganización, reestructuración, etc.), fueron clasificadas en unos intervalos según el tamaño de los activos como se muestra en la Tabla II.

TABLA II
PROPORCIÓN DE EMPRESAS EN REORGANIZACIÓN SEGÚN TAMAÑO
(CIFRAS EN MILES DE PESOS)

| Límite inferior | Límite superior | Frecuencia absoluta |
|-----------------|-----------------|---------------------|
| 275.057 | 2.243.063 | 109 |
| 2.243.063 | 4.211.069 | 68 |
| 4.211.069 | 6.179.075 | 33 |
| 6.179.075 | 8.147.081 | 30 |
| 8.147.081 | 10.115.087 | 19 |
| 10.115.087 | 12.083.093 | 18 |
| 12.083.093 | 14.051.099 | 15 |
| 14.051.099 | 16.019.105 | 9 |
| Total | | 301 |

De acuerdo a la Tabla II, se procuró elegir casi el mismo tamaño de empresas según el número de intervalos, por ejemplo, elegir 109 empresas “aparentemente sanas” con activos entre, aproximadamente, 275 millones y 2.200 millones de pesos al igual que empresas en reorganización. En total se tuvo como muestra 602 empresas tanto en reorganización como aparentemente solventes.

D. Determinación del grupo de entrenamiento

En cualquier modelo de regresión se busca encontrar el valor de los coeficientes de las variables para realizar la descripción de un conjunto de datos; así como predecir el comportamiento de una nueva observación aplicando la función de regresión hallada. Para lograr evaluar el poder estimativo y predictivo de un modelo de regresión, es necesario aplicarlo en observaciones que no hayan sido

consideradas para la determinación de los coeficientes del modelo; esto con la pretensión de no evaluar relajadamente el poder del modelo de regresión, ya que si se evalúa con observaciones que fueron utilizadas para hallar la ecuación de regresión, las características de esta observación permitirían una perfecta estimación y predicción para la observación.

El grupo de entrenamiento se compuso por casi el 50% de empresas que estuvieran dentro de cada uno de los intervalos definidos en la Tabla II. Tanto para empresas en reorganización como empresas aparentemente sanas. En la Tabla III se muestra la composición del grupo de entrenamiento.

Para el grupo de entrenamiento se definieron 300 empresas; la mitad compuesta por empresas en reorganización y la otra mitad por empresas aparentemente sanas. Por consiguiente las empresas restantes fueron utilizadas en aras de evaluar el poder estimativo y predictivo del modelo de regresión logística binaria.

Con el objetivo de garantizar la aleatoriedad en la selección de cada una de las unidades del grupo de entrenamiento, le fue asignado a cada uno de los elementos del marco muestral un número identificador, con lo que se aplicó la función de "generación de números aleatorios" de la hoja de cálculo Excel de Microsoft®, y así elegir las unidades identificadas con los números aleatorios generados por la función del Excel.

TABLA III

GRUPO DE ENTRENAMIENTO COMPUESTO POR DE EMPRESAS EN REORGANIZACIÓN Y POR EMPRESAS APARENTEMENTE SANAS

| Límite inferior | Límite inferior | Re- organización | Aparente sanas |
|-----------------|-----------------|------------------|----------------|
| 275.057 | 2.243.063 | 54 | 54 |
| 2.243.063 | 4.211.069 | 34 | 34 |
| 4.211.069 | 6.179.075 | 16 | 16 |
| 6.179.075 | 8.147.081 | 15 | 15 |
| 8.147.081 | 10.115.087 | 10 | 10 |
| 10.115.087 | 12.083.093 | 9 | 9 |
| 12.083.093 | 14.051.099 | 7 | 7 |
| 14.051.099 | 16.019.105 | 5 | 5 |
| Total x grupo | | 150 | 150 |
| TOTAL | | 300 | |

E. Selección de variables

Para el modelo de regresión logística aplicado para evaluar el grado de insolvencia (variable dependiente) de las empresas PYME en Colombia, fueron definidas ocho variables independientes: dos cuantitativas y seis cualitativas.

1) La variable dependiente

La variable dependiente de respuesta dicotómica se definió como "Estado de insolvencia", la cual tomó como valores de respuesta "1" y "0"; significando "1" que la empresa era insolvente y "0" que la empresa era solvente.

Y (variable respuesta dicotómica) = "Estado de insolvencia"
 $Y=1, \rightarrow$ "Insolvente (en reorganización)"
 $Y=0, \rightarrow$ "Solvente (aparentemente sana)"

2) Las variables independientes

Cada una de las variables independientes cualitativas elegidas para el modelo tiene estados dicotómicos representados por 1 (en caso de presencia de una característica) o 0 (en caso de ausencia de la característica). Por ejemplo, la variable independiente representó la razón corriente de una empresa, tomando el valor de "1" cuando la razón corriente era inferior a 1 y tomando el valor de "0" cuando la razón corriente estaba por encima de 1. La lógica de esta variable se puede entender en el sentido de que una empresa con una razón corriente por debajo de ciertos niveles indicaba que la empresa, tal vez, presentaría problemas para afrontar sus compromisos del corto plazo; no significando que este indicador mostrara insolvencia. El criterio de que la razón fuera mayor o igual a 1 parte del hecho de que la teoría del indicador y, empíricamente hablando, tener una razón corriente igual a 1 significa que la empresa cuenta, exactamente, con los recursos para afrontar los compromisos a corto plazo; por este motivo se tomó esta referencia.

X_1 = "Razón Corriente"
 $X_1=1, \rightarrow$ "Razón corriente inferior a 1"
 $X_1=0, \rightarrow$ "Razón corriente superior o igual a 1"

X_2 = "Estructura Financiera"
 $X_2=1, \rightarrow$ "Pasivos \geq Patrimonio"
 $X_2=0, \rightarrow$ "Pasivos < Patrimonio"

X_3 = "Utilidad Neta"
 $X_3=1, \rightarrow$ "Utilidad neta Negativa"
 $X_3=0, \rightarrow$ "Utilidad neta 0 o Positiva"

X_4 = "Utilidad Operacional"
 $X_4=1, \rightarrow$ "Utilidad operacional Negativa"
 $X_4=0, \rightarrow$ "Utilidad operacional 0 o Positiva"

X_5 = "Resultado de Ejercicios Anteriores"
 $X_5=1, \rightarrow$ "Negativo"
 $X_5=0, \rightarrow$ "Positivo"

X_6 = "Cambio Flujo de Efectivo Neto"
 $X_6=1, \rightarrow$ "Efectivo neto disminuyó"
 $X_6=0, \rightarrow$ "Efectivo neto aumentó"

3) Las variables cuantitativas

X_7 = "Grado de Apalancamiento"
 $X_7=0 < x < 1$, entre más se acerca a 1 hay mayor apalancamiento
 X_8 = "Concentración de Endeudamiento en Corto Plazo"
 $X_8=0 < x < 1$, entre más se acerca a 1 hay mayor necesidad de efectivo

La Tabla IV lista las variables con sus respectivas codificaciones.

TABLA IV
CODIFICACIÓN DE VARIABLES EN EL MODELO DE REGRESIÓN

| Nombre | Identificador |
|---|---------------|
| Identificador Muestral | IDE_MUE |
| NIT | NIT |
| Razón Social | NOM_EMP |
| Estado de Insolvencia | EST_INS |
| Razón Corriente | RAZ_COR |
| Estructura Financiera | EST_FIN |
| Utilidad Neta | UTI_NET |
| Utilidad Operacional | UTI_OPE |
| Resultados de Ejercicios Anteriores | REE_ANT |
| Cambio en Flujo de Efectivo Neto | CFL_EFE |
| Grado de Apalancamiento | GRA_APA |
| Concentración Endeudamiento Corto Plazo | CON_ENC |

IV. RESULTADOS DE LA INVESTIGACIÓN

A. Modelo de riesgo de insolvencia a través de la regresión logística binaria

Este primer modelo tiene el objetivo de ver la relación existente entre las variables independientes. Los cuadros presentados a continuación son los generados para la estructura completa del modelo de regresión logística.

1) Modelo de regresión logística 01

La Tabla V indica que fueron seleccionados los 300 elementos del grupo de entrenamiento. Todos los casos fueron válidos y no se excluyó ninguna de las observaciones.

TABLA V
RESUMEN DEL PROCESAMIENTO DE LOS CASOS
REGRESIÓN LOGÍSTICA MODELO 01

| Casos no ponderados | | N | Porcentaje |
|------------------------|--------------------------|-----|------------|
| Casos seleccionados | Incluidos en el análisis | 300 | 100,0 |
| | Casos perdidos | 0 | ,0 |
| | Total | 300 | 100,0 |
| Casos no seleccionados | | 0 | ,0 |
| Total | | 300 | 100,0 |

El modelo fue estructurado con 6 variables independientes de carácter cualitativo con dos estados. Según la Fig. 1, los resultados encontrados fueron:

| CODIFICACIONES DE VARIABLES CATEGÓRICAS | | | |
|---|---|------------|----------------------------|
| | | Frecuencia | Codificación de parámetros |
| | | | 1 |
| Cambio en Flujo de Efectivo Neto | Efectivo neto disminuyó | 160 | 1 |
| | Efectivo neto aumentó | 140 | 0 |
| Estructura Financiera | Pasivos mayores o iguales que el patrimonio | 106 | 1 |
| | Pasivos menores que el patrimonio | 194 | 0 |
| Utilidad Neta | Utilidad neta negativa | 198 | 1 |
| | Utilidad neta positiva | 102 | 0 |
| Utilidad Operacional | Utilidad operacional Negativa | 193 | 1 |
| | Utilidad operacional positiva | 107 | 0 |
| Resultados de Ejercicios Anteriores | Negativo | 150 | 1 |
| | Positivo | 150 | 0 |
| Razón Corriente | Razón corriente inferior a 1 | 222 | 1 |
| | Razón corriente superior o igual que 1 | 78 | 0 |

Fig. 1. Codificaciones de variables categóricas regresión logística modelo 01

- Para la variable CFL_EFE, 160 empresas disminuyeron su efectivo neto respecto al año anterior y 140 empresas aumentaron su efectivo neto respecto al año anterior.
- La variable EST_FIN muestra que 106 empresas tenían pasivos mayores o iguales que su patrimonio; 194 empresas tenían patrimonios menores o iguales a sus pasivos.
- UTI_NET: 198 empresas presentaron utilidad neta negativa; 102 empresas, utilidad neta positiva.
- UTI_OPE muestra que 193 empresas presentaron utilidad operacional negativa; 107 empresas, utilidad operacional positiva.
- REE_ANT: 150 empresas tenían resultados de ejercicios anteriores negativos; 150 empresas, con resultados de ejercicios anteriores positivos.
- RAZ_COR muestra que 222 empresas presentaron una razón corriente por debajo de 1; 78 empresas, una razón corriente igual o por encima de 1.

Se resalta que la Fig. 1 muestra en la tercera columna la codificación de cada uno de los estados de las variables independientes.

La categoría de referencia para cada una de las variables cualitativas es la que esté codificada con el valor de "0"; con lo cual se obtiene la probabilidad de ocurrencia de la variable dependiente teniendo el estado de la variable independiente con codificación "1" respecto a la misma variable con codificación "0".

TABLA VI

CLASIFICACIÓN SIN VARIABLES INDEPENDIENTES REGRESIÓN LOGÍSTICA MODELO 01

| Observado | | Pronosticado | | Porcentaje correcto |
|-----------------------|------------|-----------------------|------------|---------------------|
| | | Estado de Insolvencia | | |
| | | Solvente | Insolvente | |
| Estado de Insolvencia | Solvente | 0 | 150 | ,0 |
| | Insolvente | 0 | 150 | 100,0 |
| Porcentaje global | | | | 50,0 |

La Tabla VI muestra un modelo preliminar, donde solo se ha tenido en cuenta una constante para el pronóstico de las observaciones. Solo con la constante, el modelo tuvo un poder de pronóstico del 50%. De las 150 empresas solventes, se pronosticó al total de estas como insolventes, lo que se tradujo en un 0% de porcentaje correcto de pronóstico para este grupo de observaciones. Para el grupo de las 150 empresas insolventes, se logró un pronóstico del 100%, es decir, las 150 empresas insolventes fueron clasificadas como insolventes.

Debido a que se tuvo un promedio de pronóstico correcto de 50%. En este punto el resultado de pronóstico no se tiene en cuenta para un nivel de análisis ya que esta es la agrupación preliminar que se definió por parte del investigador, es decir, en este punto no se ha hecho ningún tratamiento con el modelo de regresión logística, más bien se ilustra al lector que hay dos grupos de empresas con igual número de unidades en cada uno de los grupos: 150 solventes y 150 insolventes.

TABLA VII

CONSTANTE EN LA ECUACIÓN REGRESIÓN LOGÍSTICA MODELO 01

De acuerdo a la Tabla VII, el parámetro estimado en

| | B | E.T. | Wald | gl | Sig. | Exp(B) |
|-----------|------|------|------|----|-------|--------|
| Constante | ,000 | ,115 | ,000 | 1 | 1,000 | 1,000 |

este modelo preliminar que es la constante β_0 , dio como resultado 0, lo que generó un $e^{\beta_0} = 1$. Esto significa que de acuerdo a las observaciones, por cada empresa insolvente hay una empresa solvente. Esto significa que según el grupo de entrenamiento elegido, por cada empresa solvente hay un 100% de probabilidad que haya una empresa insolvente. Lo que se traduce que es igual de probable elegir del grupo de entrenamiento una empresa que sea solvente o que sea insolvente.

TABLA VIII

VARIABLES QUE NO ESTÁN EN ECUACIÓN REGRESIÓN LOGÍSTICA MODELO 01

| Variables | Puntuación | gl | Sig. |
|---------------------|------------|----|------|
| RAZ_COR(1) | 17,741 | 1 | ,000 |
| EST_FIN(1) | 7,061 | 1 | ,008 |
| UTI_NET(1) | 31,432 | 1 | ,000 |
| UTI_OPE(1) | 13,961 | 1 | ,000 |
| REE_ANT(1) | 85,333 | 1 | ,000 |
| CFL_EFE(1) | ,054 | 1 | ,817 |
| GRA_APA | 17,232 | 1 | ,000 |
| CON_ENC | 27,891 | 1 | ,000 |
| Estadísticos | 105,741 | 8 | ,000 |

En la Tabla VIII, la segunda columna contiene el valor χ^2 resultado de la prueba de independencia entre la variable dependiente “Estado de insolvencia” (EST_INS) y cada una de las variables explicativas; esto con el fin de determinar si cada una de las variables independientes tenía alguna relación con la variable dependiente y así incluirlas en el modelo de regresión. Todas las variables, según el test de independencia de Chi-cuadrado, tuvieron relación con la variable dependiente, exceptuando la variable CFL_EFE, la cual presentó un p-valor por encima del nivel de significancia del 5% (8.17%), con lo que se aceptó la hipótesis nula de independencia entre la variable “Estado de insolvencia” y la variable “Cambio en el Flujo de Efectivo Neto”.

En este punto se pudo determinar que la variable CFL_EFE no debería entrar en el modelo de regresión logística, sin embargo como se corrió este primer modelo con la inclusión de variables a través del método “introducir”, todas las variables ingresaron para la evaluación del modelo, según las siguientes tablas de resultados.

TABLA IX

PRUEBAS ÓMNIBUS SOBRE LOS COEFICIENTES DEL MODELO 01

| Desc. | Chi | gl | Sig. |
|--------|---------|----|------|
| Paso | 129,846 | 8 | ,000 |
| Bloque | 129,846 | 8 | ,000 |
| Modelo | 129,846 | 8 | ,000 |

El resultado de la prueba Ómnibus (ver Tabla IX), muestra que la inclusión de las nuevas variables, al modelo, es significativa en comparación con el modelo que tiene en cuenta solo la constante. Esta prueba contrasta la hipótesis nula de que los coeficientes de todos los términos del modelo son iguales a cero, lo que indica, en este caso, que todo el grupo es significativo en el modelo.

TABLA X

RESUMEN DEL MODELO 01

| Paso | -2 log de la verosimilitud | R cuadrado de Cox y Snell | R cuadrado de Nagelkerke |
|------|----------------------------|---------------------------|--------------------------|
| 1 | 286,042 ^a | ,351 | ,468 |

La Tabla X evalúa de forma global la validez del modelo. El estadístico “-2log de la verosimilitud” indica que la desviación del modelo es de 286,042, lo que no deja claro si el modelo se ajusta bien a los datos. R² de Cox y Snell para este caso indica que las variables independientes explican en un 35,1% la varianza de la variable dependiente; esto es una proporción de explicación de varianza pobre. El R² de Nagelkerke, que es una corrección del estadístico anterior, indica un aumento en la explicación de la varianza, aumentándose hasta 46,8%; igualmente este resultado es pobre en explicación de la varianza de la variable dependiente. Pero se debe tener en cuenta que este estadístico es radical ya que incluso en un modelo de regresión perfecto, no se alcanza el valor de 1 (100%).

TABLA XI
TABLA DE CLASIFICACIÓN REGRESIÓN LOGÍSTICA MODELO 01

| Observado | | Pronosticado | | |
|-----------------------|------------|-----------------------|------------|---------------------|
| | | Estado de Insolvencia | | Porcentaje correcto |
| | | Solvente | Insolvente | |
| Estado de Insolvencia | Solvente | 117 | 33 | 78,0 |
| | Insolvente | 37 | 113 | 75,3 |
| Porcentaje global | | | | 76,7 |

La Tabla XI muestra que de 150 empresas solventes, el modelo clasificó a 117 empresas efectivamente como solventes y a las 33 restantes como insolventes. Igualmente, de 150 empresas insolventes, clasificó a 37 empresas como solventes y a 113 empresas como insolventes. Todo lo anterior generó un poder de pronóstico del 76,7%, lo que significa que este primer modelo puede clasificar de forma correcta a 77 empresas de un total de 100 empresas.

TABLA XII
VARIABLES EN LA ECUACIÓN REGRESIÓN LOGÍSTICA DEL MODELO 01

| Variable | B | E.T. | Wald | gl | Sig. | Exp(B) |
|------------|--------|------|--------|----|------|--------|
| RAZ_COR(1) | -,665 | ,361 | 3,392 | 1 | ,066 | ,514 |
| EST_FIN(1) | 1,219 | ,508 | 5,764 | 1 | ,016 | 3,383 |
| UTI_NET(1) | -,636 | ,404 | 2,470 | 1 | ,116 | ,530 |
| UTI_OPE(1) | ,058 | ,392 | ,022 | 1 | ,883 | 1,060 |
| REE_ANT(1) | -1,705 | ,310 | 30,257 | 1 | ,000 | ,182 |
| CFL_EFE(1) | ,222 | ,299 | ,552 | 1 | ,457 | 1,249 |
| GRA_APA | 2,722 | ,889 | 9,381 | 1 | ,002 | 15,212 |
| CON_ENC | -1,486 | ,469 | 10,028 | 1 | ,002 | ,226 |
| Constante | ,516 | ,906 | ,324 | 1 | ,569 | 1,676 |

Según muestra la Tabla XII, las variables UTI_NET, UTI_OPE, CFL_EFE y la constante no son significativas en este primer modelo de regresión logística.

2) Modelo de regresión logística 02

Se corrió el modelo automáticamente hacia adelante (*forward method*). Además de los parámetros incluidos en el modelo 01 para correr el SPSS, se indicó que se generara la bondad de ajuste de Hosmer-Lemeshow, el historial de iteraciones y las correlaciones de las estimaciones, como se aprecia en la Fig. 2.

| HISTORIAL DE ITERACIONES | | | | | | |
|--------------------------|----------------------------|----------|--------------|---------|------------|--------|
| Iteración | -2 log de la verosimilitud | Constant | Coeficientes | | | |
| | | | REE_ANT(1) | CON_ENC | RAZ_COR(1) | |
| Paso 1 | 1 | 326,792 | 1,067 | -2,133 | | |
| | 2 | 325,964 | 1,186 | -2,372 | | |
| | 3 | 325,964 | 1,19 | -2,379 | | |
| | 4 | 325,964 | 1,19 | -2,379 | | |
| Paso 2 | 1 | 314,24 | 1,72 | -1,955 | -1,109 | |
| | 2 | 311,742 | 2,144 | -2,241 | -1,515 | |
| | 3 | 311,721 | 2,19 | -2,267 | -1,561 | |
| | 4 | 311,721 | 2,19 | -2,268 | -1,561 | |
| Paso 3 | 1 | 307,188 | 2,258 | -1,749 | -1,309 | -0,684 |
| | 2 | 303,823 | 2,887 | -2,018 | -1,776 | -0,924 |
| | 3 | 303,786 | 2,963 | -2,048 | -1,834 | -0,953 |
| | 4 | 303,786 | 2,964 | -2,049 | -1,835 | -0,953 |
| | 5 | 303,786 | 2,964 | -2,049 | -1,835 | -0,953 |

Fig. 2. Codificaciones de variables categóricas regresión logística modelo 01

| PRUEBAS ÓMNIBUS SOBRE LOS COEFICIENTES DEL MODELO | | | | |
|---|--------------|----|-------|--|
| | Chi cuadrado | gl | Sig. | |
| Paso | 89,925 | 1 | 0 | |
| Paso 1 Bloque | 89,925 | 1 | 0 | |
| Modelo | 89,925 | 1 | 0 | |
| Paso | 14,242 | 1 | 0 | |
| Paso 2 Bloque | 104,167 | 2 | 0 | |
| Modelo | 104,167 | 2 | 0 | |
| Paso | 7,935 | 1 | 0,005 | |
| Paso 3 Bloque | 112,102 | 3 | 0 | |
| Modelo | 112,102 | 3 | 0 | |

Fig. 3. Prueba ómnibus sobre los coeficientes del modelo 02

El resultado de la prueba Ómnibus (ver Fig. 3), muestra que la inclusión de las nuevas variables al modelo, en los tres pasos, es significativa. Esta prueba contrasta la hipótesis nula de que los coeficientes de todos los términos del modelo son iguales a cero, lo que indica, en este caso, que todo el grupo es significativo en el modelo.

| BONDAD DE AJUSTE CON TRES ESTADÍSTICOS | | | |
|--|----------------------------|---------------------------|--------------------------|
| Paso | -2 log de la verosimilitud | R cuadrado de Cox y Snell | R cuadrado de Nagelkerke |
| 1 | 325,964 ^a | 0,259 | 0,345 |
| 2 | 311,721 ^a | 0,293 | 0,391 |
| 3 | 303,786 ^b | 0,312 | 0,416 |

DISMINUYÓ ↓
AUMENTÓ ↑
AUMENTÓ ↑

Fig. 4. Prueba ómnibus sobre los coeficientes del modelo 02

En la Fig. 4, el estadístico “menos dos veces el logaritmo neperiano de la verosimilitud” muestra que al incluir las variables en paso el dos y tres, este va disminuyendo; lo que indica que el modelo efectivamente se fue ajustando en mayores proporciones a medida que le eran incluidas más variables. El “R cuadrado de Cox y Snell”, así como el “R cuadrado de Nagelkerke” muestran que el poder explicativo de las variables en conjunto, aumenta a medida que se incluyen las demás variables significativas en el modelo (ver cuadro 16), de 25,9% a 31,2% y de 34,5% a 41,6%, respectivamente.

| | | Estado de Insolvencia = Solvente | | Estado de Insolvencia = Insolvente | | Total |
|--------|---|----------------------------------|----------|------------------------------------|----------|-------|
| | | Observado | Esperado | Observado | Esperado | |
| Paso 1 | 1 | 115 | 115 | 35 | 35 | 150 |
| | 2 | 35 | 35 | 115 | 115 | 150 |
| | 1 | 51 | 52,751 | 12 | 10,249 | 63 |
| | 2 | 21 | 24,317 | 9 | 5,683 | 30 |
| Paso 2 | 3 | 24 | 21,886 | 6 | 8,114 | 30 |
| | 4 | 19 | 16,047 | 8 | 10,953 | 27 |
| | 5 | 18 | 13,56 | 21 | 25,44 | 39 |
| | 6 | 9 | 8,832 | 21 | 21,168 | 30 |
| | 7 | 4 | 6,073 | 26 | 23,927 | 30 |
| Paso 3 | 8 | 3 | 4,252 | 27 | 25,748 | 30 |
| | 9 | 1 | 2,283 | 20 | 18,717 | 21 |
| | 1 | 45 | 45,943 | 8 | 7,057 | 53 |
| | 2 | 22 | 24,976 | 8 | 5,024 | 30 |
| | 3 | 22 | 22,22 | 8 | 7,78 | 30 |
| | 4 | 21 | 18,748 | 9 | 11,252 | 30 |
| | 5 | 17 | 13,7 | 13 | 16,3 | 30 |
| | 6 | 13 | 10,011 | 24 | 26,989 | 37 |
| | 7 | 6 | 6,456 | 24 | 23,544 | 30 |
| 8 | 3 | 4,77 | 27 | 25,23 | 30 | |
| 9 | 1 | 3,177 | 29 | 26,823 | 30 | |

Fig. 5. Contingencia para prueba de Hosmer y Lemeshow modelo 02

En la Fig. 5 se compara la división en deciles de riesgo de las frecuencias observadas en cada uno de ellos, con la frecuencia esperada en deciles del estado de insolvencia. Se observa que las frecuencias en deciles del paso 3 de las empresas solventes e insolventes tienen una proporción que es aproximada por las frecuencias de los deciles esperados. Lo anterior permite aseverar, al menos visualmente, que el modelo de regresión logística estructurado está bien ajustado a las observaciones reales del grupo de entrenamiento.

| ESTADÍSTICOS DEL MODELO (FACTORES DE RIESGO Y/O DE PROTECCIÓN) | | | | | | | | | |
|--|------------|--------|------|--------|----|------|--------|----------------------|----------|
| | | B | E.T. | Wald | gl | Sig. | Exp(B) | I.C. 95% para EXP(B) | |
| | | | | | | | | Inferior | Superior |
| Paso 1a | REE_ANT(1) | 2.379 | .273 | 75.944 | 1 | .000 | 10.796 | 6.322 | 18.435 |
| | Constante | -1.190 | .193 | 37.972 | 1 | .000 | .304 | | |
| Paso 2b | REE_ANT(1) | 2.268 | .280 | 65.626 | 1 | .000 | 9.656 | 5.579 | 16.714 |
| | CON_ENC | -1.561 | .423 | 13.640 | 1 | .000 | .210 | .092 | .481 |
| | Constante | -.077 | .348 | .049 | 1 | .824 | .926 | | |
| Paso 3c | RAZ_COR(1) | .953 | .343 | 7.742 | 1 | .005 | 2.594 | 1.325 | 5.078 |
| | REE_ANT(1) | 2.049 | .289 | 50.110 | 1 | .000 | 7.758 | 4.399 | 13.680 |
| | CON_ENC | -1.835 | .439 | 17.479 | 1 | .000 | .160 | .068 | .377 |
| | Constante | -.038 | .350 | .012 | 1 | .913 | .963 | | |

Fig. 6. Estadísticos del modelo 02, factores de riesgo y/o protección

La Fig. 6 muestra que la variable REE_ANT (resultados de ejercicios anteriores), para el paso 1 del modelo 02, es un factor de riesgo ya que está por encima de uno. En el paso 2, incluyendo la variable CON_ENC (Concentración del endeudamiento en el corto plazo), la segunda variable más significativa, la variable REE_ANT sigue como factor de riesgo y CON_ENC entra como factor de protección. En el último paso del modelo por el “método hacia adelante”, quedan, finalmente, tres variables como explicadoras del comportamiento de la solvencia en una muestra representativa de las empresas PYME de Colombia. La variable REE_ANT queda como un factor de riesgo; CON_ENC queda como factor de protección y la variable RAZ_COR (Razón corriente) queda como factor de riesgo.

| TABLA DE PRONÓSTICO | | | | | |
|---------------------|----------------------------------|--------------|------------|---------------------|--|
| Observado | Estado de Insolvencia | Pronosticado | | Porcentaje correcto | |
| | | Solvente | Insolvente | | |
| Paso 1 | Estado de Insolvencia Solvente | 115 | 35 | 76,7 | |
| | Estado de Insolvencia Insolvente | 35 | 115 | 76,7 | |
| | Porcentaje global | | | 76,7 | |
| Paso 2 | Estado de Insolvencia Solvente | 115 | 35 | 76,7 | |
| | Estado de Insolvencia Insolvente | 35 | 115 | 76,7 | |
| | Porcentaje global | | | 76,7 | |
| Paso 3 | Estado de Insolvencia Solvente | 114 | 36 | 76 | |
| | Estado de Insolvencia Insolvente | 33 | 117 | 78 | |
| | Porcentaje global | | | 77 | |

Fig. 7. Pronóstico o clasificación para el modelo 02 de Regresión Logística

En la Fig. 7 queda evidenciado el poder de pronóstico del modelo Regresión Logística con las variables mostradas en el cuadro 3. Para el paso 1 se tuvo que de 150 empresas solventes se clasificaron a 115 empresas efectivamente como solventes, y que de estas 150 empresas solventes, se clasificaron 35 empresas como insolventes, dando como resultado un porcentaje correcto de pronóstico del 76,7%. De 150 empresas insolventes (en el paso 1), se clasificaron 35 como solventes y a 115 empresas, efectivamente, como insolventes; con un porcentaje de pronóstico correcto del 76,7%. Para el paso 2, con la inclusión de la variable CON_ENC, el porcentaje de pronóstico fue el mismo que en el paso 1. En el paso 3 (modelo con las tres variables: REE_ANT, CON_ENC y RAZ_COR), de 150 empresas solventes se clasificaron 114 empresas como solventes y 36 empresas como insolventes; de 150 empresas insolventes se clasificaron 33 como solventes y 117 efectivamente como insolventes, lo que generó un porcentaje promedio de pronóstico del 77%.

V. CONCLUSIONES

El modelo de regresión logística para empresas PYME en Colombia, permitió establecer que las empresas que reflejaron en sus estados financieros una utilidad acumulada de ejercicios anteriores negativa, tendrían ocho veces más probabilidad de caer en estado de insolvencia que una empresa con una utilidad acumulada positiva; así mismo, empresas con una razón corriente por debajo de uno tendrían tres veces más probabilidad de caer en estado de insolvencia que una empresa con una razón corriente por encima de uno.

Como fuente para la búsqueda de posibles variables para el análisis financiero, los estados financieros básicos son una buena herramienta para determinar las variables que determinan el comportamiento de las empresas, ya que los registros de las cuentas del balance general, el estados de resultados y el flujo de efectivo reflejan la situación estática y dinámica de estas en un momento determinado, a lo cual se le suma las aplicaciones de técnicas estadísticas y matemáticas para el análisis exhaustivo desde el punto de vista cuantitativo que pueden generar criterios para la toma de decisiones, análisis de correlaciones, pronósticos, optimización, etc. Para el caso de este estudio se logró

manipular estadísticamente las variables relacionadas con la estructura financiera, con las utilidades y con el flujo de efectivo, de manera que en vez de tener solamente variables cuantitativas, se pudieran considerar algunas de estas como variables categóricas; esto permitió tener un referente tangible y atractivo para la comprensión de la determinación de probabilidades de una forma intuitiva.

El riesgo de insolvencia puede ponerse en perspectiva de acuerdo a la definición de riesgo de crédito que es la expresión que se utiliza para analizar el riesgo, desde el punto de vista de las entidades financieras de que sus clientes adeudados caigan en estado de insolvencia y no puedan cumplir con el pago de su capital e intereses.

Entre el riesgo de insolvencia y el riesgo de crédito no se debe discernir sino en la perspectiva de los que utilizan la maquinaria para la estructuración, análisis e interpretación de los modelos de riesgo; para el caso de la utilización de la expresión “riesgo de insolvencia” es la que se utiliza para determinar el riesgo de que la empresa caiga en estado de insolvencia y no pueda pagar su deuda; el “riesgo de crédito” es lo mismo, pero es utilizado con el objetivo de determinar si una empresa está en la capacidad de endeudarse con la entidad; un valor agregado del riesgo de crédito es que se genera una puntuación en letras, aunque esto son puntuaciones que utilizan las calificadoras de riesgo en sus programas automatizados para el análisis de riesgo de crédito.

En el presente estudio de investigación se buscó estructurar un modelo que permitiera determinar la probabilidad de que una empresa PYME cayera en estado de insolvencia de acuerdo a factores dicotómicos, representados por las cuentas de los estados financieros. El resultado fue un modelo de regresión logística binaria estructurado en SPSS 19, con ocho variables independientes categóricas (dicotómicas) y cuantitativas, y un variable dependiente dicotómica, el cual fue resuelto por medio del método de máxima verosimilitud para la determinación de los parámetros o coeficientes de la ecuación matemática; estos parámetros solo pudieron ser hallados por iteraciones, por lo que la utilización del paquete estadístico mencionado fue imprescindible para la consecución de los resultados.

El modelo de regresión logística para la medición del riesgo de insolvencia tuvo un poder de pronóstico del 77%, lo que indica que el modelo construido tuvo un poder de explicación relativamente alto, teniendo en cuenta que para la explicación de este evento quedaron finalmente en el modelo sólo tres variables: la razón corriente, resultados de ejercicios anteriores y la concentración del endeudamiento en el corto plazo.

Selección de Variables

Como para cualquier otro modelo estadístico, la selección de variables es una de las etapas en la investigación cuantitativa que requiere una gran atención. Para el caso

de los modelos de riesgo de crédito, se han establecido a través del tiempo, múltiples criterios o perspectivas para la selección de las variables independientes:

- Factores que las entidades financieras consideran para medir el nivel de riesgo en las operaciones de crédito que conceden a sus clientes.

En esta investigación la selección de variables tuvo como fuente los estados financieros de las empresas objeto de estudio; la clave estuvo en determinar cuáles variables harían parte del modelo de regresión logística.

Se estructuró un modelo de análisis de componentes principales con : Multicolinealidad y Oscilación en el tiempo.

REFERENCIAS

- [1] O. G. Martínez. Determinantes de fragilidad en las empresas colombianas: banco de la república. [En línea], [citado el 8 de febrero de 2013]. Disponible en: <http://www.banrep.gov.co/document/ftp/borra259.pdf>
- [2] Luis m. Molinero. La regresión logística. [En línea], 2001 disponible en: <http://www.seh-lha.org/pdf/rlogis1.pdf>
- [3] R. P. Botella, b. M. Martínez & g. M. Alacreu. Regresión logística. Universidad cardenal herrera. [En línea], págs. 10 Y 11. 2003 Disponible en: <http://maphysco.ceuc.es/~pbotella/introduccionalaregresionlogisticaconr-commander.pdf>
- [4] C. Beltrán. Análisis de regresión logística aplicado a la clasificación textos académicos: biometría y filosofía. 2010 [En línea]. Pp. 2. Disponible en: <http://www.Revistaepistemologi.Com.Ar/biblioteca/analisisderegresionlogisticaaplicadoalaclasificaciondetextosacademicosbiometriayfilosofia.pdf>
- [5] L. Nava & s. Surendra. Ajuste e interpretación de los modelos de regresión logística con variables categóricas y continuas. 2007 [En línea], pp. 10. Disponible en: <http://med.javeriana.edu.co/publicaciones/universitas/serial/v49n1/4-ajustemodelo.pdf>
- [6] C. M. Aguayo. Cómo hacer una regresión logística con spss© “paso a paso”. [En línea], 2007 pp. 4. Disponible en: http://www.fabris.org/ht/ml/archivos/ocuw/regres_log_1r.pdf
- [7] H. A. De lara. Capítulo 9: modelos de riesgo de crédito. En: medición y control de riesgos financieros. (3 Ed.) México: editorial limusa s.A., 2005, pp. 163
- [8] P. Jorion. Capítulo 13: riesgo crédito. En: valor en riesgo: el nuevo paradigma para el control de riesgos con derivados. México: editorial limusa s.A., 2004, pp. 267
- [9] R. Knop, r. Ordoñas, & j. Vidal. Capítulo 3: riesgo de crédito. En: medición de riesgos de mercado y crédito.(1 Ed.) Barcelona: editorial ariel s.A., 2004, Pp. 187
- [10] M. L. Berenson, d. M. Levine & t. C. Krehbiel. Capítulo 9: regresión lineal simple y correlación. En: estadística para administración. (2 Ed.) México: editorial pearson educación, 2001, pp. 504
- [11] D. R. Anderson, d. J. Sweeney & t. A. Williams. Capítulo 9: aplicaciones de programación lineal. En: métodos cuantitativos en los negocios.(7 Ed.) México: editorial thomson editores s.A., 1999. P. 367. isbn: 968-7529-56-3
- [12] A. L. Webster. Capítulo 12: regresión múltiple y correlación. En: estadística aplicada a los negocios y la economía. (3 Ed.) Bogotá: editorial irwin mc graw hill, 2000, pp. 382
- [13] R. Larson, & b. H. Edwards. Introducción al álgebra lineal. México df: limusa noriega editores, 2002. isbn: 968-118-4886-1
- [14] Dirección de impuestos de aduanas nacionales dian. Resolución (00432): clasificación internacional industrial uniforme revisión 3.1. (19 De noviembre de 2008). Recuperado de http://empr.esas.micodensa.com/bancomedios/documentos%20pdf/resolucion_00432_actividades_economicas_2008.pdf

- [15] M. T. Tascón f., F. J. Castaño g. Predicción del fracaso empresarial: una revisión p. 12.
- [16] T. Shumway. Predicción más precisa de la quiebra empresarial: modelo simple hazard (2001). Vol. 74 No. 1 Pp. 101-124



Eduardo Arturo Cruz Trejos nació en Pereira, Colombia, el 17 de junio de 1959. Se graduó en Ingeniería Industrial en la Universidad Tecnológica de Pereira – Colombia, en septiembre de 1983. Se graduó en la misma Universidad en mayo del 2000, en la Maestría de Administración Económica y Financiera.

Ejerce profesionalmente en la Universidad Tecnológica de Pereira – Colombia, ubicada en la vereda la Julita de Pereira, y ha ejercido en el Ministerio de Hacienda y Crédito Público, Comunicación Celular Comcel, Banco Central Hipotecario, Contacto Ltda. (Fundador y propietario), Genéricos del café Ltda. (Fundador y propietario) y La Empresa de Transportes el Proveedor Ltda. Entre los campos de interés están las finanzas cuantitativas y la optimización financiera aplicada a los mercados de capitales.

El ingeniero Cruz Trejos ha publicado cerca de cincuenta artículos en revistas indexadas, es coautor de once libros resultado de proyectos de investigación y es coautor de seis software, resultado de proyectos de investigación, entre los que se destacan: Financiamiento Optimizador, Análisis Técnico de renta variable, Administración de inventarios, Portafolio óptimo del mercado accionario.

Actualmente se desempeña como profesor titular en la Universidad Tecnológica de Pereira – Colombia, en la facultad de Ingeniería Industrial en las asignaturas del área financiera. Además orienta asignaturas en la maestría de Administración económica y financiera, y es el director del grupo de investigación en finanzas categoría B en Colciencias. Está en la categoría de investigador asociado en Colciencias.

Jaime Espinosa Peña nació en Pereira, Colombia, el 15 de mayo de 1949. Se graduó en Economía en la Universidad La gran Colombia - Bogotá – Colombia, en Febrero de 1971. Se graduó en la Universidad Central en septiembre de 1998, en la Maestría de Ciencias financieras y de sistemas.

Ejerce profesionalmente en la Universidad Libre sede de Pereira – Colombia, ubicada en Belmonte - Pereira, y ha ejercido en la Corporación autónoma regional de Risaralda como economista planificador, Universidad Tecnológica de Pereira como docente, Manganesos de Colombia como Jefe de departamento, Los tres elefantes s.a. como jefe de departamento administrativo, Entre los campos de interés está El análisis económico regional y las finanzas empresariales.

El Economista Espinosa Peña ha sido director del programa de Ingeniería financiera en la Universidad Libre de Pereira.

Actualmente se desempeña como profesor en la Universidad Libre de Pereira, la Universidad del Valle sede de Cartago y la Universidad del Quindío. Es el director del grupo de investigación en Ingeniería Financiera categoría D en Colciencias. Está en la categoría de investigador Junior en Colciencias.



Sergio Aristizábal Hernández nació en Cartago, Valle del Cauca, Colombia, el 13 de febrero de 1989. Se graduó como ingeniero financiero en la Universidad Libre de Colombia.

Ejerció profesionalmente en la Universidad Libre en el departamento de investigación del programa de ingeniería financiera. Entre sus campos de interés están: el análisis de riesgo de crédito empresarial, y el diseño de instrumentos financieros sintéticos.

El ingeniero Aristizábal Hernández recibió una mención de meritorio por su trabajo de grado en la Universidad Libre de Colombia. Actualmente, es estudiante de la Maestría en Finanzas de la Universidad de Los Andes.