

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS POLÍTICAS Y SOCIOLOGÍA
Departamento de Economía Financiera y Contabilidad I



TESIS DOCTORAL

**Análisis de quiebra empresarial: modelo de ecuaciones de estimación
generalizadas sobre datos panel**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

José Guillermo Contreras Frías

Directores

María Jesús Segovia Vargas
María del Mar Camacho Miñano

Madrid, 2016

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

Departamento de Economía Financiera y Contabilidad I



TESIS DOCTORAL

ANÁLISIS DE QUIEBRA EMPRESARIAL: MODELO DE ECUACIONES DE ESTIMACIÓN GENERALIZADAS SOBRE DATOS PANEL

DIRECTORES

Dr. María Jesús Segovia Vargas

Dr. María del Mar Camacho Miñano

PRESENTADA POR:

José Guillermo Contreras Frías

MADRID, 2015

AGRADECIMIENTOS

Dedico esta Tesis:

Con todo mi amor para Lú por siempre estar a mi lado: Gracias por tu amor, paciencia y apoyo incondicional.

Con todo mi amor y cariño para mi mamá y mi papá por sus enseñanzas, amor, apoyo y confianza plena en mí y en todo lo que emprendo. Gracias por darme la vida y por darme tanto amor.

A mi hermana, a mi sobrina y a Kity por ser piezas clave en todo lo que me planteo. Gracias por ser parte de mi vida.

A mis tías y primos por ser parte de mi familia. Gracias por su confianza.

A mis amigos por sus consejos, desvelos, regaños, enojos y alegrías. Gracias por nunca dejarme solo.

A mis tutoras de tesis por su incondicional apoyo, consejos, correcciones y confianza en esta investigación. Así como a todos los profesores, investigadores y demás personas que intervinieron de una u otra manera en la consecución del presente trabajo.

A México y España

ÍNDICE

RESUMEN.....	5
ABSTRACT	7
CAPITULO 1. INTRODUCCIÓN.....	9
1.1 Objetivos e Importancia de la Investigación	16
1.2 Descripción General de la Metodología.....	19
1.3 Limitaciones y Alcances de la Investigación	21
CAPITULO 2. REVISIÓN DE LA LITERATURA SOBRE QUIEBRA EMPRESARIAL	23
2.1 Modelos con Variables Microeconómicas	23
2.1.1 Etapa Descriptiva.....	33
2.1.2 Etapa Predictiva	35
2.1.2.1 Modelos Univariados.....	35
2.1.2.2 Modelos Multivariados.....	37
2.1.3 Modelos con Inteligencia Artificial.....	42
2.2 Modelos de Predicción con Variables Macroeconómicas.....	48
CAPITULO 3. HIPÓTESIS Y METODOLOGÍA	54
3.1 Hipótesis a Contrastar	54
3.2 Metodología de Ecuaciones de Estimación Generalizadas	55
3.2.1 Datos Longitudinales o Datos Panel.....	56
3.2.2 Análisis de Componentes Principales.....	65
3.2.3 Marco Teórico Ecuaciones de Estimación Generalizadas.....	66
3.3 Metodología de Árboles de Decisión	73
CAPITULO 4. ESTUDIO EMPÍRICO	79
4.1 Tratamiento de la Información.....	79
4.1.1 Determinación de la Muestra	82
4.1.2 Determinación de la Variable Dependiente	83
4.1.3 Determinación de las Variables Independientes	87
4.1.4 Definición y Análisis Descriptivo de las Variables Independientes	93
4.1.4.1 Variables Microeconómicas (Ratios Financieros).....	93
4.1.4.2 Variables Macroeconómicas.....	103
4.2 Aplicación Empírica de las Ecuaciones de Estimación Generalizadas.....	119
4.2.1 Análisis Preliminar (ACP).....	119
4.2.2 Iteraciones del Modelo (EEG)	132

4.2.3 Interpretación de Resultados del Modelo (EEG).....	135
4.3 Aplicación Empírica de los Árboles de Decisión	141
4.3.1 Corte Transversal Año 2010.....	142
4.3.1.1 Validación de Resultados.....	142
4.3.1.2 Análisis de las Principales Ramas	144
4.3.1.3 Resumen de las Principales Reglas de Decisión	146
4.3.2 Corte Transversal Año 2014.....	146
4.3.2.1 Validación de Resultados.....	146
4.3.2.2 Análisis de las Principales Ramas	148
4.3.2.3 Resumen de las Principales Reglas de Decisión.....	151
4.4 Resultados y Discusión	152
CAPITULO 5.CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN	156
BIBLIOGRAFÍA	160
ÍNDICE DE FIGURAS.....	172
ÍNDICE DE TABLAS.....	173
ÍNDICE DE GRAFICOS.....	174

RESUMEN

Análisis de Quiebra Empresarial: Modelos de Ecuaciones de Estimación Generalizadas sobre Datos Panel.

En la creciente turbulencia del entorno económico y social que impera a nivel mundial, es precisamente la predicción oportuna de la quiebra empresarial uno de los problemas a los que se ha venido enfrentando la economía debido principalmente a las consecuencias que provoca a todos los agentes relacionados con la empresa.

En época de crisis financiera, los índices de mortalidad empresarial incrementan. Muchos y muy diversos han sido los estudios que se han elaborado relacionados con la predicción de los factores para prevenir situaciones de insolvencia financiera. Hoy en día y a pesar de los esfuerzos realizados, aún no existe una teoría única sobre la quiebra empresarial sin embargo, los datos contables provenientes de los estados financieros publicados por las empresas siguen siendo válidos para predecir dichas situaciones de riesgo.

En el dinámico entorno en el que actualmente se desempeña el mundo de las empresas a nivel global, resulta fundamental disponer de herramientas y metodologías que permitan al empresario anticipar situaciones no deseadas que puedan desencadenar situaciones de insolvencia.

La presente tesis doctoral tiene como objetivo general identificar aquellas variables tanto micro como macroeconómicas que resulten más significativas al momento de analizar las probabilidades de fracaso empresarial anticipadamente.

Para conseguir el objetivo planteado se han utilizado 19 variables explicativas, calculadas para todas las empresas cotizadas en los índices IPC de México, IBEX-35 de España y EURO STOXX50 de Europa.

Al ser empresas de diferentes países se analiza el papel que tiene tanto la información contable como la información macroeconómica en la predicción de crisis empresarial.

Además es importante mencionar que se ha considerado un horizonte temporal de 5 años para cada una de las empresas consiguiendo así una estructura de datos panel o datos longitudinales.

La metodología empleada para conseguir nuestro objetivo se basa en la contrastación empírica de dos técnicas distintas. Por un lado, la técnica paramétrica pertenecientes al análisis estadístico multivariado conocida como Ecuaciones de Estimación Generalizadas (EEG) y por otro lado, la técnica no paramétrica perteneciente al campo de la inteligencia artificial (IA) conocida como Árboles de Decisión.

Los resultados obtenidos muestran que las variables más significativas en cuanto a predicción de quiebra o fracaso empresarial son las provenientes de la información contable (Ratios Financieros) en concreto: el Ratio Coeficiente de Liquidez Circulante, el Ratio de Solvencia Global y finalmente el Ratio de Rentabilidad Económica (ROA). Estos resultados fueron similares bajo ambos enfoques (EEG e IA), lo cual muestra que no existe diferencia significativa en la aplicación de estas dos técnicas.

En cuanto a las variables macroeconómicas, los resultados muestran que no tienen la misma presencia que las variables financieras. En este sentido las condiciones macroeconómicas diversas que presentan cada uno de los países podrían originar problemas de comparabilidad de la información y diluir el efecto que éstas pudieran presentar. Esto último muestra que en un contexto de empresas cotizadas, las variables macroeconómicas no condicionarían la quiebra empresarial.

Los modelos de predicción de quiebra empresarial proveen información valiosa y adecuada para los distintos usuarios de la información contable así como las los tomadores de decisiones dentro de las empresa sin embargo, es muy importante tener en cuenta sus limitaciones y considerar los resultados como un indicador y como una herramienta más de apoyo para el diagnóstico profesional.

ABSTRACT

Corporate Bankruptcy Analysis: Generalized Estimating Equations on Panel Data.

In the growing turbulence of the economic and social environment prevailing all around the world, it is precisely the timely prediction of corporate bankruptcy the economy mainly due to the consequences caused to all the agents involved with the company.

In times of financial crisis, corporate mortality rates increase. Several studies have been conducted regarding predicting factors to prevent financial distress situations. Today despite the efforts made, there is still no single theory on corporate solvency; however the financial data from the financial statements published by the company are still valid to predict such risk.

In the dynamic environment in which today's world of business works globally, it is essential to have tools and methodologies that allow the employer to anticipate unwanted situations that may trigger insolvency.

This thesis has the general objective to identify both, micro and macroeconomic variables which are more significant when analyzing the likelihood of business failure in advance.

To achieve the stated objective has been used 19 explanatory variables calculated for all companies listed on the Mexico IPC index, Spanish IBEX-35 index and Euro zone EURO STOXX50 Index.

As there are companies from different countries the role of the accounting information and macroeconomic data in predicting business crisis is analyzed. It is also important to mention that it has been considered a time horizon of five years for each of the companies, obtaining a data panel or longitudinal data structure.

The methodology used to achieve our goal is based on the empirical testing of two different techniques. On the one hand, parametric technique belonging to multivariate statistical analysis known as generalized estimating equations (GGE) and secondly the nonparametric technique belonging to artificial intelligence (AI) known as Decision Trees.

The results show that the most significant variables in terms of bankruptcy prediction are those from the financial information (Financial Ratios): the Liquidity Circulating Ratio, the overall Solvency Ratio and finally the Return on Assets Ratio (ROA). Those results were similar under both approaches (GGE and IA), which shows no significant difference in the application of these techniques.

Regarding macroeconomic variables, the results show that they do not have the same presence than microeconomic or financial variables. In this sense the various macroeconomic conditions presented from each of the countries could lead with information comparability problems and dilute the effect these might have. The latter shows that in a context of stock exchange companies, macroeconomic variables would not condition the corporate bankruptcy.

Prediction models of corporate bankruptcy, provide valuable and accurate information for different users of accounting information as well as decision makers within companies, however it is very important to note its limitations and consider the results as an indicator and as a support tool for professional diagnosis.

CAPITULO 1. INTRODUCCIÓN

Desde hace varias décadas hay muchas líneas de investigación abiertas que siguen sin tener una respuesta clara por parte de los investigadores. Una de esas líneas es la predicción de la quiebra empresarial. Después de más de setenta años tratando de estudiar por qué las empresas fracasan y más de cincuenta años realizando modelos de comportamiento de empresas con problemas financieros comparadas con aquellas que no los tienen, todavía no existe una teoría única y clara sobre la quiebra empresarial (Balcaen, & Ooghe, 2006). De hecho, cada vez que hay alguna crisis empresarial parece que se incrementan los estudios sobre el tema. En el año 1966, cuando se publicó el primer estudio empírico de Beaver, se publicaron 9 artículos en la Web of Knowledge (ISI), primera base de artículos de prestigio a nivel mundial. El número de artículos sobre quiebra han alcanzado casi los 420 en el año 2013.

Muchas empresas entran en proceso concursal a raíz de crisis económicas. Por ejemplo, en México desde hace 10 años, 468 empresas de todos los ramos han estado en concurso mercantil en el país. La directora del IFECOM¹ en México afirmó que como efecto inmediato a una crisis económica se incrementan los litigios entre las compañías.

Así sucedió con empresas como “Comercial Mexicana”, que en 2008 empezó a sufrir el acoso de los bancos con quienes contrato instrumentos financieros derivados para supuestamente mejorar sus finanzas (Contreras, Segovia & Camacho, 2014). En la siguiente tabla se muestran algunas de las empresas mexicanas que reportaron pérdidas relacionadas con el manejo de Productos Financieros Derivados, en millones de dólares:

¹ Instituto Federal de Especialistas de Concursos Mercantiles: Creado por disposición de la Ley de Concursos Mercantiles, publicada en el Diario Oficial de la Federación el 12 de mayo del año 2000. Es un Órgano Auxiliar del Consejo de la Judicatura Federal con autonomía técnica y operativa, cuya principal finalidad es la de autorizar el registro de las personas que acrediten cubrir los requisitos necesarios para realizar las funciones de Visitador, Conciliador o Síndico, quienes apoyarán a la justicia en Materia Concursal en los aspectos técnicos involucrados en los procedimientos de concurso mercantil.

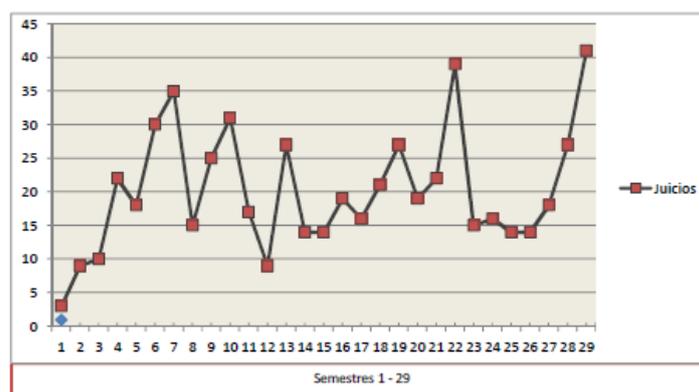
Tabla 1. Empresas en Crisis Financiera por Pago de Productos Financieros Derivados

Empresa	Pérdida o pago por derivados
Comercial Mexicana	1,080 yd
Gruña	684 yd
Vitro	227 yd
Alfa	191 yd
Grupo Industrial Saltillo	600 yd
Bachaco	50 yd
Autlán	45 yd
Grupo Posadas	25 yd

Fuentes: BMV, Byname Archival, Fitch Ratings

Como se observa en el gráfico 1, el número de empresas que entran en concurso mercantil en México se ve incrementado durante periodos de crisis económicas. Por ejemplo en el año 2010 (crisis económica a nivel mundial) y en el año 2014 (depreciación de la moneda).

Gráfico 1. Evolución del número de empresas en concurso mercantil en México



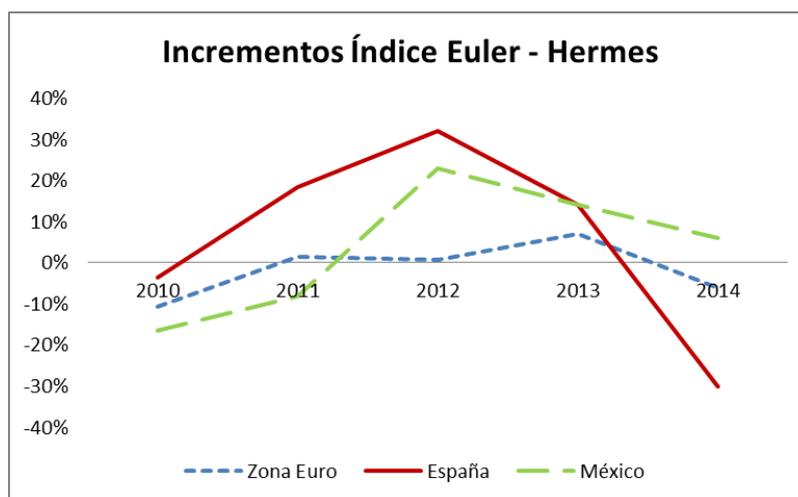
Fuente: Informe Labores IFECOM Junio-Noviembre 2014

Otro ejemplo lo encontramos en España donde el número de empresas que han entrado en proceso concursal en los últimos años ascienden casi a la cifra de 9.000 compañías. Comparando esta cifra con los años previos a la crisis, la tendencia se ha multiplicado por. De hecho, el índice global “Euler-Hermes”² sobre insolvencia de países, para España en el año 2012 sufrió un incremento de 32% con relación a 2011, tal y como se observa en el gráfico 2, mientras que el promedio de la zona euro se mantuvo alrededor del 7%.

² Euler Hermes es una compañía de seguros de crédito que ofrece una amplia gama de fianzas, garantías y servicios de cobranza para la gestión de los créditos por ventas de negocio a negocio.

En el mismo periodo dicho indicador para México fue de 23%. Este hecho es relevante porque se presupone que el estigma social del concurso de acreedores penaliza a las empresas que entran en el proceso legal y, las que entran, es porque realmente se encuentran en una situación de “coma inducido”, lo que implica que 9 de cada 10 de estas empresas terminen en liquidación y no reorganizándose financieramente para volver al mercado (Segovia-Vargas, M. J., & Camacho-Miñano, M. D. M., 2013).

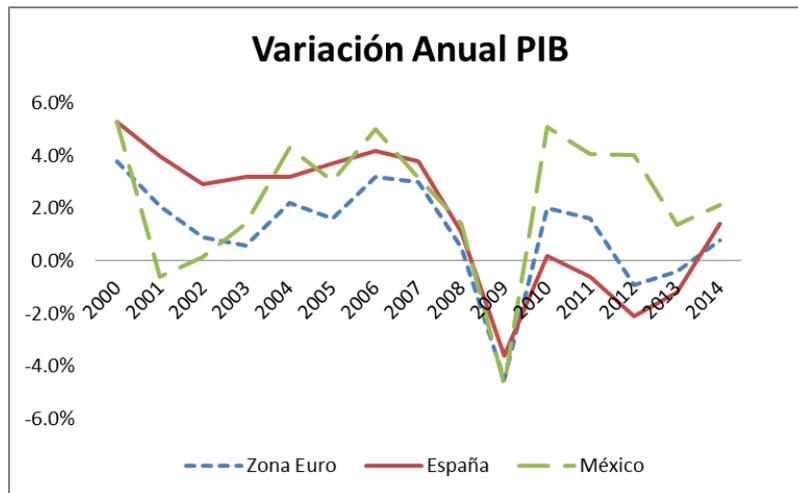
Gráfico 2. Índice Euler - Hermes en España, México y Zona Euro



Fuente: Informe “Economic Outlook” (www.eulerhermes.com)

Estos datos microeconómicos también tienen su reflejo a nivel macroeconómico. Por ejemplo, en el crecimiento del producto interior bruto (PIB), la tasa de desempleo, la deuda pública o el riesgo de insolvencia de crédito. Los siguientes gráficos (gráfico 3, 4 y 5) muestran, respectivamente, la tendencia de la variación interanual del PIB, de la tasa de desempleo y de la deuda pública en el caso español y mexicano así como las cifras de la Zona Euro.

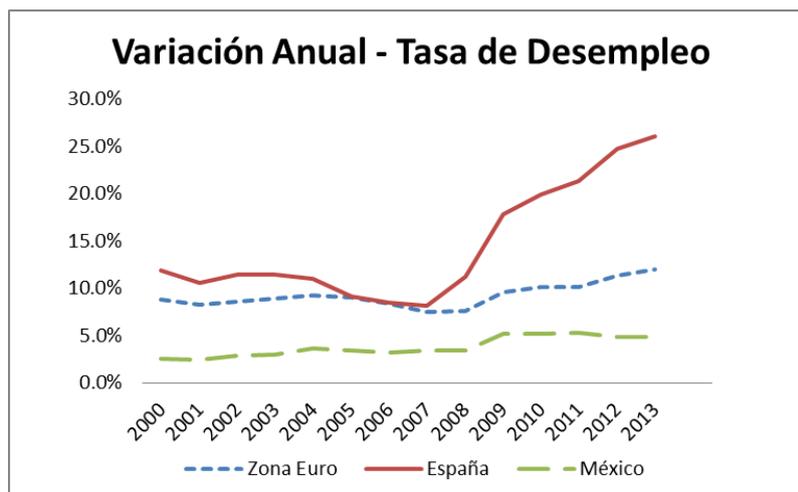
Gráfico 3. Tasa de Variación Interanual del PIB en España, México y Zona Euro



Fuente: Banco Mundial (www.datos.bancomundial.org)

El gráfico de la variación interanual del PIB (gráfico 3) muestra una caída brusca de este indicador macroeconómico en el año 2009, tanto en España como en el resto de la Zona Euro, año de fuerte impacto de la crisis económica. Este hecho se debe, en gran medida, al cierre de muchas empresas por la recesión económica sufrida en ese período. En el año 2009 también se observa la tendencia de la tasa de desempleo (gráfico 4) que despunta, aumentando hasta llegar al 26% en España. En esa misma fecha, la tasa de paro era de un 12% en el total de la Zona Euro.

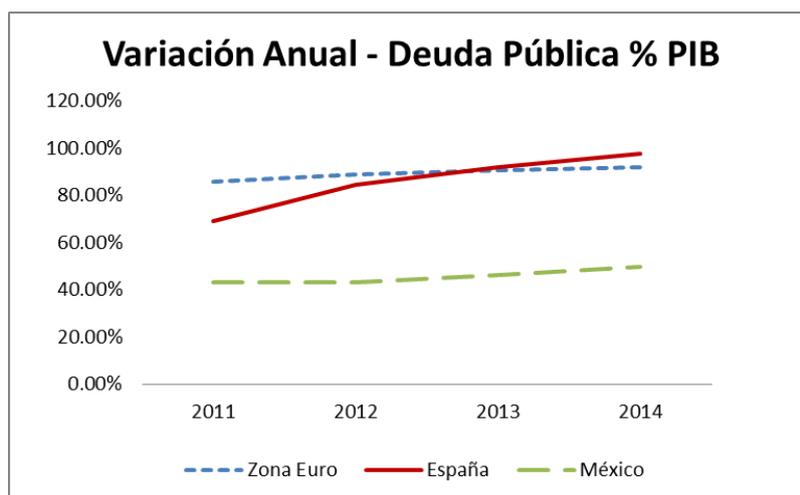
Gráfico 4. Tasa de Variación Interanual de Desempleo en España, México y Zona Euro



Fuente: Banco Mundial (www.datos.bancomundial.org)

Con respecto a la deuda pública, en el gráfico 5 se observa un crecimiento del peso de la deuda llegando a ser, al final de 2013, el 92.1% del PIB en España y casi el 90.9% en la Zona Euro, mientras que en México ha mantenido un crecimiento de más de 3 puntos porcentuales desde el año 2012 pasando de 43.17% y hasta alcanzar niveles de 50.08% a finales de 2014.

Gráfico 5. Tasa de Variación Interanual de la Deuda Pública en España, México y Zona Euro



Fuente: Banco Mundial (www.datos.bancomundial.org)

Estas variables macroeconómicas analizadas son sólo algunos ejemplos, pero la realidad es que existen otras muchas que se ven afectadas en períodos de recesión económica y que podrían ser utilizadas para predecir la quiebra empresarial.

Del mismo modo, características específicas de los distintos países, como es el caso de la legislación concreta de cada uno, también condicionan las crisis económicas. En relación con la legislación, se produce por ejemplo, un impacto en la solución de las situaciones de fracaso empresarial de las empresas, porque la resolución de los concursos de acreedores es diferente dependiendo del país del que se trate, ya que existen diferentes leyes concursales en función de cada uno. Esto se ha demostrado al comparar distintos países pertenecientes a la Unión Europea (Weijs, 2012). Por lo tanto, el tipo de regulación concursal que se establezca en un país no es una decisión baladí. Hay estudios económicos que justifican cómo las leyes sobre la insolvencia de empresas pueden condicionar el desarrollo del emprendimiento en un país (Lee et al., 2011), el crecimiento económico nacional (Laporta et al., 1998), la estabilidad económica (Beraho & Elisu, 2010) e incluso el desarrollo de los mercados de valores (Levine, 1998).

De hecho, se demuestra empíricamente cómo una ley concursal u otra puede cambiar las inversiones extranjeras en un país (Pindado et al., 2008). De ahí que sea de vital importancia un correcto desarrollo de la normativa sobre insolvencias para promover el crecimiento económico y la seguridad jurídica de los inversores.

Por todo lo expuesto anteriormente, como parte de los objetivos de este trabajo es analizar si las variables macroeconómicas pueden mejorar o no el poder predictivo de los modelos tradicionales de predicción de quiebras. Ya que son las empresas cotizadas las que se supone que más reflejan el día a día de cada país, sector y compañía, vamos a utilizar una muestra de empresas cotizadas en los índices mundiales más relevantes para testar nuestras hipótesis de partida. Se supone que nuevas variables macroeconómicas podría mejorar los modelos basados sólo en variables contables.

El presente trabajo de investigación fue estructurado y desarrollado en un total de 5 capítulos distribuidos de la siguiente manera:

- 1) El primer capítulo expone una introducción al contenido de la tesis. Se estipula el objetivo, limitaciones y alcance de la investigación para finalmente proporcionar una breve descripción de los enfoques que se han aplicado en este tipo de investigaciones.
- 2) El segundo capítulo presenta el resultado de las investigaciones realizadas sobre los antecedentes de las investigaciones en torno a los distintos modelos que se han estudiado para predecir quiebras, abarcando desde los primeros estudios desarrollados por Fitzpatrick (1932), pasando por los modelos predictivos realizados por Beaver, W. (1966) y Altman E. (1968) y finalizando por estudios con metodología como la inteligencia artificial.
- 3) El tercer capítulo se basa en el desarrollo del marco teórico de las metodologías que subyacen en este trabajo de investigación. Por una parte las Ecuaciones de Estimación Generalizadas (EEG), metodología que amplía los Modelos Lineales Generalizados (MLG) para estimar los parámetros de regresión más eficientes e imparciales relativos a mínimos cuadrados ordinarios, porque permiten la

especificación de una matriz de correlaciones que toma en cuenta la posible correlación de las medidas repetidas de cada individuo. Por otra parte, también se utilizan los árboles de decisión, técnica perteneciente al campo de la Inteligencia Artificial (IA). Estas herramientas son de gran utilidad, un complemento y en algunos casos una alternativa a los métodos estadísticos ya que mejoran en algunos aspectos, parte de sus limitaciones.

- 4) El cuarto capítulo utiliza primeramente la técnica estadística multivariada de los Componentes Principales con el objetivo de ver cuáles de las variables disponibles serán las candidatas para el modelo predictivo perdiendo la menor cantidad de información posible. Con base en los resultados obtenidos, se aplica el Modelo de Ecuaciones de Estimación Generalizadas y finalmente se contrastan los resultados con mediante la aplicación de los Árboles de Decisión.
- 5) El quinto y último capítulo presenta las conclusiones, ventajas, desventajas y limitaciones que fueron encontradas y que son el resultado de la investigación y del modelo propuesto así como las probables líneas de investigación a desarrollar.

1.1 Objetivos e Importancia de la Investigación

El desarrollo y utilización de modelos de predicción de dificultades financieras o quiebra empresarial han sido objeto de estudio de al menos las últimas cinco décadas debido, principalmente, a su habilidad para predecir si una empresa pudiera presentar dificultades financieras o eventualmente un estado de quiebra en los siguientes periodos.

Korol (2013) afirma que los problemas empresariales no son repentinos sino que se vislumbran entre cinco y seis años antes de su entrada en concurso, por lo que tienen que poder predecirse o, al menos, explicarse. La problemática económica que puede tener una empresa en dicha situación es muy grave y es por ello que contar con modelos de predicción que permitan identificar posibles dificultades resulta de gran importancia para empresarios, directivos, accionistas, inversionistas así como para el ámbito académico. Si bien es cierto lo anterior, lo que resulta aún de mayor interés, es poder tomar las acciones necesarias evitando consecuencias que impliquen mayores pérdidas.

Los modelos de predicción de dificultades financieras usualmente se construyen a partir de información proveniente de los estados financieros que permiten obtener ratios sobre rentabilidad, solvencia, liquidez, endeudamiento, etc. El análisis financiero a través de los ratios, índices o razones financieras, sirve como insumo principal en diversos modelos existentes para poder clasificar y predecir la quiebra de empresas.

El objetivo general de esta investigación consiste en identificar dentro de un conjunto de variables explicativas tanto microeconómicas como macroeconómicas, las que resulten más significativas al momento de intentar definir el posible fracaso de una empresa de manera anticipada.

El objetivo específico de esta investigación es contrastar o comprar los resultados obtenidos mediante la aplicación empírica de dos técnicas distintas. Por un lado, técnicas paramétricas pertenecientes al análisis estadístico multivariado como lo son las ecuaciones de estimación generalizadas (EEG) y, por otro lado, técnicas no paramétricas pertenecientes a la inteligencia artificial (IA) como los árboles de decisión.

El primero de los dos enfoques mencionados (EEG) requiere previamente de un análisis riguroso de los datos. Según Hair (2000), la influencia de datos atípicos, violaciones de los supuestos y la pérdida de datos puede agravarse a través de la pérdida de varias variables y tener efectos sustancialmente diferentes. Con la recogida y estructuración de los datos, el primer análisis del modelo predictivo no tiene que consistir en estimar principalmente dicho modelo, sino en evaluar los supuestos subyacentes que resultan ser tan importantes como el resultado final.

El examen de la base de datos es complejo y lleva tiempo, sin embargo, un análisis eficiente de las bases llevará siempre a una mejor predicción y, por lo tanto, a una mejor evaluación de las empresas en estudio. Los datos ausentes tienen efectos negativos en cualquier investigación y son producto de la introducción o estimación de los datos. Por otro lado, los datos atípicos son magnitudes extremas que puede también tener influencia negativa o ilógica en los resultados. Es importante resaltar que la efectividad estadística de las técnicas multivariadas requiere partir de grandes bases de datos y supuestos más complejos que los exigidos por el análisis estadístico tradicional.

Por otro lado, también es necesario obtener información contable cuya utilidad sea estable a través del tiempo, pues el análisis requiere de la observación de datos correspondientes a varios ejercicios económicos para constatar que pueden ser comparados. La gran cantidad de ratios que actualmente se presentan en la literatura contable y financiera, ha originado que los analistas se refieran a un mismo ratio con diferentes nombres, o bien, se refieran con un mismo nombre a un conjunto de ratios que son distintos entre sí y en sus componentes.

El segundo de los enfoques de árboles de decisión, no parte de hipótesis previamente establecidas, considerando los datos de partida de una forma plenamente exploratoria. Según Schaefer (1987, p.41) “los sistemas expertos son la técnica adecuada para realizar el análisis de la solvencia empresarial y poder anticiparse ante una caída o degradación en la posición financiera de la empresa”. Algunos autores han validado estas técnicas aplicándolas a datos reales para la realización de estudios (Cinca & del Brío, 1993; Díaz Martínez, Fernández Menéndez & Segovia Vargas, 2004).

Los algoritmos de inducción de reglas y árboles de decisión suponen un enfoque diferente ya que su aplicación implica realizar particiones sucesivas en el espacio de variables explicativas, empleando en cada partición una sola variable. Algunas de estas técnicas tienen carácter explicativo, como lo son la inducción de reglas y los árboles de decisión.

1.2 Descripción General de la Metodología

Con relación a las investigaciones de carácter empírico que se han desarrollado en este ámbito, según Mora Enguítanos (1994), se pueden destacar dos grandes enfoques o clases de trabajos empíricos:

- 1) Trabajos empíricos sobre los datos que analizan el comportamiento de los precios del mercado basado en la publicación de la información de estados contables. En estas investigaciones, se parte de la teoría de la eficiencia del mercado de capitales que establece la hipótesis de que “la información proveniente de estados contables tiene contenido informativo si provoca reacción a los precios”. A partir de la contrastación empírica de ésta hipótesis, los resultados positivos de la investigación científica tienden a reforzar la eficiencia del mercado de capitales.
- 2) Trabajos empíricos que desarrollan modelos de predicción sobre la solvencia y beneficio de las empresas tomando como base los datos contables. Estas investigaciones no presentan una única metodología debido principalmente, a que no existe un acuerdo generalizado sobre cuáles son los “óptimos”. Sin embargo, la evidencia empírica ha demostrado que dichos modelos explican de forma significativa cuáles son y de qué forma evolucionan las principales variables que intervienen en la composición del beneficio y en la determinación de la solvencia.

Este último tipo de investigaciones es la que se ha considerado para el desarrollo del presente trabajo. Su desarrollo no parte de una teoría económica particular, ni tampoco sobre una teoría financiera específica sobre la solvencia, sino que elaboran modelos muy variados entre sí, tomando como fundamento los datos contables con el fin de verificar su confiabilidad así como su capacidad de predicción.

Los estudios que se han realizado considerando ratios financieros para medir la solvencia y evitar el fracaso o quiebra empresarial, se han enfocado en la comparación entre empresas que no han quebrado y empresas en quiebra (sana y no sana). Para algunos investigadores, este tipo de estudios han permitido ofrecer pruebas de que las empresas que no quebraron presentaban ratios más “firmes” o “estables” con relación a las que si

quebraron, aunque si bien es cierto, no se ha demostrado de manera concluyente la capacidad de los ratios por sí solos para la predicción.

Las limitaciones de los ratios han ocasionado que no pocas investigaciones utilicen técnicas más globales y potentes como las desarrolladas en los Modelos Univariados y Multivariados. Los modelos predictivos, que integran un conjunto de métodos estadísticos avanzados, se enfocan hacia el contenido de la información contable expresada en forma de ratios.

1.3 Limitaciones y Alcances de la Investigación

Los modelos para predecir quiebras requieren de un conjunto de elementos cuya definición y aplicación técnica debe ser precisa con el fin de conseguir resultados empíricos verdaderamente valiosos. De acuerdo con los especialistas en el tema, si se consigue obtener estos elementos, entonces es posible integrarlos a todos ellos a una segunda fase, en la cual, a través de la aplicación de una metodología, se puede llegar a obtener, “ex – post”, porcentajes de capacidad y exactitud predictiva sobre un posible fracaso empresarial. Los elementos básicos a los que se refieren los especialistas en la mayoría de los casos son:

- i) Obtención y estructuración de la base de datos
- ii) Definición de la variable dependiente
- iii) Selección y definición de las variables independientes
- iv) Unidad de análisis
- v) Unidad geográfica
- vi) Unidad temporal
- vii) Selección de la muestra

En consecuencia, la aplicación de un modelo predictivo no es ninguna tarea fácil, el problema que existe para obtener las bases de datos, la comparación de los datos entre las diferentes empresas de la muestra seleccionada, las limitaciones actuales para obtener modelos armonizados para la elaboración de los estados financieros básicos (balance y estado de resultados), son solo algunos ejemplos de las limitaciones que se presentaron para conseguir los elementos necesarios que requieren los modelos predictivos.

En el caso específico de empresas que cotizan en Bolsas de Valores, García y Sánchez (2006) señalan que dada la compleja gama de circunstancias que afectan a sus títulos, los problemas o limitaciones del análisis financiero, pueden ser clasificados en los siguientes grupos: los derivados del entorno macroeconómico en general; los que se refieren a la empresa en su relación con el sector, las convenciones propias de cada mercado y finalmente, el componente psicológico consecuencia de la combinación de los factores descritos.

Los elementos que limitan y dificultan el análisis financiero son los de tipo contable, tales como la comparación de los estados financieros, la reconstrucción de los conceptos contables, la reclasificación de las partidas en función de criterios temporales a corto y largo plazo, la imprecisión y reformulación de conceptos, entre otros. La mayor limitación del análisis financiero es que en la actualidad representa únicamente un componente del análisis general de una empresa, debiendo ser complementado con el análisis técnico, el análisis fundamental, el análisis bursátil, entre otros.

Recientes investigaciones sobre la eficiencia de los modelos se han entregado por Agarwal y Taffler (2008) y Bauer y Agarwal (2014), teniendo en cuenta el rendimiento de los modelos basados en la contabilidad, los modelos basados en el mercado y los modelos basados en el riesgo. Según Agarwal y Taffler (2008), hay poca diferencia en la precisión predictiva de los modelos basados en la contabilidad y basados en el mercado, sin embargo, el uso de modelos basados en la contabilidad permite un mayor nivel de rentabilidad ajustada al riesgo de la actividad crediticia.

CAPITULO 2. REVISIÓN DE LA LITERATURA SOBRE QUIEBRA EMPRESARIAL

2.1 Modelos con Variables Microeconómicas

A lo largo de los años muchos han sido los investigadores que han centrados sus esfuerzos en intentar determinar el nivel de solvencia con el objetivo de poder predecir o, en su caso, evitar la quiebra de una empresa a través de la utilización de ratios, métodos estadísticos y análisis financiero sobre factores que impactan directamente en el resultado de las empresas tales como la liquidez, el apalancamiento, la rentabilidad, entre otros.

En el año 1932 Fitzpatrick realizó los primero trabajos dando origen a lo que se conoce como etapa descriptiva. Su objetivo primordial consistió en intentar detectar las quiebras empresariales a través de la utilización de ratios únicamente. En la misma línea se sitúa el trabajo de Winakor y Smith (1935), aplicando técnicas de análisis univariante básico, analizando las tendencias de varios ratios financieros. Sin embargo, no será hasta la década de los sesenta cuando se empiecen a utilizar técnicas estadísticas más complejas, como el análisis discriminante, univariante y múltiple. A partir de estos estudios, Beaver (1966), Altman (1968) y otros muchos, han intentado elaborar una teoría general del fracaso empresarial, aunque todavía sin éxito. En ellos la metodología se basa en un diseño apareado de empresas.

Inicialmente, los estudios que intentaban predecir y, por tanto explicar el fracaso empresarial, se basaban en ratios provenientes de información contable. Consistían en analizar los casos de fracasos empresariales reales y, utilizando un método inductivo, aprender las características comunes de las empresas fracasadas comparándolas con empresas “sanas”. Sin embargo, aún hoy en día, y con más de medio siglo de investigaciones posteriores, todavía no existe unanimidad entre los investigadores sobre cuáles son los ratios contables que mejor explican la situación de insolvencia. En la literatura contable, muchos investigadores han utilizado los principales ratios del análisis financiero o de los documentos de los estados financieros (balance de situación, cuenta de pérdidas y ganancias o estado de flujos de efectivo) para explicar la quiebra. Con carácter general, tres son los tipos de ratios más utilizados por los académicos sobre el tema: ratios de

rentabilidad, ratios de endeudamiento y ratios de equilibrio económico-financiero (entre otros, véase: Tascón Fernández y Castaño Gutiérrez (2012); Korol (2013)). Parece lógica la relación entre rentabilidad y liquidez, la idea es que empresas con problemas financieros son menos capaces de acceder a financiación, a recursos financieros externos, por ejemplo, los bancarios, lo que supone desequilibrios de caja relevantes.

Dependiendo de la metodología utilizada en la predicción de insolvencias, el modo de selección de los factores explicativos puede variar. Esto es, en el caso de que se utilicen técnicas estadísticas tradicionales, el modo de selección de las variables que expliquen la muestra puede condicionar los resultados, ya que no pueden incluirse todos los ratios contables existentes debido a los problemas de multicolinealidad entre ellos. Además, se requiere también que las variables cumplan ciertas hipótesis de partida, como son los criterios de normalidad y heterocedasticidad, para poder aplicar determinadas técnicas estadísticas.

En la década de los 80's aparecen los primeros cuestionamientos a éstos últimos modelos por ser no aleatorios Zmijewski (1984) y se avanza en la metodología con regresión logística o modelo Logit. Motivados por la importancia de incorporar la historia de cada empresa, se comenzó con la aplicación de modelos para datos longitudinales, como lo es el modelo lineal mixto o el modelo logístico mixto, que incorporan en su análisis los estados financieros de cada empresa en un horizonte temporal. A la hora de elaborar éste tipo de modelos, también es relevante la selección de las variables a utilizar. Para ello se utilizan diferentes técnicas: componentes principales, grado de significación estadística de las variables (hacia adelante o hacia atrás), juicio de investigadores o profesionales, análisis clúster, etc.

Si se decide utilizar metodologías no tradicionales (De Andrés et al., 2012; Ravi Kumar & Ravi, 2007) como métodos de inteligencia artificial (rough set, arboles de decisión como los algoritmos PART, C4.5, "random forest", redes neuronales...), se pueden considerar todas las variables disponibles siempre que clasifiquen bien el problema a estudiar, es decir, que su nivel de significación sea aceptable. Sin embargo, aunque se han reducido los errores en la clasificación de las empresas fracasadas, no se han conseguido clasificar completamente.

Según un estudio de Aziz y Dar (2006) hasta esa fecha, los estudios de predicción del riesgo de insolvencia habían utilizado en un 64% de los casos, modelos estadísticos, en un 25% habían hecho uso de técnicas de “soft computing” o inteligencia artificial, y un 11% habían estudiado el fenómeno a través de otro tipo de modelos.

Independientemente de la metodología utilizada, los ratios contables implican una serie de limitaciones inherentes a la contabilidad. En primer lugar, los datos contables son datos históricos, basados en hechos pasados, lo que dificulta en ocasiones la adopción de decisiones de predicción futura (Yeh et al., 2012). Otra limitación de los datos contables es que son objeto de manipulación o sesgos, en función de los incentivos que tengan sus gestores: el menor pago de impuestos, posponer o evitar un proceso legal de insolvencia, mostrar una imagen saneada para conseguir crédito, etc. (Campa & Camacho, 2014).

Además, las políticas contables de las empresas, e incluso los sectores a los que pertenecen, pueden tener cierta influencia sobre las variables contables (Balcaen & Ooghe, 2006). Así mismo, el hecho de la pertenencia o no a un grupo de empresas y el tipo de grupo podría modificar dichas variables (Korol, 2013).

Otro problema es la existencia de normativas contables diferentes, lo que hace difícil utilizar datos de empresas a nivel mundial. Por último, existen estudios académicos que destacan el papel que la información contable ejerce sobre la probabilidad de insolvencia de las empresas (Meeks & Meeks, 2009), ya que se plantea la paradoja de que la simple probabilidad de insolvencia afecta a la valoración contable de los activos y pasivos y que, a su vez, la valoración de éstos condiciona dicha probabilidad.

A pesar de todos estos problemas propios de la contabilidad, los modelos con ratios financieros (datos contables) son reconocidos y utilizados a nivel mundial. Recientemente se han entregado opiniones valiosas por Agarwal y Taffler (2008); Das, Hanouna y Sarin (2009) y Bauer y Argawal (2014), teniendo en cuenta la rentabilidad de los modelos basados en datos contables, los modelos basados en mercados y los modelos de riesgo. Estos tres tipos de modelos prevalecen en la literatura de finanzas. Para Argawal y Taffler (2008) hay poca diferencia predictiva de los modelos basados en datos contables y los basados en el mercado, sin embargo el uso de los modelos basado en la contabilidad permite un mayor

nivel de rentabilidad ajustada al riesgo, el utilizar ratios financieros, tiene la ventaja de no verse abrumado por el gran volumen de datos que contienen los estados financieros.

Otra causa importante para su aplicación es que para la gran mayoría existe un criterio pragmático generalizado en cuanto a su utilización, medición e interpretación para la toma de decisiones. Aun cuando se han realizado diversos tipos de investigaciones relacionadas con la predicción del fracaso empresarial, el modelo "Z-Score" original introducido por Altman (1968), ha sido el modelo dominante y de aplicación a nivel mundial. Aunque ha estado en existencia por más de 45 años, todavía se sigue utilizando como herramienta de predicción en quiebras o dificultades financieras.

A pesar de todo lo expuesto anteriormente, se plantean dos opciones por parte de los investigadores o el uso de variables bursátiles o el uso de otras variables que no sean financieras. Marais et al. (1984) fue el primero en señalar que la cotización de las acciones mejoraba la predicción del fracaso, en comparación con la utilización de variables contables únicamente. Barniv et al. (1997) comprobó que había resultados anormales acumulados, ponderados por precios de mercado, antes de la entrada a concurso de las empresas. Hillegeist et al. (2004) y Christidis y Gregory (2010) señalaron que el hecho de añadir las variables de mercado a los modelos mejoraba la oportunidad de información.

Chava y Purnanandam (2010) encontraron una relación positiva entre la rentabilidad de las acciones y el riesgo de fracaso. Por último, un reciente estudio de Bauer y Agarwal (2014) señala que los modelos de riesgos ("hazard models") añaden mayor capacidad predictiva que los modelos tradicionales de predicción de posibles insolvencias. Investigaciones recientes se han centrado en la capacidad predictiva de los estados financieros (Collins et al., 1997; Francis & Shipper, 1999 entre otros).

Los resultados obtenidos son variables y están sujetos a diversas interpretaciones. El uso de ratios financieros para predecir la quiebra empresarial presenta una larga historia (Beaver, 1966). Está bien establecido que los ratios financieros tienen poder predictivo hasta al menos 5 años antes de que suceda la quiebra.

Con relación a las variables no financieras se añadieron a los modelos variables tales como la edad de la empresa desde su fundación, el tamaño (como algoritmo del total de activos, número medio de empleados y cifra de ventas) y el sector (Tascón y Castaño, 2012).

Entre los primeros estudios que intentaron crear una teoría del fracaso y el éxito empresarial, destaca la teoría desarrollada por Lussier (1995), basada en quince variables internas de la empresa, como fueron el nivel de capitalización inicial del negocio, la experiencia de la empresa en el sector o la formación de los gestores. Sin embargo, el principal problema de este estudio era la dificultad para conseguir este tipo de datos de empresas fracasadas, procedentes de entrevistas a directivos.

Modelo a modelo se han ido añadiendo otras muchas variables. Cheng et al. (2007) realizaron un estudio en el que demostraron que las variables no financieras eran más significativas y aumentaban el grado de predicción de modelos con rough set (técnica de inteligencia artificial). En la misma línea, se encuentran los trabajos de Back (2005) y su comportamiento de pago de las empresas, Laitinen y Laitinen (2009) con información sobre los informes de auditoría, y De Andrés et al. (2012), que hacen uso de normas de los diferentes sectores económicos.

Noga y Schnader (2013) utilizan diferencias temporales de impuestos, Kallunki y Pyykkö (2013) analizan la experiencia pasada de los gestores de empresas en concurso y Chiu et al. (2013) explican la probabilidad de fracaso empresarial en función del grado de concentración del sector, basándose en la idea de que cuanto más se incrementa la competencia en un sector, más aumentará la probabilidad de fracaso.

Por último, cabe señalar que la última tendencia en investigaciones sobre la quiebra empresarial es la comparativa entre países, línea que sigue esta tesis doctoral. Son pocas las investigaciones que han realizado estos estudios comparativos debido principalmente a la falta de acceso a los datos de diferentes países. El desarrollo de las bases de datos comerciales han propiciado estos estudios.

Uno de los primeros estudios fue realizado por Altman y Narayanan (1997). Realizaron una revisión de los modelos de predicción del fracaso empresarial en 22 países. Otro estudio de Ravid y Sundgren (1998) compararon la eficiencia de los códigos legales de Finlandia y de

Estados Unidos, bajo los que se rige el fracaso empresarial. Analizando 70 empresas de cada país encontraron que, aunque los factores económicos que afectan a los procesos concursales coinciden en ambos países, la legislación finlandesa favorece que se llegue a un proceso de liquidación en mayor medida que la americana, en lugar de conseguir la reorganización de la empresa en cuestión. Laitinen (2002) también analizó empresas de 17 países europeos y de Estados Unidos, concluyendo que existen diferencias entre países con respecto al grado de fiabilidad de los modelos. A partir del análisis de su muestra, los países que mayor grado de fiabilidad en su clasificación obtuvieron fueron Alemania, Bélgica, Italia, Finlandia y Grecia, y, resultaron ser los peor clasificados, Suiza, Irlanda y Portugal.

De igual modo, Bellovary et al. (2007) se refieren a modelos en 18 países. Davydenko y Franks (2008), con una muestra de aproximadamente 1.500 empresas de Alemania, Francia y Reino Unido, llegaron a la conclusión de que la legislación vigente en cada país afectaba a las crisis empresariales aunque también el papel de los bancos o la competencia afectan al concurso de empresas.

Un estudio realizado por Korol (2013) compara datos de empresas polacas cotizadas, sanas y en concurso, con empresas latinoamericanas (de México, Argentina, Perú, Brasil y Chile) utilizando metodologías tradicionales y de inteligencia artificial. Concluye que son más difíciles de explicar las empresas latinoamericanas en concurso que las europeas, ya que el contexto normativo y macroeconómico de las latinas condiciona el concurso. Otro estudio de Laitinen y Suvas (2013) compara 30 países europeos, señalando que, a pesar de las diferencias entre países, es posible predecir el fracaso empresarial con algunos errores de clasificación aceptables.

Un reciente estudio aún en “working paper” es el trabajo de Altman et al. (2014). En este estudio se realiza un revisión de la literatura sobre la importancia y eficacia del modelo Z-Score de Altman de predicción de la quiebra a nivel mundial y sus aplicaciones en fianzas y otras área relacionadas. La revisión se basa en un análisis de 33 artículos científicos publicados desde el año 2000 en las principales revistas financieras y contables. El resultado del análisis muestra que mientras un modelo internacional general funciona razonablemente bien, con niveles de precisión de predicción que van desde 75% y hasta

90%, la precisión de la clasificación se puede mejorar de manera considerable con estimaciones específicas del país, especialmente con el uso de variables adicionales.

Resumiendo, la línea de investigación sobre predicción de quiebra entre países es clave debido a la globalización de los mercados internacionales y a la existencia de un inversor global. Por tanto, la existencia de un modelo predicción de quiebra o fracaso común para distintos países con un elevado grado de fiabilidad sigue siendo relevante y es uno de los propósitos de esta tesis. Innumerables trabajo se siguen generando en todo el mundo con el fin de “perfeccionar” los modelos predictivos con la adición en la aplicación de técnicas tanto paramétricas como no paramétricas más eficientes que han intentado obtener mayor exactitud en la predicción.

Tabla 2. Evolución Histórica de los Modelos de Predicción de Quiebra

Año	Autor	Etapas	Técnica Utilizada
1932	Fitzpatrick Paul	Descriptiva	Análisis financiero a través del método de ratios
1935	Winakor Arthur y Smith Raymond		
1942	Mervin Charles		
1965	Horrigan James		
1966, 1968	William Beaver	Inicio etapa predictiva	Estudio de modelos univariados basados en información contable
1968, 1977	Altman Edward	Desarrollo etapa predictiva	Estudio de modelos multivariados basados en información contable
1972, 1977	Deakin Edward		
1972	Edmister Robert		
1974	Blum Mare		
1980	Ohlson James		
1984	Rose Peter y Giroux Garay		
1984	Taffler Richard		
1985	Zavgren Christine	Desarrollo de etapa	Estudio de modelos multivariados
1980, 1988	Dambolena Ismael y Khory		
1980	Largay James y Stickney Clyde		

Año	Autor	Etapas	Técnica Utilizada
1980	Casey Cornelius	predictiva	basados en Cash Flow
1984, 1985	Casey Cornelius y Bartczack Norman		
1985	Kurokawa Yukiharu y Takahashi Kishinosuke		
1988	Ghalon James y Vigelan Robert		
1989	Azis Abdul y Lawson Gerald		
1973, 1975	Pinches George y Mingo Kent	Desarrollo de etapa predictiva	Estudio de modelos multivariados basados en el análisis factorial
1981	Chen Kung y Shimerda Thomas		
1985	Casey Cornelius y Bartczack Norman		
1987	Gambola Michael, Haskins Mark, Kents Edward		
1980	Olshon	Etapa Predictiva	Estudio de modelos multivariados basados en datos longitudinales
1984	Zmijewski		
1993	Marco Giancarlo y Franco Varetto		Estudio basados en Métodos de Inteligencia Artificial (Redes Neuronales)
1997	Barniv, R. , Agarwal, A y Leach, R		
2001	Atiya, Amir F	Etapa Predictiva	Estudio basados en Métodos de Inteligencia Artificial

Año	Autor	Etapas	Técnica Utilizada
2001	Swicegood, P., y Clark, J. A.		(Redes Neuronales)
2001	Mar-Molinero, C., y Serrano-Cinca		Estudio de modelos multivariados (Escalamiento Multidimensional)
2004	Cielen, A., Peeters, L. y Vanhoof, K.		Estudio de modelos multivariados (Data envelopment analysis)
2005	Beaver, W. H., McNichols, M. F., y Rhie, J. W.		Estudio de modelos multivariados (Modelos Hazard)
2008	Miranda, M. E. G., de la Torre Martínez, J. M., y Martínez, I. R.		Estudio de modelos multivariados (Modelo logístico)
2008	Agarwal, V., y Taffler, R.		Estudio de modelos multivariados
2009	Beaver, W. H., McNichols, M. F., y Rhie, J. W.		Estudio de modelos multivariados basados en datos longitudinales
2014	Bauer, J., & Agarwal, V.	Etapas Predictiva	Estudio de modelos multivariados (Modelos Hazard)

Año	Autor	Etapas	Técnica Utilizada
2014	Acosta-González, E., & Fernández-Rodríguez		Estudio basados en Métodos de Inteligencia Artificial (Algoritmo Genético)

Fuente: Elaboración Propia del Autor

Es importante destacar que, en cada una de las etapas sobre el desarrollo de las investigaciones predictivas, se ha puesto mayor atención a determinados elementos que están implícitos en todas las investigaciones (tales como la unidad temporal, la selección de la muestra, la determinación de la variable dependiente, etc.) y que consideramos son de mucha importancia para observar correctamente el desarrollo o limitaciones que presentan los estudios sobre el análisis de la solvencia.

2.1.1 Etapa Descriptiva

En la etapa descriptiva las investigaciones tuvieron como principal característica el intentar demostrar que los ratios de las empresas que fracasaban presentaban una tendencia favorable conforme se acercaba la fecha del fracaso definitivo de la empresa. Hoy en día, ese pensamiento se mantiene en muchos de los trabajos que tratan sobre este tema y que sin embargo continúan sin considerar que existen algunas premisas que son importantes antes de depositar la confianza en este criterio. Por ejemplo:

- i) Que puede existir un grado de manipulación importante en los estados financieros y en especial en aquellas partidas financieras que forman parte de algunos de los ratios con mayor capacidad predictiva.
- ii) Que a través del tiempo se producen importantes cambios en los criterios de amortización y valuación de los activos que se toman en cuenta al elaborar los estados financieros.
- iii) Falta de calidad en la auditoría externa .
- iv) Aceptación o rechazo de los directivos de las empresas para adoptar los nuevos criterios contables propuestos por los colegios de contadores.

Con respecto a las limitaciones que presenta el método de ratios, en una primera aproximación e interpretación de información financiera interna de las empresas puede realizarse por medio de la selección y síntesis adecuada de los datos. Si consideramos sus limitaciones, el uso de ratios de forma simple, rápida y significativa, proporciona información primaria y reveladora sobre la evolución de la empresa, y hasta cierto punto, sin la necesidad de recurrir a modelos matemáticos o estadísticos que, de no ser bien utilizados, se convierten en instrumentos meramente teóricos. Sin embargo, dicho método tiene insuficiencias y debilidades por la relativamente fácil manipulación de sus componentes.

Los ratios³ parten de la idea de la comparación de magnitudes. Hace referencia a dos datos aislados (partidas contables) que tienen cada uno valor por sí mismo y que frecuentemente adquieren mayor interés cuando se les combina en un cociente. Como es bien sabido la comparación de dos magnitudes se puede realizar en dos sentidos:

³ La etimología de la palabra ratio significa relación, razón, división, proporción.

- a) Restando, lo cual nos dicen en cuanto excede una cantidad a la otra.
- b) Dividiendo, lo cual nos dice cuantas veces contiene una cantidad a la otra.

La metodología de los ratios o razones financieras utiliza el cociente, cuyo fundamento teórico consiste en elegir relaciones racionales entre magnitudes significativas. La aplicación práctica de las razones y proporciones se debe a los matemáticos italianos del renacimiento, principalmente a Luca de Paccioli (1440 – 1515) considerado como el inventor de la partida doble e iniciador del desarrollo de la técnica contable.

Como se mostró en la tabla 2, los estudios empíricos con ratios financieros se inició en el año 1932 por *Paul Fitzpatrick* quién seleccionó una muestra de 19 empresas en quiebra y un grupo de control de 19 empresas sin quiebra en el periodo 1920 – 1929, con el fin de analizar las tendencias de 13 ratios a través del tiempo (3 – 5 años). Concluyó que todos los ratios predecían, en mayor o menor medida la quiebra, debido a la existencia de tendencias favorables y estables de los ratios pertenecientes a la empresas sanas, comprandola con la desfavorable evolución de los ratios pertenecientes a las empresas en quiebra.

Los modelos desarrollados durante esta etapa presentan el común denominador que aunque en la actualidad existe poca vigencia de sus resultados, la trascendencia de su aportación al desarrollo de la teoría de la solvencia ha sido muy importante, ya que representaron el primer intento para sistematizar un conjunto de procedimientos tendientes a describir el fracaso empresarial.

2.1.2 Etapa Predictiva

La predicción de la quiebra constituye el tema central que ha dirigido la mayoría de los trabajos empíricos. Durante esta etapa, la idea que se desarrolló fue la significancia de cada ratio, así como la capacidad predictiva de los ratios más relevantes. En esta segunda etapa es cuando surgen los modelos univariados.

2.1.2.1 Modelos Univariados

El análisis estadístico univariado consiste en el análisis de cada una de las variables estudiadas por separado, es decir, el análisis está basado en una sola variable. Dentro de las técnicas más frecuentes en este tipo de análisis se encuentran la distribución de frecuencias para una tabla univariada y el análisis de las medidas de tendencia central de la variable. La mayoría de los análisis univariados enfatizan la descripción mientras que los métodos multivariados hacen hincapié en las pruebas de hipótesis y explicaciones. Aunque la mayoría de investigaciones en el mundo real analiza el impacto de múltiples variables independientes sobre una variable dependiente, muchas de las técnicas multivariadas, como el caso de la regresión lineal, pueden ser utilizadas de forma univariada examinando el efecto de una sola variable independiente sobre una variable dependiente. Tal vez la mayor similitud entre ambas técnicas de análisis radica en que las dos son importantes para entender y analizar una gran cantidad de datos estadísticos.

Modelo de Beaver

William H. Beaver (1966) fue el pionero en esta etapa demostrando que los ratios financieros pueden ser de utilidad en la predicción individual de un fallo de la empresa, de las dificultades financieras y de la quiebra. Encontró que podría discriminar un número de indicadores con el simple mapeo entre muestras de empresas fracasadas y no fracasadas hasta cinco años antes del fracaso. Beaver utilizó dicha técnica, para explicar una variable

dependiente a través de la clasificación dicotómica que entendió como capacidad de predicción.

Cabe mencionar que el principal objetivo del trabajo de *Beaver* no fue encontrar el mejor predictor de quiebra, sino investigar la capacidad predictiva de los ratios financieros. Por eso en sus conclusiones él mismo indica que el sentido real del título de su artículo clásico de 1966 no debería haber sido “*los ratios financieros como predictores de una quiebra*”, sino más bien el de: “*Los datos contables como predictores de la quiebra*”, debido a que para *Beaver*, los ratios son datos contables comprimidos que pueden ser evaluados en términos de su utilidad, definiendo a su vez esta utilidad en términos de su capacidad predictiva.

En su primer estudio empírico se enfocó en la identificación de un único ratio que tuviera la capacidad predictiva para lo cual considero una muestra de 79 empresas en quiebra contra 79 empresas sin quiebra durante el periodo 1954 -1964. El ratio que obtuvo los mejores resultados dentro de su estudio fue el ratio de Cash Flow (*Cash Flow / Total Pasivo*), seguido de los ratios de Resultado Neto (Beneficio Neto / Total Activo) y Ratio de Endeudamiento (Total Pasivo / Total Activo).

Beaver concluyó que ciertos ratios son excelentes predictores de quiebras ya que permiten detectar la falta de solvencia al existir una enorme diferencia entre los ratios de las dos sub muestras que utilizó y que los ratios de rentabilidad seguidos de los ratios de liquidez representaron las variables explicativas más significativas sobre la situación futura de una empresa. Puso en duda el uso del análisis multivariado.

Los estudios de *Beaver* fueron muy importantes en el desarrollo de la teoría de la solvencia, ya que logró separar y analizar los componentes de los ratios mediante el uso de métodos estadísticos y el cálculo de la media de los valores de dichos componentes, tanto de empresas en quiebra como sanas.

Otra conclusión de gran relevancia a la que llegó *Beaver* fue relacionada con el tema de las variables independientes y fue que los ratios de rentabilidad seguidos de los ratios de liquidez representaron las variables explicativas más significativas sobre la situación futura de una empresa.

2.1.2.2 Modelos Multivariados

El análisis multivariado tiene sus orígenes en las primeras generalizaciones de la correlación y regresión, en donde se establecieron las primeras ideas del análisis de componentes principales (Pearson, 1901; Spearman, 1904); su establecimiento definitivo se ubica en los años treinta con los estudios de Hotelling (1931), Fisher (1935) entre otros. En general, cualquier análisis que considere simultáneamente dos o más variables forma parte de lo que se conoce como análisis multivariado.

La teoría de la solvencia originalmente consideró como fundamento la metodología de los ratios y se vio seriamente enriquecida con los métodos estadísticos multivariados⁴ que dieron a los investigadores instrumentos más potentes y eficaces para llegar a predicciones más fiables. De manera paralela a este desarrollo, en gran parte de los países se experimentó un progreso en la armonización de los principios de contabilidad generalmente aceptados así como en las normas y procesos de auditoría.

Poco a poco los modelos multivariados fueron superando la exactitud de las clasificaciones univariadas al ser más precisos los ratios y obtener porcentajes más altos de exactitud, alcanzando entre 95% y 97% comparados con los univariados que alcanzaban el 87%. El análisis multivariado parte de la idea de que la predicción de quiebra es un tema complejo y como tal, requiere de un análisis en el que intervienen una considerable cantidad de variables ($n > 1$). Para Kendal (1975) este tipo de análisis se refiere al conjunto de técnicas estadísticas que analizan de manera simultánea más de dos variables en una muestra de observaciones dadas. Para Cuadras (1981) esta técnica estudia, interpreta y elabora el material estadístico sobre la base de un conjunto de $n > 1$ variables, las cuáles podrían ser de tipo cualitativo o cuantitativo o bien, una combinación de ambas.

⁴ Los métodos estadísticos multivariados cobraron fuerza gracias al desarrollo de softwares estadísticos tales como SPSS, SAS, BMPD, LISREL, SPAD, etc.

Modelo de Altman

Edward Altman (1968) amplió el análisis multivariado al introducir por primera vez múltiples predictores de quiebra mediante el Análisis Discriminante Múltiple (MDA). A lo largo de los años, este autor ha sido considerado por muchos como el investigador que más ha contribuido al desarrollo de la teoría de la solvencia mediante la creación del modelo conocido como “Z-Score”. Para el desarrollo de su investigación Altman (1968) seleccionó una sub muestra de 33 empresas que fueron a la quiebra y otra de igual tamaño de empresas que no fueron a la quiebra del sector manufacturero que cotizaban en la bolsa de valores durante el periodo 1946 al 1965. Para la selección de las variables independientes Altman integró inicialmente un grupo de 22 ratios que fueron aplicados a ambas sub muestras de empresas. Estos 22 ratios fueron disminuidos a cinco factores que median: la rentabilidad, actividad, liquidez, apalancamiento y solvencia. Para Altman estos 5 factores resultaban ser la mejor combinación para el discriminante entre empresas en quiebra y empresas sin quiebra.

La función discriminante que construyó y conocida como “Z-Score” es considerada por un gran número de investigadores y académicos como uno de los mejores modelos teóricos de predicción de quiebras. Se basa en unas ponderaciones sobre cinco ratios.

$$Z = 1.2 X_1 + 1.4 X_2 + 3.3 X_3 + 0.6 X_4 + 0.99 X_5$$

Donde:

$Z = \text{Indicador Global } Z - \text{Score}$

$$X_1 = \text{Ratio de liquidez} = \frac{\text{Capital circulante neto}}{\text{Activo Total}}$$

$$X_2 = \text{Ratio de rentabilidad acumulada} = \frac{\text{Beneficios no distribuidos}}{\text{Activo Total}}$$

$$X_3 = \text{Ratio de rentabilidad} = \frac{\text{Beneficios antes de intereses e impuestos}}{\text{Activo Total}}$$

$$X_4 = \text{Ratio de estructura financiera} = \frac{\text{Valor de mercado de fondos propios}}{\text{Valor Contable de los Activos}}$$

$$X_5 = \text{Tasa de rotación de capital} = \frac{\text{Ventas Netas}}{\text{Activo Total}}$$

La experiencia de este modelo llevó a su autor a la conclusión de que puntuaciones Z-Score por debajo de 1,81 indicaban una alta probabilidad de quiebra. En cambio puntuaciones superiores a 3,00 indicaban una escasa probabilidad de quiebra. El rango comprendido en 1,81 y 2,99 fue denominado como “zona de ignorancia”. Las empresas con este z-score deberían ser analizadas en profundidad para determinar su probabilidad de quiebra.

Modelo de Deakin

Edward B. Deakin (1972) combinó en su trabajo las investigaciones realizadas por Beaver y Altman dentro de un único estudio desarrollando un modelo de fracaso empresarial. Consideró que Beaver en su modelo univariado obtuvo mejores resultados predictivos a través de la eficiencia de algunos de los ratios contrastados con el Modelo Multivariado de Altman, sin embargo, el método de este último tenía más atracción intuitiva.

En su trabajo intentó combinar primeramente los aspectos más interesantes de ambos modelos y partiendo de la misma hipótesis teórica, reprodujo el estudio de Beaver (1968), utilizando las mismas variables explicativas. Posteriormente, utilizó la metodología de MDA de Altman para investigar las combinaciones de los 14 ratios de Beaver, los cuales podrían ser los mejores predictores de una quiebra dentro de una combinación lineal de ratios. En su muestra Deakin seleccionó 11 empresas en quiebra incluidas en el “*Mody’s Industrial Manual*” y 23 empresas sin quiebra durante el periodo 1963 – 1964. Su estudio partió de un test de “significación individual” y “clasificación dicotómica” de cada variable independiente. Luego llevó a cabo el cálculo de las combinaciones multivariantes de los ratios a través de las funciones discriminantes. Deakin llegó a dos interesantes conclusiones basadas sobre el “*scales vector*” que indicaban la contribución relativa de cada variable a la función discriminante y fueron:

- 1) Consiguió disminuir el número de variables, eliminando aquellas que proporcionan una contribución relativamente pequeña a la función y que daban como resultado el incremento sustancial dentro del número de errores de clasificación.
- 2) Descubrió que un modelo único puede ser suficiente para predecir una quiebra con un alto grado de probabilidad, pero insuficiente para predecir a largo plazo la misma probabilidad de que ocurra.

Sus resultados coincidieron con los de Beaver en cuanto al ratio de cash flow / deuda total que obtuvo la mayor exactitud de clasificación, y cuyos porcentajes fueron altos, incluso hasta 5 años antes de la quiebra. También los ratios de resultado neto / activo total y deuda total / activo total, proporcionaron excelentes resultados.

Modelo de Ohlson

James A. Ohlson (1980) desarrolló su trabajo presentando como cambio principal con respecto a otras investigaciones la utilización de un método estadístico diferente de los que tradicionalmente se habían venido utilizando. Utilizó el modelo econométrico de probabilidad condicional de regresión logística (Logit) en lugar del MDA intentando mejorar las diferencias que éste último presentaba. Según Ohlson el MDA tenía poco valor interpretativo al contrario del Logit. Antes de la década de los 80's, aun cuando ya se habían detectado problemas metodológicos y estadísticos aplicando el MDA, ningún otro investigador intentó probar otras alternativas para mejorar estos aspectos.

Ohlson detectó, con respecto a la variable dependiente, que la definición de fracaso tenía significativas variaciones y arbitrariedades teniendo en cuenta la diversidad de estudios; La dicotomía entre "quiebra vs no quiebra" era normalmente la más utilizada, pero para él representaba una aproximación muy "cruda". Además, señaló que también existía un gran número de dificultades estadísticas y problemas metodológicos que debían ser discutidos.

La base de datos utilizada por Ohlson para el desarrollo de su investigación, difirió de los estudios anteriores en que no seleccionó empresas fracasadas, pues optó por un procedimiento diferente, que según él, tenía una mayor ventaja al considerar que los

reportes indicaban una situación en un momento del tiempo en que éstos fueron hechos públicos, y por lo tanto, el analista no podía verificar si la empresa había llegado a la quiebra antes o después de que los datos fueran publicados.

Este problema de temporalidad se daba porque se creía que los informes financieros de las empresas estaban disponibles a la fecha del cierre anual, cosa que en la práctica no sucedía, y menos aún en las empresas “fracasadas” las cuales en ocasiones llegaban a producir su información después de ésta y ya no era “oportuna”. Ohlson prefirió buscar otra fuente para la extracción de sus bases de datos que le indicara la fecha exacta de su emisión pública para en realidad poder predecir relaciones económicas, pues si una empresa solicitaba el procedimiento concursal entre el posterior cierre de ejercicio y antes de la emisión de los estados financieros finales, la información perdía validez, y por consiguiente el análisis predictivo.

Los estudios realizados por Ohlson sentaron un precedente en los estudios sobre la posibilidad de elección de metodologías estadísticas alternativas, debido a que se comenzaron a utilizar modelos más flexibles en cuanto a sus requerimientos, proponiendo en general otros dos modelos de probabilidad condicional (Logit y Probit) que resultaban tan eficientes como el MDA.

Cash Flow

A partir de la década de los 80's se iniciaron estudios conducidos principalmente para investigar si existía evidencia empírica sobre una mayor capacidad predictiva al utilizarse el cash flow operativo, para aplicarlo posteriormente a ratios individuales o en combinación con otros ratios basados en el devengo. La utilización de datos basados únicamente en el devengo, queda parcialmente justificada por la falta de información y normatividad sobre el cash flow en aquella década. Éste concepto fue considerado en general como la suma de los ingresos netos más las depreciaciones o el capital de trabajo operativo.

Los resultados de los estudios predictivos, que consideraron al cash flow operativo, mostraron que tales medidas podían utilizarse solo como una contribución marginal para diferenciar entre las empresas en quiebra y las empresas sanas sin embargo, dicha exactitud marginal dentro del modelo podía ser mínima y siempre debía tenerse un gran cuidado para seleccionar a las variables independientes.

2.1.3 Modelos con Inteligencia Artificial

En el momento actual la Inteligencia Artificial se aplica a numerosas actividades humanas y como líneas de investigación más explotadas destacan el razonamiento lógico, la traducción automática y comprensión del lenguaje natural, la robótica, la visión artificial y, especialmente, las técnicas de aprendizaje y de ingeniería del conocimiento. Estas dos últimas ramas son las más directamente aplicables al campo de las finanzas pues, desde el punto de vista de los negocios, lo que interesa es construir sistemas que incorporen conocimiento y, de esta manera, sirvan de ayuda a los procesos de toma de decisiones en el ámbito de la gestión empresarial.

En el ámbito específico del Análisis Contable, según Bonsón y Sierra (1996) la Inteligencia Artificial constituye una de las líneas de actuación futura más prometedoras, con posibilidades de aplicación tanto en el ámbito de la investigación como en el diseño de sistemas de información inteligentes, que no solamente proporcionen datos al decisor sino que recomienden el mejor curso de actuación a seguir.

Según Altman y Saunders (1998, p. 1722) el análisis de la solvencia empresarial ha sufrido una gran evolución a lo largo de los últimos 20 años, debido a factores tales como el aumento en el número de quiebras, la desintermediación creciente que se observa en los mercados financieros, la disminución de los tipos de interés o el desarrollo de nuevos instrumentos financieros. Todo ello ha impulsado el desarrollo de nuevos y más sofisticados métodos de análisis de la solvencia, y entre este tipo de sistemas ocupan un papel destacado aquellos que están basados en técnicas de Inteligencia Artificial.

La determinación de la solvencia futura de una empresa puede ser entendida en la mayoría de los casos como una operación de clasificación, es decir, dada una información inicial o conjunto de atributos asociados a una empresa, y extraídos en su mayor parte de los estados contables de la misma, lo que pretende el analista es tomar la decisión de clasificar a esa empresa dentro de una categoría concreta de riesgo financiero, de entre varias posibles.

Aplicando la clásica división que hizo Simon (1960) de los procesos de decisión entre estructurados y no estructurados, es claro que esa decisión es de tipo no estructurado ya que no existe un procedimiento definido para abordarla, siendo necesario el juicio y la propia evaluación del decisor. Tal y como señalan diversos autores (Ball y Foster, 1982, p. 217; Martín Marín, 1984, p. 630), no existe una teoría comúnmente aceptada que explique el fenómeno del fracaso empresarial por lo que a priori, no es posible establecer qué variables financieras ni qué valores en las mismas determinan la futura solvencia o insolvencia de una firma.

Debido a lo anterior, el estudio de la solvencia implica una investigación selectiva dentro de un espacio de alternativas inmenso pues, como se ha comentado, no existe un procedimiento que conduzca de forma inequívoca a la solución óptima. Por lo tanto, la selección ha de estar basada en reglas prácticas o heurísticas, debiendo fijarse también un criterio de suficiencia para determinar cuándo las soluciones encontradas son satisfactorias. Todo ello concuerda plenamente con el paradigma de la racionalidad limitada, que gobierna los procesos de decisión en el ámbito económico. Ese análisis heurístico se ha implementado tradicionalmente a través de la aplicación de técnicas estadísticas, tales como el análisis multidiscriminante lineal o los diversos modelos de variable de respuesta cualitativa (Logit, Probit, etc.). Sin embargo todas estas técnicas presentan limitaciones, pues parten de hipótesis más o menos restrictivas, que por su propia naturaleza la información económica, y en especial los datos extraídos de los estados financieros de las empresas, no van a cumplir, perjudicando así los resultados.

La aplicación de técnicas procedentes del campo de la Inteligencia Artificial surge como un intento de superar esta limitación, pues estas últimas no parten de hipótesis preestablecidas y se enfrentan a los datos de una forma totalmente exploratoria,

configurándose como procedimientos estrictamente no paramétricos. Dentro de los diversos sistemas de Inteligencia Artificial que han sido aplicados al campo del análisis de la solvencia se encuentran:

- i) Los Sistemas Expertos. Según Schaefer (1987, p. 41), los sistemas expertos son una técnica adecuada para cumplimentar muchas de las tareas que engloba el estudio de la solvencia empresarial, entre las que se puede destacar la evaluación de la gestión de la empresa, la vigilancia para anticipar una degradación en la posición financiera y la medida de la eficacia de la empresa en relación con su entorno competitivo. La mayor parte de los sistemas expertos en el campo del análisis de la solvencia están formados por dos módulos: un módulo informático convencional que realiza los cálculos con ayuda de una hoja de cálculo electrónica y un módulo de sistema experto, que efectúa los procesos de análisis e interpretación de los datos y de emisión del informe, normalmente por medio de un procesador de texto.

- ii) Las Técnicas de Aprendizaje Automático. Cuando se trata de inferir conocimiento a partir de los datos contenidos en una base, es posible hablar de dos líneas de investigación principales:
 - a. Una de ellas trata de realizar una aproximación simbólica/algorítmica, y simula el razonamiento lógico mediante la manipulación de símbolos. Dentro de esta corriente, los algoritmos de inducción de reglas y árboles de decisión son las técnicas más adecuadas para el análisis de la solvencia.
 - b. La otra rama de la investigación se ha dirigido a intentar construir estructuras que aprendan y se auto organicen, imitando a las de la vida biológica, siendo las redes neuronales el producto de esta filosofía que más se ha utilizado en los trabajos sobre predicción de la insolvencia.

La utilización de algoritmos de inducción de reglas y árboles de decisión para el pronóstico de la insolvencia ofrece interesantes ventajas:

- Casi todos los algoritmos inductores presentan un buen comportamiento en presencia de variables nominales, lo cual permite enriquecer los modelos aportando nuevas perspectivas a los enfoques tradicionales.
- Al contrario de lo que ocurre con los modelos predictivos basados en redes neuronales, las reglas producto de los sistemas de inducción son entendibles por un analista humano, y además las variables sin poder clasificar son eliminadas del modelo, pues no figuran en los árboles/reglas. Todo esto permite la interpretación económica del análisis, y por lo tanto un aumento progresivo de las capacidades del equipo de analistas de riesgos de la entidad que implante el sistema.
- Asimismo, otro punto en favor de estos algoritmos es que la mayor parte de ellos pueden programarse para asignar correctamente todos los casos de ejemplo lo cual, si bien es poco útil a efectos predictivos pues implica sobre ajustar el modelo, es útil si se quiere realizar una caracterización detallada de las empresas que fracasaron en un entorno e intervalo de tiempo determinados.
- En adición, los algoritmos de aprendizaje suelen incorporar procedimientos de limitación de la complejidad del sistema, lo cual es útil para evitar el sobreajuste de los modelos a los datos de ejemplo, lo cual es muy importante en un problema como el del pronóstico de la insolvencia, donde los patrones económico-financieros que caracterizan a las empresas pertenecientes a cada grupo no van a ser estables.
- Por último, ciertos estudios (Liang et al., 1992, p. 326) apuntan que podrían conseguirse muy buenos resultados combinando las técnicas de inducción con otros procedimientos, por ejemplo, usando los árboles y reglas clasificadoras como técnica exploratoria para filtrar las variables, que después serían analizadas con procedimientos estadísticos.

Una vez revisados los principales modelos de predicción de quiebra podemos mencionar, a manera de resumen, que estos trabajos comparten una metodología estática común, consistente en la estimación transversal de los modelos. Concretamente, la

estimación se efectúa separadamente con datos de los estados financieros de uno, dos, tres, etc., años antes de que la quiebra tenga lugar. Modelos que, después, se aplican a muestras externas de validación para determinar los errores de clasificación, y que proporcionan la probabilidad de que una empresa quiebre o no. Prácticamente todos los resultados empíricos de los distintos modelos coinciden en señalar que las empresas con dificultades tienen características significativamente distintas a las sanas para períodos anteriores al fracaso, en lo que a rentabilidad y endeudamiento se refiere.

Por otra parte, es preciso señalar que los modelos de predicción presentan ciertos problemas en relación con los datos. En efecto, las estimaciones y los indicadores de distancia a la quiebra, se obtienen a partir de los últimos estados financieros publicados por la empresa en cuestión. Estados financieros que se suelen dar a conocer con posterioridad a la finalización del ejercicio. Las investigaciones que constituyen el área de estudio del fracaso financiero tienen como característica común la escasa existencia de aportaciones teóricas en las que, a partir de un modelo formalmente enunciado, se describa el proceso por el que una empresa se ve conducida a esta situación.

Giancarlo y Varetto (1993) realizaron un trabajo empírico en el que utilizaban redes neuronales. Para éstos dos autores, las redes neuronales no representaban una teoría matemática exacta y clara en comparación con las técnicas estadísticas tradicionales, sin embargo, y a pesar de estas limitaciones, subrayaron la importancia de que una de las partes integrales de su proyecto consistía en probar nuevas metodologías de análisis para las empresas crediticias a través de construir y mantener una base de datos específica que incluyera compañías con problemas financieros. Esto con el objetivo de desarrollar investigaciones mediante procesos dinámicos (en tiempo real) de quiebras y fracasos.

El método de las redes neuronales consistía en interconectar un gran número de unidades con otras, para ver si son capaces de dar resultados que son relativamente fáciles de calcular. Es decir, la idea se basaba en observar como un gran número de unidades interactúan entre sí. Altman, Giancarlo y Varetto (1994) apuntaban que las redes neuronales no requerían una previa especificación de una forma funcional lineal, ni la adopción de supuestos restrictivos acerca de las características de las distribuciones estadísticas de las variables y errores del modelo. Es decir, las redes neuronales daban la posibilidad de

trabajar con variables imprecisas, cambiando los modelos a través del tiempo para que éstos fueran capaces de adaptarse gradualmente a cada nuevo caso que se presentara. Una novedad en el estudio de estos investigadores consistió en probar por primera vez un método alternativo al análisis discriminante cuya aplicación estuviera basada en la inteligencia artificial, a través de las redes neuronales. Sus resultados les confirmaron que este sistema proporciona un diagnóstico automático y puede ser utilizado para preseleccionar negocios que deban examinarse posteriormente con más profundidad para otorgarles créditos.

Concluyeron que aunque este sistema tenía un poder predictivo considerable y era flexible, se requerían numerosas pruebas para obtener la estructura de las redes neuronales. Además, también existía la posibilidad de obtener resultados con un comportamiento ilógico dado las diferentes variaciones de los “input-valores”. Esto desde un punto de vista financiero representaba un serio problema. Por eso reconocieron la necesidad de utilizar de manera simultánea el análisis discriminante múltiple y el método de redes neuronales.

2.2 Modelos de Predicción con Variables Macroeconómicas

Además de las variables microeconómicas, también existen investigaciones relativas a estudios sobre fenómenos macroeconómicos que condicionan también el concurso de acreedores. Como se ha comentado anteriormente, la predicción de la quiebra de las empresas y del proceso concursal ha recibido gran atención por parte de los investigadores a nivel mundial por el fuerte impacto negativo que produce el cierre de empresas, tanto para los accionistas de las empresas en cuestión, como para sus empleados, prestamistas, acreedores o proveedores, suponiendo todo esto un fuerte agravio para la economía en general (Campa y Camacho-Miñano, 2014).

Los estudios macroeconómicos sobre predicción de quiebras existen, aunque en menor medida que los modelos con variables microeconómicas. Uno de los pioneros en estudiar el ámbito de los factores macroeconómicos que rodean al concurso de acreedores fue Altman (1983), que afirmó que, hasta aquel momento, pocos investigadores se habían centrado en la cara macroeconómica del conflicto de la quiebra de las empresas.

En sus resultados, pone de manifiesto que la propensión de una empresa a quebrar puede verse afectada por el efecto acumulado de cuatro variables: una reducción en el crecimiento económico del país, la bajada en el rendimiento del mercado de capitales, la caída en la disponibilidad al crédito y la baja edad de las empresas, pues a medida que aumenta provoca un decremento de su tendencia hacia la quiebra (Altman, 1983; Liu, 2004; Salman et al., 2011). Después de sus hallazgos, son muchos los autores que han continuado con esta línea de investigación (Hernández Tinoco y Wilson, 2013; Nam et al., 2008; Qu, 2008), centrándose sobre todo en el estudio de estas y otras variables macroeconómicas en países concretos. Por lo tanto, aunque la afectación de variables macroeconómicas al problema de la quiebra es un tema tratado en la literatura académica, pocos han sido los estudios comparativos a nivel de países, objetivo de esta tesis.

Una investigación ya comentada en el desarrollo de los factores microeconómicos que condicionan el fracaso empresarial es la de Laitinen y Suvas (2013). Estos autores han elaborado un estudio actual transnacional en 30 países europeos, que pone de manifiesto la posibilidad de predecir los problemas financieros de las empresas en diferentes países, así

como la comparación del grado de predicción entre países y la creación de un modelo genérico de predicción. Su estudio se lleva a cabo a través del uso de variables microeconómicas, y los autores afirman que las diferencias entre países no se producen exclusivamente por cifras de los estados contables o estados financieros, si no que su modelo se ve, en parte, deteriorado por factores como la situación económica, la legislación y la cultura, entre otros. Los autores animan a continuar por esta línea de investigación, añadiendo nuevos países o continentes o intentando minimizar los factores que deterioren la creación de un modelo de predicción transnacional común.

Son muchos los factores económicos ya analizados en la literatura previa. Por ejemplo, el sector al que pertenecen las empresas quebradas condicionan el concurso (Chava y Jarrow, 2004), y la inclusión en los modelos de predicción de algún índice de producción sectorial provoca incrementos explicativos de los modelos (Hol, 2007; Mare, 2013). De acuerdo con Acharya et al. (2007), los acreedores de empresas con problemas suelen recuperar una cantidad menor de sus créditos si dichas empresas pertenecen a un sector con poca liquidez y particularmente si además el sector posee activos muy específicos y difícilmente reubicables. También demostraron que las empresas concursadas pertenecientes a sectores que atraviesan situaciones financieras complicadas suelen finalizar su proceso concursal con una reorganización, en mayor medida que con la liquidación o la compra por parte de otra empresa. En el caso del sector manufacturero o industrial, otros estudios relacionan la quiebra empresarial con la actividad industrial o productiva, indicando que a medida que la actividad se incrementa, disminuyen los problemas de financiación de las empresas (Salman et al., 2011).

Un ejemplo evidente lo encontramos en el sector de la construcción español a raíz de la “burbuja” inmobiliaria en el año 2008. Según datos del INE (Instituto Nacional de Estadística) en España el número de empresas concursada pertenecientes al sector construcción representó casi el 40% del total de empresas de otros sectores y con relación a su mismo sector pero con respecto al año 2007 creció un 316%.

Gráfico 6. Evolución del número de Empresas Concursada del Sector Construcción en España (2005-2014)



Fuente: Elaboración Propia del Autor

La inflación es otra variable que afecta significativamente a la quiebra empresarial, sobre todo en el largo plazo (Halim et al., 2008). A modo de ejemplo, según Buehler et al. (2012), cuando el franco suizo se apreció un 1%, la ratio de quiebra se vio incrementada 8 puntos porcentuales. Estos dos movimientos traen consigo la caída de la competitividad de las firmas suizas en los mercados extranjeros, así como el consiguiente aumento de la competitividad de firmas extranjeras en ese país.

Otra variable macroeconómica es el emprendimiento y su relación con las leyes sobre el concurso de acreedores. Principalmente, los análisis defienden que las leyes no deben ser excesivamente duras con los emprendedores, ya que se pueden llegar a perder oportunidades de negocio muy rentables por el alto miedo al fracaso y a sus consecuencias (Lee et al., 2007). Si las leyes apoyan el emprendimiento, las barreras de entrada al mercado se verán reducidas y el camino se verá allanado para los emprendedores (Peng et al., 2010).

En un estudio realizado en 29 países entre los años 1990 y 2008, se ha demostrado que las leyes sobre el concurso de acreedores que no penalizan al emprendedor, favorecen el desarrollo del emprendimiento de cada país, medido a través de la ratio de entrada de nuevas empresas (Lee et al., 2011). Armour y Cumming (2008) avalaron ya esta afirmación en su estudio, añadiendo además que tanto leyes concursales más permisivas como un requerimiento de capital inicial bajo para la formación de una nueva empresa fomentan la actividad emprendedora.

Otra variable es el ciclo macroeconómico. Existe una relación entre las empresas que quiebran y el ciclo macroeconómico, con mayor impacto en los momentos de inestabilidad del ciclo (Bhattacharjee et al., 2009). Testado empíricamente con empresas que cotizan en bolsa, los cambios en el ciclo macroeconómico afectan de forma distinta a empresas que acaban de comenzar a cotizar en bolsa frente a las que llevan años en el mercado bursátil. La propensión hacia la liquidación se incrementa en el caso de empresas que han comenzado a cotizar en bolsa en momentos de reciente prosperidad económica en cuanto la situación económica empeora. Las sociedades que consiguen sobrevivir a una recesión económica suelen ser adquiridas, en vez de liquidadas, en cuanto se vislumbra algo de mejora en la economía. Otro estudio corrobora que un número más elevado de empresas tienden a liquidarse durante años de inestabilidad económica, mientras que en años de bonanza las adquisiciones son más frecuentes (Salman et al., 2011).

Otra variable es la tendencia a la exportación o no y su relación con la quiebra empresarial. Lógicamente, aquellas empresas que exportan tienen un porcentaje de quiebra más bajo que las que no exportan (Buehler et al., 2012; Melitz, 2003; Melitz y Ottaviano, 2008). Concretamente, el porcentaje disminuye un 20% si las empresas exportan menos de un tercio o más de dos tercios de sus ventas, es decir, cuando la exportación es muy baja o muy alta.

Cumming y Saini (1981) establecen que una política monetaria restrictiva provoca un incremento en el número de empresas que entran en concurso. Como se ha visto con anterioridad, Altman (1983) corrobora dichos resultados afirmando que la disponibilidad al crédito favorece la recuperación de las empresas y, por el contrario, si la oferta monetaria disminuye, aumenta el riesgo de quiebra (Salman et al., 2011). Del mismo modo, autores posteriores también defienden que la oferta monetaria es un buen predictor de la quiebra, junto con otras variables macroeconómicas, como son el output GAP y los índices de producción sectorial (Hol, 2007). También relacionada con la política monetaria, los tipos de interés se consideran un instrumento a utilizar para la toma de medidas de reducción de quiebra empresarial (Liu, 2004). La quiebra de las empresas y los tipos de interés están positivamente correlacionados, lo que implica que una disminución de los tipos de interés provoca una caída también de las empresas en situaciones financieras complicadas.

Relativo a lo anterior, encontramos el acceso al crédito de las empresas con problemas financieros, otra variable macroeconómica que influye sobre la quiebra empresarial. La disponibilidad de crédito y la quiebra empresarial mantienen una correlación negativa (Altman, 1983; Liu 2004). Cuando es difícil acceder al crédito bancario, las empresas podrían verse obligadas a financiarse internamente, lo que produciría una bajada de la producción y las inversiones y el casi inevitable camino hacia la liquidación (Liu, 2004).

Otra variable con influencia en la liquidación de las empresas es su situación geográfica. En relación a investigaciones previas, se puede afirmar que las empresas que se localizan en zonas con importante actividad empresarial, es decir, zonas metropolitanas, tienden menos a liquidarse que aquellas que se encuentran en regiones con una actividad empresarial menor o zonas no metropolitanas (Buehler et al., 2012). Este mismo autor también señala que las “condiciones económicas” de un país a su cifra de inversión pública, su porcentaje del impuesto sobre sociedades y sus datos sobre el paro, se ha demostrado que afectan a la quiebra de empresas. Cuando dichas condiciones económicas son favorables, el número de empresas que terminan en liquidación disminuye considerablemente. Esto es, con impuestos bajos, tasa de paro baja y alta inversión pública, las cifras de quiebra de empresas mejoran.

El producto interior bruto (PIB) es una variable incluida también en los estudios sobre el comportamiento de las empresas en quiebra. Según Halim et al. (2008), que llevan a cabo un estudio sobre quiebra empresarial con variables macroeconómicas, encuentran nexos muy significativos entre el PIB y la supervivencia de las entidades. En concreto, la subida del producto interior bruto afecta de forma inversa sobre la quiebra de las empresas, es decir, cuanto mayor sea el PIB mayor será la rentabilidad de las empresas y menos de ellas llegarán a verse en situaciones financieras complicadas. Adicionalmente, no sólo el PIB es una cifra que puede condicionar la situación de quiebra de las empresas. La diferencia entre el PIB potencial y el observado es una variable que afecta al riesgo de quiebra de las firmas (Hol, 2007; Carling et al., 2007) y tiene un poder explicativo muy alto sobre la quiebra. Según Hol (2007), la relación existente entre la quiebra de entidades y el GDP GAP es inversa; a medida que el GAP decrece, aumenta el riesgo de quiebra. En el estudio que elaboran Carling et al. (2007), partiendo únicamente de información financiera específica de

empresas, es decir, de variables contables y ratios financieros, el incremento en el poder explicativo del modelo al añadir variables macroeconómicas, tales como el output gap, la curva de rendimiento y las expectativas de los consumidores, es muy relevante.

Otra variable macroeconómica a tener en cuenta para la predicción de insolvencias es la regulación sobre quiebra empresarial. Se ha verificado que las leyes influyen fuertemente sobre las situaciones de concurso. Esta conexión está relacionada, por ejemplo, con la gestión bancaria, ya que los bancos establecerán mayores garantías o mayores tipos de interés sobre los préstamos que concedan si la legislación del país donde operen protege fuertemente al acreedor (Davydenko y Franks, 2008). Otra relación entre la legislación y la quiebra empresarial tratada en el mundo académico está referida a las ventajas o perjuicios de la ley sobre el emprendimiento (Lee et al., 2007; Lee et al., 2011; Peng et al., 2010).

Hasta la fecha, los estudios sobre la quiebra se han centrado principalmente en distinguir la salud de las empresas según sus características financieras, medidas a través de sus datos contables. No obstante resulta difícil explicar el fenómeno de la quiebra empresarial sin emplear otro tipo de variables que tengan en cuenta el estado de la economía en su conjunto, y no únicamente los ratios contables internos de cada empresa.

Foster (1986) sugirió que un modelo multivariantes podría incrementar su poder de predicción incorporando variables macroeconómicas. También Rosé et al. (1982) en su intento por descubrir que variables macroeconómicas están más relacionadas con la quiebra, examinaron una serie de indicadores económicos sugeridos en su mayoría de la teoría económica. Como afirma Jones (1987), podría ser útil el hecho de incorporar indicadores regionales o sectoriales en la muestra.

Varios estudios han analizado indicadores macroeconómicos para predecir el riesgo de insolvencias en los países. Sin embargo, y teniendo en cuenta la revisión de la literatura realizada para esta propuesta, consideramos que existen muchos otros factores macroeconómicos que podrían ayudarnos a mejorar la predicción global de la insolvencia. El inversor internacional necesita herramientas globales, de ahí que sea importante el estudio que proponemos. Algunos posibles factores a tener en cuenta que no han sido estudiados en este tema serían el nivel de transparencia de los países, el índice de corrupción, el nivel de “enforcement” de la legislación vigente, entre otros.

CAPITULO 3. HIPÓTESIS Y METODOLOGÍA

3.1 Hipótesis a Contrastar

Una vez analizada la literatura previa y en línea con los objetivos propuestos, se exponen dos hipótesis a contrastar en esta tesis doctoral:

H1: La aplicación de técnicas paramétricas y no paramétricas en el análisis de la quiebra empresarial no muestran diferencias significativas para las variables que mejor capacidad predictiva presentan.

H2: La presencia de variables macroeconómicas no condiciona la quiebra empresarial en un contexto global de empresas cotizadas.

3.2 Metodología de Ecuaciones de Estimación Generalizadas

El proceso de toma de decisiones a menudo se ha visto obligado a depender de métodos sub óptimos de análisis de datos debido a que las variables de respuesta generalmente no presentan una distribución normal. En este sentido los investigadores tienen que, o bien realizar transformaciones sobre la variable de respuesta antes de analizarlas, o bien usar un método de agregación de su variable de respuesta con el fin de aproximar a la normalidad la distribución de las respuestas. Sin embargo este tipo de enfoques sacrifican tanto la precisión en el análisis como la claridad en la interpretación de los resultados.

Para el investigador supone un reto adicional el analizar datos que están correlacionados entre sí, como los previstos en los estudios longitudinales en donde los datos se agrupan dentro de los subgrupos. La no incorporación de la correlación que existe entre los sujetos dentro de un modelo predictivo puede producir estimaciones incorrectas de los parámetros del modelo de regresión. La estimación de los parámetros de regresión (β s) resulta menos eficiente, es decir, se encuentran ampliamente dispersos alrededor del valor real de la población, de lo que sería si se toma en cuenta la correlación intrasujeto y se incorpora en el modelo.

A menudo las empresas se enfrentan a la evaluación de variables dependientes obtenidas de la colección de datos longitudinales. En este caso pueden utilizar un método que proporciona estimaciones eficientes e imparciales de los parámetros para el adecuado análisis de los datos sin la necesidad de que estos sean transformados y que producen resultados de fácil interpretación

Cuando el investigador se enfrenta a datos que consiste en medidas repetidas que pueden estar correlacionadas dentro de un sujeto a través de medidas repetidas dentro de una agrupación de observaciones en un determinado grupo, se debe tener en cuenta la correlación dentro de las respuestas al estimar los parámetros de regresión, de lo contrario se pueden hacer inferencias incorrectas sobre los coeficientes de regresión (debido a la estimación incorrecta de las varianzas), que podrían conducir a conclusiones equívocas.

Fitzmaurice (1995) demostró que cuando se enfrenta a una variable independiente que varía dentro de un grupo (referido como una covariable dependiente del tiempo en estudios longitudinales o datos panel), “la eficiencia de estimadores disminuye con el aumento de la correlación y el descenso es más notable cuando la correlación es mayor que 0.4”. Las pérdidas de eficiencia eran mayores según aumentaba la correlación. Los errores son especialmente grandes para los casos en que la correlación intrasujeto es muy positiva o muy negativa.

3.2.1 Datos Longitudinales o Datos Panel

En muchas ramas de la ciencia se realizan experimentos para poder medir distintas características de cierto fenómeno. Dichos experimentos pueden consistir en mediciones repetidas a lo largo de un periodo de tiempo sobre un mismo individuo, obteniendo así un historial que muestra el desarrollo o evolución de las características que se miden. Las observaciones de este tipo de experimentos constituyen una serie de tiempo. Entonces se puede definir una serie de tiempo como el conjunto de datos resultado de las observaciones repetidas durante un periodo de tiempo de alguna(s) característica(s) de interés sobre una sola unidad experimental (individuo, empresa, etc.), siguiendo a Polvo Vuelvas (2007).

Por otro lado, en los llamados estudios de corte transversal se obtiene una sola observación para cada uno de los individuos de una muestra; cada una de estas observaciones está conformada por una variable respuesta (o simplemente respuesta), que es el foco de atención en el estudio y por un conjunto de variables explicativas que son datos de las características del individuo estudiado, las cuales sirven para explicar el comportamiento de la respuesta; en otras palabras, el comportamiento de la respuesta depende de los distintos valores que tomen las variables explicativas.

Por ejemplo:

- 1) El número de horas que un niño pasa viendo la tele a la semana (variable explicativa) puede ser determinante en el riesgo de padecer problemas de lento aprendizaje (respuesta).
- 2) Las mediciones de peso, nivel de triglicéridos y colesterol en la sangre (variables explicativas), pueden constituir factores muy significativos en el desarrollo de problemas cardiacos (respuesta).
- 3) La condición socioeconómica, nivel de estudios e ingresos promedio de una persona (variables explicativas) puede ser factores determinantes para la renovación o no de una póliza de seguros (respuesta).

Los datos de corte transversal permiten modelar la respuesta de la muestra, caracterizada por los valores de las variables explicativas a fin de estimar los parámetros de población que relacionan a las variables explicativas con la respuesta.

Los datos longitudinales pueden ser vistos como la fusión del enfoque de series de tiempo y el enfoque de corte transversal. Son arreglos en los cuales se consideran varias unidades experimentales (personas, empresas, ciudades, etc.) de las cuales se registran repetidamente a lo largo de un periodo de estudio las observaciones de las respuestas de interés conjuntamente con sus variables explicativas.

Debido a la naturaleza misma de este tipo de datos, una característica que los distingue, y que debe tenerse en cuenta en la modelación, es la posible correlación dada entre las mediciones repetidas de la variable respuesta en cada individuo, considerando las mediciones entre los individuos independientes. Esto es, las mediciones son correlacionadas dentro e independientes entre individuos. Así, se pretende identificar la evolución de la variable respuesta y determinar cómo es afectada por las covariables o variable independientes. La técnica se encuadra en el análisis de regresión incluida en el conjunto de herramientas multivariantes destinadas al análisis de la dependencia entre variables, medidas todas ellas (endógena y exógenamente) preferentemente en una escala estrictamente cuantitativa.

El término de modelo de datos panel se aplica en este contexto a aquel modelo de regresión que utiliza, para la estimación de los parámetros de interés, la variabilidad temporal y transversal de los datos.

Estudio de Datos Longitudinales en Estadística

Los datos longitudinales implican mediciones repetidas en los sujetos a lo largo del tiempo, proponiendo un nuevo tipo de análisis para ese tipo de datos. En el libro de Rothman y Greenland, dentro del capítulo Introducción a la Modelación de la Regresión, afirman que los datos longitudinales son las mediciones repetidas en los sujetos a lo largo de un periodo de tiempo y que se pueden realizar para exposiciones dependientes del tiempo.

Justificación por el Interés por los Datos de Panel⁵

La utilización de datos de panel en lugar de series temporales se justifica por aprovechar la variabilidad transversal. La identificación y estimación de los parámetros de una función de respuesta explota la variación de las variables incluidas. Si las variables no presentan excesiva variabilidad temporal pero sí transversal, la aproximación con datos de panel aportaría capacidad extra para esa estimación. En algunos casos específicos, la introducción de la variabilidad transversal en procedimientos clásicos de análisis meramente temporal favorece las propiedades estadísticas de los procedimientos de inferencia tradicionales. Así, por ejemplo, la potencia y propiedades asintóticas de los test de integración y cointegración se ven claramente favorecidas cuando se combinan datos de corte temporal y transversal, lo que justifica el actual interés por las líneas metodológicas que combinan análisis de series temporales y utilización de variabilidad transversal.

⁵ Siguiendo a Mahía (2000)

Características particulares de los estudios longitudinales

Al realizar mediciones a lo largo del tiempo, el control de la calidad de los datos juega un papel fundamental. La larga duración de algunos estudios obliga a prestar especial atención al cambio de personal, al deterioro de los equipos, al cambio de tecnologías y a las inconsistencia de las respuestas de los participantes a los largo del tiempo.

Los estudios longitudinales requieren una gran cantidad de esfuerzo, pero proporcionan varios beneficios entre los que podemos destacar:

- i) Puede registrarse la reincidencia del evento.
- ii) Evaluación prospectiva de exposición.
- iii) Medición del cambio individual en la respuesta.
- iv) Separación de efectos temporales.
- v) Control por efecto de cohorte.

Retos en los estudios longitudinales

- 1) Seguimiento de los participantes. Existe un riesgo de tener un sesgo motivado por un seguimiento incompleto o abandono de los participantes en el estudio. Si los sujetos que se siguen hasta el final convenido del estudio, difieren de los sujetos con seguimiento truncado, un análisis simple puede dar resúmenes que no sean representativos de la población objetivo original.
- 2) Análisis de datos correlacionados. El análisis estadístico de datos longitudinales, requiere métodos que puedan tomar en cuenta de forma apropiada la correlación intra sujeto de las respuestas medidas. Si se ignora dicha correlación, las inferencias como pruebas de hipótesis estadísticas o intervalos de confianza, pueden carecer de validez.
- 3) Covariables dependientes del tiempo. Aunque los diseños longitudinales ofrecen la oportunidad de asociar cambios en la exposición con cambios en la respuesta de interés, la dirección de causalidad puede ser complicada por “interacción” entre las respuestas y la interacción.

Características de los datos longitudinales

La característica principal que define a un estudio de datos longitudinales, a diferencia de un estudio de corte transversal, es que los individuos son observados repetidamente a lo largo del tiempo, registrando los valores que toman las variables explicativas que intervienen en el estudio para así poder entender el comportamiento de la respuesta, la cual también es registrada a lo largo del tiempo del estudio. En este sentido, es posible estudiar los efectos del tiempo en cada individuo, así como las características globales de la población estudiada.

Los datos longitudinales requieren métodos estadísticos especiales, debido a que el conjunto de respuestas en un sujeto, comúnmente está intercorrelacionado, es decir, los resultados de cierta medición pueden depender (es decir estar correlacionados) de los resultados de alguna medición pasada. Esta correlación necesita tomarse en cuenta para obtener inferencias válidas.

La principal ventaja de un análisis longitudinal radica en su efectividad para estudiar los cambios en las respuestas. Respecto a un estudio de corte transversal, los estudios longitudinales tienen la ventaja de predecir más acertadamente el comportamiento futuro pues al tener observaciones repetidas para los individuos se puede obtener con mayor facilidad la razón instantánea de cambio de los datos, además de que se toma en cuenta la correlación entre observaciones como parte de la información que explica el comportamiento de las respuestas.

Otro mérito del estudio longitudinal es su habilidad para distinguir el grado de variación en la respuesta a lo largo del tiempo para un individuo en particular, con respecto a la variación de la respuesta poblacional.

Es importante mencionar que la elección del modelo estadístico depende del tipo de respuesta y del tipo de inferencia que se pretende hacer. Existen varias metodologías usadas para el modelo de datos longitudinales.

Las tres aproximaciones más importantes (Diggle, Heagerty, Liang y Zeger, 2004):

- 1) Modelos Marginales o análisis marginal. Sólo interesa la respuesta promedio.
- 2) Modelos de transición. Enfocados en determinar cómo la respuesta, Y_{it} , depende de los valores anteriores de ella misma y de otras variables (covariables) (i.e., un modelo condicional).
- 3) Modelos de efectos aleatorios. Aquí el enfoque es determinar cómo los coeficientes de regresión del modelo, cambian sobre los individuos.

En términos matemáticos, los datos panel (o datos longitudinales) consiste en observaciones de un corte transversal de unidades individuales (hogares, empresas, países, etc.) repetidas sobre el tiempo (Albarrán Pérez, 2011). La especificación general de un modelo de regresión con datos de panel es la siguiente:

$$Y_{it} = \alpha_{it} + X_{it} \beta + u_{it} \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T$$

Donde i se refiere al individuo o a la unidad de estudio, t a la dimensión en el tiempo, α es un vector de interceptos de n parámetros, β es un vector de K parámetros y X_{it} es la i -ésima observación al momento t para las K variables explicativas.

- En general, los datos se observan a intervalos regulares de tiempo.
- Los datos panel pueden ser balanceados ($T_i = T$ para todo i) o no balanceados $T_i \neq T$ para algún i .
- Se pueden tener paneles :
 - De muchos individuos y pocos periodos (Paneles Micro o “Short Panels”).
 - De pocos individuos y muchos periodos temporales (Paneles Macro o “Long Panels”).
 - De muchos individuos y muchos periodos temporales (Campo Aleatorio).

- Se puede hacer inferencia asintótica:
 - $NT \rightarrow \infty$
 - $N \rightarrow \infty, T \rightarrow \infty$
 - $N \rightarrow \infty, T \rightarrow \text{fijo}$
 - $T \rightarrow \infty, N \rightarrow \text{fijo}$
- Los errores estarán probablemente correlacionados (en el tiempo para un individuo y/o entre individuos).
- Se pueden tener regresores invariantes en el tiempo $x_{it} = x_i$, que no varían con los individuos $x_{it} = x_t$ o bien que varían tanto con el tiempo como con los individuos x_{it}
- Permiten la estimación de modelos dinámicos.
- Para cada observación debe conocerse el individuo i y el periodo temporal t al que se refiere.
- Para paneles balanceados describir el número de observaciones implica:
 - Número de individuos distintos N .
 - Total de periodos cubiertos por el panel T .
 - Número total de observaciones NT .
 - Para paneles no balanceados el total de observaciones $\sum_{i=1}^N T_i$
- Las variables pueden tener variación tanto en el tiempo como entre los individuos.

Como se mencionó anteriormente, los estudios longitudinales son diseñados para investigar cambios sobre el tiempo de una característica, la cual es medida repetidamente para cada uno de los participantes del estudio. En estudios médicos, las medidas podrían ser presión sanguínea, nivel de colesterol, volumen pulmonar, etc. No siempre es posible tener control completo de las circunstancias bajo las cuales se toman las medidas, y existe la posibilidad de alguna variación considerable entre los individuos respecto al número y calendarización de las observaciones, por ello se puede presentar una situación en la que el número de observaciones para un individuo es diferente del número de observaciones para otro.

Cuando el objetivo de un estudio longitudinal es analizar las características individuales, la mejor suposición es que algunos parámetros de regresión, como la ordenada y el efecto de una que otra variable explicativa podrían variar de un sujeto a otro, como por ejemplo, el efecto de un tratamiento médico varía aleatoriamente de un individuo a otro.

A menudo, los parámetros individuales tienen una interpretación natural la cual es relevante para los objetivos del estudio, y sus estimaciones pueden ser usadas para análisis exploratorio. Es aquí donde la heterogeneidad de los individuos debe tomarse en cuenta.

Los modelos de efectos aleatorios pueden ser usados fácilmente pues tienen varias características deseables como poder trabajar con datos desbalanceados, ya sea que éste desbalance sea en el número de observaciones entre individuos, o bien, en los intervalos entre las observaciones. Otra característica de estos modelos es que permite el modelado explícito y el análisis de variación entre individuos y sobre cada individuo, entre otras.

Los modelos de efectos aleatorios son más usados cuando el objetivo es hacer inferencias sobre los individuos más que sobre la población. En estos modelos se asume que existe un conjunto de variables explicativas que tienen un efecto aleatorio en la respuesta de diferentes individuos, i. e., el efecto de dichas variables, varía aleatoriamente de un individuo a otro. Los datos panel permiten controlar variables que no pueden ser observadas o medidas (por ejemplo los factores culturales o las diferencias existentes en prácticas de negocio entre compañías); o también variables que cambian conforme el tiempo pero no entre individuos (por ejemplo acuerdos internacionales, regulaciones federales, políticas nacionales, etc).

Ventajas y Desventajas de la Utilización de Datos Panel

La técnica de datos de panel presenta una serie de ventajas y desventajas en comparación con los modelos de series de tiempo y de corte transversal. Las más relevantes son las siguientes:

Ventajas:

- ❖ La técnica permite al investigador económico disponer de un mayor número de observaciones incrementando los grados de libertad y reduciendo la colinealidad entre las variables explicativas y, en última instancia, mejorando la eficiencia de las estimaciones econométricas.

- ❖ Tal y como se mencionó anteriormente, la técnica permite capturar la heterogeneidad no observable ya sea entre unidades individuales de estudio como en el tiempo. Con base en lo anterior, la técnica permite aplicar una serie de pruebas de hipótesis para confirmar o rechazar dicha heterogeneidad y cómo capturarla.
- ❖ Los datos en panel suponen, e incorporan en el análisis, el hecho de que los individuos, firmas, bancos o países son heterogéneos. Los análisis de series de tiempo y de corte transversal no tratan de controlar esta heterogeneidad corriendo el riesgo de obtener resultados sesgados.
- ❖ Permite estudiar de una mejor manera la dinámica de los procesos de ajuste. Esto es fundamentalmente cierto en estudios sobre el grado de duración y permanencia de ciertos niveles de condición económica (desempleo, pobreza, riqueza).
- ❖ Permite elaborar y probar modelos relativamente complejos de comportamiento en comparación con los análisis de series de tiempo y de corte transversal. Un ejemplo claro de este tipo de modelos, son los que se refieren a los que tratan de medir niveles de eficiencia técnica por parte de unidades económicas individuales (empresas, bancos, etc.).

Desventajas:

- ❖ En términos generales, las desventajas asociadas a la técnica de datos de panel se relacionan con los procesos para la obtención y el procesamiento de la información estadística sobre las unidades individuales de estudio, cuando esta se obtiene por medio de encuestas, entrevistas o utilizando algún otro medio de levantamiento de los datos. Ejemplos de este tipo de limitaciones son: cobertura de la población de interés, porcentajes de respuesta, preguntas confusas, distorsión deliberada de las respuestas, etc.

3.2.2 Análisis de Componentes Principales

Es habitual que las empresas comuniquen a los usuarios más de una decena de indicadores, cuando en realidad no todos ellos son necesarios. Una apropiada selección de los indicadores financieros puede ayudar a identificar con mayor facilidad las directrices posibles de la política a seguir.

Desde finales del siglo pasado se ha extendido la aplicación de la técnica estadística conocida como Análisis de Componentes Principales (ACP) y cuyo objetivo consiste en sintetizar la información, o bien la reducción de la dimensión (número de variables). Dicho de otra manera, ante un banco de datos con muchas variables, el objetivo será reducirlas a un menor número perdiendo la menor cantidad de información posible.

Un problema central en el análisis de datos multivariantes es la reducción de la dimensionalidad: si es posible describir con precisión los valores de p variables por un pequeño subconjunto $r < p$ de ellas, se habrá reducido la dimensión del problema a costa de una pequeña pérdida de información.

El ACP consiste en encontrar transformaciones ortogonales de las variables originales (ratios financieros) para conseguir un nuevo conjunto de variables no correlacionadas (componentes).

La esencia matemática de esta técnica radica en el cálculo de los autovalores y los correspondientes autovectores de las matrices cuadradas denominadas de correlaciones o de covarianzas de la matriz original

Análisis de la Matriz de Correlaciones

Un análisis de componentes principales tiene sentido si existen altas correlaciones entre las variables, ya que esto es indicativo de que existe información redundante y, por tanto, pocos factores explicarán gran parte de la variabilidad total.

Selección de los Factores

La elección de los factores se realiza de tal forma que el primero recoja la mayor proporción posible de la variabilidad original; el segundo factor debe recoger la máxima variabilidad posible no recogida por el primero, y así sucesivamente. Del total de factores se elegirán aquéllos que recojan el porcentaje de variabilidad que se considere suficiente. A éstos se les denominará componentes principales.

Análisis de la Matriz Factorial

Una vez seleccionados los componentes principales, se representan en forma de matriz. Cada elemento de ésta representa los coeficientes factoriales de las variables (las correlaciones entre las variables y los componentes principales). La matriz tendrá tantas columnas como componentes principales y tantas filas como variables.

En el presente trabajo de investigación se desarrolla un análisis de componentes principales con el objetivo de encontrar “pistas” sobre las variables que introduciremos en las iteraciones que se realizarán en la modelación de las ecuaciones de estimación generalizadas.

3.2.3 Marco Teórico Ecuaciones de Estimación Generalizadas

El enfoque Ecuaciones de Estimación Generalizadas (EEG) fue desarrollado por *Liang y Zeger (1986)* como un medio de prueba de hipótesis con respecto a la influencia de los factores en variables de respuesta con distribución binaria y otras (Poisson / Gamma / Binomial Negativa) recogida dentro de los sujetos a través del tiempo. Los Modelos de Ecuaciones de Estimación Generalizadas son una extensión de los Modelos Lineales Generalizados, que facilitan el análisis de regresión sobre variables dependientes que no están normalmente distribuidas.

El enfoque de este estudio es aquel en el que las EEG desarrolla una población modelo de media o marginal. Los modelos marginales dan una respuesta promedio para observaciones que comparten las mismas covariables como una función de las mismas. Dicho de otra manera, por cada aumento de una unidad en una covariable a través de la población, los modelos EEG le dicen al usuario cuanto cambiaría la respuesta promedio, estiman los coeficientes de regresión y errores estándar con las distribuciones de muestreo que son normales asintóticamente (*Liang & Zeger, 1986*), pueden ser aplicados para probar principales efectos e interacciones, y pueden ser usadas para evaluar variables independientes categóricas o continuas.

Los estimadores de las EEG son los mismos que se producirían para regresión por mínimos cuadrados ordinarios, cuando la variable dependiente presenta una distribución normal y se supone no correlación entre la respuesta.

El método de estimación de parámetros empleando las EEG requiere la selección de una función de enlace, una función de varianza y una matriz de correlación de trabajo. Una vez seleccionadas estas componentes, según lo indique la naturaleza de los datos, la estimación de los parámetros del modelo por EEG, corresponde a la solución para α del sistema de ecuaciones:

$$\sum_{i=1}^n \left(\frac{\partial \mu_i}{\partial \alpha} \right)^T V_i^{-1} (y_i - \mu) = 0$$

Dónde:

$\mu_i = [\mu_{i1}, \dots, \mu_{ini}]^T$ es el vector de medias asociadas con las medidas repetidas del *i-ésimo* individuo, esto es, $\mu_i = \mathbb{E}(y_{ij}), i = 1, \dots, n, j = 1, \dots, n_i, \theta$ es un estimador consistente de θ , el vector de parámetros asociado con la matriz de correlación de trabajo del *i-ésimo* individuo $R_i = R_i(\theta)$,

$$V_i = V_i(\theta) = \phi A_i^{\frac{1}{2}} R_i(\theta) A_i^{\frac{1}{2}}, \quad i = 1, \dots, n$$

es la matriz de covarianzas de trabajo, con \emptyset un parámetro de escala posiblemente desconocido, y

$$A_i = \text{diag} [V(\mu_{i1}), \dots, V(\mu_{ini})], \quad i = 1, \dots, n$$

con $V(\cdot)$ La función de varianza correspondiente.

Ajustar un modelo de EEG requiere la especificación de los siguientes tres parámetros:

- 1) La elección de la función de enlace
- 2) La distribución de la variable dependiente
- 3) La estructura de correlación de la variable dependiente

Elección de la función de enlace

Para modelar el valor esperado de la respuesta marginal para la población $\mu_i = \mathbb{E}(y_i)$ como una combinación lineal de las covariables, el usuario debe especificar una función de transformación de enlace que permitirá a la variable dependiente ser expresada como un vector de parámetros estimados (β) en la forma de un modelo aditivo. Para nuestro estudio se ha utilizado la función de enlace Logit utilizado para variables de respuesta binaria.

La distribución de una variable dependiente generalmente limita las opciones del usuario con referencia a la función de enlace utilizada, así tenemos que:

- 1) El enlace Logit es la función de enlace estándar para variables dependientes binarias. Este enlace permite a la ecuación de regresión mapear el intervalo (0,1) y es expresado como:

$$g(x) = \log\left[\frac{\mu}{(1-\mu)}\right]$$

Distribución de la Variable Dependiente

Como ya se mencionó, un segundo paso consiste en especificar la distribución de la variable respuesta o variable dependiente, de tal modo que la varianza puede ser calculada como una función de la respuesta media calculada anteriormente. EEG permiten la

especificación de distribuciones de la familia exponencial, que incluyen, la distribución normal, la normal inversa, la Binomial, la Poisson, la Binomial Negativa y la Gamma.

Al igual que en los Modelos Lineal Generalizados, la varianza necesita ser expresada como una función de la media; esto se incorpora a continuación en el cálculo de la matriz e covarianza multiplicando los componentes contra una matriz de $N \times N$ (W) con valor W_i en la diagonal que está determinado por la función de la varianza; Para distribuciones de Poisson esta cifra es μ , para datos binarios esta cifra es $\mu(1-\mu)$ y para datos con distribución normal es 1.

En el ajuste de una GEE (o cualquier modelo lineal generalizado) el usuario debe hacer todos los esfuerzos razonables para especificar correctamente la distribución de la variable de respuesta de manera que la varianza se pueda calcular eficientemente como una función de la media y los coeficientes de regresión puedan ser interpretados adecuadamente.

Generalmente el usuario tendrá algún conocimiento previo sobre la distribución de la variable de respuesta. Como regla general:

- 1) Si las respuestas son datos binarios, los usuarios deberán especificar la distribución binomial.
- 2) En los casos de respuestas de conteo, los usuarios deberán especificar la distribución Poisson.
- 3) Para analizar datos estructurados, Binomial Negativa, en los casos de gran dispersión.

Estructura de correlación de la Variable Dependiente

Un tercer paso consiste en la especificación de la forma de correlación de las respuestas intrasujetos o en el grupo al que pertenecen en la muestra. Los usuarios de EEG tienen varias opciones para especificar la forma de la matriz de correlación. Esta especificación será diferente según la naturaleza de los datos. Según *Pan (2001)* "el objetivo de la selección de una estructura de correlación es una estimación eficiente de los parámetros β ".

Aun cuando los modelos EEG son generalmente robustos a errores de la estructura de correlación, en los casos en los que dicha estructura no incorpore toda la información sobre la correlación de las mediciones dentro del panel, se podría esperar que los estimadores resultaran deficientes. A continuación se enuncian las opciones más comunes para la especificación de la estructura de correlación de los datos.

Una estructura de correlación intercambiable (exchangeable) es común utilizar este tipo de estructura cuando no existe un orden aparente para las observaciones dentro de un grupo. Este tipo de estructura resulta apropiada en situaciones en las que los datos se agrupan dentro de un tema pero no se dispone de datos en serie de tiempo.

Se puede permitir el uso de una estructura de correlación no estructurada (unstructured) para la estimación libre en la correlación intra-sujeto. Este tipo de matriz estima todas las posibles correlaciones entre las respuestas y los incluye en la estimación de las varianzas. Finalmente se puede permitir una estructura de correlación independiente en la cual se asume que las respuestas entre los sujetos son independientes entre sí; este enfoque sacrifica uno de los beneficios del uso de las EEG, pero sigue siendo útil en los modelos apropiados.

Validación del Modelo

Una de las cuestiones principales que surge durante la aplicación de un modelo es la discrepancia o desviación que existe entre éste y las observaciones de la muestra tratada; de ahí la necesidad de considerar un proceso de ajuste o validación del modelo estadístico. De aquí la lógica que siguen las distintas técnicas de validación de un modelo de datos. Una de éstas técnicas es la prueba o test Wald χ^2

Test Wald χ^2

Cada vez que hay una relación dentro o entre los datos, se puede expresar un modelo estadístico con los parámetros a ser estimados a partir de una muestra. Pues bien, el Test de

Wald es una prueba estadística paramétrica que se utiliza para poner a prueba el verdadero valor del parámetro basado en la estimación de la muestra.

En este test, la estimación de Máxima Verosimilitud $\hat{\theta}$ de cierto parámetro determinado θ , se compara con el valor propuesto θ_0 bajo la suposición de que la diferencia entre ambos seguirá aproximadamente una distribución Normal.

Normalmente, el cuadrado de la diferencia se compara con una distribución de Chi-Cuadrada; siendo el estadístico de Wald a comparar:

$$\frac{(\hat{\theta} - \theta_0)^2}{var(\hat{\theta})}$$

Alternativamente, la diferencia también puede ser comparada con una distribución Normal; por lo que el estadístico de Wald quedaría:

$$\frac{\hat{\theta} - \theta_0}{se(\hat{\theta})}$$

donde $se(\hat{\theta})$ es el error estándar de la estimación de Máxima Verosimilitud.

El Test de Wald sobre varios parámetros a la vez se lleva a cabo utilizando una Matriz de Varianza. Así mismo, se puede utilizar en una gran variedad de diferentes modelos, incluyendo modelos que consideren variables tanto dicotómicas como variables continuas

Aplicación de Ecuación de Estimación Generalizadas

Para la aplicación de un modelo de EEG, se pueden definir los siguientes pasos:

1) Análisis Preliminar: Esta es la primera fase en la construcción de un modelo de Ecuaciones de Estimación Generalizadas, es decir la etapa de preparación de los datos. En esta etapa se realiza un análisis exploratorio de las variables a considerar dentro del modelo. Los análisis a los que nos referimos son:

- Análisis de Distribuciones (Estadística Descriptiva)

- Análisis Bivariados (Análisis de Correleaciones)
- Análisis de Componentes Principales

2) Iteraciones del Modelo: En esta fase, se recurre a la selección adecuada de factores que mejor se ajustan a los datos; y por tanto, hacer uso recurrente del diagnóstico de las hipótesis y parámetros del modelo. Es importante tener en cuenta que no existe un único modelo válido que se pueda ajustar a la muestra de datos analizados. Es decir, la mayoría de las veces, existe más de un modelo posible. Por ese motivo, el tema más complicado es saber y comprobar cuál se ajusta mejor y por lo tanto, es el más adecuado de todos ellos. Cuando existen una gran cantidad de variables posibles que considerar dentro de un modelo, puede ser complicado iniciar el proceso de iteración considerando todos y cada una de las variables explicativas. En estos casos, es recomendable seleccionar un conjunto de factores que se consideren importantes; y de manera inversa, incluir uno a uno de los factores excluidos al modelo y medir el nivel de significancia que se gana con dicha inclusión. Es aquí donde la aplicación de la técnica de componentes principales resulta de gran utilidad proporcionando “pistas” sobre las variables significativas.

Iteraciones manuales del modelo analizando:

- La significancia de cada factor en el modelo, eliminando cada vez el menos significativo, considerando un cierto rango o límite mínimo de aceptación determinado por el usuario
- La significancia de cada factor no incluido en el modelo, comparándolo con un nuevo modelo que sí contenga el factor potencial a incluir

Repetir estos 2 pasos tantas veces como sea necesario hasta que todos los factores resulten significativos y todos los factores no incluidos en el modelo sean los de poca significancia

Además de tener en cuenta la significancia de las variables modelados, existen otras pruebas de diagnóstico del modelo, las cuales permiten la adecuación de otros supuestos del modelo que deben evaluarse, tal es el caso del antes mencionado test Wald χ^2 . Entre mayor resulte este estadístico, mayor será el poder predictivo que presente el modelo en su conjunto con las variables seleccionadas.

3.3 Metodología de Árboles de Decisión

La Inteligencia Artificial es una disciplina que se dedica a la construcción de programas informáticos capaces de realizar trabajos inteligentes. Hoy en día esta técnica es aplicada a numerosas actividades del quehacer humano. Dichos programas informáticos manifiestan cierta forma de inteligencia, por ejemplo:

- Sistemas que aprenden nuevos conceptos y tareas
- Sistemas que pueden razonar y derivar conclusiones útiles acerca del mundo que nos rodea
- Sistemas que pueden comprender un lenguaje natural o percibir y comprender una escena visual
- Sistemas que realizan otro tipo de actividades que requieren de inteligencia humana

Las técnicas de Inteligencia artificial se basan en el aprendizaje a partir de los datos a través de su semejanza con un pensamiento estructurado similar al comportamiento humano. Siendo tan amplio su campo de acción, reúne varias áreas de investigación donde ha sido utilizada; donde uno de éstos es el reconocimiento de patrones con el propósito de extraer información que permita establecer propiedades y características de cierto conjunto de objetos.

O' Leary (1995) indica que los sistemas inteligentes pueden construirse a partir de dos enfoques:

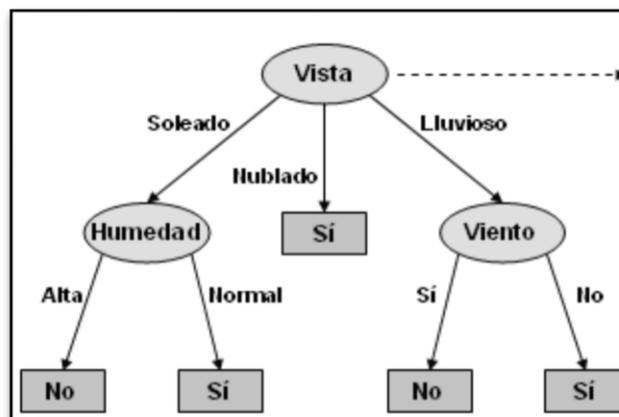
- **Sistemas Expertos:** Consiste en introducir en el ordenador, el conocimiento que los expertos humanos han ido acumulando a lo largo de su vida profesional; la mayor limitación de este enfoque radica en el proceso de captación de la información, la cual, ha de hacerse mediante una serie de entrevistas a los expertos.
- **Aprendizaje Automático (Machine Learning).** Este enfoque consiste en la elaboración de programas de ordenador que sean capaces de generar conocimiento a través del análisis de los datos y posteriormente utilizar dicho conocimiento para realizar inferencias sobre nuevos datos (Mena, 1996). Dentro de las técnicas aplicables de este enfoque encontramos:

- i. Redes Neuronales Artificiales.
- ii. Algoritmos de Inducción de Reglas
- iii. Árboles de Decisión

Un árbol de decisión puede interpretarse (Molina y García, 2006) como una serie de reglas compactadas para su representación en forma de árbol y lo que diferencia un árbol de decisión de otro es el algoritmo que lo genera.

De manera sencilla podríamos decir que los árboles de decisión permiten representar de forma gráfica una serie de reglas sobre la decisión que se debe tomar en la asignación de un valor de salida a un determinado registro.

Figura 1. Ejemplo de resultados de un árbol de decisión



Fuente: Molina y García (2006)

Seguendo a Gervilla et al. (2009) citaremos a continuación algunas de las ventajas más importantes de los árboles de decisión, que son:

- i) **Carácter Descriptivo:** Permite entender e interpretar fácilmente las decisiones tomadas por el modelo, ya que tenemos acceso a las reglas que se utilizan en la tarea predictiva. Es posible derivar fácilmente reglas de decisión (para cada rama terminal) siguiendo las rutas marcadas en la estructura del árbol que llevan a un determinado nodo hoja (la decisión del modelo). Las reglas de decisión son un conjunto de sentencias lógicas de la siguiente forma:

“Si....”se verifican unas condiciones.....”entonces” ...se toma determinada decisión

- ii) Las reglas de decisión proporcionadas por un modelo de árbol pueden usarse con fines predictivos y no sólo descriptivos. Esto es así porque se puede evaluar su precisión a partir de unos datos independientes (datos test o muestra de test que se puede utilizar para validar el modelo) a los utilizados en la construcción del modelo (datos o muestra de entrenamiento utilizada para estimar el modelo).
- iii) Permiten trabajar con un número elevado de variables de entrada y no necesitan para trabajar un número elevado de datos.
- iv) Su carácter estrictamente no paramétrico presenta una indudable ventaja para su aplicación al ámbito financiero ya que muchas de las variables económico-financieras, utilizadas habitualmente en los análisis, no suelen cumplir las hipótesis requeridas por las técnicas estadísticas.

La principal desventaja que presenta este tipo de algoritmos es que el grado de precisión de los mismos no es tan elevado como en otros algoritmos o que pueden a veces generarse árboles muy grandes y difíciles de interpretar.

Existen una serie de algoritmos desarrollados desde los principios de los años sesenta para la construcción de árboles de decisión. Así algoritmos que desarrollan árboles de decisión son: CART (Breiman et al., 1984), ACLS (Paterson y Niblett, 1982), ASSISTANT (Cestnik et al., 1987) etc. Muchos de estos desarrollos también se han convertido en herramientas comerciales, por ejemplo, RuleMaster (1984), Ex-Tran (1984) y Expert-Ease (1983).

En el presente estudio se ha seleccionado como técnica explicativa el árbol de decisión C4.5. El algoritmo desarrollado por Quinlan (Quinlan, 1986; 1993) e implantado en el C4.5 (o su versión comercial C5.0) es probablemente el más popular de entre todos los algoritmos de árboles de decisión y uno de los más eficientes. Hoy en día las herramientas comerciales como SAS, Clementine, etc. o herramientas libres como WEKA incorporan este algoritmo o variantes del mismo lo que justifica, junto con los motivos anteriormente expuestos, su elección.

El algoritmo C4.5

A continuación se expone el probablemente más popular de entre todos los algoritmos de aprendizaje de árboles de clasificación a partir de un conjunto de datos de ejemplo: el algoritmo C4.5. Seguiremos en nuestra exposición en su mayor parte el trabajo de Díaz et al. (2009).

En él, el criterio utilizado para hacer las particiones se apoya en una serie de conceptos procedentes de la Teoría de la Información y ha experimentado a lo largo del tiempo una serie de notables mejoras. La idea central que comparte con otros algoritmos similares es la de tomar en cada rama del árbol para hacer la correspondiente partición aquella variable que proporciona más información de cara a clasificar los elementos que constituyen el conjunto de entrenamiento, conjunto de datos usados para construir el árbol.

Dicho algoritmo permite:

1. Empleo del concepto razón de ganancia (Gain Ratio). Nos entretendremos detalladamente en esta cuestión.

La información que proporciona un mensaje o la realización de una variable aleatoria x es inversamente proporcional a su probabilidad (Reza, 1994). Resulta habitual en Ingeniería de Comunicaciones o en Estadística medir esta cantidad en bits, que se obtiene como $\log_2 \frac{1}{p_x}$. El promedio de esta magnitud para todas las posibles ocurrencias de la variable aleatoria x recibe el nombre de entropía de x :

$$H(x) = \sum_x p(x) \log_2 \frac{1}{p_x}$$

En consecuencia, la entropía es una medida de la aleatoriedad o incertidumbre de x o de la cantidad de información que, en promedio, nos proporciona el conocimiento de x .

De forma análoga se puede definir la entropía conjunta de dos variables aleatorias x e y como la cantidad de información que, en promedio, nos proporciona el conocimiento de x e y :

$$H(x, y) = \sum_{x,y} p(x, y) \log_2 \frac{1}{p(x,y)}$$

La entropía condicional de x dada y , $H(x/y)$, se define como:

$$H(x/y) = \sum_{x,y} p(x,y) \log_2 \frac{1}{p(x|y)}$$

La entropía condicional es una medida de la incertidumbre respecto a x cuando se conoce y . Representa la cantidad de información que se necesita para conocer plenamente x cuando ya se tiene la información suministrada por y . Se cumplirá siempre que $H(x/y) \leq H(x)$, pues al conocer y tenemos más información que nos puede ayudar a reducir la incertidumbre sobre x .

A esa reducción de incertidumbre se la denomina información mutua entre x e y : $I(x; y) = H(x) - H(x/y)$, que es la información que una de las variables nos transmite sobre la otra. Además se verifica que $I(x; y) = I(y; x)$, siendo la información mutua una magnitud similar a la covarianza.

Inicialmente Quinlan seleccionaba para hacer cada partición aquella variable y_i que proporcionaba la máxima información sobre x , es decir, maximizaba $I(x; y_i)$ (magnitud denominada Gain). Sin embargo, maximizar "gain" proporciona buenos resultados, pero introduce un sesgo en favor de las y_i con muchos valores distintos. Para corregir este sesgo, en las versiones posteriores del algoritmo se selecciona aquella y_i que maximiza la magnitud $\frac{I(x; y_i)}{H(y_i)}$ (denomina Gain Ratio). Se define como el porcentaje de la información proporcionada por y_i que es útil para conocer x .

Adicionalmente, para evitar seleccionar un atributo simplemente porque su entropía $H(y_i)$ sea pequeña, lo que aumentaría el valor del cociente anterior. Para evitarlo se exige además que $I(x; y_i)$ sea razonablemente grande.

Si el procedimiento descrito, se aplica de manera reiterada se va construyendo el árbol de decisión; hasta que se alcanza la pureza del nodo entendiendo como nodo puro aquél al que sólo corresponden casos pertenecientes a una de las clases del problema, o cuando la ramificación del árbol ya no suponga ninguna mejora.

En el algoritmo C4.5, para facilitar la comprensión del árbol, se puede realizar una *poda* del mismo. El proceso de *poda* comienza en los *Nodos Hoja* y recursivamente continúa hasta llegar al *Nodo Raíz*. En consecuencia, tras la *poda* del árbol, éste ganará en capacidad de generalización, a costa de reducir el grado de pureza de sus hojas (*Hernández et al., 2004; Larose, 2005*). Es decir, se obtienen modelos más generales pero aumentando el error de clasificación.

Para aplicar el algoritmo C4.5 a nuestra muestra tenemos varias posibilidades. Se puede descargar gratuitamente la versión C4.5 Release 8 desde la página de Ross Quinlan (<http://www.rulequest.com/Personal/>). Este programa, una vez compilado, puede ser ejecutado en sistemas operativos Unix. Además, existen nuevas versiones comerciales del algoritmo (C5.0 para Unix y See5 para Windows) que implementan mejoras y funcionalidades adicionales y se comercializan por su creador, Quinlan, (RULEQUEST RESEARCH) o a través de paquetes de minería de datos como Clementine, aunque también hay versiones de demostración gratuitas limitadas a bases de datos pequeñas (<http://www.rulequest.com/>).

En cuanto al sistema de inducción de árboles de decisión del C5.0, parece ser esencialmente el mismo que en C4.5. Sin embargo, la inducción de reglas con las nuevas versiones es diferente y más rápida. En nuestro estudio no induciremos reglas de decisión (donde sustancialmente están las mejoras del C5.0) por lo que la utilización del C4.5 es suficiente.

Concretamente el análisis lo realizaremos con el algoritmo J48 es la implementación en Java de libre acceso del algoritmo C4.5 y que contiene la herramienta de minería de datos WEKA. WEKA es el paquete de minería de datos desarrollado por la Universidad de Waikato (Witten y Frank, 2005).

CAPITULO 4. ESTUDIO EMPÍRICO

4.1 Tratamiento de la Información

La base de datos resulta ser uno de los elementos más importantes para el éxito del desarrollo de los modelos. Este capítulo muestra el desarrollo empírico del modelo analizado con base en las 119 empresas que conformaron los índices IPC, IBEX35 y EURO STOXX 50 respectivamente durante los años 2010, 2011, 2012, 2013 y 2014 y que fueron extraídos de la aplicación financiera Bloomberg.

Tabla 3. Relación de Empresas Consideradas en el Estudio

Id	Nombre de la Empresa	Sector	Índice
1	Anheuser-Busch InBev NV	Consumo, No Cíclico	EUROSTOXX50
2	Air Liquide SA	Materias Primas	EUROSTOXX50
3	Allianz SE	Financiero	EUROSTOXX50
4	ASML Holding NV	Tecnológico	EUROSTOXX50
5	BASF SE	Materias Primas	EUROSTOXX50
6	Bayer AG	Consumo, No Cíclico	EUROSTOXX50
7	Banco Bilbao Vizcaya Argentaria SA	Financiero	EUROSTOXX50
8	Bayerische Motoren Werke AG	Consumo, Cíclico	EUROSTOXX50
9	Danone SA	Consumo, No Cíclico	EUROSTOXX50
10	BNP Paribas SA	Financiero	EUROSTOXX50
11	Carrefour SA	Consumo, No Cíclico	EUROSTOXX50
12	CRH PLC	Industria	EUROSTOXX50
13	AXA SA	Financiero	EUROSTOXX50
14	Daimler AG	Consumo, Cíclico	EUROSTOXX50
15	Deutsche Bank AG	Financiero	EUROSTOXX50
16	Vinci SA	Industria	EUROSTOXX50
17	Deutsche Telekom AG	Comunicaciones	EUROSTOXX50
18	Essilor International SA	Consumo, No Cíclico	EUROSTOXX50
19	Enel SpA	Servicio Público	EUROSTOXX50
20	Eni SpA	Energía	EUROSTOXX50
21	E.ON SE	Servicio Público	EUROSTOXX50
22	TOTAL SA	Energía	EUROSTOXX50
23	Orange SA	Comunicaciones	EUROSTOXX50
24	Assicurazioni Generali SpA	Financiero	EUROSTOXX50
25	Societe Generale SA	Financiero	EUROSTOXX50
26	GDF Suez	Servicio Público	EUROSTOXX50
27	Iberdrola SA	Servicio Público	EUROSTOXX50
28	ING Groep NV	Financiero	EUROSTOXX50
29	Intesa Sanpaolo SpA	Financiero	EUROSTOXX50
30	Inditex SA	Consumo, Cíclico	EUROSTOXX50

Id	Nombre de la Empresa	Sector	Índice
31	LVMH Moet Hennessy Louis Vuitton SE	Consumo, Cíclico	EUROSTOXX50
32	ArcelorMittal	Materias Primas	EUROSTOXX50
33	Muenchener Rueckversicherungs-Gesellschaft AG in Muenchen	Financiero	EUROSTOXX50
34	Nokia OYJ	Comunicaciones	EUROSTOXX50
35	L'Oreal SA	Consumo, No Cíclico	EUROSTOXX50
36	Koninklijke Philips NV	Industria	EUROSTOXX50
37	Repsol SA	Energía	EUROSTOXX50
38	RWE AG	Servicio Público	EUROSTOXX50
39	Sanofi	Consumo, No Cíclico	EUROSTOXX50
40	Banco Santander SA	Financiero	EUROSTOXX50
41	SAP SE	Tecnológico	EUROSTOXX50
42	Cie de Saint-Gobain	Industria	EUROSTOXX50
43	Siemens AG	Industria	EUROSTOXX50
44	Schneider Electric SE	Industria	EUROSTOXX50
45	Telefonica SA	Comunicaciones	EUROSTOXX50
46	UniCredit SpA	Financiero	EUROSTOXX50
47	Unibail-Rodamco SE	Financiero	EUROSTOXX50
48	Unilever NV	Consumo, No Cíclico	EUROSTOXX50
49	Vivendi SA	Comunicaciones	EUROSTOXX50
50	Volkswagen AG	Consumo, Cíclico	EUROSTOXX50
51	Abertis Infraestructuras SA	Consumo, No Cíclico	IBEX35
52	Abengoa SA	Industria	IBEX35
53	ACS Actividades de Construccion y Servicios SA	Industria	IBEX35
54	Acerinox SA	Materias Primas	IBEX35
55	Amadeus IT Holding SA	Tecnológico	IBEX35
56	Acciona SA	Industria	IBEX35
57	Banco Bilbao Vizcaya Argentaria SA	Financiero	IBEX35
58	Bankinter SA	Financiero	IBEX35
59	Bolsas y Mercados Espanoles SHMSF SA	Financiero	IBEX35
60	CaixaBank SA	Financiero	IBEX35
61	Distribuidora Internacional de Alimentacion SA	Consumo, No Cíclico	IBEX35
62	Endesa SA	Servicio Público	IBEX35
63	Enagas SA	Servicio Público	IBEX35
64	Fomento de Construcciones y Contratas SA	Industria	IBEX35
65	Ferrovial SA	Industria	IBEX35
66	Gas Natural SDG SA	Servicio Público	IBEX35
67	Grifols SA	Consumo, No Cíclico	IBEX35
68	International Consolidated Airlines Group SA	Consumo, Cíclico	IBEX35
69	Iberdrola SA	Servicio Público	IBEX35
70	Indra Sistemas SA	Tecnológico	IBEX35
71	Inditex SA	Consumo, Cíclico	IBEX35
72	Mapfre SA	Financiero	IBEX35
73	ArcelorMittal	Materias Primas	IBEX35
74	Obrascon Huarte Lain SA	Industria	IBEX35
75	Banco Popular Espanol SA	Financiero	IBEX35

Id	Nombre de la Empresa	Sector	Índice
76	Red Electrica Corp SA	Servicio Público	IBEX35
77	Repsol SA	Energía	IBEX35
78	Banco de Sabadell SA	Financiero	IBEX35
79	Banco Santander SA	Financiero	IBEX35
80	Sacyr SA	Industria	IBEX35
81	Telefonica SA	Comunicaciones	IBEX35
82	Mediaset Espana Comunicacion SA	Comunicaciones	IBEX35
83	Tecnicas Reunidas SA	Industria	IBEX35
84	Viscofan SA	Consumo, No Cíclico	IBEX35
85	Arca Continental SAB de CV	Consumo, No Cíclico	IPC
86	Alfa SAB de CV	Diversos	IPC
87	Alpek SAB de CV	Materias Primas	IPC
88	Alsea SAB de CV	Consumo, Cíclico	IPC
89	America Movil SAB de CV	Comunicaciones	IPC
90	Grupo Aeroportuario del Sureste SAB de CV	Industria	IPC
91	TV Azteca SAB de CV	Comunicaciones	IPC
92	Grupo Bimbo SAB de CV	Consumo, No Cíclico	IPC
93	Bolsa Mexicana de Valores SAB de CV	Financiero	IPC
94	Cemex SAB de CV	Industria	IPC
95	Grupo Comercial Chedraui SA de CV	Consumo, Cíclico	IPC
96	Gentera SAB de CV	Financiero	IPC
97	Grupo Elektra SAB DE CV	Consumo, Cíclico	IPC
98	Fomento Economico Mexicano SAB de CV	Consumo, No Cíclico	IPC
99	Grupo Aeroportuario del Pacifico SAB de CV	Industria	IPC
100	Corp GEO SAB de CV	Industria	IPC
101	Grupo Financiero Inbursa SAB de CV	Financiero	IPC
102	Grupo Financiero Banorte SAB de CV	Financiero	IPC
103	Grupo Mexico SAB de CV	Materias Primas	IPC
104	Grupo Modelo SAB de CV	Consumo, No Cíclico	IPC
105	Gruma SAB de CV	Consumo, No Cíclico	IPC
106	Desarrolladora Homex SAB de CV	Industria	IPC
107	Empresas ICA SAB de CV	Industria	IPC
108	Industrias CH SAB de CV	Industria	IPC
109	Kimberly-Clark de Mexico SAB de CV	Consumo, No Cíclico	IPC
110	Coca-Cola Femsa SAB de CV	Consumo, No Cíclico	IPC
111	Genomma Lab Internacional SAB de CV	Consumo, No Cíclico	IPC
112	El Puerto de Liverpool SAB de CV	Consumo, Cíclico	IPC
113	Mexichem SAB de CV	Materias Primas	IPC
114	Minera Frisco SAB de CV	Materias Primas	IPC
115	OHL Mexico SAB de CV	Consumo, No Cíclico	IPC
116	Industrias Penoles SAB de CV	Materias Primas	IPC
117	Grupo Televisa SAB	Comunicaciones	IPC
118	Urbi Desarrollos Urbanos SAB de CV	Industria	IPC
119	Wal-Mart de Mexico SAB de CV	Consumo, Cíclico	IPC

Fuente: Elaboración Propia del Autor

4.1.1 Determinación de la Muestra

En estadística inferencial elegir una muestra representativa de la población permite con los resultados llevar a cabo generalizaciones sobre el conjunto total de sus elementos. De manera general en los modelos predictivos se procura que sus características sean definidas lo más específica y claramente con el fin de conocer el alcance de los resultados con mayor precisión. Zmijewski (1984) considera que en los casos de los modelos Logit y Probit, todos los coeficientes están afectados por la no aleatoriedad de la muestra, y por lo tanto, el hecho de que la selección de la muestra sea no aleatoria no supone una variación significativa en los resultados obtenidos en las investigaciones.

Este mismo autor examinó dos vías para estimar las muestras en los modelos predictivos:

La primera alternativa se refiere a seleccionar a las empresas de una muestra, observando primero la variable respuesta o dependiente y basándose en el conocimiento de la probabilidad de que una empresa entre en dicha muestra con base en los atributos de la variable dependiente.

La segunda alternativa es utilizando “datos completos”. Los resultados obtenidos mostraban claramente que la utilización de cualquiera de esta vías en general no afecta la inferencia estadística o los porcentajes de clasificación del modelo financiero predictivo.

La unidad de análisis en los modelos predictivos comprende a las empresas muestrales. En este punto lo que resulta más importante es seleccionar las empresas comparables entre sí. Para la selección de la empresas en el presente estudio, y dada la escasez de información y las dificultades enfrentadas para la obtención de los datos, se consideró que al principio de cualquier trabajo no se debe establecer como requisito indispensable efectuar un proceso de emparejamiento entre empresas sanas y fracasadas, pues se ha observado a través de la literatura que varios estudios revelan que el realizar dicho emparejamiento, aun siendo del mismo sector y tamaño, no necesariamente proporcionan mayor representatividad ni incorporan elemento significativos o variable explicativas determinantes en los modelos predictivos.

Para la realización del presente trabajo se consideró el segundo de los enfoques y fueron elegidas la totalidad de las empresas (119) que conformaban los índices de referencia en el sin conocimiento previo de su situación financiera, es decir:

IPC de México (35)

IBEX35 de España (34 empresas)

EURO STOXX 50 de la Zona Euro (50 empresas)

La unidad geográfica se refiere a la selección de la región, o país al que pertenecen las unidades de análisis. En la mayoría de las investigaciones realizadas se ha optado por incluir muestras pertenecientes a un único país o región, sin embargo en esta investigación se ha ampliado la unidad geográfica a varios países considerando el lugar donde las empresas realizan su actividad económica.

La unidad temporal comprende el periodo de tiempo de la base de datos. Normalmente estos periodos son anuales y varían desde los 3 y hasta los 10 años. En el presente estudio se contaba con los datos acumulados al último trimestre de cada año desde 2010 y hasta 2014 por lo que se dispone de un horizonte temporal de 5 años.

4.1.2 Determinación de la Variable Dependiente

Cualquier término que sea seleccionado como variable dependiente o variable respuesta, representará siempre una serie de dificultades conceptuales, jurídicas, técnicas y de criterios financieros, por solo mencionar algunos, que no permiten conseguir una definición única que resulte satisfactoria y sea aceptada generalizadamente. La determinación de la variable dependiente es un tema de múltiples divergencias y contradicciones. El análisis de la quiebra o fracaso empresarial recibe varios nombres y definiciones que originan desacuerdos en cuanto al momento y a los indicadores que son utilizados para declarar dicho estado.

Según Altman E. (1968) podría considerarse que el fracaso económico de una empresa tiene su origen en el momento en que los propietarios reciben una rentabilidad por sus inversiones menor que la rentabilidad del mercado, ante una misma situación de riesgo.

La primera etapa y quizá la más clara dentro del fracaso o quiebra empresarial es la *insolvencia o quiebra técnica*. Este término se presenta cuando no se tiene liquidez para cumplir con los pasivos vencidos. De no lograr superarse este estado, los pasivos exigibles tienden a ser superiores a los activos totales, y por consecuencia el capital contable será negativo, llegando así a un estado de insolvencia definitiva en la que el total de los pasivos de la empresa es superior a los activos totales.

En la siguiente tabla se muestran diferentes definiciones de fracaso o quiebra empresarial, utilizados por los autores estadounidenses, y que son representativas de las que se han utilizado en los distintos trabajos realizados sobre la predicción de quiebra siguiendo a Mora Enguidanos (1994).

Tabla 4. Definiciones de Quiebra / Fracaso Empresarial

Autor	Término	Definición
Altman	Quiebra	Aquellas empresas que se encuentran legalmente en quiebra.
Beaver	Fracaso	La incapacidad de la empresa para atender sus obligaciones financieras a su vencimiento.
Blum	Fracaso	Incapacidad de pagar las deudas por parte de la empresa, entrado en un proceso de quiebra o en un acuerdo para reducir dichas deudas.
Deakin	Fracaso	Empresas que se encuentran en situación de quiebra, insolvencia, o fueron liquidadas a beneficio de los acreedores.
Taffler	Fracaso	Liquidación voluntaria, orden legal de liquidación o intervención estatal.
Zmijewski	Fracaso	Solicitar la quiebra

Fuente: Elaboración Propia del Autor

Según Segovia-Vargas (2003), las causas del fracaso de las empresas son muy diversas y generalmente son varias las que concurren para llegar a una situación de crisis. En ocasiones el origen del problema es externo (indicios del entorno); también puede ser

interno (indicios internos) y a menudo influyen ambos tipos, aunque siempre hay que señalar que con independencia del número y origen de los problemas el equipo directivo de la empresa tiene la responsabilidad de abordarlos y encontrar soluciones. Dado que las causas que pueden originar el fracaso empresarial son muchas y de diversa índole, la intención del presente trabajo será detectar indicios o indicadores provenientes de la información contenida en los estados financieros. De acuerdo con Gelashvili, Camacho Miñano y Segovia Vargas (2015), los ratios o datos contables eran la base de los estudios que intentaban predecir y, por tanto, explicar el fracaso empresarial.

Los primeros trabajos consistían en analizar los casos de fracasos empresariales reales y, utilizando un método inductivo, aprender las características de las empresas fracasadas comparándolas con otras “sanas”. Sin embargo, aún hoy en día, y con más de medio siglo de investigaciones posteriores, todavía no existe unanimidad entre los investigadores sobre cuáles son las variables que mejor explican la situación de insolvencia ni la reorganización de empresas.

Dada la problemática expuesta, y ante la diversidad de definiciones en el presente estudio hemos clasificado a las empresas en dos categorías, sanas e insolventes⁶, en función de la puntuación obtenida al utilizar el Z-score de Altman (1968), indicador que sigue vigente desde hace casi cincuenta años.

Tras aplicar este indicador a las 119 empresas del estudio y utilizando la información proveniente de sus estados financieros, obtenemos una clasificación según se muestra en la siguiente tabla.

Tabla 5. Clasificación de empresas (Según Puntuación Z-SCORE ALTMAN)

Año	NO Quiebra	Quiebra	No Clasificada
2010	46	38	35
2011	43	45	31
2012	43	46	30
2013	46	41	32
2014	46	41	32

Fuente: Elaboración Propia del Autor

⁶ Empresas que presentan dificultades financieras según este indicador y que por lo tanto tienen un elevado riesgo de fracaso.

Un promedio del 27% de las empresas no han podido ser clasificadas según el Z-score debido a la falta de datos contables. Sin embargo, del restante 73% de empresas que sí han sido clasificadas, se obtiene un 51% de empresas Sanas y un 49% de empresas en Quiebra. Con esto se concluye que partimos de una muestra equilibrada o emparejada entre empresas sanas y no sanas para cada uno de los años dentro de la serie de tiempo analizada.

4.1.3 Determinación de las Variables Independientes

Sin lugar a dudas la selección de las variables que serán utilizadas en cualquier modelo resulta de especial importancia y atención por parte de los investigadores, ya que con base en ellas se podrán realizar conclusiones adecuadas y certeras. Los análisis contables tradicionales⁷, proponen el uso de ratios financieros para la evaluación de la situación y evolución económica de las empresas, sin embargo tienen una limitada capacidad para dar solución al problema sobre el cálculo de la liquidez, la rentabilidad, el endeudamiento, etc., y en general sobre la determinación del éxito o fracaso de una empresa. Para Brealey y Myers (1999), el utilizar ratios o razones financieras tiene la ventaja de no verse abrumado por el gran volumen de información y datos que contienen los estados financieros.

Según Segovia Vargas, M.J. (2003), una ventaja que presenta la utilización de los ratios es que reducen la dispersión en las cifras de los estados financieros de las empresas derivada del tamaño de las mismas. Esto facilita la comparación interempresas. Las empresas de gran tamaño producen grandes números contables y las de pequeño tamaño números contables menores. Por tanto, la utilización de ratios presenta la ventaja de reducir el sesgo que podría producirse derivado del tamaño de las empresas.

El incluir ratios en los primeros estudios predictivos tuvo como objetivo principal, investigar en qué medida éstos representaban herramientas valiosas y adicionales para el análisis de la solvencia a largo plazo para la posible predicción de una quiebra. Dentro del esquema multivariado, en general se han basado la selección de variables explicativas o independientes (ratios), en un fundamento teórico que permita en mayor medida una interpretación lógica de los resultados para la mejor comprensión de las causas del proceso de quiebra. Un problema que se presenta con frecuencia es cuando hay que realizar la selección y clasificación de cada ratio, pues es importante considerar en primera instancia que es lo que se desea medir y darle un nombre a la magnitud medida.

Según Tascón y Casado (2012), la ausencia de un modelo teórico formal de las relaciones entre el fracaso de la empresa, los factores económico-financieros internos o

⁷ Son los utilizados normalmente por la mayor parte de los analistas financieros tales como el análisis vertical y el análisis horizontal.

externos a la empresa, los intereses económicos de los distintos agentes relacionados y las formas de actuar de los gestores, ha sido la causa de que las variables que se incluyen en los modelos se seleccionen cada vez menos a partir del razonamiento económico y cada vez más a partir de la experiencia empírica previa.

En cuanto a la selección de los ratios financieros, en el presente estudio se realizó tomando en consideración los siguientes elementos:

- 1) Los datos extraídos de los estados financieros que disponíamos para realizar el estudio.
- 2) La revisión bibliográfica sobre un considerable número de artículos en esta línea de investigación.
- 3) El número de trabajos en los que fueron utilizados con mayor frecuencia los ratios financieros.

Con el fin de determinar el poder determinar el papel que tienen las variables macroeconómicas en la explicación y la predicción de la quiebra empresarial, el presente trabajo de investigación considera dos tipos de variables explicativas:

- i) Ratios contables clásicos (véase tabla 6)

Tabla 6. Variables Microeconómicas Seleccionadas (Ratios Financieros)

Variable	Nº Trabajos	Factor
Activo Circulante / Total Pasivo	1	Liquidez
Total Pasivo / Total Activo	18	Endeudamiento
Total Pasivo / Total Capital Contable	3	Endeudamiento
EBIT / Ingresos	1	Rentabilidad
EBIT / Total Capital Contable	5	Rentabilidad
Utilidad Neta / Ingresos Netos	2	Rentabilidad
Utilidad Neta / Total Capital Contable	6	Rentabilidad
Utilidad Neta / Total Activos	14	Rentabilidad
Ingresos / Activo Fijo	5	Eficiencia

Fuente: Elaboración Propia del Autor

ii) Variables macroeconómicas

En relación a las variables macroeconómicas (véase tabla 7) dadas las características de los datos del presente estudio (datos longitudinales) donde contamos con series temporales (2010, 2011, 2012, 2013, 2014) y un conjunto de empresas que desarrollan su actividad económica en diversas geografías y pertenecen a distintos índices bursátiles⁸, resulta apropiado incluirlas y analizar el efecto que desempeñan en el modelo predictivo. El interés principal de incluir a estas variables dentro del estudio radica en poder aportar evidencia empírica de la posible mejora en la capacidad predictiva de los modelos al incluir este tipo de variables dentro de los modelos.

Tabla 7. Variables Macroeconómicas Seleccionadas

Variable	Significado
Sector	Clasificación según la especialización de su actividad
GII	Global Insolvency Index (Euler Hermes)
TEA	Early - Stage Entrepreneurial Activity
PIB	Producto Interno Bruto
Desempleo	Tasa de Desempleo o Paro
Corrupción	Índice de Percepción de la Corrupción
Inflación	Inflación
Situación Legal	Clasificación según derecho civil o derecho común

Fuente: Elaboración Propia del Autor

Factores de Liquidez

Los objetivos a largo plazo para el éxito empresarial se basan en la maximización de los beneficios y del valor de la empresa mientras que los objetivos a corto plazo se enfocan principalmente en:

- 1) Obtener una liquidez óptima o al menos apropiada para la empresa
- 2) Evitar las dificultades financieras

⁸ Un índice bursátil es un indicador de la evolución de un mercado en función del comportamiento de las cotizaciones de los títulos más representativos. Se compone de un conjunto de instrumentos, acciones o deuda, y busca capturar las características y los movimientos de valor de los activos que lo componen. También es una medida del rendimiento que este conjunto de activos ha presentado durante un periodo de tiempo determinado. (Bolsa de Valores de Colombia).

Con el propósito de evitar las dificultades financieras es necesario alcanzar el primer objetivo, es decir, conseguir la liquidez adecuada. De no lograrse esto último, es seguro que la empresa entra en un proceso de enfermedad financiera, que comienza con un vencimiento y falta de pago de las obligaciones a corto plazo para posteriormente entrar en un proceso de falta de liquidez afectando de esta manera al ciclo de explotación de la empresa, lo cual la obliga a entrar en una suspensión de pagos, que es sin lugar a dudas la antesala de la quiebra técnica.

Para Bernstein (1989) la liquidez es la facilidad, velocidad y el grado de merma para convertir los activos circulantes en tesorería, y la escasez de esta la considera uno de los principales síntomas de dificultades financieras. En principio la liquidez se puede definir de forma simple como la capacidad de pago que presenta la empresa en el corto plazo, o bien como la capacidad de convertir los activos e inversiones en dinero o instrumentos líquidos, que es lo que se conoce como tesorería.

Factores de Endeudamiento

El factor de endeudamiento puede ser definido como aquel que presenta la situación de la estructura de capital, la cual se determina en principio a partir de la relación entre los fondos ajenos y los fondos propios. Mientras que los fondos ajenos tienen un rendimiento garantizado y además se sabe con exactitud el tiempo de su reembolso y el tipo de interés devengado, esta situación se ignora en el caso de los fondos propios.

El comportamiento del factor de endeudamiento indica que a medida que sus proporción es más significativa dentro de la estructura de capital, más altos son los gastos fijos, situación que puede llevar a un exceso de dependencia del exterior y por lo tanto a la pérdida del control de la empresa por parte de los accionistas. En cambio, el endeudamiento en proporciones adecuadas representa una sana financiación para el desarrollo de las actividades de la empresa, pues representa un origen de fondos más barato con respecto al capital propio. Otra ventaja que representa el mantener un endeudamiento razonable, es que en épocas de inflación la empresa distribuye en riesgo sistemático entre los accionistas y los acreedores.

En general los ratios de endeudamiento miden los diferentes componentes de la estructura del capital con respecto al total de esta magnitud. También se utilizan para determinar que otras deudas tienen la empresa con el fin de saber el grado de apalancamiento financiero que existe.

Al igual que muchos autores en la literatura sobre modelo predictivos, en el presente estudio hemos optado por incluir en un mismo factor al endeudamiento y a la solvencia considerando que el objetivo de ambos factores es medir el grado de liquidez a largo plazo que presenta la empresa.

Factores de Rentabilidad

La rentabilidad es el incremento porcentual de riqueza e implica la ganancia que es capaz de brindar una inversión, estrictamente es la relación que existe entre el rendimiento generado por una inversión y el método de la misma. Las razones de rentabilidad miden la efectividad de la gerencia para generar utilidades contables sobre las ventas y la inversión, mediante el control de costos y gastos, y, del correcto uso de los recursos de la empresa. Mientras mayores sean los resultados, mayor será la prosperidad para la misma empresa.

Así, estos indicadores comparan las ganancias de un período con determinadas partidas del Balance General y del Estado de Resultados. La ganancia por unidad invertida por los socios marca cuántos pesos se ganan por cada peso invertido. Las razones de rentabilidad son de dos tipos: aquellas que muestran la rentabilidad con relación a las ventas y aquéllas que muestran la rentabilidad en relación con la inversión. En conjunto, señalan la eficiencia de operación de la empresa y ayudan a medir la eficacia en la obtención del objetivo básico financiero (aumentar el valor de la compañía).

Factores de Productividad / Eficiencia

La productividad se entiende como la relación entre la producción y sus factores. Según el diccionario económico de Oxford: “la productividad en economía es la razón entre el producto y los factores para producirlo”. En el marco del IV Congreso Mundial de

Productividad (1984) se estableció que la productividad es un concepto universal que aspira a proporcionar más bienes y servicios para un mayor número de personas, con cada vez menor número de recursos reales; para ello considera que es necesaria la aplicación integrada de habilidades y esfuerzos humanos, capital, tecnología, etc. Con respecto al término eficiencia, se dice que se da en una empresa cuando se busca lograr el mínimo de costes.

En cuanto a los ratios de eficiencia, éstos miden primordialmente dos factores: a) evalúan como está funcionando la empresa con respecto a las políticas preestablecidas; y b) evalúan con que efectividad la empresa maneja sus recursos.

4.1.4 Definición y Análisis Descriptivo de las Variables Independientes

4.1.4.1 Variables Microeconómicas (Ratios Financieros)

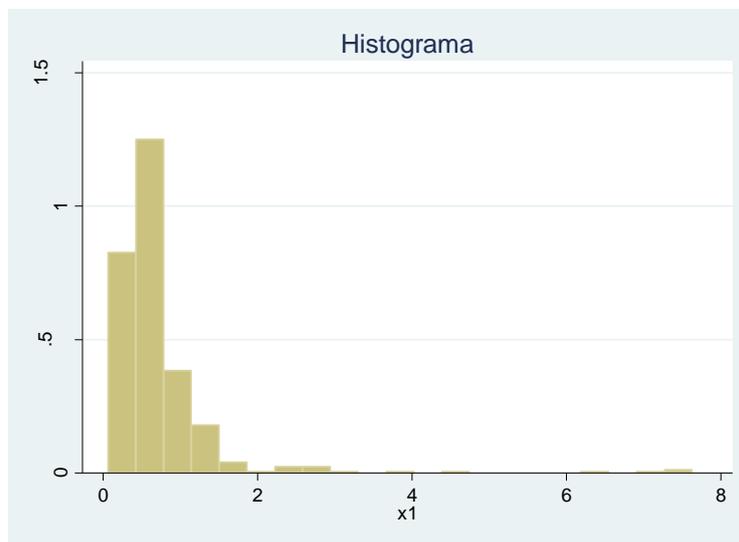
$$X_1 = \frac{\text{Activo Circulante}}{\text{Total Pasivo}}$$

Definición

Este ratio es conocido como ratio de Capital Circulante o Coeficiente de Liquidez Circulante y mide la participación del financiamiento total en las inversiones de corto plazo.

Estadística Descriptiva

Gráfico 7. Histograma de Frecuencias de las Variable X1



Fuente: Elaboración Propia del Autor

Tabla 8. Resumen de Estadísticas de la Variable X1

periodo	N	max	min	mean	p50	sd
2010	93	7.0064	.0711	.7462753	.5847	.7988958
2011	96	6.3651	.0622	.7459865	.5799	.8086429
2012	96	7.5517	.0625	.7333646	.5418	.8319814
2013	96	7.6307	.1075	.7325219	.5639	.8751108
2014	96	2.9518	.0785	.6513427	.5243	.485277
Total	477	7.6307	.0622	.7217449	.5628	.7701546

Fuente: Elaboración Propia del Autor

$$X_2 = \frac{\text{Total Pasivo}}{\text{Total Activo}}$$

Definición

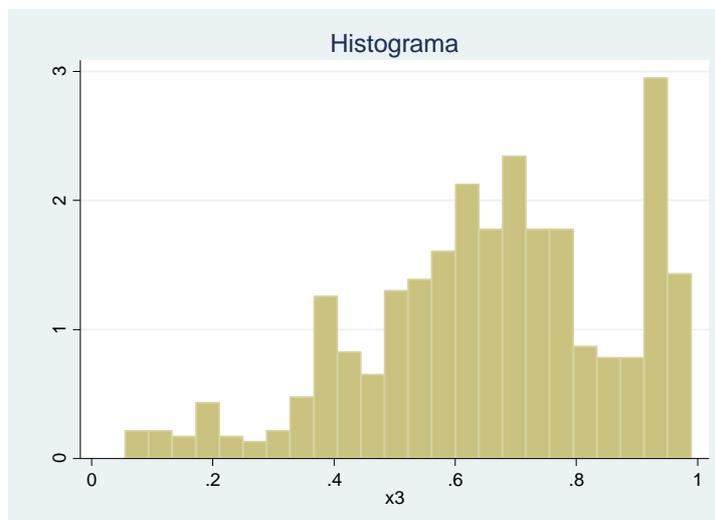
Este indicador es conocido en la literatura como ratio de solvencia global y se considera el más importante de los ratios de endeudamiento y mide la proporción de activos financiados por recursos de terceros. El coeficiente indica que recursos ajenos tiene invertidos la empresa en cada unidad monetaria del activo total. Cuanto mayor es el coeficiente más depende la empresa de los recursos ajenos.

Una relación alta de deuda-activos significa que la empresa tiene demasiada deuda. Aunque la deuda tiende a ser menos cara que los recursos obtenidos mediante capital, se crea mayor riesgo. Una empresa endeudada tiene la obligación de pagar su deuda en fechas específicas. Si la empresa no logra cubrir esas obligaciones a tiempo, los acreedores podrían confiscar sus activos y la empresa podría quebrar.

En el caso contrario, es decir una empresa con relación deuda-activo baja tiene muy poca deuda en comparación con su capital. Esto significa que tiene muy pocas obligaciones de pago. Sin embargo la empresa debe repartir sus utilidades con los accionistas.

Estadística Descriptiva

Gráfico 8. Histograma de Frecuencias de las Variable X2



Fuente: Elaboración Propia del Autor

Tabla 9. Resumen de Estadísticas de la Variable X2

periodo	N	max	min	mean	p50	sd
2010	116	.9857	.0595	.6551586	.6627	.2089883
2011	119	.986	.0708	.657605	.6793	.2118635
2012	119	.9887	.0553	.659521	.675	.2114277
2013	119	.9899	.0562	.6588773	.6723	.2129642
2014	119	.9877	.1235	.6664042	.6971	.2051385
Total	592	.9899	.0553	.6595353	.6757	.2094221

Fuente: Elaboración Propia del Autor

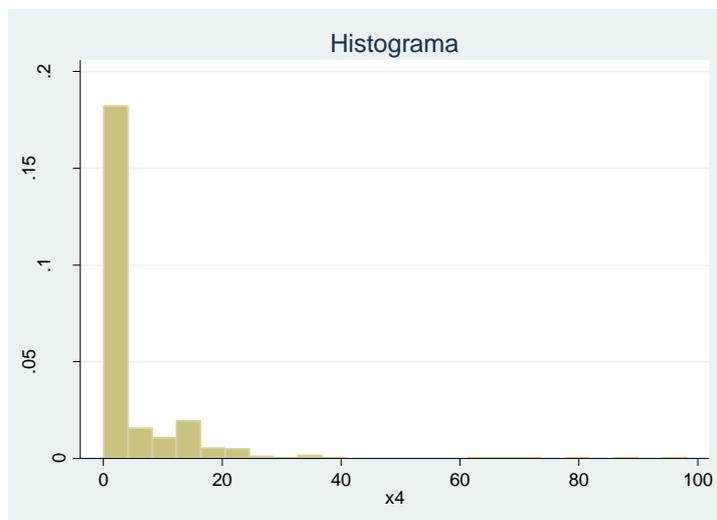
$$X_3 = \frac{\text{Total Pasivo}}{\text{Total Capital Contable}}$$

Definición

Este indicador es conocido en la literatura como razón de apalancamiento externo e indica cuantas unidades monetarias han sido aportadas al financiamiento de la inversión por lo accionistas, por cada unidad monetaria tomada de terceros.

Estadística Descriptiva

Gráfico 9. Histograma de Frecuencias de las Variable X3



Fuente: Elaboración Propia del Autor

Tabla 10. Resumen de Estadísticas de la Variable X3

periodo	N	max	min	mean	p50	sd
2010	116	68.9967	.0633	5.03564	1.96485	8.59922
2011	119	70.3704	.0762	5.258013	2.1183	8.979209
2012	119	87.6127	.0586	5.248584	2.0766	9.739949
2013	119	98.167	.0596	5.858249	2.0653	11.83254
2014	119	80.0361	.141	5.223291	2.3016	9.215436
Total	592	98.167	.0586	5.32622	2.0835	9.716814

Fuente: Elaboración Propia del Autor

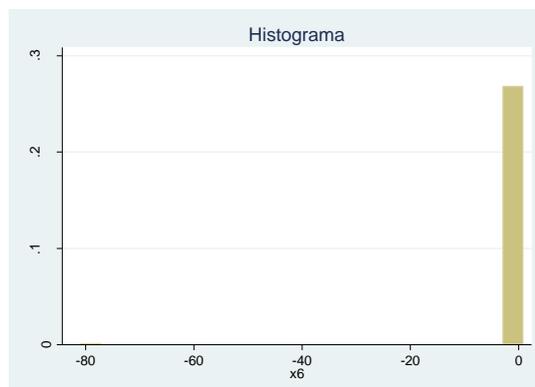
$$X_4 = \frac{EBIT}{Ingresos}$$

Definición

Este indicador es conocido en la literatura como margen de beneficio neto o rentabilidad de los ingresos e indica cuanto beneficio se obtiene por cada peso de ventas, o dicho de otra manera, cuánto gana la empresa por cada peso que vende. El valor de este indicador estará en relación directa al control de los gastos, pues por mucho que la empresa venda, si los gastos aumentan, el resultado se verá reducido.

Estadística Descriptiva

Gráfico 10. Histograma de Frecuencias de las Variable X4



Fuente: Elaboración Propia del Autor

Tabla 11. Resumen de Estadísticas de la Variable X4

periodo	N	max	min	mean	p50	sd
2010	95	.7983	-.4763	.1924063	.1412	.1882858
2011	98	.6639	-.0646	.1682214	.12735	.151985
2012	98	.6966	-.1421	.1589541	.1118	.1639877
2013	99	.8033	-80.9048	-.6865061	.1053	8.150692
2014	99	.7988	-80.9048	-.7073616	.0999	8.148066
Total	489	.8033	-80.9048	-.1792456	.1127	5.183963

Fuente: Elaboración Propia del Autor

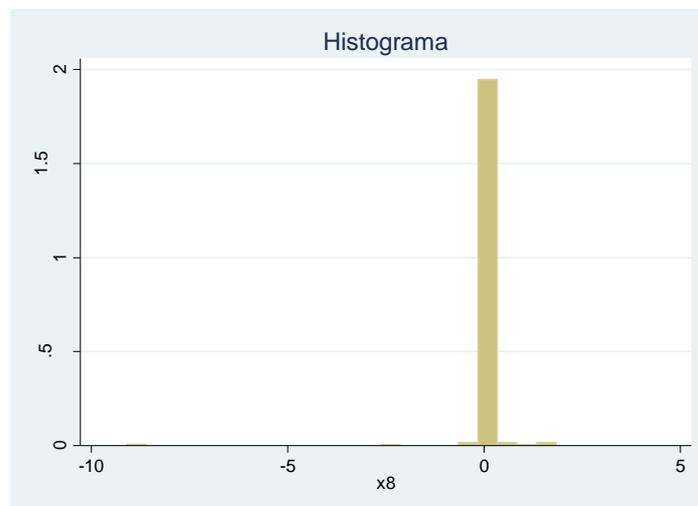
$$X_5 = \frac{EBIT}{Total\ Capital\ Contable}$$

Definición

Este indicador es conocido como rendimiento del capital y mide la capacidad que tiene una empresa para generar dinero con base en la inversión de los accionistas. Es decir, cuánta utilidad se genera por cada unidad monetaria de los accionistas.

Estadística Descriptiva

Gráfico 11. Histograma de Frecuencias de las Variable X5



Fuente: Elaboración Propia del Autor

Tabla 12. Resumen de Estadísticas de la Variable X5

periodo	N	max	min	mean	p50	sd
2010	95	1.8412	-.3352	.1060968	.0755	.1978564
2011	98	1.706	-.2006	.0924173	.06495	.179162
2012	98	1.3942	-.161	.0822735	.0613	.1478166
2013	99	1.1243	-9.0978	-.038602	.0588	.9556367
2014	99	1.5489	-9.0978	-.0134212	.058	.9380635
Total	489	1.8412	-9.0978	.0450892	.0629	.6180148

Fuente: Elaboración Propia del Autor

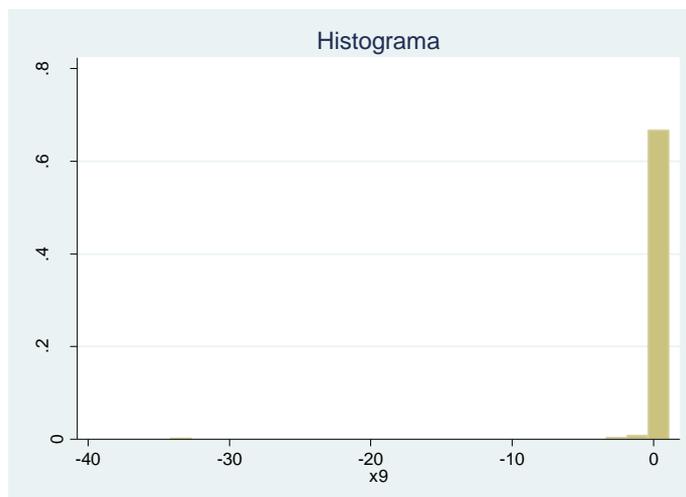
$$X_6 = \frac{\text{Utilidad Neta}}{\text{Ingresos}}$$

Definición

Este indicador es conocido como margen neto de utilidad o rentabilidad de las ventas y mide el porcentaje que está quedando a los propietarios para poder operar la empresa, es decir, el porcentaje que se convierte en ganancia por cada unidad monetaria vendida.

Estadística Descriptiva

Gráfico 12. Histograma de Frecuencias de las Variable X6



Fuente: Elaboración Propia del Autor

Tabla 13. Resumen de Estadísticas de la Variable X6

periodo	N	max	min	mean	p50	sd
2010	113	1.0945	-.2674	.1219549	.0821	.1672652
2011	116	.6869	-1.1141	.0771517	.0549	.2014243
2012	117	.742	-1.6891	.0610385	.0494	.2251959
2013	117	.7133	-34.142	-.2766308	.0602	3.182855
2014	117	.6705	-34.142	-.2208564	.0547	3.178781
Total	580	1.0945	-34.142	-.0488517	.05895	2.026189

Fuente: Elaboración Propia del Autor

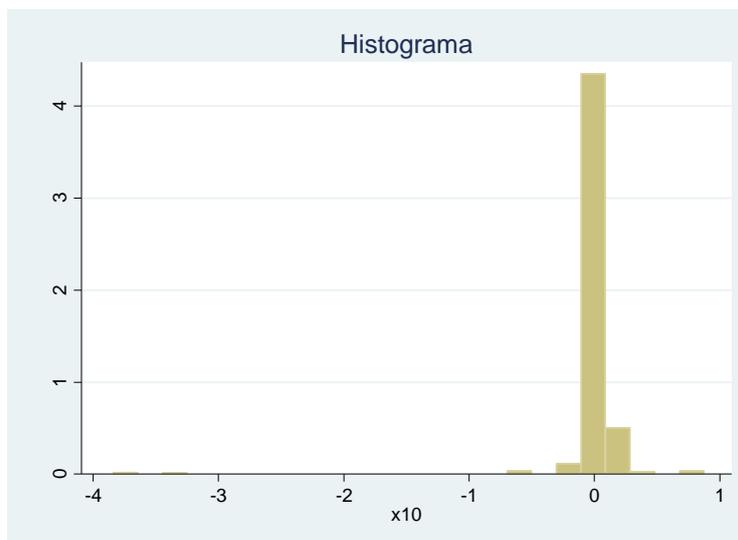
$$X_7 = \frac{\text{Utilidad Neta}}{\text{Total Capital Contable}}$$

Definición

Este indicador es conocido como margen de utilidad neto y mide el porcentaje que está quedando a los propietarios para poder operar la empresa.

Estadística Descriptiva

Gráfico 13. Histograma de Frecuencias de las Variable X7



Fuente: Elaboración Propia del Autor

Tabla 14. Resumen de Estadísticas de la Variable X7

periodo	N	max	min	mean	p50	sd
2010	115	.7946	-.1882	.0524243	.0366	.087157
2011	117	.8765	-.6703	.0400308	.0332	.1154089
2012	118	.7353	-.6294	.0285051	.02545	.1118276
2013	118	.873	-3.8393	-.0395068	.02845	.4920986
2014	118	.3124	-3.8393	.001778	.02765	.3632134
Total	586	.8765	-3.8393	.0164232	.0308	.2872445

Fuente: Elaboración Propia del Autor

$$X_8 = \frac{\text{Utilidad Neta}}{\text{Total Activos}}$$

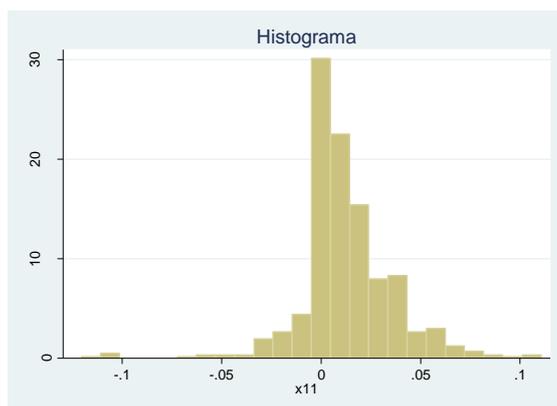
Definición

Este indicador es conocido en la literatura como Rendimiento de los Activos y es mejor conocido por sus siglas ROA (Return on Assets), toma en cuenta el aporte de los activos a las utilidades de la empresa. Cuanto más alto es el ratio, más rentables son los activos. Un ratio

elevado es sinónimo de activos muy eficientes y productivos, en tanto que un ratio bajo se asocia a la baja productividad o bien a la ineficiencia en su manejo.

Estadística Descriptiva

Gráfico 14. Histograma de Frecuencias de las Variable X8



Fuente: Elaboración Propia del Autor

Tabla 15. Resumen de Estadísticas de la Variable X8

periodo	N	max	min	mean	p50	sd
2010	115	.0792	-.0271	.0173261	.013	.0188567
2011	117	.1107	-.1022	.0146462	.0111	.0232009
2012	118	.088	-.0542	.0127381	.0084	.0220616
2013	118	.0892	-.1204	.0088339	.00935	.0274863
2014	118	.1023	-.1058	.0109593	.00555	.0256636
Total	586	.1107	-.1204	.0128751	.00965	.0237645

Fuente: Elaboración Propia del Autor

$$X_9 = \frac{\text{Ingresos}}{\text{Activos Fijos}}$$

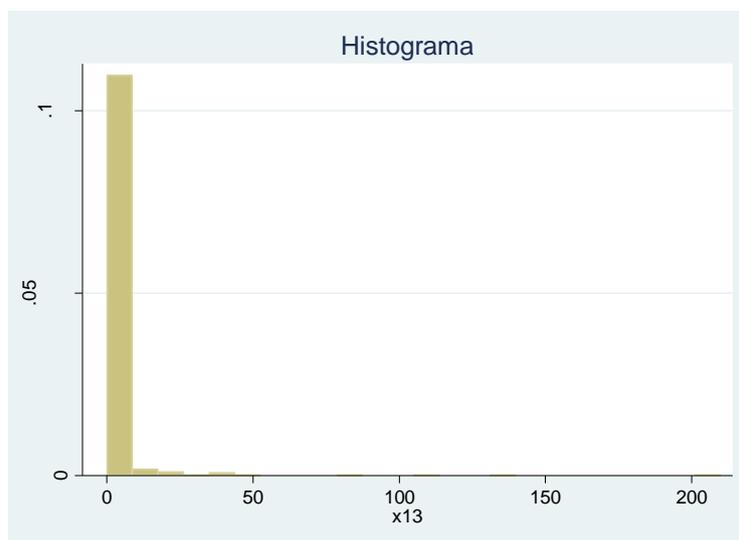
Definición

Este indicador se conoce en la literatura como rotación de activos fijos. Mide la eficiencia de las ventas (ingresos) con relación a los activos permanentes. Algunos autores lo clasifican como un ratio de rentabilidad, productividad o eficiencia.

Aunque esta magnitud es a largo plazo, se pudiera a estimar en el corto plazo para observar el comportamiento de factores temporales que afectan esta relación como por ejemplo, el exceso en la capacidad de producción, o bien, aquellos activos ineficientes u obsoletos. Debido a que los incrementos en la capacidad de fabricación no son graduales, debe buscar una medida adecuada entre la inversión de activos fijos y la capacidad productiva.

Estadística Descriptiva

Gráfico 15. Histograma de Frecuencias de las Variable X9



Fuente: Elaboración Propia del Autor

Tabla 16. Resumen de Estadísticas de la Variable X9

periodo	N	max	min	mean	p50	sd
2010	113	40.7811	.0445	2.529227	.9804	5.807037
2011	117	80.6444	.0366	2.803536	.9643	8.235468
2012	117	106.098	.0363	2.987707	.9939	10.53397
2013	117	209.9444	.0329	3.923162	.9258	19.82441
2014	117	133.7733	.0538	3.386776	.8515	13.04481
Total	581	209.9444	.0329	3.130191	.9608	12.45553

Fuente: Elaboración Propia del Autor

4.1.4.2 Variables Macroeconómicas

Sector

Definición

La actividad económica está dividida en sectores económicos, cada uno de los cuales, hace referencia a una parte de la actividad económica cuyos elementos presentan características comunes. Su división se hace generalmente de acuerdo a los procesos de producción que ocurren al interior de cada uno de ellos. Es común que las actividades económicas se diferencien dependiendo de su especialización, lo cual da origen a la siguiente clasificación⁹:

1. Sector Consumo, No Cíclico: La diferencia entre las industrias cíclicas y no cíclicas es simplemente la diferencia entre necesidad y lujo. Hay ciertos elementos que no podemos vivir sin él y que probablemente no vamos a recortar incluso cuando los tiempos son difíciles. Las acciones de las empresas que producen estas cosas son no-cíclico y son "defendidos" en contra de los efectos de la recesión económica, proporcionando buenos lugares para invertir cuando el panorama económico es amargo. Por ejemplo, artículos para el hogar no duraderos - un término de lujo para las cosas que usa rápidamente alrededor de la casa - como la pasta de dientes,

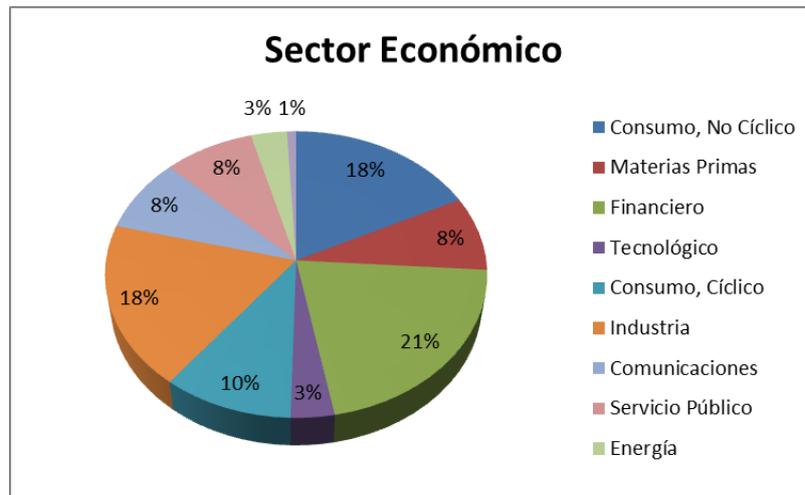
⁹ Definiciones de los sectores económicos según Investopedia: <http://www.investopedia.com/>

jabón, champú y detergente para lavar platos pueden no parecer como esencial, pero no se puede realmente sacrificarlos. La mayoría de las personas no sienten que pueden esperar hasta el próximo año para hacer espuma con el jabón en la ducha.

2. **Materias Primas:** Categoría de acciones que representa para las empresas que participan en el descubrimiento, desarrollo y transformación de materias primas. El sector de los materiales de base incluye la extracción y refinación de metales, fabricantes de productos químicos y productos forestales.
3. **Financiero:** Categoría de acciones que representa a las firmas que prestan servicios financieros a los clientes comerciales y minoristas. Este sector incluye los bancos, fondos de inversión, compañías de seguros y bienes raíces.
4. **Tecnológico:** Categoría de acciones relacionadas con la investigación, el desarrollo y / o distribución de bienes y servicios de base tecnológica. Este sector contiene empresas de fabricación de la electrónica, la creación de software, computadoras o productos y servicios relacionados con la tecnología de la información.
5. **Consumo, Cíclico:** Este sector incluye los bienes de consumo que se consideran compras opcionales y no incluyen alimentos, ropa, combustible o cualquier otra compra que es necesario para la supervivencia. Este sector también se conoce como el sector de consumo discrecional.
6. **Industrial:** Categoría de acciones que se relacionan con la producción de bienes utilizados en la construcción y la manufactura. Este sector incluye a las empresas que participan en la industria aeroespacial y de defensa, maquinaria industrial, herramientas, producción de madera, la construcción, el cemento y la fabricación de metal.
7. **Comunicaciones:** Incluyen proveedores integrados gigantes de servicios de líneas inalámbricas y de cable, así como las empresas más pequeñas que se especializan en servicios inalámbricos o equipos de comunicaciones de fabricación.

8. Utilidades: Incluye a las empresas de servicios públicos como el gas y la energía. El sector de servicios públicos contiene compañías como electricidad, gas y las empresas de agua y proveedores integrados.
9. Energía: Incluye a las empresas que se relacionan con la producción o el suministro de energía. Este sector incluye a las empresas involucradas en la exploración y desarrollo de reservas de petróleo o gas, la extracción de petróleo y gas, o las empresas eléctricas integradas
10. Diversificado: Una empresa que tiene varios negocios no relacionados. Negocios no relacionados son los que (1) requiere conocimientos de gestión única, (2) tienen diferentes clientes finales y (3) producir diferentes productos o proporcionar diferentes servicios. Uno de los beneficios de ser una compañía diversificada es que amortigua una empresa de las fluctuaciones dramáticas en cualquier sector de una industria. Sin embargo, este modelo también es menos probable que los accionistas permitirá darse cuenta de las ganancias o pérdidas significativas debido a que no está enfocada singularmente en un negocio.

Gráfico 16. Distribución de la Variable Sector



Fuente: Elaboración Propia del Autor

País

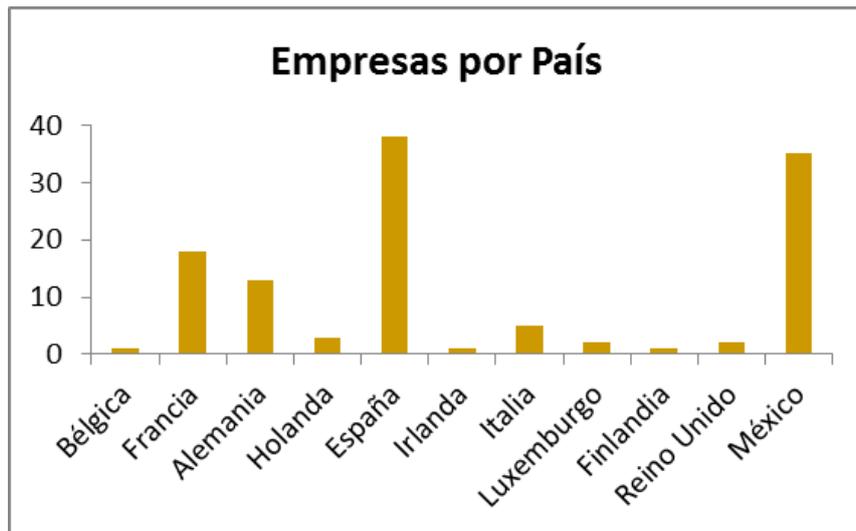
En el presente trabajo de investigación se analizó una muestra de 119 empresas que desempeñaban su actividad económica en distintas geografías y que pertenecían a tres índices de referencia distintos: IPC, IBEX35 y EURO STOXX 50

Tabla 17. Código de la Variable País

Valor	País
1	Bélgica
2	Francia
3	Alemania
4	Holanda
5	España
6	Irlanda
7	Italia
8	Luxemburgo
9	Finlandia
10	Reino Unido
11	México

Fuente: Elaboración Propia del Autor

Gráfico 17. Distribución de la Variable País



Fuente: Elaboración Propia del Autor

GII

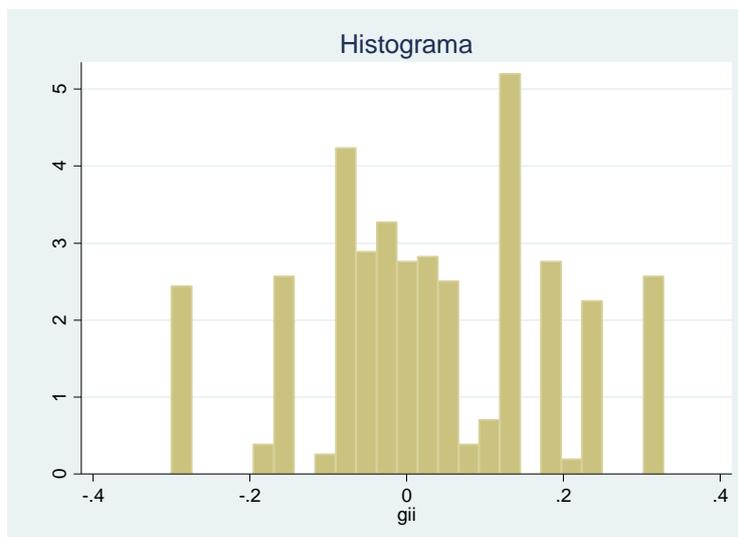
Definición

El índice global de insolvencia es un indicador que fue creado por la empresa Euler Hermes dado que el concepto de insolvencia empresarial varía de un país a otro y se calcula como la suma ponderada de los índices nacionales.

Cada país se pondera con su participación en el PIB total de los países incluidos en el estudio. De acuerdo con la empresa Euler Hermes este indicador es la mejor manera de controlar la salud de la economía real mundial.

Estadística Descriptiva

Gráfico 18. Histograma de Frecuencias de las Variable GII



Fuente: Elaboración Propia del Autor

Tabla 18. Resumen de Estadísticas de la Variable GII

periodo	N	max	min	mean	p50	sd
2010	119	.3286	-.1648	-.0607958	-.0357	.0950808
2011	119	.1844	-.0813	.0356689	-.003	.1128763
2012	119	.3196	-.0849	.1721244	.2284	.1424627
2013	119	.14	-.19	.087395	.14	.0844771
2014	119	.11	-.3	-.0931933	-.07	.1558896
Total	595	.3286	-.3	.0282398	.0187	.1549384

Fuente: Elaboración Propia del Autor

TEA

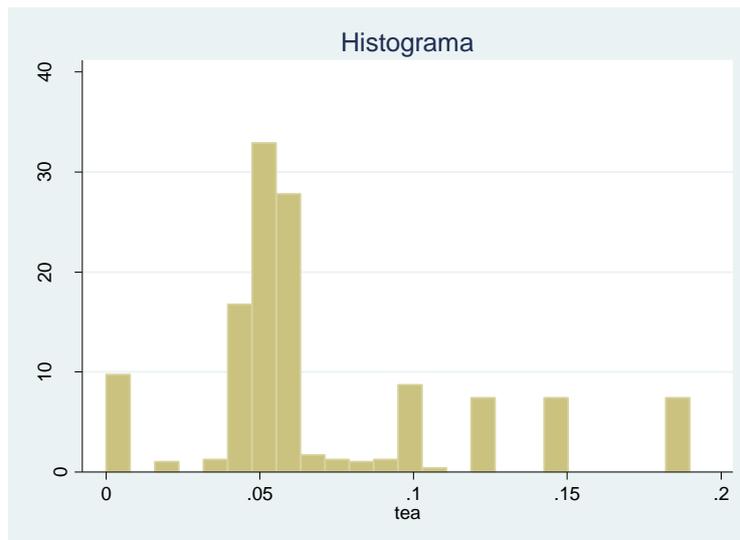
Definición

El Global Entrepreneurship Monitor (GEM) es un estudio global realizado por un consorcio de universidades desde 1999 y que tiene como objetivo principal analizar el nivel de iniciativa empresarial que ocurre en una amplia gama de países seleccionados (70 en 2013). El principal indicador contenido en este estudio es conocido como TEA (Total Early-

Stage Entrepreneurial Activity) y evalúa el porcentaje de población en edad laboral, de aquellos que están a punto de iniciar una actividad empresarial como los que han comenzado desde un año y hasta máximo tres y medio años.

Estadística Descriptiva

Gráfico 19. Histograma de Frecuencias de las Variable TEA



Fuente: Elaboración Propia del Autor

Tabla 19. Resumen de Estadísticas de la Variable TEA

periodo	N	max	min	mean	p50	sd
2010	119	.072	0	.0323109	.043	.0234957
2011	119	.096	0	.0664034	.058	.0240329
2012	119	.12	0	.0746218	.06	.0318309
2013	119	.148	.034	.0806134	.052	.0448421
2014	119	.19	.044	.095958	.055	.0617401
Total	595	.19	0	.0699815	.056	.0450707

Fuente: Elaboración Propia del Autor

PIB

Definición

El PIB es el valor monetario de los bienes y servicios finales producidos por una economía en un período determinado. EL PIB es un indicador representativo que ayuda a medir el crecimiento o decrecimiento de la producción de bienes y servicios de las empresas de cada país, únicamente dentro de su territorio. Este indicador es un reflejo de la competitividad de las empresas.

¿Por qué es importante que crezca el PIB?

- Indica la competitividad de las empresas. Si la producción de las empresas mexicanas no crecen a un ritmo mayor, significa que no se está invirtiendo en la creación de nuevas empresas, y por lo tanto, la generación de empleos tampoco crece al ritmo deseado.
- Si el PIB crece por abajo de la inflación significa que los aumentos salariales tenderán a ser menores que la misma.
- Un crecimiento del PIB representa mayores ingresos para el gobierno a través de impuestos. Si el gobierno desea mayores ingresos, deberá fortalecer las condiciones para la inversión no especulativa, es decir, inversión directa en empresas; y también fortalecer las condiciones para que las empresas que ya existen sigan creciendo.

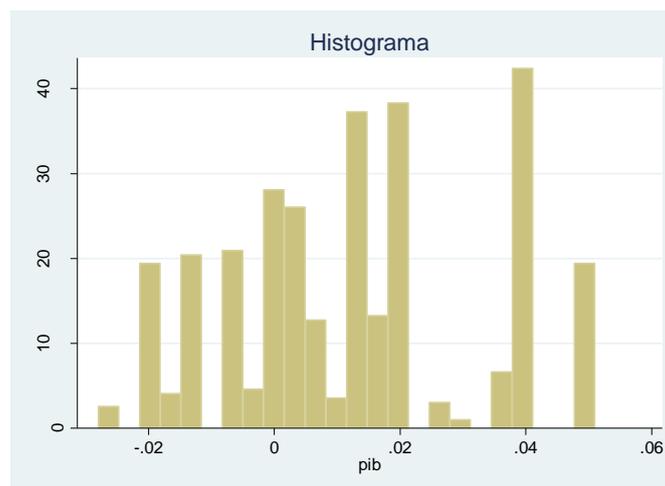
Algunas aclaraciones sobre el PIB

- El PIB de una país aumentará si el gobierno o las empresas dentro del mismo toman préstamos en el extranjero, obviamente, esto disminuirá el PIB en períodos futuros.
- No toma en cuenta la depreciación del capital (Aquí se incluyen tanto maquinaria, fábricas, etc., como así también recursos naturales, y también se podría incluir al "capital humano"). Por ejemplo, un país puede incrementar su PIB explotando en forma intensiva sus recursos naturales, pero el capital del país disminuirá, dejando para generaciones futuras menos capital disponible.

- No tiene en cuenta externalidades negativas que algunas actividades productivas generan, por ejemplo, la contaminación ambiental.
- No tiene en cuenta la distribución del ingreso. Los pobladores de un país con igual PIB per cápita que otro pero con una distribución más equitativa del mismo disfrutarán de un mayor bienestar que el segundo.
- La medida del PIB no tiene en cuenta actividades productivas que afectan el bienestar pero que no generan transacciones, por ejemplo trabajos de voluntarios o de amas de casa.
- Actividades que afectan negativamente el bienestar pueden aumentar el PIB, por ejemplo divorcios y crímenes.

Estadística Descriptiva

Gráfico 20. Histograma de Frecuencias de las Variable PIB



Fuente: Elaboración Propia del Autor

Tabla 20. Resumen de Estadísticas de la Variable PIB

periodo	N	max	min	mean	p50	sd
2010	119	.051	-.003	.0251092	.02	.021457
2011	119	.04	-.006	.0189328	.021	.0193376
2012	119	.04	-.028	.0041681	.002	.0253736
2013	119	.02	-.017	.0011176	.001	.011728
2014	119	.048	-.004	.0136807	.014	.0082912
Total	595	.051	-.028	.0126017	.014	.020369

Fuente: Elaboración Propia del Autor

Desempleo

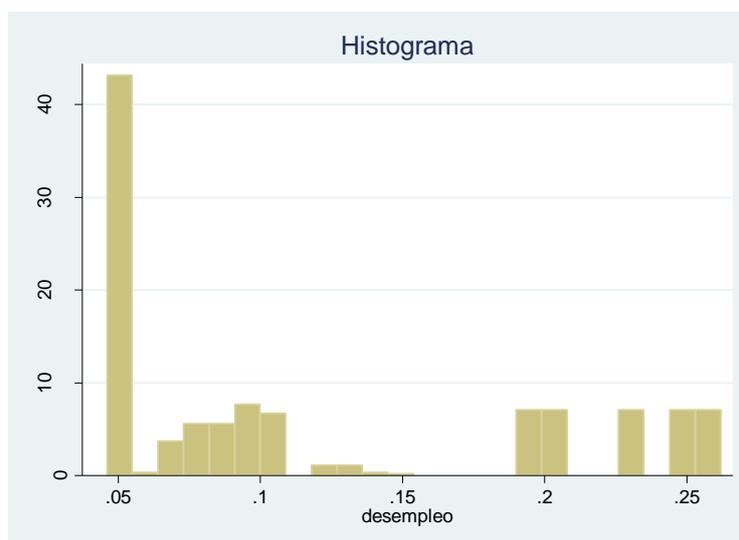
Definición

El Desempleo (o paro) se produce cuando la gente está sin trabajo y buscando trabajo activamente. La tasa de desempleo es una medida de la extensión del desempleo y se calcula como un porcentaje dividiendo el número de personas desempleadas por todas las personas que se encuentran en la fuerza laboral. Durante los períodos de recesión, la economía experimenta generalmente una tasa de desempleo relativamente alta.

La tasa de desempleo es una de las estadísticas más observadas por los analistas y es vista como un signo de debilidad de la economía que puede requerir reducir la tasa de interés. Una tasa decreciente, de manera similar, indica una economía en crecimiento, que suele ir acompañada por una tasa de inflación más alta y requerir así un aumento en las tasas de interés.

Estadística Descriptiva

Gráfico 21. Histograma de Frecuencias de las Variable Desempleo



Fuente: Elaboración Propia del Autor

Tabla 21. Resumen de Estadísticas de la Variable Desempleo

periodo	N	max	min	mean	p50	sd
2010	119	.192	.048	.1086723	.084	.059351
2011	119	.204	.047	.110437	.081	.0661495
2012	119	.232	.049	.1190588	.094	.0796962
2013	119	.262	.049	.1308487	.103	.0927903
2014	119	.252	.046	.1267899	.101	.089124
Total	595	.262	.046	.1191613	.091	.0787109

Fuente: Elaboración Propia del Autor

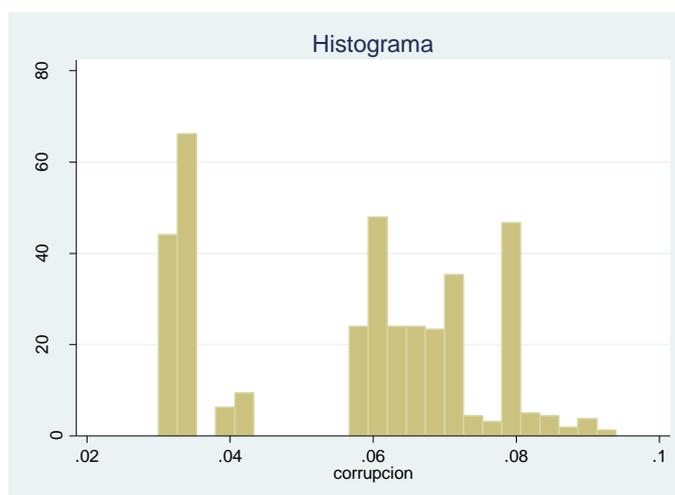
Corrupción

Definición

El Índice de Percepción de Corrupción¹⁰ establece cada año un ranking de países basado en los niveles de corrupción en el sector público percibidos por empresarios, analistas de riesgo, y ciudadanos comunes. Se centra en la corrupción del sector público y la define como el abuso del servicio público para el beneficio particular, por tanto, define como corruptos actos como sobornos a funcionarios, pagos irregulares en las contrataciones y malversación de fondos.

Estadística Descriptiva

Gráfico 22. Histograma de Frecuencias de las Variable Corrupción



¹⁰ <http://www.transparency.org/>

Fuente: Elaboración Propia del Autor

Tabla 22. Resumen de Estadísticas de la Variable Corrupción

periodo	N	max	min	mean	p50	sd
2010	119	.092	.031	.0561176	.061	.0187283
2011	119	.094	.03	.0566218	.062	.0196421
2012	119	.09	.034	.058563	.065	.0177031
2013	119	.089	.034	.056605	.059	.0170419
2014	119	.089	.035	.0571176	.06	.0167176
Total	595	.094	.03	.057005	.061	.0179583

Fuente: Elaboración Propia del Autor

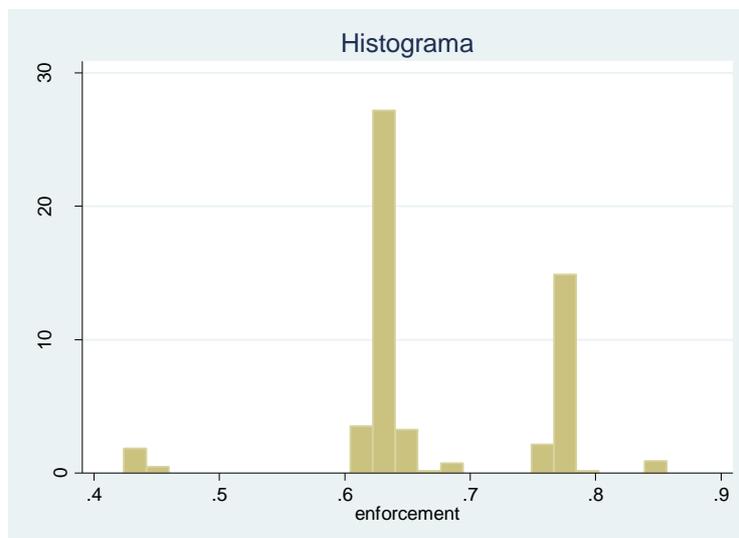
Enforcement

Definición

El índice de Cumplimiento de Contratos mide la eficiencia del sistema judicial en la resolución de una disputa comercial. Los datos se construyen siguiendo la evolución paso a paso de una disputa comercial ante los tribunales locales. Los datos se recogen a través del estudio de los códigos de procedimiento civil y otras regulaciones judiciales, así como cuestionarios completados por abogados y jueces locales.

Estadística Descriptiva

Gráfico 23. Histograma de Frecuencias de las Variable Enforcement



Fuente: Elaboración Propia del Autor

Tabla 23. Resumen de Estadísticas de la Variable Corrupción

periodo	N	max	min	mean	p50	sd
2010	119	.857	.4238	.6656378	.627	.0881844
2011	119	.857	.4238	.6693101	.63	.0865938
2012	119	.857	.4238	.6690823	.63	.0862696
2013	119	.857	.4306	.6725412	.6369	.0838448
2014	119	.857	.4561	.6745681	.6461	.0803705
Total	595	.857	.4238	.6702279	.6314	.0848651

Fuente: Elaboración Propia del Autor

Situación Legal

Definición

Los sistemas legales de todo el mundo varían mucho, pero por lo general siguen el derecho civil o el derecho común. Nuevas investigaciones tales como Beck, Demirgüç-Kunt y Levine. (2003) sugieren que las diferencias entre países en origen legal ayuda a explicar las diferencias en el desarrollo financiero. En el sistema legal que prevalece en países anglosajones, "derecho común" (common law) es un sistema legal donde se da preeminencia a la decisión de un juez más que a lo establecido en las leyes. Y se suele contraponer a lo que en inglés se denomina "civil law", que significa "Derecho Civil" pero entendido como el sistema legal donde ocurre lo inverso, es decir, donde la base del sistema es la ley y no tanto la jurisprudencia o precedentes que establezcan jueces o tribunales.

Inflación

Definición

La inflación es el aumento generalizado y sostenido de los precios de bienes y servicios en un país. Para medir el crecimiento de la inflación se utilizan índices, que reflejan el crecimiento porcentual de una canasta de bienes ponderada. El índice de medición de la inflación es el Índice de Precios al Consumidor (IPC) que en México se le llama INPC (Índice Nacional de Precios al Consumidor). Este índice mide el porcentaje de incremento en los precios de una canasta básica de productos y servicios que adquiere un consumidor típico en el país.

Existen otros índices, como el Índice de Precios al Productor, que mide el crecimiento de precios de las materias primas.

Causas de la Inflación

Existen tres tipos de inflación:

- Inflación por consumo o demanda. Esta inflación obedece a la ley de la oferta y la demanda. Si la demanda de bienes excede la capacidad de producción o importación de bienes, los precios tienden a aumentar.
- Inflación por costos. Esta inflación ocurre cuando el precio de las materias primas (cobre, petróleo, energía, etc.) aumenta, lo que hace que el productor, buscando mantener su margen de ganancia, incremente sus precios.
- Inflación autoconstruida. Esta inflación ocurre cuando se prevé un fuerte incremento futuro de precios, y entonces se comienzan a ajustar éstos desde antes para que el aumento sea gradual.
- Inflación generada por expectativas de inflación (círculo vicioso). Esto es típico en países con alta inflación donde los trabajadores piden aumentos de salarios para contra restar los efectos inflacionarios, lo cual da pie al aumento en los precios por parte de los empresarios, originando un círculo vicioso de inflación

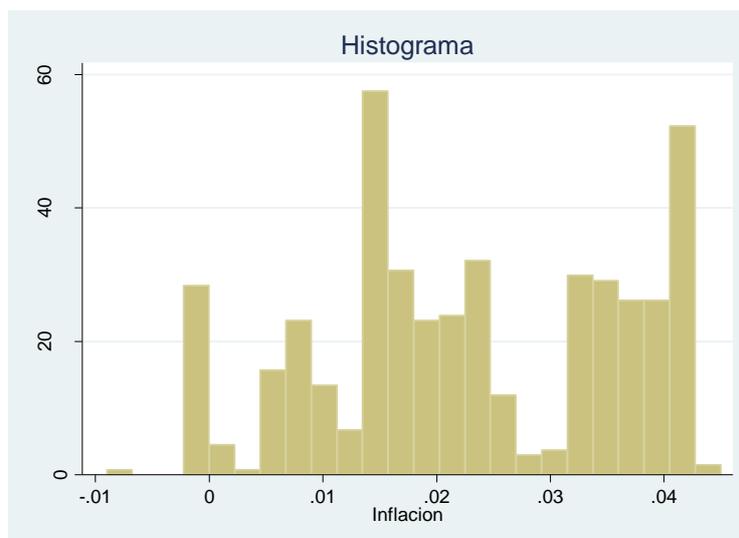
¿Cómo se detiene la inflación?

Para detener la inflación, los bancos centrales tienden a incrementar la tasa de interés de la deuda pública. De esta manera se incrementan las tasas de interés en los préstamos al consumo (tarjetas de crédito, hipotecas, etc.). Al aumentar las tasas de interés del consumo, se frena la demanda de productos.

El lado negativo de este control es que al frenar la demanda de productos, se frena a la industria que los produce, lo cual puede llevar a un estancamiento económico y desempleo.

Estadística Descriptiva

Gráfico 24. Histograma de Frecuencias de las Variable Inflación



Fuente: Elaboración Propia del Autor

Tabla 24. Resumen de Estadísticas de la Variable Inflación

periodo	N	max	min	mean	p50	sd
2010	119	.042	-.009	.0236807	.018	.0125279
2011	119	.045	.021	.0295294	.032	.0057693
2012	119	.041	.017	.0283613	.024	.0085675
2013	119	.038	.005	.0207647	.014	.0116289
2014	119	.04	-.001	.014	.005	.0172646
Total	595	.045	-.009	.0232672	.021	.0130383

Fuente: Elaboración Propia del Autor

4.2 Aplicación Empírica de las Ecuaciones de Estimación Generalizadas

4.2.1 Análisis Preliminar (ACP)

Matriz de Correlaciones

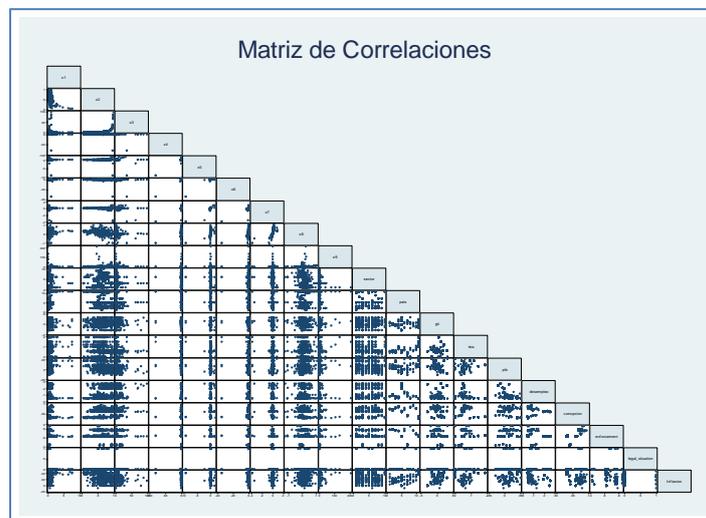
$X_1, \dots, X_9, \text{sector}, \text{pais}, \text{gii}, \text{tea}, \text{pib}, \text{desempleo}, \text{corrupcion}, \text{enfocement}, \text{legal}, \text{inflacion}$

Figura 2. Matriz de Correlaciones (Representación Numérica)

Matriz Corr	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	sector	pais	gii	tea	pib	desempleo	corrupcion	enfocement	legal	inflacion
X_1	1.0000																		
X_2	-0.5377	1.0000																	
X_3	-0.0304	0.4211	1.0000																
X_4	-0.0193	-0.1336	-0.2178	1.0000															
X_5	-0.0370	-0.0795	-0.2058	0.9515	1.0000														
X_6	-0.0084	-0.1514	-0.2152	0.9970	0.9490	1.0000													
X_7	-0.0075	-0.1347	-0.2842	0.7958	0.9027	0.8060	1.0000												
X_8	0.1801	-0.3716	-0.2099	0.3365	0.4102	0.3746	0.5463	1.0000											
X_9	-0.0547	-0.0320	-0.0184	0.0225	0.0206	0.0245	0.0236	0.0612	1.0000										
sector	-0.1350	0.2308	-0.0224	-0.0378	-0.0852	-0.0429	-0.1110	-0.3022	-0.1355	1.0000									
pais	0.2421	-0.2675	-0.0715	-0.0794	-0.0730	-0.0791	-0.0517	0.0155	0.1020	-0.1513	1.0000								
gii	-0.0018	0.0059	0.0521	-0.0312	-0.0429	-0.0368	-0.0630	-0.0941	0.0310	0.0262	0.0737	1.0000							
tea	0.1164	-0.1700	-0.0467	-0.1350	-0.1416	-0.1382	-0.1140	-0.0492	0.1089	-0.1297	0.5179	0.3058	1.0000						
pib	0.1887	-0.2830	-0.1697	-0.0075	-0.0005	-0.0017	0.0402	0.1418	0.0435	-0.2201	0.5322	-0.4048	0.1038	1.0000					
desempleo	-0.1551	0.3072	0.2407	0.0578	0.0675	0.0559	0.0086	-0.0686	-0.0560	0.2475	-0.4240	0.1072	-0.3424	-0.6646	1.0000				
corrupcion	-0.2226	0.2780	0.0495	0.0724	0.0656	0.0703	0.0507	-0.0540	-0.1001	0.0999	-0.8345	-0.0714	-0.4542	-0.4865	0.2871	1.0000			
enfocement	-0.0958	0.0019	-0.0921	0.0225	0.0109	0.0214	0.0227	0.0035	-0.0432	-0.1619	-0.5495	-0.1669	-0.1619	0.0157	0.2836	0.6909	1.0000		
legal	-0.0575	0.0533	-0.0188	0.0098	0.0302	0.0100	0.0422	0.1002	-0.0167	-0.0485	0.1071	-0.1070	0.0189	0.0168	-0.0399	0.1882	0.0853	1.0000	
inflacion	0.2010	-0.2470	-0.0866	-0.0701	-0.0628	-0.0700	-0.0342	0.0430	0.0870	-0.1814	0.7715	0.3634	0.4393	0.4332	-0.5125	-0.6924	-0.3579	-0.0259	1.0000

Fuente: Elaboración Propia del Autor

Figura 3. Matriz de Correlaciones (Representación Gráfica)



Tras analizar y observar que existen algunas correlaciones entre las variables (financieras y macroeconómicas), se realizará el análisis de componentes principales para intentar reducir la dimensión.

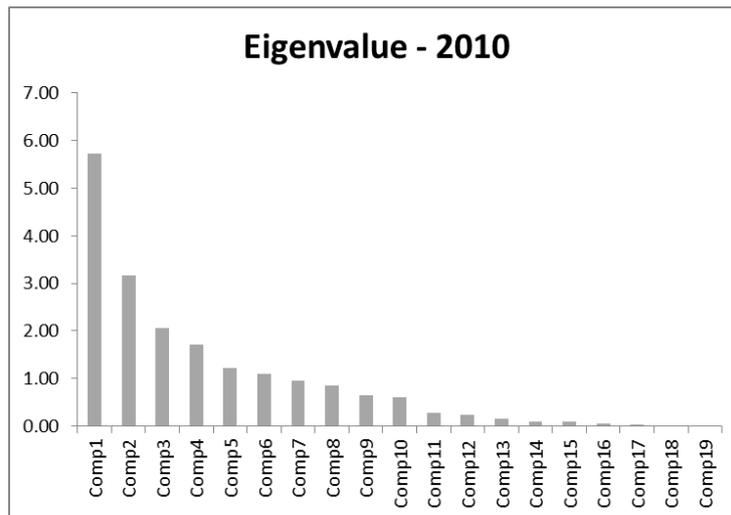
Análisis de Componentes Principales - Año 2010

Primeramente, se inicia el análisis de Componentes Principales para el Año 2010, donde se obtienen los siguientes resultados:

Tabla 25. Valores Propios de la Matriz de Covarianza Año 2010

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	5.73	2.57	0.30	0.30
Comp2	3.16	1.09	0.17	0.47
Comp3	2.07	0.36	0.11	0.58
Comp4	1.70	0.48	0.09	0.67
Comp5	1.23	0.12	0.06	0.73
Comp6	1.10	0.15	0.06	0.79
Comp7	0.95	0.11	0.05	0.84
Comp8	0.84	0.21	0.04	0.88
Comp9	0.64	0.03	0.03	0.92
Comp10	0.61	0.33	0.03	0.95
Comp11	0.28	0.05	0.01	0.96
Comp12	0.23	0.07	0.01	0.98
Comp13	0.17	0.06	0.01	0.98
Comp14	0.10	0.01	0.01	0.99
Comp15	0.09	0.03	0.00	0.99
Comp16	0.06	0.03	0.00	1.00
Comp17	0.03	0.02	0.00	1.00
Comp18	0.01	0.01	0.00	1.00
Comp19	0.00	.	0.00	1.00

Gráfico 25. Sedimentación Año 2010



Fuente: Elaboración Propia del Autor

Figura 4. Matriz de Variabilidades Máximas Año 2010

Variable	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Comp8	Comp9	Comp10	Comp11	Comp12	Comp13	Comp14	Comp15	Comp16	Comp17	Comp18	Comp19
X ₁	-0.1709	0.0085	0.1725	-0.4301	-0.0177	0.1169	0.0820	-0.3362	0.6310	0.2953	-0.2097	0.2728	-0.0706	0.0667	-0.0868	0.0341	-0.0025	-0.0176	0.0039
X ₂	0.1884	0.1044	-0.3052	0.4836	0.1207	0.0433	0.0478	0.0770	0.1513	-0.0757	-0.0904	0.7211	-0.0865	0.0586	-0.1298	0.0507	-0.1008	0.0666	0.0169
X ₃	0.0561	0.1887	-0.2633	0.0733	0.4363	0.2916	0.5263	-0.1778	0.2281	-0.0623	0.3880	-0.3005	0.0617	0.0126	0.0216	-0.0226	0.0248	0.0140	-0.0070
X ₄	-0.1064	0.4262	-0.0500	-0.2654	0.2468	-0.0046	0.0558	0.2526	-0.2530	0.0178	-0.2319	0.0200	-0.1480	0.6583	0.0131	0.0567	0.0033	-0.1698	0.0106
X ₅	-0.0832	0.4142	0.1007	0.3724	0.0072	-0.1948	-0.1069	-0.1702	0.1175	0.1999	-0.2828	-0.3188	-0.0406	0.0406	0.0103	-0.0165	-0.0668	0.5911	0.0389
X ₆	-0.1195	0.3789	-0.0346	-0.3804	0.1747	0.0280	0.1036	0.3445	-0.1459	-0.0061	-0.0785	0.1899	0.0476	-0.6442	-0.0297	-0.0280	-0.0441	0.2257	0.0104
X ₇	-0.0886	0.4665	0.1291	0.3079	-0.0700	-0.1459	-0.0552	-0.0695	0.1125	0.1367	-0.0288	-0.0585	0.0641	-0.2646	0.0420	0.0168	0.1204	-0.7069	-0.0554
X ₈	-0.1171	0.3734	0.2949	-0.0670	-0.2756	0.0116	-0.1651	-0.0832	-0.0112	-0.1581	0.6871	0.2540	0.0916	0.1908	-0.0451	-0.0638	-0.0346	0.1763	0.0109
X ₉	-0.0446	0.0032	-0.0624	0.0948	0.1626	0.6663	-0.5980	0.2400	0.1069	0.2635	0.0723	-0.1102	0.0135	-0.0089	-0.0046	-0.0123	0.0295	-0.0092	-0.0068
sector	0.1232	-0.0054	-0.3327	-0.0973	-0.1888	-0.4016	-0.0405	0.5537	0.4721	0.1911	0.2056	-0.1605	0.1273	0.1050	-0.0476	0.0198	0.0395	0.0224	-0.0078
pais	-0.3735	-0.0669	-0.1350	0.0839	-0.1603	0.0812	0.1522	0.0384	-0.0620	0.2001	0.0033	0.0706	0.3084	0.0450	0.4910	0.1135	-0.5599	-0.0503	0.2405
gii	0.2897	-0.0529	0.0427	-0.0371	0.1581	-0.1554	0.0493	-0.1412	-0.3432	0.7316	0.2515	0.1046	0.0019	-0.0228	-0.1707	0.2076	0.0221	0.0067	0.1952
tea	0.3704	0.1234	0.1344	-0.0645	-0.1394	0.1231	-0.0144	0.0602	0.1435	-0.2411	-0.0327	-0.1204	-0.2922	-0.0726	0.1075	0.3679	-0.0695	-0.0459	0.6695
pib	-0.3194	-0.1853	0.2225	0.1653	0.2531	-0.1175	0.0821	0.1927	0.0600	0.0560	0.0558	0.1050	-0.0612	0.0307	0.0748	-0.4790	0.3453	0.0067	0.5340
desempleo	0.2557	0.1772	-0.3603	-0.1744	-0.1869	0.0790	-0.1243	-0.2809	-0.0936	-0.0498	-0.1931	0.0175	0.5417	0.0585	-0.0742	-0.3556	0.1766	-0.0157	0.3137
corrupcion	0.3699	0.0325	0.2503	-0.0060	0.0603	0.0446	0.0677	0.0870	0.0567	0.0690	-0.0519	0.1584	0.1968	0.0497	0.7030	0.0620	0.3697	0.1243	-0.2335
enforcement	0.1914	-0.0731	0.5119	0.0752	0.2835	-0.0050	0.0541	0.2031	0.1118	-0.1228	-0.1176	-0.0378	0.5589	0.0812	-0.3294	0.0627	-0.2990	-0.0449	0.0112
legal	0.0478	0.0424	0.1569	0.1476	-0.5532	0.4062	0.4924	0.2685	-0.0943	0.2123	-0.1206	-0.0430	-0.0510	0.0499	-0.2212	-0.1397	0.1208	0.0411	-0.0612
inflacion	-0.3965	-0.0561	-0.0943	0.0800	-0.0353	0.0334	0.0353	-0.0023	-0.0393	-0.0968	-0.0325	0.0357	0.3125	0.0240	-0.1532	0.6419	0.5048	0.1020	0.0924

Fuente: Elaboración Propia del Autor

Análisis de Componentes Principales - Año 2011

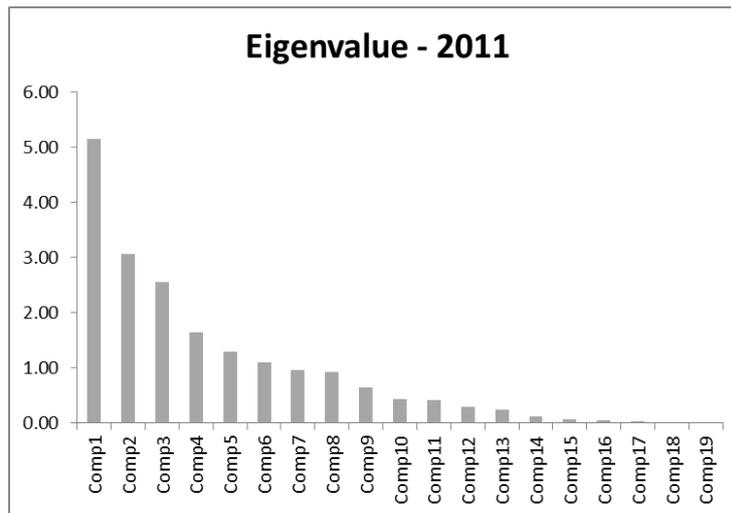
Se procede a realizar el análisis de Componentes Principales pero ahora para el Año 2011, donde se obtiene:

Tabla 26. Valores Propios de la Matriz de Covarianza Año 2011

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	5.16	2.09	0.27	0.27
Comp2	3.06	0.51	0.16	0.43
Comp3	2.55	0.91	0.13	0.57
Comp4	1.64	0.36	0.09	0.65
Comp5	1.29	0.19	0.07	0.72
Comp6	1.09	0.14	0.06	0.78
Comp7	0.96	0.03	0.05	0.83
Comp8	0.93	0.29	0.05	0.88
Comp9	0.64	0.20	0.03	0.91
Comp10	0.44	0.02	0.02	0.93
Comp11	0.42	0.13	0.02	0.96
Comp12	0.29	0.04	0.02	0.97
Comp13	0.25	0.13	0.01	0.99
Comp14	0.12	0.04	0.01	0.99
Comp15	0.07	0.03	0.00	1.00
Comp16	0.04	0.01	0.00	1.00
Comp17	0.03	0.01	0.00	1.00
Comp18	0.02	0.02	0.00	1.00
Comp19	0.00	.	0.00	1.00

Fuente: Elaboración Propia del Autor

Gráfico 26. Sedimentación Año 2011



Fuente: Elaboración Propia del Autor

Figura 5. Matriz de Variabilidades Máximas Año 2011

Variable	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Comp8	Comp9	Comp10	Comp11	Comp12	Comp13	Comp14	Comp15	Comp16	Comp17	Comp18	Comp19
X ₁	0.2132	0.0297	0.0530	-0.3776	0.0455	-0.1059	-0.4844	-0.0203	0.5678	0.2853	0.0923	0.0049	0.3768	-0.0186	0.0642	-0.0329	-0.0053	-0.0022	0.0018
X ₂	-0.2694	0.0910	-0.0894	0.4134	-0.1905	0.0240	0.1488	0.2615	-0.0837	0.2220	0.1790	0.0772	0.7102	0.0447	0.0852	0.0346	-0.0367	-0.0491	0.0056
X ₃	-0.1015	0.2825	0.0590	0.2261	-0.3397	0.1019	-0.4409	0.4201	0.0074	0.2889	-0.0747	-0.2020	-0.4104	0.2384	-0.0523	-0.0142	0.0192	0.0156	0.0025
X ₄	0.1789	0.2419	0.1357	-0.2754	-0.2787	0.2967	-0.1161	0.2653	-0.1324	-0.6369	-0.0511	0.0426	0.2583	0.0338	0.0273	-0.0328	0.2614	0.0102	-0.0009
X ₅	-0.0066	0.3222	0.2074	0.4218	-0.0112	-0.1953	-0.1456	-0.2587	0.3075	-0.4420	-0.0365	0.0364	0.0129	0.0285	-0.0396	0.1659	-0.4731	-0.0490	-0.0059
X ₆	0.1909	0.2683	0.2920	-0.2679	-0.0659	0.0075	0.2714	0.3227	-0.1259	0.2500	-0.0864	0.2101	-0.0640	-0.4273	0.0082	0.1030	-0.4715	-0.0522	-0.0087
X ₇	0.0600	0.3528	0.3627	0.2584	0.0768	-0.2402	0.1651	-0.0875	0.1492	0.1170	-0.0758	0.1222	-0.0576	-0.2733	0.0730	-0.2401	0.6106	0.0687	0.0044
X ₈	0.2034	0.2245	0.4681	-0.1552	0.2131	-0.0898	0.2776	-0.0654	-0.1452	0.1464	0.1149	-0.2140	0.1011	0.6954	-0.0721	0.1143	-0.0099	-0.0040	0.0081
X ₉	0.0697	0.0501	0.0090	0.0790	-0.2134	0.7361	0.3340	-0.2769	0.4247	0.1507	-0.0272	-0.0281	-0.0640	0.0409	-0.0024	-0.0108	-0.0111	0.0077	0.0012
sector	-0.1514	-0.0505	-0.2317	-0.1488	-0.1232	-0.3284	0.4391	0.4269	0.5283	-0.1862	-0.0391	0.0946	-0.1557	0.2239	0.0060	-0.0302	0.0278	0.0130	0.0007
pais	0.3538	0.0880	-0.3024	0.1373	0.0914	-0.0017	-0.0241	0.0372	0.0008	0.0740	-0.2046	0.2175	-0.0336	0.0453	0.0597	0.5840	0.2474	-0.4913	-0.0383
gii	-0.3334	0.3059	-0.1270	-0.1910	0.1276	0.0558	-0.0259	-0.1026	-0.0242	0.0067	0.0117	0.0606	-0.0355	0.0029	0.0717	0.0107	0.0095	-0.1557	0.8226
tea	0.3468	0.0270	-0.1175	0.1283	-0.0337	-0.0056	0.0229	0.0908	0.0161	-0.0653	0.8296	0.0312	-0.1624	-0.1147	-0.2644	-0.0705	0.0266	-0.0925	0.1462
pib	0.3415	-0.2940	0.0956	0.2057	-0.0794	-0.0351	0.0081	0.1208	0.0260	0.0035	-0.0941	0.0061	0.0232	-0.0019	0.1627	0.3315	0.0233	0.6178	0.4411
desempleo	-0.3065	0.3207	-0.1324	-0.1947	0.0450	0.0263	-0.0077	-0.1118	-0.0154	-0.0036	0.3556	-0.0036	-0.1240	-0.0798	0.3613	0.4567	0.0951	0.3737	-0.3102
corrupcion	-0.3421	-0.1351	0.2749	-0.0443	0.1706	0.1341	-0.0874	0.1017	0.0944	0.0038	0.0036	0.2942	0.0304	-0.0218	-0.6894	0.3380	0.1503	0.1121	-0.0243
enforcement	-0.1263	-0.3244	0.4002	0.0633	0.0948	0.1431	-0.1218	0.1070	0.0248	-0.0418	0.2075	0.4969	-0.1619	0.2089	0.4903	-0.0697	-0.0262	-0.2252	0.0007
legal	-0.0025	0.0254	0.0016	0.1656	0.7050	0.2790	-0.0050	0.4157	0.1490	-0.1267	-0.0165	-0.3806	0.0243	-0.1539	0.1176	-0.0278	-0.0308	-0.0186	-0.0261
inflacion	0.1737	0.3873	-0.3666	0.0725	0.2878	0.1215	-0.0860	0.0233	-0.0798	0.0476	-0.1115	0.5524	-0.0024	0.2382	-0.0883	-0.3197	-0.1190	0.3578	-0.0900

Fuente: Elaboración Propia del Autor

Análisis de Componentes Principales - Año 2012

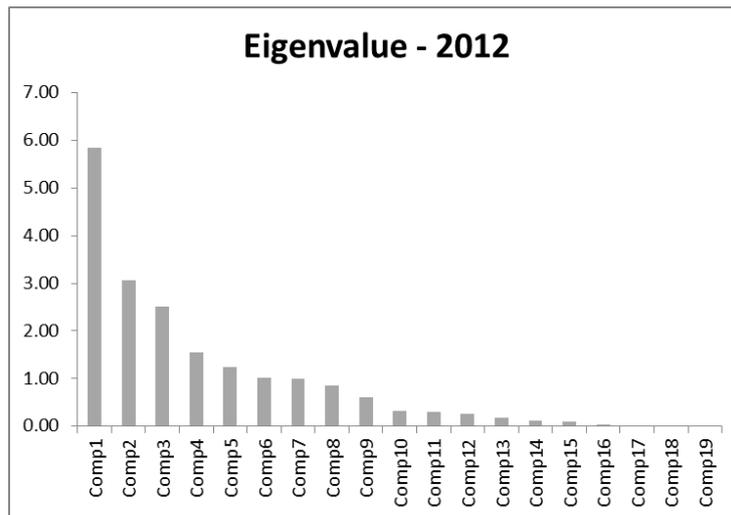
Se continúa con el análisis de Componentes Principales pero ahora para el Año 2012, donde se tiene:

Tabla 27. Valores Propios de la Matriz de Covarianza Año 2012

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	5.85	2.78	0.31	0.31
Comp2	3.07	0.56	0.16	0.47
Comp3	2.51	0.95	0.13	0.60
Comp4	1.56	0.31	0.08	0.68
Comp5	1.24	0.22	0.07	0.75
Comp6	1.02	0.03	0.05	0.80
Comp7	1.00	0.16	0.05	0.85
Comp8	0.84	0.24	0.04	0.90
Comp9	0.60	0.28	0.03	0.93
Comp10	0.33	0.04	0.02	0.95
Comp11	0.29	0.03	0.02	0.96
Comp12	0.27	0.10	0.01	0.98
Comp13	0.17	0.06	0.01	0.99
Comp14	0.11	0.01	0.01	0.99
Comp15	0.09	0.06	0.00	1.00
Comp16	0.03	0.01	0.00	1.00
Comp17	0.02	0.01	0.00	1.00
Comp18	0.01	0.01	0.00	1.00
Comp19	0.00	.	0.00	1.00

Fuente: Elaboración Propia del Autor

Gráfico 27. Sedimentación Año 2012



Fuente: Elaboración Propia del Autor

Figura 6. Matriz de Variabilidades Máximas Año 2012

Variable	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Comp8	Comp9	Comp10	Comp11	Comp12	Comp13	Comp14	Comp15	Comp16	Comp17	Comp18	Comp19
X ₁	0.1682	0.0365	-0.0888	-0.4115	-0.1615	0.3894	-0.2545	-0.1902	0.6099	0.0310	0.3265	-0.1679	0.0114	0.0304	-0.0742	-0.0049	-0.0156	0.0050	0.0007
X ₂	-0.2164	-0.0349	0.2389	0.4560	0.2217	-0.0359	-0.0751	0.1462	0.0227	0.1251	0.6649	-0.3537	0.0637	0.0892	-0.0884	0.0320	-0.0689	-0.0203	-0.0115
X ₃	-0.0544	0.1586	0.2315	0.1957	0.3939	0.4385	-0.4846	0.0880	0.0186	-0.3191	-0.1419	0.3860	-0.0740	0.0974	0.0246	-0.0521	0.0142	0.0126	0.0071
X ₄	0.1922	0.3410	0.0169	-0.2126	0.4044	0.0713	-0.0111	0.1180	-0.1554	0.3766	-0.1981	-0.4306	-0.2161	0.1165	-0.0134	-0.3891	0.1246	-0.0208	0.0104
X ₅	0.0303	0.3665	0.0417	0.4623	-0.2769	-0.0286	-0.0718	-0.0935	0.3162	0.2342	-0.2906	-0.1321	-0.3453	0.0411	0.2527	0.3250	-0.1008	-0.0246	-0.0044
X ₆	0.2024	0.3994	-0.0296	-0.2138	0.2014	0.0610	0.0641	0.3287	-0.1158	0.0097	0.0738	-0.0159	0.3273	-0.2966	0.0367	0.5875	-0.1865	-0.0076	-0.0289
X ₇	0.0876	0.3509	-0.0424	0.2714	-0.2838	-0.1185	-0.0126	0.1637	0.1824	-0.0591	-0.0278	0.1008	0.4417	-0.1838	-0.2995	-0.4468	0.1578	0.0332	0.0094
X ₈	0.1777	0.4043	-0.1392	-0.0902	-0.2035	-0.0649	0.1829	0.0754	-0.2210	-0.3052	0.3997	0.1930	-0.2998	0.5007	0.0979	-0.0425	-0.0369	-0.0107	0.0143
X ₉	0.0854	0.1112	-0.0075	0.0281	0.5709	-0.3860	0.3464	-0.3268	0.4719	-0.1494	0.0203	0.1781	-0.0004	0.0208	-0.0049	0.0026	-0.0024	0.0108	0.0048
sector	-0.1213	-0.1831	0.2289	-0.1659	-0.0466	-0.0793	0.1953	0.7454	0.4075	0.1076	-0.0964	0.1738	-0.1369	0.1772	0.0198	-0.0396	-0.0108	0.0138	0.0116
pais	0.3596	-0.1742	0.0934	0.1479	0.0029	0.1280	0.0907	-0.0333	0.0149	0.0862	-0.1088	-0.0298	0.4354	0.4846	0.1364	0.1571	0.3637	-0.3585	-0.1875
gii	0.1413	0.0559	0.5273	-0.1263	-0.1039	-0.0595	-0.0093	-0.2294	-0.1028	0.3853	0.0780	0.3237	0.1069	0.0332	-0.0153	-0.1131	-0.3874	-0.2214	0.3483
tea	0.3759	-0.1120	0.0425	0.1059	0.0028	0.0440	0.0343	0.0691	-0.0226	0.1604	0.3012	0.2654	-0.2292	-0.4913	0.4130	-0.1811	0.2451	-0.0409	-0.2808
pib	0.3400	-0.1600	-0.2302	0.1574	0.0653	-0.0547	-0.1226	0.0995	-0.0194	0.1538	0.0658	0.1401	-0.2144	-0.0192	-0.4071	0.2345	0.3240	0.0612	0.5701
desempleo	-0.1959	0.1806	0.4594	-0.1550	-0.0990	0.0753	0.1713	-0.1951	-0.0782	0.1256	0.0366	0.1032	-0.1080	-0.0026	-0.3004	0.2341	0.4894	0.3459	-0.2477
corrupcion	-0.3749	0.1309	-0.1423	-0.0736	0.0350	0.0689	0.0274	-0.0327	0.0233	0.1691	0.0939	0.0855	0.2615	0.0568	0.5654	-0.0286	0.3249	0.2153	0.4687
enforcement	-0.2167	0.0424	-0.4577	0.0127	0.1061	-0.0345	-0.1689	-0.0167	-0.0157	0.5450	0.0764	0.4087	0.0555	0.1962	-0.1604	0.0301	-0.1302	0.0204	-0.3804
legal	-0.0328	-0.0284	-0.1355	0.2333	0.0374	0.6592	0.6386	-0.0358	-0.0010	0.0726	-0.0239	0.0558	-0.0253	-0.0589	-0.1309	-0.0801	-0.1844	0.0311	0.0939
inflacion	0.3877	-0.1599	0.0644	0.1093	0.0035	-0.0264	-0.0521	0.0051	-0.0163	0.0486	-0.0471	-0.0374	0.1740	0.1978	0.1145	-0.0533	-0.2541	0.8040	-0.0593

Fuente: Elaboración Propia del Autor

Análisis de Componentes Principales - Año 2013

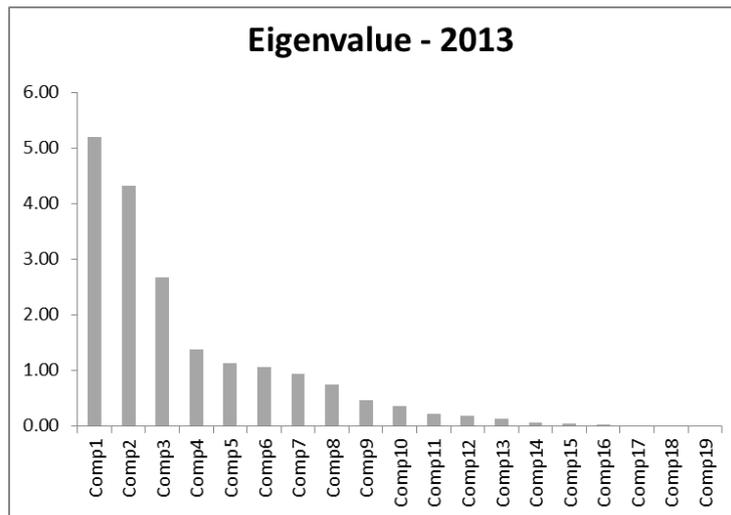
Ahora bien, el análisis de Componentes Principales pero ahora para el Año 2013, se obtiene los resultados:

Tabla 28. Valores Propios de la Matriz de Covarianza Año 2013

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	5.20	0.87	0.27	0.27
Comp2	4.33	1.65	0.23	0.50
Comp3	2.68	1.29	0.14	0.64
Comp4	1.39	0.26	0.07	0.72
Comp5	1.13	0.06	0.06	0.77
Comp6	1.07	0.13	0.06	0.83
Comp7	0.93	0.18	0.05	0.88
Comp8	0.75	0.29	0.04	0.92
Comp9	0.47	0.11	0.02	0.94
Comp10	0.36	0.13	0.02	0.96
Comp11	0.23	0.04	0.01	0.97
Comp12	0.19	0.05	0.01	0.99
Comp13	0.14	0.07	0.01	0.99
Comp14	0.07	0.02	0.00	1.00
Comp15	0.05	0.03	0.00	1.00
Comp16	0.02	0.02	0.00	1.00
Comp17	0.00	0.00	0.00	1.00
Comp18	0.00	0.00	0.00	1.00
Comp19	0.00	.	0.00	1.00

Fuente: Elaboración Propia del Autor

Gráfico 28. Sedimentación Año 2013



Fuente: Elaboración Propia del Autor

Figura 7. Matriz de Variabilidades Máximas Año 2013

Variable	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Comp8	Comp9	Comp10	Comp11	Comp12	Comp13	Comp14	Comp15	Comp16	Comp17	Comp18	Comp19
X ₁	0.1377	0.0740	-0.0157	-0.5851	-0.1246	0.3306	0.1222	0.4606	-0.0101	0.4965	0.0113	0.1552	-0.0995	0.0123	0.0044	-0.0044	0.0107	0.0052	-0.0008
X ₂	-0.1387	-0.2502	0.0632	0.5163	0.1037	-0.0642	0.2245	-0.0912	0.0820	0.6462	0.0087	0.3595	-0.1249	0.0311	-0.0097	-0.0435	0.0406	0.0117	0.0055
X ₃	-0.0202	-0.2580	0.0962	0.1235	0.3332	0.2665	0.5427	0.3316	0.3620	-0.2102	0.0764	-0.3712	0.0527	0.0085	0.0061	0.0095	-0.0126	-0.0032	0.0038
X ₄	-0.1409	0.4025	0.1460	0.1543	-0.0188	-0.0381	0.1835	0.2446	-0.1182	-0.1581	-0.0975	0.2183	-0.0239	0.0003	0.0028	0.0028	-0.0019	0.0139	0.7621
X ₅	-0.1354	0.3291	0.1321	0.1475	-0.0186	-0.0339	0.1133	0.1009	-0.0783	0.0836	0.0500	-0.0848	-0.0147	0.0469	-0.0202	0.0127	-0.7711	0.0266	-0.3357
X ₆	-0.1406	0.4072	0.1429	0.1421	-0.0120	-0.0283	0.1803	0.2283	-0.0821	-0.1723	-0.1012	0.2151	-0.0174	-0.0162	0.0073	-0.0147	0.5520	-0.0326	-0.5373
X ₇	-0.0867	0.4173	0.0603	0.0584	-0.0473	-0.0341	-0.0629	-0.1726	0.1173	0.4256	0.2395	-0.6377	0.0805	0.0014	-0.0115	0.0210	0.3014	-0.0140	0.1311
X ₈	0.0012	0.4341	-0.0125	-0.1591	0.2177	0.2448	-0.0644	-0.3932	0.6856	-0.0656	-0.0251	0.3284	-0.0651	-0.0550	0.0370	-0.0551	-0.0466	0.0111	0.0108
X ₉	0.0637	0.0468	0.0046	0.1181	0.6241	-0.1257	-0.6021	0.4482	0.0570	0.0830	0.0222	0.0057	-0.0155	0.0031	0.0030	0.0046	0.0038	0.0019	-0.0008
sector	-0.0977	-0.1391	0.1774	0.1356	-0.6023	-0.2126	-0.2592	0.3582	0.5291	-0.0752	0.1674	0.0282	-0.0452	-0.0138	0.0202	0.0094	-0.0231	0.0034	-0.0059
pais	0.4017	0.0557	0.0694	0.1975	-0.0796	0.1429	-0.0153	0.0243	-0.0506	0.0389	0.0494	0.0486	0.3796	-0.2116	0.6951	-0.2547	-0.0322	-0.1316	-0.0021
gii	0.2054	-0.0338	0.4807	-0.1508	0.1077	-0.1421	0.0524	-0.1186	-0.0601	-0.0672	0.4327	0.1357	-0.0063	0.6526	0.0763	0.0102	0.0374	-0.0853	0.0094
tea	0.4196	0.0792	-0.0346	0.1025	-0.0315	-0.0481	0.0676	0.0312	0.0001	-0.0246	0.2127	0.0673	0.1736	-0.1065	-0.4557	-0.4130	0.0227	0.5713	-0.0067
pib	0.2926	0.1153	-0.3763	0.1104	-0.0069	-0.1265	0.1214	0.0325	0.0042	-0.0744	0.2502	-0.0188	-0.5655	0.0042	0.3386	0.3493	0.0280	0.2993	-0.0063
desempleo	-0.2373	-0.1133	0.4178	-0.0669	0.0688	0.2388	-0.0994	-0.0930	-0.1997	-0.0650	0.4724	0.0987	-0.0961	-0.5673	0.0140	0.1980	0.0078	0.1552	-0.0018
corrupcion	-0.3859	-0.0270	-0.2407	-0.0759	0.0209	0.0159	-0.0072	0.0411	0.0254	0.0315	0.0963	0.1258	0.5558	0.2427	0.2258	0.2906	0.0192	0.5037	-0.0127
enforcement	-0.1883	0.0441	-0.4747	-0.1088	0.0789	-0.2572	0.1587	0.0759	0.0016	-0.0467	0.5574	0.1615	0.1270	-0.1317	-0.0787	-0.2290	-0.0203	-0.4364	0.0100
legal	-0.0356	0.0347	-0.2356	0.3576	-0.1687	0.7121	-0.2492	0.0311	-0.1305	-0.0999	0.2106	0.0266	-0.0657	0.3098	-0.1567	-0.0265	0.0224	-0.1317	0.0086
inflacion	0.4184	0.0742	0.0005	0.1212	-0.0315	-0.0247	0.0612	0.0094	0.0674	0.0498	0.0149	0.1204	0.3535	-0.1308	-0.3228	0.6780	-0.0133	-0.2673	-0.0034

Fuente: Elaboración Propia del Autor

Análisis de Componentes Principales - Año 2014

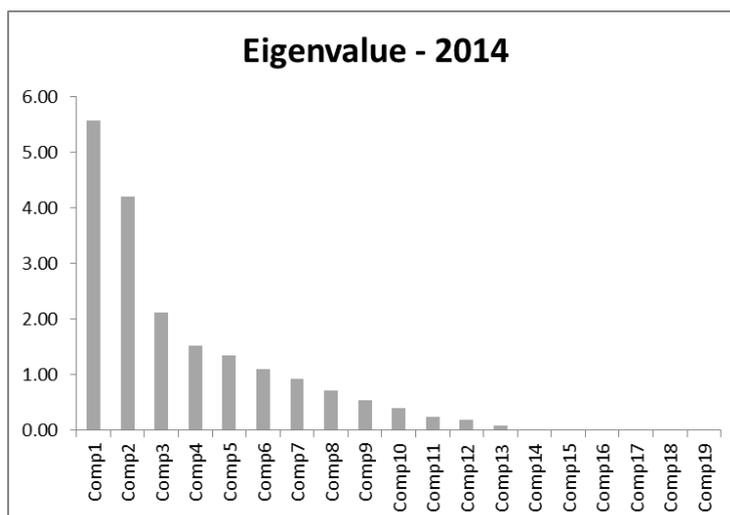
El análisis de Componentes Principales pero ahora para el Año 201, se tendrían los siguientes resultados:

Tabla 29. Valores Propios de la Matriz de Covarianza Año 2014

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	5.57	1.37	0.29	0.29
Comp2	4.21	2.08	0.22	0.51
Comp3	2.13	0.60	0.11	0.63
Comp4	1.52	0.18	0.08	0.71
Comp5	1.35	0.24	0.07	0.78
Comp6	1.11	0.18	0.06	0.84
Comp7	0.92	0.22	0.05	0.88
Comp8	0.70	0.17	0.04	0.92
Comp9	0.54	0.14	0.03	0.95
Comp10	0.40	0.16	0.02	0.97
Comp11	0.24	0.06	0.01	0.98
Comp12	0.18	0.09	0.01	0.99
Comp13	0.09	0.07	0.00	1.00
Comp14	0.02	0.00	0.00	1.00
Comp15	0.02	0.01	0.00	1.00
Comp16	0.01	0.00	0.00	1.00
Comp17	0.00	0.00	0.00	1.00
Comp18	0.00	0.00	0.00	1.00
Comp19	0.00	.	0.00	1.00

Fuente: Elaboración Propia del Autor

Gráfico 29. Sedimentación Año 2014



Fuente: Elaboración Propia del Autor

Figura 8. Matriz de Variabilidades Máximas Año 2014

Variable	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Comp8	Comp9	Comp10	Comp11	Comp12	Comp13	Comp14	Comp15	Comp16	Comp17	Comp18	Comp19
X ₁	-0.0957	0.0629	-0.0630	-0.5354	0.4063	0.0981	0.2120	0.4422	-0.1634	-0.0849	0.4905	-0.0031	0.0727	0.0213	0.0127	-0.0136	-0.0213	0.0040	-0.0031
X ₂	0.0643	-0.2534	0.1951	0.4899	-0.2708	0.0174	0.1936	0.0072	0.0031	0.0166	0.7189	0.0817	0.1006	0.0416	0.0554	-0.0133	-0.0582	0.0128	0.0020
X ₃	-0.0386	-0.2443	0.1710	0.1330	0.0348	-0.2081	0.6276	0.4642	0.3113	0.0357	-0.3697	-0.0207	-0.0461	0.0026	0.0194	0.0015	0.0082	0.0086	-0.0015
X ₄	0.2381	0.3819	0.0962	0.0437	-0.1221	0.0559	0.0738	0.1145	-0.1090	-0.0410	-0.0523	-0.0291	-0.0032	0.1032	0.2753	0.0212	-0.0737	0.7905	-0.1237
X ₅	0.2427	0.3707	0.1170	0.0751	-0.1057	0.0426	0.1194	0.1147	-0.0683	-0.0507	0.0052	0.0046	0.0607	-0.2819	-0.7780	0.0459	-0.2105	-0.0423	-0.0019
X ₆	0.2413	0.3822	0.0906	0.0359	-0.1082	0.0442	0.0733	0.1100	-0.0819	-0.0248	-0.0619	-0.0298	-0.0207	0.1593	0.4316	0.0770	-0.4802	-0.5391	0.0616
X ₇	0.2511	0.3789	0.0921	0.0625	-0.0535	0.0232	0.0947	0.0707	0.0218	-0.0015	0.0525	0.0256	0.0062	0.0147	0.0888	-0.1485	0.8177	-0.2365	0.0705
X ₈	0.1801	0.2404	-0.0884	-0.0457	0.3273	-0.1973	0.0010	-0.2466	0.7094	0.3236	0.2273	0.1019	-0.0551	0.0319	-0.0059	-0.0036	-0.1226	0.0609	-0.0121
X ₉	-0.0396	0.0659	0.0043	0.0949	-0.0661	-0.7256	-0.4958	0.4468	-0.0388	-0.0156	0.0821	0.0017	0.0121	-0.0017	-0.0071	-0.0066	0.0040	-0.0005	-0.0006
sector	0.0607	-0.1379	0.2022	-0.0999	-0.2180	0.5398	-0.4358	0.4512	0.3540	0.2375	-0.0447	0.0935	0.0352	-0.0084	-0.0088	-0.0144	0.0010	-0.0012	-0.0024
pais	-0.3473	0.1791	0.2090	0.0504	0.0372	0.0241	0.0046	-0.0157	-0.0292	-0.1161	-0.0378	0.6547	-0.2341	0.5029	-0.1762	0.0742	0.0027	0.0159	0.1001
gii	-0.2764	0.1793	-0.3323	0.0740	-0.2166	0.0934	0.0182	0.0821	0.3019	-0.2917	0.0850	-0.4129	0.0681	0.2597	-0.0896	0.4243	0.0724	0.0367	0.3038
tea	-0.3740	0.2052	0.0556	0.0356	-0.0432	0.0002	0.0534	-0.0155	-0.0071	0.1455	-0.0177	0.2311	0.2248	-0.5529	0.2192	-0.0359	-0.0469	0.0908	0.5638
pib	-0.2245	0.1154	0.2719	0.2915	0.3590	0.0900	-0.0559	0.0395	-0.2553	0.5451	-0.0063	-0.4395	-0.1470	0.1836	-0.0893	0.0633	0.0215	0.0119	0.1111
desempleo	0.2498	-0.1711	0.4286	-0.1960	0.1159	-0.1305	-0.0441	-0.1550	-0.0461	0.0087	-0.0655	0.0615	0.4744	0.0913	0.0044	0.6014	0.0992	0.0121	0.1200
corrupcion	0.3268	-0.1544	-0.2847	0.1477	0.1529	0.0704	-0.0287	0.1329	-0.1396	0.0470	0.0320	0.2069	-0.5574	-0.2419	0.0608	0.4834	0.0682	0.0313	0.2048
enforcement	0.1218	-0.0188	-0.5852	0.1828	0.0207	0.0464	0.0743	0.1454	-0.1808	0.3843	-0.1146	0.2589	0.4975	0.2569	-0.0729	-0.0203	0.0017	-0.0027	0.0024
legal	0.0090	0.0347	0.0205	0.4801	0.5796	0.2010	-0.1888	0.0976	0.1041	-0.4848	-0.0873	0.0439	0.2402	-0.1130	0.0722	-0.0554	-0.0235	-0.0115	-0.1067
inflacion	-0.3733	0.2094	-0.0407	0.0582	-0.0921	0.0249	0.0594	0.0211	0.0166	0.1522	0.0337	0.1008	0.0377	-0.2751	0.1077	0.4233	0.0951	-0.1072	-0.6896

Fuente: Elaboración Propia del Autor

Para elegir con que componentes nos quedaremos habrá que seguir algunos criterios:

- 1) Criterio de Kaiser: Este criterio menciona que los autovalores (Eigenvalue) han ser mayores que 1, ya que son estos los que mayor varianza explican. La media de todos estos valores es igual a 1.
- 2) Elegir un mínimo de varianza que se desea explicar. Para este estudio fijamos un objetivo entre 65% y 75%.

El gráfico de sedimentación representa la proporción de información que ganamos con cada factor extra y nos ayuda a distinguir con mayor claridad los puntos de inflexión. Por ejemplo en el años 2010 se pueden distinguir dos puntos uno entre el Comp2 y el Comp3 y el segundo entre el Comp4 y Comp5.

Tomando en cuenta estos criterios, hemos elegido 4 componentes para cada uno de los años analizados y finalmente se procede ahora a su interpretación observando que variables saturan a cada factor en cada uno de los años (los que más peso tienen dentro de cada uno de ellos).

En el espacio de las variables, el análisis tiene sentido si existen variabilidades positivas de las variables, ya que esto es indicativo de su mayor incidencia sobre la variabilidad absoluta total, y por tanto los demás factores tendrán poca incidencia.

El siguiente esquema muestra las variables que saturan cada factor por cada uno de los años y para los 4 componentes principales seleccionados:

Figura 9. Saturación de los Componentes por la Variables Explicativas por Año de Observación

Componente1						Componente2						Componente3						Componente4					
Variable	2010	2011	2012	2013	2014	Variable	2010	2011	2012	2013	2014	Variable	2010	2011	2012	2013	2014	Variable	2010	2011	2012	2013	2014
X ₁						X ₁						X ₁						X ₁					
X ₂						X ₂						X ₂						X ₂	✓	✓	✓	✓	✓
X ₃						X ₃						X ₃						X ₃					
X ₄						X ₄	✓				✓	X ₄						X ₄					
X ₅						X ₅	✓					X ₅						X ₅		✓	✓		
X ₆						X ₆			✓	✓		X ₆						X ₆					
X ₇						X ₇	✓			✓	✓	X ₇						X ₇					
X ₈						X ₈			✓	✓		X ₈		✓				X ₈					
X ₉						X ₉						X ₉						X ₉					
sector						sector						sector						sector					
pais				✓		pais						pais						pais					
gii				✓		gii						gii			✓	✓		gii					
tea				✓		tea						tea			✓			tea					
pib						pib						pib						pib					
desempleo						desempleo						desempleo			✓	✓	✓	desempleo					
corrupcion						corrupcion						corrupcion						corrupcion					
enforcement						enforcement						enforcement	✓	✓				enforcement					
legal						legal						legal						legal					✓
inflacion			✓	✓		inflacion		✓				inflacion						inflacion					

Fuente: Elaboración Propia del Autor

Tabla 30. Frecuencia de Aparición de las Variables Explicativas en la Saturación de los Componentes

Variable	Comp 1	Comp 2	Comp 3	Comp 4	Total
X ₁	0	0	0	0	0
X ₂	0	0	0	5	5
X ₃	0	0	0	0	0
X ₄	0	2	0	0	2
X ₅	0	1	0	2	3
X ₆	0	2	0	0	2
X ₇	0	2	0	0	2
X ₈	0	2	1	0	3
X ₉	0	0	0	0	0
sector	0	0	0	0	0
pais	1	0	0	0	1
gii	0	0	2	0	2
tea	1	0	0	0	1
pib	0	0	0	0	0
desempleo	0	0	3	0	3
corrupcion	0	0	0	0	0
enforcement	0	0	2	0	2
legal	0	0	0	1	1
inflacion	2	1	0	0	3

Fuente: Elaboración Propia del Autor

Toda vez que ya se realizó el ACP, contamos con algunas “pistas” de las variables independientes que podrían formar parte del modelo sin perder información del resto de ellas. Recordemos que para los GEE hay que especificar:

- 1) La elección de la función de enlace
- 2) La distribución de la variable dependiente
- 3) La estructura de correlación de la variable dependiente

Con este objetivo primero observamos gráficamente el comportamiento anual que presenta la variable de respuesta y tal como lo muestra la siguiente gráfica, no parece haber diferencias a lo largo del tiempo. Para el presente estudio y después de haber analizado tato la variable respuesta como las variables independientes, se propone en un primer modelo:

- 1) Función de enlace: Log-Log
- 2) Distribución de la Variable Dependiente: Binomial
- 3) Estructura de Correlación: No estructurada

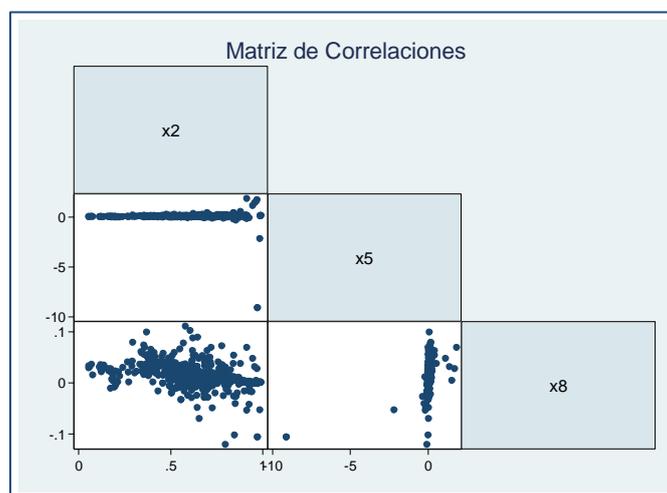
Después de realizar algunas pruebas, presentamos los dos escenarios que mejor ajustan al modelo de EEG bajo los supuestos mencionados y los datos utilizados en las siguientes secciones.

4.2.2 Iteraciones del Modelo (EEG)

Iteración 1: Variables: $X_2 / X_5 / X_8$

Figura 10. Matriz de Correlación Variables ($X_2 / X_5 / X_8$)

	x2	x5	x8
x2	1.0000		
x5	-0.0759	1.0000	
x8	-0.3742	0.4083	1.0000



Fuente: Elaboración Propia del Autor

Figura 11. Resultados Iteración 1 Modelo EEG

GEE population-averaged model		Number of obs	=	435
Group and time vars: empresa periodo		Number of groups	=	90
Link:	cloglog	Obs per group: min	=	1
Family:	binomial	avg	=	4.8
Correlation:	unstructured	max	=	5
		Wald chi2(3)	=	93.13
Scale parameter:	1	Prob > chi2	=	0.0000

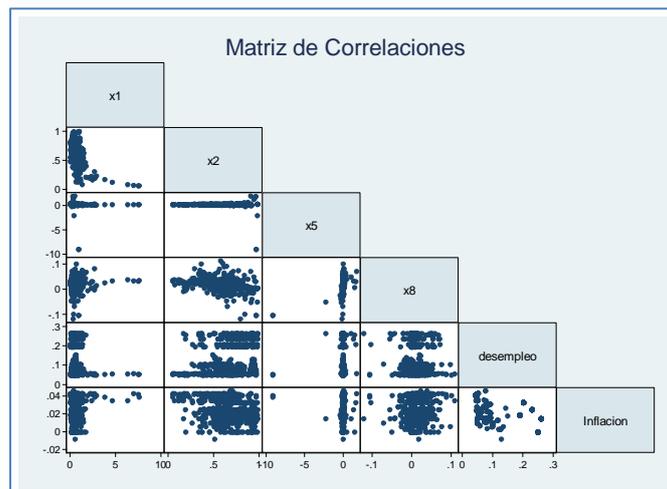
y_altman	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
x2	7.114805	.9087602	7.83	0.000	5.333668	8.895943
x5	-1.073765	.5443223	-1.97	0.049	-2.140617	-.0069126
x8	-30.90444	6.10992	-5.06	0.000	-42.87966	-18.92922
_cons	-4.390566	.6055193	-7.25	0.000	-5.577362	-3.20377

Fuente: Elaboración Propia del Autor

Iteración 2: Variables: X₁ / X₂ / X₅ / X₈ / Desempleo / Inflación

Figura 12. Matriz de Correlación Variables (X₂ / X₅ / X₈ / Desempleo / Inflación)

	x1	x2	x5	x8	desemp-o	inflac-n
x1	1.0000					
x2	-0.5377	1.0000				
x5	-0.0370	-0.0795	1.0000			
x8	0.1801	-0.3716	0.4102	1.0000		
desempleo	-0.1551	0.3072	0.0675	-0.0686	1.0000	
inflacion	0.2010	-0.2470	-0.0628	0.0430	-0.5125	1.0000



Fuente: Elaboración Propia del Autor

Figura 13. Resultados Iteración 2 Modelo EEG

GEE population-averaged model		Number of obs	=	435
Group and time vars: empresa periodo		Number of groups	=	90
Link:	cloglog	Obs per group: min	=	1
Family:	binomial	avg	=	4.8
Correlation:	unstructured	max	=	5
		Wald chi2(6)	=	107.83
Scale parameter:	1	Prob > chi2	=	0.0000

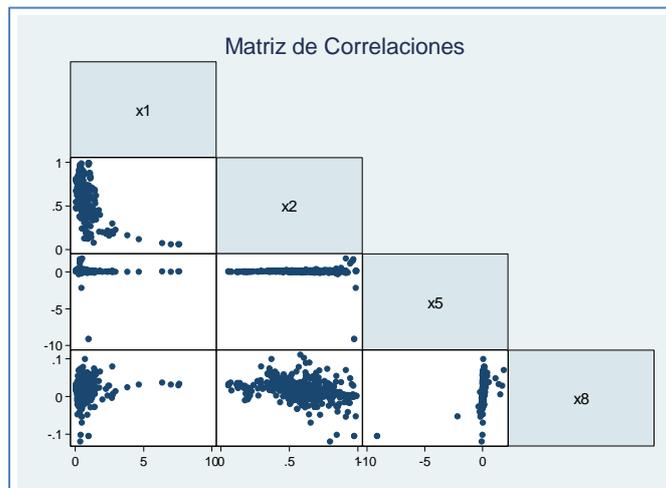
y_altman	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
x1	-1.932693	.3964458	-4.88	0.000	-2.709713 -1.155674
x2	6.742163	.9759349	6.91	0.000	4.829365 8.65496
x5	-1.501229	.5540472	-2.71	0.007	-2.587141 -.415316
x8	-33.82928	6.589504	-5.13	0.000	-46.74447 -20.91409
desempleo	3.56673	1.33255	2.68	0.007	.9549787 6.17848
inflacion	13.18571	7.896662	1.67	0.095	-2.291459 28.66289
_cons	-3.759169	.7505671	-5.01	0.000	-5.230253 -2.288084

Fuente: Elaboración Propia del Autor

Iteración 3: Variables: $X_1 / X_2 / X_5 / X_8$

Figura 14. Matriz de Correlación Variables (X2 / X2/ X5 / X8)

	x1	x2	x5	x8
x1	1.0000			
x2	-0.5377	1.0000		
x5	-0.0370	-0.0795	1.0000	
x8	0.1801	-0.3716	0.4102	1.0000



Fuente: Elaboración Propia del Autor

Figura 15. Resultados Iteración 3 Modelo EEG

GEE population-averaged model		Number of obs	=	435
Group and time vars: empresa periodo		Number of groups	=	90
Link:	cloglog	Obs per group: min	=	1
Family:	binomial	avg	=	4.8
Correlation:	unstructured	max	=	5
		Wald chi2(4)	=	116.96
Scale parameter:	1	Prob > chi2	=	0.0000

y_altman	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
x1	-1.914055	.3610832	-5.30	0.000	-2.621765	-1.206345
x2	6.880732	.8730574	7.88	0.000	5.169571	8.591893
x5	-1.240081	.5304855	-2.34	0.019	-2.279814	-.2003486
x8	-31.39007	6.326883	-4.96	0.000	-43.79053	-18.98961
_cons	-3.215395	.6216582	-5.17	0.000	-4.433823	-1.996967

Fuente: Elaboración Propia del Autor

4.2.3 Interpretación de Resultados del Modelo (EEG)

Iteración 1: Variables: X_2 / X_5 / X_8

Tabla 31. Variables Significativas Iteración 1

Variable en el Modelo	Definición	Factor
X2	Total Pasivo / Total Activo	Endeudamiento
X5	EBIT / Total Capital Contable	Rentabilidad
X8	Utilidad Neta / Total Activo	Rentabilidad

Fuente: Elaboración Propia del Autor

En una primera iteración se han introducido en el modelo de ecuaciones de estimación generalizadas las variables microeconómicas que presentaron mayor saturación de los componentes principales, obteniendo que todas salen significativas con un nivel de confianza del 99.9%, excepto X_5 con un 95% de confianza. La prueba Wald χ^2 arroja un valor de 93.13

Tabla 32. Significancia Estadística Iteración 1

Wald χ^2 (3) = 93.13		
Variable	P- Value (P > z)	Coficiente
X ₂	0.000	7.114
X ₅	0.049	-1.073
X ₈	0.000	-30.904

Fuente: Elaboración Propia del Autor

Interpretación de los coeficientes en términos financieros

- Coeficiente positivo de la variable X₂ (Solvencia Global): Indica que es más probable que una empresa quiebre cuando más depende de su deuda. Es decir por cada unidad monetaria del activo total más depende de recursos ajenos.
- Coeficiente negativo de la variable X₅ (Rendimiento del Capital): Indica que es más probable que una empresa quiebre cuando por cada unidad monetaria de capital invertido por los accionistas genere menos utilidad antes de impuestos e intereses.
- Coeficiente negativo de la variable X₈ (Return on Assets): Indica que es más probable que una empresa quiebre cuando por cada unidad monetaria invertido en activo se genere menos utilidad neta.

Iteración 2: Variables: X₁ / X₂ / X₅ / X₈ / Desempleo / Inflación

Tabla 33. Variables Significativas Iteración 2

Variable en el Modelo	Definición	Factor
X1	Activo Circulante / Total Pasivo	Liquidez
X2	Total Pasivo / Total Activo	Endeudamiento
X5	EBIT / Total Capital Contable	Rentabilidad
X8	Utilidad Neta / Total Activo	Rentabilidad
Desempleo	Variable Macroeconómica	
Inflación	Variable Macroeconómica	

Fuente: Elaboración Propia del Autor

En una segunda iteración se han introducido en el modelo de ecuaciones de estimación generalizadas tanto variables microeconómicas (añadiendo el ratio de liquidez X_1) como variables macroeconómica, de igual manera, siguiendo el criterio de ser las que presentaron mayor saturación de los componentes principales. Se obtiene que todas salen significativas con un nivel de confianza arriba del 99% excepto la inflación con un 90%. La prueba Wald χ^2 arroja un valor de 107.83, superior al 93.13 obtenido en la primera iteración sin las variables macroeconómicas, lo cual evidencia que se ha mejorado el poder predictivo del modelo al incluir estas variables.

Tabla 34. Significancia Estadística Iteración 2

Wald χ^2 (6) = 107.83		
Variable	P- Value (P > z)	Coficiente
X_1	0.000	-1.932
X_2	0.000	6.742
X_5	0.007	-1.501
X_8	0.000	-33.829
Desempleo	0.007	3.566
Inflación	0.095	13.185

Fuente: Elaboración Propia del Autor

Interpretación de los coeficientes en términos financieros

- Coeficiente negativo de la variable X_1 (Coeficiente de Liquidez Circulante): Indica que es más probable que una empresa quiebre cuando por cada peso de duda tiene menos activos líquidos para hacer frente a sus obligaciones.
- Coeficiente positivo de la variable X_2 (Solvencia Global): Indica que es más probable que una empresa quiebre cuando más depende de su deuda. Es decir por cada unidad monetaria del activo total más depende de recursos ajenos.
- Coeficiente negativo de la variable X_5 (Rendimiento del Capital): Indica que es más probable que una empresa quiebre cuando por cada unidad monetaria de capital invertido por los accionistas genere menos utilidad ante de impuestos e intereses.

- Coeficiente negativo de la variable X_8 (Return on Assets): Indica que es más probable que una empresa quiebre cuando por cada unidad monetaria invertido en activo se genere menos utilidad neta.
- Coeficiente positivo del Desempleo: Indica que es más probable que una empresa quiebre cuanto mayor sea la tasa de desempleo del país, es decir la generación de empleos no crece al ritmo deseado y por lo tanto la producción de la empresa no crece a un ritmo mayor.
- Coeficiente positivo de la Inflación: Indica que entre mayor sea esta tasa, más probable será que una empresa quiebre, es decir, si suben los precios, podremos adquirir menos bienes y servicios con los mismo ingresos, por lo que se dice que hay pérdida de *poder adquisitivo*

Iteración 3: Variables: $X_1 / X_2 / X_5 / X_8$

Tabla 35. Variables Significativas Iteración 3

Variable en el Modelo	Definición	Factor
X1	Activo Circulante / Total Pasivo	Liquidez
X2	Total Pasivo / Total Activo	Endeudamiento
X5	EBIT / Total Capital Contable	Rentabilidad
X8	Utilidad Neta / Total Activo	Rentabilidad

Fuente: Elaboración Propia del Autor

En una tercera iteración se han eliminado las variables macro económicas y se ha añadido la variable microeconómica X_1 . Con esta última iteración se mejora la prueba Wald χ^2 arrojando un valor de 116.96, superior al 107. 83 obtenido en la segunda. En esta iteración todas las variables presentan un nivel de confianza arriba de 99%.

Tabla 36. Significancia Estadística Iteración 3

Wald χ^2 (4) = 116.96		
Variable	P- Value (P > z)	Coficiente
X ₁	0.000	-1.91
X ₂	0.000	6.88
X ₅	0.019	-1.24
X ₈	0.000	-31.39

Fuente: Elaboración Propia del Autor

Interpretación de los coeficientes en términos financieros

- Coeficiente negativo de la variable X₁ (Coeficiente de Liquidez Circulante): Indica que es más probable que una empresa quiebre cuando por cada peso de duda tiene menos activos líquidos para hacer frente a sus obligaciones.
- Coeficiente positivo de la variable X₂ (Solvencia Global): Indica que es más probable que una empresa quiebre cuando más depende de su deuda. Es decir por cada unidad monetaria del activo total más depende de recursos ajenos.
- Coeficiente negativo de la variable X₅ (Rendimiento del Capital): Indica que es más probable que una empresa quiebre cuando por cada unidad monetaria de capital invertido por los accionistas genere menos utilidad antes de impuestos e intereses.
- Coeficiente negativo de la variable X₈ (Return on Assets): Indica que es más probable que una empresa quiebre cuando por cada unidad monetaria invertido en activo se genere menos utilidad neta.

Después de la tercera iteración se mantiene las mismas variables como significativas dentro del modelo de EEG, además de que se fue mejorando la capacidad predictiva del mismo añadiendo la variable X₁.

Así podemos concluir que las variables dentro de la tercera iteración son las que presentarían mayor poder predictivo con relación a la variable dependiente Y_{Altman} , es decir a la posible quiebra de una empresa.

4.3 Aplicación Empírica de los Árboles de Decisión

Presentaremos y analizaremos los resultados obtenidos tras la aplicación de un árbol C4.5. En el presente trabajo, se cuenta con un conjunto de empresas caracterizadas por una serie de valores cuantitativos, es decir variables independientes y que tienen origen tanto micro como macroeconómico que intentamos asignar a alguna de las dos categorías: quiebra o no quiebra. La distinción entre estas dos categorías en la muestra está representada por el valor que toma la variable respuesta o variable dependiente (Y) siendo: 1 en el caso de quiebra y 0 en el caso contrario, es decir empresa sana.

Si bien los datos que disponemos presentan la característica de datos panel, para la aplicación de estos modelos únicamente se realizarán dos cortes transversales, uno para el año 2010 y otro para el año 2014, con lo cual podremos analizar el principio y el final de la serie de tiempo. En la siguiente tabla se muestra el porcentaje de empresas en situación de quiebra y no quiebra los dos años analizados, según la variable respuesta.

Tabla 37. Distribución de la Muestra en Función de las Variable Respuesta

Año	NO Quiebra	Quiebra
2010	55%	45%
2014	53%	47%

Fuente: Elaboración Propia del Autor

Estas dos categorías han sido evaluadas de conformidad con los valores que toman las 19 variables respuesta para el año 2010 y 2014 respectivamente y para el total de las empresas. Esta base de datos se ha introducido y programado en WEKA, que como mencionamos en el capítulo anterior, es uno de los softwares informáticos que ha desarrollado el análisis del algoritmo C4.5

4.3.1 Corte Transversal Año 2010

4.3.1.1 Validación de Resultados

A continuación presentamos la salida de los resultados obtenidos en el programa WEKA para el año 2010:

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: BANKRUPTCY_MACRO_2010-

Instances: 81

Attributes: 20

X1_AC_TP
X2_TP_TA
X3_TP_TCC
X4_EBIT_INGR
X5_EBIT_TCC
X6_UN_INGRN
X7_UN_TCC
X8_UN_TA
X9_INGR_AF
SECTOR
PAIS
GII
TEA
PIB
DESEMPLEO
CORRUPCION
ENFORCEMENT
LEGAL_SITUATION
INFLATION
Y_ALTMAN

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

```
-----  
X2_TP_TA <= 0.6054: 0 (37.0/2.0)  
X2_TP_TA > 0.6054  
| X1_AC_TP <= 0.5563  
| | GII <= -0.1031  
| | | X2_TP_TA <= 0.6338: 0 (2.0)  
| | | X2_TP_TA > 0.6338: 1 (4.0)  
| | GII > -0.1031: 1 (24.0)  
| X1_AC_TP > 0.5563  
| | X5_EBIT_TCC <= 0.0736: 1 (6.0/1.0)  
| | X5_EBIT_TCC > 0.0736: 0 (8.0)
```

Number of Leaves: 6
Size of the tree: 11
Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	63	77.7778 %
Incorrectly Classified Instances	18	22.2222 %
Kappa statistic	0.5472	
Mean absolute error	0.2488	
Root mean squared error	0.4369	
Relative absolute error	50.5919 %	
Root relative squared error	88.0524 %	
Total Number of Instances	81	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.743	0.196	0.743	0.743	0.743	0.806	1
	0.804	0.257	0.804	0.804	0.804	0.806	0
Weighted Avg.	0.778	0.231	0.778	0.778	0.778	0.806	

=== Confusion Matrix ===

```

a b <-- classified as
26 9 | a = 1
9 37 | b = 0

```

La validación cruzada es un procedimiento que consiste en hacer numerosas particiones de igual tamaño en los datos, dejando unas para estimar el modelo y las restantes para validar. El proceso se repite tantas veces como particiones hayamos hecho, y vamos cambiando las que sirven para estimar de las que sirven para validar. El resultado final es la media de todos los resultados obtenidos. El resultado obtenido de la validación cruzada para 10 particiones (que es la más habitual) de este modelo es de 77.77%.

Observando la diagonal de la matriz de confusión, se tiene unos valores superiores a los elementos a_{21} y a_{12} . Esto es, $a_{11}=26$ es mayor que $a_{21}=9$; y por otro lado, $a_{22}=37$ es mayor que $a_{12}=9$. En concreto, se observa que un 80.4% de las empresas sanas (no quiebra) son clasificadas correctamente y el 74.3% de las empresas en quiebra son clasificadas correctamente.

El resultado del índice Kappa mide si la concordancia establecida se debe exclusivamente al azar. Esto es, un índice Kappa 0 es el que se espera cuando la concordancia de los datos se debe al azar. En el caso contrario, si este índice es mayor que cero, se interpreta cuando los datos no son exclusivamente aleatorios. En el caso del presente estudio se tiene un índice de 0.5472, lo cual nos indica que la clasificación de las empresas en no quiebra y quiebra, no es aleatoria.

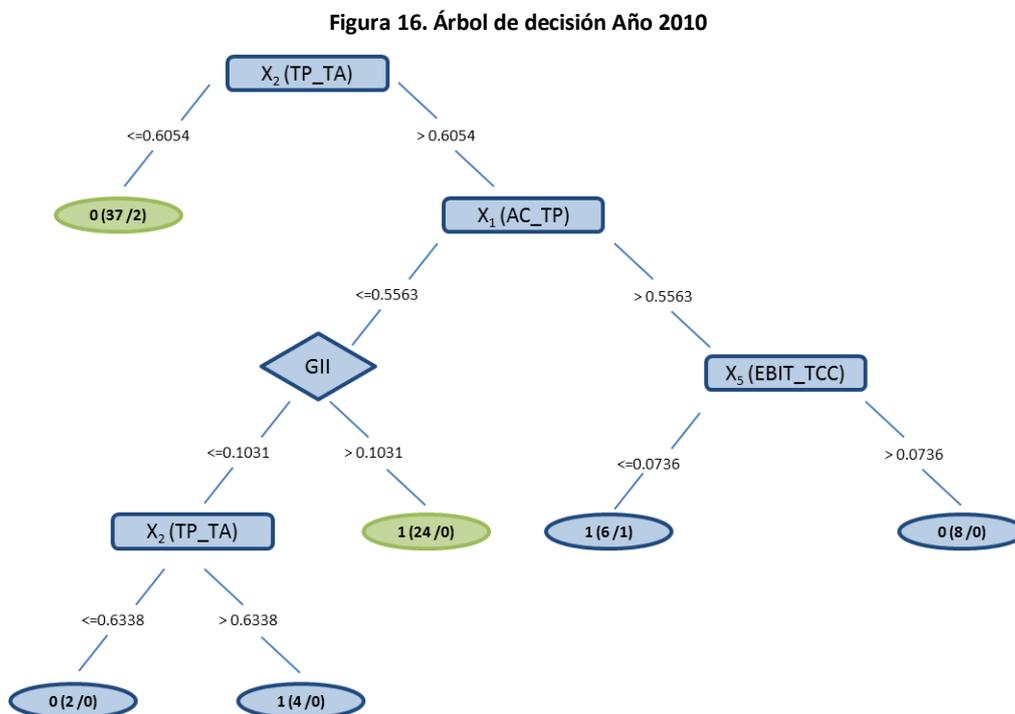
Para analizar el modelo bajo la metodología C4.5 y poder interpretar el árbol, hay que ir descendiendo, hasta completar la totalidad de sus hojas, lo cual nos proporcionara las reglas de decisión.

Al final de cada hoja del árbol aparece un valor n o n/m , siendo su interpretación:

- n : representa el número de pólizas en la muestra que se clasifican de acuerdo a las condiciones que nos llevan hasta esa hoja
- m : representa el número de pólizas mal clasificadas

4.3.1.2 Análisis de las Principales Ramas

Para comprender mejor el árbol de decisión obtenido para el año 2010, a continuación se presenta un esquema gráfico. Dentro del árbol se han identificado las ramas correspondientes a las reglas con mayor fuerza para ambas categorías de clasificación: quiebra y no quiebra (color verde).



Fuente: Elaboración Propia del Autor

Reglas de la Clase 1 = Quiebra

❖ Regla 1 de la clase 1.

- Rama 1. Todas las empresas, según el criterio de clasificación proporcionado por la variable de respuesta (quiebra o no quiebra), pueden ser clasificadas atendiendo en primer lugar a la variable microeconómica conocida como razón de solvencia global X_2 . Como se mencionó en el Capítulo 3, cuanto mayor es el resultado de este cociente más depende la empresa de los recursos ajenos. Para valores superiores a 0.6054 en este ratio, el árbol nos sugiere el análisis de otra variable adicional, es decir la variable microeconómica X_1 .
- Rama 2. Corresponde al análisis de la variable X_1 , el ratio financiero conocido como coeficiente de liquidez circulante que mide la participación del financiamiento total en las inversiones de corto plazo, cuanto mayor sea el ratio, mayor liquidez a corto plazo presenta la empresa. El árbol de decisión sugiere el análisis de la variable Macroeconómica GII para valores inferiores a 0.5563.
- Rama 3. Corresponde al análisis de la variable conocida como GII (Global Insolvency Index) calculada por la empresa Euler-Hermes. Entre mayor sea este indicador, resulta menor el nivel de solvencia de una empresa. Para valores mayores a 0.1031 la empresa se clasificaría como quiebra, situación que se cumple en 24 casos, es decir un total de 30% de la muestra.

Reglas de la Clase 0 = No Quiebra

❖ Regla 1 de la clase 0.

- Rama 1. Siguiendo el mismo criterio, todas las empresas, pueden ser clasificadas atendiendo en primer lugar a la variable X_2 . Analizada de manera inversa, entre menor es el resultado de este cociente menos depende la empresa de los recursos ajenos. Es por ello que el árbol sugiere que para valores inferiores a 0.6054 de esta variable la empresa se clasificaría directamente como sana sin necesidad de alguna otra variable adicional, situación que se cumple en 37 casos, es decir un total de 46% de la muestra.

4.3.1.3 Resumen de las Principales Reglas de Decisión

Así se tiene que las principales Reglas de Decisión obtenidas son:

Figura 17. Principales Reglas de Decisión Año 2010

Año	2010
Categoría	1 - Quiebra
Muestra	81

	Rama 1	Rama 2	Rama 3	Casos	Fuerza
Regla 1	X ₂ - Razón de Solvencia Global	X ₁ - Coeficiente de Liquidez Circulante	GII - Global Insolvency Index	24	30%

Año	2010
Categoría	0 - No Quiebra
Muestra	81

	Rama 1	Casos	Fuerza
Regla 1	X ₂ - Razón de Solvencia Global	37	46%

Fuente: Elaboración Propia del Autor

4.3.2 Corte Transversal Año 2014

4.3.2.1 Validación de Resultados

A continuación presentamos la salida de los resultados obtenidos en el programa WEKA para el año 2014:

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: BANKRUPTCY_MACRO_2014-

Instances: 85

Attributes: 20

X1_AC_TP
 X2_TP_TA
 X3_TP_TCC
 X4_EBIT_INGR
 X5_EBIT_TCC
 X6_UN_INGRN
 X7_UN_TCC
 X8_UN_TA
 X9_INGR_AF
 SECTOR
 PAIS
 GII
 TEA
 PIB
 DESEMPLEO
 CORRUPCION
 ENFORCEMENT
 LEGAL_SITUATION

```

INFLATION
Y_ALTMAN
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===
J48 pruned tree
-----
X8_UN_TA <= 0.0141
| GII <= -0.3: 1 (18.0)
| GII > -0.3
| | X1_AC_TP <= 0.6263: 1 (24.0/5.0)
| | X1_AC_TP > 0.6263: 0 (8.0)
X8_UN_TA > 0.0141
| X4_EBIT_INGR <= 0.5008: 0 (33.0/2.0)
| X4_EBIT_INGR > 0.5008: 1 (2.0)

Number of Leaves:      5
Size of the tree:      9
Time taken to build model: 0 seconds

```

```

=== Stratified cross-validation ===
=== Summary ===

```

Correctly Classified Instances	61	71.7647 %
Incorrectly Classified Instances	24	28.2353 %
Kappa statistic	0.4355	
Mean absolute error	0.3015	
Root mean squared error	0.4778	
Relative absolute error	60.3281 %	
Root relative squared error	95.5572 %	
Total Number of Instances	85	

```

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.732	0.295	0.698	0.732	0.714	0.767	1
	0.705	0.268	0.738	0.705	0.721	0.767	0
Weighted Avg.	0.718	0.281	0.719	0.719	0.718	0.718	0.767

```

=== Confusion Matrix ===

```

```

a b <-- classified as
30 11 | a = 1
13 31 | b = 0

```

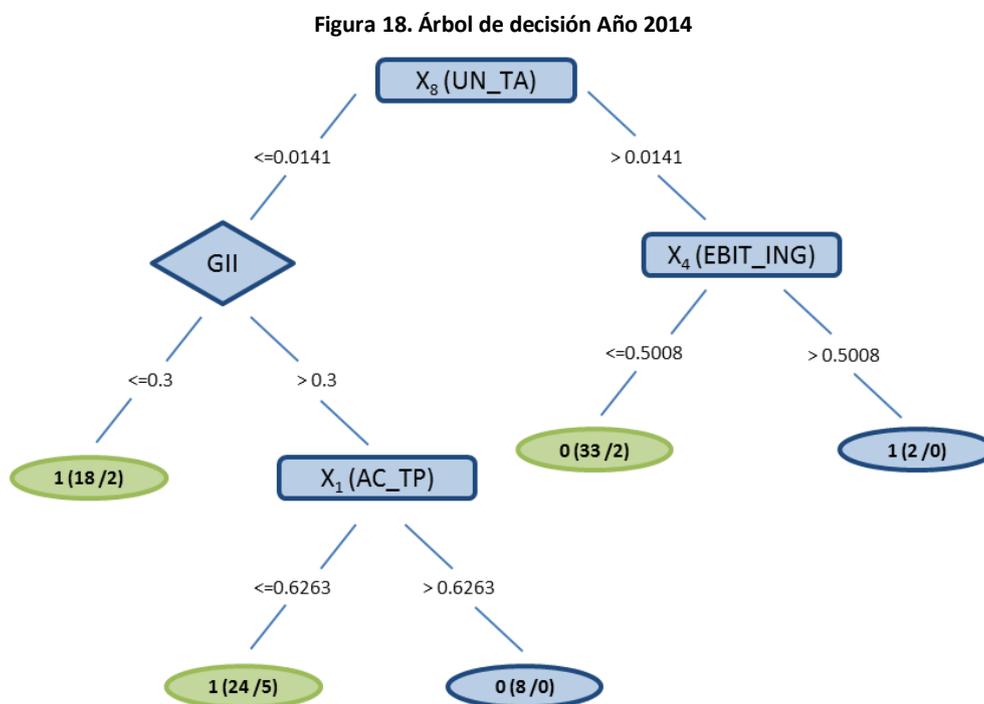
El resultado obtenido de la validación cruzada para 10 particiones de este modelo es de 71.76%.

Observando la diagonal de la matriz de confusión, se tiene unos valores superiores a los elementos a_{21} y a_{12} . Esto es, $a_{11}=30$ es mayor que $a_{21}=13$; y por otro lado, $a_{22}=31$ es mayor que $a_{12}=11$. En concreto, se observa que un 70.5% de las empresas sanas (no quiebra) son clasificadas correctamente y el 73.2% de las empresas en quiebra son clasificadas correctamente.

En el caso del años 2014 el índice Kappa es de 0.4355, lo cual nos indica que la clasificación de las empresas en no quiebra y quiebra, no es aleatoria.

4.3.2.2 Análisis de las Principales Ramas

Para comprender mejor el árbol de decisión obtenido para el año 2014, a continuación se presenta un esquema gráfico. Dentro del árbol se han identificado las ramas correspondientes a las reglas con mayor fuerza para ambas categorías de clasificación: quiebra y no quiebra (color verde).



Fuente: Elaboración Propia del Autor

Reglas de la Clase 1 = Quiebra

❖ Regla 1 de la clase 1.

- Rama 1. Todas las empresas, según el criterio de clasificación proporcionado por la variable de respuesta (quiebra o no quiebra), pueden ser clasificadas atendiendo en primer lugar a la variable microeconómica conocida como razón de solvencia global X_8 . Este indicador es conocido en la literatura como ROA (Return on Assets). Un ratio bajo se asocia a la baja productividad o bien a la ineficiencia en su manejo. Para valores inferiores a 0.0141 en este ratio, el árbol nos sugiere el análisis de otra variable adicional, es decir la variable macroeconómica GII.
- Rama 2. Corresponde al análisis de la variable GII, el Global Insolvency Index de Euler-Hermes. Para valores inferiores a 0.3 la empresa se clasificaría como quiebra, situación que sucede en 18 casos, es decir en un total de 21% de la muestra.

❖ Regla 2 de la clase 1.

- Rama 1. Todas las empresas, según el criterio de clasificación proporcionado por la variable de respuesta (quiebra o no quiebra), pueden ser clasificadas atendiendo en primer lugar a la variable microeconómica conocida como razón de solvencia global X_8 . Este indicador es conocido en la literatura como ROA (Return on Assets). Un ratio bajo se asocia a la baja productividad o bien a la ineficiencia en su manejo. Para valores inferiores a 0.0141 en este ratio, el árbol sugiere el análisis de otra variable adicional, es decir la variable macroeconómica GII.
- Rama 2. Corresponde al análisis de la variable GII, el Global Insolvency Index de Euler-Hermes. Entre mayor sea este indicador, resulta menor el nivel de solvencia de una empresa. Para valores mayores a 0.3 en este indicador macroeconómico, el árbol sugiere el análisis de otra variable adicional, es decir la variable microeconómica X_1 .

- Rama 3. Corresponde al análisis de la variable X_1 , el ratio financiero conocido como coeficiente de liquidez circulante que mide la participación del financiamiento total en las inversiones de corto plazo. Para valores inferiores a 0.6263 en esta variable, la empresa se clasificaría como quiebra, situación que se cumple en 24 casos, es decir un total de 28% de la muestra.

Reglas de la Clase 0 = No Quiebra

❖ Regla 1 de la clase 0.

- Rama 1. Todas las empresas, según el criterio de clasificación proporcionado por la variable de respuesta (quiebra o no quiebra), pueden ser clasificadas atendiendo en primer lugar a la variable microeconómica conocida como razón de solvencia global X_8 . Este indicador es conocido en la literatura como ROA (Return on Assets). Cuanto más alto es el ratio, más rentables son los activos. Un ratio elevado es sinónimo de activos muy eficientes y productivos. Para valores mayores a 0.0141, el árbol sugiere el análisis de otra variable adicional, es decir la variable microeconómica X_4 .
- Rama 2. Corresponde al análisis de la variable microeconómica X_4 . Este indicador es conocido en la literatura como margen de beneficio neto o rentabilidad de los ingresos e indica cuanto beneficio se obtiene por cada peso de ventas, entre menor sea el indicador menos beneficio se obtiene por la ventas realizadas, es por ellos que para valores inferiores a 0.5008, la empresa se clasificaría como quiebra, situación que se cumple en 33 casos, es decir un total de 39% de la muestra.

4.3.2.3 Resumen de las Principales Reglas de Decisión

Así se tiene que las principales Reglas de Decisión obtenidas son:

Figura 19. Principales Reglas de Decisión Año 2014

Año	2014
Categoría	1 - Quiebra
Muestra	85

	Rama 1	Rama 2	Rama 3	Casos	Fuerza
Regla 1	X ₈ - Return on Assets	GII - Global Insolvency Index		18	21%
Regla 2	X ₈ - Return on Assets	GII - Global Insolvency Index	X ₁ - Coeficiente de Liquidez Circulante	24	28%
				42	49%

Año	2014
Categoría	0 - No Quiebra
Muestra	85

	Rama 1	Rama 2	Casos	Fuerza
Regla 1	X ₈ - Return on Assets	X ₄ - Margen de Beneficio	33	39%

Fuente: Elaboración Propia del Autor

4.4 Resultados y Discusión

El incremento de la capacidad o poder predictivo de los modelos de predicción de quiebra empresarial puede ser abordado desde dos enfoques distintos. El primero de estos enfoques relacionado con la adecuada elección de las variables y el segundo relacionado con la adecuada elección de la metodología o técnica de aplicación utilizada.

En términos generales, los trabajos de investigación realizados en este sentido señalan que la precisión de los modelos de quiebra no puede ser apreciablemente mejorada por la elección de uno u otro algoritmo de clasificación. Por ejemplo en el trabajo realizado por Reznakova, Maria y Karas Michael (2014) para el caso de empresas en República Checa, se probaron un método paramétrico vs uno no paramétrico aplicado a la misma muestra inicial y obteniendo como resultados para el caso de la metodología paramétrica (Análisis Discriminante) 8 ratios significativos y 7 para el caso de la no paramétrica (Árboles) obteniendo 3 ratios en común.

Olmeda & Fernández (1997) comparan la precisión de clasificadores paramétricos y no paramétricos sugiriendo que un sistema óptimo para la clasificación del riesgo debe combinar dos o más técnicas distintas.

Los resultados presentados en este trabajo de investigación dejan ver que la aplicación de técnicas paramétricas y no paramétricas no muestra diferencias significativas para las variables que mejor capacidad predictiva presentan, por lo que no se rechaza la hipótesis H_1 planteada en el apartado 3.1.

Como se observa en la Tabla 38, bajos estos dos enfoques, las variables microeconómicas o ratios financieros que presentan mayor poder predictivo X_1 , X_2 , X_8 resultan ser coincidentes. La variable X_2 (ratio de solvencia global) y la variable X_8 (Return on Assets / ROA) han estado presentes en la mayoría de los trabajos de investigación relacionados con la quiebra empresarial (Ver Tabla 6). Es de esperar que estas variables tengan gran relevancia en el estudio ya que por un lado el ratio de solvencia global relaciona la totalidad de activos que tiene una empresa para hacer frente al total de sus obligaciones, misma relación que nos indica que entre más depende la empresa de los recursos ajenos mayor posibilidad existe de presentar un estado de quiebra.

Por otro lado el ROA o retorno sobre los activos que es quizá el ratio de rentabilidad más conocido en la literatura ya que muestra que tan efectivos están siendo los activos de la empresa y, por lo tanto, generar mayor rentabilidad ocasionando a su vez que se tenga la capacidad para solventar el financiamiento que tiene la empresa por parte de recursos ajenos.

En cuanto al ratio X_1 (coeficiente de liquidez circulante) forma parte de los factores de liquidez y su presencia como ratio significativo dentro de los modelos está más que justificada debido a que estos indicadores demuestran de manera global la capacidad que tiene una empresa para pagar sus deudas (a corto y largo plazo), si es necesario mediante la liquidación de los activos de forma rápida o al convertirlos en dinero en efectivo.

Tabla 38. Coincidencia de Variables Micro y Macroeconómicas Resultantes

Técnica / Variable	Microeconómicas (Ratios Financieros)					Macroeconómicas
	X_1	X_2	X_4	X_5	X_8	GII
EEG (Paramétrico) 2010 - 2014	✓	✓		✓	✓	
Árboles (No Paramétrico) 2010	✓	✓			✓	✓
Árboles (No Paramétrico) 2014	✓		✓		✓	✓

Fuente: Elaboración Propia del Autor

Según Banet (2001) estas dos disciplinas (Inteligencia Artificial y Estadística) se han desarrollado una a la espaldas de la otra, mientras que la IA se ha enfocado más en ofrecer soluciones algorítmicas con un coste informático aceptable, la estadística se ha preocupado más por el poder de generalización de los resultados obtenidos, es decir en poder inferir los resultados a situaciones más generales. A diferencia de la IA, que es una ciencia joven, en Estadística se viene aprendiendo de los datos desde hace más de un siglo, la diferencia consiste que ahora existe la potencia de cálculo suficiente. Finalmente y debido a la gran amplia variedad de técnicas que hoy en día se muestran como una opción para la determinación de la posible quiebra o fracaso empresarial (para el caso que nos compete) y para muchos otros temas de investigación.

En el contexto globalizado actual en el que gobierna la incertidumbre y los cambios continuos en las políticas macroeconómicas, la estructura social y en el tejido empresarial, resulta de gran valor contar con modelos de predicción que generen valor para las empresas y más concretamente en el mundo corporativo de empresas cotizadas (que cotizan en alguna bolsa de valores). En años recientes la relación entre el crecimiento macroeconómico y el desarrollo de los mercados ha sido objeto de atención por parte de los tomadores de decisión empresarial.

Si bien es cierto que hasta el día de hoy la mayoría de las investigaciones sobre quiebra empresarial se han centrado en gran medida en distinguir la salud de las empresas con base en sus características financieras medida a través de sus datos contables en forma de ratios, no obstante, debido a la crisis económica actual a nivel mundial y el aumento exponencial de la insolvencia de las empresas, resulta difícil poder estudiar y explicar este fenómeno sin emplear otro tipo de variables que tengan en consideración el estado de la economía en su conjunto y no exclusivamente los ratios contables internos de las empresas.

Los indicadores macro económicos son una serie de datos que se representan en valores estadísticos teniendo como objetivo principal indicarnos la situación que guarda la economía de un país. La tendencia a ajustar los modelos a grupos específicos de empresas con características similares, por situación geográfica, tamaño, fase del ciclo de vida de la empresa, actividad económica etc., es factible cuando se dispone de una gran cantidad de información para tener muestras realmente representativas con que trabajar los modelos.

Por esta razón una alternativa es la que fue utilizada en el presente trabajo de investigación y consiste en realizar estudios con empresas de distintos sectores y localizados en distintas geografías con el fin de aumentar el tamaño de la muestra.

Los resultados obtenidos aplicando ambas técnicas (Ver Tabla 38) muestran que las variables macroeconómicas no tienen la misma presencia que las variables financieras y esto puede deberse en gran medida a que las condiciones macroeconómicas diversas que presentan cada uno de los países puede plantear problemas de comparabilidad de la información y, por lo tanto, recaer mayor peso en las variables internas.

Las variables macroeconómicas se diluyen en un entorno de empresas cotizadas. Tras realizar varias iteraciones en el modelo de EEG el poder predictivo se vio incrementado según la prueba Wald χ^2 (Ver Tablas 32, 34 y 36) al dejar únicamente las variables financieras lo cual nos lleva a no rechazar la hipótesis H_2 planteada en este trabajo de investigación en el apartado 3.1.

Dicho lo anterior se demuestra que en un contexto de empresas cotizadas, las variables macroeconómicas no condicionarían la quiebra empresarial.

CAPITULO 5.CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

Los resultados obtenidos en las aplicación empírica de las dos técnicas, permiten concluir que los ratios más significativos en cuanto a predicción de quiebra o fracaso empresarial son el ratio Coeficiente de Liquidez Circulante X_1 , el ratio Razón de Apalancamiento X_2 y el ratio Margen de Utilidad Neta X_8 , que a su vez están asociados a factores de Liquidez, Endeudamiento y Rentabilidad y receptivamente.

Los trabajos empíricos analizados dejan claro que no existe una teoría económica propiamente dicha sobre la solvencia empresarial y, en consecuencia, no se cuenta con una metodología única ya que los modelos existentes son muchos y muy variados en cuanto a su desarrollo metodológico y en cuanto a la técnica aplicada, es por ello que el problema puede ser abordado siguiendo distintas aproximaciones.

La pregunta que puede ser formulada como consecuencia lógica de la aplicación empírica de estas dos técnicas sería ¿Cuál de estas dos técnicas resulta más apropiada para predecir la quiebra o fracaso empresarial? En este sentido la experiencia muestra que no existe una técnica más apropiada sino formas apropiadas de utilizar una técnica utilizando cada uno de mejor manera los conocimientos de los que dispone.

Desde la década de 1960, se han desarrollado diferentes estudios para predecir crisis empresariales, aplicando diversos modelos sobre la información que proporcionan los ratios financieros, correspondiendo la mayoría a la aplicación de modelos de corte transversal, por lo que la construcción de modelos para datos de panel resulta pertinente en tanto incorporan la dimensión temporal en el estudio.

Los modelos que son utilizados con mayor frecuencia para predecir crisis financieras son el discriminante lineal y el modelo logístico. El segundo permite levantar la normalidad, un supuesto clave que se mantiene es la independencia entre las observaciones, lo que afecta la variabilidad de las estimaciones. No obstante, es evidente que el poder predictivo de un modelo está determinado por las características de la muestra empleada para su elaboración.

El contar con una estructura de datos longitudinales (datos panel), es decir varias mediciones de las mismas unidades para un mismo individuo (empresa) permiten captar su heterogeneidad no observable induciendo correlación entre las respuestas, resultando necesario trabajar con modelos que incorporen esa estructura de correlación, así mismo permite realizar un análisis más dinámico al incorporar la dimensión temporal de los datos, lo que enriquece el estudio, particularmente en períodos de grandes cambios.

Las EEG ofrecen la oportunidad de utilizar diseños longitudinales para variables de respuesta que no necesariamente presentan una distribución normal, situación con la que suele enfrentarse muy comúnmente tanto investigadores, académicos y analistas dentro de las organizaciones.

El presente trabajo de investigación presenta como novedad respecto a otros estudios empíricos la aplicación de un modelo que permite trabajar con datos panel que considera la especificación de la forma de correlación de las variables así como la inclusión de variables macroeconómicas, cuyos resultados han sido contrastados con la metodología de inteligencia artificial conocida como árboles de decisión enfoque que cuenta con demostrada experiencia en la predicción de la insolvencia empresarial.

Con respecto a la variable dependiente, uno de los elementos clave en la aplicación de los modelos predictivos es la determinación del concepto de fracaso o insolvencia. Sin embargo, en ocasiones esta determinación resulta un tanto arbitraria.

El enfoque jurídico es el que presenta menos arbitrariedad por cuanto existe un presupuesto objetivo concreto, que en la mayoría de casos se traduce en el momento en que se ha declarado judicialmente la situación de insolvencia de la empresa. El enfoque económico está sometido a una arbitrariedad mayor por cuanto pueden ser muchos y diferentes los criterios utilizados para realizar la separación entre las empresas sanas y las fracasadas.

En nuestro caso, dadas las dificultades señaladas y al tipo de muestra que disponíamos en la que intervenían empresas de diversos países y por lo tanto los criterios de clasificación jurídica varían dentro de cada país, nos decantamos por la aplicación del enfoque económico, estableciendo como definición del mismo para clasificar a las empresas

fracasadas la puntuación obtenida mediante la aplicación del modelo conocido como Z-score de Altman.

Con relación a las variables del modelo, y pese a las limitaciones que diversos autores han señalado a la utilización de la información financiera, entendemos que ésta continúa constituyendo la fuente más accesible y objetiva para alimentar un modelo predictivo. La utilización de estados financieros anuales auditados y publicados en el software financiero Bloomberg, ofrecerá la fiabilidad necesaria para sustentar la validez del modelo.

Este trabajo presenta algunas limitaciones. Si bien se trabajó con todas las empresas para las cuales se disponía de estados financieros para el periodo 2010 – 2014 incrementando así el tamaño de la muestra, debido a las propias limitaciones de la información, existían partidas contables para las cuales no se disponía de dato y por lo tanto no fue posible estimar el ratio para ese periodo en particular, sin embargo esta limitación en la estimación del modelo es atenuada a través de la consideración de varios periodos para cada empresa.

A pesar de las limitantes expuestas, consideramos que todos los modelos financieros para la predicción de quiebra empresarial deben tomarse únicamente como un parámetro de referencia, es decir sólo como un indicador y como una herramienta más de apoyo para el diagnóstico empresarial, debiéndose complementar con lo diferentes tipos de análisis financieros correspondientes así como con la experiencia observada a través de los años y el entorno socioeconómico en el que se desempeña la empresa en estudio.

Con lo expuesto hasta este momento y después de haber abordado el problema de la quiebra o fracaso empresarial bajo dos ópticas distintas podemos concluir que existen grandes retos a los que se enfrentan las compañías a nivel mundial y que serían objeto de futuras líneas de investigación.

Uno de estos retos es conseguir modelos globales que permitan identificar con claridad que factores internos (intrínsecos a la operación propia de la empresa) y externos (tanto económicos como de índole cultural, político y social) juegan un papel importante en el desarrollo de una teoría consolidada sobre la solvencia empresarial, teniendo como objetivo ser capaces de establecer adecuados mecanismos de administración de riesgos que

les permitan, no solamente detectar cuales son, sino saber cuáles serían las posibles consecuencias en sus finanzas ante posibles cambios a en cualquiera de los factores, es decir realizar análisis de sensibilidad de las variables involucrada previendo así la posibilidad de entrar en procesos concursales que terminen con la posible quiebra de la empresa.

Según Bosón Ponte y Sierra Molina (1996) en el ámbito específico del Análisis Contable, la Inteligencia artificial constituye una de la líneas de actuación futura más prometedoras, con posibilidades de aplicación tanto en el ámbito de la investigación como en el diseño de sistemas de información “inteligentes”, que no solamente proporcionen datos al decisor sino que recomienden el mejor curso de actuación a seguir.

Los esfuerzos realizados en el ámbito de quiebra empresarial se han enfocado en investigar las causas del fracaso empresarial intentando construir modelos para anticipar situaciones de insolvencia orientándose a probar el contenido informativo de los estados financieros dejando de lado el impacto social. Es por ello que consideramos que otro reto es conseguir estudios empíricos que involucren la medición del impacto social de la quiebra de una empresa utilizando indicadores de bienestar social.

Sin lugar a dudas en las últimas décadas el análisis de la solvencia empresarial se ha convertido en una pieza clave y un tema de preocupación a nivel mundial, principalmente debido al aumento en el número de quiebras empresariales (que día con día son más frecuentes sin importar ubicación geográfica o actividad económica desarrollada), al desarrollo de nuevos instrumentos financieros y en general a la globalización que tiene gran importancia para las empresas ya que permite el libre comercio y da apertura a la competencia en mercados internacionales.

BIBLIOGRAFÍA

- Acharya, V. V., Baharat, S. T., & Srinivasan, A. (2007): Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries. *Journal of Financial Economics*, 85(3), 787-821.
- Agarwal, V., & Taffler, R. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance*, 32(8), 1541-1551.
- Albarrán Pérez, P. (2011) Modelos para datos de panel. <http://rua.ua.es/dspace/bitstream/10045/15809/6/Tema6p.pdf>
- Altman, E. I. (1968): Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, XXIII, September 1968
- Altman, E. I., & Narayanan, P. (1997). An international survey of business failure classification models. *Financial Markets, Institutions & Instruments*, 6(2), 1-57.
- Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E.K. and Suvas, A. (2014): "*Distressed Firm and Bankruptcy Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model*" (August 10, 2014). Disponible en SSRN: <http://ssrn.com/abstract=2536340> o <http://dx.doi.org/10.2139/ssrn.2536340>.
- Altman, E. I., & Saunders, A. (1998). Credit risk measurement: Developments over the last 20 years. *Journal of Banking and Finance*, 21.
- Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of banking & finance*, 18(3), 505-529.
- Armour, J., & Cumming, D. (2008). Bankruptcy law and entrepreneurship. *American Law and Economics Review*, 10(2), 303-350.
- Aziz, M. A., & Dar, H. A. (2006). Predicting corporate bankruptcy: where we stand? *Corporate governance*, 6(1), 18-33.

- Back, P. (2005). Explaining financial difficulties based on previous payment behavior, management background variables and financial ratios. *European Accounting Review*, 14(4), 839-868.
- Balcaen, S., & Ooghe, H. (2006). 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, 38(1), 63-93.
- Ball, R. y Foster, G. (1982): "Corporate financial reporting: a methodological review of empirical research", *Journal of Accounting Research*, supplement, pp. 161-234.
- Banet, T. (2001). La minería de datos, entre la estadística y la inteligencia artificial. *Questií: Quaderns d'Estadística, Sistemes, Informatica i Investigació Operativa* Nº 3, Vol.25. Universitat Politecnica de Catalunya.
- Barniv, R., Agarwal, A., & Leach, R. (1997). Predicting the outcome following bankruptcy filing: a three-state classification using neural networks. *Intelligent Systems in Accounting, Finance and Management*, 6(3), 177-194.
- Bauer, J., & Agarwal, V. (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance*, 40, 432-442.
- Beaver, W. H. (1966): *Financial Ratios as Predictors of Failure. Empirical Research in Accounting: Selected Studies, 1966*, University of Chicago, 1967
- Beck, T., Demirgüç-Kunt, A., & Levine, R. (2003). Law and finance: why does legal origin matter?. *Journal of comparative economics*, 31(4), 653-675.
- Bellovary, J. L., Giacominio, D. E., & Akers, M. D. (2007). A review of bankruptcy prediction studies: 1930 to present. *Journal of Financial Education*, 1-42.
- Beraho, E. K., & Elisu, R. (2010). Influence of country culture on bankruptcy and insolvency legal reform management. *International Journal of Management & Information Systems*, 14(2).

- Bernstein, L. A. (1989). *Financial statement analysis theory, application and interpretation*.
- Bhattacharjee, A., Higson, C., Holly, S., & Kattuman, P. (2009). Macroeconomic instability and business exit: Determinants of failures and acquisitions of UK firms. *Economica*, 76(301), 108-131.
- Bonsón Ponte, E., & Sierra Molina, G. (1996). Intelligent Accounting: impact of Artificial Intelligence on accounting research and accounting information. *Proceedings of the ITHURS, León*, 361-368.
- Brealey, R. and Myers, B. (1999): *Principles of corporate finance*. USA, Edit. McGraw-Hill.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. *Classification and Regression Trees*. *Proceedings of the Thirteenth International Conference, Bari, Italy*. 1996.
- Buehler, S., Kaiser, C., & Jaeger, F. (2012). The geographic determinants of bankruptcy: evidence from Switzerland. *Small Business Economics*, 39(1), 231-251.
- Camacho-Miñano, M.M. & Segovia-Vargas, M. (2013). *Las variables económicas que condicionan el éxito del pre concurso. Conservación de empresas en crisis*. Wolters-Kluwer. La Ley.
- Campa, D., & Camacho-Miñano, M. M. (2014). Earnings Management among Bankrupt Non-Listed Firms: Evidence from Spain. *Revista Española de Financiación y Contabilidad / Spanish Journal of Finance and Accounting*, 43(1), 3-20.
- Carling, K., Jacobson, T., Lindé, J., & Roszbach, K. (2007). Corporate credit risk modeling and the macro economy. *Journal of Banking & Finance*, 31(3), 845-868.
- Cestnik, B., Kononenko, I., & Bratko, I. (1987, May). ASSISTANT 86: A Knowledge-Elicitation Tool for Sophisticated Users. In *EWSL* (pp. 31-45).
- Chava, S., & Jarrow, R. A. (2004). Bankruptcy prediction with industry effects. *Review of Finance*, 8(4), 537-569.
- Chava, S., & Purnanandam, A. (2010). Is default risk negatively related to stock returns? *Review of Financial Studies*, 23(6), 2523-2559.

- Cheng, J. H., Yeh, C. H., & Chiu, Y. W. (2007). Improving business failure predication using rough sets with non-financial variables. In Adaptive and Natural Computing Algorithms (pp. 614-621). Springer Berlin Heidelberg.
- Chiu, W. C., Peña, J. I., & Wang, C. W. (2013). Do structural constraints of the industry matter for corporate failure prediction? *Investment Analysts Journal*, (78), 65-81.
- Christidis, A. C. Y., & Gregory, A. (2010). Some new models for financial distress prediction in the UK (Working Paper).
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1687166 (acceso el 30 de agosto de 2015)
- Cinca, C. S., & del Brío, B. M. (1993). Predicción de la quiebra bancaria mediante el empleo de redes neuronales artificiales. *Revista Española de financiación y contabilidad*, 153-176.
- Collins, D.W.; MA YDEW, E.L.; WEISS, I.S. (1997): Changes in the value-relevance of earnings and book values over the past forty years. *Journal of Accounting and Economics*, vol. 24, pp. 39 - 67.
- Contreras, J., Segovia, M., & del Mar Camacho, M. (2014). Origen del endeudamiento en empresas mexicanas y su entrada en concurso mercantil-Propuesta de detección anticipada de la insolvencia. *Crecimiento Empresarial*, RECOFAN, 23.
http://www.ecorfan.org/series/riesgos/Serie_Topicos%20selectos%20de%20Riesgos_3.pdf (Acceso el 30 de mayo de 2015)
- Cuadras, C. M. (1981). *Análisis multivariantes*. Engibar. Barcelona.
- Cumming, C., & Saini, K. (1981). The macroeconomic determinants of corporate bankruptcies in Japan and the United Kingdom. Federal Reserve Bank of New York.
- Das, S. R., Hanouna, P., & Sarin, A. (2009). Accounting-based versus market-based cross-sectional models of CDS spreads. *Journal of Banking & Finance*, 33(4), 719-730.
- Davydenko, S. A., & Franks, J. R. (2008). Do bankruptcy codes matter? A study of defaults in France, Germany, and the UK. *The Journal of Finance*, 63(2), 565-608.

- De Andrés, J., Landajo, M., & Lorca, P. (2012). Bankruptcy prediction models based on multinomial analysis: An alternative to accounting ratios. *Knowledge-Based Systems*, 30, 67-77.
- De Andrés Suárez, J. (2000). Técnicas de Inteligencia Artificial aplicadas al análisis de la solvencia empresarial. Documentos de trabajo (Universidad de Oviedo. Facultad de Ciencias Económicas), (206), 1-31.
- Deakin, E. B. (1972). A discriminant analysis of predictors of business failure. *Journal of accounting Research*, 167-179.
- Díaz Martínez, Z., Fernández Menéndez, J., & Segovia Vargas, M. J. (2004) Sistemas de inducción de reglas y árboles de decisión aplicados a la predicción de insolvencias en empresas aseguradoras. Universidad Complutense de Madrid. XII Jornadas de ASEPUMA.
- Diggle, P., Heagerty, P., Liang, K. Y., & Zeger, S. (2002). *Analysis of longitudinal data*. Oxford University Press. UK.
- Fernández, A. I., (1986). El diagnóstico financiero de la empresa. Nuevas tendencias en el análisis. *Revista española de financiación y contabilidad*, XVI, 49, 113-132.
- Fisher, R. A. (1935). The logic of inductive inference. *Journal of the Royal Statistical Society*, 39-82.
- Fitzmaurice, G. M., & Laird, N. M. (1995). Regression models for a bivariate discrete and continuous outcome with clustering. *Journal of the American statistical Association*, 90(431), 845-852
- Fitzpatrick, P. J. (1932). A comparison of ratios of successful industrial enterprises with those of failed firms. *Certified Public Accountant*, 10(2), 598-605, 656-662 y 727-731.
- Foster, G. (1978). *Financial Statement Analysis*, 2/e. Pearson Education India.
- Fraire, L. A., Vázquez, R. D., & Valdés, A. L. Influencia macroeconómica y contable en los rendimientos accionarios en México. *ESCUELA SUPERIOR DE ECONOMÍA*, 97.

- Francis, J. and Shipper, K. 1999. "Have Financial Statements Lost Their Relevance?". *Journal of Accounting Research*. 37 (2): 319-352.
- García, E., & Sánchez, J. P. (2006). Un estudio meta-analítico de los factores determinantes de la revelación de información. *Spanish Journal of Finance and Accounting/Revista Española de Financiación y Contabilidad*, 35(132), 761-788.
- Ganga, H., Acosta González, E., Fernández Rodríguez F. Influencia de las variables macroeconómicas sobre la supervivencia de las empresas constructoras e inmobiliarias españolas. Facultad de Economía, Empresa y Turismo, Universidad Las Palmas de Gran Canaria.
- Gelashvili, V.; Camacho-Miñano, M.M. & Segovia-Vargas, M.J. (2015) Patrones de supervivencia para la gestión de los centros especiales de empleo. *Revista de Estudios Empresariales*, 1. https://www.google.es/?gws_rd=ssl#q=Patrones+de+supervivencia+para+la+gesti%C3%B3n+de+los+centros+especiales+de+empleo.
- Gervilla García, E., & Palmer Pol, A. (2009). Predicción del consumo de cocaína en adolescentes mediante árboles de decisión. *Revista de Investigación en Educación*, 6, 7-13.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (Vol. 6). Upper Saddle River, NJ: Pearson Prentice Hall.
- Halim, A., Ahmad, H., Daud, S. N. M., & Marzuki, A. (2008). Macroeconomic determinants of corporate failures in Malaysia. *International Journal of Business and Management*, 3(3), P3.
- Hernández Tinoco, M., & Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis*, 30, 394-419.
- Hernández, J. Ramírez, M^a J & Ferri, C. (2004) *Introducción a la Minería de Datos*. Pearson Educación. Editorial Pearson Prentice Hall. España.

- Hillegeist, S. A., Keating, E. K., Cram, D. P., & Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9(1), 5-34.
- Hol, S. (2007). The influence of the business cycle on bankruptcy probability. *International transactions in operational research*, 14(1), 75-90.
- Hotelling, H. (1931). The economics of exhaustible resources. *The journal of political economy*, 137-175.
- Ibarra Mares, A., & Montllor i Serrats, J. (2002). Análisis de las dificultades financieras de las empresas en una economía emergente. Universitat Autònoma de Barcelona,.
- Jones, F. L. (1987). Current techniques in bankruptcy prediction. *Journal of accounting Literature*, 6(1), 131-164.
- Kallunki, J. P., & Pyykkö, E. (2013). Do defaulting CEOs and directors increase the likelihood of financial distress of the firm? *Review of Accounting Studies*, 18(1), 228-260.
- Kendall M. (1987). *Multivariate Analysis*; Edit. Griffin, Londres.
- Korol, T. (2013). Early warning models against bankruptcy risk for Central European and Latin American enterprises. *Economic Modelling*, 31, 22-30.
- Laitinen, E. K. (2002). Financial rating of European technology companies: an international comparison. *Journal of Business Finance & Accounting*, 29(7-8), 871-901.
- Laitinen, E. K., & Laitinen, T. (2009). Audit report in payment default prediction: A contingency approach. *International Journal of Auditing*, 13(3), 259-280.
- Laitinen, E. K., & Suvas, A. (2013). International Applicability of Corporate Failure Risk Models Based on Financial Statement Information: Comparisons across European Countries. *Journal of Finance & Economics*, 1(3), 1-26.
- Larose, D. T. (2005) *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons, Inc. USA.

- Lee, S. H., Yamakawa, Y., Peng, M. W., & Barney, J. B. (2011). How do bankruptcy laws affect entrepreneurship development around the world? *Journal of Business Venturing*, 26(5), 505-520.
- Lee, S. H., Peng, M. W., & Barney, J. B. (2007). Bankruptcy law and entrepreneurship development: A real options perspective. *Academy of Management Review*, 32(1), 257-272.
- Levine, R. (1998). The Legal Environment, Banks, and Long-Run Economic Growth. *Journal of Money, Credit & Banking* (Ohio State University Press), 30(3).
- Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 13-22.
- Liang, T. P., Chandler, J., Han, I., & Roan, J. (1992). An Empirical Investigation of Some Data Effects on.
- Liu, J. (2004). Macroeconomic determinants of corporate failures: evidence from the UK. *Applied Economics*, 36(9), 939-945.
- Lussier, R. N. (1995). A nonfinancial business success versus failure prediction model for young firms. *Journal of Small Business Management*, 33(1), 8-20.
- Mahía R. (2000). Introducción a la especificación y estimación de modelos con datos panel. http://www.uam.es/personal_pdi/economicas/fphernan/EconometriaTVIII.pdf (Acceso 2 de mayo de 2015).
- Marais, M. L., Patell, J. M., & Wolfson, M. A. (1984). The experimental design of classification models: An application of recursive partitioning and bootstrapping to commercial bank loan classifications. *Journal of accounting Research*, 87-114.
- Mare, D. S. (2013). Contribution of macroeconomic factors to the prediction of small bank failures. Working paper. 26th European Conference on Operational Research, Rome, Italy.

- Martín Marín, J.L. (1984): "El diagnóstico de las dificultades empresariales mediante modelos de mercado de capitales", *Boletín de Estudios Económicos*, nº 123, diciembre, pp.621-635.
- Meeks, G., & Meeks, J. G. (2009). Self-Fulfilling Prophecies of Failure: The Endogenous Balance Sheets of Distressed Companies. *Abacus*, 45(1), 22-43.
- Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6), 1695-1725.
- Melitz, M. J., & Ottaviano, G. I. (2008). Market size, trade, and productivity. *The Review of Economic Studies*, 75(1), 295-316.
- Mena, J. (1996). Machine-learning the business: Using data mining for competitive intelligence. *Competitive Intelligence Review* Nº 4, Vol 7.
- Mena, J. (1999) Data mining your website. Digital Press.
- Mora Enguñados, A. (1994). Los modelos de predicción del fracaso empresarial: una aplicación empírica del Logit. *Revista Española de Financiación y Contabilidad*, 203-233.
- Nam, C. W., Kim, T. S., Park, N. J., & Lee, H. K. (2008). Bankruptcy prediction using a discrete-time duration model incorporating temporal and macroeconomic dependencies. *Journal of Forecasting*, 27, 493–506.
- Noga, T. J., & Schnader, A. L. (2013). Book-Tax Differences as an Indicator of Financial Distress. *Accounting Horizons*, 27(3), 469-489.
- O'Leary, D.E. (1995): "Artificial intelligence in business", AI/ES Section of the American Accounting Association, (Internet, <http://www.rutgers.edu/accounting/raw/aaa/aiet>).
- Olmeda, I., & Fernández, E. (1997). Hybrid classifiers for financial multicriteria decision making: The case of bankruptcy prediction. *Computational Economics*, 10(4), 317-335.

- Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, 109-131.
- Pan, W. (2001), 'Akaike's Information Criterion in Generalized Estimating Equations', *Biometrics* 57(1), 120–125.
- Paterson, A., & Niblett, T. B. (1982). *ACLS manual*. Edinburgh: Intelligent Terminals Ltd.
- Pearson, K. (1901). Principal components analysis. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 6(2), 559
- Peng, M. W., Yamakawa, Y., & Lee, S. H. (2010). Bankruptcy Laws and Entrepreneur-Friendliness. *Entrepreneurship Theory and Practice*, 34(3), 517-530.
- Pindado, J., Rodriguez, L., & de la Torre, C. (2008). How do insolvency codes affect a firm's investment? *International Review of Law and Economics*, 28(4), 227-238.
- Polvo Vuelvas, MA. (2007). *Análisis de Modelos de Datos Longitudinales*". Tesis Maestría. Universidad Autónoma Metropolitana.
- Press, S. J., & Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association*, 73(364), 699-705.
- Qu, Y. (2008). Macroeconomic factors and probability of default. *European Journal of Economics, Finance and Administrative Sciences*, 192-215.
- Quinlan, J.R. (1985) *Induction of Decision Trees*. *Revista Machine Learning* Volº 1-1.
- Quinlan, J.R. (1993) *C4.5: Programs for machine learning*. Morgan Kaufmann Publishers, Inc. California.
- Ravi Kumar, P., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review. *European Journal of Operational Research*, 180(1), 1-28.
- Ravid, S. A., & Sundgren, S. (1998). The comparative efficiency of small-firm bankruptcies: A study of the US and Finnish bankruptcy codes. *Financial Management*, 27(4), pp. 28-40.

- Reza, F. M. (1961) An introduction to information theory. Courier Corporation.
- Reznakova, María, Michal Karas (1997). A parametric or nonparametric approach for creating a new bankruptcy prediction model: The evidence from the Czech Republic. International Journal of mathematical models and methods in applied science.
- Rosé, P. S., Andrews, W. T., & Giroux, G. A. (1982). Predicting business failure: A macroeconomic perspective. Journal of Accounting, Auditing and Finance, 6(1), 20-31.
- Rothman KJ, Greenland S, editores. Modern Epidemiology. 2ª ed. Filadelfia: Lippincott-Raven; 1998.
- Salman, A. K., Friedrichs, Y. V., & Shukur, G. (2011). The Determinants of Failure of Small Manufacturing Firms: Assessing the Macroeconomic Factors. International Business Research, 4(3), 22-32.
- Schaefer, V. (1987). Le systemes expert et l'analyse financiere. Revenue Francaise de Comptabilite. , n°177, pp.39-47.
- Segovia Vargas, M.J. (2003) Predicción de Crisis Empresarial en Seguros No Vida mediante la Metodología Rough Set (Tesis Doctoral). Universidad Complutense de Madrid. Madrid. España.
- Segovia-Vargas, M.J (2005). Análisis de la solvencia en entidades aseguradoras mediante la metodología Rough Set, Thomson-Civitas, España
- Simon, H. A. (1960): The new science of management decision, Harper&Row, New York (USA).
- Spearman, C. (1904). "General Intelligence," objectively determined and measured. The American Journal of Psychology, 15(2), 201-292.
- Tascon Fernández, M. T. & Castaño Gutiérrez, F. J. (2012). Variables y modelos para la identificación y predicción del fracaso empresarial: revisión de la investigación empírica reciente. Revista de Contabilidad-Spanish Accounting Review, 15(1), 7-58.

- Weijs, R. J. (2012). Harmonisation of European Insolvency Law and the Need to Tackle Two Common Problems: Common Pool and Anticommons. *International Insolvency Review*, 21(2), 67-83.
- Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Winakor, C.H. & Smith, R.F. (1935). Changes in Financial Structures of Unsuccessful Industrial Companies. Bulletin n° 51. Bureau of Economic Research. University of Illinois Press.
- Yeh, C. C., Lin, F., & Hsu, C. Y. (2012). A hybrid KMV model, random forests and rough set theory approach for credit rating. *Knowledge-Based Systems*, 33, 166-172.
- Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting research*, 59-82.

ÍNDICE DE FIGURAS

Figura 1. Ejemplo de resultados de un árbol de decisión	74
Figura 2. Matriz de Correlaciones (Representación Numérica)	119
Figura 3. Matriz de Correlaciones (Representación Gráfica)	119
Figura 4. Matriz de Variabilidades Máximas Año 2010	121
Figura 5. Matriz de Variabilidades Máximas Año 2011	123
Figura 6. Matriz de Variabilidades Máximas Año 2012	125
Figura 7. Matriz de Variabilidades Máximas Año 2013	127
Figura 8. Matriz de Variabilidades Máximas Año 2014	129
Figura 9. Saturación de los Componentes por la Variables Explicativas por Año de Observación	130
Figura 10. Matriz de Correlación Variables (X2 / X5 / X8).....	132
Figura 11. Resultados Iteración 1 Modelo EEG.....	133
Figura 12. Matriz de Correlación Variables (X2 / X2/ X5 / X8 / Desempleo / Inflación)	133
Figura 13. Resultados Iteración 2 Modelo EEG.....	134
Figura 14. Matriz de Correlación Variables (X2 / X2/ X5 / X8)	134
Figura 15. Resultados Iteración 3 Modelo EEG.....	135
Figura 16. Árbol de decisión Año 2010	144
Figura 17. Principales Reglas de Decisión Año 2010.....	146
Figura 18. Árbol de decisión Año 2014	148
Figura 19. Principales Reglas de Decisión Año 2014.....	151

ÍNDICE DE TABLAS

Tabla 1. Empresas en Crisis Financiera por Pago de Productos Financieros Derivados	10
Tabla 2. Evolución Histórica de los Modelos de Predicción de Quiebra	29
Tabla 3. Relación de Empresas Consideradas en el Estudio	79
Tabla 4. Definiciones de Quiebra / Fracaso Empresarial	84
Tabla 5. Clasificación de empresas (Según Puntuación Z-SCORE ALTMAN)	85
Tabla 6. Variables Microeconómicas Seleccionadas (Ratios Financieros)	88
Tabla 7. Variables Macroeconómicas Seleccionadas.....	89
Tabla 8. Resumen de Estadísticas de la Variable X1	94
Tabla 9. Resumen de Estadísticas de la Variable X2	95
Tabla 10. Resumen de Estadísticas de la Variable X3	96
Tabla 11. Resumen de Estadísticas de la Variable X4	97
Tabla 12. Resumen de Estadísticas de la Variable X5	98
Tabla 13. Resumen de Estadísticas de la Variable X6	99
Tabla 14. Resumen de Estadísticas de la Variable X7	100
Tabla 15. Resumen de Estadísticas de la Variable X8	101
Tabla 16. Resumen de Estadísticas de la Variable X9	103
Tabla 17. Código de la Variable País	106
Tabla 18. Resumen de Estadísticas de la Variable GII.....	108
Tabla 19. Resumen de Estadísticas de la Variable TEA	109
Tabla 20. Resumen de Estadísticas de la Variable PIB	111
Tabla 21. Resumen de Estadísticas de la Variable Desempleo.....	113
Tabla 22. Resumen de Estadísticas de la Variable Corrupción	114
Tabla 23. Resumen de Estadísticas de la Variable Corrupción.....	115
Tabla 24. Resumen de Estadísticas de la Variable Inflación.....	118
Tabla 25. Valores Propios de la Matriz de Covarianza Año 2010	120
Tabla 26. Valores Propios de la Matriz de Covarianza Año 2011	122
Tabla 27. Valores Propios de la Matriz de Covarianza Año 2012	124
Tabla 28. Valores Propios de la Matriz de Covarianza Año 2013	126
Tabla 29. Valores Propios de la Matriz de Covarianza Año 2014	128
Tabla 30. Frecuencia de Aparición de las Variables Explicativas en la Saturación de los Componentes.....	131
Tabla 31. Variables Significativas Iteración 1	135
Tabla 32. Significancia Estadística Iteración 1	136
Tabla 33. Variables Significativas Iteración 2	136
Tabla 34. Significancia Estadística Iteración 2	137
Tabla 35. Variables Significativas Iteración 3	138
Tabla 36. Significancia Estadística Iteración 3	139
Tabla 37. Distribución de la Muestra en Función de las Variable Respuesta	141
Tabla 38. Coincidencia de Variables Micro y Macroeconómicas Resultantes	153

ÍNDICE DE GRAFICOS

Gráfico 1. Evolución del número de empresas en concurso mercantil en México	10
Gráfico 2. Índice Euler - Hermes en España, México y Zona Euro	11
Gráfico 3. Tasa de Variación Interanual del PIB en España, México y Zona Euro	12
Gráfico 4. Tasa de Variación Interanual de Desempleo en España, México y Zona Euro	12
Gráfico 5. Tasa de Variación Interanual de la Deuda Pública en España, México y Zona Euro	13
Gráfico 6. Evolución del número de Empresas Concursada del Sector Construcción en España (2005-2014)	50
Gráfico 7. Histograma de Frecuencias de las Variable X1	93
Gráfico 8. Histograma de Frecuencias de las Variable X2	95
Gráfico 9. Histograma de Frecuencias de las Variable X3	96
Gráfico 10. Histograma de Frecuencias de las Variable X4	97
Gráfico 11. Histograma de Frecuencias de las Variable X5	98
Gráfico 12. Histograma de Frecuencias de las Variable X6	99
Gráfico 13. Histograma de Frecuencias de las Variable X7	100
Gráfico 14. Histograma de Frecuencias de las Variable X8	101
Gráfico 15. Histograma de Frecuencias de las Variable X9	102
Gráfico 16. Distribución de la Variable Sector	106
Gráfico 17. Distribución de la Variable País.....	107
Gráfico 18. Histograma de Frecuencias de las Variable GII	108
Gráfico 19. Histograma de Frecuencias de las Variable TEA	109
Gráfico 20. Histograma de Frecuencias de las Variable PIB.....	111
Gráfico 21. Histograma de Frecuencias de las Variable Desempleo	112
Gráfico 22. Histograma de Frecuencias de las Variable Corrupción	113
Gráfico 23. Histograma de Frecuencias de las Variable Enforcement.....	114
Gráfico 24. Histograma de Frecuencias de las Variable Inflación	117
Gráfico 25. Sedimentación Año 2010	121
Gráfico 26. Sedimentación Año 2011	123
Gráfico 27. Sedimentación Año 2012	125
Gráfico 28. Sedimentación Año 2013	127
Gráfico 29. Sedimentación Año 2014	129