

# MODELOS ESTADÍSTICOS MULTIVARIANTES DE PRONÓSTICO Y DE CLASIFICACIÓN NO PARAMÉTRICOS PARA EL ANÁLISIS DE RIESGO BANCARIO<sup>1</sup>

## MULTIVARIATE STATISTICAL FORECASTING AND NONPARAMETRIC CLASSIFICATION MODELS FOR THE ANALYSIS OF BANKING RISK

**Grupo Banca:** Gerardo Colmenares (*gcolmen@ula.ve*)<sup>2</sup>, Giampaolo Orlandoni (*orlandon@ula.ve*)<sup>4</sup>, Rafael Eduardo Borges<sup>5</sup> (*borgesr@ula.ve*), Ruth Guillén (*guillenr@ula.ve*)<sup>2</sup>, Alexis Melo (*amelo@ula.ve*)<sup>6</sup>, María Alejandra Ayala (*marialej@ula.ve*)<sup>5,7</sup>, Zuleima Andreina Durán<sup>8</sup>, Carlos Martínez<sup>9</sup>, Daniel Paredes<sup>6</sup>, José Valera<sup>7</sup>

### Summary

Often the decision tools related to banking risk control are being more affected in their outcomes due to the dynamic influence of the global markets to the local and international financial policy. The financial institutions require specialized methodologies and tools to measure and estimate critical values of banking risk and therefore, *multivariate statistics and artificial intelligence* techniques stand out the construction of prognosis and classification models. These techniques allow to know variables and underlying risk factors, which they clearly define the risk patterns and therefore, to give answer to needs of consistent information. Consequently, there will be models where data processing is a determining factor for building categorical (financial risk) and quantitative (financial reasons) variables. The identification of those variables and the acceptable data quality are used in hybrid models based on a *parametric* and *non parametric* source. Those models are able to recognize patterns of the behavior risk in the banking system by prognosis and classification options. Techniques involved for the conformation of such hybrid models are: a) Principal Component Analysis (PCA) like variable reduction, b) Discriminant Analysis (DA) like output (dependent) variable pre-processor, c) Survival Analysis (SA) as a method to predict banking survival, and, d) Non-supervised Models to predict and classify patterns based on Neural Networks Algorithms (NN), fundamentally back-propagation training model (NN-BP) and the radial basis function training model (NN-RBF).

### Resumen

Los mecanismos de decisión pertinentes al control de riesgo bancario cada vez se ven más afectados en sus mediciones, debido a la dinámica impuesta por los mercados globales y su implicación, de impacto local e internacional, en las políticas financieras. Para las mediciones y estimaciones de las exposiciones al riesgo por parte de las instituciones financieras es necesario utilizar metodologías y herramientas especializadas, destacándose la construcción de modelos de pronóstico y clasificación mediante técnicas *estadísticas multivariantes y de inteligencia artificial*. Estas técnicas permiten la consideración de variables y factores definitorios de los riesgos que se van a estimar; esto es, el reconocimiento de comportamientos que claramente definen los patrones de riesgo para dar respuesta a las necesidades de información

---

<sup>1</sup> El presente proyecto de investigación está circunscrito al proyecto sobre Gestión del Riesgo Bancario auspiciado por el Banco Central de Venezuela (BCV) y desarrollado en el Instituto de Investigaciones Económicas y Sociales de la Universidad de Los Andes (IIES – ULA); Mérida, Venezuela, el cual propone aplicar metodologías estadísticas y no estadísticas al problema de la gestión de riesgos en la banca. Convenio ULA-BCV , grupo transdisciplinario *Grupo Banca*.

<sup>2</sup> Instituto de Investigaciones Económicas y Sociales (IIES). Universidad de Los Andes. Mérida. Venezuela

<sup>3</sup> Instituto de Estadística Aplicada y Computación (IEAC). Universidad de Los Andes. Mérida. Venezuela

<sup>4</sup> Instituto de Estadística Aplicada y Computación (IEAC). Universidad de Los Andes. Mérida. Venezuela

<sup>5</sup> Departamento de Estadística, Escuela de Estadística. Universidad de Los Andes. Mérida. Venezuela

<sup>6</sup> Postgrado en Estadística. IEAC. Universidad de Los Andes. Mérida. Venezuela

<sup>7</sup> Postgrado en Economía. IIES. Universidad de Los Andes. Mérida. Venezuela

<sup>8</sup> Tesis en Estadística. Escuela de Estadística. Universidad de Los Andes. Mérida. Venezuela

<sup>9</sup> Tesis en Ingeniería de Sistemas. Facultad de Ingeniería. Universidad de Los Andes. Mérida. Venezuela

consistente. Se construirán, por tanto, modelos donde el preprocesamiento de datos es un factor determinante para la generación de variables categórica (riesgo financiero) y cuantitativas (razones financieras). La identificación de estas variables y la aceptable calidad de los datos que ellas definen, son usadas en modelos híbridos de origen *paramétrico* y *no paramétrico* conducentes al reconocimiento de los patrones de comportamiento del riesgo en el sistema bancario mediante las opciones de pronóstico y clasificación. Las técnicas involucradas para la conformación de los modelos híbridos son: a) Análisis de Componentes Principales (ACP) como mecanismo de reducción de variables, b) Análisis Discriminante (AD), como preprocesador de la variable dependiente, c) Análisis de Supervivencia (AS), como método de pronóstico de supervivencia para cada banco, y, d) modelos no supervisados, para pronóstico y clasificación de Redes Neuronales (RN), fundamentalmente la función logística con propagación del error hacia atrás (RN-RP) y la función de base radial (RN-BR).

# 1. MODELO DE CLASIFICACIÓN PARAMÉTRICA Y NO PARAMÉTRICA DE LOS FENÓMENOS DE RIESGO BANCARIO EN VENEZUELA

*Alexis A. Melo T., Gerardo Colmenares*

Se trata de un modelo híbrido en el que se aprovecha las bondades de la técnica multivariante lineal, conocida como Análisis Discriminante (AD) y el aprendizaje cualitativo no supervisado mediante la función base radial, bajo la topología de redes neuronales para la gestión de riesgo bancario. Este modelo se configura en dos etapas: la primera, consiste en determinar los patrones de salida mediante AD y la segunda, en determinar un modelo de pronóstico y a su vez, de reconocimiento del comportamiento de la banca de acuerdo a una agrupación que surge en la composición automática de los grupos, luego de disponer de los patrones de entrenamiento, después de haber sido transformada las razones financieras a una escala. El desempeño de este modelo está siendo evaluado mediante el uso del conjunto de datos construido para tales fines mediante treinta muestras de iguales características y bajo una topología exactamente igual para cada uno de los treinta modelos entrenados y probados.

El objetivo fundamental de este proyecto, es conjugar ambas técnicas para obtener un mecanismo automático de clasificación de riesgo global en la banca y la eventual capacidad de medir la consistencia de generalización mediante el comportamiento de un banco en particular, dado que es conocido un nuevo ejemplo.

## 1.1. Análisis Discriminante aplicado al riesgo bancario

Se tiene un amplio conjunto de datos bancarios. A partir de los balances generales de cada banco recopilados mensualmente, fueron construidas las razones financieras que en definitiva son las observaciones empleadas de la banca comercial y universal durante ocho años. Son un conjunto de mediciones repetidas con una dimensión de 16 bancos, 23 razones financieras y 108 observaciones por banco, 39744 observaciones. Guillén R., (2005). El problema del AD se traduce en la habilidad que se tiene, mediante un modelo lineal, de poder clasificar este conjunto de datos de entrada, en función de una regla de discriminación o de clasificación predeterminada. La aplicabilidad de este método tiene como objetivos, el de identificar un conjunto de variables que *mejor* discrimine o separe linealmente entre los grupos. La nueva variable **Z**, dada por la proyección de las variables originales sobre este eje, provee la máxima separación o discriminación entre 2 grupos, vale decir, un hiperplano que permita la máxima separación entre grupos para el caso multidimensional. De este modo, se puede clasificar futuras observaciones dentro de uno de los grupos. Es decir su capacidad de pronóstico. Sharma, (1996)

Al Análisis Discriminante se le puede considerar también como una técnica de Clasificación Supervisada. Dada una muestra de observaciones previamente procesadas, se permite la clasificación de datos adicionales.

Este método es de amplia aplicación en diversos campos. Para el área financiera, tiene especial aplicabilidad en el análisis del credit-scoring y del estudio del problema de la turbulencia financiera en el sistema bancario. Altman et al, (1994); Atiya, (2001); Foster et al, (1999).

Esta relación de discriminación lineal viene dada por la expresión  $Z = \mathbf{X}'\mathbf{y}$ , donde **X** representa el vector de variables observadas, **y** son los pesos de la función discriminante y **Z** es la función discriminante. **Z** resultará una variable dependiente indicadora y no numérica. Por ejemplo, la variable **Z** podría ser una variable indicadora, que denota la ocurrencia o no de un evento.

En este contexto, el Análisis Discriminante se emplea para determinar cuál o cuáles variables contribuyen a discriminar entre dos o más grupos.

En resumen, la idea básica que subyace en el Análisis Discriminante es determinar si unos grupos difieren en función de la media de una variable, y emplear luego esa variable para predecir la pertenencia de una nueva observación a determinado grupo.

Visto desde el análisis de varianza, consiste en responder a la pregunta de si dos o más grupos son significativamente diferentes uno de otro respecto a la media de una variable en particular.

En el caso particular de función discriminante para dos grupos, tal función puede verse como un caso particular de análisis de regresión lineal múltiple. Si se codifican los dos grupos como 1 y 2, y se emplea

tal variable como dependiente pueden obtenerse resultados similares a los que se obtendrían de un análisis discriminante.

Para dos grupos se ajusta una ecuación lineal del tipo  $GRUPO = a + b_1x_1 + b_2x_2 + \dots + b_mx_m$ , donde,  $a$  es una constante y  $b_1$  a  $b_m$  son coeficientes de regresión o pesos de la función discriminante. La interpretación de estos resultados es similar a la de un modelo de regresión múltiple. Los más significativos son los que contribuyen más a la predicción de pertenencia a un grupo.

Cuando es posible identificar más de dos grupos, puede estimarse más de una función discriminante similares a la presentada anteriormente. Por ejemplo, cuando se tienen tres grupos, puede estimarse: a) una función para discriminar entre grupo 1 y grupos 2 y 3 combinados, y b) otra función para discriminar entre grupo 2 y grupo 3.

Es importante tener presente unos supuestos implícitos a este tipo de análisis: a) Se asume que los datos para las variables representan una muestra proveniente de una distribución normal multivariable. No obstante, el no cumplimiento de este supuesto no es problema para el análisis. b) Se supone que las matrices de varianzas y covarianzas son homogéneas entre grupos. c) El principal obstáculo para la validez de las pruebas de significancia se presenta cuando la media de las variables entre grupos están correlacionadas con las varianzas. d) Se supone que las variables empleadas para discriminar entre grupos no son completamente redundantes; por ejemplo que una variable no sea la suma de otras dos que también están en el modelo.

En la figura siguiente se puede observar un esquema funcional del preprocesamiento de los datos financieros de la banca comercial basados en la aplicabilidad del análisis discriminante.

Figura 1. Discriminación de los bancos en dos grupos

La aplicabilidad de AD consiste en dado que se ha construido en una escala homogénea mediante la estandarización de los datos originales (razones financieras), y posteriormente categorizándolos de acuerdo a sus valores extremos, como patrones de entrada, El resultado serán los patrones de salida en el entrenamiento del modelo de red neuronal no supervisado usando función de base radial. El propósito de esta construcción híbrida es el de lograr un modelo con un alto grado de aprendizaje que se pueda generalizar a todo el dominio del problema real (la banca comercial y universal) bajo el mismo contexto de las razones financieras empleadas para el entrenamiento. Un buen desempeño de pronóstico será evaluado con muestras no utilizadas en el entrenamiento por la red y que puedan ser separadas en una de las categorías con un alto grado de aceptación.

## 1.2. Alerta temprana y agrupación automática mediante la función de base radial (RBF)

La topología regular de una red neuronal RBF consiste en tres capas: a) la capa de entrada que sirve para los ejemplos o patrones de entrenamiento y prueba, b) la capa oculta completamente interconectada entre todos sus nodos con la capa de entrada y activada a través de la función radial (de origen gaussiano) y, c) la capa de salida, también completamente interconectada a la capa oculta o intermedia y activada a través de una función lineal continua. El objetivo principal es el de ejecutar una *correspondencia no lineal* entre los patrones de entrenamiento que conforman el espacio de entrada al espacio oculto definido por la capa oculta y una *correspondencia lineal* desde este espacio al espacio de salida. Es decir, definir a la salida una superficie que describa las entradas mediante clases o categorías.

El entrenamiento no supervisado, a diferencia de la red usando retro-propagación del error (backpropagation) (RN-RP), es solamente hacia adelante. De este modo, la salida  $z$  de una red RBF, en general, está influenciada por una transformación no lineal originada en la capa oculta, a través de la función radial y la función lineal en la capa de salida mediante función lineal continua.

Tal como lo muestra la figura 2, los nodos ocultos contienen una función base radial o función gaussiana, cuyos parámetros son un centro y un ancho. El centro es único para cada función radial involucrada en la capa oculta, y el ancho identifica la amplitud de la campana de gauss originada por la función radial, es decir, la desviación estándar de la función radial. Algunos autores, Broomhead y Lowe, (1988), consideran a este ancho como un valor constante para cada una de las funciones radiales

en la capa oculta para simplificar los pasos de construcción del modelo de entrenamiento de la red. El primer cálculo efectuado en la capa oculta consiste en hallar para cada nodo de la capa oculta, la distancia radial (distancia euclidiana)  $d$  entre el vector de entrada (patrón o ejemplo de entrenamiento)  $\mathbf{x}$  y el centro de gravedad  $\mathbf{c}$  de ese mismo nodo, para cada una de las  $n$  observaciones. Es decir:

$$d = \|\mathbf{x} - \mathbf{c}\| = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + \dots + (x_n - c_n)^2} = \sum_{i=1}^n \sqrt{(x_i - c_i)^2}$$

Este valor  $d$  es calculado para cada nodo oculto y es un componente de la entrada para activar la función radial  $G(\bullet)$  correspondiente a cada nodo. La función radial  $G(\bullet)$  más comúnmente empleada es  $\exp(-r^2)$ , donde  $r$  es el contenido evaluado en cada nodo de la capa oculta. En este caso particular, este contenido es la distancia euclidiana  $d$ . De ahí que la expresión anterior sería  $\exp(-d^2)$ .

Una de las derivaciones del modelo RBF es emplear el ancho (desviación estándar) para activar la función  $G(\bullet)$ . En este caso se estaría trabajando con algo similar a  $\exp(-d^2/a)$ , donde  $a$  es el ancho para ese nodo oculto.

Entre la capa oculta y la capa de salida se derivan un conjunto de pesos  $\mathbf{w}$  que se verían afectados de acuerdo al algoritmo de aprendizaje. En este caso particular sería la combinación lineal entre los pesos y la resultante de cada función radial para determinar la salida  $\mathbf{z}$ .

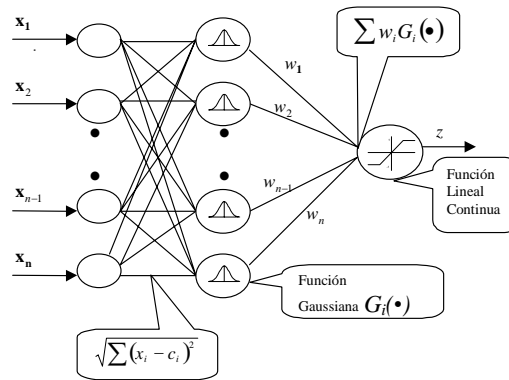


Figura 2. Componentes de una RBF: función de activación en ambas capas, pesos sinápticos

En definitiva, sería  $\mathbf{z} = \sum w_i G(\bullet)$ , donde  $G(\bullet)$  es la salida de la capa oculta y se corresponde con la función radial aplicada a la distancia euclidiana en cada una de las unidades ocultas.  $\mathbf{z}$  es una combinación lineal entre la función no lineal  $G(\bullet)$  y los pesos.  $G(\bullet)$  es de la forma  $\exp(-\|\mathbf{x}_j - \mathbf{c}\|)$  y la expresión anterior, en forma matricial sería  $\mathbf{G}\mathbf{w} = \mathbf{z}$ , donde:

$$g_{j,i} = g(\|\mathbf{x}_j - \mathbf{c}\|), j, i = 1 \dots N; \mathbf{z} = [z_1, z_2, z_3, \dots, z_N]^T \text{ y } \mathbf{w} = [w_1, w_2, w_3, \dots, w_N]^T$$

Dado que todas las observaciones son distintas, entonces  $\mathbf{G}$  se podría decir que es positiva definida y por lo tanto los pesos podrían ser calculados mediante la inversa de  $\mathbf{G}$ . Es decir,  $\mathbf{w} = \mathbf{G}^{-1}\mathbf{z}$ . Sin embargo, se puede correr el riesgo de que la inversa de la matriz de interpolación  $\mathbf{G}$  pueda ser singular. En este caso se procedería mediante la teoría de la regularización para perturbar la matriz mediante  $\mathbf{G} = \mathbf{G} + \lambda \mathbf{I}$ . Haykin, (1995). De esta manera sería un aprendizaje directo, provocando cambio a los pesos que están ubicados entre la capa oculta y la capa de salida.

Del resultado de este tipo de entrenamiento podemos observar que: a) Los valores de entrada se recomiendan que previamente sean de algún modo transformados a una escala. b) Se recomienda un proceso de entrenamiento usando ejemplos representativos tanto para la entrada como para la salida. Es decir, muestras que incluyan ejemplos de todo el dominio del problema. c) En la capa oculta, en la medida que los valores de entrada se parezcan más a un centro, su distancia tenderá a *cero* y de este modo la función gaussiana se dispararía a las vecindades de *uno*. Por otro lado, en la medida que los valores de entrada no se parezcan a su centro, la distancia será mayor y la función radial parecería

tender a *cero*. Este proceso es una *clasificación no lineal* de las entradas. d) En la capa de salida del modelo RBF, los valores obtenidos en las salidas de la capa oculta serían transformados por la función lineal que permite aproximar los valores  $\mathbf{z}$  a los valores deseados, mediante la combinación lineal que se sucede en esta capa entre sus pesos y el resultado de aplicar la función radial. Es decir,  $\mathbf{z} = \sum w_i G(\bullet)$ .

De acuerdo a este concepto, la red RBF ha originado variantes de cálculo como producto fundamentalmente de las siguientes limitaciones que se han presentado en algunos casos: a) de no conocer los centros (a veces el ancho) para cada función radial; b) de situaciones de singularidad presentes en la implementación del algoritmo con problema de dimensionalidad; c) de un gran volumen de entradas haciendo inmanejable la aplicación del algoritmo. Se presentan problemas de regularización Haykin, (1995).

### 1.3. Diseño experimental.

La aplicabilidad de ambas técnicas, análisis discriminante y redes neuronales de funciones de base radial, describen un modelo híbrido que incluye los datos construidos en una escala homogénea como patrones de entrada definidos por las razones financieras, y el resultado de utilizar el análisis discriminante para obtener los patrones de salida. La figura 4 describe funcionalmente la aplicabilidad de la RBF y en la figura 1 la construcción de los patrones de salida. Los patrones de entrada se construyen transformando las razones financieras en un nuevo conjunto de datos discretizados, tal que se pueda unificar las escalas de medida. Se aplicará el método de discretización supervisado CAIM (Class-Attribute Interdependence Maximization) Ching et al., (1995), Kurgan et al., (2001). CAIM es uno de los más reciente métodos de discretización dando muy buenos resultados en un tiempo de cómputo bastante reducido. Este método divide el rango de la variable, en este caso la razón financiera, en un número muy pequeño de intervalos que podría ser encontrados automáticamente. Si la escala de intervalos se desea dejar fija, entonces se puede aplicar una variante del método. Campos et al., (2004).

Figura 3. Pronóstico de riesgo clasificado en categorías

## 2. MODELO DE IDENTIFICACIÓN DE INDICADORES DE GESTIÓN DE RIESGO FINANCIERO.

*Ruth Guillén, Gerardo Colmenares, Giampaolo Orlandoni*

El Análisis de Componentes Principales (ACP) puede disminuir el efecto de redundancia o dispersión de los fenómenos observados originalmente mediante la obtención de nuevas características con poca pérdida de la información. Esta reducción de variables, además de crear nuevas características para el análisis, permite el desarrollo de modelos de pronóstico donde el trabajo de cómputo se disminuye notablemente. En algunos trabajos referidos a el uso de ACP en inteligencia artificial se ha demostrado que se producen excelente resultados y más aún, un excelente desempeño de pronóstico cuando estas nuevas características son los patrones de entrenamiento de los modelos de redes neuronales mediante el uso del algoritmo no lineal de retropropagación del error (RN-RP). Adicionalmente, y como aporte particular de esta investigación, se está incorporando como patrones de salida de este modelo híbrido (ACP y RN-RP), los valores categorizados de las razones financieras representados en una sola variable de salida. El desempeño de este modelo está siendo evaluado mediante el uso del conjunto de datos construido para tales fines. El logro fundamental de esta iniciativa es la de vincular ambas técnicas para obtener un mecanismo automático de pronóstico de riesgo y la eventual capacidad de generalizar el comportamiento de un banco en particular en un período dado

De acuerdo al tipo de medición que se haga de las variables observadas, puede variar la utilidad de las técnicas que permiten este tipo de transformación. En los métodos multivariantes más comúnmente utilizados tales como el Análisis Discriminante (AD), el Análisis de Correspondencia (AC) y el Análisis de Componentes Principales (ACP), para cada técnica se requiere un diferente tipo de medición de las características de las variables. Tal es el caso de AD donde su fuente de datos lo conforman variables

de tipo categórico o discreto, en ACP las variables son continuas o cuantitativas y en AC son una mezcla de ambos tipos de medición. Sharma, (1996).

En este caso particular, además de aprovechar las bondades de la técnica ACP para preparar información latente mediante la interacción de las variables originales por de combinación lineal entre ellas, se hace énfasis en el significado de reducir significativamente las variables originales sin sacrificar mayor cantidad de información. En este contexto y en el caso de lo que se investiga, se realizó una buena contribución con un muy excelente resultado de pronóstico de riesgo de quiebra desde el punto de vista estadístico. La interpretación de las variables en el nuevo espacio permitió de manera alternativa identificar individuos (banca) propensos a la quiebra como consecuencia de analizar los resultados de los componentes principales obtenidos. Sin embargo, se corre con el riesgo de no poder explicar el dominio original del problema debido a la pérdida de porciones de la varianza explicada presente en los datos originales, como consecuencia de la presencia de variables originales con un alto grado de interdependencia y que no representan un aporte de información significativo. Ayesterán, et al., (1996)

## **2.1 El ACP aplicado al riesgo bancario.**

El ACP como técnica de reducción de variables ha mostrado excelentes resultados. El beneficio alcanzado reutilizando fuentes históricas de datos, es el de permitir la disminución del tiempo de cómputo empleado en la construcción y uso de los modelos debido al empleo de las técnicas de preprocesamiento que seleccionan confiablemente la menor cantidad de observaciones y variables posible. Por ejemplo, el método Stratified/PCA, Colmenares, (2004), como técnica de preprocesamiento, puede reducir una matriz original de observaciones y variables a un subconjunto apreciablemente menor que sus originales con un nivel de confiabilidad y consistencia bastante altos como para poder ser aplicados sobre fuentes históricas de datos. Adicionalmente, estos conjuntos reducidos de datos empleados en la construcción de modelos no lineales usando RNA, han mostrado tener similar capacidad de predicción y/o clasificación que los construidos con el conjunto total de los datos originales. De igual modo, mostraron ser mucho más consistente en sus resultados después de varias réplicas con diferentes conjuntos reducidos de datos. Dong et al. (1995), Tan et al. (1995).

En este estudio se enfatiza la utilidad práctica de la técnica al hacer reducción de las variables originales a un nuevo conjunto, tal que sean las variables explicativas o patrones de entrada, en los modelos de pronósticos. Hay varias experiencias que han demostrado un mejor y eficiente desempeño de los modelos usando estas variables reducidas al disminuir el tiempo de cómputo y el de mantener o mejorar la capacidad de pronóstico. Colmenares, (2004). Atiya hace un buen inventario de resultados evaluativos de la aplicación de este tipo de técnicas, e inclusive algunas no paramétricas, en el análisis del riesgo bancario, fundamentalmente para la clasificación y la predicción. Atiya, (2001). Las técnicas multivariantes, en las mayorías de los estudios revisados, fueron aplicadas a grupos de variables asociadas a las razones financieras, y a su vez, comparadas con métodos numéricos emergentes demostrando su utilidad.

Por otro lado, los patrones de salida, similar a los que conforman la entrada, son estimados mediante análisis discriminante, indicando la característica cualitativa de riesgo dependiente del patrón de entrada que lo origina.

Figura 4. ACP como técnica de reducción de variables

Sea  $\mathbf{X}$ , la matriz de orden  $p$ -dimensional conformada por  $p$  variables (23 razones financieras) con  $N$  observaciones repetidas (cada banco por mes) para cada variable.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdot & \cdot & x_{1p} \\ x_{21} & x_{22} & x_{23} & \cdot & \cdot & x_{2p} \\ x_{31} & x_{32} & x_{33} & \cdot & \cdot & x_{3p} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{N1} & x_{N2} & x_{N3} & \cdot & \cdot & x_{Np} \end{bmatrix} \quad \mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & z_{13} & \cdot & \cdot & z_{1k} \\ z_{21} & z_{22} & z_{23} & \cdot & \cdot & z_{2k} \\ z_{31} & z_{32} & z_{33} & \cdot & \cdot & z_{3k} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ z_{N1} & z_{N2} & z_{N3} & \cdot & \cdot & z_{Nk} \end{bmatrix}$$

Mediante ACP, el conjunto total de  $p$  variables podría ser reducido a un nuevo conjunto *enteramente independiente* de nuevas variables (algunas veces conocidas como variables latentes) expresada en un matriz resultante  $\mathbf{Z}$  de orden  $k$ .

Del nuevo conjunto de variables define un espacio  $k$ -dimensional, mucho más reducido que el original, donde  $k \ll p$ , y se pueden realizar algunas consideraciones: a) las nuevas variables en  $\mathbf{Z}$  son *completamente independientes*; b) la utilidad práctica de la matriz  $\mathbf{Z}$  se reduce a una matriz mas manejable ( $k \ll p$ ); c) para eliminar el efecto de la varianza relativa de las variables originales, la entrada principal a ACP es la matriz de correlación  $\mathbf{R}$ ; d) cada  $z_j$  es una variable transformada de las  $x_i(s)$  variables originales. Siendo  $i=1,2,3,\dots,p$ . Estas nuevas variables contienen los valores de las variables transformadas; e) ACP captura solamente linealidad entre las variables: las  $z_i$  son combinación lineal de los componentes y las variables originales  $x_i$ ; f) los primeros componentes obtenidos mediante ACP explican la mayor cantidad de la varianza total de las variables originales.

Existe un método matricial directo conocido como Valor de Descomposición Singular (VDS) que puede ayudar a resolver ACP. Supóngase que se tiene la misma matriz original  $\mathbf{X}$  de  $N$  observaciones (filas) y  $p$  variables (columnas). La descomposición espectral de esta matriz  $\mathbf{X}$  es  $\mathbf{X} = \mathbf{QSE}'$ , donde  $\mathbf{E}$  es una matriz ortogonal  $p \times p$  con los autovectores de  $\mathbf{X}$ ,  $\mathbf{S}$  es una matriz diagonal  $p \times p$  formada por los autovalores de  $\mathbf{X}$ , y  $\mathbf{Q}$  es una matriz ortogonal  $N \times p$ . Además,  $\mathbf{E}$  y  $\mathbf{Q}$  son ortonormales. De igual modo, el producto  $\mathbf{X}'\mathbf{X}$  forma una matriz simétrica y puede ser escrito mediante la descomposición espectral de la siguiente manera,  $\mathbf{X}'\mathbf{X} = (\mathbf{QSE}')'(\mathbf{QSE}') = \mathbf{ES}'\mathbf{SE}' = \mathbf{EAE}'$ , donde  $\mathbf{E}$  sigue siendo la matriz ortogonal  $p \times p$  con la  $i$ -ésima columna correspondiente al  $i$ -ésimo autovalor. Las variables transformadas quedan definidas por  $\mathbf{Z}$  y  $\mathbf{Z} = \mathbf{XE}$ , que no son más que los valores de  $\mathbf{X}$  proyectados sobre la nueva base ortonormal con ejes definidos por los autovectores  $\mathbf{E}$ . Si las variables originales son estandarizadas, entonces se usa la matriz de correlación  $\mathbf{R}$  en sustitución de la matriz de covarianza  $\Sigma$ . En este caso, la descomposición espectral para  $\mathbf{R}$  sería  $\mathbf{R} = \mathbf{EAE}'$ . De igual manera que para la matriz de covarianza,  $\mathbf{Z} = \mathbf{XE}$  y  $E(\mathbf{Z}'\mathbf{Z}) = \mathbf{\Lambda}$ . Adicionalmente, para este caso en particular, la suma de los autovalores en  $\mathbf{\Lambda}$  es igual a  $p$  y es igual a la suma de de los elementos de la diagonal de  $\mathbf{R}$ . Lo que es lo mismo  $\text{tr}(\mathbf{R}) = \text{rango}(\mathbf{R}) = \text{tr}(\mathbf{\Lambda}) = p$ . Es importante tener en cuenta que los autovectores y los autovalores son sensibles a los cambios de escala. Así,  $\Sigma$  podría mostrar severos cambios por esta misma razón, mientras que  $\mathbf{R}$  omite este problema y de ahí que los autovalores y autovectores originados por  $\mathbf{R}$  o por  $\Sigma$  sean diferentes.

## 2.2 Las Redes Neuronales y el algoritmo de retropropagación del error.

En el entendido de que es un procesador paralelo distribuido en los que se almacena el conocimiento, emula su símil biológico en dos aspectos fundamentalmente: a) el conocimiento es adquirido a través de un proceso de aprendizaje y b) la fortaleza de la interconexión de las neuronas artificiales está descrita en los pesos sinápticos.

Bajo estas consideraciones esenciales, toda red artificial ajusta permanentemente sus parámetros libres (pesos y umbrales) mediante mecanismos de aprendizaje que podrían ser supervisados o no. El caso particular que se aplica a este modelo híbrido es un modelo de red neuronal multicapa que activa el cambio en las interconexiones de los pesos sinápticos a través de una función de excitación no lineal (logística) y un proceso de corrección de estos pesos mediante la constante evaluación de su convergencia a un umbral dado. Este proceso no supervisado se le conoce con el nombre de red neuronal con un algoritmo de retropropagación del error (RN-RP).



Tal como lo muestra la figura 5, la topología de una red neuronal RN-RP muestra una capa de nodos que capturan los componentes de los patrones de entrada uno a la vez y del mismo ejemplo, el patrón de salida correspondiente a la variable categorizada capturado por el nodo de salida. En el intermedio se incluye una capa de nodos que se le conoce como capa oculta y en cada uno de sus nodos se incluye la función de excitación  $\mathbf{G}$ . Esta función puede variar de acuerdo a las características del dominio del problema real. Puede ser funciones asimétricas como la hiperbólica tangencial, la lineal, y la función no simétrica conocida como logística. La salida calculada  $s_c$  en la red durante la fase de entrenamiento se traduce en una expresión del tipo  $s_c = f(\mathbf{W}, \mathbf{G}(\bullet))$ , donde se construye una función no lineal  $f$  al combinar la función de activación  $\mathbf{G}(\bullet)$  con los parámetros libres  $\mathbf{W}$ .

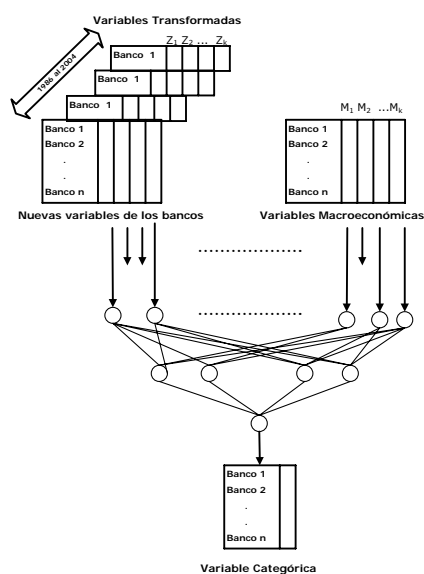


Figura 5. ACP y RN-RP para la construcción del modelo híbrido de pronóstico

El proceso de aprendizaje en este tipo de redes es básicamente en dos fases (Werbos 1974; Rumelhart et al. 1986): a) *Fase hacia delante*. Los parámetros libres  $\mathbf{W}$  de la red son establecidos y el patrón de entrada  $i$  es propagado a través de la red capa por capa. Esta fase finaliza con el cálculo del error  $e_i$  entre la salida calculada  $s_c$  y el patrón de salida dado  $s_d$ ,  $e_i = s_d - s_c$ . b) *Fase hacia atrás*. El error  $e_i$  es propagado a través de la red en dirección hacia atrás. De ahí el nombre del algoritmo. Es en esta fase que se realizan los ajustes a los parámetros libres de la red, así que se minimice el error en un sentido estadístico, mediante el error cuadrático medio. Este algoritmo de retropropagación se procesa en el modo lote, regularmente. Los pesos se ajustan sobre la base de ciclo a ciclo, donde cada ciclo consiste de el conjunto completo de patrones de entrada y salida disponible para el entrenamiento.

El número de nodos en la capa oculta estará directamente vinculado con la cantidad de parámetros libres que incluiría la red neuronal y el nivel de convergencia deseado. De este modo, existirán una fase inicial de entonación de la red, ajustando nodos ocultos, función de activación, separación de los conjuntos de datos de entrenamiento y verificación, ajuste de la tasa de aprendizaje, selección del momento ideal para escapar lo más de las soluciones locales, nivel del error de convergencia, selección de un conjunto de pesos iniciales. La segunda fase que es totalmente experimental comprende solamente la construcción y prueba de los modelos. El criterio de parada empleado es el de la verificación cruzada (Haykin, 1999) que consiste en: a) Dividir el conjunto de entrenamiento en dos partes: subconjunto de estimación usado para el entrenamiento y subconjunto de verificación para evaluar el desempeño del modelo previamente entrenado; b) Entonación final de la red usando el

conjunto entero de entrenamiento y luego probar la generalización con patrones nunca conocidos por el modelo.

### **2.3 Sección experimental**

Haciendo referencia al diagrama funcional mostrado en la figura 3 y para el caso particular de los datos de la banca comercial compuesto por las razones financieras y la agrupación obtenida en el modelo de análisis discriminante en una variable categórica, se dividirá en dos subconjuntos al conjunto original, antes de iniciar la construcción de los modelos. El conjunto original de datos de aproximadamente 8000 observaciones, se dividirá en uno para realizar el entrenamiento, verificación y puesta a tono del modelo, de aproximadamente 6000 ejemplos, y el otro, formado por un subconjunto con aproximadamente 2000 ejemplos nunca usados por el modelo serán usados para verificar el nivel de desempeño o generalización. En la construcción de la red neuronal y como un primer intento de estimación del número de nodos presentes en la capa oculta, se empleará el criterio heurístico que determina el orden de acuerdo a una relación entre el nivel de error deseado y el número limitado de parámetros libres involucrados (pesos y umbrales).

Los modelos serán replicados una cantidad aceptable de veces con diferentes grupos de datos de entrenamiento y manteniendo constante las consideraciones iniciales asumidas para la topología y arquitectura de la red neuronal. El resultado de obtener varios modelos es para que permita medir la consistencia de los valores en las tres fases fundamentales de la construcción de los modelos: entrenamiento, verificación y generalización. Las medidas de evaluación que garantizarían resultados aceptables y consistentes serán el error medio cuadrático calculado entre los valores deseados y calculados en la fase de entrenamiento y verificación, el error medio cuadrático (RMSE) obtenido durante la fase de generalización. Ambos resultados serán medidos para cada modelo y luego de su tabulación, serán observadas las variaciones en los resultados del RMSE obtenidos en las dos fases para determinar el grado de consistencia conseguido en la construcción de los modelos bajo condiciones uniformes de diseño. Una alta consistencia conseguida entre los modelos entrenados y probados garantiza una aceptable robustez en los procesos de pronóstico bajo las condiciones de construcción que se siguieron al inicio.

## **3. CONTROL DE CALIDAD DE INDICADORES DE ALERTA TEMPRANA DETERMINADOS MEDIANTE MODELOS DE ECUACIONES ESTRUCTURALES. CASO: RIESGO FINANCIERO EN LA BANCA NACIONAL**

*Zuleima A. Durán M, Giampaolo Orlandoni*

El vínculo entre las crisis financieras y la actividad económica de un país, es un aspecto trascendental, que permite comprender, más allá de los intereses de las propias instituciones, la importancia de identificar y medir la exposición al riesgo de las entidades bancarias. En muchos casos, estas crisis han significado una pérdida de capacidad de los mercados financieros de canalizar eficientemente fondos hacia oportunidades de inversión productivas. Los bancos dejan de funcionar como orientadores de fondos y transformadores de plazos dificultando la puesta en marcha de proyectos y viabilidad de los ya existentes. Giampaolo Orlandoni, Colmenares Gerardo y otros (2005).

Debido a la frecuencia con las que se han presentado las crisis financieras en las últimas dos décadas, la oportuna evaluación del riesgo se ha convertido en una herramienta de estudio vital para las instituciones financieras, ya que ello permite implementar medidas correctivas para restaurar la solidez de las entidades y limitar las pérdidas potenciales que ellas arraigan. Una de las razones que explican la aparición de las crisis financieras es la elevada exposición que poseen las instituciones bancarias a riesgos muy diversos y al inadecuado manejo de los mismos. El riesgo, puede definirse como la volatilidad de los flujos financieros, generalmente derivada del valor de los activos o los pasivos de la organización en estudio. Las entidades bancarias están expuestas, en general, a tres tipos de riesgo: riesgos de negocios, riesgos estratégicos y riesgos financieros. Guillén Ruth (2005).

Cada uno de los tipos de riesgo relacionados con el riesgo financiero, puede ser explicado por alguna de las razones financieras. La metodología CAMEL ha sido la más utilizada en la determinación de dichas razones, la cual considera varios factores en la evaluación de los bancos. Evaluando dichos indicadores, y observando la estructura de aquellos que salen de los parámetros “normales”, se determina a que tipo de riesgo está sometida la banca y se toman las medidas correctivas.

Sin embargo, este proyecto de investigación pretende dar un enfoque diferente. Aplicando Modelos de Ecuaciones Estructurales se busca obtener algunos indicadores claves de cada tipo de riesgo financiero (crédito, liquidez, mercado y operacional). Seguidamente, y para darle utilidad a la investigación, se pretende evaluar mediante técnicas de control estadístico de calidad, la situación de algunas entidades de la banca universal y comercial ante cada tipo de riesgo asociado al riesgo financiero en el periodo 1996-2004. De este modo se pretende: a) Determinar cuáles son las razones financieras que pueden incidir en la determinación de los tipos de riesgo financiero y su relación con las variables macroeconómicas detonantes de riesgo bancario. b) Evaluar el comportamiento conjunto de los indicadores de cada tipo de Riesgo Financiero, en las entidades bancarias comerciales y universales más importantes del país, en el periodo 1996-2004.

Así, la investigación implicará la evaluación del proceso desencadenado por indicadores de alerta temprana, previamente determinados mediante MEE, para cada tipo de Riesgo Financiero.

### **3.1 La banca, los modelos de ecuaciones estructurales y los sensores de calidad.**

La función principal de las instituciones financieras es administrar activamente los riesgos financieros. Dichas instituciones se han percatado que deben identificar y medir con precisión los riesgos para posteriormente controlarlos y evaluarlos en forma apropiada. Un entendimiento cabal del riesgo permite que los administradores financieros puedan estar en condiciones de planear adecuadamente la forma de anticiparse a posibles resultados adversos y sus consecuencias y, de este modo, estar mejor preparados para enfrentar la incertidumbre futura sobre las variables que puedan afectar sus resultados. La administración del Riesgo, que puede definirse como el proceso mediante el cual se identifica, se mide y se controla la exposición al riesgo, se ha convertido en una herramienta esencial para la supervivencia de cualquier actividad empresarial. Guillén Ruth (2005).

Las instituciones bancarias, bien sea por factores como el entorno macroeconómico, los shocks agregados a la economía (modificaciones importantes en las principales variables macroeconómicas que afectan la economía en su conjunto), la naturaleza propia de los intermediarios financieros o por deficiencias de la propia gerencia bancaria, pueden verse expuestas a diversos tipos de riesgo: Crédito, Mercado, Operacional, Liquidez, Legal y Reputación, todos ellos contenidos en lo que podemos llamar: Riesgo financiero. Pastor, Eduardo (2003). Para detectar cuándo un banco está expuesto a cada tipo de riesgo, se han creado sistemas de indicadores de alerta temprana, basados principalmente en el método CAMEL. Estos indicadores evalúan la banca en base a: Capital (C), Activos (A), Administración (M), Ganancias (E), y Liquidez (L). Sin embargo, en pocas ocasiones se ha pensado en determinar un sistema de indicadores específico de cada tipo de riesgo.

Jaramillo, A., (2002) aplica la metodología CAMEL para analizar la sensibilidad del sector cambiario frente la coyuntura. El procedimiento que sigue consiste en definir los indicadores y ponderar para cada una de las variables CAMEL, y luego, calificar las variables CAMEL en donde el (1) es considerado el mejor puntaje, mientras que (5) el peor. La determinación de rangos para asignar valores entre 1 y 5 se realiza a partir de las desviaciones estándar con respecto a la media de cada indicador del método CAMEL. Finalmente, obtiene el indicador CAMEL final agrupado multiplicando el resultado final de cada variable por su ponderación. Así, propone revisar la evolución de aquellas instituciones que se ubican por encima de (3). Más adelante el autor aplica dicho procedimiento para evaluar la banca colombiana.

Así, una primera etapa de la investigación consistirá en determinar teóricamente, cuáles son los indicadores correspondientes a cada tipo de riesgo financiero y cuáles son las variables macroeconómicas (reales y nominales) cuyo comportamiento puede generar presiones de crisis bancaria de forma más relevante.

Una vez definidas las razones financieras correspondientes a cada tipo de riesgo, se pretende construir un Modelo de Ecuaciones Estructurales (MEE). Esta herramienta permitirá determinar si los indicadores seleccionados para cada tipo de riesgo son los apropiados, si las relaciones establecidas entre los indicadores y los tipos de riesgo son óptimas, y si las relaciones que supondremos existen entre las variables macroeconómicas y los distintos tipos de riesgos e indicadores son las apropiadas. Una vez se tenga la selección óptima de indicadores para cada tipo de Riesgo, se aplicará Control de Calidad Multivariado para evaluar la situación de un grupo de entidades de la banca comercial y universal en el periodo 1996-2004, ante cada tipo de riesgo asociado al riesgo financiero.

La utilidad de esta investigación radicaría, en primer lugar, en conocer la aplicabilidad de la técnica de modelado con ecuaciones estructurales, al medir la eficiencia con que las razones financieras explican los riesgos o aspectos asociados a los riesgos en que puede incurrir la banca. En segundo lugar, conocer la utilidad de las técnicas de control de calidad multivariado al evaluar la situación de los bancos, y finalmente, conocer la situación de los bancos más importantes del país (Comerciales y Universales), en el periodo 1996-2004, ante cada uno de los tipos de riesgo financiero.

### **3.1.1 Modelo ecuaciones estructurales**

El IIES, en su desarrollo del proyecto de gestión de riesgos auspiciado por el BCV, estimó un conjunto de razones financieras para el monitoreo de la banca. Este proceso se realizó revisando las razones financieras publicadas en el boletín trimestral de la Superintendencia de Bancos, estudiando las metodologías de estimación, e incluso revisando las metodologías usadas por las superintendencias de otros países para la construcción de indicadores de alerta temprana. Así, se logró estimar las razones financieras de manera tal que existiera continuidad e invariación en los conceptos y cuentas asociados con las mismas.

Las razones estimadas por el IIES serán utilizadas para identificar y evaluar las razones determinantes de cada tipo de riesgo financiero y determinar las relaciones existentes entre cada tipo de riesgo y las variables macroeconómicas que puedan incidir en ellos.

Para establecer las relaciones estadísticas entre variables y cuantificar la incidencia de unas variables sobre otras, han sido utilizados modelos clásicos como el modelo de regresión lineal, simple o múltiple, y los modelos del análisis de varianza y covarianza. Sin embargo, estos modelos no reconocen la simultaneidad de diferentes relaciones entre las variables, ni la naturaleza no observable o latente que puedan tener. Por consiguiente, se han desarrollado modelos más generales, como el Modelo de Ecuaciones Estructurales, que sí cumple con mencionadas características, y cuya generalidad, así como la existencia de software estadístico para su análisis (LISREL, AMOS, EQS), ha conducido a la proliferación de aplicaciones en ámbitos diversos.

Los modelos ecuaciones estructurales son usados en presencia de variables latentes, es decir, variables no observables, y que por tanto requieren para su medición de una serie de indicadores que capten en su conjunto las distintas facetas de su propia estructura [1]. Los distintos tipos de riesgo financiero (liquidez, crédito, mercado, operacional, legal y reputación) pueden ser considerados variables latentes, ya que no pueden ser explicados por sí solos, sino que necesitan de un conjunto de indicadores para ser medidos.

Los MEE permiten evaluar si los indicadores que se seleccionen para “medir” las variables latentes (endógenas o exógenas), conforman un conjunto de medidas válidas y confiables de dicha variable, y por otro lado, permiten estudiar si las relaciones predictivas y/o causales que se supone existen entre ellas tienen algún apoyo de los datos muestrales.

Este enfoque del MEE es básicamente una herramienta para el contraste de hipótesis. Las hipótesis que se contrastan recogen las relaciones que el investigador supone existen entre las variables y se expresan en términos de un modelo. Por tanto, dos hipótesis pueden modelarse, la que analiza las características de validez y confiabilidad de los indicadores que miden a las variables latentes y la que recoge las relaciones entre las variables latentes. El primero de estos modelos se conoce como Modelo de Medida y el segundo como Modelo de variables latentes. Durán R., Mayorga M. y Montero R., (1999).

Por tanto, los MEE serán usados para determinar si los indicadores que se seleccionen para medir cada tipo de riesgo financiero son medidas válidas y confiables de cada uno de ellos y, por otro lado, determinar la incidencia de las variables macroeconómicas en cada uno de los tipos de riesgo.

En este caso, los tipos de riesgo (el riesgo de liquidez, de crédito, el operacional y el de mercado) se considerarán como variables latentes endógenas y las variables macroeconómicas reales y nominales serán consideradas como dos variables latentes exógenas, entre las cuales se supondrá existen ciertas relaciones. A cada variable latente (endógena o exógena) le corresponderán un conjunto de indicadores y partiremos de la hipótesis de que son los indicadores que realmente las describen.

Así, y a manera de ilustración, el modelo debería tener una forma similar a la siguiente:

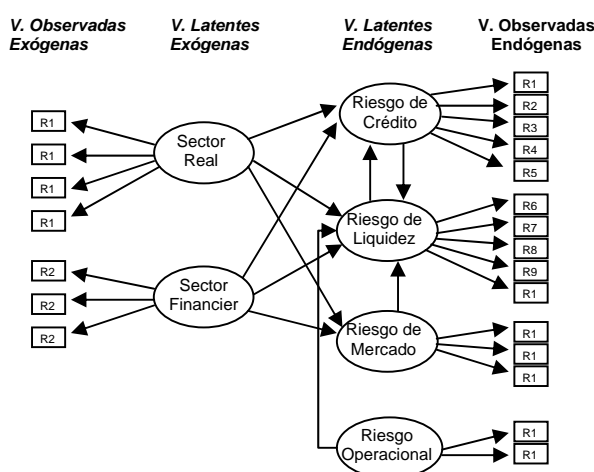


Figura 6. Esquema del MEE

Las variables entre óvalos son las variables latentes. Ellas constituyen el modelo estructural. Las flechas rectas señalarán efectos causales directos que se supondrá existen entre ellas. Las variables en rectángulos representarán los indicadores correspondientes a cada variable latente. Durán R., Mayorga M. y Montero R., (1999).

Serán contruidos cuatro modelos, dos para la Banca Comercial y dos para la Universal, uno para la banca privada y otro para la pública respectivamente, esto debido al comportamiento heterogéneo que tienen la banca comercial y la universal, la privada y la pública. Todas las hipótesis que se construyan serán contrastadas con esta metodología. Además, podrán formularse modelos equivalentes y seleccionar el que muestre una mejor combinación de ajuste y sencillez.

Sin embargo, se han considerado las acotaciones de algunos trabajos dirigidos a la determinación de los puntos críticos que pueden conducir a soluciones finales inadecuadas en el desarrollo de Modelos de Ecuaciones Estructurales. Todos ellos coinciden en establecer unas características generales (tamaño de muestra suficiente, distribuciones normales, valores atípicos, etc.) que permiten la obtención de estimaciones que no se vean afectadas por elementos ajenos a la definición del modelo. Por ello, se ha considerado de vital importancia utilizar técnicas propiciadas por el análisis exploratorio de datos como fase previa a la estimación de los parámetros. Giampaolo Orlandoni, Colmenares Gerardo y Guillén Ruth (2005).

Así mismo, ya que las teorías causales son susceptibles de ser estadísticamente rechazadas si se contradicen con los datos, es decir, con las covarianzas o correlaciones entre las variables, la definición de los indicadores apropiados para explicar cada tipo de riesgo serán determinadas bajo un enfoque teórico, pero también bajo un enfoque estadístico. Por tanto, se aplicarán técnicas de análisis de datos previo a la construcción del modelo.

### 3.1.2 Control de calidad multivariado

Por otro lado, y para darle la utilidad final a los resultados obtenidos, se evaluarán algunas entidades de la banca universal y comercial en el periodo 1996-2004, respecto a los diferentes tipos de riesgo financiero. La evaluación se realizará aplicando técnicas de control de calidad multivariado.

La ventaja de utilizar métodos estadísticos de control multivariado esta en la posibilidad de controlar el efecto de las interacciones entre las diversas variables, así como su variación a lo largo del tiempo, lo que en términos estadísticos significa no obviar la información contenida en las covarianzas y su estabilidad. Dentro de las técnicas de control multivariado caben destacar por su desarrollo y mayor uso tres grupos: métodos basados en distancias ( $T^2$  de Hotelling), métodos MEWMA y MCUSUM. Estas técnicas presentan ventajas entre ellas según su aplicación práctica. Salafranca, Luis (1992).

El  $T^2$  de Hotelling permitirá detectar grandes cambios en la media del proceso, pero cambios inferiores a dos desviaciones estándar normalmente no serán detectados. Esto se debe a que el  $T^2$  de Hotelling no considera la información pasada del proceso. Sin embargo, los gráficos MCUSUM y EWMA sí consideran la información pasada, razón por la cual son más susceptibles a detectar pequeños cambios. Sin embargo, el  $T^2$  de Hotelling será más efectivo en la detección de grandes cambios en la media del proceso que las otras dos metodologías.

El cumplimiento de las hipótesis de partida, normalidad e independencia de las observaciones, representa una limitante en la aplicación de estos métodos. Sin embargo, diversos estudios han confirmado que la no existencia de normalidad no produce grandes alteraciones en el modelo, pero la existencia de correlación sí es imprescindible tomarla en cuenta. En este sentido el gráfico MEWMA tiene ventajas sobre el gráfico  $T^2$  de Hotelling, ya que el cálculo de los valores EWMA se puede considerar como un proceso auto regresivo que tiende a disminuir e incluso eliminar el efecto que la correlación en las observaciones (autocorrelación) tiene sobre el gráfico de control. Salafranca, Luis (1992).

Por tanto, la elección de una de las metodologías dependerá del comportamiento de los datos y las necesidades que se tengan de detectar grandes o pequeños cambios en la media del proceso.

La idea es observar el comportamiento conjunto de los indicadores para evaluar la situación de un banco, en distintos instantes del tiempo, respecto a un riesgo particular, estudiando las señales del proceso mediante gráficos de control y estudiando las causas que provocaron, en dado caso, una salida de los límites.

Así, el control de calidad implicará la evaluación del proceso continuo a través de indicadores de alerta temprana, determinados mediante MEE, para cada tipo de Riesgo Financiero.

### 3.1.3 Variables

Como ya se ha dicho, se pretende construir un Modelo de Ecuaciones Estructurales que contribuya a determinar cuáles son los indicadores que explican cada uno de los tipos de riesgo financiero: riesgo de crédito, riesgo de liquidez, riesgo de mercado, riesgo operacional y riesgo legal. Además, se busca establecer las relaciones que existen entre los tipos de riesgo, y las relaciones existentes entre ellos y las principales variables macroeconómicas.

Tal como se especificó en el punto anterior, cada tipo de riesgo representa una variable no observable endógena, cada uno de las cuales será explicada a través de un conjunto de indicadores. Cada riesgo con sus respectivos indicadores representan el entorno microeconómico del modelo. Los indicadores considerados responden a los aspectos analizados por el método CAMEL. El listado de indicadores con que serán asignados a cada tipo de riesgo se presenta a continuación. Navarro, José L (2001):

- 1.- Patrimonio/Activo Total
- 2.- (Patrimonio + Gestión Operativa)/Activo Total
- 3.- Otros Activos/Patrimonio
- 4.- Activos Improductivos / Patrimonio + Gestión Operativa
- 5.- Provisión de Cartera de Crédito/Cartera Inmovilizada Bruta.
- 6.- Provisión de Cartera de Crédito /Cartera de Crédito Bruta.
- 7.- Otros Activos/Activo Total
- 8.- Activo Improductivo/Activo Total
- 9.- Cartera Inmovilizada Bruta / Cartera de Crédito Bruta

- 10.- (Disponibilidad-Rendimiento por cobrar por disponibilidades)/(Captaciones del público-Gastos por pagar).
- 11.- A/ Captaciones del público - Gastos por pagar.
- 12.- Cartera crédito neta/Captaciones del público – gastos por pagar.
- 13.- Cartera vigente bruta/Captaciones totales
- 14.- Cartera de crédito neta/captaciones totales
- 15.- Cartera de inversiones/captaciones totales
- 16.- Gastos de transformación/activo promedio
- 17.- Gastos de personal/captaciones del público promedio.
- 18.- Gastos operativos/margen financiero bruto
- 19.- Margen de intermediación/Ingresos Financieros
- 20.- Ingresos x cartera de crédito/cartera de crédito neta promedio.
- 21.- Ingresos x cartera de crédito/cartera de crédito bruta promedio.
- 22.- Margen financiero bruto/activo total promedio
- 23.- Ingreso financiero/margen financiero

En cuanto a las variables macroeconómicas a incluir en el Modelo, éstas serán separadas en nominales y Reales. El listado de variables macroeconómicas a considerar se muestra a continuación Navarro, José L. (2001): Tasa de Interés, Producto interno Bruto, Reservas Internacionales, Inflación, Base Monetaria. Este modelo será construido considerando los bancos más importantes del país para el presente año, separando Banca Universal y Banca Comercial en entidades privadas y públicas. Los mismos serán también usados para realizar la evaluación mediante control estadístico de calidad.

Tabla 1. Banca Universal pública y privada considerar en el estudio

BANCOS UNIVERSALES	
Bancos Privados	Bancos Públicos
Banco de Venezuela	Banfoandes
Banco Provincial	
Banco Mercantil	
Banesco	
Citibank	
Banco Exterior	
Banco del Caribe	

Tabla 2. Banca Comercial pública y privada considerar en el estudio

BANCOS COMERCIALES	
Bancos Privados	Bancos Públicos
Banco Federal	Banco Industrial de Venezuela
Banco de Guayana	
Banco Confederado	
Banco Plaza	

Los bancos, las razones financieras seleccionadas para cada tipo de riesgo y las variables macroeconómicas nominales y reales, todas establecidas en forma mensual para el periodo 1996-2004, constituirán la base de datos a utilizar en el presente estudio.

#### 4. ANÁLISIS DE SUPERVIVENCIA APLICADO A LA BANCA COMERCIAL VENEZOLANA. 1996 – 2004.

*María Alejandra Ayala, Rafael Eduardo Borges*

La discusión sobre la situación y perspectivas de la economía internacional se ha visto fuertemente influenciada por una serie de crisis cambiarias que han afectado diversos países. Particularmente, en Latinoamérica las crisis bancarias han alcanzado a México, Brasil, Chile, Colombia, Venezuela y Argentina. Así como también han ocurrido en países industrializados como Japón y Estados Unidos.

Estas crisis se originaron primordialmente ante la falta de evaluación de las condiciones financieras de los bancos y la débil legislación en materia bancaria. A partir de los años 70 se introducen modelos para la asignación eficiente de los recursos de supervisión. Dichos modelos permiten entre otras cosas observar la evolución de una entidad en el lapso de tiempo de dos inspecciones en el sitio, determinar los principales factores de riesgo que asume la banca y realizar una mejor medición de ellos. Estos modelos se conocen como “*modelos de alerta temprana para el sistema financiero*”, y permiten predecir la probabilidad de cambios en las condiciones de solvencia y liquidez de los bancos.

En el presente trabajo se propone la utilización del Análisis de Supervivencia, como técnica estadística para investigar *cuándo* es probable que una entidad financiera cambie de estado. La variable aleatoria de interés, es el tiempo que tarda la entidad en cambiar de estado. Específicamente, interesa el tiempo que tarda en ocurrir el cambio de estado y cuales son las variables que más influyen en el cumplimiento del cambio de estado.

Se lleva a cabo un análisis de supervivencia clásico, mediante el método de Kaplan y Meier y ajustado por el modelo de Cox, utilizando los datos de la banca comercial venezolana.

En el período 1996-2004, se observa la fusión bancaria como el evento de interés para el cambio de estado. Como resultado se logra identificar un grupo de cuatro razones financieras de riesgo de un grupo original de 15, que son referidas históricamente en la revisión realizada como determinantes de los modelos de alerta temprana.

#### **4.1 Modelos de análisis de supervivencia en la gestión del riesgo bancario**

La utilización de los modelos de supervivencia en el área financiera es reciente, particularmente González-Hermosillo, Pazarbasioglu and Billings (1996), Cole and Gunther (1995), Marcelo Dabós (2000) entre otros, han comunicado resultados para la predicción de crisis bancarias utilizando esta técnica. En el caso venezolano García Herrero (2003) empleó la metodología de supervivencia en dirección al estudio de las variables determinantes de la crisis bancaria en Venezuela a mediados de los 90.

En el año 1994, el sistema financiero Venezolano experimento una crisis bancaria producto de la intervención del Banco Latino, por parte de la Superintendencia de Bancos. Esta intervención, marco el inicio de una ola de cierres de entidades bancarias. Así en Junio del 2004, 7 instituciones fueron cerradas<sup>10</sup>, en agosto del mismo año, El Banco Venezuela y el Banco Consolidado fueron intervenidos, posteriormente entre diciembre de 1994 y febrero de 1995, 5 instituciones<sup>11</sup> más son intervenidas.

Producto de esta crisis, se propuso una nueva legislación bancaria, acorde con la situación que vivía el país. Un resultado fundamental de la transformación de la ley general de bancos fue la inclusión de las normas sobre regulación financiera y la ley de Fusiones Bancarias. A finales de julio del año 1999 se aprobó el "Proyecto de ley de estímulo a la competitividad, el fortalecimiento patrimonial y la reactivación de los gastos de transformación en el sector bancario", mejor conocido como Ley de Fusiones Bancarias (Medina 2005). Esta ley induce a las instituciones financieras a emprender fusiones y adquisiciones con un propósito fundamental, reducir la gran cantidad de entidades bancarias y hacer que el sistema financiero se fortalezca financieramente.

Los modelos de alerta temprana para el sistema financiero, permiten predecir la probabilidad de cambios en las condiciones de solvencia y liquidez de los bancos. Lo importante es que permiten realizar un análisis más profundo de los factores relacionados con quiebras de las instituciones financieras de tal manera que se pueda dirigir y regular a los bancos más eficientemente. Por otro lado, la habilidad para diferenciar entre los bancos solventes y aquellos con problemas permite reducir el costo esperado de las crisis bancarias. Si los problemas pueden ser detectados con prontitud, pueden tomarse acciones preventivas que permitan minimizar los costos de un eventual cierre de la institución. Diferentes han sido las técnicas estadísticas que las entidades bancarias han utilizado para describir su comportamiento, más importante aún, para intentar identificar situaciones irregulares que pueden desembocar en crisis financieras. Aunque los modelos *probit* han demostrado buen desempeño en este

---

<sup>10</sup> La Guaira, Maracaibo, Amazonas, Bancor, Barinas, Construcción, y Metropolitano

<sup>11</sup> Progreso, República, Italo Venezolano, Principal y Profesional



sentido, estos modelos, sólo permiten estimar la probabilidad de que un banco cambie de “estado”<sup>12</sup>, pero no informan acerca del tiempo que las entidades vulnerables podrían demorar en demostrar problemas.

El objetivo general de este trabajo es investigar *cuándo* es probable que la entidad financiera cambie de estado. Dicho de otra forma, la variable aleatoria de interés, es el tiempo que tarda en ocurrir el cambio de estado y cuales son las variables que más influyen en el cumplimiento del cambio de estado.

La técnica que permite describir el comportamiento de datos que corresponden al tiempo o duración desde un origen bien definido hasta la ocurrencia de un cambio de estado o punto final se denomina “*análisis de supervivencia*”. Klein – Moeschberger (1997).

#### 4.2 Análisis de Supervivencia.

El concepto central de un modelo de supervivencia no es la probabilidad de que un cambio de estado ocurra (por ejemplo, probabilidad que un banco se fusione), sino más bien la probabilidad condicional de que ocurra un cambio de estado, dado que tenía en el tiempo anterior otro estado, (por ejemplo, que un banco se fusione dado que en el periodo anterior no lo había hecho). Este tipo de análisis permite además incluir factores explicativos constantes y variables en el tiempo. Otra ventaja de esta técnica es la existencia de censura en los datos. La censura ocurre cuando el resultado o evento de interés (cambio de estado) no se observa para todos los individuos dentro del periodo en que se realiza la recolección de los datos. Por lo tanto, muchas de las observaciones representan la duración registrada hasta el momento de la medición y no el lapso de tiempo transcurrido hasta la ocurrencia del evento. Para el caso particular del estudio del tiempo que podría tardar una entidad financiera en cambiar de estado, el tipo de censura que se presenta, se conoce como, censura por la derecha.

El análisis de supervivencia tiene por objetivo encontrar la función de riesgo, esta, describe el riesgo de cambio de estado en diferentes periodos de tiempo y representa una secuencia de probabilidades condicionales:  $f(t) = P(\text{banco se fusione en el momento } t \text{ dado que en } t-1 \text{ no lo había hecho})$ .

Si  $T$  es una variable aleatoria discreta no negativa que representa el tiempo de vida de un banco<sup>13</sup>,  $T$  puede tomar los valores  $0 \leq t_0 \leq t_1 \leq t_2 \leq \dots$ , entonces la función de probabilidad asociada a la variable aleatoria  $T$  está dada por

$$f(t_i) = P(T = t_i) \text{ para } i = 1, 2, 3, \dots$$

Luego, la función de supervivencia se define como:

$$S(t) = \sum_{j:t_j \geq t} f(t_j)$$

y representa la probabilidad de que  $T$  sea mayor o igual a un valor  $t$ , es decir, la probabilidad de que la supervivencia del banco sea  $T \geq t$ . En este punto debe aclararse, que si no existiese censura, la estimación de esta función sería simplemente contar cuantas entidades financieras llegaron hasta el momento  $t_j$  respecto al total.

La función de riesgo se define como:

$$\lambda_j = P(T = t_j | T \geq t_j) = \frac{f(t_j)}{S(t_j)}$$

Esta función representa la probabilidad de que una entidad financiera cambie de estado en el momento  $t_j$  dado que no lo había hecho. Una forma alternativa de escribir la función de supervivencia es:

$$1 - \lambda_j = 1 - \frac{f(t_j)}{S(t_j)}$$

de donde

<sup>12</sup> En el presente trabajo, un cambio de estado se define como la fusión de una entidad financiera, puesto que a partir de la crisis de 1994 la nueva Ley de Bancos promueve la fusión para evitar las quiebras.

<sup>13</sup> Consideraremos en este caso que un Banco vive mientras no se fusione.

$$S(t_j) = \prod_{i=0}^{j-1} (1 - \lambda_j).$$

Una estimación no paramétrica de la función de supervivencia es:

$$\hat{S}(t_j) = \prod_{i=0}^{j-1} (1 - \hat{\lambda}_j)$$

$\hat{\lambda}_j$  puede obtenerse como:  $\hat{\lambda}_j = \frac{d_j}{n_j}$ , donde  $d_j$  representa el número de individuos o elementos que registraron el evento de interés en el momento  $j$ . y  $n_j$  representa el número total de elementos presentes en el tiempo  $j$ .  $n_j$  se conoce como, el conjunto de riesgo, o simplemente el número al riesgo

(Hosmer et al., 1999), y a  $\hat{\lambda}_j = \frac{d_j}{n_j}$  se le conoce como el estimador no paramétrico de Kaplan-Meier o

también estimador del producto límite (Hosmer et al., 1999). Este estimador incorpora información de todas las observaciones disponibles, sean o no censuradas. Además, es una función que depende solo de los valores de la muestra y permite describir la relación existente entre la función de riesgo y el tiempo de supervivencia hasta que ocurre el evento de interés. La validez del estimador de Kaplan-Meier descansa en los supuestos de que los individuos que se retiran del estudio tienen un destino semejante al de los individuos que permanecen, y que el tiempo durante el cual un individuo entra al estudio no tiene efecto independiente en la respuesta.

Una vez que se tiene la función de supervivencia, es de interés investigar si el riesgo  $\lambda$  difiere sistemáticamente entre los individuos, esto significa, identificar variables explicativas de la heterogeneidad observada en función del riesgo. La forma de estudiar la heterogeneidad observada es introduciendo en el modelo  $p$  variables explicativas  $Z_p$  que caracterizan a un individuo o su entorno.

Se define entonces un vector de variables  $\mathbf{Z}_{ij} = [\mathbf{z}_{1ij}, \mathbf{z}_{2ij}, \dots, \mathbf{z}_{pij}]$  donde cada elemento del vector representa la característica  $p$  para el individuo  $i$ -ésimo en el momento  $j$ . En este trabajo se utiliza el modelo de riesgo proporcional de Cox para la incorporación de variables explicativas.

#### 4.2.1 Modelo de Riesgo proporcional propuesto por Cox<sup>14</sup>.

En 1972, Cox introduce el modelo de regresión más utilizado en análisis de supervivencia, este modelo puede escribirse mediante:

$$\lambda_i(t | z) = e^{(\beta_1 z_{1i} + \beta_2 z_{2i} + \dots + \beta_p z_{pi})} \lambda_0(t)$$

donde  $Z_i(t)$  es el vector de covariables para el  $i$ -ésimo individuo en el tiempo  $t$ . Este modelo incluye una parte paramétrica  $e^{\beta Z_i(t)}$ , llamada puntaje de riesgo y otra parte no paramétrica  $\lambda_0(t)$ , llamada función de riesgo base.

Para dos individuos con variables explicativas  $\mathbf{Z}$  y  $\mathbf{Z}^*$ , el cociente de las respectivas funciones de riesgo viene dado por (Hosmer et al., 1999):

---

<sup>14</sup> Propuesto por Cox en 1972

$$\frac{\lambda(t|z)}{\lambda(t|z^*)} = \frac{e^{\left[\sum_{p=1}^p \beta_p z_p\right]} \lambda_0(t)}{e^{\left[\sum_{p=1}^p \beta_p z_p^*\right]} \lambda_0(t)} = e^{\left[\sum_{p=1}^p \beta_p (z_p - z_p^*)\right]}$$

Esta razón se conoce como riesgo relativo, es constante en el tiempo y las tasas de riesgo son proporcionales. La interpretación de los coeficientes está dada por la siguiente derivada:

$$\frac{\partial \ln \lambda(t|z)}{\partial z_p} = \beta_p$$

$\beta_p$  representa el cambio proporcional en la función de riesgo que resulta de un cambio marginal en la p-ésima variable explicativa. Si  $\mathbf{Z}$  y  $\mathbf{Z}^*$  difieren en la p-ésima variable explicativa, la cual es una variable binaria, se tiene (Hosmer et al., 1999):

$$\frac{\lambda(t|z)}{\lambda(t|z^*)} = e^{\beta_p}$$

Una de las principales hipótesis del modelo de Cox es precisamente que la función de riesgo es proporcional dados dos perfiles de cambio de estado distintos, y por tanto se debe mantener a lo largo del tiempo. Los supuestos del modelo de riesgo proporcional son cuatro: a) no existe influencia de las entidades financieras en la estimación del modelo, b) no existe influencia de las entidades financieras en la estimación de cada parámetro del modelo, c) se supone inexistencia de heterogeneidad no observada<sup>15</sup>, d) se supone adecuación de la forma funcional.

#### 4.3 Análisis de supervivencia Aplicado a la Banca Comercial Venezolana.

A partir del año 1996, la nueva ley de bancos incentiva las fusiones, para evitar la quiebra de las entidades financieras. En este trabajo se considera el evento de interés a la fusión del banco. Dicho en otras palabras, el banco muere cuando se fusiona de lo contrario sobrevive.

Se analizaron 34 entidades bancarias de tipo comercial, entre enero de 1996 y diciembre de 2004, estableciendo 108 meses de observación. Como variables independientes, se utilizaron 15 razones financieras que describen el comportamiento de las entidades bancarias<sup>16</sup>

Se estimó la Función de Supervivencia utilizando el estimador de Kaplan y Meier, observándose algunas características resaltantes que contribuyeron al análisis conclusivo y reflejan parte del fenómeno histórico en la banca. Posteriormente, se estimó el modelo de Cox y resultaron varias razones financieras estadísticamente significativas. Estas razones han sido reconocidas como influyentes en la detección de riesgo en parte de la literatura disponible de alerta temprana. Finalmente se comprobó estimación de los coeficientes del modelo y gráficamente, la tendencia del ajuste.

<sup>15</sup> toda la heterogeneidad presente en las entidades financieras es recogida en las variables explicativas.

<sup>16</sup> Estas razones corresponden al análisis de los datos realizados por Guillen Ruth, Gerardo Colmenares y Giampaolo Orlandony en el trabajo Algoritmo Seguido para el Pre-procesamiento de datos de la Banca y Construcción de Modelos no Lineales. Presentado en BCV en Septiembre del 2005.

Tabla 2. Razones financieras utilizadas

Código	Razón Financiera
R1	Patrimonio/ Activo Total
R2	Patrimonio + Gestión Operativa) / Activo Total
R3	Otros Activos / Patrimonio
R4	Activos Improductivos Brutos / (Patrimonio + gestión operativa)
R5	Provisión de cartera de crédito / Cartera inmovilizada bruta
R6	Provisión de cartera de crédito / Cartera crédito bruta
R7	Otros activos / Activo total
R8	Activo improductivo / activo total
R9	Cartera inmovilizada bruta / cartera de crédito bruta
R10	(Disponibilidades - Rendimiento por cobrar por disponibilidades) / (Captaciones del público - gastos por pagar)
R11	(Disponibilidades - Rendimiento por cobrar por disponibilidades+ inversiones temporales (títulos valores) -rendimiento por cobrar por inversiones temporales) / (Captaciones del público - gastos por pagar)
R12	Cartera de crédito neta / (Captaciones del público - gastos por pagar)
R13	Cartera vigente bruta / Captaciones totales
R14	Cartera crédito neta / Captaciones totales
R15	Cartera de inversión (en títulos valores y en sucursales)/ captaciones totales.

El Análisis de Supervivencia ofrece una metodología robusta, para la detección de variables indicadoras de riesgo en las instituciones bancarias. En este trabajo, cuatro razones financieras fueron identificadas como razones de riesgo a través del modelo de regresión de Cox. La verificación de los supuestos del modelo de Cox se realizó, sin encontrar violaciones de dichos supuestos, lo que garantiza la representatividad de los resultados.

## 5. USO DE LAS TÉCNICAS DE PREPROCESAMIENTO DE DATOS E INTELIGENCIA ARTIFICIAL (LÓGICA DIFUSA) EN LA CLASIFICACIÓN DEL RIESGO BANCARIO. CASO DE ESTUDIO: LA BANCA COMERCIAL Y UNIVERSAL. *Carlos A Martínez Ortiz, Gerardo Colmenares*

El Comité de Supervisión Bancaria de Basilea sugirió un nuevo marco para la regulación y control del sistema Bancario (BASILEA II). El Banco Central de Venezuela, circunscrito a esta meta ha marcado como objetivo la formulación de un modelo en el cual logre la integración de todos los bancos a este acuerdo e para finales de 2007. Este acuerdo consta de tres pilares: Capital mínimo exigible, Examen por parte del Supervisor y Disciplina del mercado. Guillén Ruth (2005)

Por otro lado, actualmente existe un conjunto de factores que favorecen la orientación actual de las entidades que les permite emplear métodos cuantitativos que contribuyen en la medición del riesgo bancario. El actual entorno económico en el país, caracterizado por una mayor incertidumbre y volatilidad en los mercados, propicia una mejora en los sistemas de medición y control de los riesgos. Otro de las razones que motiva la realización de este proyecto en particular, es la construcción de un modelo que permita la comparación y evaluación de los resultados obtenidos con otros modelos prototipos, y así lograr evaluar su desempeño y reforzar la identificación, medición y control de las variables que afectan el Riesgo Bancario.

En este proyecto se estudiará la construcción del modelo clasificador/predicador del riesgo bancario mediante lógica difusa: El primer paso consistirá en la recopilación de los datos; gran parte de la información es proporcionada por la Superintendencia de Bancos (SUDEBAN) en formato electrónico, sin embargo, para el período de estudio no está completa y en algunos casos sus estructuras difieren. De este modo, en el IIES esta disponible una base de datos de estructura homogénea, la cual contiene de las instituciones financieras, los movimientos suscitados en sus Balances de Publicación y Estados de Resultados desde junio de 1996 hasta diciembre del 2004. Esta disposición de datos se

encuentra a períodos semestrales. Por otro lado, el IIES construyó a partir de la fuente de SUDEBAN otra base de datos de periodicidad mensual en el mismo período. Estos datos representan una mayor continuidad en el tiempo. Sin embargo, los resultados conseguidos por los modelos al aplicar ambas bases de datos serán comparados para determinar la consistencia de los datos construidos y la confiabilidad de los valores de pronóstico. Giampaolo Orlandoni, Colmenares Gerardo y Guillén Ruth (2005)

### 5.1 Aplicación de la lógica difusa al Riesgo Bancario

Para la construcción del modelo, estos datos serán convertidos en un conjunto de variables difusas, a este paso se le conoce como clasificación difusa “Fuzzificación”. Esto se hace asignando valores (que serán los valores difusos), a partir de un conjunto de funciones de membresía o pertenencia. Es decir, un clasificador difuso divide los posibles rangos en los cuales puede clasificarse los valores de entrada. Luego, se crean un conjunto de reglas difusas, con la ayuda de los expertos, lo cual generará un nuevo conjunto de variables.

El último paso en la construcción de un sistema de lógica difusa es convertir de nuevo las variables difusas generadas por la base de reglas en valores con interpretación real. El proceso se denomina “Desfuzzificación”, el cual se encarga de obtener un solo valor de salida (expresado en lógica convencional o no difusa).

Una representación esquemática de lo anterior es como sigue:

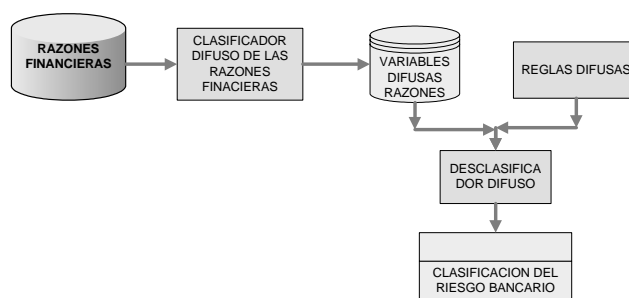


Figura 7. Esquema metodológico del modelo de lógica difusa

#### 5.1.1 Utilidad de la lógica difusa en el riesgo bancario

Hasta el siglo XX, la lógica clásica trabajaba con dos valores: Verdadero y Falso, no obstante durante el último siglo se empezaron a crear nuevas lógicas que incorporaban más valores de certeza, entre ellas la lógica difusa.

Esta rama de la Inteligencia Artificial, utiliza expresiones que no son ni totalmente ciertas ni completamente falsas. Por consiguiente, es la lógica aplicada a conceptos que pueden tomar un valor cualquiera de veracidad dentro de un conjunto de valores que oscilan entre dos extremos: la verdad absoluta y la falsedad total.

El adjetivo difuso, aplicado a estas lógicas, se debe a que en ellas los valores de verdad no-deterministas utilizados tienen, por lo general, una connotación de incertidumbre. Ejemplo: Cuando se dice que un vaso está medio lleno o medio vacío es un elemento de incertidumbre, es decir, de difusidad.

Las características más atractivas que justifican el uso de la lógica difusa son: flexibilidad, tolerancia con la imprecisión, capacidad para modelar problemas no-lineales, base en el lenguaje natural, para la toma de decisiones en presencia de datos o conocimientos inciertos, reconocimientos de patrones ambiguos y como un componente de sistemas expertos difusos.

### 5.2 Diseño Experimental

- **Identificación de las variables de entrada y salida.** Se utilizarán como variables de entrada las 23 razones financieras previamente seleccionadas mediante ACP, provenientes de los Balances de

Publicación y Estado de Resultados de los 18 bancos considerados (comerciales y universales). Estas razones serán tomadas de forma semestral desde el periodo B96 (segundo período del año) hasta el periodo A04. De igual manera, se construirá el modelo donde los valores de entrada serán las razones financieras de forma mensual desde enero del 1996 hasta diciembre del 2004. Como variable de salida tendremos el nivel de riesgo, calculado como una salida cualitativa.

- **Identificación de los conjuntos difusos.** En este paso se determina la función característica asociada a los conjuntos de entrada y salida; bien sea, a través del conocimiento de los expertos o utilizando una colección de datos para construir la misma función de acuerdo a la distribución de los datos. Entre las funciones características más conocidas están: Triangular, Trapezoidal, Gaussiana, Sigmoidal, etc.
- **Selección del método de clasificación difusa o difusificación.** En este paso se transforman las razones financieras en variables difusas. Esto se realiza asignando grados de membresía o pertenencia con respecto a los conjuntos difusos, a las variables de entrada (razones financieras). Este clasificador difuso dividirá los rangos posibles en los cuales pueden clasificarse los valores de las razones financieras:  $x_1, \dots, x_{23}$ . En este caso se hará una fuzzificación de cinco niveles (Alto, Medio Alto, Mediano, Bajo, Muy Bajo).
- **Creación de la base de Conocimiento Experto utilizando reglas del tipo sí \_ entonces.** En esta parte se creará un conjunto de reglas con la ayuda de los expertos que se ajusten a las características del problema.
- **Diseño del mecanismo de inferencia.** En esta parte se realiza la traducción matemática de las reglas difusas que relacionan los conjuntos difusos de entrada y salida y que representan a las reglas que definen el sistema. En este proceso las entradas son conjuntos difusos (grados de pertenencia) y las salidas son también conjuntos difusos asociados a la variable de salida.
- **Aplicación del método de desdifusificación.** Finalmente, del conjunto difuso obtenido del paso anterior y mediante métodos matemáticos de desdifusión (método del máximo, del centroide, o el método de la altura), se obtiene el valor concreto de la variable de salida, es decir el resultado, en nuestro caso el nivel de riesgo al que se expone el banco (Alto, Medio Alto, Medio, Bajo, Muy bajo).

## **6. ESTUDIO E IMPLEMENTACIÓN DE HERRAMIENTAS DE ANÁLISIS DE DATOS PARA VALORES ATÍPICOS E IMPUTACIÓN DE DATOS FALTANTES APLICADO A LA BANCA NACIONAL.** *Daniel E. Paredes Moreno, Gerardo Colmenares.*

En la actualidad, la investigación científica genera gran cantidad de datos en todas sus actividades. Estos datos guardan información, y para acceder a esta se necesita hacer uso de técnicas de análisis que permitan visualizar relaciones, patrones o comportamientos que no pueden ser apreciados a primera vista.

Este es el caso del Análisis Exploratorio de Datos, cuya filosofía consiste en describir las características estructurales de los datos sin involucrar Modelos Estadísticos o la Inferencia Clásica. Su propósito es el de examinar los datos, previamente a cualquier tratamiento formal, para detectar posibles anomalías que pudieran llevar a conclusiones equivocadas.

La necesidad de información y conocimiento de nuestros días, implica la recopilación de grandes cantidades de datos, por lo que se hace necesario tener herramientas que nos permitan lidiar con grandes volúmenes de datos. Es aquí cuando las técnicas que conforman el Análisis Exploratorio de Datos son útiles, pues permiten reducir la dimensionalidad del problema y facilita la obtención de información y la interpretación que le podamos dar a la misma.

No importa que técnica o método de análisis se utilice, la calidad de los resultados y conclusiones que se obtengan dependerá de la calidad de los datos con los que se haya trabajado. En este sentido, aun en los datos recolectados con mayor cuidado, algunas veces pueden presentarse observaciones que parecieran no pertenecer al conjunto de datos. En este caso se puede estar tratando con observaciones

Espurias o Atípicas. También, se presenta el caso donde es imposible recopilar la información completa para todos los individuos o para todos los períodos de tiempo a los que se quiere referir la investigación o para todas las variables en consideración; es decir, no se tiene ningún valor para esa observación.

Se hace necesario disponer de herramientas y técnicas que nos permitan revisar, describir y explorar la matriz de datos para detectar y solucionar este tipo de problemas, y de esta manera evitar que las conclusiones obtenidas sean distorsionadas o inválidas. Existen herramientas de Análisis de Datos que pueden aplicarse como un preprocesamiento a los datos para identificar Valores Atípicos y completar Datos Faltantes. En muchos casos, los Valores Atípicos pueden deberse a condiciones especiales del problema en consideración, o a errores por lo que deben recibir un tratamiento diferente. Desconocer la existencia de Valores Atípicos puede llevar a conclusiones que se alejan de la realidad.

Cuando no se dispone de la información completa para cualquiera de las variables involucradas, se presenta otro tipo de problema, pues, dependiendo de la metodología a utilizar para el análisis, en muchos casos no es posible procesar o analizar los datos y la carencia de esa información, impide realizar un análisis completo del problema.

En el caso específico de la Banca Comercial Venezolana, el preprocesamiento de los datos se hace necesario para adecuarlos a análisis posteriores donde los Valores Atípicos deben ser tratados con especial atención y los Valores Faltantes representan un serio problema. Por esta razón, se desea estudiar las herramientas de Análisis de Datos para realizar la detección de Valores Atípicos y la Imputación de Datos Faltantes para tener un conjunto de métodos y técnicas que nos permitan enfrentar estos problemas, establecer comparaciones entre estas técnicas, evaluar su aplicabilidad en cada caso e implementarlas para resolver estos problemas.

#### **6.1 La aplicación estadística en la gestión de riesgo bancario.**

El objetivo principal consiste en evaluar e implementar las Técnicas de Análisis de Datos que permiten tratar con Valores Atípicos y Datos Faltantes observados en los Balances de Comprobación y Estados Financieros de la Banca Comercial Venezolana en el período 1994-2004. Este proyecto está siendo desarrollado actualmente por el Instituto de Investigaciones Económicas y Sociales (IIES). Se calcularon un conjunto de razones financieras a partir de los Balances de Comprobación y Estados de Resultados que se recopilaron utilizando la información suministrada por la Superintendencia de Bancos (SUDEBAN) para los bancos comerciales y universales del país.

Las razones calculadas obedecen a la metodología utilizada por SUDEBAN para la administración de riesgos publicada en su Boletín Trimestral. Estas razones serán utilizadas para evaluar el riesgo financiero de los bancos: riesgo de crédito, de mercado, operacional, de liquidez, legal. Las razones calculadas comprenden un índice específico para evaluar aspectos de uno de los tipos de riesgos mencionados anteriormente como sugiere el método CAMEL, metodología para evaluar aspectos relevantes en la sanidad bancaria.

Figura 8. Esquema de AED

Dichas razones se calcularon para un conjunto de bancos comerciales y universales bajo las modalidades mensuales entre los años 1996 y 2004 y semestrales desde el semestre B-1996 hasta el A-2004. La disponibilidad de los datos semestrales representa mayor confianza en su información debido a que proviene de cifras confiables y auditadas. Sin embargo, la construcción de los datos mensuales permite a los modelos disponer de una periodicidad mayor de los datos y en consecuencia, una mayor capacidad para poder crear conjuntos de entrenamiento, pruebas y generalización de los modelos construidos. Se pueden presentar asimetrías en las observaciones que se correspondan a un mismo sujeto observado (banco) y la garantía de contar con una alta consistencia de los datos es a mediante un Análisis Exploratorio de los datos para cada uno de los fenómenos estadísticos que puedan ocurrir: series incompletas, valores extremos incongruentes, colinealidad entre las razones construidas, incoherencia entre observaciones construidas y publicadas, etc. Díaz M. Luís. (2002)

Utilizando técnicas de Análisis de Datos se efectuará un Análisis Exploratorio con el fin de obtener una descripción del conjunto de datos y observar el comportamiento de estos durante el periodo en estudio y para cada una de las razones financieras.

El pre-procesamiento de los datos deberá comenzar con:

- La depuración de la Base de datos para observar posibles inconsistencias o errores en su formación. Se realizará comparando con la información existente, y con las fuentes originales de la información, como son los Balances de Comprobación y Estados de Resultados y revisando la metodología utilizada, además se observará el comportamiento de las razones financieras.
- Se realizará un Análisis Exploratorio de los Datos utilizando técnicas de Análisis de Datos para observar: a) el comportamiento de las razones financieras en cuanto a su estructura o distribución respecto a los bancos para cada periodo, b) el comportamiento de las razones financieras a través del tiempo para cada banco, c) el comportamiento de los bancos respecto al conjunto de razones financieras que sirven de indicadores para cada tipo de riesgo en particular en los distintos períodos y d) el grado de asociación que tienen las razones financieras entre sí. Esto permitirá adecuar los datos para un posterior análisis, reconociendo patrones y comportamientos que nos permitan describir mejor la realidad del problema y hacer sugerencias acerca de los métodos de análisis que se usarán posteriormente. Johnson Richard y Wichern Dean (1998)
- Utilizar Técnicas para Detección de Valores Atípicos que permitan reconocer observaciones que tienen un comportamiento distinto al del resto e inclusive distinto al que se esperaba a nivel teórico. Además, se utilizarán técnicas que permitan analizar de manera adecuada este tipo de observaciones ante las condiciones particulares de nuestro problema. La detección y posterior tratamiento de este tipo de datos es importante al momento de realizar el análisis de riesgo financiero. López M. Thomas A. Julio (2001)
- Aplicación de Técnicas de Imputación de Datos Faltantes debido a que la información no se dispone en su totalidad ni para todos los bancos, ni para todas las razones ni en todos los periodos de tiempo debido al comportamiento que ha tenido la banca comercial y universal en nuestro país y a las escasas recopilación de información en el pasado y aún hoy en día, esta es un obstáculo importante pues no se cuenta con la información completa para estudiar en plenitud en problema de la gestión de riesgo financiero. Se compararán las técnicas de Imputación de Datos para utilizar la que mas se adecue a escenario de la banca comercial y universal que permita completar la base de datos y que los valores imputados no representen ninguna distorsión de la realidad de los bancos en análisis posteriores.

## 6.2 Resultados esperados

Realizadas las pruebas a las técnicas e implementadas a los datos de Balances de Comprobación y Estados de Resultados de la Banca Comercial Venezolana se espera:

- Tener una recopilación de herramientas y técnicas que permitan resolver el problema de pre-procesamiento de datos (detección de valores atípicos e imputación de datos faltantes) en distintos casos.
- Tener un criterio de selección de las técnicas más apropiadas evaluando la aplicabilidad de la técnica en cada caso y la necesidad a resolver.
- Resolver el problema de Pre-procesamiento de los datos de la Banca Comercial Venezolana en miras al análisis del que será objeto.



## Referencias.

- Agresti, A. (1990). Categorical Data Analysis. John Wiley and Sons Ltd., New York.
- Allison, P. D. (1982). Discrete time methods for the analysis of event histories. Sociological Methodology. Bass Publishers, San Francisco.
- Amir F. Atiya, Senior Member, IEEE. (2001). Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results. IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 12, NO. 4, JULY 2001
- Ayesterán, José R., Ramoni, Josefa y Orlandoni Giampaolo. (1996). Crisis y Poder: El caso del Sistema Financiero. Economía Nueva Etapa. No. 11.
- Bell, J. y D. Pain, (2000), Leading Indicator Models of Banking Crises - a Critical Review, Bank of England. Review of Financial Stability No. 9:113-129.
- Borges, R. (2005). Análisis de supervivencia de pacientes con diálisis peritoneal. Revista Colombiana de Estadística 28 (2), 243-259
- Borges, R.E. (2002). Análisis de Supervivencia Aplicado a un Caso de Diálisis Renal: Diálisis Peritoneal en el Hospital Clínico Universitario de Caracas y Hemodiálisis en el Hospital de Clínicas Caracas, 1980-2000. Tesis de M.Sc. en Estadística Aplicada, Mérida, Venezuela: Instituto de Estadística Aplicada y Computación, Universidad de Los Andes.
- Broomhead, D. S., and D. Lowe, (1988). Multivariable functional Interpolation and adaptive networks. Complex Systems, vol. 2, pp. 321-355.
- Campos, R., Ruiz, F., Agell, N and Angulo, C. (2004). Financial Credit Risk Measurement Prediction Using Innovative Soft-computing Techniques. International Conference on Computational Finance And its Applications. Bologna, Italy, 2004.
- Ching, J.Y., Wong, A.K.C. and Chan, K.C.C. (1995). Class-Dependent Discretization for Inductive Learning from Continuous and Mixed Mode Data, IEEE Transactionson Pattern Analysis and Machine Intelligence, 17 (7), pp. 641-651, 1995.
- Colmenares G. and Pérez R. (1999). A Reliable Method to Reduce Observations and Variables when Building Neural Network Models. CAIP'99. San José. Costa Rica.
- Colmenares Gerardo, (2004). Reducing Archives to Build Non Linear Models Using Neural Networks. AMSE Periodicals. Lion. France
- Comisión Económica para América Latina (2000), Situación del Sistema Financiero en América Latina. CEPAL, Chile.
- Díaz M. Luís. (2002) Estadística Multivariada. Inferencia y Métodos.
- Dong D. and McAvoy T. (1996). Nonlinear Principal Component Analysis -Based on Principal Curves and Neural Networks. Computer Chem. Engng., vol 20, no. 1, pp. 65-78.
- Durán, Rodolfo, Mayorga, Mauricio y Montero, Renato (1999). Propuesta de indicadores macroeconómicos y financieros de alerta temprana para la detección de crisis bancaria. Banco Central de Costa Rica. División Económica. Costa Rica.
- Giampaolo Orlandoni, Colmenares Gerardo y Guillén Ruth (2005). Algoritmo Seguido para el Pre-Procesamiento de Datos de la Banca y Construcción de los Modelos No Lineales. Instituto de Investigaciones Económicas y Sociales. ULA, Mérida.
- Giampaolo Orlandoni, Colmenares Gerardo y otros (2005). Análisis Integral del Riesgo en la Banca Comercial y Entidades Financieras. Instituto de Investigaciones Económicas y Sociales. ULA, Mérida.
- Guillén Ruth (2005). Consideraciones Preliminares de la Gestión del Riesgo en el Sistema Bancario - El Caso de las Crisis Bancarias en Venezuela. Instituto de Investigaciones Económicas y Sociales. ULA, Mérida.
- Hair, J., Anderson, R., Tatham, R. and Black, W. (1998). Multivariate Data Analysis. Prentice Hall, 5th Ed.
- Haykin Simon. (1995). Neural Networks. A comprehensive foundation. Macmillan College Publishing Company, Inc.
- Haykin Simon. (1998). Feedforward: Neural Networks: An Introduction. Chapter 1. Manuscrito
- Hosmer y Lemeshow. Applied Survival Analysis. John Wiley & SONS, INC. (1999)
- Johnson Richard y Wichern Dean (1998). Applied Multivariate Statistical Analysis.

- Klein, J. and M. Moeschberger. Survival Analysis. Techniques for Censored and Truncated Data. Springer - Verlag. New York Inc. 1997
- Kramer M. (1992). Autoassociative Neural Networks. Computer Chem. Engng., vol. 16, no. 4, pp. 313-328.
- Kurgan, L. and Cios, K.J. (2001). Discretization Algorithm that Uses Class-Attribute Interdependence Maximization. Proc. of the 2001 International Conference on Artificial Intelligence (ICAI-2001): Las Vegas, pp.980-987, 2001
- López M. Thomas A. Julio (2001) Estudio de Técnicas de Análisis de Datos para Selección de Variables, Detección de Valores atípicos y Estimación de Valores Faltantes en Entradas al Sistema Neuronmaster.
- Medina, Yeny J. Borgucci Emmanuel. Desempeño de la rentabilidad de los bancos fusionados en Venezuela entre los años 1998 - 2002. Revista Venezolana de Análisis de Coyuntura, 2005, Vol. XI, No. 1 (ene-jun), pp. 119- 154.
- Muñoz S., Evelyn. (1998). La técnica de Análisis Discriminante: Una aplicación para el área bancaria. Banco Central de Costa Rica, DIE-NT-03-98
- Navarro, José L (2001). Métodos Multivariantes en el Control Estadístico de la Calidad. Facultad de Ciencias Económicas y Empresariales de Albacete. Universidad de Castilla - La Mancha, España.
- Pastor, Eduardo (2003). Sistema de Alerta Temprana de Fragilidad Financiera. Universidad de Pacífico. Departamento de Economía. Perú.
- Rangel, Javier G (2005). Creación de un Credit Scoring mediante Modelos de Ecuaciones Estructurales para la Selección de Clientes de Tarjeta de Crédito del Banco Sofitasa, Banco Universal C.A. Tesis de Pregrado, Escuela de Estadística, ULA, Mérida.
- Salafranca, Lluís (1992). Aplicación del análisis exploratorio de datos a los sistemas de ecuaciones estructurales. Universidad de Barcelona. Anales de psicología, 1992, 8 (1-2), 123-130.
- Sharma Subash. (1996). Applied Multivariate Techniques. John Wiley. USA
- Tan S. and Mavrouniotis M. (1995). Reducing data dimensionality through optimizing neural network inputs. AIChE Journal, vol 41, no. 6, pp. 1471-1480.
- Wang Z. Xue (1999). Data Mining and Knowledge Discovery for Process Monitoring and Control. Springer-Verlag. London.
- Xari Rovira, Núria Agell, Mónica Sánchez, Francesc Prats and Xavier Parra. (2002). An Approach to Qualitative Radial Basis Function Networks over Orders of Magnitude. This work was supported by the MCyT (Spanish Ministry of Science and Technology) MERITO project (TIC2002-04371-C02).