

CONAC

ACTUARIOS TRABAJANDO

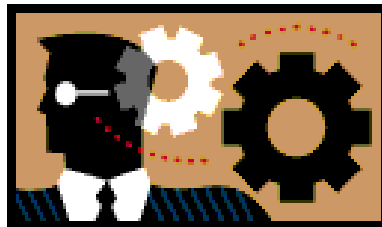
**REVISTA MEXICANA DE
INVESTIGACION ACTUARIAL
APLICADA**

$$q_x$$

$$\mu_x$$

$$d_x$$

$${}_tV_x$$



$$A_x$$

$$l_x$$

$$\ddot{a}_x$$

$$p_x$$

AÑO 2 NUM. 4

ENERO 2010

ACTUARIOS TRABAJANDO

REVISTA MEXICANA DE INVESTIGACION ACTUARIAL APLICADA

COORDINADOR Y EDITOR:

M.E.I.O. Gabriel Núñez Antonio

gab.nuneza@gmail.com

REVISORES ASOCIADOS:

Enrique de Alba

Ma. de los Angeles Yáñez

Juan José Fernández

Luis Enrique Nieto Barajas

Jesús Alfonso Zúñiga San Martín

Jorge Rendón Elizondo

Leovigildo Leandro López García

Diego Hernández

Ricardo Nava

Sofía Romano

José Luis Salas

Oliva Sánchez

Rodica Simón

Carlos Soto

José Luis Suárez

Crisóforo Suárez Tinoco

Gabriel Núñez Antonio

CONTENIDO

<u>Carta del Editor</u>	4
<u>Comentarios al cuestionario ampliado (preliminar) del Censo de Población y Vivienda del año 2010*</u> <i>Editado por Aremis Villalobos Hernández</i>	5
<u>Modelo estocástico de ciclones tropicales para México</u> <i>Luisarturo Castellanos Hernández</i>	13
<u>Algoritmo basado en principios de asignación eficiente de recursos para unidades en un mapa de escala multidimensional.</u> <i>Héctor J Alvarado Valdivia</i>	37
<u>Evolución de la Eficiencia en el Sector Asegurador Mexicano</u> <i>Villarreal Azúa Ricardo Erick</i>	62
<u>Segmentación Binaria Circular: Una técnica para detectar regiones dañadas del ADN</u> Sofía López Lizarraga y Luis Enrique Nieto Barajas	94
<u>Análisis de la Mortalidad Diferencial por Regiones Socioeconómicas del INEGI</u> Alejandra Cevallos Nava	111
<u>Caracterización de los depósitos a la vista de la banca comercial Mexicana y administración de su riesgo de liquidez a partir de un Modelo de Tasas de Interés de Mercado</u> <i>Leticia Sigrid González Méndez y Mercedes Gregorio D.</i>	145

Estimados Colegas:

Es un gusto presentarles el cuarto número de “Actuarios Trabajando”. El objetivo de contar con esta revista de tipo electrónico es que en el futuro cercano se convierta en uno de los medios relevantes de difusión del trabajo del actuario en México, incluyendo en ella temas de interés para todas las áreas del ejercicio profesional.

El hacer del actuario es diverso. Por un lado, su práctica abarca desde el estudiante con aplicaciones metodológicas relevantes hasta el profesionalista con propuestas innovadoras. Por otro lado, las áreas de desarrollo del actuario en México son muy amplias. En este número he decidido dar espacio a esta enorme diversidad con el fin de promover, ampliar y difundir el interés en las diversas áreas de desarrollo del actuario en México y en el mundo.

Por lo anterior, en esta edición contamos con artículos que van desde el discurso del profesor *Manuel Ordorica Mellado* pronunciado en la Tercera Reunión sobre el XIII Censo General de Población y Vivienda 2010 hasta un trabajo referente a técnicas para detectar regiones dañadas del ADN.

Continuamos exhortándolos a que nos envíen sus colaboraciones para futuros números, así como sus opiniones a los artículos ya publicados, además de sus sugerencias para mejorar la revista.

Quiero agradecer la preferencia por *Actuarios Trabajando* a todos los autores que enviaron su trabajo, especialmente a la Dra. Angeles Yañez su entusiasmo y compromiso sin los cuales este proyecto no sería posible.

Espero que disfruten la lectura y les envío un cordial saludo.

M.E.I.O. Gabriel Núñez Antonio

Editor.

Comentarios al cuestionario ampliado (preliminar) del Censo de Población y Vivienda del año 2010*

Palabras pronunciadas por *Manuel Ordorica Mellado* en el Tercera Reunión sobre el XIII Censo General de Población y Vivienda 2010, México, D.F., 15 de diciembre de 2009.

Editado por:

Aremis Villalobos Hernández.

Nuestro país es rico en información estadística. Tenemos censos modernos desde 1895, registros administrativos y encuestas que permiten profundizar en el conocimiento de nuestra realidad. Sin embargo también hay que reconocer que nuestras estadísticas vitales tienen deficiencias que impiden conocer, por ejemplo, el número real de nacimientos. Esto significa no conocer fielmente la dinámica demográfica nacional.

El censo que se va a levantar en el 2010 no es cualquier censo, con él iniciaremos la historia censal del nuevo milenio. Será el primero del siglo XXI y con él conmemoraremos el centenario de nuestra Revolución y el bicentenario de nuestra Independencia. Deberemos tener la oportunidad de comparar los cambios sociodemográficos ocurridos en los últimos cien y doscientos años. Deberá ser en pocas palabras, el mejor censo de población de la historia.

* Palabras pronunciadas en el Tercera Reunión sobre el XIII Censo General de Población y Vivienda 2010, México, D.F., 15 de diciembre de 2009.

Su éxito depende del INEGI, de los demógrafos, de los legisladores, de todos, incluyendo a la población en general, que tendrá que apoyar su levantamiento y dar información veraz. El censo del 2000 fue de elevada calidad, levantado por quienes hoy también lo van hacer. Miguel y Elsa, como lo mencioné en el taller anterior, tienen una gran experiencia en la generación de datos censales y de encuestas, y un amplio reconocimiento nacional e internacional en este campo. Su capacidad profesional está fuera de toda duda.

Este censo al igual que el de 1990 ha sido de gran valor para conocer la realidad social y para la planeación de las políticas gubernamentales. Recordemos la desafortunada historia del censo de 1980. Un censo al que se le realizaron varios cambios y que debido a la falta de experiencia de los encargados de su realización llevaron a la mayor catástrofe censal de la historia de México. Además tuvo un alto costo económico. Nunca pudo ser analizado. En su lugar fueron utilizadas las estimaciones y proyecciones de la población elaboradas por el Consejo Nacional de Población. Que eran exactamente eso, estimaciones.

Aunque los censos puedan tener errores, los censos son la verdad. Un ejemplo de lo anterior, es que las cifras que del censo se derivan son la base para la construcción de la distritación electoral. Se elabora a partir del último censo. En la más reciente construcción de los 300 distritos electorales se empató la sección electoral con el área geoestadística básica (AGEB). Las cifras censales son como dije antes, la verdad. Es la fuente de información más importante en la planeación.

Si bien es cierto que el INEGI es la institución encargada de su realización, la responsabilidad es de todos.

Deseo señalar que si bien realizaré comentarios al cuestionario ampliado, considero que ambos instrumentos, el básico y el ampliado, deben estar íntimamente vinculados desde un punto de vista metodológico y conceptual. No es posible aislarlos, ya que conforman un todo integrado, ubicado también en la necesidad de avanzar en el conocimiento de los principales temas sociodemográficos nacionales, actuales y por venir.

Desde mi punto de vista, la natalidad, la fecundidad adolescente, la migración interna e internacional, la concentración y dispersión de la población, la familia, la población indígena y el empleo son algunos de los temas que marcarán las agendas de investigación demográfica en los próximos años. Existen otros como el envejecimiento demográfico, la pobreza, la marginación, la educación y el medio ambiente que se encontrarán en el análisis de los demógrafos, pero salen fuera de mis comentarios.

Los nacimientos en el país en el 2008 según las Estadísticas Vitales se calculan en dos millones seiscientos mil, mientras que las cifras oficiales del CONAPO son de un millón novecientos mil. Es una diferencia de 700 mil nacimientos. No sabemos si son registros dobles o son nacimientos ocurridos en otros países pero registrados en México. La cuestión es que no sabemos cómo explicar estas diferencias. Esto se viene observando desde hace varios lustros. El CONAPO

conjuntamente con el Colegio de México y el INEGI hemos realizado la mejor conciliación posible. Sin embargo, sólo son estimaciones. Realmente no tenemos la certeza de estos datos. Esto significa que no sabemos con claridad cuál es la dinámica demográfica del país. Mucho menos sabemos qué pasa en los municipios y en las localidades. Si bien es cierto que la política de población y en general las políticas públicas se han planteado sólo en los niveles nacional y estatal, es necesario bajar a niveles menores como el municipio, por ejemplo. Los Consejos de Población y las áreas de planeación del desarrollo de los estados y municipios requieren de información que les permitan diseñar sus políticas a esos niveles.

En el país se están presentando cambios en la fecundidad de las mujeres jóvenes. Parecería que existe un repunte de la fecundidad en la adolescencia. Sabemos por otra parte que muchos de los embarazos que se presentan entre las mujeres menores de 20 años, terminan en ocasiones en un aborto. La fecundidad acumulada de las mujeres de este grupo puede darnos pistas de cómo evitar este hecho que nadie quiere. ¿En qué localidades están ocurriendo estos cambios? Las Estadísticas Vitales no nos permiten aproximarnos al conocimiento de estas transformaciones, debido a sus deficiencias. No es posible calcular tasas confiables a niveles de mayor desagregación geográfica.

Un hecho que se encuentra en discusión es el de la migración internacional, que desde mi punto de vista está resuelto en el cuestionario ampliado. Recordemos que el dato de la población total estuvo sujeto a un examen riguroso en el Conteo

del 2005. Se presentó un número elevado de viviendas pendientes de censar en 2005. Una imputación de cuatro, de tres, o de dos personas por vivienda pendiente de contar, tiene repercusiones, tanto en el número de habitantes, como en el número de migrantes internacionales. La cifra de aproximadamente 300 mil migrantes anuales a los Estados Unidos pasó a cerca de 600 mil. Este es un cambio que no sabemos si es real o se debe únicamente a la cifra usada en la imputación. El punto es que no conocemos con exactitud cuál es el volumen real de la migración a los Estados Unidos.

En nuestro país se ha cumplido, más o menos, la famosa meta del 1% al año 2000 en el crecimiento demográfico planteada en 1977. Sin embargo, prácticamente nada hemos avanzado en el desarrollo de una política de migración interna. Hace algunos años se realizó un intento que derivó en las tres erres: retención, reorientación y reubicación. Los planteamientos que se hicieron se basaron en la pregunta sobre migración incluida en el censo de 1970, que permitía conocer los movimientos ocurridos en los últimos años antes del censo. Las tabulaciones sobre el tiempo de residencia fueron clave para conocer los flujos migratorios. Sin esta pregunta es imposible hacer un planteamiento de la migración y su política. La pregunta sobre lugar de nacimiento nos permite acercarnos a la medición de la migración, pero es insuficiente para lograr tener una visión completa de este componente demográfico. A la fecha no se ha bajado a otros niveles geográficos, pero es una etapa de la Política de Población que se tendrá que cubrir en el futuro.

Un especial interés del gobierno federal es diagnosticar el tema del empleo y analizar los efectos de la crisis sobre esta variable. Es importante responder a la pregunta de cómo está resolviendo la población sus estrategias de sobrevivencia, en qué actividades económicas y en qué ramas de actividad se presentan los principales problemas. La información sobre la actividad económica es uno de los principales elementos para medir la calidad de vida de los habitantes. Esto es de especial interés a nivel de AGEB.

En nuestro país continúa el binomio dispersión- concentración. Este tema sigue preocupando a los especialistas y funcionarios. Es imposible realizar acciones de política en los cientos de miles de pequeñas localidades dispersas que existen en el país. Es en estas localidades donde viven los más pobres entre los pobres, muchos de ellos son indígenas. Es necesario diseñar políticas que atiendan a las zonas dispersas. Además de disponer de instrumentos de análisis como el índice de marginación para ubicar las zonas más pobres, es necesario disponer de otros indicadores que posibiliten las interrelaciones entre las diferentes áreas geográficas, a fin de organizar la dispersión poblacional. Esto se puede realizar mediante la pregunta que mide la movilidad entre áreas geográficas.

El cuestionario ampliado se compone de 19 preguntas sobre vivienda, 13 sobre migración internacional, 12 vinculadas al tema del empleo, 6 sobre fecundidad y mortalidad, 4 sobre alfabetismo y educación, entre otras. Esto muestra las prioridades de información que se tiene en el país. Hay que reconocer que es un cuestionario cargado de preguntas, lo que se explica por el interés de las

instituciones de disponer de datos que no producen y si lo hacen son de deficiente calidad. Los países desarrollados realizan censos con pocas preguntas porque disponen de registros administrativos confiables y completos sobre diversos temas. Lo que no ocurre en México.

Me preocupa que no tengamos información completa y confiable sobre datos básicos, como los nacimientos y la migración internacional. Si no sabemos cuál es el número de nacimientos, no sabemos cuál es el crecimiento de la población. Si no conocemos el crecimiento demográfico actual, no podemos establecer hipótesis adecuadas sobre la evolución futura de este componente. Las proyecciones tendrán un problema de origen. Las cifras de población proyectadas luego se usan para realizar las expansiones de datos para medir otras variables, como la pobreza, el desempleo, etcétera.

El cuestionario básico por su parte contiene 23 preguntas: 9 sobre vivienda, 3 sobre indígenas, 3 sobre alfabetismo y educación, 1 sobre discapacidad, etcétera. Entiendo que la fecundidad hoy en día no sea considerada ya como un tema de análisis, porque nos encontramos en el nivel de reemplazo poblacional. Sin embargo, es necesario conocer este dato de la ecuación compensadora para tener una idea clara de los elementos que componen la dinámica demográfica.

En este censo de vivienda y población entiendo que se le dé gran importancia a la vivienda para construir el índice de marginación. Este índice tiene deficiencias. No se puede comparar en el tiempo y no es de fácil interpretación. Además se ha

visto que es más que suficiente con tener unas cuantas variables debido a que muchas de las que integran el índice están altamente correlacionadas. El índice de desarrollo humano es más adecuado, porque es comparable en el tiempo y es de más fácil interpretación.

Me parece importante este taller para realizar una reflexión conjunta del censo del 2010. El INEGI, el CONAPO, la Academia, las otras instituciones de los gobiernos federal y estatal, estamos en el mismo barco: INEGI generando información, el CONAPO haciendo las políticas y la Academia analizando y evaluando datos sociodemográficos y proponiendo información faltante. Es un triángulo indisoluble. Recordemos que una de las pocas políticas de estado exitosas surgió de esta institución: la Política de Población. Aquí se realizaron las primeras proyecciones demográficas y los primeros planteamientos de política con Urquidi y Cabrera.

Todo lo que hacemos tiene como fin buscar mejorar las condiciones de vida de las personas. La población pobre, marginada, desempleada debe de marcar la orientación del censo. La crítica constructiva es necesaria hoy más que nunca en México. El INEGI, tiene una excelente oportunidad para recuperar la experiencia del pasado, reconociendo errores y aciertos para con ello realizar el levantamiento del mejor y más completo censo de población y vivienda de la historia de nuestro país.

Manuel Ordorica Mellado
Secretario General de El Colegio de México
Doctor en Ingeniería (Investigación de Operaciones), UNAM.
Investigador Nacional, nivel II.
mordori@colmex.mx

Modelo estocástico de ciclones tropicales para México

M. en C. Luisarturo Castellanos Hernández
Director Actuarial
CatRiskMéxico, S.C.

Paseo del Marqués de la Villa del Villar 2051,
Col. Clautros de la Corregidora II, Querétaro, Qro., C.P. 76090

lcastellanos@catriskmexico.com

luisart47@hotmail.com

Teléfono: (442)229-1091

Resumen

México, por su posición geográfica, está expuesto ante la amenaza que representan los ciclones tropicales. Éstos son fenómenos naturales que si bien ocurren todos los años, la frecuencia con la que azotan la República Mexicana así como la severidad de los daños que provocan son variables dependiendo de su intensidad y de su trayectoria. Es por esto que se plantea un modelo estocástico que simule un gran número de temporadas distintas de ciclones. Este modelo puede ser de gran utilidad para el sector asegurador.

Palabras clave: huracán, modelación, estocástica, seguro, ciclón tropical.

I. Introducción

México se encuentra situado geográficamente en una zona en la que está altamente expuesto a los ciclones tropicales. Tiene amplios litorales tanto en el Océano Pacífico como en el Atlántico, y al ubicarse en el Trópico de Cáncer es altamente vulnerable ante los efectos de ciclones tropicales que se generan en ambos Océanos.

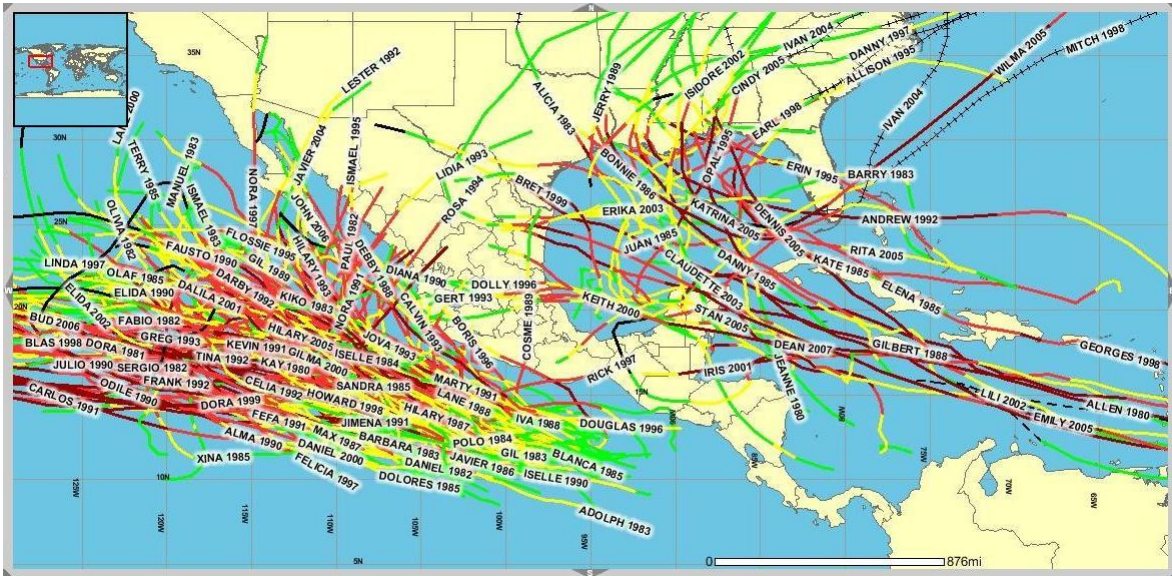


Figura 1. Trayectorias de huracanes que han pasado a 60 millas de la Costa Mexicana. Fuente: NOAA (1980-2007)

El mapa de la figura 1 muestra las trayectorias de aquellos ciclones tropicales que por su intensidad fueron catalogados como “huracanes” por el *National Hurricane Center* (NHC) de la *National Oceanic and Atmospheric Administration* (NOAA) de Estados Unidos, y que, además, pasaron a 60 millas o menos de la costa mexicana en el período que comprende los años 1980-2007 inclusive. Es evidente que México está altamente expuesto a este fenómeno natural, y que si bien, ocurre año tras año, presenta variabilidad en cuanto a la frecuencia y severidad con las que azota a la República.

Durante la temporada 2005 se presentó una actividad anormal en la cuenca del Atlántico donde huracanes como el “Wilma” y el “Stan” provocaron graves daños en México así como el “Katrina” y el “Rita” en los Estados Unidos. Sin embargo, en años más recientes la actividad ciclónica ha regresado a la normalidad. Ante tal variabilidad tanto en la frecuencia como en la severidad se desprende la necesidad de modelar el fenómeno relacionado con los ciclones tropicales para cuantificar, de manera probabilística, el impacto que puede tener en diversas carteras de seguros.

El NHC, considerada la fuente oficial respecto a los huracanes, cuenta con una base de datos de más de 100 años de historia, sin embargo, la calidad de los datos no es la misma a lo largo del tiempo debido a la tecnología satelital con la que hoy se cuenta y con la que se pueden obtener mediciones de mejor calidad en zonas del océano donde antes era imposible acceder. Por esta razón se vuelve relevante simular un número mayor de temporadas sintéticas para tener mayor precisión en la estimación de los daños potenciales.

Ciclones tropicales

Para poder modelar el fenómeno de los ciclones tropicales es necesario entender qué son y cómo se comportan para poder tratar de replicar, en la medida de lo posible, el comportamiento natural del fenómeno. El Servicio Meteorológico Nacional define a un ciclón tropical como un remolino gigantesco que cubre cientos de kilómetros cuadrados y tiene lugar, primordialmente, sobre los espacios oceánicos tropicales. Cuando las condiciones oceánicas y atmosféricas propician que se genere un ciclón tropical, la evolución y desarrollo de éste puede llegar a convertirlo en huracán. Según su etapa de evolución se clasifican en:

- Perturbación tropical
- Depresión tropical
- Tormenta tropical
- Huracán

El vocablo huracán deriva del maya “*Jun Raqan*” que significa “el de una sola pierna”. Era el dios maya del viento, las tormentas y el fuego, y fue una de las deidades que participaron en la creación de la humanidad a partir del maíz. También fue el causante de “El Gran Diluvio” enviado para destruir a los primeros hombres que habían enfurecido a los dioses. Supuestamente vivió en las nubes sobre la Tierra inundada mientras la rehacía empujando el agua hasta las costas originales. Es representado como un ser de una sola pierna, con cola de serpiente y de aspecto reptiloide, porta un objeto humeante (posiblemente una antorcha) y una gran corona. En otras zonas del mundo reciben otros nombres, como “tifón” en el sureste asiático, pero el fenómeno meteorológico es el mismo.

El proceso físico que genera un ciclón tropical consiste en los siguientes pasos:

1. La superficie caliente del mar favorece la evaporación del agua oceánica. Este vapor de agua tiende a ascender y se enfría. El enfriamiento provoca que el agua se condense y forme nubes.
2. Durante el proceso de condensación se libera calor. Este calor aumenta la temperatura del aire y la hace más ligera, lo cual implica que ascenderá.
3. Mientras asciende, el aire superficial se mueve para tomar su lugar, que es el viento fuerte asociado a los huracanes.
4. En principio este movimiento debería ser en línea recta, pero como la Tierra está en rotación, el movimiento tiene forma circular (fuerza de Coriolis).

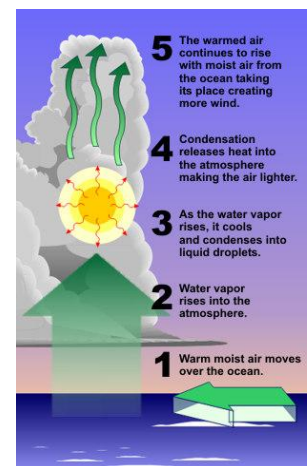


Figura 2.
Proceso físico

La estructura de un ciclón tropical consiste en un ojo que es la zona de relativa calma en el centro del ciclón, y suele tener un diámetro de entre 20 y 50 kilómetros. El radio de las nubes, en cambio, puede ser de varios cientos de kilómetros. La presión más baja se registra en el ojo del ciclón. Mientras más baja es la presión, más intenso será el movimiento de aire hacia el ojo así como el movimiento ascendente. La pared del ojo es la zona en la que se registran los vientos más intensos. Con base en esta medición se asigna la magnitud.

Los huracanes se forman a partir de una pequeña inestabilidad que genera una circulación con sentido de rotación contrario al de las manecillas del reloj, este vórtice puede ir creciendo. Las siguientes condiciones favorecen su formación e intensificación: a) Temperaturas en la superficie del mar relativamente altas: las altas temperaturas proveen la energía en forma de calor latente (evaporación) que se transforma en la energía cinética (movimiento) que provoca la circulación ciclónica, b) Latitudes bajas (dentro de los trópicos), y c) débil gradiente de vientos.

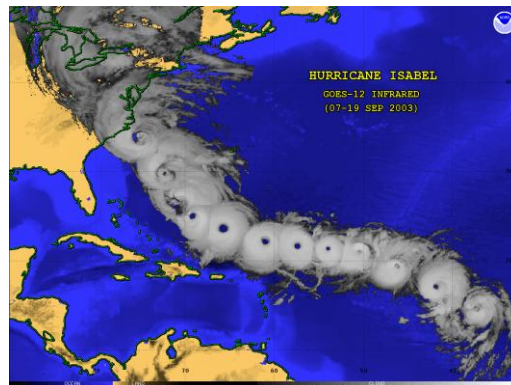


Figura 3. Ciclogénesis. Fuente: NOAA

II. Modelación

El modelo se centra exclusivamente en el análisis de tormentas tropicales y huracanes (*i.e.* no contempla depresiones tropicales) por que se basa en la estadística provista por el NHC para ambos dominios (Océanos Atlántico y Pacífico) que contiene las siguientes variables: fecha de ocurrencia del fenómeno e información subsecuente en lapsos de 6 horas, coordenadas, velocidad del viento y presión atmosférica.

Dada esta información se propone modelar estocásticamente los siguientes aspectos:

- Número de tormentas por temporada.
- Duración de cada tormenta (en lapsos de 6 horas).
- Posición inicial de cada tormenta (lugar de gestación).
- Trayectoria de cada tormenta.

- Velocidad máxima del viento de cada lapso de cada tormenta.
- Presión atmosférica mínima de cada lapso de cada tormenta.

Un modelo es una representación parsimoniosa de la realidad. La intención de este modelo es replicar de forma estocástica el comportamiento de una temporada hipotética de tormentas tropicales, tanto para el Atlántico como para el Pacífico independientemente, para posteriormente simular un número mucho mayor (10,000) de diferentes temporadas para evaluar el impacto que puede tener en el futuro este fenómeno aplicado a cualquier cartera de seguros. De ninguna forma se pretende usar este modelo con fines de pronóstico meteorológico.

Tormentas por Temporada

Mucho se ha dicho con respecto al cambio climático y su efecto en los ciclones tropicales, por ejemplo una de las hipótesis es que cada año habrá más huracanes y serán de mayor intensidad. La siguiente gráfica (Figura 5) muestra el número de tormentas tropicales por temporada (ciclones que por lo menos llegaron a convertirse en tormentas) que se han observado en ambas cuencas en los últimos 60 años (1949-2008).

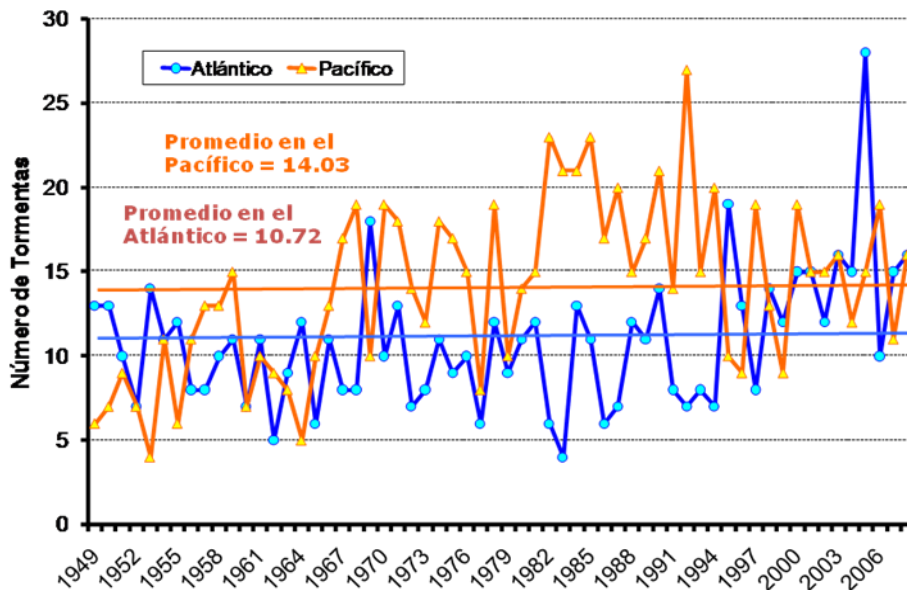


Figura 4. Tormentas por temporada en los últimos 60 años.

A simple vista se puede concluir que no hay evidencia de que cada año se incrementa de manera sustancial el número de tormentas por temporada. No parece existir una tendencia creciente ni para el Océano Pacífico ni para el Atlántico. Aparentemente en la década de los ochentas se incrementó el número de tormentas en el Pacífico comparado con el Atlántico, pero en años más recientes (noventas en adelante) pareciera que vuelven a ubicarse aproximadamente en el mismo nivel. Salta la atención el año 2005, donde evidentemente el número de tormentas (28) en el Atlántico sobrepasa cualquier norma. Fue el año en el que se presentaron huracanes tan poderosos como el “Katrina”, el “Wilma” y el “Stan” en la misma temporada todos ellos en la cuenca

del Atlántico. Sin embargo, para las siguientes 3 temporadas (2006-2008) se observa como el número de tormentas ha regresado al nivel normal.

El número de tormentas por temporada evidentemente es una variable discreta y entera y el fenómeno puede ser visto como el número de llegadas a un sistema en un lapso de tiempo determinado. El lapso de tiempo es la temporada que si bien se refiere a todo un año calendario, en estricto sentido las temporadas se limitan al verano (Mayo-Noviembre) debido a la temperatura del mar. Dadas estas condiciones se propone modelar al número de tormentas con base en una distribución Poisson, la cual precisamente tiene como fin modelar el número de llegadas a un sistema. La función de densidad de la distribución Poisson es la siguiente:

$$f(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}$$

donde

x es la variable aleatoria que mide el número de tormentas por temporada

λ es el parámetro que representa tanto la media como la varianza de la distribución

e es la constante exponencial $\sim 2.7182818...$

Las siguientes gráficas muestran el histograma comparativo entre los datos reales de tormentas por temporada en ambos océanos comparado contra 10,000 iteraciones de una simulación de distribuciones Poisson con el parámetro λ estimado bajo el criterio de máxima verosimilitud.

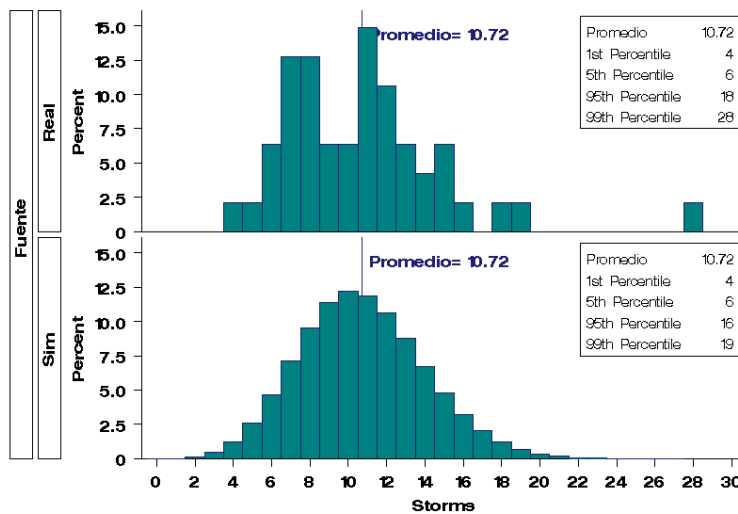


Figura 5. Histograma de tormentas reales por temporada en el Atlántico (n=58 años), contra 10,000 simulaciones de una Poisson ($\lambda=10.72$)

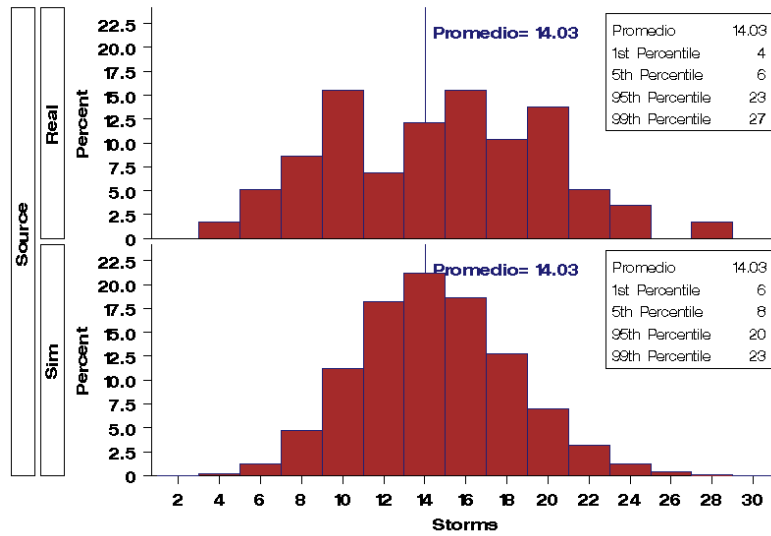


Figura 6. Histograma de tormentas reales por temporada en el Pacífico (n=58 años), contra 10,000 simulaciones de una Poisson ($\lambda=14.03$)

Número de lapsos por tormenta

La duración de cada tormenta se puede aproximar como el número de lapsos de 6 horas usando la información disponible de la base de datos. Vista la duración de esa forma, puede ser modelada usando una distribución discreta. Se propone modelar la duración de cada tormenta usando una función Binomial Negativa, cuya función de densidad está dada por la siguiente expresión:

$$f(k; r, p) = \binom{k+r-1}{r-1} p^r (1-p)^k$$

donde

k es la variable aleatoria que cuenta el número de lapsos de 6 horas de cada tormenta

r es un parámetro discreto no-negativo

p es un parámetro continuo que representa la probabilidad de éxito, $0 < p < 1$

La siguiente gráfica muestra un histograma comparativo entre las observaciones reales de las duraciones de cada tormenta en el Atlántico contra la simulación de una Binomial Negativa con parámetros $r = 4$, $p = 0.11252$.

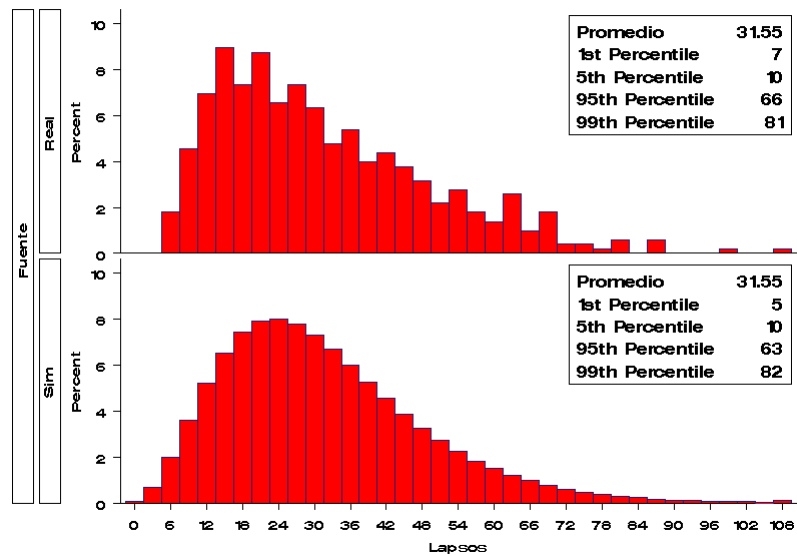


Figura 7. Histograma de duración de tormentas reales en el Atlántico (n=504 tormentas), contra 10,000 simulaciones de una Binomial Negativa (r=4, p=0.11252)

La duración depende de varios factores, siendo sin duda el primordial el hecho de contar con la fuente de energía que las alimenta. El diferencial de temperatura en el agua del océano es lo que en mayor medida puede alimentar la fuerza de los huracanes. Es decir que cuando las tormentas tocan tierra se debilitan y suelen desaparecer poco tiempo después. Asimismo, al encontrarse en latitudes altas pierden energía por la temperatura más fría de las aguas. En futuras versiones del modelo se pretende incorporar estos aspectos como parámetros de la función de duración.

Velocidad del viento

La base de datos histórica del NHC cuenta con información respecto a la velocidad sostenida del viento en el centro de la tormenta para cada lapso de 6 horas, *i.e.* se cuenta con 4 mediciones diarias, hechas a las 0:00 GMT (hora del meridiano de Greenwich), a las 6:00 GMT, 12:00 GMT y 18:00 GMT respectivamente. Por tanto, se podría modelar la velocidad del viento en cada lapso de 6 horas, lo cual será objeto de futuros trabajos de investigación. Para efecto de este documento solamente se modeló la velocidad máxima del viento de cada tormenta.

La velocidad del viento está medida en nudos (kts.) que equivale a 1 milla náutica y a su vez, a 1.852 kilómetros por hora (km/h). En este caso se trata de una variable continua no negativa que puede ser modelada usando una distribución Weibull, la cual es una generalización de la distribución Rayleigh que es usada precisamente para representar la velocidad del viento. Tiene dos parámetros (α , β) y su función de densidad está definida por la siguiente expresión:

$$f(x; \alpha, \beta) = \frac{\alpha x^{\alpha-1}}{\beta^\alpha} e^{-\left(\frac{x}{\beta}\right)^\alpha}$$

La siguiente gráfica muestra un histograma comparativo entre las velocidades máximas del viento para cada tormenta comparado contra simulaciones de una distribución Weibull con parámetros estimados por máxima verosimilitud.

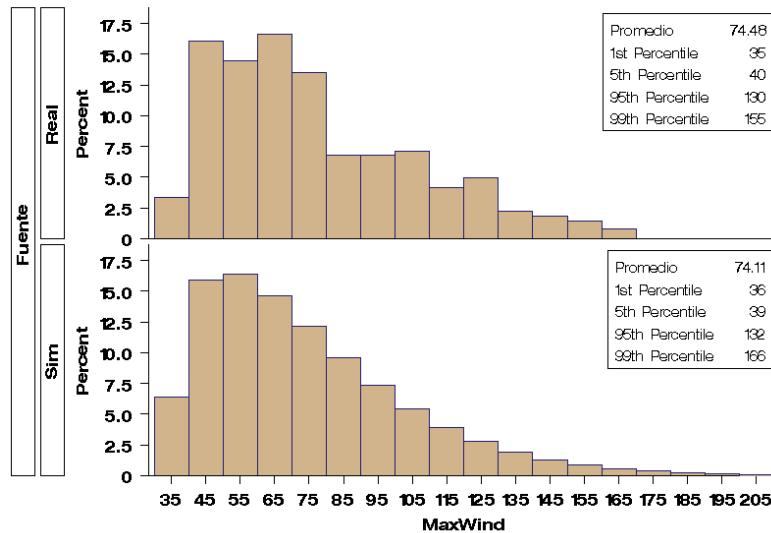


Figura 8. Histograma de velocidades máximas del viento de cada tormenta (n=621 tormentas) en el océano Atlántico contra 10,000 simulaciones de una distribución Weibull ($\alpha=1.35$, $\beta=43.984$)

Como se puede apreciar en el histograma la distribución Weibull replica de manera adecuada el comportamiento real de la velocidad máxima del viento. En el anexo se presentan las pruebas de bondad de ajuste realizadas.

De esta forma se le puede asignar de manera aleatoria, un nivel de velocidad máxima del viento a cada tormenta. Esta aplicación presenta un problema ya que se desconocería el momento en el cual esa velocidad máxima se presenta durante la trayectoria de la tormenta, lo cual es importante, ya que podría representar daños diferentes si esa velocidad máxima se presenta cuando hace contacto con la superficie continental a que si se presenta cuando la tormenta todavía se encuentra en medio del océano. Como se dijo anteriormente en futuras mejoras al modelo está contemplado modelar la distribución de la velocidad del viento durante toda la trayectoria de cada tormenta.

Sin embargo, esta propuesta de modelación de la velocidad máxima del viento es consistente con la escala Saffir-Simpson, con la cual se cataloga cada tormenta de acuerdo a su intensidad. Esta escala surgió en 1971 de manera conjunta por el meteorólogo Bob Simpson y el ingeniero civil Herbert Saffir con el fin de medir el daño e inundación potenciales que provoca un huracán al hacer contacto con tierra firme. La siguiente tabla muestra los parámetros sobre los cuáles una tormenta se considera huracán y se clasifica en categorías de acuerdo con la escala.

Saffir-Simpson Categoría	Velocidad máxima del viento		
	mph	m/s	kts
1	74-95	33-42	64-82
2	96-110	43-49	83-95
3	111-130	50-58	96-113
4	131-155	59-69	114-135
5	156+	70+	136+

Tabla 1. Escala Saffir-Simpson para categorizar los huracanes. Fuente: NOAA

Presión atmosférica

Así como la base de datos cuenta con información correspondiente a la velocidad del viento para cada momento de la trayectoria de las tormentas, también cuenta con información relativa a la presión atmosférica en el ojo de la tormenta en cada lapso. La presión atmosférica se define como el peso por unidad de superficie ejercida por la atmósfera y actualmente se mide en milibares que son equivalentes a los hectopascales (1 mb = 1 hPa). Esta variable también es continua sin embargo, por su naturaleza de que a menor presión mayor intensidad, su cola se presenta en la parte izquierda de la distribución en vez de la derecha como comúnmente ocurre. Por esta razón se propone modelar la presión atmosférica usando una distribución Weibull transformada. La siguiente gráfica muestra el histograma de presión atmosférica mínima de cada tormenta en el océano atlántico contrastado contra 10,000 simulaciones de una distribución Weibull transformada.

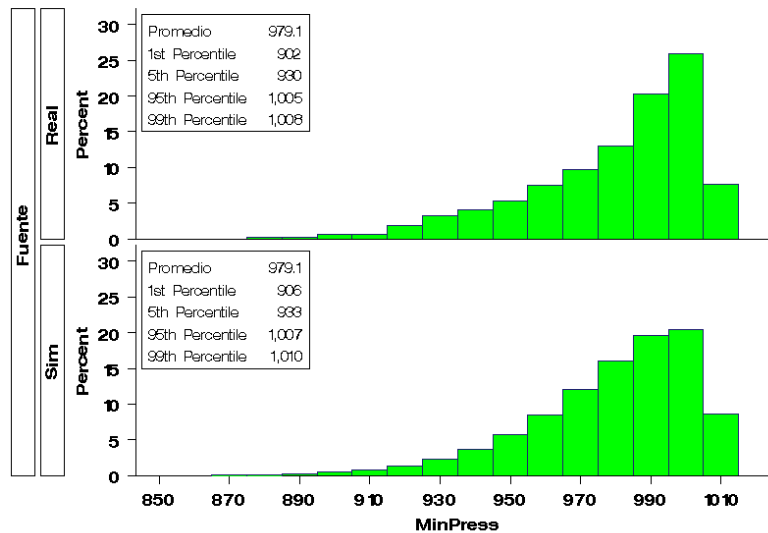


Figura 9. Histograma de presión atmosférica mínima para cada tormenta (n=621 tormentas), contra 10,000 simulaciones de una 1024 - Weibull ($\alpha=1.37$, $\beta=35.03$).

Lugar de gestación

La base de datos cuenta con la ubicación geográfica en coordenadas (latitud y longitud) donde se encuentra el centro de la tormenta en cada lapso de 6 horas, por lo cual se puede construir una distribución Kernel bivariada de la posición inicial.

La distribución Kernel es una distribución no paramétrica que en términos prácticos lo que hace es sobreponer una distribución definida encima de cada uno de los puntos donde se encuentre una observación, normalmente una distribución gaussiana. El ensamble de todas las distribuciones dispuestas en cada una de las observaciones corresponde a la distribución Kernel. La siguiente gráfica muestra un ejemplo teórico de lo anterior.

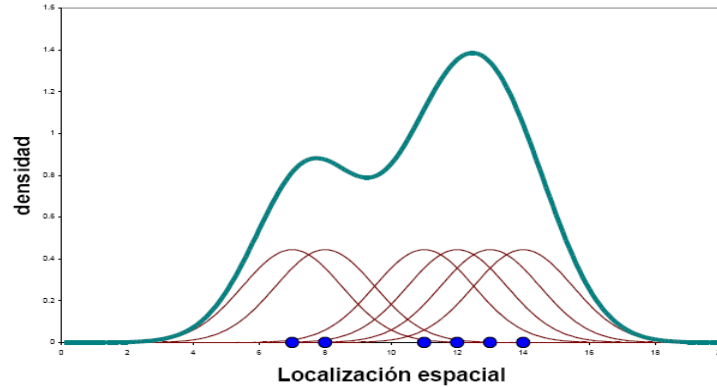


Figura 10. Ejemplo de distribución Kernel

Sin embargo, para el caso de la distribución geográfica del lugar de gestación, no se trata de una variable simple sino una distribución bivariada que consiste en la longitud y la latitud al mismo tiempo, por lo cual la propuesta es generar un Kernel en dos dimensiones como sigue:

$$f(x, y) = \frac{1}{n} \sum_{i=1}^n \varphi_{\mathbf{h}}(x - X_i, y - Y_i) = \frac{1}{n h_X h_Y} \sum_{i=1}^n \varphi\left(\frac{x - X_i}{h_X}, \frac{y - Y_i}{h_Y}\right)$$

Donde $\varphi(x, y)$ es una función de distribución Gaussiana bivariada. En estricto sentido, como se dijo anteriormente, el lugar de gestación de un ciclón tropical suele darse en latitudes bajas intertropicales, pero debido a que el NHC agrega un ciclón a la base de datos hasta que éste se convierte en tormenta tropical, el registro de lugar de gestación se refiere exclusivamente al momento en el que se convierten en tormentas y no al momento en que se gestan como ciclón. El siguiente mapa representa las ubicaciones geográficas donde históricamente se han convertido en tormentas.

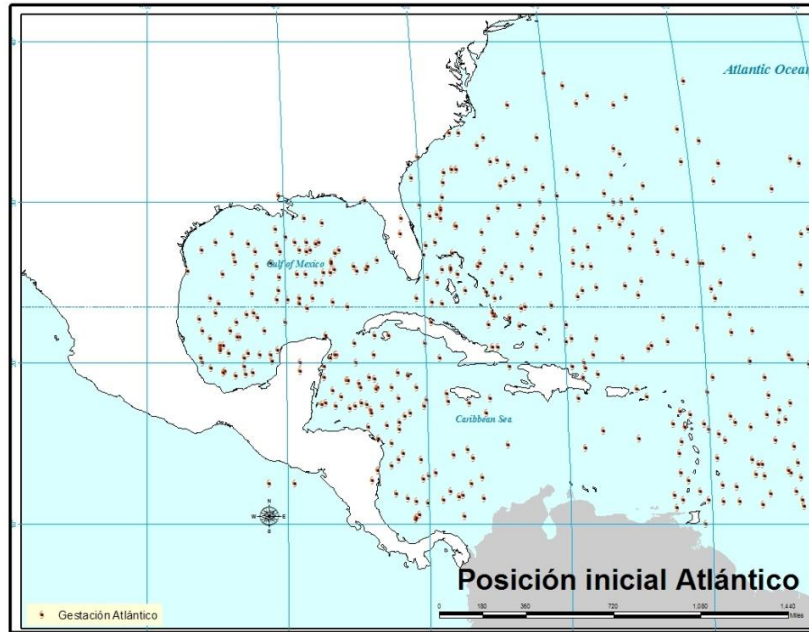


Figura 11. Mapa de lugar de gestación histórica de las tormentas en el Atlántico.

La aplicación del ajuste de un Kernel Gaussiano bivariado a las observaciones del mapa anterior genera una malla de densidades de probabilidad. En el siguiente mapa se puede apreciar las curvas de nivel expresadas en forma de malla, donde los puntos con colores rojos y anaranjados tienen la más alta probabilidad de ser seleccionados aleatoriamente como lugar de gestación, mientras que los puntos morados y rosas representan las zonas con baja densidad de probabilidad. El modelo genera, de manera aleatoria, un punto de la malla como ubicación inicial de cada tormenta usando la malla para la probabilidad de selección.

Otro punto que deberá corregirse en futuras versiones del modelo consiste en restringir los puntos de la malla a zonas oceánicas del Atlántico exclusivamente, es decir, quitar la posibilidad de que una tormenta se origine en tierra firme o en el Océano Pacífico.

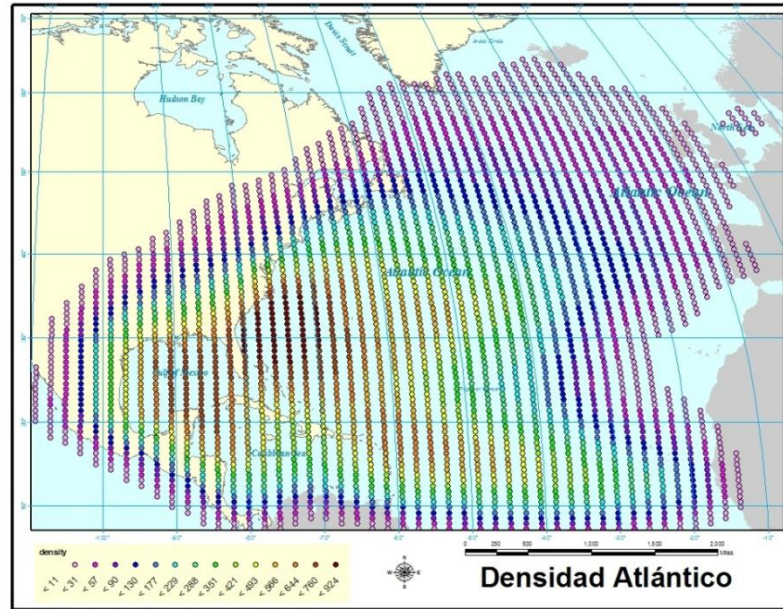
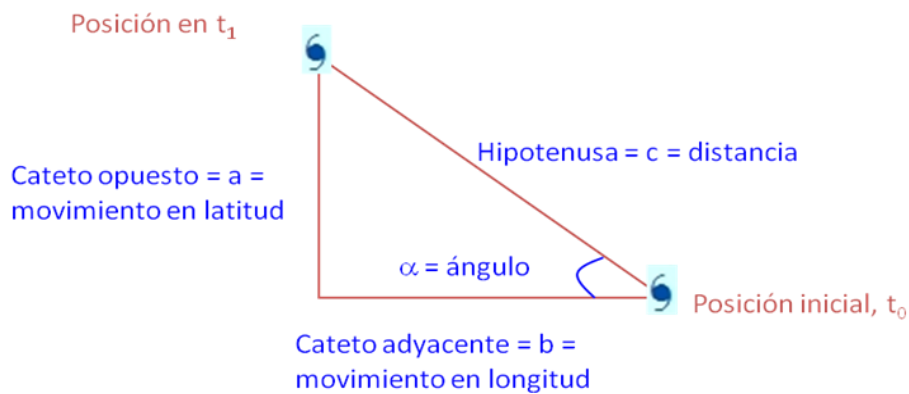


Figura 12. Mapa con curvas de nivel expresadas en forma de malla de densidades de probabilidad del Kernel bivariado.

Trayectoria

Para modelar el desplazamiento que sigue cada tormenta dado el punto de gestación simulado de la sección anterior, se propone el uso de trigonometría básica. Supóngase que el centro de la tormenta se encuentra en una posición inicial en el tiempo t_0 , y se desplaza hacia una nueva posición en el siguiente momento en t_1 . A través de la construcción de triángulos rectángulos y el uso de funciones trigonométricas se puede calcular la distancia que recorre, así como el ángulo de desplazamiento en cada momento. Los siguientes esquemas resumen la propuesta.



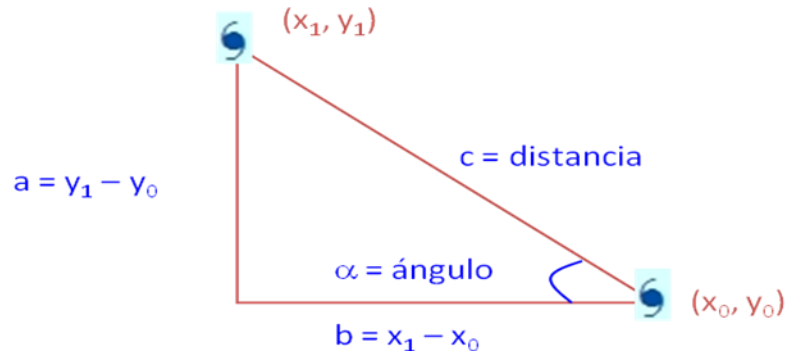


Figura 13. Esquema geométrico propuesto para modelar el desplazamiento.

De esta forma, la distancia que recorre en cada lapso es representada por la hipotenusa c , que una vez conocidas las coordenadas (x_0, y_0) y (x_1, y_1) se puede calcular con la siguiente fórmula:

$$c = \sqrt{a^2 + b^2} = \sqrt{(y_1 - y_0)^2 + (x_1 - x_0)^2}$$

Mientras que el ángulo de desplazamiento se puede calcular con diferentes funciones trigonométricas, por ejemplo la siguiente:

$$\alpha = \tan^{-1} \frac{a}{b}$$

Una vez que se tiene calculada la distancia que cada tormenta recorrió en cada lapso de 6 horas se puede construir una función de distribución de probabilidad que modele la distancia recorrida. Para esto se propone el uso de la distribución Kernel no paramétrica que tiene la siguiente forma:

$$f(x) = \frac{1}{\sum_{i=1}^n W_i} \sum_{i=1}^n W_i \varphi_h(x - X_i)$$

Donde: $\varphi_h(x) = \frac{1}{\sqrt{2\pi h}} e^{-\frac{x^2}{2h^2}}$ es la distribución Normal.

La siguiente gráfica muestra un histograma con la distribución real de distancias recorridas por todas las tormentas en cada lapso de 6 horas contra una simulación del Kernel propuesto. El uso de una distribución no paramétrica se fundamenta principalmente en que la cola que presenta la distribución es muy pesada, y ninguna de las distribuciones paramétricas es capaz de replicar ese comportamiento de una manera tan adecuada como el Kernel lo hace. Nótese, sobretodo, la similitud en la cola derecha (percentiles altos), que sería muy difícil de replicar con una distribución paramétrica.

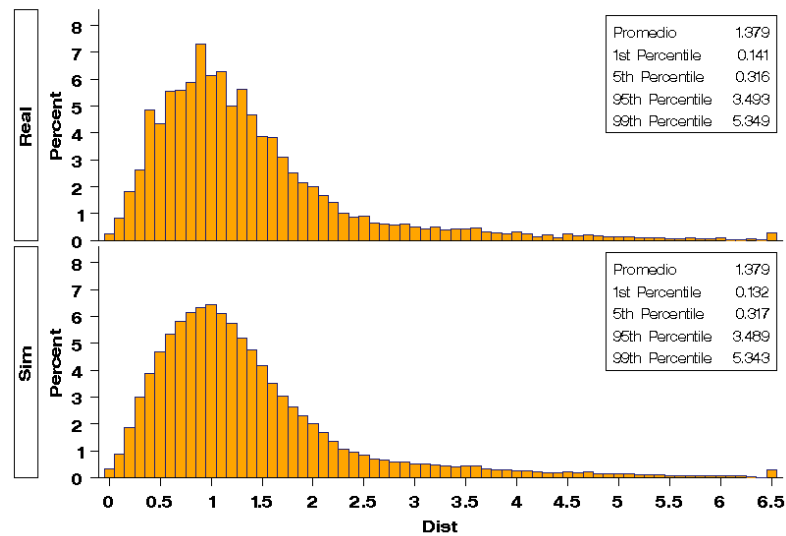


Figura 14. Histograma con distancias reales recorridas vs. Simulación del Kernel.

Asimismo, se puede modelar el ángulo de desplazamiento en cada lapso de 6 horas, la propuesta nuevamente consiste en usar una distribución no paramétrica por la complejidad de su comportamiento. Cabe hacer notar que el ángulo está transformado de radianes a grados para hacer más fácil su interpretación. La siguiente gráfica muestra el histograma comparativo entre los ángulos de desplazamiento reales contra la simulación del Kernel propuesto.

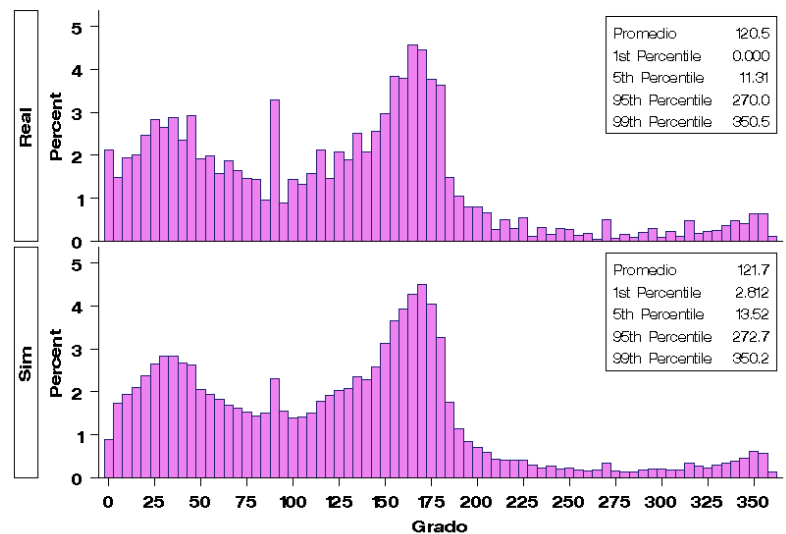


Figura 15. Histograma con ángulos de desplazamiento reales vs. Simulación del Kernel.

En particular para el caso del ángulo de desplazamiento es evidente que ninguna función paramétrica sería capaz de modelar las múltiples modas que presenta la distribución original. El Kernel en cambio, replica casi como un sastrero el comportamiento del ángulo de desplazamiento.

Las tormentas, sin embargo, no siguen una trayectoria caótica a lo largo de su vida. Generalmente siguen un camino estable en el cual el ángulo no cambia

drásticamente de un lapso a otro. Tampoco la distancia recorrida es muy diferente a la observada en el momento previo. Si se simularan ambas variables sin tomar en consideración dicha característica observaríamos movimientos caóticos que no replicarían con precisión el comportamiento real. Es por eso que se propone simular de manera correlacionada en el tiempo tanto el ángulo como la distancia, de manera que no se presenten dichos saltos caóticos. La siguiente gráfica señala la función de autocorrelación del ángulo descrito por las trayectorias en el Océano Atlántico con la cual se corrobora que no pueden ser considerados independientes.

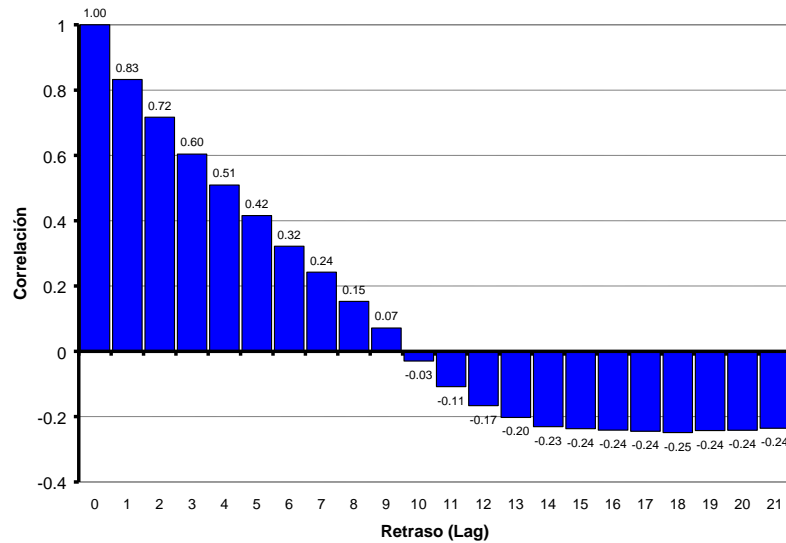


Figura 16. Función de autocorrelación del ángulo de desplazamiento

El siguiente mapa muestra algunas de las trayectorias simuladas dado un punto de gestación al mismo tiempo que se observan algunas trayectorias de tormentas reales con fines comparativos, en el cual se hace notar que si bien las trayectorias simuladas no replican a la perfección el comportamiento de trayectorias reales, al menos no presentan el comportamiento caótico descrito anteriormente.

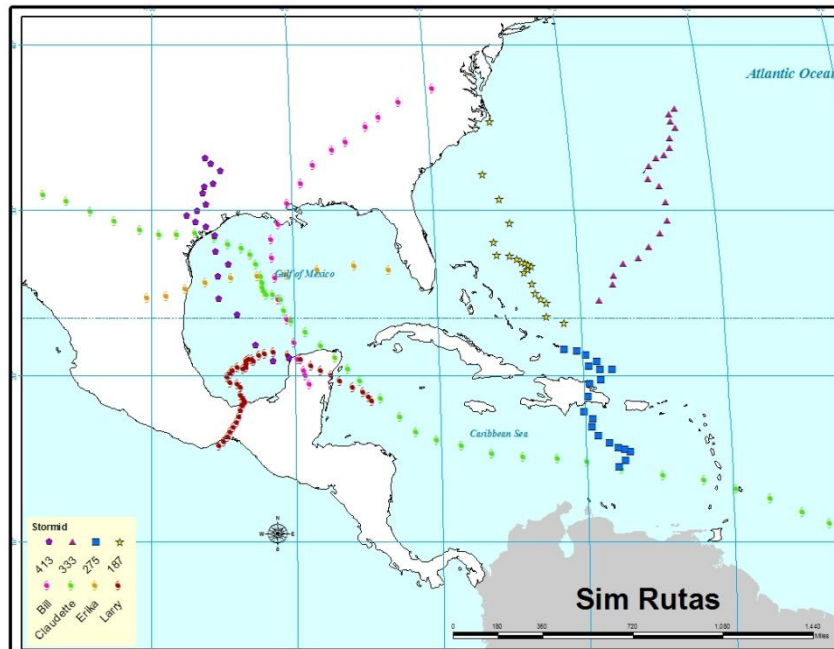


Figura 17. Mapa de trayectorias reales (símbolo de huracán) contra simuladas (otros símbolos).

El Modelo construido es capaz de simular un número grande (10,000 o más) de diferentes temporadas hipotéticas de tormentas tropicales, de modo que en lugar de contar con 50 años de historia real, contamos con 10,000 “años” de datos simulados.

Cada iteración representa una temporada hipotética. Para cada temporada simulada se cuenta con un número de tormentas aleatorio, y, a su vez, cada tormenta cuenta con un lugar de gestación y una duración aleatorias y únicas para cada tormenta. Una vez que se conoce ese lugar de gestación, la trayectoria de cada tormenta es única y aleatoria también, simulando tanto el ángulo de desplazamiento como la distancia recorrida en cada lapso de 6 horas. Asimismo, cada tormenta tendrá asignado, de manera aleatoria, un nivel de velocidad máxima así como un nivel de presión atmosférica mínimo. El último paso en la cadena sería asignar un valor de daño a la cartera que cada tormenta generaría, dependiendo de la exposición de los lugares por donde pase cada tormenta (en caso de tocar suelo mexicano), y de su intensidad.

Cabe hacer notar que para que el modelo funcione de una manera más exacta habría que asignar aleatoriamente un nivel de velocidad del viento así como de presión atmosférica para cada momento de la trayectoria, ya que el daño potencial sería mucho menor si el momento de velocidad máxima del viento se presenta cuando la tormenta se encuentra en el Océano lejos de hacer contacto con tierra firme a que si ese momento se encuentra cerca de los lugares con mayor exposición en la cartera. Se prevé incorporar la modelación de la velocidad del viento y la presión atmosférica para todo el trayecto de la tormenta en siguientes versiones.

La siguiente gráfica resume esquemáticamente el proceso de simulación de una sola iteración (una temporada hipotética) propuesto:

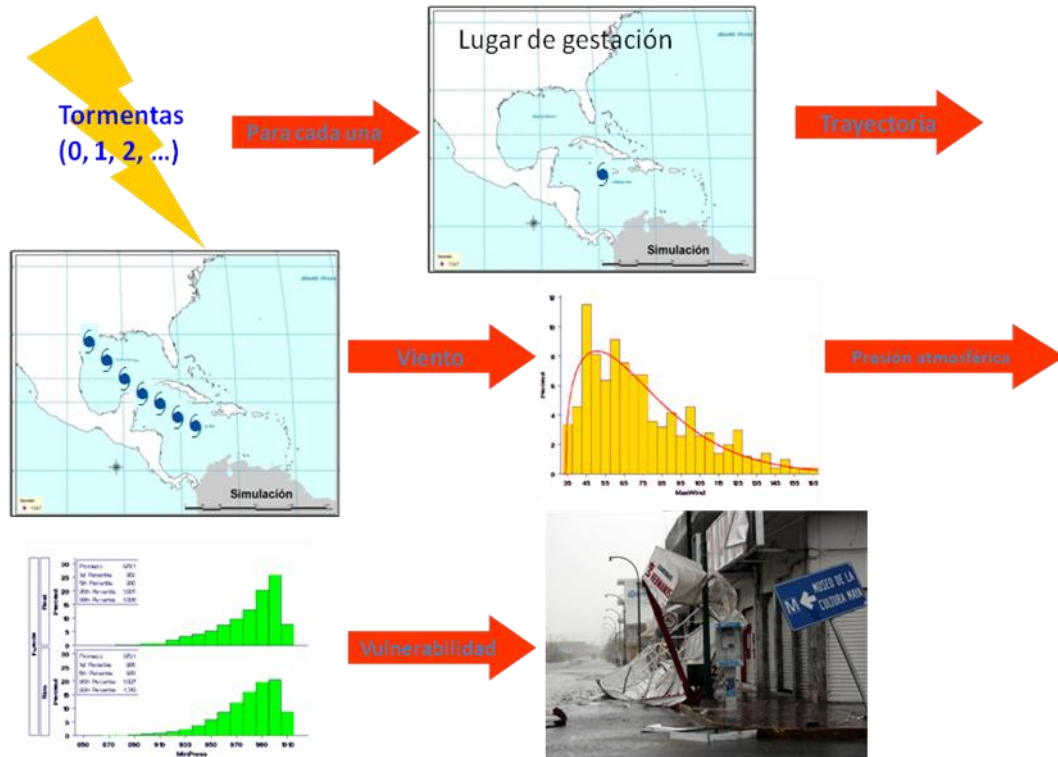


Figura 18. Proceso de simulación representativo de una temporada.

III. Ejemplo

Como ejemplo del funcionamiento del modelo a continuación se presenta el resultado de simular una temporada completa. Por principio de cuentas se simuló el número de ciclones tropicales que llegarían a convertirse en tormentas usando el Proceso Poisson descrito en el documento. En este ejemplo en particular el número de tormentas aleatorias fue de doce, lo cual es un poco más que el número esperado de tormentas (10.72), pero claramente dentro de lo esperado. El segundo paso fue seleccionar aleatoriamente el punto de gestación de cada una de ellas, usando para ello el Kernel bivariado. El siguiente mapa muestra los puntos de gestación de cada una de las 12 tormentas (puntos rojos).

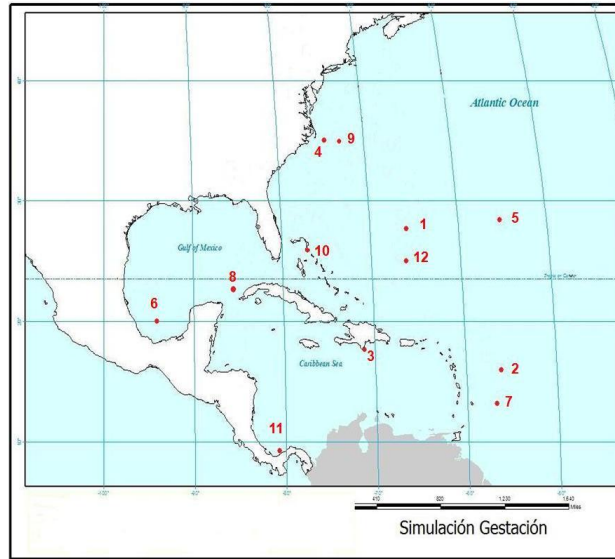


Figura 19. Puntos de gestación de las 12 tormentas simuladas de una temporada.

Una vez que se conoce el número de tormentas y su lugar de gestación se simula la trayectoria de cada tormenta, utilizando para esto el proceso descrito con trigonometría, es decir, para cada lapso de 6 horas de cada tormenta se obtiene un ángulo y una distancia respectivos. El siguiente mapa ejemplifica de manera gráfica el proceso para las 12 tormentas del mapa anterior. Cada tormenta tiene un color diferente para hacer más fácil el seguimiento de su trayectoria.

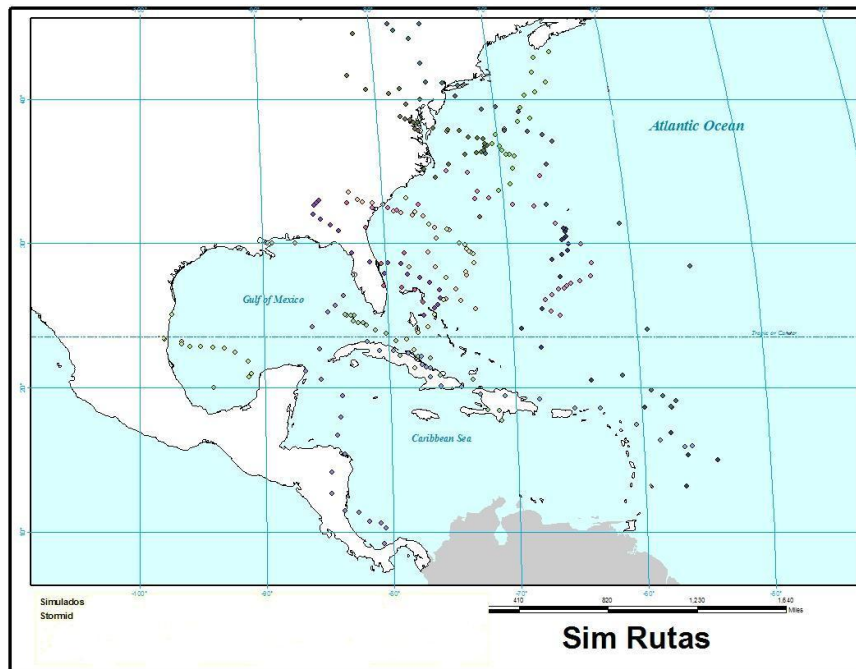


Figura 20. Trayectorias simuladas de las 12 tormentas de una temporada.

En este caso en particular se puede apreciar que la tormenta que se gestó en Centroamérica llegó a hacer contacto con tierra mexicana precisamente en la zona de Cancún. A su vez la tormenta que se gestó en el Golfo de México, la simulación de su trayectoria la llevó a hacer contacto con el estado de Tamaulipas.

Por último es necesario generar aleatoriamente la velocidad máxima del viento y la presión atmosférica mínima que cada tormenta tendrá. Para esto se utilizan las distribuciones Weibul descritas anteriormente. La siguiente tabla resume la velocidad del viento así como la presión atmosférica simuladas para cada tormenta. Asimismo se representa con colores la categoría de huracán, en su caso, al que cada tormenta llegó de acuerdo a la escala Saffir-Simpson.

Tormenta#	Viento	Presión	Categoría
1	123	940	Huracán 4
2	44	1,010	Tormenta
3	100	979	Huracán 3
4	66	998	Huracán 1
5	67	992	Huracán 1
6	60	1,004	Tormenta
7	54	1,011	Tormenta
8	49	1,014	Tormenta
9	80	981	Huracán 1
10	82	961	Huracán 1
11	35	1,019	Tormenta
12	112	990	Huracán 3

Tabla 2. Velocidad del viento máxima, presión atmosférica mínima y categoría de Huracán de las 12 tormentas.

IV. Conclusiones

La ventaja del uso del modelo comparado contra valuar una cartera directamente contra los datos históricos radica en que la valuación se basaría en un número mucho mayor de temporadas (10,000 o más contra 50 años de historia).

El modelo no ha sido aplicado aún como instrumento de valuación de riesgos a ninguna cartera debido a los asuntos pendientes principalmente los relacionados a las funciones de vulnerabilidad, que han sido expuestos en este documento, sin embargo, es evidente que el uso del modelo hará más precisos los cálculos de las pérdidas potenciales esperadas pero sobre todo aquellos referentes a las colas de la distribución de pérdidas (y por ende la pérdida máxima probable (PML) de la cartera).

Por otro lado, al simular el fenómeno se obtiene la posibilidad de encontrar tormentas que no han ocurrido en el pasado, tanto en trayectoria como en intensidad como en la combinación de ambas (trayectoria e intensidad), por lo cual la simulación puede arrojar escenarios relativos a un nivel de daño que si bien no se ha presentado en la historia pasada, es susceptible de ocurrir en el futuro.

V. Trabajos futuros

El modelo descrito en este documento replica estocásticamente el fenómeno de tormentas tropicales en general, sin embargo existen algunos elementos que necesitan ser mejorados que están previstos a ser resueltos en futuras versiones del modelo:

Modelación del Océano Pacífico. Fuera del número de tormentas por temporada, las demás variables fueron modeladas usando información

exclusivamente del Océano Atlántico, pero el Océano Pacífico tiene sus propias peculiaridades con respecto a los ángulos de desplazamiento así como al lugar de gestación por obvias razones. Éstos tendrán que ser modelados de manera independiente considerando la estadística propia de los sucesos históricos en esta cuenca oceánica.

Distribución para el número de tormentas por temporada. Si bien es cierto que el número de tormentas puede ser modelado con un proceso de llegadas a un sistema a través de la distribución Poisson, también es cierto que esta distribución tiene un solo parámetro, lo cual hace imposible ajustar media y varianza de manera separada cuando éstas no coinciden. El ajuste de otras distribuciones discretas podría suplir este defecto por lo que se propone probar otras distribuciones en trabajos posteriores.

Vínculo entre intensidad de viento o presión atmosférica con precipitación y mediciones de viento en estaciones terrestres. Cuando una tormenta presenta velocidades del viento altas la presión atmosférica observada es baja y viceversa lo cual implica que ambas variables están relacionadas entre sí. Hace falta incorporar en el modelo la correlación existente entre ambas variables si las funciones de vulnerabilidad dependen de ambos factores. Asimismo, es necesario establecer un vínculo entre los niveles de ambas variables con las mediciones de precipitación observadas en estaciones meteorológicas terrestres.

Buscar relaciones entre distancia-ángulo y latitud. Históricamente se observa que cuando una tormenta llega a latitudes altas (alejándose del Ecuador), la distancia que recorre así como el ángulo de desplazamiento tienden a dispararse de los parámetros comunes en latitudes bajas. Actualmente el modelo contempla ambas variables independientemente de la latitud en la cual se ubica el centro de la tormenta en cada momento. Se propone analizar esta relación e incorporarla dentro del modelo en versiones futuras.

Atenuar la intensidad de las tormentas al hacer contacto con tierra firme. La energía que alimenta las tormentas se encuentra en los océanos por lo que al hacer contacto con tierra firme éstas pierden fuerza y tienden a debilitarse. Actualmente el modelo no considera este hecho por lo cual se sugiere incorporar una función de atenuamiento de la intensidad de las tormentas (así como de su trayectoria) al reconocer que éstas han hecho contacto con las costas. Asimismo, si volvieran a hacer contacto con el océano podrían recuperar su fuerza como ha sido el caso de algunas tormentas (Gilbert, 1988 y Stan, 2005), lo cual deberá ser incorporado en el modelo en versiones futuras.

Perfeccionamiento del Kernel del lugar de gestación. Al momento de construir y simular el Kernel para el lugar de gestación no se considera si el punto se encuentra dentro del océano o en tierra firme, por lo cual algunas tormentas simuladas se gestaron aleatoriamente en tierra firme, lo cual es casi imposible de ocurrir. El modelo tendrá que ser revisado para evitar, en la medida de lo posible, este hecho en versiones futuras.

Modelación del radio de la tormenta. Hasta el momento solo se ha modelado el desplazamiento del centro de las tormentas. Pero en ocasiones no es el centro de la tormenta lo que provoca el daño sino las nubes que giran alrededor de él. Estas nubes suelen cubrir radios de varios cientos de kilómetros por lo cual el daño que provocan se puede extender a zonas alejadas del centro de las tormentas. Se ha documentado que el radio de las nubes depende de la presión atmosférica del centro por lo cual es factible modelar este aspecto en el futuro en función de la presión atmosférica modelada.

Funciones de vulnerabilidad. Dependerá de la cartera a la cual se le desee aplicar el modelo para cuantificar pérdidas, pero un punto vital a considerar es la función del daño observado en el bien dependiendo de la intensidad e ubicación de la tormenta. Muchos esfuerzos muy valiosos se han hecho en este aspecto por especialistas en la materia en el país, los cuales vale la pena incorporar para hacer una óptima valuación del riesgo.

Incorporación de modelos meteorológicos. Existen algunos modelos de pronóstico meteorológico desarrollados en el país que se basan en las condiciones físicas actuales observadas para modelar el estado futuro de la atmósfera. La combinación de ambos modelos (el meteorológico y el estocástico) podría ser de gran utilidad si lo que se desea es con fines de pronóstico para una temporada en particular.

Anexo

Pruebas de bondad de ajuste

Distribución Poisson para el número de tormentas por temporada en el Atlántico:

$$\chi^2 = 2.2157$$

Valor p = 0.8988

Distribución Poisson para el número de tormentas por temporada en el Pacífico:

$$\chi^2 = 8.1027$$

Valor p = 0.3236

Distribución Binomial Negativa para la duración de cada tormenta:

$$\chi^2 = 30.359$$

Valor p = 0.0642

Distribución Weibull para velocidad máxima del viento:

$$\text{Kolmogorov-Smirnov} = 0.0608$$

Valor p = 0.047

Anderson-Darling = 2.1081
Valor p = 0.084

Distribución Weibull para presión atmosférica:

Kolmogorov-Smirnov = 0.0567

Valor p = 0.084

Anderson-Darling = 2.2581

Valor p = 0.071

VI. Bibliografía

1. Ali, A. (1998). Nonparametric Spatial Rainfall Characterization Using Adaptive Kernel Estimator. *Journal of Geographic Information and Decision Analysis*, Vol. 2, No. 2, pp. 34-43.
2. Boissonade, A.C., Heitkemper L. J., and Whitehead D. (2002). Weather Data: Cleaning and Enhancement. In: Dischel, Robert (ed). *Climate Risk and the Weather Market: Financial Risk Management with Weather Hedges*. RISK Books. London.
3. Conde, C.; Ferrer R.M.; Araujo R.; Gay C.; Magaña V.; Pérez J.L.; Morales T. y Orozco S. (1999). El Niño y la Agricultura. *Los Impactos de El Niño en México*. Secretaría de Gobernación, México, 2004. Primera reimpression.
4. Delgadillo, J.; Aguilar T. y Rodríguez D. (1999). Los aspectos económicos y sociales de El Niño (1999). *Los Impactos de El Niño en México*. Secretaría de Gobernación, México, 2004. Primera reimpression.
5. Emmanuel, K.A., (2005): *Divine Wind: The history and science of hurricanes*. Oxford Univ. Press, New York, 304 pp.
6. Hess, U.; Richter K., Stoppa A. (NA). *Weather Risk Management for Agriculture and Agri-Business in Developing Countries*. IFC, World Bank and Procom Agr, Rome.
7. Morduch, J. (2001). *Rainfall insurance and vulnerability: Economic principles and cautionary notes*. New York University.
8. Mosiño, P.; García E. (1981). *The variability of rainfall in Mexico and its determination by means of the gamma distribution*. *Geografiska Annaler* 63A.
9. NOAA, <http://www.nhc.noaa.gov/>

10. Base de datos HURDAT Atlántico:
http://www.nhc.noaa.gov/tracks1851to2008_atl_reanal.txt
11. Base de datos HURDAT Pacífico:
http://www.nhc.noaa.gov/tracks1949to2008_epa.txt
12. Servicio Meteorológico Nacional, <http://smn.cna.gob.mx/>
13. Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall. London.
14. Vaughan, E. (1989). *Fundamentals of risk and insurance*. New York. John Wiley & Sons.
15. Wilks, D.S. (1995). *Statistical Methods in the Atmospheric Sciences (An Introduction)*. International Geophysics Series. Vol. 59. Academic Press. 1995.

Algoritmo basado en principios de asignación eficiente de recursos para unidades en un mapa de escala multidimensional
Caso 1: Fuente de bienes simple – nodos receptores múltiples

*MSc. Héctor J Alvarado Valdivia ASA.
Profesor de tiempo parcial del departamento de estadística del ITAM*

Kepler 134 Departamento 2, Col. Anzures
Del. Miguel Hidalgo 11590, México DF
Número telefónico: 50.12.79.79
alvahect@gmail.com

Traducción parcial del trabajo de tesis de maestría en estadística aplicada presentado en Septiembre de 2009 al panel examinador del departamento de estadística de la Universidad de Oxford, Inglaterra.

Resumen

En un mundo donde diversos puntos de compra están al alcance del consumidor, una asignación eficiente de los recursos es esencial. En este trabajo, es propuesto que dicha asignación eficiente puede ser lograda a través de adquirir grupos de bienes del mismo punto de compra o alternativamente agrupar teóricamente los puntos de compra entre si. Usando los principios de escala multimensional es posible identificar semejanzas y diferencias entre los puntos de compra (o bien entre los bienes) y construir un mapa de ellos en un plano bidimensional. Con base en la información del mapa, se pueden usar algoritmos para aproximar el

costo total mínimo necesario para adquirir todos los bienes deseados. Este trabajo presenta algoritmos mediante los cuales se puede aproximar dicho costo total mínimo en un mundo donde hay una sola fuente de bienes y múltiples nodos receptores de los mismos.

I. Introducción

El caso que trataré en este trabajo está definido en un mundo donde sólo hay una fuente de bienes (o proveedor) y varios nodos receptores (o consumidores)¹; en el cual, los consumidores demandan una cantidad entera finita de unidades de cada bien y donde las unidades demandadas pueden ser compartidas por dos o más consumidores. El número de unidades disponibles de cada bien se asumirá entero finito e igual al número total de unidades demandadas por los *consumidores desagrupados*; por lo cual, una asignación de recursos siempre es posible.

La asignación eficiente de los recursos presentada en este texto, será modelada desde la perspectiva de la entidad simple (en este caso la fuente de bienes). Por lo tanto, la solución escogerá la relación de correspondencia bien-consumidor que minimice el costo total para el *proveedor*.

¹ Para efectos del caso presentado en este trabajo, el lector deberá considerar los conceptos “fuente de bienes”, “entidad simple” y “proveedor” como sinónimos; así como también los conceptos “nodos receptores”, “entidades múltiples” y “consumidores” como sinónimos.

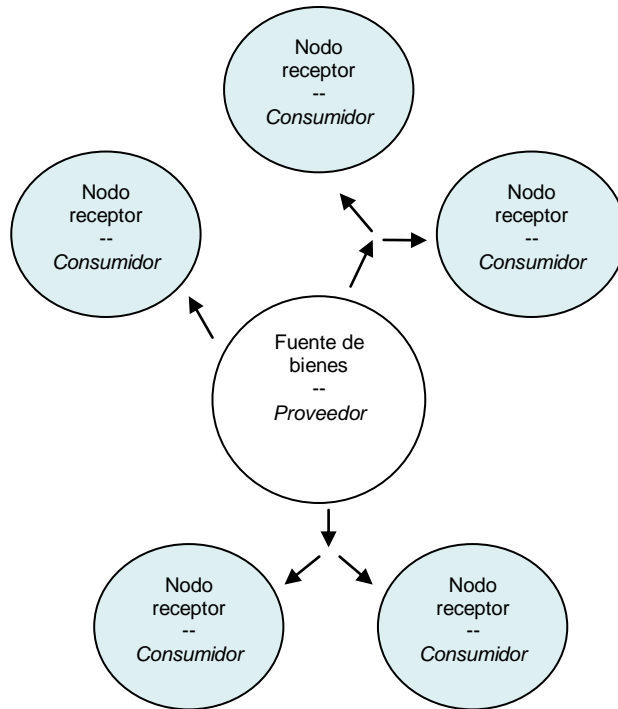


Figura 1. Fuente de bienes simple – nodos receptores múltiples

La idea principal sobre la cual se construyó la solución presentada en este trabajo es la agrupación teórica de las entidades múltiples (en este caso los nodos receptores). Basado en un mapa bidimensional construido vía el método de escala multidimensional, la distancia euclidiana entre las entidades múltiples puede ser usada para agrupar a los consumidores, ya sea por diferencias o por similitudes dependiendo del contexto del problema. La asignación de los recursos es hecha en paralelo con la agrupación de las entidades múltiples y se hace de la fuente de bienes hacia los nodos receptores (de ahí la dirección de las flechas en la Figura 1), buscando minimizar el costo total necesario para adquirir todos los bienes deseados.

Una característica esencial del método presentado es que busca explotar la naturaleza subaditiva de los modelos de costo de adquisición de bienes. Es decir, para un conjunto S de consumidores, asumiendo que las unidades pueden ser

compartidas entre ellos, la cantidad total de recursos $r(S)$ demandados por S satisface

$$r(S) \leq \sum_{i \in S} r(i)$$

y por lo tanto podemos decir que la demanda es subaditiva. En este caso, el modelo buscará explotar dicha característica de la demanda y encontrar la partición S_1, \dots, S_k que minimice $\sum_{i=1}^l r(S_i)$ a partir de un conjunto de particiones probables P de $\{1, \dots, l\}$ donde l es el número total de bienes en cuestión.

El objetivo principal de este trabajo es construir algoritmos que puedan encontrar asignaciones eficientes de recursos agrupando teóricamente las entidades múltiples (o consumidores) – graficadas a través de un mapa de escala multidimensional, basado en distancia euclidiana y considerando el modelo de costos del proveedor en cuestión.

El modelo presentado en este texto es un primer intento por modelar la asignación eficiente de recursos considerando los beneficios derivados de las interacciones entre los consumidores de bienes. Este trabajo está enfocado en citar las aplicaciones potenciales de estas ideas originales y, por lo tanto, deja espacio para más desarrollo teórico del modelo.

II. Consideraciones preliminares

II. 1 Planteamiento del problema

El tipo de problema considerado puede ser definido como un intento por minimizar la cantidad total de recursos² necesarios para satisfacer la demanda de todos los consumidores. Matemáticamente esto es buscar R tal que

$$R = \min \left(\sum_{i=1}^k r(i) \right)$$

donde

R = total de recursos necesarios; k = número total de nodos receptores; $r(i)$ = recursos demandados por el nodo receptor i , donde $r(i)$ es un vector con una entrada para cada bien del cual el consumidor i demanda al menos una unidad.

La teoría convencional de asignación de recursos busca encontrar un vector de ponderadores $w = (w_1, \dots, w_k)$ tales que el nodo receptor i reciba una cantidad igual a $E \cdot w_i$ de los recursos disponibles (ver referencia bibliográfica [2]) – donde E denota el total de recursos disponibles. El vector de ponderadores w puede ser calculado con base en diferentes criterios tales como maximizar los beneficios, minimizar la razón costo/beneficio, minimizar los costos, etc (ver referencia bibliográfica [2]). Sin embargo, estas ideas fallan en considerar los beneficios que pueden surgir al asignar recursos a un grupo de nodos receptores similares.

² El término “recursos” debe ser entendido por el lector como un conjunto de bienes.

La solución propuesta se desenvuelve alrededor de reducir la cantidad de recursos $r(i)$ al considerar las interacciones entre los diferentes nodos receptores y los beneficios que surgen de asignar recursos a grupos de ellos. Dadas estas consideraciones, el problema de asignación de recursos se convierte en un problema de dos pasos:

1. Reducir la cantidad de recursos $r(i)$ lo más que se pueda.
2. Calcular los ponderadores óptimos dados los costos que resultan de ejecutar el paso 1.

Este trabajo está enfocado en agrupar a los consumidores de una forma tal que el paso 1 se aproxime a la solución óptima.

Los problemas que pueden ser resueltos usando esta herramienta son muy diversos. Por ejemplo, los algoritmos pueden: construir los equipos más parejos a partir de una base de n jugadores, encontrar los grupos de planes de pensiones tales que la tasa de rendimiento sobre los activos necesaria para pagar los beneficios del grupo sea mínima, encontrar la mejor manera de asignar alumnos de diferentes programas a salones tal que el número de profesores y asistentes requeridos sea mínimo, encontrar la mejor manera de distribuir pacientes en un hospital tal que el número necesario de doctores, enfermeras y equipo de trabajo sea mínimo, construir equipos de trabajo tal que todos los proyectos puedan ser terminados eficientemente dados trabajadores con distintas características y niveles de experiencia, etc.

Para resolver estos problemas, se buscará encontrar diferencias y similitudes entre los consumidores. Por lo tanto, usaré conceptos de análisis multivariado

(escala multidimensional) y basaré la fase de creación de los grupos en las distancias punto a punto que resulten de aplicar escala multidimensional a la descripción original de los consumidores.

II. 2 Escala Multidimensional

Escala Multidimensional o Multidimensional Scaling (MDS) es una clase de métodos usados para hacer mapas de datos multidimensionales en espacios más reducidos, tal que la distancia entre los puntos sea preservada lo mejor posible [1].

En otras palabras, para cada par de puntos $(x_i, x_j) \in R^p$, se puede obtener una matriz D de diferencias punto a punto usando una medida de distancia $d(x_i, x_j)$;

MDS encuentra representaciones $z_1, \dots, z_n \in R^k$ tales que

$$d(x_i, x_j) \approx \tilde{d}(z_i, z_j)$$

donde d representa la distancia punto a punto en el espacio original de dimensión p y \tilde{d} representa la distancia punto a punto en el espacio reducido de dimensión k .

Los puntos (z_i, z_j) son aquellos que minimizan una “función de stress” que depende de cómo se defina la proximidad entre dos puntos. Para este trabajo, la proximidad entre dos puntos será considerada geoméricamente; por lo tanto, la función de stress a minimizar es

$$S_{metric}(d_{ij}, \tilde{d}_{ij}) = \frac{1}{\sum_{i \neq j} d_{ij}^2} \sum_{i \neq j} (d_{ij} - \tilde{d}_{ij})^2$$

definido como en la referencia bibliográfica [1].

Para agrupar a los consumidores usando MDS en este caso, es necesario conocer las características de cada consumidor. Sea γ_t un vector de dimensión p tal que cada una de sus entradas describe una característica del consumidor t y Γ_p una matriz con un renglón por cada consumidor (k consumidores) y una columna por cada característica.

Usando MDS, todos los consumidores pueden ser trasladados a un mapa bidimensional preservando la distancia entre los puntos originales lo mejor posible. Las coordenadas en la matriz resultante Γ_2 pueden ser usadas para identificar a los consumidores que son similares entre si y como la base para agruparlos.

II.3 Supuestos

Este trabajo busca crear algoritmos para *aproximar* la cantidad mínima de recursos necesarios para satisfacer a todos los consumidores bajo los siguientes supuestos:

1. Todos los bienes tienen el mismo costo
2. Los grupos de consumidores serán construidos con base en una “*guía de agrupación*”
3. Todos los nodos tienen la misma estructura y solo bienes similares pueden ser compartidos entre ellos.
4. La oferta de bienes no está restringida (no hay restricción presupuestal)
5. No hay restricción en cuanto al número de grupos de consumidores que pueden ser formados por el algoritmo.

II.4 Definiendo el nodo receptor

Para efectos de este trabajo, todos los nodos receptores serán definidos por dos vectores: uno para describir al nodo y el otro para listar los recursos demandados por el mismo (con cardinalidades δ y ε respectivamente). Un ejemplo de nodo descrito por 6 dimensiones ($\delta = 6$) tal que requiere 3 recursos ($\varepsilon = 3$) está ilustrado en la Figura 2.

Es posible estructurar un nodo através de un solo vector que sirva como descriptor y como lista de recursos requeridos al mismo tiempo; sin embargo, crear el algoritmo basado en un nodo estructurado como se ilustra en la Figura 2 es más flexible y permite llegar a una solución más general.

Node - W Rooney				
	Description	Required resources		
Position ->	Fwd	10	Fwd	<- Coach hours type
Goals ->	80			
Games Played ->	181	40	4	<- Primary field % section
Goal/Games Ratio ->	0.35			
Age ->	23	10	3	<- Secondary field % section
Weight ->	12.08			

Figura 2. Estructura del nodo receptor

II.5 Consideraciones y definiciones

Guía de agrupación. Dependiendo del contexto del problema, el usuario deberá “decirle” al algoritmo si debe agrupar por similitudes o por diferencias

Similitudes y diferencias entre los grupos. Cuando se agrupa por similitudes, todos los nodos en el mismo grupo serán similares entre si; por lo tanto, los grupos

serán muy diferentes entre ellos (todos los individuos altos estarán agrupados con individuos altos y todos los individuos pequeños estarán agrupados con individuos pequeños). Cuando se agrupa por diferencias, todos los nodos del mismo grupo serán diferentes entre si; por lo tanto, los grupos se parecerán entre ellos (todos los grupos tendrán individuos altos y pequeños).

Localización de los centros. MDS creará un mapa bidimensional donde unidades similares estarán cerca unas de las otras; por lo tanto, si se pretende agrupar por diferencias, los centros iniciales deberán estar lo más cerca posible – pues estamos buscando construir los grupos más parejos posibles. Asimismo, cuando se pretenda agrupar por similitudes, los centros iniciales deberán estar lo más lejos posible – pues estamos buscando crear los grupos más disparejos posibles.

Definición de centro. Un “centro” es un nodo alrededor del cual el algoritmo intentará construir un grupo. Un nodo se convierte en centro cuando se le da una “*etiqueta de centro*” y deja de serlo cuando dicha etiqueta es eliminada – las etiquetas son asignadas como números en orden ascendente. Inicialmente, habrá tantos centros como numero mínimo de grupos, el primer centro es escogido aleatoriamente entre los nodos disponibles y los centros subsecuentes son asignados con base en la regla de *localización de centros*.

Definición de vacío. El término “vacío” en el algoritmo será usado cuando no se pueda seleccionar ningún nodo. Por ejemplo, si el algoritmo necesita seleccionar el siguiente centro pero no queda ningún centro disponible, la selección será “vacío”; asimismo, si el algoritmo necesita seleccionar el siguiente nodo más cercano pero no quedan nodos disponibles, la selección será “vacío”.

Definición de nodo agrupado. Un nodo pertenece a un grupo si tiene una etiqueta de grupo. Las etiquetas de grupo son puestas igual al número de la “etiqueta de centro” del nodo que se encuentre al centro del grupo.

Cerrar un grupo. En el algoritmo, la expresión “cerrar el grupo” significa: Sólo para el nodo al centro del grupo, poner la etiqueta de grupo igual a la etiqueta de centro y eliminar la etiqueta de centro.

III. Solución

El algoritmo propuesto como solución al problema de asignación de recursos está dividido en dos secciones:

1. Escoger el siguiente nodo disponible – respetando la numeración de los centros.
2. Revisar si la agrupación es eficiente.

Paso 0. Definir la capacidad y las reglas de combinación de los nodos – para ser usadas en la sección dos del algoritmo.

III.1 Escoger el siguiente nodo disponible

Esta sección está escrita asumiendo que los nodos serán agrupados por similitudes – los ajustes necesarios para agrupar por diferencias están en paréntesis donde es necesario.

Paso 1.0 Hacer el análisis MDS, producir el mapa bidimensional y calcular las distancias entre los puntos en la matriz Γ_2 .

Paso 1.1 Inicializar. Definir si se agrupará por similitudes o por diferencias, escoger los nodos centrales, etiquetarlos con números en orden ascendente y definir el número máximo de nodos por grupo (ν). Al terminar este paso, los

únicos nodos con etiqueta deberán ser los nodos centrales y las únicas etiquetas asignadas deberán ser las etiquetas de centro.

Paso 1.2 Selecciona el siguiente centro. Si la selección es “vacío” ve al Paso 1.3; si la selección no es “vacío” ve al paso 1.4.

Paso 1.3 Si todos los nodos tiene una etiqueta de grupo entonces Detenerse; si no todos los nodos tienen etiquetas de grupo entonces escoger aleatoriamente un nodo sin etiqueta de grupo, etiquetarlo como el siguiente centro y ejecutar de nuevo el paso 1.2

Paso 1.4 Seleccionar el nodo disponible más cercano (más lejano si el agrupamiento es por diferencias).

- Si la selección es “vacío” entonces no hay más nodos disponibles para ese grupo → Cerrar el grupo y regresar al Paso 1.2
- Si la selección no es “vacío” entonces revisar si la agrupación es eficiente
 - Si la agrupación es eficiente entonces
 - ★ Poner la etiqueta de grupo del nodo seleccionado igual a la etiqueta de centro del nodo central.
 - ★ Si el número de nodos en el grupo es igual a v (el grupo está lleno) → Cerrar el grupo y regresar al Paso 1.2
 - ★ Si el grupo no está lleno → Regresar al Paso 1.2
 - Si la agrupación no es eficiente entonces
 - ★ Etiquetar el nodo seleccionado como “no disponible” para ese grupo – el nodo permanece disponible para todos los grupos que no lo han etiquetado como “no disponible”.

III.1 Revisar si la agrupación es eficiente

Este paso es para revisar si la inclusión del nodo seleccionado al grupo reduce los recursos demandados por el grupo aumentado. Si este es el caso, entonces la agrupación es marcada como eficiente; si no es el caso, entonces la agrupación es marcada como ineficiente.

Esta revisión será modelada como una red de dos nodos (uno es el grupo y otro el nodo a incluir) donde el arco que conecta a los nodos tiene cierta capacidad y la pregunta a responder es si el flujo es aceptado o no – el flujo es aceptado si y solo si es menor o igual a la capacidad del arco. Bajo este modelo, hay dos grupos de reglas que serán necesarias (estas se determinan en el Paso 0 del algoritmo).

1. Reglas para definir la capacidad del arco – inicialmente iguales a la suma de los recursos demandados por los nodos.
2. Reglas para definir el flujo del arco – estas son las reglas de inclusión del nodo seleccionado.

Paso 2.0 Poner el flujo actual igual a cero, borrar la *lista auxiliar de recursos requeridos por el grupo aumentado* y calcular la capacidad del arco con base en las reglas determinadas en el Paso 0 del algoritmo.

Paso 2.1 Escanear cada recurso requerido por el grupo comparándolo con todos los recursos requeridos por el *nodo a incluir* de la siguiente manera:

- Si el mismo tipo de recurso es encontrado:
 - Flujo actual = Flujo actual + Flujo basado en las reglas de inclusión
 - Añadir el recurso escaneado a la *lista auxiliar de recursos requeridos por el grupo aumentado* con el flujo correspondiente

- Etiquetar el recurso en el nodo a incluir como “escaneado”
- Seleccionar el siguiente recurso a escanear
- Si el mismo tipo de recurso no es encontrado:
 - $\text{Flujo actual} = \text{Flujo actual} + \text{Flujo del recurso que se está escaneando}$
 - Añadir el recurso escaneado a la *lista auxiliar de recursos requeridos por el grupo aumentado* con el flujo correspondiente
 - Seleccionar el siguiente recurso a escanear

Paso 2.2 Escanear todos los recursos del nodo a incluir que no tengan etiqueta de “escaneado” de la siguiente manera:

- $\text{Flujo actual} = \text{Flujo actual} + \text{Flujo del recurso que se está escaneando}$
- Añadir el recurso escaneado a la *lista auxiliar de recursos requeridos por el grupo aumentado* con el flujo correspondiente

Paso 2.3 Si el “Flujo actual” es mayor que la capacidad entonces el resultado es “Agrupación no eficiente”; de otra forma el resultado es “Agrupación eficiente” y los recursos requeridos por el grupo aumentado son iguales a la información en la *lista auxiliar de recursos requeridos por el grupo aumentado*.

Este algoritmo deberá ser usado solamente si la meta principal es construir grupos con base en medidas de similitud/diferencia y no buscando minimizar la cantidad total de recursos requeridos (el algoritmo sólo aproxima el mínimo, se enfoca en considerar la relación entre los nodos receptores). Tales escenarios pueden surgir cuando se construyan portafolios de inversión en los cuales la diversificación y la administración del riesgo sean tan importantes como maximizar el retorno de la inversión; también, dicho algoritmo puede ser útil cuando se intente construir

equipos de trabajo donde se agrega valor al hacer que trabajadores experimentados transmitan conocimiento a trabajadores con poca experiencia a expensas de una reducción potencial en la eficiencia del proyecto, etc.

Una vez que el mapa de escala multidimensional es construido, el algoritmo encuentra, en un número de iteraciones, los grupos en

$$\tau_1 = k \leftarrow \text{Máximo } k \text{ centros}$$

* $\nu \leftarrow$ Selecciona r nodos hasta que el grupo esté lleno

* $3 \cdot \varepsilon^2 \leftarrow$ Revisar si la agrupación es eficiente

$$\approx O(k \cdot \nu \cdot \varepsilon^2)$$

IV. Aplicaciones prácticas

IV.1 Clases a nivel universitario

La meta es encontrar grupos de estudiantes tales que el total de recursos requeridos sea calculado considerando las similitudes y la colaboración entre los diferentes departamentos.

Datos de 17 clases a nivel licenciatura fueron tomados de las referencias bibliográficas [3], [4], [5] y los vectores descriptores fueron capturados en R como se muestra en la Figuras 3 y 4.

Name	Description					
	Division	Department/ School/Faculty	Program	Lecture Classifier 1	Lecture Classifier 2	Lecture Classifier 3
Chem Calculus	10	5	1	0	0	0
Chem Matrix Alg	10	5	1	2.5	2.5	2.5
Bio Stats & Data H	10	10	2	5	5	5.0
Bio Quant Methods	10	10	2	7.5	7.5	7.5
Math Applied Stats	10	15	3	4.5	5.5	4.0
Math Linear Algebra	10	15	4	3	2	2.0
Math Calculus	10	15	4	0.5	0.5	0
Math Stats Inference	10	15	3	10	10	10.0
Stats S Methods	10	20	5	5.5	6	5.0
Stats FS Methods	10	20	5	7	8	6.5
Stats S Theory	10	20	5	9	9.5	11.0
Antr Human Gen & Evol	20	25	6	12.5	12.5	12.5
Arch Human Evol	20	30	6	12	13.5	12.0
Class Ancient Greek	30	35	7	15	15	15.0
Class Latin	30	35	7	17.5	17.5	17.5
Hist Medieval Greek	30	40	8	16.5	15	14.0
Hist Medieval Latin	30	40	9	16.5	17	16.5

Figura 3. Descriptores de las clases a nivel universitario

Name	Professor		Teacher assistant		Classroom	
	#	type	#	type	#	type
Chem Calculus	1	A	3	A	193	L
Chem Matrix Alg	1	A	3	A	193	L
Bio Stats & Data H	1	B	1	B	90	S
Bio Quant Methods	1	B	1	B	90	S
Math Applied Stats	1	B	1	B	26	S
Math Linear Algebra	1	A	3	A	169	L
Math Calculus	1	A	3	A	169	L
Math Stats Inference	1	B	1	B	26	S
Stats S Methods	1	B	1	B	28	S
Stats FS Methods	1	B	1	B	28	S
Stats S Theory	1	B	1	B	28	S
Antr Human Gen & Evol	1	C	1	C	21	S
Arch Human Evol	1	C	1	C	21	S
Class Ancient Greek	1	D	2	D	122	L
Class Latin	1	E	2	E	122	L
Hist Medieval Greek	1	D	3	D	233	L
Hist Medieval Latin	1	E	3	E	233	L

Figura 4. Recursos requeridos por las clases de nivel universitario

Algunos comentarios preliminares:

- Los profesores son clasificados como tipo: A para las clases de teoría matemática, B para las clases de matemática aplicada, C para humanidades, D para el lenguaje griego y E para el latín.
- El número de alumnos por salón es igual al número de alumnos admitidos al programa en 2008-2009 [5].

- Los clasificadores de las clases fueron asignados con base en el análisis detallado del syllabus de la materia buscando que materias con syllabus similares estuvieran juntas unas de las otras.

La estructura del nodo receptor para este caso particular está ilustrada en la Figura 5.

Node - Chem Calculus				
Description		Required resources		
Division ->	M, Ph & L	1	A	<- Professor type
Program department ->	Chem			
Program ->	Chem	3	A	<- Teacher Assistant type
Lecture classifier 1 ->	0			
Lecture classifier 2 ->	0			
Lecture classifier 3 ->	0	193	L	<- Classroom type

Figura 5. Estructura del nodo receptor

Las reglas de inclusión son:

- Para el mismo tipo de profesor → Número resultante de profesores = 1. Un profesor puede enseñar la clase indistintamente del departamento al cual la clase pertenezca o el número de estudiantes que la cursan.
- Para el mismo tipo de asistente
 - Si es tipo A → Número resultante de asistentes = 1.5 por cada 100 estudiantes
 - Si no es tipo A → Número resultante de asistentes = 1 por cada 100 estudiantes
- Para el mismo tipo de salón → Número resultante de estudiantes en el salón = Suma de los estudiantes de ambos nodos

Reglas de capacidad:

- El arco puede sostener un profesor como máximo.
- El arco puede sostener un máximo de 6 asistentes tipo A, 3 asistentes tipo B ó C o bien 5 asistentes tipo D ó E.
- El arco puede sostener un máximo de 120 estudiantes en un salón pequeño (S) ó 400 estudiantes en un salón grande (L).
- El arco puede sostener estudiantes de un mismo programa solo una vez (de otra forma estaríamos contando estudiantes dos veces).

Con estas especificaciones y los datos en las Figuras 3 y 4, el mapa de escala multidimensional (Figura 6) y la matriz de distancias entre los puntos de la matriz Γ_2 fueron calculados.

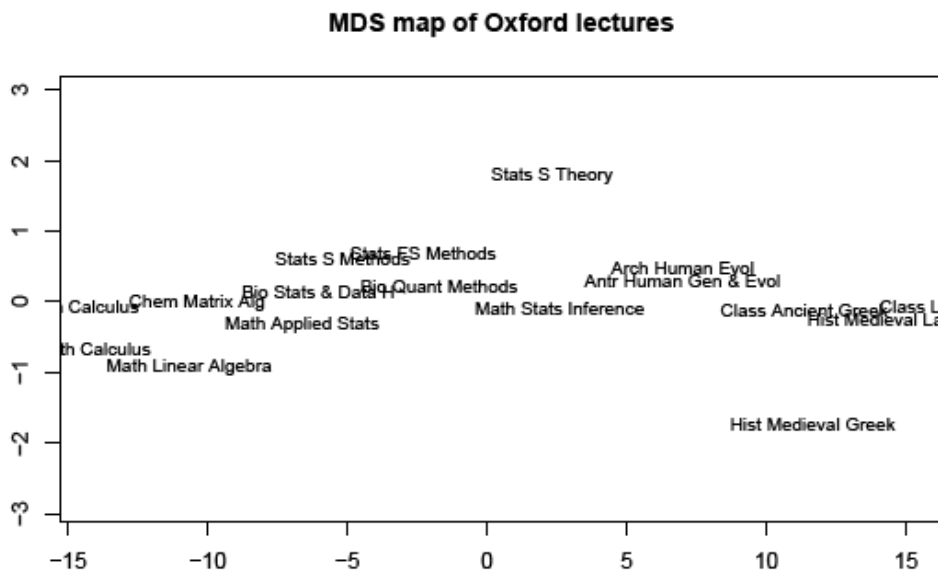


Figura 6. Mapa de las clases a nivel universitario

Dado que busco clases especializadas, el algoritmo debe ser inicializado con centros lo más alejados posible unos de los otros. Para este estudio, escogí

empezar subjetivamente con 5 centros y sin limitar el número de nodos en cada grupo. Seleccionando el nodo "Stats S Methods" aleatoriamente, los siguientes cuatro centros son: "Class Latin", "Chem Calculus", "Hist Medieval Greek" and "Math Linear Algebra" – numerados 1 al 5 respectivamente.

Ejecutando el algoritmo respetando el orden de las etiquetas de centro y de acuerdo con las reglas de inclusión, resultan los pasos siguientes:

- Centro *Stats S Methods* añade *Bio Stats & Data H* (1) → rechaza todos los nodos adicionales hasta crear un nuevo centro seleccionando "*Bio Quant Methods*" aleatoriamente (41)
 - Centro *Bio Quant Methods* añade *Stats FS Methods* (42) → rechaza todos los nodos adicionales hasta crear un nuevo centro seleccionando "*Math Applied Stats*" aleatoriamente (48)
 - Centro *Math Applied Stats* añade *Stats S Theory* (49) añade *Math Stats Inference* (50) → rechaza todos los nodos adicionales hasta crear un nuevo centro seleccionando "*Antr Human Gen & Evol*" aleatoriamente (53)
 - Centro *Antr Human Gen & Evol* añade *Arch Human Evol* (54) → **Detiene el algoritmo**
- Centro *Class Latin* añade *Hist Medieval Latin* (2) → rechaza todos los nodos adicionales
- Centro *Chem Calculus* añade *Math Calculus* (3) → rechaza todos los nodos adicionales

- Centro *Hist Medieval* añade *Class Ancient Greek* (4) → rechaza todos los nodos adicionales
- Centro *Math Linear Algebra* añade *Chem Matrix Algebra* (5) → rechaza todos los nodos adicionales

El algoritmo crea los grupos en las Figuras 7 y 8, donde se puede ver que agrupar las clases de esta forma reduce considerablemente la cantidad de recursos requeridos.

	Group 1	Group 2	Group 3
Lectures	Bio Stats & Data H Stats S Methods	Class Latin Hist Medieval Latin	Chem Calculus Math Calculus
Professors	1	1	1
Teacher Assistants	2	4	6
Students in small classroom	118	-	-
Students in large classroom	-	355	362
	Group 4	Group 5	Group 6
Lectures	Class Ancient Greek Hist Medieval Greek	Chem Matrix Alg Math Linear Algebra	Bio Quant Methods Stats FS Methods
Professors	1	1	1
Teacher Assistants	4	6	2
Students in small classroom	-	-	118
Students in large classroom	355	362	-

Figura 7. Grupos de clases 1

	Group 7	Group 8	Totals
Lectures	Math Applied Stats Math Stats Inference Stats S Theory	Antr Human Gen & Evol Arch Human Evol	After / Before the algorithm
Professors	1	1	8 / 17
Teacher Assistants	1	1	26 / 31
Students in small classroom	80	42	4 / 9
Students in large classroom	-	-	4 / 8

Figura 8. Grupos de clases 2

El resultado puede ser interpretado como los grupos de clases tales que, cuando con agrupados con base en colaboración interdepartamental y similitudes en el syllabus, mejor reducen la cantidad total de recursos requeridos por la universidad.

IV.2 Pensiones

La meta principal es construir 5 grupos de 3 planes de pensiones cada uno tales que los beneficios a pagar sean lo más nivelados posible entre ellos.

De acuerdo con la legislación de pensiones, en Estados Unidos, todas las compañías publicas que coticen en la bolsa deben presentar, cada año, el estado de cambios en sus pasivos de pensiones, el estado de cambios en el valor corriente de sus activos (MVA) y los de una proyección de los siguientes 10 años de pagos de beneficios.

Para este ejemplo, el MVA y los 10 años de pagos de beneficios fueron tomados de la sección “Notes to financial statements” del estado de resultados de 2008.

Los reportes de 15 compañías en la lista “Fortune 500” fueron capturados y clasificados de la siguiente manera:

- Las compañías fueron codificadas como 100 para bancos, 200 para compañías en telecomunicaciones, 300 para banca de inversión, 400 para manufactureros de computadoras, 500 para manufactureros de vehículos, 600 para compañías petroleras, 700 para compañías de seguros y 800 para compañías manufactureras de artículos para el hogar.
- La tasa interna de interés de equilibrio (TIIE) fue estimada asumiendo que el último pago de beneficios se mantiene constante por los siguientes 75 años y tomando el MVA como la única fuente de activos para pagar todos los beneficios futuros.

El nodo para este caso tiene la siguiente estructura:

Node - Ford		Required resources	
Description			
Year 1 BP ->	\$ 3,910.00	7.27%	500
Year 2 BP ->	\$ 3,850.00		
Year 3 BP ->	\$ 3,770.00		
Year 4 BP ->	\$ 3,620.00		
Year 5 BP ->	\$ 3,530.00		
Year 6 BP ->	\$ 3,415.00		
Year 7 BP ->	\$ 3,360.00		
Year 8 BP ->	\$ 3,278.00		
Year 9 BP ->	\$ 3,234.00		
Year 10 BP ->	\$ 3,103.00		

← IRR* | type

*based on a MVA of \$ 45,759

Figura 9. Estructura del nodo receptor

Dado que todas las compañías tienen diferentes niveles de activos, los datos fueron estandarizados de la siguiente manera:

- Los 10 años de pagos de beneficios fueron sumados para calcular los “Pagos acumulados de beneficios”

- El vector descriptor cargado en R es el vector de porcentajes donde el porcentaje para el “Año 1” es la fracción de los “Pagos acumulados de beneficios” que corresponde a 2009:

$$\frac{\text{Año 1}}{\text{Pagos Acumulados de beneficios}}$$

El vector de porcentajes fue usado para construir el mapa MDS (abajo).

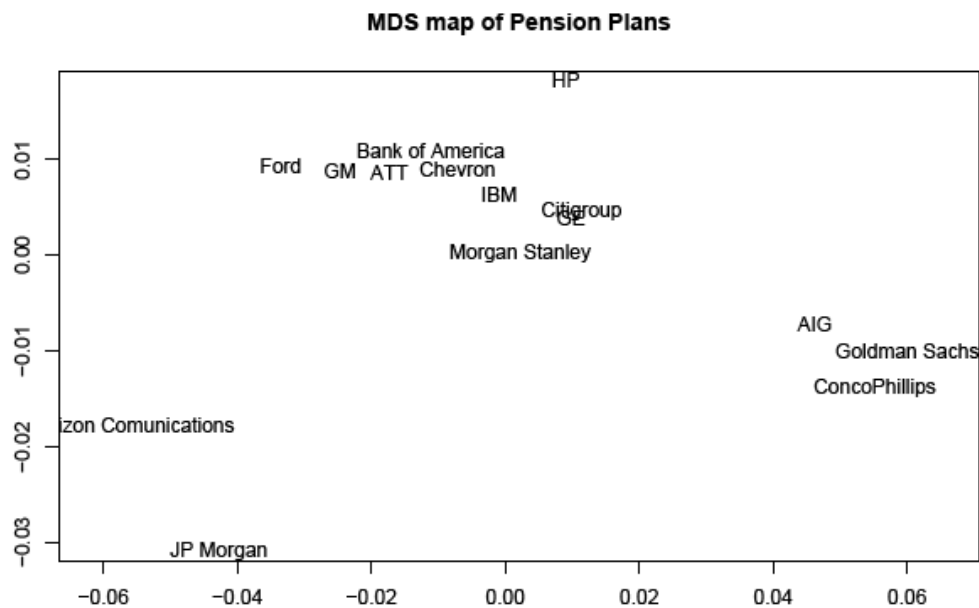


Figura 10. Mapa de las compañías consideradas en este ejemplo

La regla de inclusión en este caso es: TIIE resultado = TIIE calculada con la suma de los pagos de beneficios y la suma de los MVA.

La regla de capacidad es: TIIE resultado = Promedio ponderado de las TIIEs de los planes considerados. Los ponderadores son calculados con base en el MVA de los nodos en consideración.

Dado que quiero al menos 5 grupos de planes de pensiones con beneficios lo más parejo posible, el algoritmo es inicializado con 5 centros lo más cercano posible los unos a los otros. Tomando el nodo “Bank of America” aleatoriamente, los otros cuatro centros son: “Chevron”, “AT&T”, “IBM” y “Morgan Stanley” – numerados del 1 al 5 respectivamente.

El algoritmo resulta en los grupos siguientes:

	Group 1	Group 3	Group 2	Group 4	Group 5	Totals	
Companies	Bank of America	AT&T	Chevron	IBM	Morgan Stanley		
	Goldman Sachs	Citigroup	AIG	Verizon	Ford		
	JP Morgan	GM		HP	GE		
Market Value of Assets	\$ 28,971.00	\$ 187,720.00	\$ 11,047.00	\$ 107,163.00	\$ 107,745.00	AVG	STD DEV
IRR	6.10%	9.43%	6.61%	6.30%	5.99%	6.4957%	
Year 1	11.0608%	10.4681%	9.3720%	11.6057%	10.1869%	9.8373%	1.9630%
Year 2	11.2404%	10.3321%	9.4107%	10.8289%	10.1371%	9.6167%	1.6948%
Year 3	9.3070%	10.2935%	9.5755%	9.6177%	10.0884%	9.2647%	0.9998%
Year 4	9.6892%	10.2078%	9.7500%	9.7227%	10.0169%	9.5607%	0.7370%
Year 5	9.8016%	10.0498%	9.9438%	9.5845%	9.9606%	9.9273%	0.6864%
Year 6	9.6667%	9.9232%	10.2636%	9.8235%	9.9393%	10.0680%	0.5676%
Year 7	9.7847%	9.8304%	10.4284%	10.0337%	9.9317%	10.5100%	1.2584%
Year 8	9.7735%	9.7203%	10.4575%	9.8134%	9.9743%	10.2013%	0.9204%
Year 9	9.7173%	9.6016%	10.3993%	9.7616%	9.9971%	10.5447%	1.3175%
Year 10	9.9590%	9.5733%	10.3993%	9.5082%	9.7977%	10.4733%	1.3750%

Figura 11. Grupos de planes de pensiones

donde podemos ver que los pagos relativos de beneficios para todos los grupos están bastante cerca del promedio para los 10 años de pago.

Cabe mencionar que en este ejemplo, los grupos fueron construidos de acuerdo con las similitudes y diferencias entre los pagos de beneficios y sin tomar en cuenta la TII para la creación del mapa MDS. Por lo tanto, los grupos resultantes tienen TIIes que varían significativamente unas de las otras. Si se quisiera evitar esto, la TII debería ser considerada en el vector de descriptores para así incluir dicha información en el mapa MDS.

Dada la regla de capacidad, los grupos resultantes pueden ser interpretados como

los grupos de planes de pensiones con los pagos relativos de beneficios mas

balanceados posible y para los cuales, la TIE es menor que el promedio ponderado de las TIEs de los miembros del grupo³.

V. Conclusiones

Los algoritmos ilustrados aquí presentan formas de agrupar consumidores de bienes basadas en eficiencias de asignación de recursos. Yo creo que los algoritmos (o versiones más desarrolladas de los mismos) son suficientemente flexibles como para poder resolver una gran variedad de problemas. De hecho, cualquier problema de asignación de recursos en donde haya valor agregado derivado de agrupar a los consumidores potenciales, puede ser resuelto con los algoritmos presentados en este trabajo.

VI. Bibliografía

1. Dr. Nicolai Meinshausen (2009). Notes to accompany lectures in Statistical Data Mining. MSc in Applied Statistics lecture notes for Hilary Term 2009, Oxford, UK, Department of Statistics, Oxford University.
2. Chapter 8. Resource Allocation. Web page
<http://www.sciencedownload.net/demodownload/chapter8.pdf>
3. Oxford University Weblearn website. Web page -
<http://weblearn.ox.ac.uk/site/?campus>
4. Oxford University website. Web page - <http://www.ox.ac.uk>
5. Oxford University Gazette 2008-2009 admission statistics. Oxford
6. University Gazette web page - http://www.ox.ac.uk/gazette/2008-9/supps/2_vol139.pdf

³ El nodo para “Conco Phillips” no pudo ser integrado a ningún grupo dada la estructura de sus pagos de beneficios. De hecho no hay ningún grupo de 3 planes de pensiones que pueda incluir a “Conco Phillips” y a la vez obedecer la regla de capacidad – todas las combinaciones posibles fueron chequeadas a mano.

Evolución de la Eficiencia en el Sector Asegurador Mexicano

Act. Ricardo Erick Villarreal Azúa

Calle Hidalgo Mz 11 Lote 30 Col. Miguel Hidalgo, Tlalpan, D.F.

Tel. oficina 5621 1941.

erick.villarreal@gmail.com

Resumen

La Investigación de Operaciones es una rama de las Matemáticas que consiste en el uso de modelos y algoritmos para el apoyo en la toma de decisiones, su objetivo principal es optimizar o mejorar sistemas habitualmente reales.

Dentro de los modelos de Investigación de Operaciones existen los llamados Modelos de Frontera Eficiente que, como su nombre lo indica, permiten medir y comparar de manera objetiva la eficiencia para las distintas unidades analizadas.

En este trabajo se plantea la medición de la eficiencia comparando las compañías aseguradoras participantes en el mercado mexicano con dos objetivos principales:

-Evolución de la eficiencia del sector asegurador. Hipótesis principal del presente trabajo que consiste en medir la eficiencia del mercado de forma longitudinal mediante un panel de datos con el fin de evaluar si las compañías aseguradoras implementaron mejoras en sus procesos ante la apertura del mercado derivada del Tratado de Libre Comercio de América del Norte.

-Análisis entre compañías. Este análisis de carácter transversal identifica los factores que definen a las compañías más eficientes e identifica las áreas de oportunidad para que las empresas menos eficientes puedan mejorar su desempeño.

I. INTRODUCCIÓN

La apertura de los mercados internacionales demanda a las empresas enfrentar condiciones de competencia más estrictas y mejorar su desempeño para el cumplimiento de sus objetivos. La obtención de información integral para la toma de decisiones en la dirección es de suma importancia. En este sentido los métodos para la medición de la eficiencia son una herramienta que lo facilita.

Derivado del TLCAN se estableció la apertura paulatina del sector para implementar mejoras y enfrentar la competencia con las empresas estadounidenses y canadienses. Las fechas establecidas para la apertura en el sector se presentan a continuación:

Tabla I.1 Participación máxima de Capital Extranjero en el Sector Asegurador (TLCAN)

Año (1° Enero)	Participación (%)	Año (1° Enero)	Participación (%)
1990	10	1997	45
1994	30	1998	51
1995	35	1999	75
1996	40	2000	100

El presente análisis se enfoca en dicha apertura, ya que mediante un panel que cubre de 1998 a 2005 se evalúa la eficiencia del sector asegurador y su relación con la apertura del mercado.

II. MEDICIÓN DE LA EFICIENCIA

II. 1 Definición Eficiencia y Productividad

La eficiencia se define como “una situación en la que se minimiza el costo de producir una cantidad dada”. La eficiencia económica implica eficiencia tecnológica (producir la máxima cantidad de producto posible a partir de factores de producción dados).

II. 2 Métodos para medir la eficiencia.

Índices parciales de productividad.

Los índices parciales de productividad son cocientes entre el producto obtenido y la cantidad empleada de uno de los factores empleados en su producción.

Las razones financieras: Estas medidas son generalmente utilizadas por los administradores derivadas de los estados financieros. Entre ellas encontramos: las razones de rentabilidad que miden la eficiencia en la administración de los recursos a través de los rendimientos generados sobre las ventas, el activo y la inversión; las razones de liquidez o solvencia que evalúan la capacidad de la empresa para satisfacer sus obligaciones a corto plazo; las razones de apalancamiento o estabilidad que miden el grado en el cual la empresa ha sido financiada mediante endeudamiento, y las razones de actividad o productividad son aquellas que miden la eficacia con que son utilizados los recursos.

Indicadores Parciales de Productividad: Básicamente son una generalización de los indicadores anteriores y se utilizan para medir la productividad mediante la comparación del crecimiento del producto con el crecimiento de alguno de los insumos. Estos índices solo consideran uno de los factores utilizados, por lo que la mejora de un índice de productividad puede deberse a un proceso de sustitución de insumos y no a una mejora de su eficiencia.

Métodos de Frontera.

En esta metodología es necesario definir la frontera eficiente que se compone a partir de la combinación de procesos óptimos. La forma en que se mide la eficiencia es mediante la distancia entre la empresa y la frontera, aquellas

empresas que producen sobre la frontera son empresas eficientes mientras que las que no lo hacen tienen posibilidades de mejorar sus prácticas.

Métodos paramétricos: Estos métodos utilizan una función de producción o de costos para estimar los parámetros que permiten caracterizar la frontera eficiente. Se utilizan procesos estocásticos que permiten considerar la existencia de ruido aleatorio generado por errores de medición o por factores que están fuera de control de la firma. Los enfoques que existen son; Stochastic Frontier Approach (SFA) y Free Distribution Approach (FDA).

Métodos no paramétricos: En este caso la frontera se construye a partir de los datos observados y no a partir de una forma funcional específica. Así mismo, se utiliza la programación matemática para calcular la frontera eficiente. El modelo más conocido es el del Análisis de la Envolvente de Datos (DEA).

II. 3 El modelo de frontera y los tipos de orientación.

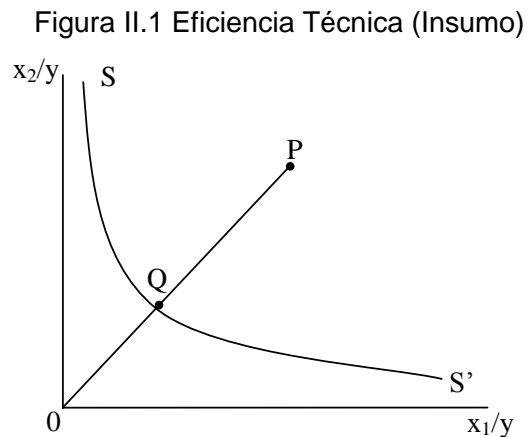
Orientación del insumo. La función describe el menor número de insumos necesario para alcanzar cierto nivel de producción.

Orientación de los resultados obtenidos. La función describe cuánto se pueden aumentar los resultados obtenidos sin modificar los insumos dados.

A continuación solo se presenta el detalle del modelo bajo orientación del insumo que es la utilizada en el presente trabajo. Si se requiere mayor detalle se recomienda consultar el trabajo de Tim Coelli "A guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program".

Orientación del insumo.

Para ilustrar esta idea, tomemos el siguiente ejemplo. Sean dos unidades de medición (DMU) que utilizan dos insumos (x_1, x_2) para obtener un solo resultado (y) , bajo el supuesto de rendimientos constantes a escala.



En la figura II.1 definimos SS' como la isocuanta unitaria de una firma completamente eficiente, para permitir la medición de la eficiencia técnica.

Si una firma usa ciertos insumos, definidos en el punto P , para producir una unidad, la ineficiencia técnica de la firma se puede representar mediante la distancia QP que es la cantidad en que los insumos pueden ser proporcionalmente reducidos para obtener el mismo nivel de producto.

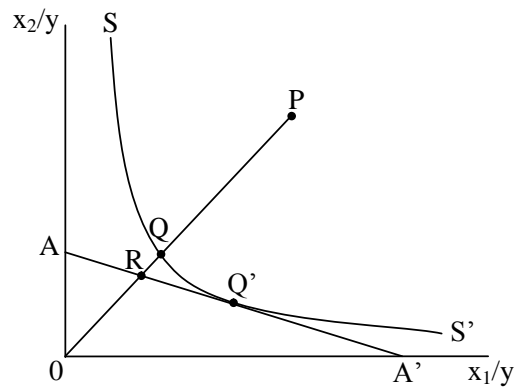
La forma de expresar esta reducción en forma de porcentaje mediante el cociente QP/OP se le llama *eficiencia técnica* (ET), que representa el porcentaje en que todos los insumos pueden ser reducidos.

$$ET_i = \frac{OQ}{OP} = 1 - \frac{QP}{OP}$$

La ET toma valores entre cero y uno, otorgando un indicador del nivel de ineficiencia técnica de la firma. El valor 1 indica que la firma es eficiente técnica.

Por ejemplo, el punto Q cumple con esta condición, ya que se encuentra en la isocuanta unitaria.

Figura II.2 Eficiencia Asignativa



Observando la figura II.2 la línea AA' representa el cociente entre los precios de los insumos, en este caso también se puede calcular la eficiencia asignativa. La *eficiencia asignativa* (EA) en una firma que opera en el punto P es definida por el cociente

$$EA_i = \frac{OR}{OQ}$$

La distancia RQ representa la reducción en costo de producción que se podría tener si la producción se realizara en el punto de eficiencia asignativa Q' , en lugar del punto de eficiencia técnica, pero ineficiencia asignativa Q .

La *eficiencia económica* (EE) total se define por el cociente

$$EE_i = \frac{OR}{OP}$$

Donde la distancia RP se puede interpretar en términos de reducción de costos.

Es importante observar que el valor de los tres indicadores de eficiencia toman valores entre cero y uno. Así mismo el producto de la eficiencia técnica y asignativa da como resultado el valor de la eficiencia económica.

$$ET_i * EA_i = \left(\frac{OQ}{OP}\right) * \left(\frac{OR}{OQ}\right) = \frac{OR}{OP} = EE_i$$

Esta medición de la eficiencia parte del supuesto que la función de producción de aquella firma completamente eficiente es conocida. En la práctica esta función no es conocida y la curva de eficiencia isocuanta debe ser estimada a partir de los valores observados. Farrell sugiere dos formas de estimar la función de producción, mediante una frontera lineal no paramétrica convexa construida de tal forma que ningún punto observado se encuentre a la izquierda o debajo de la misma, o una función paramétrica ajustada a los datos tal que ningún punto observado se localice a la izquierda o debajo de ella.

Orientación de los resultados obtenidos.

La medición de la eficiencia técnica medida mediante la orientación del insumo se enfoca en calcular cuánto se pueden reducir los insumos proporcionalmente sin cambiar el nivel de producción. Una forma alternativa de plantear el problema es responder cuánto se puede aumentar el nivel de producción sin modificar los insumos. Bajo este supuesto se plantea la medición de la eficiencia orientada a los resultados obtenidos.

Rendimientos de Escala.

Los rendimientos a escala expresan cómo varía la cantidad producida a medida que varían todos los factores que intervienen en la producción en la misma proporción.

Existen tres tipos de rendimientos a escala, los rendimientos constantes a escala, se presentan cuando variando en una proporción determinada la cantidad de factores utilizados, la cantidad producida varía en la misma proporción, es decir,

$kf(x) = f(kx)$. Los rendimientos crecientes a escala se presentan cuando multiplicando los factores de producción por una cantidad determinada se obtiene una cantidad producida mayor a k , $f(kx) > kf(x)$. Los rendimientos decrecientes a escala ocurren cuando aumentando todos los factores de producción en la misma proporción, la cantidad producida aumenta en una proporción menor, $f(kx) < kf(x)$.

Cuando se presentan rendimientos crecientes o decrecientes a escala se dice que se tienen rendimientos variables a escala.

III. METODOLOGÍA DEL ANÁLISIS DE LA ENVOLVENTE DE DATOS

El análisis de la envolvente de datos (DEA) es un modelo no paramétrico para calcular la estimación de la frontera. Este modelo supone rendimientos constantes o variables a escala, a continuación se detalla cada uno de ellos.

III. 1 Modelo de rendimientos a constantes escala (RCE) desde el punto de vista del Insumo.

Supongamos que se tienen:

n unidades de medición (DMU's Decision Making Unit)

k insumos

m resultados obtenidos

x_i los insumos para la i -ésima DMU

y_i los resultados obtenidos para la i -ésima DMU

X la matriz de tamaño $k \times n$ de los insumos para las n DMU

Y la matriz de tamaño $m \times n$ para los resultados para las n DMU

El propósito del DEA es construir una frontera envolvente no-paramétrica de tal forma que todos los valores observados caen sobre o por debajo de la frontera de

producción. El caso de la frontera donde se tiene un solo resultado y dos insumos, se puede interpretar como la intersección de los planos que forma la cubierta más ajustada para todos los puntos observados en el espacio tridimensional.

Supongamos que $x_{ig} \geq 0$ y que $y_{ih} \geq 0$, $i = 1, 2, \dots, n$; $g = 1, 2, \dots, k$; $h = 1, 2, \dots, k$; y que cada DMU utiliza al menos un insumo y genera al menos un resultado.

Se desea obtener para cada DMU una medición de eficiencia mediante el cociente de todos los resultados entre todos los insumos, de la forma $u' y_i / v' x_i$ donde, u es un vector de tamaño $m \times 1$ con los pesos de los resultados y v es un vector de tamaño $k \times 1$ con los pesos de los insumos.

La idea es obtener los pesos óptimos, el planteamiento del problema de programación es el siguiente:

$$\begin{aligned} & \max_{u,v} (u' y_i / v' x_i), \\ \text{s.a.} \quad & u' y_j / v' x_j \leq 1, j = 1, 2, \dots, n, \\ & u, v \geq 0. \end{aligned}$$

Esto involucra encontrar los valores de u y v tal que la medición de eficiencia de la i -ésima DMU es máxima, sujeto a la restricción que la medición de eficiencia debe ser menor o igual a uno. Un problema es que el planteamiento tiene un número infinito de soluciones ya que si (u^*, v^*) son solución entonces $(\alpha u^*, \alpha v^*)$, $\alpha > 0$ también será solución. Para evitar esta situación se puede realizar la siguiente transformación⁴ $v' x_i = 1$ que resulta en:

$$\begin{aligned} & \max_{\mu,\gamma} (\mu' y_i), \\ \text{s.a.} \quad & \gamma' x_j = 1, \end{aligned}$$

⁴ Transformación desarrollada por Charnes y Cooper que selecciona una solución representativa. (1962)

$$\mu' y_j - \gamma' x_j \leq 0, j = 1, 2, \dots, n.$$

$$\mu, \gamma \geq 0,$$

donde el cambio de variables de u a μ y de v a γ reflejan la transformación.⁵

Utilizando el programa dual, se deriva una forma envolvente equivalente para el problema.

$$\min_{\theta, \lambda} \theta,$$

$$\text{s.a.} \quad -y_i + Y\lambda \geq 0,$$

$$\theta x_i - X\lambda \geq 0,$$

$$\lambda \geq 0,$$

Donde θ es un escalar y λ es un vector de tamaño $n \times 1$ donde cada entrada es constante. Esta forma envolvente involucra menos restricciones que la forma anterior y es la que se utiliza generalmente.

El valor obtenido para θ será la calificación para la i -ésima DMU. Este valor satisface que $\theta \leq 1$, donde un valor de $\theta = 1$ indica un punto que pertenece a la frontera eficiente, sin embargo una DMU que está en la frontera puede ser eficiente o ineficiente.

III. 2 Modelo de rendimientos variables a escala (RVE) desde el punto de vista del Insumo.

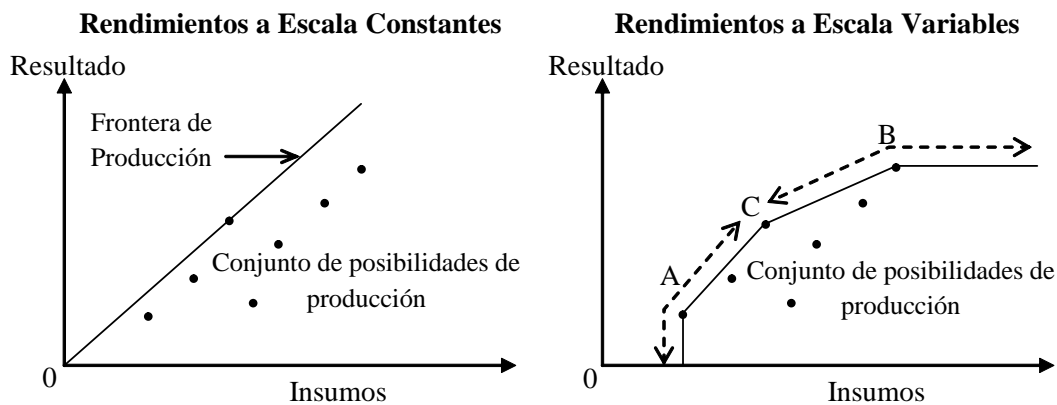
Los rendimientos variables a escala permiten abarcar las distintas formas de la función de posibilidades de producción. Generalmente se presentan cuando existen restricciones financieras o una competencia imperfecta. En el caso de que existieran condiciones y se realizara la medición bajo el supuesto de rendimientos

⁵ Esta forma es conocida como la forma “multiplicativa” del problema de programación lineal. Los pesos representados en las variables v y μ se conocen como la Normalización de los Precios.

constantes a escala se tendría que la medición de la *ET* se confundiría con la eficiencia de escala. La utilización de rendimientos constantes a escala permite la medición de la eficiencia técnica libre del efecto de la eficiencia de escala.

El modelo de rendimientos variables a escala fue propuesto por Banker, Charnes y Cooper (1984). La frontera de producción del modelo de rendimientos variables a escala se compone por la superficie convexa más cercana a las DMUs.

Figura III.1



Como se puede observar en la Figura III.1 las características que presenta la frontera del modelo de rendimientos variables a escala son, (A) rendimientos crecientes a escala en el primer segmento, (B) rendimientos decrecientes a escala en el segundo segmento y (C) rendimientos constantes a escala en el punto donde se realiza la transición del primer al segundo segmento. Al igual que en el modelo de rendimientos constantes a escala pueden existir deficiencias.

El modelo de rendimientos variables a escala se obtiene agregando la restricción de convexidad $\sum \lambda = 1$ para obtener:

$$\min_{\theta, \lambda} \theta,$$

$$\text{s.a.} \quad -y_i + Y\lambda \geq 0,$$

$$\theta x_i - X\lambda \geq 0,$$

$$N1' \lambda = 1$$

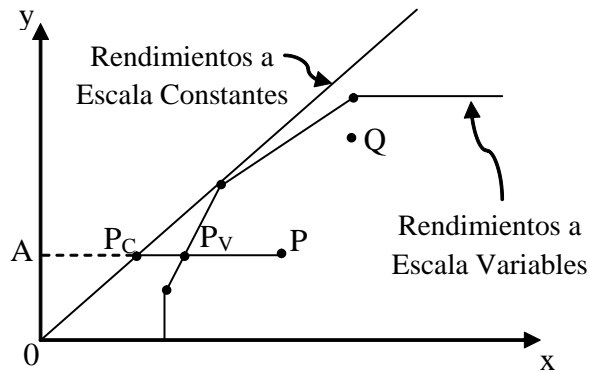
$$\lambda \geq 0,$$

donde $N1$ es un vector de unos de tamaño $N \times 1$. Esta restricción genera la estimación de la corteza convexa más estrecha generada a partir de los planos que envuelven las observaciones. El modelo de rendimientos variables a escala genera mediciones de la eficiencia técnica mayores o iguales a los obtenidos utilizando rendimientos constantes a escala.

La medición de la eficiencia de escala se realiza mediante la descomposición de la eficiencia técnica en dos componentes, la ineficiencia de escala y la eficiencia técnica "pura". Para obtener esta medición se requiere aplicar los modelos de rendimientos constantes a escala y rendimientos variables a escala sobre la mismas DMU's. Si existe diferencia entre las mediciones de la eficiencia técnica para una DMU implica que dicha DMU presenta ineficiencia de escala, la cual puede ser calculada como la diferencia entre la eficiencia técnica bajo rendimientos variables a escala y la eficiencia técnica bajo rendimientos constantes a escala.

$$ES = ET_v - ET_c$$

Figura III.2 Rendimientos a escala



En la Figura III.2 se observa el ejemplo donde se tiene un solo insumo y un solo producto y se grafican el caso de rendimientos constantes a escala y de rendimientos variables a escala. La ineficiencia técnica para el punto P bajo rendimientos constantes a escala es la distancia de PP_C mientras que bajo rendimientos variables a escala la ineficiencia técnica solo sería PP_V . La diferencia entre las dos eficiencias técnicas da como resultado la ineficiencia de escala. Estas medidas de la eficiencia también se pueden definir como cocientes delimitados entre cero y uno de la siguiente manera:

$$ET_{I,CRS} = \frac{AP_C}{AP}$$

$$ET_{I,VRS} = \frac{AP_V}{AP}$$

$$SE_I = \frac{AP_C}{AP_V}$$

Obsérvese que también se cumple

$$ET_{I,CRS} = ET_{I,VRS} * SE_I \Rightarrow AP_C/AP = (AP_V/AP) * (AP_C/AP_V)$$

Esta es la razón por la que la medición de la eficiencia técnica se descompone en eficiencia técnica “pura” y eficiencia de escala.

III. 3 Ventajas de la DEA.

Después de detallar los modelos de frontera y revisar la metodología del modelo DEA encontramos que existen varias ventajas respecto a las aproximaciones paramétricas. Algunas de ellas son que; no impone una forma particular de la frontera de producción, permite utilizar múltiples productos y múltiples insumos, emplea únicamente información empírica por lo que no necesita establecer ningún supuesto sobre la forma funcional de la frontera eficiente, no se requiere ningún supuesto acerca del error estocástico, es menos demandante en términos de grados de libertad que las aproximaciones paramétricas y finalmente permite evaluar la eficiencia para múltiples productos a la vez, lo que todavía no se puede realizar de manera satisfactoria desde una perspectiva paramétrica o mediante el uso de índices parciales de productividad.

IV. DESCRIPCIÓN DEL PANEL DE DATOS (1998-2005)

Para el análisis que se realiza para el sector asegurador en México, el panel que se utiliza se compone de 44 compañías que han operado de manera continua a lo largo del período comprendido entre 1998 y 2005. Según las cifras publicadas por la CNSF el segmento de mercado que cubren dichas compañías a diciembre de 2005 comprende el 92% del sector asegurador lo que permite cubrir un espectro importante del mercado. Se cuenta con la información de las 17 compañías más importantes, en su conjunto cubren el 85% del mercado (Tabla IV.1).

Tabla IV.1 Participación de Mercado en el Sector Asegurador según las primas emitidas a diciembre de 2005	
Nombre Compañía	Participación %
Grupo Nacional Provincial, S.A.	18.1
Metlife México, S.A.	15.4
Seguros Axa, S.A. de C.V.	13.36
Seguros Inbursa, S.A.	6.02
Seguros Monterrey New York Life, S.A. de C.V.	4.74
Seguros BBVA Bancomer, S.A. de C.V.	4.1
Quálitas, Compañía de Seguros, S.A.	3.16
Seguros Banorte Generali, S.A. de C.V.	2.91
Aba Seguros, S.A. de C.V.	2.46
Mapfre Tepeyac, S.A.	2.35
Seguros Banamex , S.A. de C.V.	2.32
Zurich, Compañía de Seguros, S.A.	2.14
Seguros Atlas, S.A.	1.96
Allianz México, S.A., Compañía de Seguros.	1.67
Ace Seguros, S.A.	1.59
AlG México, Seguros Interamericana, S.A. de C.V.	1.43
Aseguradora Interacciones, S.A.	1.21
General de Seguros, S.A.	0.69
Metropolitana, Compañía de Seguros, S.A.	0.65
Reaseguradora Patria, S.A.	0.62
Seguros Santander Serfin, S.A.	0.6
Seguros Afirme, S.A. de C.V.	0.53
Patrimonial Inbursa, S.A.	0.48
Grupo Mexicano de Seguros, S.A. de C.V.	0.46
GE Seguros, S.A. de C.V.	0.43
Zurich Vida, Compañía de Seguros, S.A.	0.35
Ana Compañía de Seguros, S.A. de C.V.	0.31
QBE del Istmo México, Cia. de Reaseguros, S.A. de C.V.	0.26
Agroasemex, S.A.	0.23
Tokio Marine Compañía de Seguros, S.A. De C.V.	0.23
Protección Agropecuaria, Compañía de Seguros, S.A.	0.21
La Latinoamericana Seguros, S.A.	0.2
Chubb de México, Compañía de Seguros, S.A. de C.V.	0.18
Seguros el Potosí, S.A.	0.14
Atradius Seguros de Crédito, S.A.	0.12
La Peninsular Seguros, S.A.	0.12
El Águila, Compañía de Seguros, S.A. de C.V.	0.09
Sompo Japan Insurance de Mexico, S.A. de C.V.	0.05
Royal & Sunalliance Seguros (México), S.A. de C.V.	0.03
Hir Compañía de Seguros, S.A. de C.V.	0.03
Principal México, Compañía de Seguros, S.A. de C.V.	0.02
Gerling de México Seguros, S.A.	0.02
Torreón, Sociedad Mutualista de Seguros	0.01
Skandia Vida, S.A. de C.V.	0.001

Los datos utilizados provienen de tres fuentes disponibles en la CNSF; el estado de resultados, el balance general (publicados trimestralmente) y la información correspondiente a la “Revista: Actualidad en Seguros y Fianzas” (publicada anualmente).

La condición para que una empresa entrara en el análisis es que su información no presentara faltantes en el período de interés para ninguna de las tres fuentes mencionadas anteriormente.

Para el análisis, las variables obtenidas del balance general y del estado de resultados se deflactan para obtener sus valores a precios constantes mediante el INPC, este índice es publicado por el Banco de México, la serie se obtiene de la página del Servicio de Administración Tributaria (SAT) con los valores indicados en la Tabla IV.2.

Tabla IV.2 Índice Nacional de Precios al Consumidor

Año	Índice Nacional de Precios al Consumidor											
	Ene	Feb	Mar	Abr	May	Jun	Jul	Ags	Spt	Oct	Nov	Dic
2005	112.554	112.929	113.438	113.842	113.556	113.447	113.891	114.027	114.484	114.765	115.591	116.301
2004	107.661	108.305	108.672	108.836	108.563	108.737	109.022	109.695	110.602	111.368	112.318	112.550
2003	103.320	103.607	104.261	104.439	104.102	104.188	104.339	104.652	105.275	105.661	106.538	106.996
2002	98.253	98.190	98.692	99.231	99.432	99.917	100.204	100.585	101.190	101.636	102.458	102.904
2001	93.765	93.703	94.297	94.772	94.990	95.215	94.967	95.530	96.419	96.855	97.220	97.354
2000	86.730	87.499	87.984	88.485	88.816	89.342	89.690	90.183	90.842	91.467	92.249	93.248
1999	78.119	79.169	79.904	80.637	81.122	81.655	82.195	82.658	83.456	83.985	84.732	85.581
1998	65.638	66.787	67.569	68.201	68.745	69.557	70.228	70.903	72.053	73.085	74.380	76.195

Las cifras correspondientes a enero de 1998 y meses posteriores han sido publicadas en el DOF. conforme a las base 1994=100. La serie fue obtenida de la página del SAT.

El Balance General es un caso particular ya que la información presentada corresponde a un intervalo de tiempo “Balance General del... al...” (en este caso trimestral), mientras las otras fuentes se presentan de forma anual. Para resolver esta situación se suman los 4 trimestres del año. Tomar la suma o el promedio anual no genera diferencia en los resultados ya que el modelo DEA es invariante ante transformaciones lineales en las variables utilizadas.

V. SELECCIÓN DE VARIABLES

Uno de los principales retos al medir la eficiencia de las aseguradoras mediante la DEA, y en general de las empresas del sector financiero, radica en definir los insumos y los productos que se utilizan para el análisis. Esto se debe a que, la

mayoría de las veces, tanto los insumos como los servicios que proveen son intangibles y más aún, en algunas ocasiones, difíciles de medir.

V. 1 Clasificación de las variables en el sector asegurador.

Los insumos los podemos clasificar de manera general en tres grandes grupos, variables de costo laboral y operación, reservas y capital.

Para el caso de las variables de salida o productos nos enfrentamos a la dificultad de encontrar una unidad de medición homogénea, esto es consecuencia de la naturaleza intangible de los servicios financieros. Para el presente estudio se parte del hecho que las compañías están preparadas para ofrecer bienes y servicios y que la venta de los mismos permite en consecuencia generar las utilidades. Para ello se utilizó como variable las primas emitidas.

V. 2 Selección y definición de variables.

Según Cooper, Seiford y Tone (2000) *ceteris paribus* es preferible insumos pequeños y productos grandes, ya que esta relación se refleja en las calificaciones que arroja el modelo DEA.

Para la selección de variables se tomaron varios elementos en consideración, en primer lugar se seleccionaron variables consistentes con la definición de insumos y productos, además se buscaron variables cuyo valor no fuera demasiado volátil para permitir un análisis del panel de datos más estable a lo largo del tiempo, que la definición sea intuitiva y que hayan sido utilizadas en estudios similares.

Del universo de variables disponibles y en función de los elementos de selección, se utilizaron las siguientes variables:

Tabla V.1 Variable Definitivas

Fuente	Variable	Tipo
Estado de Resultados	Primas Emitidas	Resultado
Balance General	Pvo. Reservas Técnicas	Insumos
Balance General	Capital	Insumos
Actualidad en Seguros y Fianzas	Empleados	Insumos

Primas Emitidas: se define como la suma de las primas directas más las primas tomadas, donde las primas directas son el monto total de primas netas correspondientes a pólizas y endosos expedidos a los Asegurados y las Primas tomadas son las primas generadas por las responsabilidades que una compañía acepta como reaseguro de otras aseguradoras o reaseguradoras.

Reservas Técnicas: las reservas técnicas representan las provisiones requeridas por la regulación para hacer frente a los riesgos y a las obligaciones futuras con los asegurados.

Capital: Indica la inversión de los accionistas en la empresa, reflejada en el patrimonio.

Empleados: se compone del reporte de la fuerza productiva y el personal ocupado según la circular S-20.3; la fuerza productiva son los agentes ya sean personas físicas o morales, y el personal ocupado se compone de los funcionarios y de los empleados.

Finalmente se valida la dimensionalidad del modelo, entendida como la diferencia entre las variables y el de unidades, ya que si esta dimensionalidad no es suficiente el modelo de DEA calificará como eficientes unidades que realmente no lo son. Como regla general Charnes y Cooper (1990) sugieren que el número de unidades sea al menos tres veces el número de insumos y productos combinados, es decir, $n \geq 3(p+q)$. En el modelo se utilizaron 44 unidades (compañías) y 4 variables por lo que se cumple la condición de dimensionalidad sugerida.

VI. DESARROLLO Y ANALISIS DE LOS RESULTADOS

VI. 1 Planteamiento de la Hipótesis.

La hipótesis que plantea el presente trabajo es que, a partir del anuncio de la apertura del mercado asegurador y el establecimiento de las fechas para ello, las compañías tomaron medidas para incrementar su eficiencia y productividad, y así poder competir en el mercado internacional.

Para validar la hipótesis de la presente tesis se realiza la prueba t para dos muestras pareadas comparando el promedio del año previo respecto al del año corriente. Para cada par de años se plantea la siguiente hipótesis:

Ho: El promedio de las calificaciones de un año a otro es igual.

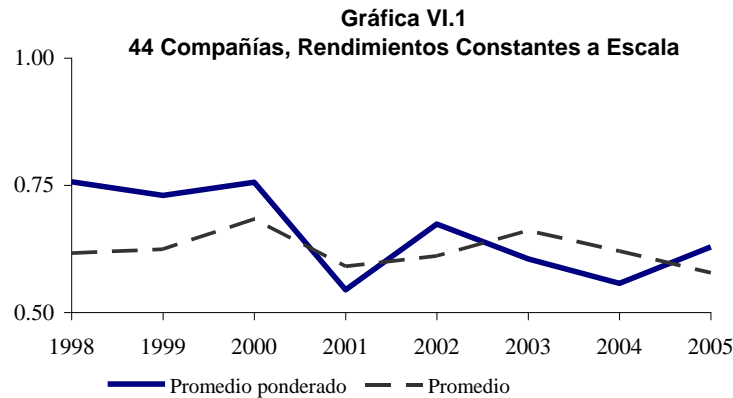
Ha: El promedio de las calificaciones de un año a otro es distinto.

En caso de que la hipótesis nula no sea rechazada estaríamos ante el escenario de que la eficiencia en las aseguradoras no ha incrementado, en caso contrario se debe plantear si el promedio de las calificaciones aumenta o disminuye ya que en función de esto la eficiencia del mercado estará incrementando o disminuyendo.

VI. 2 Resultados.

Para realizar el análisis de la evolución de las calificaciones otorgadas por el DEA a lo largo del período de 1998 a 2005 para 44 compañías con lo que se cubre el 92% del mercado al cierre de 2005. Se obtienen los resultados bajo la orientación del insumo y se contrastan los resultados bajo rendimientos constantes y rendimientos variables a escala.

VI. 3 Rendimientos constantes a escala.



En el caso de RCE para las 44 compañías observamos que el promedio ponderado es mayor que el promedio durante los tres primeros años mientras que en el resto del panel los valores se van intercalando, exceptuando 2004, lo que indica que no existe una diferencia clara entre la calificación obtenida por las compañías con mayor participación y aquellas con menor calificación.

En la tabla VI.2 podemos observar los resultados de las calificaciones para el total de las compañías. La interpretación práctica de la calificación otorgada por la DEA es mediante la reducción proporcional de todos los insumos tal que la compañía forme parte de la frontera eficiente. Por ejemplo, si tomamos a GNP en el primer año del análisis observamos una calificación de 0.86 y que está en posibilidades de reducir sus insumos sin modificar las cantidades producidas. Para formar parte de la frontera tendría que modificar sus insumos mediante la reducción de un 14% (100% - 86%) en todos sus insumos tal que forme parte de la frontera definida por el mercado.

Tabla VI.2 Rend. Const. 44 compañías.

Nombre Compañía	Participación de Mercado	Rendimientos Constantes a Escala							
		1998	1999	2000	2001	2002	2003	2004	2005
GNP	18.10	0.86	0.834	0.757	0.474	0.665	0.687	0.65	1
Metlife	15.40	1	1	1	0.607	1	0.444	0.378	0.4
Axa	13.36	0.851	0.765	0.784	0.533	0.568	0.428	0.415	0.479
Inbursa	6.02	0.3	0.287	0.346	0.321	0.588	0.446	0.372	0.367
Mty NYL	4.74	0.425	0.438	0.379	0.309	0.299	0.835	0.56	0.729
BBVA Bancomer	4.10	1	1	1	1	1	1	1	1
Quálitas	3.16	0.599	0.473	0.605	0.601	0.608	0.623	0.6	0.583
Banorte	2.91	0.661	0.713	0.784	0.571	0.547	0.772	0.445	0.591
ABA Seg.	2.46	0.887	0.598	0.579	0.442	0.499	0.541	0.452	0.492
Mapfre	2.35	0.563	0.439	0.485	0.365	0.408	0.463	0.376	0.369
Banamex	2.32	0.928	1	1	0.6	0.577	0.761	0.942	0.277
Zurich	2.14	1	0.67	1	0.753	0.839	0.824	0.905	0.904
Atlas	1.96	0.461	0.52	0.639	0.495	0.541	0.438	0.501	0.484
Allianz	1.67	0.528	0.665	0.814	0.512	0.588	0.795	0.832	0.492
ACE	1.59	0.676	0.831	1	1	1	1	1	1
AIG	1.43	0.445	0.45	0.687	0.619	0.521	0.527	0.519	0.691
Interacciones	1.21	0.549	0.489	0.815	0.813	1	0.936	0.636	0.787
General de Seg.	0.69	0.488	0.394	0.482	0.357	0.435	0.409	0.386	0.249
Metropolitana	0.65	0.423	0.498	0.793	0.489	0.469	0.452	0.437	0.45
Patria	0.62	0.452	0.565	0.731	0.737	0.477	0.752	0.564	0.313
Santander	0.60	0.534	0.678	0.594	0.116	0.128	0.701	0.528	0.437
Afirme	0.53	0.485	0.698	0.906	0.783	0.426	0.679	0.96	1
Patrim. Inbursa	0.48	0.278	0.421	0.191	0	0	0	1	1
Gpo. Mex. de Seguros	0.46	0.395	0.774	0.93	0.616	0.506	0.814	0.646	0.585
GE	0.43	0.606	0.452	0.586	0.573	0.568	0.533	0.537	0.471
Zurich Vida	0.35	0.466	0.699	0.931	0.893	1	0.483	0.44	0.665
ANA	0.31	0.51	0.414	0.561	0.684	0.532	0.811	0.894	0.767
QBE del Itsmo	0.26	1	0.935	0.331	1	1	1	1	1
Agroasemex	0.23	0.641	0.396	0.56	0.267	0.285	0.288	0.327	0.201
Tokio Marine	0.23	0.506	0.582	0.782	0.661	0.669	0.381	0.417	0.538
Protección Agropecuaria	0.21	1	1	1	0.527	0.994	0.976	1	0.91
La Latinoamericana	0.20	0.274	0.28	0.337	0.335	0.36	0.547	0.376	0.472
Chubb de México	0.18	0.519	0.501	0.58	0.613	0.706	0.659	0.631	0.477
El Potosí	0.14	0.441	0.428	0.531	0.533	0.505	0.535	0.519	0.431
Atradius	0.12	0.294	0.527	0.658	0.671	0.671	0.794	0.537	0.299
La Peninsular	0.12	0.47	0.405	0.558	0.765	0.464	0.539	0.588	0.133
El Águila	0.09	0.798	0.58	0.702	0.632	0.638	0.669	0.58	0.441
Sompo Japan	0.05	0.78	1	1	1	0.766	0.863	0.612	0.4
Royal & Sunalliance	0.03	1	0.986	0.859	0.734	0.666	1	0.818	0.888
HIR	0.03	1	0.736	0.607	0.321	0.856	1	0.589	0.627
Principal	0.02	0.108	0.611	0.437	1	0.145	0.151	0.088	0.033
Gerling	0.02	0.458	0.573	0.536	0.444	0.477	0.539	0.36	0.264
Mut. Torreón	0.01	0.464	0.169	0.201	0.228	0.894	1	0.899	0.732
Skandia Vida	0.00	1	1	1	1	1	1	1	1
Promedio ponderado		0.76	0.73	0.76	0.54	0.67	0.61	0.56	0.63
Promedio		0.62	0.62	0.68	0.59	0.61	0.66	0.62	0.58

En estos resultados resalta el caso de MetLife que aparentemente es una empresa eficiente de 1998 a 2001 y a partir del 2003 se vuelve una empresa con

importantes posibilidades de mejorar su eficiencia. Profundizando en la información observamos durante el segundo trimestre de 2003 la compañía estadounidense MetLife gana la licitación de la empresa estatal Aseguradora Hidalgo (Ahisa). Como resultado de esta fusión al cierre de 2003 MetLife México se convierte en la segunda empresa con mayor participación de mercado 16.7%, solo superada por Seguros Axa. En este contexto podemos interpretar que después de la fusión MetLife se vuelve una empresa ineficiente posiblemente derivado de las formas administrativas y organizacionales heredadas por Ahisa.

Tabla VI.3 Cambios de Metlife

	MetLife			
	Primas Emitidas*	Reservas Técnicas	Capital*	Empleados
Año 2002	4,894	15,370	1,423	1,445
Año 2003	14,047	89,975	15,842	4,588
Diferencia	9,152	74,605	14,419	3,143
Cambio %	187%	485%	1013%	218%
Año 2004	21,709	127,800	28,788	4,700
Año 2005	21,698	137,879	33,582	4,700

*Cifras en Millones de MXP

Analizando a detalle los datos de Metlife (tabla VI.3) observa que la principal razón para el decremento en su eficiencia responde a que a pesar de tener un incremento importante en el volumen de primas de 2002 a 2003 (187%) el incremento en los insumos es mucho mayor en término porcentuales que el de las primas emitidas, donde sobresale el incremento del 1013% del capital, lo que en términos del modelo se traduce en una menor eficiencia.

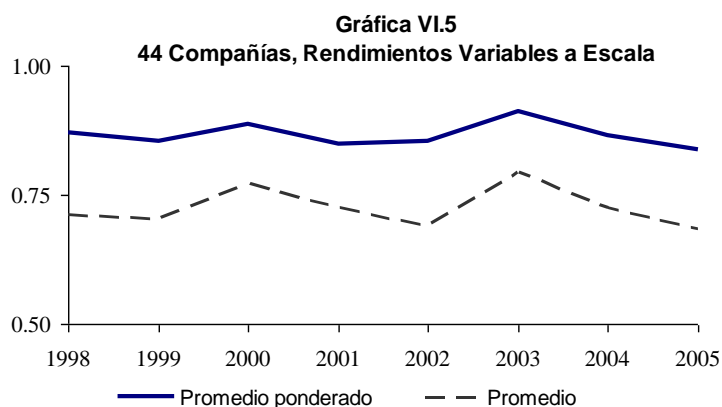
Para mostrar formalmente si el promedio de la eficiencia en el sector asegurador mexicano ha incrementado a raíz de la apertura del mercado, se presentan los resultados de la Prueba-t para la media de dos muestras dependientes.

Tabla VI.4 Prueba de Hipótesis Rend. Const. 44 Cias.

Prueba-t: Para la Media de dos Muestras Dependientes.	Rendimientos Constantes a Escala (44 Compañías)							
	1998-1999	1999-2000	2000-2001	2001-2002	2002-2003	2003-2004	2004-2005	1998-2005
Media primer año	0.616	0.624	0.683	0.591	0.611	0.661	0.621	0.616
Varianza primer año	0.061	0.051	0.053	0.061	0.065	0.060	0.056	0.061
Observaciones primer año	44	44	44	44	44	44	44	44
Media segundo año	0.624	0.683	0.591	0.611	0.661	0.621	0.578	0.578
Varianza segundo año	0.051	0.053	0.061	0.065	0.060	0.056	0.072	0.072
Observaciones segundo año	44	44	44	44	44	44	44	44
Correlación de Pearson	0.738	0.747	0.575	0.555	0.650	0.625	0.764	0.419
Hipótesis de la diferencia de medias	0	0	0	0	0	0	0	0
Grados de Libertad	43	43	43	43	43	43	43	43
Estadística t	-0.307	-2.394	2.778	-0.569	-1.592	1.284	1.619	0.918
P(T<=t) dos-colas	0.761	0.021	0.008	0.573	0.119	0.206	0.113	0.364
t Crítica de dos-colas	2.017	2.017	2.017	2.017	2.017	2.017	2.017	2.017
Ho: Medias son iguales.								
Prueba Ho: dos colas	No rechaza	No rechaza	No acepta	No rechaza	No rechaza	No rechaza	No rechaza	No rechaza

Los resultados de las pruebas de hipótesis indican que solo entre los años 2000 y 2001 no aceptamos la hipótesis nula, entre dichos años la media de la eficiencia del panel disminuye. Si tomamos la prueba entre el año inicial y el año final observamos que la hipótesis no se rechaza por lo que no existe evidencia suficiente para suponer que la media de la calificación del panel es distinta entre el año inicial y el año final y por tanto que en promedio el mercado ha disminuido o aumentado su eficiencia.

VI. 4 Rendimientos variables a escala.



Para los RVE observamos que el promedio y el promedio ponderado se mueven en línea a lo largo del panel y no presentan cruzamientos como lo hace bajo RCE.

De manera gráfica no se observa alguna tendencia clara (creciente o decreciente) en el promedio de calificaciones.

Tabla VI.6 Rend. Var. 44 compañías.

Nombre Compañía	Participación de Mercado	Rendimientos Variables a Escala							
		1998	1999	2000	2001	2002	2003	2004	2005
GNP	18.10	1	1	1	1	1	1	1	1
Metlife	15.40	1	1	1	1	1	0.979	1	1
Axa	13.36	1	1	1	1	1	1	1	0.837
Inbursa	6.02	0.535	0.621	0.676	0.645	1	0.696	0.594	0.546
Mty NYL	4.74	0.684	0.713	0.538	0.538	0.386	1	0.668	0.729
BBVA Bancomer	4.10	1	1	1	1	1	1	1	1
Quálitas	3.16	0.7	0.584	0.75	0.676	0.627	0.87	0.855	0.971
Banorte	2.91	0.807	0.776	0.91	0.576	0.549	0.774	0.456	0.81
ABA Seg.	2.46	1	0.687	0.642	0.591	0.546	0.722	0.596	0.635
Mapfre	2.35	0.897	0.773	0.729	0.54	0.487	0.682	0.507	0.552
Banamex	2.32	1	1	1	0.787	0.715	1	1	0.401
Zurich	2.14	1	0.675	1	1	1	1	1	1
Atlas	1.96	0.554	0.558	0.71	0.599	0.57	0.55	0.618	0.66
Allianz	1.67	0.563	0.667	0.822	0.538	0.62	1	1	0.519
ACE	1.59	0.678	0.839	1	1	1	1	1	1
AIG	1.43	0.644	0.528	0.784	0.779	0.538	0.578	0.578	0.837
Interacciones	1.21	0.667	0.505	0.829	0.978	1	0.936	0.637	0.91
General de Seg.	0.69	0.58	0.478	0.63	0.357	0.436	0.507	0.49	0.298
Metropolitana	0.65	0.437	0.499	0.819	0.489	0.483	0.581	0.474	0.45
Patria	0.62	0.518	0.572	0.736	0.743	0.705	0.761	0.574	0.323
Santander	0.60	0.554	0.686	0.595	0.133	0.129	0.715	0.612	0.447
Afirme	0.53	0.55	0.718	0.923	0.795	0.431	0.808	0.967	1
Patrim. Inbursa	0.48	0.307	0.424	0.199	1	1	1	1	1
Gpo. Mex. de Seguros	0.46	0.932	0.852	1	0.617	0.509	0.903	0.689	0.586
GE	0.43	0.639	0.476	0.647	0.582	0.57	0.67	0.587	0.474
Zurich Vida	0.35	0.671	0.7	1	0.901	1	0.58	0.442	0.669
ANA	0.31	0.521	0.449	0.58	0.713	0.54	0.959	0.951	0.783
QBE del Itsmo	0.26	1	1	0.45	1	1	1	1	1
Agroasemex	0.23	0.771	0.445	0.67	0.27	0.309	0.335	0.37	0.204
Tokio Marine	0.23	0.64	0.62	0.788	0.663	0.683	0.41	0.419	0.546
Protección Agropecuaria	0.21	1	1	1	0.571	1	1	1	0.949
La Latinoamericana	0.20	0.298	0.285	0.369	0.337	0.362	0.626	0.379	0.48
Chubb de México	0.18	0.521	0.529	0.603	0.631	0.708	0.77	0.637	0.497
El Potosí	0.14	0.458	0.434	0.534	0.546	0.514	0.553	0.527	0.468
Atradius	0.12	0.336	0.54	0.708	0.686	0.676	0.978	0.667	0.355
La Peninsular	0.12	0.486	0.417	0.563	0.768	0.467	0.577	0.596	0.161
El Águila	0.09	0.801	0.621	0.752	0.675	0.65	0.72	0.588	0.505
Sompo Japan	0.05	0.985	1	1	1	0.784	1	0.645	0.497
Royal & Sunalliance	0.03	1	1	1	0.735	0.667	1	0.818	0.939
HIR	0.03	1	1	1	0.988	1	1	1	1
Principal	0.02	0.119	0.611	0.453	1	0.156	0.153	0.092	0.038
Gerling	0.02	0.463	0.643	0.583	0.52	0.509	0.56	0.877	1
Mut. Torreón	0.01	1	1	1	1	1	1	1	1
Skandia Vida	0.00	1	1	1	1	1	1	1	1
Promedio ponderado		0.87	0.85	0.89	0.85	0.85	0.91	0.86	0.84
Promedio		0.71	0.70	0.77	0.73	0.69	0.79	0.73	0.68

Si observamos las calificaciones obtenidas bajo RVE las tres empresas con mayor participación de mercado forman parte de la frontera eficiente en la mayoría de los años. Dentro de estas empresas se encuentra MetLife que en el análisis de RCE concluimos que su eficiencia disminuye luego de la adquisición de Ahisa en 2003, en este caso la conclusión es distinta ya que solo en el 2003 la eficiencia de Metlife disminuye y a partir del año 2004 vuelve a ser eficiente. La explicación de esta situación es que Metlife forma parte de las compañías con rendimientos no constantes, en específico con rendimientos decrecientes de escala lo que genera que bajo RVE sea una empresa eficiente. El análisis se hace bajo las dos orientaciones, dado que no sabemos cual forma funcional tiene la frontera de posibilidades de producción del mercado asegurador en México.

Las compañías bancarias que mejor aprovechan su posición en términos de eficiencia dentro del mercado bancario son Bancomer y Banamex. En el caso de Bancomer es una empresa eficiente en la totalidad del panel, esto responde a que Bancomer reporta a la CNSF que tiene un solo empleado trabajando, esto derivado de su estrategia de nómina y de las ventajas que le da que los empleados del sistema bancario puedan ofrecer los servicios de seguros en la misma sucursal que se ofrecen los servicios financieros. En el caso de Banamex es una empresa mucho más eficiente bajo RVE que bajo RCE ya que para el primer caso es eficiente en 5 de los 8 años del análisis solo siendo ineficiente en 2001, 2002 y 2005 (Tabla VI.6) mientras que en el caso de RCE solo es eficiente durante 1999 y 2000. Estas dos empresas son líderes en el mercado bancario y en consecuencia esta posición estratégica se ve reflejada en su eficiencia dentro del sector asegurador.

El resto de los bancos que ofrecen seguros (Inbursa, Banorte y Santander) no resultan ser empresas tan eficientes dentro del sector asegurador. Banorte es la empresa con mejor calificación promedio con 71% para RVE seguida de Inbursa con 66% y por Santander con 48%. Sin embargo la única empresa que en alguno de los años forma parte de la frontera es Inbursa en el 2002. El caso de HSBC no se analiza ya que al momento de adquirir Bital y darse la fusión del área de Seguros, Bital es dividida en Vida HSBC y Daños HSBC por lo que la información anual deja de ser comparable y se excluye del panel.

Tabla VI.7 Prueba de Hipótesis Rend. Var. 44 Cias.

Prueba-t: Para la Media de dos Muestras Dependientes.	Rendimientos Variables a Escala (44 Compañías)							
	1998-1999	1999-2000	2000-2001	2001-2002	2002-2003	2003-2004	2004-2005	1998-2005
Media primer año	0.712	0.703	0.773	0.727	0.689	0.794	0.725	0.712
Varianza primer año	0.061	0.047	0.044	0.053	0.067	0.049	0.058	0.061
Observaciones primer año	44	44	44	44	44	44	44	44
Media segundo año	0.703	0.773	0.727	0.689	0.794	0.725	0.684	0.684
Varianza segundo año	0.047	0.044	0.053	0.067	0.049	0.058	0.078	0.078
Observaciones segundo año	44	44	44	44	44	44	44	44
Correlación de Pearson	0.814	0.711	0.396	0.733	0.641	0.837	0.754	0.500
Hipótesis de la diferencia de medias	0	0	0	0	0	0	0	0
Grados de Libertad	43	43	43	43	43	43	43	43
Estadística t	0.409	-2.853	1.255	1.367	-3.379	3.438	1.485	0.706
P(T<=t) dos-colas	0.684	0.007	0.216	0.179	0.002	0.001	0.145	0.484
t Crítica de dos-colas	2.017	2.017	2.017	2.017	2.017	2.017	2.017	2.017
Ho: Medias son iguales.								
Prueba Ho: dos colas	No rechaza	No rechaza	No rechaza	No rechaza	No rechaza	No acepta	No rechaza	No rechaza

Los resultados obtenidos para las pruebas de hipótesis en el caso de RVE para las 44 compañías indican que sólo entre los años 2003 y 2004 no aceptamos la prueba de hipótesis, entre dichos años el promedio de la eficiencia disminuye 6.9 puntos porcentuales. En el resto de los años del panel no hay evidencia suficiente para concluir que la media es distinta. Asimismo en la prueba donde se compara el año inicial y el año final no rechazamos la prueba de hipótesis por lo que no podemos concluir que la media de la eficiencia sea diferente y que el mercado sea más o menos eficiente derivado de la apertura del mercado asegurador en México.

La dinámica del sector asegurador en México ha cambiado de manera importante en los últimos años luego de la apertura del mercado y la inversión de capital extranjero, las exigencias al interior de las empresas de generar utilidades acordes a las expectativas de los inversionistas y la competencia con el resto del mercado exige a las empresas mejoras continuas en su operación, es en este sentido el análisis mediante la DEA permite evaluar desde una nueva perspectiva la eficiencia entre las compañías aseguradoras.

VII. Conclusiones

La apertura comercial con América del Norte derivada del TLC deja atrás las políticas proteccionistas del gobierno mexicano sobre el mercado y permite al capital extranjero invertir en compañías mexicanas o establecerse en el país. Las compañías aseguradoras con mejor eficiencia serán las que obtengan mayores beneficios y oportunidades generadas por la liberación financiera y la integración al mercado internacional.

Entre 1990 y 2000, años entre los que se establece la apertura total del mercado, la participación extranjera pasa de un mercado cerrado a un mercado donde 29 de los 70 participantes eran filiales de instituciones del exterior.⁶ Para el cierre de 2005 de las 89 compañías en el mercado 39 presentaban capital mayoritariamente extranjero.⁷

El Análisis de la Envolvente de Datos permite analizar la eficiencia de las empresas utilizando múltiples variables a la vez y es una alternativa a la utilización de los índices parciales de productividad y a los métodos de frontera paramétricos. Para que una empresa formara parte del panel que va de 1998 a 2005 se requiere

⁶ Anuario estadístico de seguros y fianzas 2000, CNSF

⁷ Anuario estadístico de seguros y fianzas 2005, CNSF

que la empresa esté presente en el total de los años, es decir, que no desapareciera o fuera de creación posterior a 1998, y que no presenta faltantes de información en el periodo para ninguna de las variables utilizadas. En el caso de fusiones o adquisiciones la información que prevalece antes de la fusión / adquisición es la de las compañías que la realiza sin importar cual de ellas es más grande.

Una vez definido el panel de datos, se selecciona la orientación del insumo dado que las compañías tienen mayor control sobre los insumos que sobre los resultados, por ejemplo, si todas las compañías utilizan los mismos insumos entre un año y otro no necesariamente el volumen de primas emitidas será la misma, en este sentido podemos pensar que existen condiciones del mercado que influyen sobre el volumen de primas emitidas.

Dado que se desconoce la forma en que cambia el volumen producido a medida que varían en la misma proporción todos los insumos de la producción, el análisis del panel se realiza desde rendimientos constantes y rendimientos variables a escala. En el primer caso se supone que los factores utilizados y el volumen producido varían en la misma proporción mientras que en el caso de rendimientos variables no sucede así.

Las primas emitidas son utilizadas como variable de resultado (producto) mientras que las reservas técnicas, los empleados y el capital se utilizan como variables de entrada (insumos). Lo que se busca es producir la misma cantidad de producto reduciendo proporcionalmente todos los insumos según la calificación otorgada por la DEA.

La hipótesis que se plantea originalmente es que, derivado de la apertura del sector asegurador por el TLCAN, la eficiencia de las empresas debe incrementar ante la necesidad de enfrentar las nuevas condiciones de competencia. En este caso la prueba se hace sobre el promedio de las calificaciones de un año al otro e igualmente se contrastan los resultados entre el año inicial y final.

Los resultados obtenidos según las pruebas de hipótesis no permiten concluir que entre 1998 y 2005 la eficiencia promedio del mercado aumente.

La idea de realizar la prueba de hipótesis entre el año inicial y final y no solo entre años consecutivos radica en tomar en consideración que la implementación de la eficiencia no necesariamente se realiza en un periodo breve de tiempo y que es un proceso constante y de mejora continua. Otra posibilidad que se cubre con el análisis entre al año inicial y final es que una vez que las empresas son adquiridas se les permite continuar laborando de manera regular para generar el menor impacto posible y no es hasta años después de adquirida la empresa cuando se definen las mejoras.

Para el presente estudio no se cuenta con el total de los años que corresponden al periodo de apertura del mercado que va de 1994 a 2000 por lo que no podemos concluir que no se tomaron medidas para la mejora de la eficiencia en el sistema asegurador ante la apertura del mercado, ya que existe la posibilidad de que se hubieran realizado antes de 1998. Lo que es un hecho es que a partir de la apertura total del sector asegurador en el año 2000 la eficiencia de las empresas no ha cambiado de manera clara hasta 2005.

Adicional a la conclusión principal del presente trabajo podemos obtener otros resultados importantes. La mejora en la eficiencia de las empresas permite un

mejor rendimiento en el sector asegurador, el caso de Metlife-Ahisa muestra que las empresas en el mercado que no mantienen un buen desempeño tenderán a fusionarse mientras que las empresas que las adquieren consideran que pueden mejorar su participación en el mercado sin disminuir su eficiencia. Al final las dos compañías deberán verse beneficiadas por la fusión en términos de su eficiencia.

Otro resultado implícito es que las empresas más eficientes no son necesariamente las más grandes, la razón de que exista en México un mercado tan disperso es que las empresas pequeñas son capaces de enfocarse en un nicho de mercado que les otorga una ventaja competitiva respecto a sus competidores. Es decir, vemos que las empresas se vuelven especialista en un sector, situación que vemos en otros intermediarios financieros como los bancos o las casas de bolsa.

El modelo de la DEA nos permite explorar nuevas alternativas antes los resultados obtenidos, ya sea agregando nuevas variables al análisis o modificando la perspectiva de análisis. Otra alternativa es considerar variables cualitativas que nos permitan evaluar la calidad en el servicio y la satisfacción del cliente para lo cual también existen variaciones de la DEA.

Adicionalmente la DEA otorga la posibilidad de enfocarse en la eficiencia al interior de las compañías según los intereses de cada una mediante la definición de sus insumos y sus productos o hasta la evaluación del desempeño de sus empleados como puede ser la fuerza de venta o los ajustadores, por lo que el presente trabajo queda abierto a análisis adicionales.

VIII. Bibliografía

1. Alvarado Rodrigo. "Eficiencia productiva en Municipios de Chile en la prestación de servicios de atención primaria: A dos décadas de la descentralización" Universidad Autónoma de Barcelona, 2005.
2. Álvarez Alejandro y Martínez Sandra. "Significados del Tratado de Libre Comercio de América del Norte para México". Gobierno de España, Información Comercial Española, 2001.
3. Cheng Yuk-shing. "Productivity Growth, Technical Progress and Efficiency Change in Chinese Agriculture". Hong Kong Baptist University, 1998.
4. CNSF. "Revista Actualidad en Seguros y Fianzas". SHCP, Comisión Nacional de Seguros y Fianzas, publicación trimestral de 1998 a 2005.
5. CNSF. "Manual del Sistema Estadístico del Sector Afianzador". CNSF Circular F13.3, 2006.
6. CNSF. "Anuario Estadístico de Seguros y Fianzas". SHCP, Comisión Nacional de Seguros y Fianzas, ediciones de 1998 a 2005.
7. CNSF. "Compañías Aseguradoras con participación extranjera". SHCP, Comisión Nacional de Seguros y Fianzas, datos de 1998-2005.
8. Coelli Tim. "A guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program". University of New England. CEPA Working Paper 96/08, 1996.
9. Cooper William, et al. "Data Envelopment Analysis, a Comprehensive text with Models, Applications, References and DEA-Solver Software". Kluwer Academic Publishers, 2000.

10. Cummins David, et al. "Organizational Form and Efficiency: An Analysis of Stock and Mutual Property-Liability Insurers" The Wharton School, University of Pennsylvania, 1996.
11. Cummins David y Venard Bertrand. "Handbook of international Insurance: Between Global Dynamics and Local Contingencies". Springer, 2007, pp. 205-233.
12. Diacon Stephen. "The Efficiency of UK General Insurance Companies". Nottingham University Business School. CRIS Discussion Paper Series, 2001.
13. Eugene Brigham, Joel Houston. "Fundamentos de administracion financiera". Cengage Learning Editores, 2006.
14. García Denise. "Razones Financieras del Sector Asegurador". CNSF Serie Documento de Trabajo, 1993.
15. Kirwan Ana. "El análisis de la Envoltente de Datos: Un Enfoque estadístico". Tesis ITAM, 1994.
16. Sinha Tapen, et al. "Insurance Regulation in North America". Kluwer Law International, 2003.

Segmentación Binaria Circular: Una técnica para detectar regiones dañadas del ADN

LÓPEZ-LIZARRAGA, S. Y NIETO-BARAJAS, L. E.¹

¹Departamento de Estadística, ITAM
sofialopezlizarraga@hotmail.com
lnieto@itam.mx

Resumen

El cáncer es una enfermedad genética que incluye defectos en los mecanismos de control celular que conducen a la formación de tumores. Las células cancerosas presentan alteraciones en el número normal de cromosomas. Se han creado numerosas técnicas para poder estudiar aquellas regiones en el ADN que se modifican cuando esta enfermedad ocurre. La técnica estadística de segmentación binaria circular nos permite acercarnos a aquellas regiones dañadas y observar cual es el gen involucrado para determinados tipos de cáncer.

Palabras clave: cáncer, ADN, cromosomas, regiones dañadas, segmentación binaria circular, pruebas de hipótesis estadísticas.

1. Introducción

Durante la división celular, cada célula duplica su material genético sintetizando una nueva copia de cada uno de sus cromosomas, usando el ADN original como patrón. En este proceso muchas alteraciones pueden ocurrir. Algunas regiones del ADN se pueden multiplicar, resultando en una ganancia del número de cadenas copiadas. Otras veces, puede haber una delección, resultando en una pérdida en el número copiado. En caso de que una alteración no sea detectada, las células llevan estas mutaciones y se proliferan.

Actualmente, hay evidencia de que es posible estudiar el cáncer por medio del número copiado de cadenas de ADN en las células cancerígenas. El número copiado se define como el número de copias de ADN en cierta región del genoma. Cuando una célula es normal, el número copiado es igual a dos; cuando la célula

está dañada, el número copiado será menor o mayor a dos. En la Figura 1 podemos observar el cariotipo de una célula normal y otra con cáncer. La célula cancerosa tiene varias anomalías como la delección del cromosoma 2 y la trisomía en el tercer cromosoma.

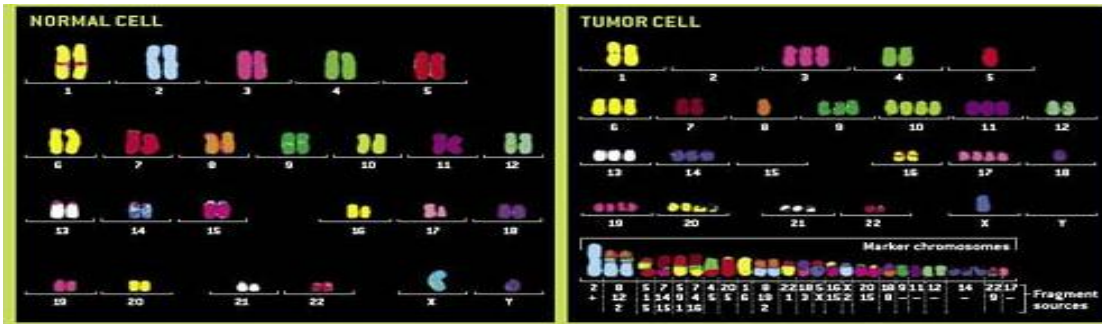


Figura 1. Cariotipo de célula normal (izq.) y célula cancerosa (der.)

Mediante la tecnología de arreglos de Hibridación Genómica Comparativa (aCGH) es posible detectar desequilibrios cromosómicos en el número copiado, ya sea pérdida o ganancia. Debido a errores de medición, para poder ubicar exactamente qué partes del ADN están alterados es necesario utilizar métodos estadísticos que permiten detectarlas y descartar aquellas regiones que aparentan estar dañadas cuando están normales.

La segmentación binaria circular es un método estadístico propuesto por Olshen et al. (2004) para analizar la información de un aCGH mediante el cual se van a poder ubicar puntos de cambio en segmentos del cromosoma por medio de pruebas de hipótesis estadísticas y de esta manera localizar en que parte del genoma hay pérdida o ganancia de ADN. Identificándose así segmentos del genoma con el mismo número copiado.

El presente es un artículo de divulgación que pretende mostrar y explicar los detalles técnicos de la técnica de segmentación binaria circular e ilustrar su implementación y uso mediante algunos ejemplos. La estructura del artículo es la siguiente: La sección 2 describe los arreglos de hibridación genómica comparativa, lo cual nos permite tener una mejor idea del experimento. La sección 3 presenta los detalles técnicos de la segmentación binaria circular. En la sección 4 se presentan algunos detalles necesarios para la implementación de la técnica. Se

ilustra el uso de la técnica en la sección 5 y finalmente se concluye en la sección 6.

2. Arreglos de Hibridación Genómica Comparativa

La Hibridación Genómica Comparativa (CGH) fue creada como un método para detectar alteraciones en el genoma. El experimento de acuerdo con el autor Molinaro et al. (2002) consiste en lo siguiente. Se purifica el ADN en dos muestras una de casos (muestras con tejido canceroso) y la otra de controles (muestras con tejido normal). De esta manera, es posible observar la secuencia de bases nitrogenadas del ADN. Cada una de las muestras se marca con un color diferente, los casos en verde y los controles en rojo. Asumiendo que el tejido normal no tiene ninguna alteración, al hibridar las muestras en sondas va a predominar un color. Esto sucede ya que, si una secuencia de bases nitrogenadas se duplica o se pierde, como los controles no tienen ninguna alteración y el número copiado permanece constante, lo único que cambiaría serían los casos. Si predomina el verde en alguna parte del ADN, hay ganancia en el cromosoma, y si predomina el rojo, hay pérdida. Si observamos un color amarillo, no hay ni pérdida ni ganancia. En la Figura 2, se pueden observar las tres diferentes tonalidades.



Figura 2. Ejemplo de las tonalidades de color en hibridación genómica comparativa

Esta técnica de hibridación presenta el problema de que sólo detecta el número copiado (intensidad del color) en segmentos grandes. Esta limitación se resolvió mediante los arreglos de Hibridación Genómica Comparativa (aCGH). Esta última permite realizar un análisis de la secuencia del número copiado en el genoma entero en un solo experimento. Los datos que se analizan son las intensidades de cada una de las sondas formadas por BACs (vectores de

clonación), las intensidades de los casos (R_s) y las intensidades de los controles (T_s) para cada sonda “s” en el cromosoma. En el artículo del autor Wicker et al. (2007), se puede obtener una descripción de los BACs. Estas mediciones se transforman en variables X_s , para $s=1, \dots, n.$, donde $X_s = \log_2(T_s/R_s)$ y n es el número de sondas. En la Figura 3 se presenta un diagrama que ilustra una hibridación en arreglos.

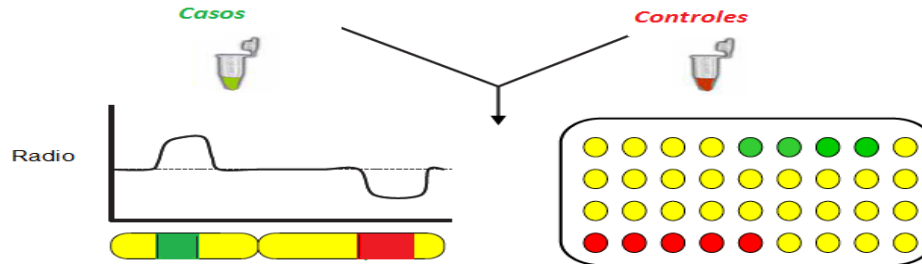


Figura 3. Diagrama de un arreglo de hibridación genómica comparativa.

El \log_2 del cociente de las intensidades es una transformación conveniente que normaliza los datos y tiene un punto de referencia que es el cero. En un escenario en el que no se observe ninguna alteración en el número copiado en una región del ADN, se tendría $X_s = \log_2(2/2) = 0$. Si hay pérdida en una de las copias, tendríamos $X_s = \log_2(1/2) = -1$. Si hubiera pérdida de ambas copias o deleción, X_s tendería a $-\infty$. En el caso de que haya una ganancia tendríamos $X_s = \log_2(3/2) = 0.58$. Si hay copias con ganancias múltiples tendríamos $X_s = \log_2(4/2)$, $X_s = \log_2(5/2)$, $X_s = \log_2(6/2)$, etc. En esta situación teórica, las alteraciones en el número copiado podrían observarse sin la necesidad de métodos estadísticos, sin embargo los valores observados de X_s están sujetos a errores de medición y a errores de la hibridación misma. Ver Figura 4.

Para poder detectar las regiones con número copiado común y los puntos de cambio, se recurre a métodos estadísticos. En particular la técnica de segmentación binaria circular (Olshen et al., 2004) es una de las primeras técnicas propuestas para detectar puntos de cambio y es una de las más populares. La finalidad de esta técnica es poder segmentar el genoma en regiones con el mismo número copiado y así caracterizar el cáncer del paciente a nivel genómico.

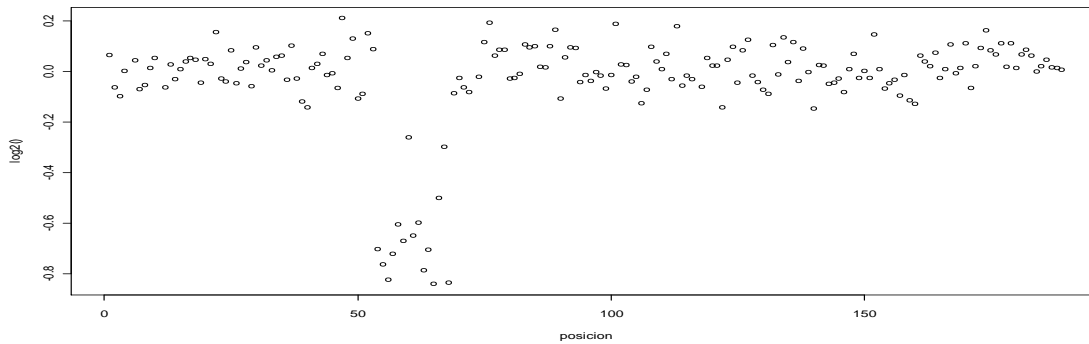


Figura 4. Ejemplo de una gráfica de X_s vs. s .

3. Segmentación Binaria Circular

Sea X_1, X_2, \dots, X_n una sucesión de variables aleatorias independientes que describen el \log_2 del cociente de los números copiados de ADN de n sondas. De manera general, un índice “ v ” se llama punto de cambio en un cromosoma si X_1, \dots, X_v , tienen la misma distribución, digamos F_0 , y X_{v+1}, \dots, X_n tienen una distribución diferente, F_1 . En el estudio de los aCGH, el objetivo consiste en identificar los puntos de cambio que van a partir el cromosoma en segmentos en los que el número copiado permanece constante y determinar si en ese segmento hay pérdida o ganancia en el número copiado.

Segmentación binaria

La técnica de segmentación binaria circular tiene sus orígenes en la técnica de segmentación binaria (Sen, A. y Srivastava, M. S., 1975), la cual describimos a continuación. En el mismo contexto del párrafo anterior, supongamos que $X_i \sim N(\mu_i, \sigma^2)$ independientemente para $i=1, \dots, n$, y sean $S_i = \sum_{r=1}^i X_r$ las sumas parciales, para $1 \leq i \leq n$.

Suponiendo que los puntos de cambio se originan únicamente por cambios en la localización, en el caso de que no hay puntos de cambio, las medias de todas las variables aleatorias serán iguales; en cambio, si hay un punto de cambio en la posición “ i ”, la media de las variables aleatorias X_1, \dots, X_i será diferente a la media de las variables aleatorias X_{i+1}, \dots, X_n , es decir, $X_k \sim N(\mu, \sigma^2)$ para $k=1, \dots, i$ y X_k

$\sim N(\mu^*, \sigma^2)$, para $k=i+1, \dots, n$. Estadísticamente, esto se reduce a probar que $\mu = \mu^*$, es decir, más formalmente

$$H_0: \mu = \mu^*, \sigma^2 > 0 \quad \text{vs.} \quad H_a: \mu \neq \mu^*, \sigma^2 > 0.$$

Note que se está suponiendo que la varianza es desconocida, por lo que el conjunto de posibles valores de σ^2 se escribe como parte de las hipótesis.

Para encontrar un procedimiento de prueba usamos la prueba de cociente de verosimilitud generalizado. No es difícil demostrar que la estadística de prueba obtenida es T_i que tiene una distribución *t*-student con $n-2$ grados de libertad, donde

$$T_i = \frac{\left(\frac{S_i}{i} - \frac{S_n - S_i}{n-i} \right)}{\sqrt{\frac{1}{i} + \frac{1}{n-i}}} \sqrt{\frac{\sum_{r=1}^i \left(X_r - \frac{S_i}{i} \right)^2 + \sum_{r=i+1}^n \left(X_r - \frac{S_n - S_i}{n-i} \right)^2}{n-2}}$$

tal que $T_i \sim t_{(n-2)}$ bajo H_0 .

Recordemos que nuestras hipótesis determinan la existencia de un punto de cambio en la posición i . Si quisiéramos probar la hipótesis nula de que no hay un punto de cambio contra la hipótesis alternativa de que hay punto de cambio, independientemente de la posición en la que ocurra, entonces la estadística de prueba que debe usarse es $T_B = \max_{1 \leq i \leq n} |T_i|$. La hipótesis nula de que no hay puntos de cambio es rechazada si la estadística T_B excede al α -ésimo cuantil de la distribución nula de T_B y el punto de cambio detectado sería i tal que $T_B = |T_i|$.

En caso de haber detectado un punto de cambio, repetiremos el método de manera recursiva dentro de los subsegmentos obtenidos para detectar otros posibles puntos de cambio.

3.2 Segmentación binaria circular

Como el método de Segmentación Binaria fue hecho para encontrar un punto de cambio en cada paso, muchas veces no se detectan puntos de cambio

adicionales. Por esta razón y para hacer más rápida la búsqueda, se modificó el método y se creó la segmentación binaria circular. De esta manera, es posible dividir el cromosoma en tres segmentos al ubicar dos puntos de cambios a la vez.

Consideramos las variables aleatorias $X_1, \dots, X_i, \dots, X_j, \dots, X_n$ donde $X_i \sim N(\mu_i, \sigma^2)$ independientemente para $i=1, \dots, n$. Formamos un círculo, de tal forma que las variables aleatorias X_1 y X_n estén unidas. Tomamos por un lado un primer bloque X_1, \dots, X_i y X_{j+1}, \dots, X_n , y por otro lado un segundo bloque X_{i+1}, \dots, X_j . En el caso de que no haya ningún punto de cambio, las medias de todas las variables aleatorias serán iguales; si hay dos puntos de cambio, la media de las variables aleatorias del primer y segundo bloques serán diferentes. Es decir, $X_k \sim N(\mu, \sigma^2)$ para $k=1, \dots, i$ y $j+1, \dots, n$ y por otro lado $X_k \sim N(\mu^*, \sigma^2)$, para $k=i+1, \dots, n$. En este caso las hipótesis estadísticas son:

$$H_0: \mu = \mu^*, \sigma^2 > 0 \text{ vs. } H_a: \mu \neq \mu^*, \sigma^2 > 0.$$

Nuevamente, usando la prueba de cociente de verosimilitud generalizado llegamos a la siguiente estadística de prueba:

$$T_{ij} = \frac{\frac{(S_j - S_i)}{j-i} - \frac{(S_n - S_j + S_i)}{n-j+i}}{\sqrt{\frac{1}{j-i} + \frac{1}{n-j+i}}}$$

$$= \frac{\sum_{r=1}^i \left(X_r - \frac{(S_n - S_j + S_i)}{n-j+i} \right)^2 + \sum_{r=j+1}^n \left(X_r - \frac{(S_n - S_j + S_i)}{n-j+i} \right)^2 + \sum_{r=i+1}^j \left(X_r - \frac{(S_j - S_i)}{j-i} \right)^2}{n-2}$$

donde $T_{ij} \sim t(n-2)$ bajo H_0 .

Si quisiéramos probar la hipótesis nula de que no hay puntos de cambio contra la hipótesis alternativa de que hay dos puntos de cambio, independientemente de la posición en la que ocurra, entonces usamos la estadística de prueba $T_C = \max_{1 \leq i \leq n} |T_{ij}|$. La hipótesis nula de que no hay puntos de cambio es rechazada si la estadística T_C excede al α -ésimo cuantil de la distribución nula de T_C y los puntos de cambio detectados serían i y j tal que $T_C = |T_{ij}|$.

Si al final del procedimiento, observamos que hay dos puntos de cambio que son i y j . Esto va a dividir nuestra cadena de Bases Nitrogenadas en tres pedazos de X_1, \dots, X_i , de X_{i+1}, \dots, X_j y de X_{j+1}, \dots, X_n . Este procedimiento, se repetirá recursivamente para cada uno de los pedazos.

Los autores Olshen et al. (2004), sugieren realizar un pre-procesamiento de los datos debido a la existencia de observaciones atípicas y un post-procesamiento debido a de tendencias locales. Estas sugerencias son las siguientes:

I. Suavizar puntos atípicos

Muchas veces por errores técnicos o humanos en los experimentos o por números copiados aberrantes en la región de una sonda, existen puntos atípicos. Antes de segmentar en regiones el cromosoma se deben de suavizar los datos. Se define una región de suavizamiento para cada posición i que va de $i-R$ a $i+R$ donde R es un número entero entre 2 y 5.

Sea m_i la mediana de los datos en la región de suavizamiento y $\hat{\sigma}$ la desviación estándar de todos los datos. Si la observación X_i es la máxima o la mínima de todas las observaciones en la región de suavizamiento, buscamos la observación X_j más cercana a X_i . Si la distancia de X_i a X_j excede $L\hat{\sigma}$, con $L=4$, entonces reemplazamos X_i por $m_i + \text{sign}(X_i - X_j)M\hat{\sigma}$, con $M=2$.

II. Eliminar tendencias locales

Muchas veces hay tendencias locales en los datos que nos llevan a considerar puntos de cambio erróneos. Una forma de solucionar este problema es por medio de la técnica "podado de árboles" de Breiman et al. (1984). Esto consiste en eliminar algunos puntos de cambio que se están detectando y que deberían de ser descartados. Suponiendo que tenemos C puntos de cambio. La decisión consiste en escoger C puntos de cambio o considerar, $C-1$, $C-2$, ..., 1 o ningún punto de cambio.

Sea $SS(C)$ la suma del cuadrado de las desviaciones de los datos con respecto a la media de cada segmento definida como:

$$SS(C) = \sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^2 + \sum_{j=1}^{n_2} (Y_{2j} - \bar{Y}_2)^2 + \dots + \sum_{j=1}^{n_k} (Y_{kj} - \bar{Y}_k)^2,$$

para un cromosoma dividido en k segmentos de tamaño n_1, n_2, \dots, n_k . Calculamos $SS(1), \dots, SS(C-1)$ para el mejor conjunto de puntos de cambio que deben ser considerados de tamaño $1, 2, \dots, C-1$. Entonces el número de puntos de cambios que se considerará es $c = \min\{c: [SS(c)/SS(C) - 1] < \lambda\}$, con $\lambda = 0.05$ ó 0.1 .

4. Implementación de la técnica

Recordemos que la estadística de prueba para la detección de un punto de cambio mediante segmentación binaria circular es T_C , la cual está definida como $T_C = \max_{1 \leq i \leq n} |T_{ij}|$. Aún cuando la distribución nula de T_{ij} sea conocida de manera analítica (*t*-student), la distribución de T_C no es fácil de obtener para todo valor de n . El primer autor de esta artículo, como parte de su trabajo de tesis de licenciatura (López Lizarraga, 2009), realizó una aproximación a la distribución nula de T_C mediante métodos de simulación. Esta aproximación se describe a continuación.

4.1 Aproximación de la distribución nula de T_C

Para poder estimar la distribución de la estadística de prueba T_C , se obtiene un número grande de simulaciones de ella de la siguiente manera. Se generan N simulaciones de muestras de tamaño n , X_1, \dots, X_n , donde, sin pérdida de generalidad, se supone $X_s \sim N(0, 1)$, para $s=1, \dots, n$. Para cada una de las N muestras se calcula $T_C = \max_{1 \leq i \leq j \leq n} |T_{ij}|$. El máximo que define a T_C involucra a todas las combinaciones posibles de n en dos, es decir, $n(n-1)/2$ valores de T_{ij} . Se calcula el valor absoluto de las T_{ij} encontradas y el máximo de esas estadísticas será un valor simulado de la estadística de prueba. De esta manera se obtienen N valores simulados de la estadística de prueba T_C , bajo la hipótesis nula. Nótese que la distribución de T_C depende del número de sondas n de cada cromosoma a analizar.

Los cuantiles de la distribución nula de T_C se pueden aproximar mediante los correspondientes cuantiles muestrales usando los N valores simulados de T_C . De la misma manera, los valores p se aproximan mediante frecuencias relativas de los valores simulados de T_C . Si se desea visualmente ver la forma de la distribución nula de T_C es posible calcular la función de distribución empírica o aproximar la densidad mediante un método de aproximación de densidades de kernel.

Para el cálculo de la función de distribución empírica, ordenamos los datos y calculamos para cada valor de t la frecuencia relativa de valores menores o iguales a t , es decir,

$$\hat{F}_N(t) = \frac{\#\text{Tc}_i \text{'s} \leq t}{N},$$

donde N es el número de simulaciones de T_C .

Alternativamente, podemos realizar una estimación de la función de densidad de T_C mediante el método de estimación de Kernel (Silverman, 1978). Este estimador es de la forma

$$\hat{f}(t) = \frac{1}{Nh} \sum_{i=1}^N \delta\left(\frac{t - T_{Ci}}{h}\right),$$

donde δ es el Kernel de probabilidad, con localización cero y dispersión uno, y h es un parámetro de suavizamiento comúnmente conocido como ancho de ventana.

Es importante escoger un Kernel adecuado en el soporte de los datos. La elección del ancho de ventana es más complicada. Si el ancho de ventana es muy pequeño, se generan picos en la función de densidad estimada, en cambio si es muy grande, se sobresuavizan las estimaciones y se obtiene una densidad casi uniforme. El valor de h realmente depende del tamaño de muestra, por lo que $h=h(N)$. Para muestras pequeñas se sugiere un valor para $h(N)$ grande para que se suavicen los datos. En muestras grandes se puede escoger un $h(N)$ pequeño. Silverman (1978) demostró que el valor del parámetro h óptimo al usar kernels Gaussianos y suponiendo que la densidad que generó los datos también es Gaussiana, es de la forma:

$$h(N) = 1.06 \min\{\hat{\sigma}, R/1.34\} N^{-\frac{1}{5}},$$

donde R es el rango intercuartílico y $\hat{\sigma}$ la desviación estándar estimada.

4.2 Determinación de puntos de cambio

Supongamos que X_1, \dots, X_n es un conjunto de valores observados para un cromosoma en particular. Supongamos además que el valor observado de la estadística de prueba es $T_C = t_0$. Con ayuda de los N valores simulados de la distribución nula de T_C , es posible aproximar el valor p como:

$$\text{Valor } p = P(T_C \geq t_0) \approx \frac{\#\text{Tc}_i\text{'s} \geq t_0}{N}.$$

Una vez que se calculó el valor- p , se debe tomar la decisión si se rechaza o no la hipótesis nula de que no hay puntos de cambio a favor de la hipótesis alternativa de que hay puntos de cambio en las posiciones i y j . La hipótesis nula es rechazada si el valor p es menor que el nivel de significancia de la prueba α , en cuyo caso los puntos de cambio serían i y j tal que $T_C = \max_{1 \leq i \leq n} |T_{ij}|$. Como el método es recursivo, se realiza el mismo procedimiento para cada segmento en el que se dividió el cromosoma.

Finalmente, una vez detectados todos los puntos de cambio, se calcula la media en cada uno de los segmentos para poder determinar el número copiado común para todas las sondas del segmento. Si la media del segmento es cero o cercana a cero se le asigna al segmento un color azul de “normal”. Si la media es un valor positivo se le asigna un color verde de “ganancia”, mientras que si es negativo se le asigna el color rojo de “pérdida”.

4.3 Estimación del valor p con permutaciones

Una forma alternativa para aproximar el valor p es mediante el uso de técnicas de permutación (Davison y Hinkley, 1997). De hecho, los autores del método de segmentación binaria circular (Olshen et al., 2004) sugieren esta forma de aproximar el valor p debido a que es más rápida que la aproximación del valor p mediante simulaciones directas como se mencionó en la sección anterior.

Los métodos de permutación consisten en obtener varias “muestras” a partir de la muestra observada, permutándola sin reemplazo. En nuestro caso, permutamos X_1, \dots, X_n obteniendo una muestra diferente a la original (recordemos que la posición de las observaciones es importante en la detección repuntos de cambio). Para cada muestra permutada se calcula la estadística de prueba T_C . Esto se repite N veces, obteniendo N valores para T_C que de acuerdo con Davison y Hinkley (1997) pueden ser usados para aproximar la distribución nula de T_C . Estos valores de T_C así obtenidos se usan para aproximar el valor p como se mencionó en la sección anterior.

5. Ejemplo numérico

Los datos que corresponden a valores de la variable $X_s = \log_2(T_s/R_s)$, se obtuvieron de una base de datos de libre acceso llamada Coriell disponible en http://www.nature.com/ng/journal/v29/n3/suppinfo/ng754_S1.html. La base de datos contiene varias células, en particular trabajamos con los datos de la célula GM05296. Como ilustración consideramos únicamente los cromosomas 10 y 11.

La Figura 5 muestra el comportamiento de las observaciones para los dos cromosomas a estudiar. Se puede observar que el número de sondas “s” utilizadas para cada cromosoma es diferente, presentando una mayor resolución (más sondas) el cromosoma 11.

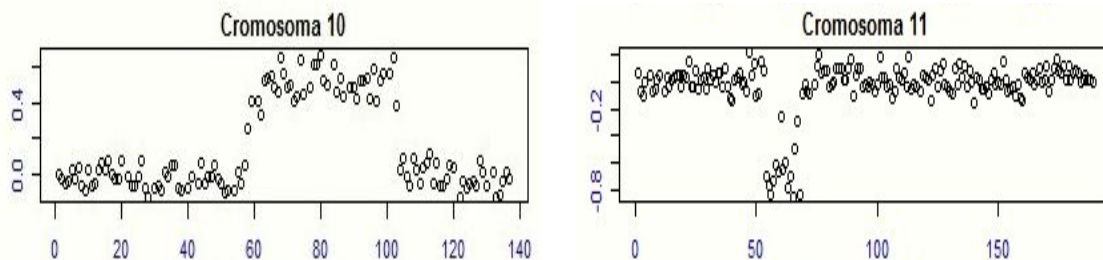


Figura 5. Gráficas de X_s vs. s para los cromosomas 10 y 11 de Coriell

Antes de analizar los datos, se realiza el suavizamiento de los mismos sugerido en la sección 3.2. Una vez realizado este paso se procede a aplicar el método segmentación binaria circular. Se realizó el cálculo de la estadística de prueba T_C y para la determinación del valor p de los valores observados de la estadística se recurrió a los métodos de simulación directa y por permutación

descritos en la sección 4. Se obtuvieron $N=5000$ valores simulados para cada tamaño de muestra, 139 del cromosoma 10, y 189 del cromosoma 11.

La implementación de los cálculos se realizó en el paquete R y Matlab. El código esta disponible solicitándolo al primer autor. Para la aproximación de la función de densidad de T_C bajo H_0 se uso un kernel Gaussiano con ancho de banda óptimo. La Figura 6 muestra los estimadores de kernel de la distribución nula de T_C con $n=139$ y $n=189$ para los cromosomas 10 y 11 respectivamente. La línea vertical corresponde a la localización del valor observado de la estadística de prueba. Como se puede observar, en ambos casos se obtuvo un valor p prácticamente de cero y con un nivel de significancia de $\alpha=0.01$ se detectaron puntos de cambio.

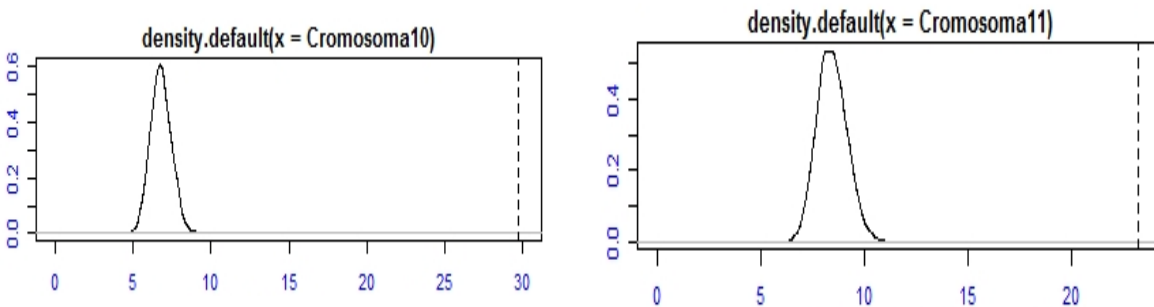


Figura 6. Estimadores de kernel de la distribución nula de T_C para los cromosomas 10 y 11. La línea vertical corresponde al valor observado de la estadística de prueba.

Los puntos de cambio que se obtuvieron para el Cromosoma 10 son las sondas 57 y 103 y para el Cromosoma 11 son las sondas 53 y 69. Estos segmentos se observan en la Figura 7.

Una vez ubicados los primeros puntos de cambio para cada cromosoma, se procede a aplicar el mismo método a cada uno de los segmentos encontrados. La Figura 6 muestra el estimador de kernel de la distribución nula para cada uno de los tres segmentos del cromosoma 10, denotados como 10.1, 10.2 y 10.3 en la Figura 6, y para el cromosoma 11 cuyos subsegmentos se denotan como 11.1, 11.2 y 11.3. Nuevamente la línea punteada vertical corresponde a la localización

de la estadística de prueba observada. La región sombreada en la cola derecha corresponde al valor p .

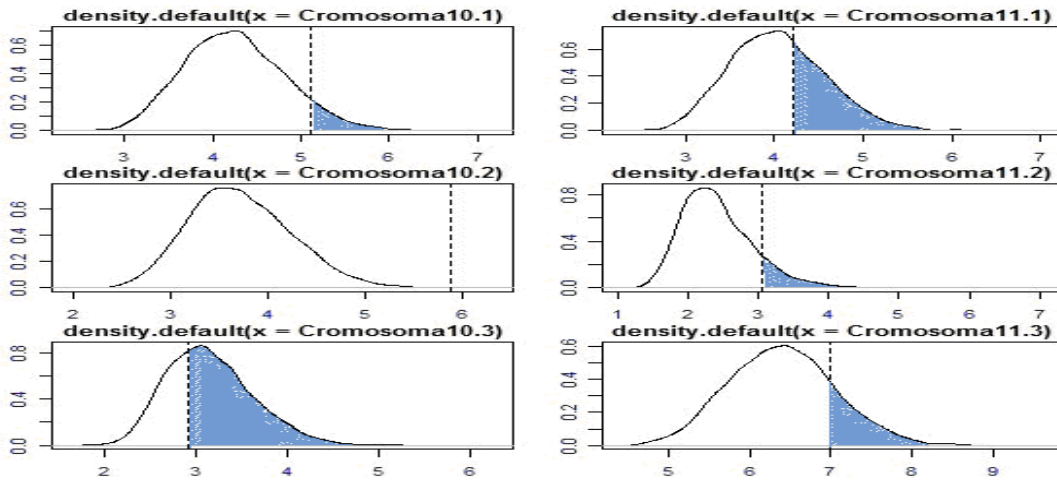


Figura 6. Estimadores de kernel de la distribución nula de T_C de los subsegmentos de los cromosomas 10 y 11. La línea vertical corresponde al valor de la estadística de prueba observado y la parte sombreada corresponde al valor p .

En la Tabla 1 se pueden observar los valores observados de las estadísticas de prueba para los segmentos en los que se dividieron los Cromosomas 10 y 11. El único segmento cuyo valor p es menor o igual al nivel de significancia de 0.01 es el segundo segmento del cromosoma 10 (segmento 10.2). Este segmento va de la sonda 54 a la 94. El punto de cambio detectado corresponde a la sonda 93, sin embargo al estar en el extremo del segmento se descarta como punto de cambio como los sugieren Olshen et al. (2004).

	T_c	Valor-p (simulación)
Cromosoma 10.1	5.1035	0.0744
Cromosoma 10.2	5.8801	0.0002
Cromosoma 10.3	2.9053	0.6706
Cromosoma 11.1	4.2209	0.0306
Cromosoma 11.2	3.0598	0.0892
Cromosoma 11.3	6.9907	0.3014

Tabla 1. Valores observados de la estadística de prueba y valor p para los segmentos de los Cromosomas 10 y 11.

Finalmente, una vez detectados los puntos de cambio, se calcula la media en cada uno de los segmentos para poder determinar el número copiado común para el segmento. En la Figura 7 se pueden observar los resultados finales y los segmentos en los que se dividen los cromosomas. En el cromosoma 10 se observa una ganancia (línea verde) en el segmento que va del la sonda 57 a la 103, mientras que en el cromosoma 11 se observa una pérdida (línea roja) en la región de la sonda 53 a la 69. En las demás partes de ambos cromosomas se observan medias cercanas a cero por lo que se les asigna un color azul de neutralidad.

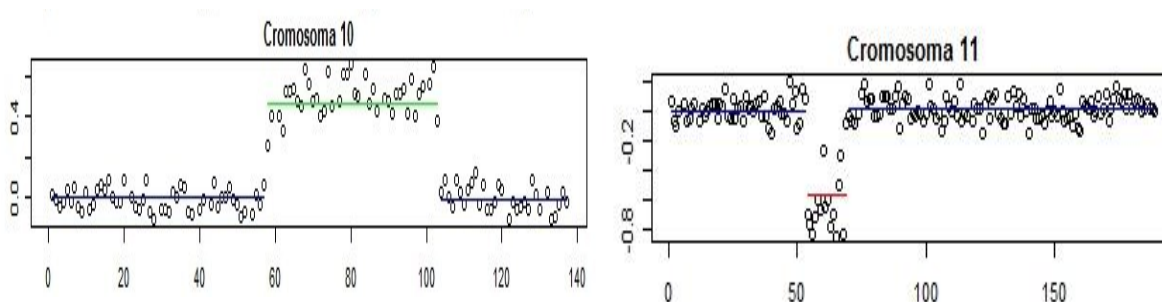


Figura 7. Segmentos detectados con segmentación binaria circular para los cromosomas 10 y 11.

Con fines de comparación, se calcularon también los valores p a través del método de permutaciones descrito en la sección 4.3. Los resultados obtenidos son los mismos obteniéndose valores p cercanos a cero para la primera partición de los cromosomas 10 y 11 y se detectaron los mismos puntos de cambio.

6. Conclusiones

En este artículo se presentaron los detalles técnicos y de implementación del método de segmentación binaria circular. El método se ilustró con un ejemplo sencillo de una base de datos pública llamada Coriell. Todos los cálculos fueron hechos por el primer autor como parte de su tesis de licenciatura. Sin embargo es importante mencionar que existe actualmente una librería del paquete R que implementa el método de segmentación binaria circular llamada DNACopy, la cual es de acceso libre en el sitio <http://r-ptoject.org>.

Vale la pena mencionar que las aplicaciones del método de segmentación binaria circular son muy diversas y de gran importancia para las ciencias biomédicas. Dentro de las aplicaciones más importantes encontramos: 1) Encontrar la expresión de genes para distintos tipos de cáncer y así determinar cuáles de ellos se involucran en cada tipo de cáncer; 2) monitorear la progresión de una enfermedad en los pacientes; 3) diagnosticar una enfermedad; 4) determinar qué tipo de quimioterapia es necesaria para un paciente; 5) encontrar las regiones en las que el orden de bases nitrogenadas está alterado; y 6) realizar estudios de laboratorio para probar cómo reacciona una célula dañada con diferentes drogas y así encontrar una cura para la enfermedad.

Referencias

1. Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984). Classification and Regression Trees. Belmont, Pacific Grove. CA: Wadsworth.
2. Davison, A. y Hinkley, D. (1997). Bootstrap Methods and their Application. Cambridge University Press, Estados Unidos. .
3. López Lizárraga, Sofía (2009). Segmentación Binaria Circular: Una técnica para detectar pérdidas y ganancias en regiones del ADN. Tesis de licenciatura en Actuaría. Instituto Tecnológico Autónomo de México, México.
4. Molinaro, A., van der Laan, M., Moore, D., (2002), Comparative Genomic Hybridization Array Analysis. University of California, Berkeley, Division of Biostatistics Working Paper Series, Working Paper 106.
5. Olshen, A., Venkatraman, E., Lucito, R., Wigler, M. (2004), Circular Binary Segmentation for the Analysis of Array- based DNA Copy Number Data. Biostatistics **5**, 557-572.
6. Sen, A. and Srivastava, M. S. (1975), On tests for detecting a change in mean. Annals of Statistics **3**, 98-108.
7. Silverman, W. (1978), Choosing the Window Width when Estimating Density. Biometrika **65**, 1-11.
8. Wicker, N., Carles, A., Mills, I., Wolf, M., Veerakumarasivam, A., Edgren, H., Boileau, F., Wasyluk, B., Schalken, J., Neal, D., Kallioniemi, O., Poch, O. (2007), A new look towards BAC-based array CGH through a comprehensive comparison with oligo-based array CGH. Genomics **8**, 8-84.

Análisis de la Mortalidad Diferencial por Regiones Socioeconómicas del INEGI

Alejandra Cevallos Nava
AOHUI Consultoría en Administración de Riesgos y Salud
Socio Consultor

Fernando Montes de Oca 17-PH, 52116819
alecevallos@hotmail.com,

Resumen

El presente artículo tiene como finalidad la construcción y el análisis de tablas de mortalidad para las distintas regiones socioeconómicas de México definidas por el INEGI. Se comentan las diferencias socioeconómicas que caracterizan a cada una de las regiones y se busca explicar como éstas inciden en el comportamiento de la mortalidad en cada una de las regiones, distinguiendo por grupos de edad y sexo.

I. Introducción

La salud constituye uno de los derechos humanos aceptados universalmente. En México se han logrado importantes avances en cuanto a la salud; sin embargo, este progreso no ha sido homogéneo a nivel estatal, lo que se ve reflejado en los niveles de mortalidad de la población a nivel regional. Para hablar del desarrollo socioeconómico que ha tenido un país o región resulta muy importante analizarlo a través del fenómeno demográfico de la mortalidad.

Se debe mencionar que son grandes las lagunas que se tienen en el conocimiento de los factores sociales, culturales, económicos y médicos que están afectando directa o indirectamente a la mortalidad; sin embargo, se sabe que variables como el sexo, nivel educativo, nivel de ingreso, edad, lugar de residencia y acceso a ciertos servicios de salud influyen en el comportamiento de la mortalidad⁸.

El objetivo de este trabajo es analizar las diferencias en la mortalidad en México entre las regiones socioeconómicas definidas por el INEGI. Las fuentes de información serán las bases de datos de mortalidad de la Secretaría de Salud para el año 2007 y la información obtenida en el XII Censo General de Población y Vivienda 2000 y el II Conteo de Población y Vivienda 2005. Con esta información se construyen las tablas de mortalidad para las 7 regiones socioeconómicas por grupos de edad y sexo. Para finalmente mostrar un análisis sobre los resultados obtenidos y presentar las conclusiones del mismo.

II. Regiones Socioeconómicas del INEGI

En el año 2004 el INEGI realizó un estudio que tuvo como objetivo presentar un resumen comparativo de las diferencias y similitudes observadas en las condiciones económicas y sociales de la población en todo el territorio nacional, mediante indicadores que abordan temas relacionados con el bienestar como son: educación, ocupación, salud, vivienda, y empleo.

Para ello se formaron 7 estratos que tienen en promedio características similares, es decir, son homogéneos. Ordenando dichos estratos tal que en el estrato 7 se

⁸Raúl Benítez Zenteno, "Los factores y el cambio demográfico en México", Siglo Veintiuno editores, S.A, primera edición, 1984.

encuentran las Entidades Federativas (Municipios o AGEBs según sea el caso) que presentan en promedio la situación relativa más favorable; en ese orden, el estrato 1 se compone de las unidades que en promedio presentan la situación relativa menos favorable. Para este estudio se ha decidido trabajar a nivel Entidad Federativa, con lo cual dejamos de lado la clasificación por municipios o por AGEB, quedando la clasificación de la siguiente manera:

Cuadro II.1 Clasificación por Regiones Socioeconómicas

Región Socioeconómica	Entidad Federativa
7	Distrito Federal
6	Aguascalientes Coahuila Jalisco Nuevo León
5	Baja California Baja California Sur Chihuahua Sonora Tamaulipas
4	Colima Estado de México Morelos Nayarit Querétaro Quintana Roo Sinaloa Yucatán
3	Durango Guanajuato Michoacán Tlaxcala Zacatecas
2	Campeche Hidalgo Puebla San Luis Potosí Tabasco Veracruz
1	Chiapas Guerrero Oaxaca

Fuente: INEGI

III. Construcción de la tabla de mortalidad

III. 1 Población por Región Socioeconómica.

Partiendo de la información del número de habitantes por entidad federativa, sexo y grupo de edad del XII Censo General de Población y Vivienda 2000 y del II

Conteo de Población y Vivienda 2005, se proyectó la información al 30 de junio de 2007.

Una vez proyectada la información de población por edad y sexo al 30 de junio de 2007, se procedió a agrupar por región socioeconómica, conforme a la clasificación que se incluye en la sección anterior.

III.2 Construcción de la información sobre la mortalidad por entidades federativas

Para el análisis de la mortalidad por regiones socioeconómicas se utilizó como base la información del Sistema Nacional de Información en Salud (SINAIS) sobre las bases de datos de defunciones⁹ para el 2007. Dicha información contiene más de 45 variables que nos ayudan a categorizar los datos dependiendo del tipo de análisis que nos interesa. Para efectos de este trabajo se decidió clasificar la información respecto a la edad de defunción, la entidad de ocurrencia y el sexo.

Para clasificar respecto a la edad se hicieron ciertas consideraciones ya que la forma en la que se registró dicha variable fue a través de la variable “cveedad”, la cual podía tomar 4 valores (A=años, M=meses, D=días y H=horas) y la variable “edad” que tomaba valores de 1 a 120 y para los casos en los que no se especificaba la edad se registraba como 998. De tal forma que todos aquellos registros que tenía en la variable “cveedad” los valores de M, D y H correspondían a las defunciones de menores de un año.

Al igual que para el caso de Población, se decidió prorratear la información de los NE, siguiendo exactamente el mismo método (ver anexo 1).

Se decidió analizar al grupo de edad de menores de un año por separado ya que la mortalidad infantil resulta ser todo un tema de análisis por sí solo y aunque el

⁹ Fuente: SINAIS, Disponible en: <http://sinais.salud.gob.mx/basesdedatos/index.html#defunciones>, Consultado el día 17 de marzo de 2009

propósito de este trabajo no es analizar la mortalidad infantil consideramos importante saber cómo se comporta este indicador en las distintas regiones socioeconómicas del país. De la misma forma se optó por agrupar a los mayores de 84 años en un solo grupo debido a la carencia de información.

III.3 Construcción de la tabla de mortalidad

Una vez obtenida la población y la mortalidad registrada para cada región socioeconómica del país, se procedió a iniciar los cálculos propios de una tabla de mortalidad. El método más utilizado para construir una tabla de mortalidad a partir de la información de estadísticas vitales e información censal consiste en estimar en primer lugar las tasas centrales de mortalidad, como el cociente de las defunciones ocurridas en el período de interés entre la población expuesta a la mitad de dicho período, es decir:

$$\tilde{m}_x = \frac{nD_x^t}{P_{x,x+n}^{t+1/2}}$$

donde:

nD_x^t = Defunciones registradas para las edades x y $x+n$ en el año t que para nuestro caso representa los grupos quinquenales antes mencionados.

$P_{x,x+n}^{t+1/2}$ = Población a mitad del período de observación.

Posteriormente, utilizando los factores de tiempo promedio vivido de los fallecidos entre edades x y $x+n$ (${}_n f_x$), se calculan las probabilidades de fallecimiento (${}_n q_x$). En general el método más usado para la estimación de dichas probabilidades es bajo el supuesto de Distribución Uniforme de Muertes (DUM), lo que nos lleva a:

$${}_n q_x = \frac{{}_n m_x}{(1 + n(1 - 1/2)) {}_n m_x}$$

Una vez obtenidas las probabilidades de fallecimiento se genera el resto de las series, para finalmente obtener una tabla de mortalidad para hombres y otra para mujeres por región socioeconómica que contenga la siguiente información:

${}_nq_x$, ${}_nl_x$, ${}_nd_x$, ${}_nL_x$, ${}_nT_x$ y e_x donde:

${}_nl_x$ que representa la función de vivos y se calcula como:

$${}_nl_x = {}_nl_{x-n} * (1 - {}_nq_{x-n})$$

Cabe mencionar que se tomó $l_0 = 1,000,000$ de habitantes, por lo que decimos que las tablas presentadas en este trabajo nos dan el número de defunciones por cada millón de habitantes

${}_nd_x$ que representa el número de defunciones en un período específico y se calcula como:

$${}_nd_x = {}_nl_x * {}_nq_x$$

${}_nL_x$ que representa los años persona vividos y se calcula como

$${}_nL_x = {}_nd_x / {}_nm_x$$

${}_nT_x$ que representa el número de años persona vividos a futuro y se calcula como:

$$T_x = \sum_{y=x}^{\infty} L_y$$

e_x que es la esperanza de vida a edad alcanzada x y se calcula como

$$e_x = T_x / l_x$$

Es importante recordar que para efectos de este estudio se tomó para el análisis de la mortalidad únicamente los datos arrojados por el SINAIS para el año 2007, cuestión que pudiera arrojar datos con una mayor varianza. Sin embargo también vale la pena resaltar que en nuestro país no se han dado eventos catastróficos de gran impacto como para alterar radicalmente los resultados de un año a otro.

Es sabido que la tasa de mortalidad infantil representa un importante indicador en los niveles económicos y de desarrollo que presenta un país o región. Para suavizar la subestimación en la mortalidad infantil se propone el estudio de ésta a través del estimador construido con el método indirecto de William Brass. Teniendo en cuenta la fuente de datos, se considera que la TMI es subestimada por problemas como:

- a) Errores por mala instrucción de los encuestadores, falta de supervisión o captura equivocada.
- b) Omisión de áreas geográficas, viviendas y/o personas.
- c) Subregistro tanto de nacimientos como de defunciones.
- d) Falta de especialistas para establecer claramente las causas de muerte.
- e) Falsedad en las declaraciones de los encuestados.
- f) Familiares que ocultan determinadas enfermedades.
- g) Registros extemporáneos.

Realizando estos cálculos llegamos a los siguientes resultados:

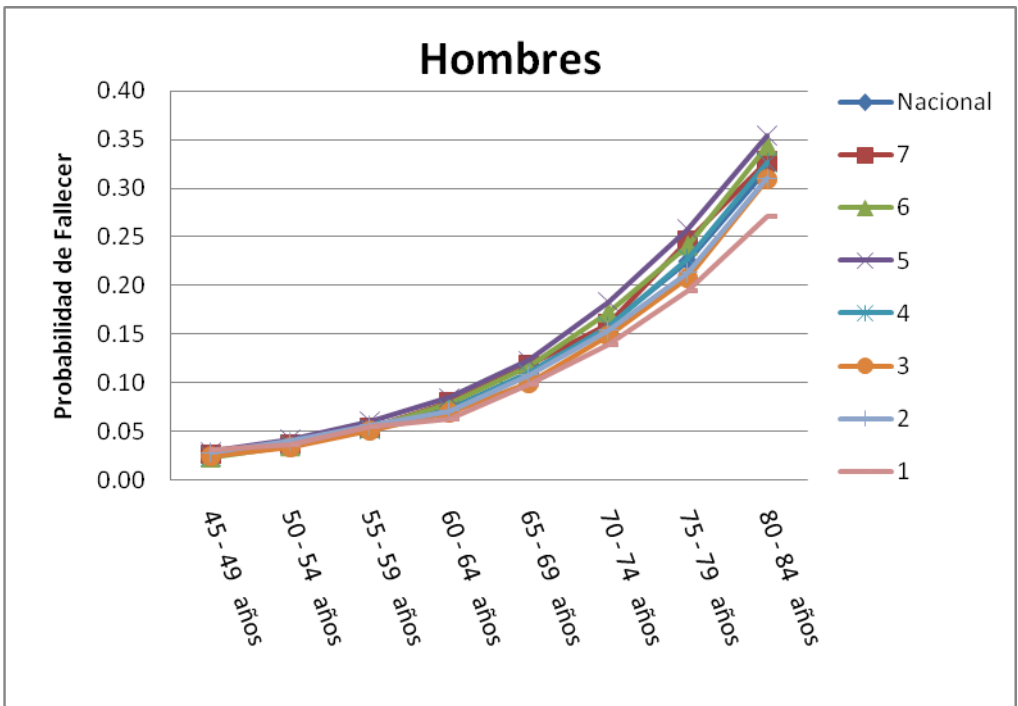
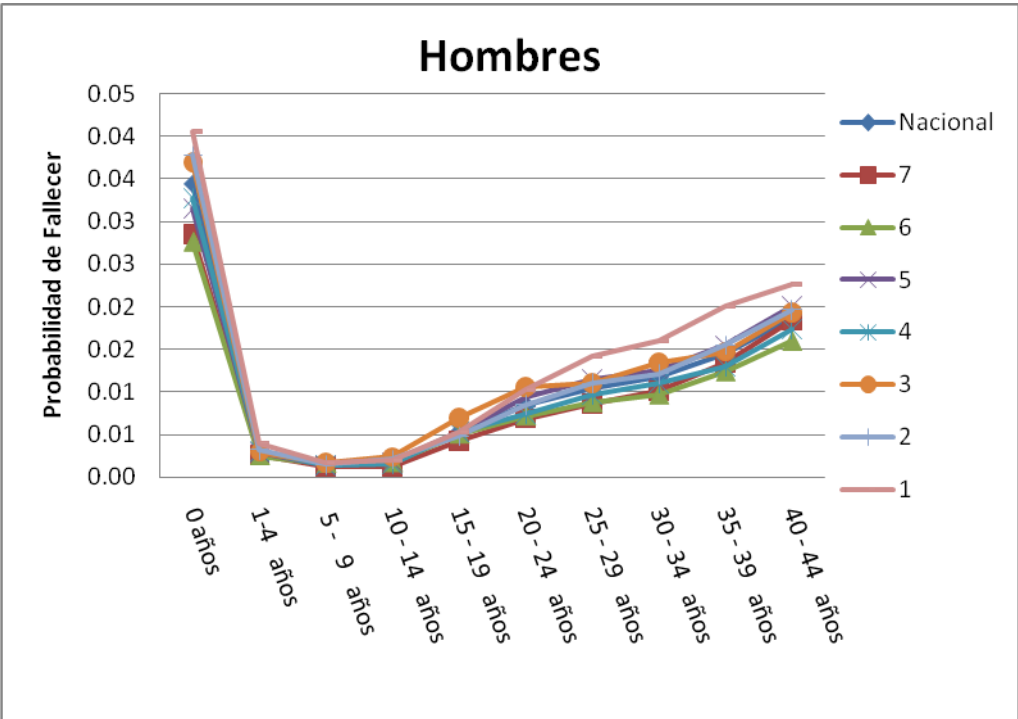
EDAD	Nacional		Región 7		Región 6	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
	nq_x	nq_x	nq_x	nq_x	nq_x	nq_x
0 años	0.03437	0.02744	0.02855	0.02311	0.02762	0.02198
1-4 años	0.00307	0.00263	0.00261	0.00235	0.00252	0.00201
5 - 9 años	0.00152	0.00128	0.00125	0.00101	0.0016	0.00111
10 - 14 años	0.00185	0.00129	0.00126	0.00103	0.00169	0.00129
15 - 19 años	0.00532	0.00231	0.00427	0.00203	0.00505	0.00209
20 - 24 años	0.00841	0.00271	0.00688	0.00239	0.00715	0.00233
25 - 29 años	0.01048	0.00327	0.00862	0.00316	0.00875	0.00275
30 - 34 años	0.01182	0.00394	0.01012	0.00317	0.00964	0.00332
35 - 39 años	0.01454	0.00547	0.01345	0.00469	0.01236	0.00503
40 - 44 años	0.01872	0.0087	0.01836	0.00808	0.01594	0.00764
45 - 49 años	0.02649	0.01364	0.02623	0.01262	0.02351	0.01266
50 - 54 años	0.03731	0.02223	0.03675	0.02171	0.03507	0.02206
55 - 59 años	0.05446	0.0368	0.05364	0.03692	0.05378	0.03783
60 - 64 años	0.07362	0.0532	0.08052	0.05495	0.07821	0.05697
65 - 69 años	0.11027	0.08262	0.11837	0.0864	0.11687	0.08559
70 - 74 años	0.15891	0.12436	0.16078	0.12103	0.17235	0.12982
75 - 79 años	0.22477	0.18999	0.24641	0.19405	0.24116	0.19778
80 - 84 años	0.31993	0.27407	0.32743	0.27843	0.34333	0.28588

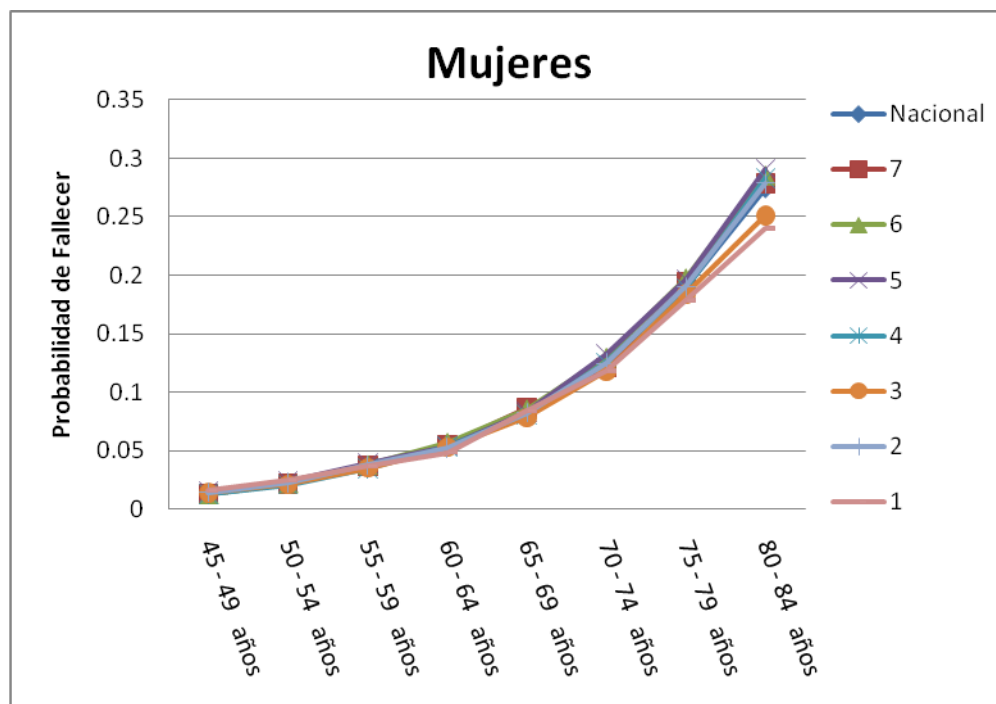
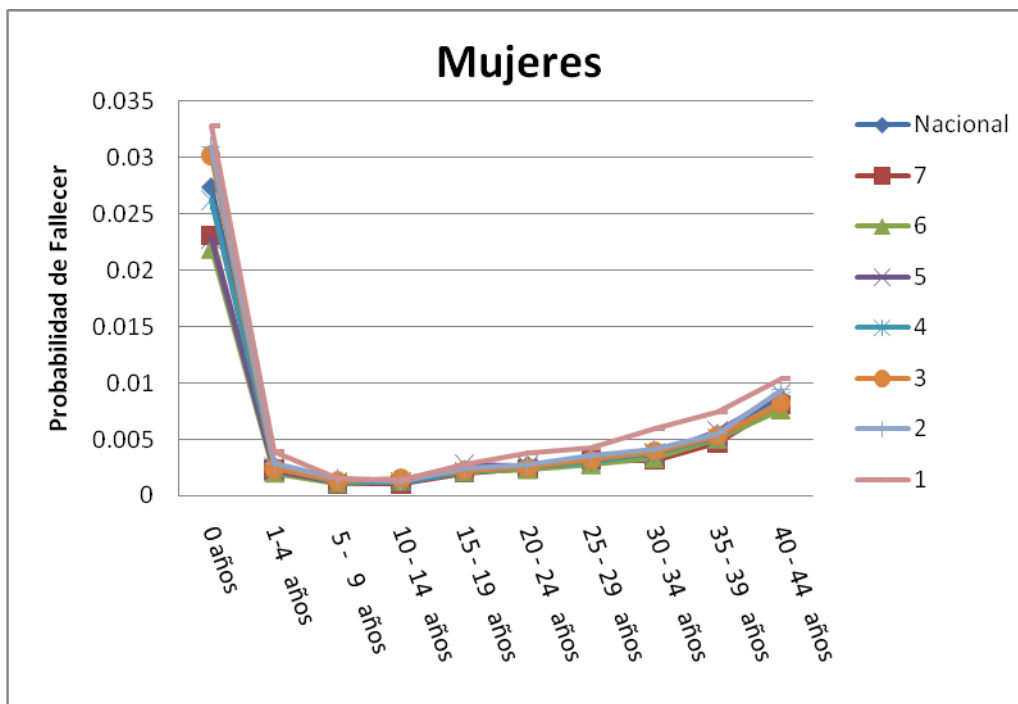
EDAD	Nacional		Región 5		Región 4	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
	nq_x	nq_x	nq_x	nq_x	nq_x	nq_x
0 años	0.03437	0.02744	0.03155	0.02283	0.03278	0.02621
1-4 años	0.00307	0.00263	0.00287	0.00218	0.00309	0.00257
5 - 9 años	0.00152	0.00128	0.00139	0.00115	0.00139	0.0012
10 - 14 años	0.00185	0.00129	0.00177	0.00118	0.00159	0.00123
15 - 19 años	0.00532	0.00231	0.00534	0.00276	0.0053	0.00206
20 - 24 años	0.00841	0.00271	0.00931	0.00267	0.00746	0.0026
25 - 29 años	0.01048	0.00327	0.0115	0.0034	0.00964	0.00296
30 - 34 años	0.01182	0.00394	0.01255	0.00369	0.01101	0.00378

EDAD	Nacional		Región 5		Región 4	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
	nq_x	nq_x	nq_x	nq_x	nq_x	nq_x
35 - 39 años	0.01454	0.00547	0.01549	0.00573	0.01299	0.00513
40 - 44 años	0.01872	0.0087	0.02009	0.00902	0.01736	0.00843
45 - 49 años	0.02649	0.01364	0.02925	0.01473	0.02441	0.01282
50 - 54 años	0.03731	0.02223	0.04212	0.02372	0.03619	0.0209
55 - 59 años	0.05446	0.0368	0.06058	0.03892	0.05229	0.03494
60 - 64 años	0.07362	0.0532	0.08467	0.05386	0.07188	0.05321
65 - 69 años	0.11027	0.08262	0.123	0.08302	0.10969	0.08074
70 - 74 años	0.15891	0.12436	0.18328	0.13263	0.15829	0.12539
75 - 79 años	0.22477	0.18999	0.25784	0.19633	0.22722	0.18842
80 - 84 años	0.31993	0.27407	0.35398	0.29103	0.32671	0.2828

EDAD	Nacional		Región 3		Región 2		Región 1	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
	nq_x	nq_x	nq_x	nq_x	nq_x	nq_x	nq_x	nq_x
0 años	0.03437	0.02744	0.03699	0.03019	0.03779	0.031	0.0406	0.03284
1-4 años	0.00307	0.00263	0.00298	0.00243	0.00322	0.00286	0.00403	0.00394
5 - 9 años	0.00152	0.00128	0.0017	0.0013	0.00149	0.0015	0.00175	0.00152
10 - 14 años	0.00185	0.00129	0.00238	0.00151	0.00208	0.00132	0.00195	0.00136
15 - 19 años	0.00532	0.00231	0.00698	0.00225	0.0048	0.00239	0.00537	0.00276
20 - 24 años	0.00841	0.00271	0.01056	0.0025	0.00851	0.00278	0.01022	0.00385
25 - 29 años	0.01048	0.00327	0.01101	0.00318	0.01103	0.00356	0.01416	0.00425
30 - 34 años	0.01182	0.00394	0.01345	0.00396	0.01212	0.00416	0.01604	0.00595
35 - 39 años	0.01454	0.00547	0.01473	0.00533	0.01549	0.00555	0.02006	0.00748
40 - 44 años	0.01872	0.0087	0.01929	0.00825	0.01967	0.00941	0.02262	0.0104
45 - 49 años	0.02649	0.01364	0.02484	0.01368	0.02833	0.01424	0.03124	0.01574
50 - 54 años	0.03731	0.02223	0.03409	0.02124	0.03984	0.02277	0.03675	0.02478
55 - 59 años	0.05446	0.0368	0.05115	0.03511	0.05626	0.03765	0.05456	0.03741
60 - 64 años	0.07362	0.0532	0.069	0.05284	0.0713	0.05202	0.06275	0.04846
65 - 69 años	0.11027	0.08262	0.09947	0.0783	0.10821	0.08232	0.09797	0.084
70 - 74 años	0.15891	0.12436	0.14833	0.11783	0.15384	0.12455	0.13874	0.11854
75 - 79 años	0.22477	0.18999	0.20807	0.18382	0.2132	0.19033	0.19425	0.17903
80 - 84 años	0.31993	0.27407	0.30938	0.25126	0.31001	0.27859	0.27153	0.23988

A continuación se muestran algunas gráficas comparativas





Los resultados obtenidos son razonables si partimos de la importancia que tiene la mortalidad infantil como indicador del desarrollo económico y social en las entidades federativas ya que conforme nos alejamos de la región socioeconómica mejor calificada (Región 7) mayor será la mortalidad infantil.

Ahora bien, si analizamos los grupos de edad 1-4 y 5-9 resulta que las diferencias entre regiones disminuyen considerablemente, presentando en ambos casos a la Región 1 (Guerrero, Chiapas y Oaxaca) como la que presenta mayor probabilidad de fallecer para estos grupos. Dicho resultado tiene un fundamento esperado desde varios puntos de vista. El primero de ellos, la situación de pobreza alimentaria ya que según las estimaciones del Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) dichas entidades son las que representan el mayor porcentaje de población en situación de pobreza alimentaria, al encontrarse por arriba del 37 %¹⁰. Es sumamente comprensible que para dichos grupos de edad la mala alimentación sea un factor agravante y determinante en la probabilidad de fallecer ya que nos referimos a una de las etapas de mayor desarrollo en un ser humano. Ahora bien, si lo analizamos desde el punto de vista de la educación, resulta ser que el estado de Chiapas ocupa el último lugar en asistencia escolar en niños de entre 5 y 12 años, es decir, nivel primaria, con una asistencia del 92.4% en el 2005; Guerrero se encuentra tan solo una posición arriba con el 94.3 % de asistencia y Oaxaca se encuentra 5 posiciones arriba de Chiapas con una asistencia del 95.5%¹¹. Resulta interesante como las entidades que muestran una mayor probabilidad de fallecer son aquellas que ocupan los lugares más desfavorecidos en los ámbitos de alimentación y educación.

Para las edades comprendidas entre los 10 y los 19 años la región que muestra una menor probabilidad de fallecer es la Región 7, mientras que la Región 3

¹⁰Fuente: estimaciones del CONEVAL con base en el II Censo de Población y Vivienda 2005 y la Encuesta Nacional de Ingreso y Gasto de los Hogares 2005. Información disponible en <http://www.coneval.gob.mx/mapas/NACIONAL/Nacional.pdf> consultada el 25 de Abril de 2009.

¹¹Fuente: estimaciones del INEGI con base en el II Censo de Población y Vivienda 2005. Información disponible en: <http://www.inegi.org.mx/est/contenidos/espanol/rutinas/ept.asp?t=medu06&s=est&c=3273> consultada el 5 de mayo de 2009.

muestra la mayor probabilidad. Si analizamos dicha información desde el enfoque de la educación, resulta que el Distrito Federal ocupa el mejor lugar en cuanto a la asistencia escolar en niños de 5 a 15 años, teniendo una asistencia por arriba del 88%¹². Por otro lado, las condiciones de vida y el rezago social pueden significar grandes diferencias en la mortalidad y vale la pena mencionar que el Distrito Federal ocupa el mejor lugar en cuanto a porcentaje de viviendas con piso de tierra ya que menos del 1%¹³ cumple con esta condición; al mismo tiempo se encuentra como el mejor ubicado en viviendas con agua entubada y drenaje; condiciones de vivienda que sin duda ayudan a la higiene y salud de los individuos.

Para los grupos comprendidos entre los 20 y los 49 años, se observa que la región que presenta mayor probabilidad de fallecer es la región 1, mientras que las regiones que presentan menor probabilidad son la Región 6 y 7, cuestión que una vez más resulta muy razonable tomando en cuenta el desarrollo social y económico que presentan dichas regiones.

Vale la pena mencionar que la Región 1 es la que presenta el mayor porcentaje de población sin derechohabencia a servicios de salud, teniendo a más del 74% de sus habitantes en dicha situación¹⁴, cuestión que sin duda alguna influye en los niveles de mortalidad de la Región. Por otro lado, si partimos de que este segmento de la población se encuentra dentro de lo que se considera como la

¹² Fuente: estimaciones del INEGI con base en el II Censo de Población y Vivienda 2005. Información disponible en: <http://www.inegi.org.mx/est/contenidos/espanol/rutinas/ept.asp?t=medu06&s=est&c=3273> consultada el 5 de mayo de 2009.

¹³ Fuente: Indicadores, índice y grado de rezago social, estimaciones del CONEVAL con base en el II Censo de Población y Vivienda 2005. Disponible en: <http://www.coneval.gob.mx/mapas/NACIONAL/Nacional.pdf> consultada el 25 de Abril de 2009.

¹⁴ Indicadores, índice y grado de rezago social, estimaciones del CONEVAL con base en el II Censo de Población y Vivienda 2005. Disponible en: <http://www.coneval.gob.mx/mapas/NACIONAL/Nacional.pdf> consultada el 25 de Abril de 2009.

Población Económicamente Activa (PEA) y que las entidades de esta región son quienes presentan mayores porcentajes de población analfabeta, tenemos entonces un mayor porcentaje de la población que se dedica a actividades agropecuarias, ya que según estudios del INEGI, para el caso de la población analfabeta el 51% de ellos se dedica a esta actividad, mientras que para el caso de la población alfabetada solo el 13.1%¹⁵ lo hace y es sabido que las condiciones de trabajo en el campo demandan grandes esfuerzos físicos que a la larga pudieran traer como consecuencia problemas de salud.

Si analizamos a la población masculina entre 50 y 59 años de edad observamos que la región que presenta una mayor probabilidad de fallecer es la Región 5, mientras que la Región 3 se encuentra ubicada como la de menor probabilidad de fallecer para este segmento de la población.

Para los mayores de 60 años se tienen patrones claros en los que se denota que la Región con mayor probabilidad de fallecer es la Región 5 y por el contrario, la Región 1 ocupa la menor probabilidad de fallecer. Es de llamar la atención que la Región más desfavorecida y con mayor rezago social sea la que presente una menor mortalidad en la vejez masculina; sin embargo, quizá podamos atribuirlo a la cuestión cultural que se vive en estas entidades federativas ya que es sabido que entre las comunidades rurales el arraigo a los padres y al consejo de los ancianos como autoridad suprema¹⁶ son muy apreciados y por ende, procuran cuidar a este grupo de la población.

¹⁵ “Estadísticas a propósito del día Internacional de la Alfabetización”, Datos Nacionales, INEGI, Disponible en:

<http://www.inegi.gob.mx/inegi/contenidos/espanol/prensa/contenidos/estadisticas/2002/alfabeti02.pdf>

Consulta realizada el 22 de mayo de 2009.

¹⁶ Navarrete Linares Federico, “Los pueblos indígenas de México”, Comisión Nacional para el Desarrollo de los Pueblos Indígenas. Disponible en: http://www.cdi.gob.mx/index.php?option=com_docman&Itemid=24,

Otra cuestión que es importante resaltar, es que las entidades que conforman la región 1 son las que mayor participación de la población ocupada en el sector primario presentan¹⁷ y dicho sector refleja una estructura radicalmente diferente a los otros sectores, ya que en él existe un repunte de la participación de la población mayor de 65 años¹⁸, lo cual nos habla de cierto arraigo al sector tradicional, en el que la edad no condiciona la permanencia en activo y consecuentemente se encuentran en posibilidad de garantizar una de las necesidades primarias como lo es la alimentación. A esto adicionamos el hecho de que la mayor parte de esta población carece de un sistema de pensiones y prestaciones, lo cual limita la jubilación y refuerza la necesidad de producir para el autoconsumo.

En el caso de las mujeres llama la atención que quienes presentan menores índices de mortalidad son las regiones más desfavorecidas, tomando en cuenta el nivel socioeconómico en el que se encuentran; y por otro lado llama la atención que las regiones 5, 6 y 7 presenten la más alta mortalidad ya que son entidades federativas con importante actividad económica, altos niveles de educación y derechohabiencia.

Si bien la diferencia entre regiones para las edades entre los 60 y los 79 años no es muy marcada, para el grupo de edad 80-84 se tienen diferencias importantes. Se tiene a la región 5 como la de mayor probabilidad de fallecer con una $nx = .29103$ y a la Región 1 como la de menor probabilidad con una $nx = .23988$,

Consultado el 9 de junio de 2009.

¹⁷ Ahidé Rivera Vázquez, "Perfil sociodemográfico de la población ocupada en el sector primario y su distribución territorial", Publicaciones Recientes del CONAPO, Disponible en: <http://www.conapo.gob.mx/publicaciones/sdm/sdm2008/10.pdf> Consultado el 22 de mayo de 2009.

¹⁸ Ibid

cuestión que una vez más podemos atribuir al arraigo familiar que existe en las zonas rurales.

A este sector de la población habrá que seguirle mucho la pista durante las próximas décadas ya que es sabido que el rol que juega la mujer en la sociedad mexicana ha sufrido una importante transformación en las últimas décadas y será muy interesante valorar el comportamiento de la mortalidad bajo estas nuevas condiciones de vida.

IV. Conclusiones

El estudio de la mortalidad es quizá uno de los indicadores que presenta mayor sensibilidad ante los cambios en el desarrollo socioeconómico de una región y a través de esto se puede tener una clara visión sobre las diferentes condiciones de vida que se presentan a lo largo y ancho de un país con la finalidad de poder orientar las políticas económicas, educativas, culturales y sociales dependiendo de las necesidades específicas de cada región, buscando como principal objetivo mejorar las condiciones de vida de la población en general.

Los cambios en la organización social combinados con los efectos de los avances tecnológicos han arrojando como resultado un abatimiento de los altos índices de mortalidad; sin embargo, estos cambios, al igual que el desarrollo social y económico se ha dado de forma desigual para las diferentes regiones. De esta manera se experimentan índices de mortalidad diversos según la situación social predominante en cada región.

Este trabajo permite conocer, según la clasificación del INEGI, como se comporta la mortalidad en las 7 diferentes regiones socioeconómicas del país. Existen algunas limitaciones en la información disponible que se traducen en resultados

muy poco creíbles como fue el caso de la mortalidad infantil, ya que la región con mayor mortalidad infantil era justamente aquella que tenía condiciones socioeconómicas más favorables.

En México se subestima la TMI básicamente por dos razones; por un lado el sub-registro de defunciones infantiles y por el otro el sobre-registro de nacimientos y ambos errores actúan en el mismo sentido, es decir subestiman la mortalidad infantil¹⁹. Por ello se optó por utilizar el método de estimación indirecta de William Brass que resulta ser la manera más práctica y conocida de corregir el problema de la subestimación que presenta la mortalidad infantil utilizando información acerca de los hijos nacidos vivos e hijos sobrevivientes que proporcionan las mujeres en edad fértil; esta información se puede obtener a través de los Censos. Los resultados arrojados por dicha estimación muestran elevados índices de mortalidad en las regiones socioeconómicas más desfavorecidas, como es el caso de Oaxaca, Chiapas y Guerrero y una mortalidad infantil mucho más controlada en las regiones más favorecidas como es el caso del Distrito Federal, Nuevo León o Jalisco.

Los avances logrados en México en cuanto a salud han sido significativos; sin embargo, no han sido equitativos entre las entidades federativas, incluso entre género en una misma entidad. Y es que hablando en términos de la esperanza de vida encontramos importantes diferencias como es el caso, para la población masculina, de la Región 5 con una esperanza de vida por debajo de los 70 años y para la región 6 una esperanza superior a los 71 años de vida es decir más de 19 meses de diferencia entre una región y otra; para el caso de las mujeres la Región

¹⁹ Aguirre Alejandro, "La Mortalidad Infantil en los años noventa", Boletín Oficial del G.A.A.P.P, Año 5, Número 28, Octubre 1998.

1 presentó la menor esperanza de vida con más de 74 años y la Región 7 la mayor esperanza de vida con 75.62 años, cuestión que deja ver una diferencia de más de 14 meses. Pero siguiendo estos datos es muy interesante observar que la diferencia entre género es aún mayor, si tomamos la región con mayor esperanza de vida para el caso de los hombres y lo comparamos con el de las mujeres resulta ser que existen más de 4 años de diferencia, cuestión que sin duda alguna para nosotros los actuarios deberá considerarse ampliamente al construir tablas de mortalidad para los seguros de vida, gastos médicos y pensiones principalmente.

Si bien se ha comentado sobre las lagunas que existen para determinar los factores específicos que influyen en el comportamiento de la mortalidad, se puede concluir que, sobre todo a edades tempranas, el nivel socioeconómico y las características de desarrollo que presentan cada una de las regiones en el país son un factor determinante en el comportamiento de la mortalidad. Sin embargo, no se puede concluir lo mismo para edades avanzadas pues resulta ser que las regiones más desfavorecidas son las que presentan una menor probabilidad de fallecer, muy probablemente por un subregistro de fallecimientos similar a lo que ocurre con la mortalidad infantil.

Otro hecho que se puede concluir es que la mortalidad por género presenta un patrón que favorece ampliamente al sexo femenino ya que para todas las edades y regiones socioeconómicas dicha mortalidad resulta ser menor que la que presenta el sexo masculino. Las diferencias entre regiones en términos de la esperanza de vida varían de 1.55 a 1.64 años para los hombres en edades de 0 a 10, de 1.56 a 1.57 años para hombres de edades de 10 a 19 años, de 1.57 a 1.85

años para la población masculina entre los 20 y los 59 años de edad y finalmente van de .41 a 1.75 años para el caso de los hombre mayores de 60; ahora bien, para el caso de las mujeres las diferencia encontradas fueron de .63 a 1.23 años para el caso de la población femenina de menores de 10 años, de .64 a .67 años para el caso de las mujeres entre los 10 y los 19 años, de .56 a .64 años para las mujeres entre los 20 y los 59 años de edad y de .26 a .57 años para el caso de la población femenina mayor de 60 años.

Las principales diferencias entre género, se encuentran en la región 5 para el caso de los menores de 10 años ya que la esperanza de vida de las mujeres es de 72.13 años contra únicamente 67 años para el caso de la población masculina de la misma edad.

Con este estudio se dejan varias líneas de investigación abiertas, las cuales valdría la pena desarrollar en un futuro para lograr información con la menor varianza posible.

Estudios como este definitivamente nos exigirán una reorientación de las políticas empleadas para mejorar las condiciones de vida de la población y un seguimiento a grupos específicos para determinar las medidas que se tomarán a futuro y así lograr una pirámide poblacional lo suficientemente balanceada.

V. Bibliografía

1. Benítez Z. (1984), "Los factores y el cambio demográfico en México", Siglo Veintiuno editores, S.A, primera edición.

2. Tuirán R., Partida V., Mojarro O. y Zúñiga E., “Tendencias y Perspectivas de la fecundidad”; CONAPO; Disponible en:
<http://www.conapo.gob.mx/publicaciones/sdm/sdm2002/03.pdf>; Consultado el 10 de enero de 2009.
3. Ordorica M.; “El perfil presente y futuro de la población”; Seminarios COLMEX; Disponible en:
<http://seminarios.colmex.mx/videoseminario/1marzo/ponencias/manuelordorica.pdf> Consultado el 9 de enero de 2009.
4. Covarrubias Martínez, Laura Edith (2006), “La Heterogeneidad en los niveles de la tasa de mortalidad infantil en México en el año 2000”. Tesis de Licenciatura en Actuaría. México, Instituto Tecnológico Autónomo de México.
5. Anzaldo C. y Prado M. (2006), Índices de marginación; Primera edición.
6. Villagómez P., Bistrain C., “Situación Demográfica Nacional”, Publicaciones Recientes del CONAPO, Disponible en:
<http://www.conapo.gob.mx/publicaciones/sdm/sdm2008/01.pdf> . Consultado el 22 de Mayo de 2009.
7. Navarrete F., “Los pueblos indígenas de México”, Comisión Nacional para el Desarrollo de los Pueblos Indígenas. Disponible en:
http://www.cdi.gob.mx/index.php?option=com_docman&Itemid=24 . Consultado el 9 de junio de 2009.
8. Rivera A., “Perfil sociodemográfico de la población ocupada en el sector primario y su distribución territorial”, Publicaciones Recientes del CONAPO,

- Disponible en: <http://www.conapo.gob.mx/publicaciones/sdm/sdm2008/10.pdf>
Consultado el 22 de mayo de 2009.
9. Aguirre A., "La Mortalidad Infantil en los años noventa", Boletín Oficial del G.A.A.P.P, Año 5, Número 28, Octubre 1998.
 10. Regiones Socioeconómicas del INEGI Disponible en: www.inegi.gob.mx
Fecha de consulta 15 de enero de 2009
 11. Bogue, D.J., Principles of Demography, New York: John Wiley & Sons, 1969.
 12. "Seguir Avanzando en la transición demográfica"; Publicaciones Recientes CONAPO; Disponible en: http://www.conapo.gob.mx/pnp/pnp_pdf/041.pdf;
Consultado el 09 de enero de 2009.
 13. Situación demográfica en México, 1998; "La Salud en México"; CONAPO;
Disponible en: <http://www.conapo.gob.mx/publicaciones/sdm/sdm1998/03.pdf> ;
Consultado el 10 de enero de 2009.
 14. SINAIS, Disponible en:
<http://sinais.salud.gob.mx/basesdedatos/index.html#defunciones>, Consultado el día 17 de marzo de 2009,
 15. INEGI, Disponible en:
http://www.inegi.org.mx/est/librerias/tabulados.asp?tabulado=tab_po02b&s=est&c=11514, Consultado el día 9 de enero de 2009,
 16. Secretaría de Salud, Dirección General de Información en Salud, Disponible en: <http://www.inegi.org.mx/inegi/default.aspx?s=est&c=124>; Consultado el 21 de junio de 2009

17. Centro de Estudios de las Finanzas Públicas de la H. Cámara de Diputados datos con base en información del Cuarto Informe de Gobierno, 2004. Poder Ejecutivo Federal, Disponible en:
http://www.cefp.gob.mx/intr/bancosdeinformacion/estatales/indicadores_socioeconomicos/indicadoresocioeconomicos.html , Consultado el 15 de mayo de 2009.
18. Estimaciones del INEGI con base en el II Censo de Población y Vivienda 2005. Disponible en:
<http://www.inegi.org.mx/est/contenidos/espanol/rutinas/ept.asp?t=medu06&s=est&c=3273> , Consultada el 5 de mayo de 2009.
19. Estimaciones del CONEVAL con base en el II Censo de Población y Vivienda 2005 y la Encuesta Nacional de Ingreso y Gasto de los Hogares 2005. Disponible en <http://www.coneval.gob.mx/mapas/NACIONAL/Nacional.pdf>, Consultada el 25 de Abril de 2009.
20. Indicadores, índice y grado de rezago social, estimaciones del CONEVAL con base en el II Censo de Población y Vivienda 2005. Disponible en:
<http://www.coneval.gob.mx/mapas/NACIONAL/Nacional.pdf> consultada el 25 de Abril de 2009.
21. “Estadísticas a propósito del día Internacional de la Alfabetización”, Datos Nacionales, INEGI, Disponible en:
<http://www.inegi.gob.mx/inegi/contenidos/espanol/prensa/contenidos/estadisticas/2002/alfabeti02.pdf> Consulta realizada el 22 de mayo de 2009.

ANEXO 1.

Proyección de Población Censal al 30 de Junio de 2007

Es importante recordar que existen fuentes de error en la información censal. Dichos errores en general tenderían a compensarse entre sí; sin embargo, puede ocurrir que existan errores sistémicos si es que hubiera una tendencia de la gente por contestar incorrectamente en el mismo sentido.

Los errores pueden venir de muchas fuentes, pero se suelen agrupar en:

- Errores de Cobertura: Por ausencia o múltiple contabilización de la información.
- Errores de Respuesta: Cuando el cuestionario está incompleto por ausencia de respuestas o de información sobre uno o varios individuos del hogar.
- Errores en la declaración de edad.
- Errores de Proceso: Propios de la captura y procesamiento de la información.
- Errores de muestreo

Existen métodos para encontrar y corregir errores en la información censal, los cuales no son limitativos y varían de investigador a investigador dependiendo las habilidades de estos para ajustar patrones de diferentes indicadores; sin embargo, existen métodos de corrección de la información universalmente aceptados como lo es el Prorrato de No Especificados.

Al hablar del estudio de la población se deben incluir aquellos procesos que determinan la formación, conservación y la desaparición de ésta; para lo cual se estudian los procesos biológicos, de cohesión y repulsión como lo son la fecundidad, la mortalidad, la migración, la nupcialidad, entre otros para el correcto análisis²⁰.

La ecuación demográfica básica describe el cambio en la población entre dos períodos utilizando los flujos de nacimientos, defunciones y migraciones y se describe de la siguiente manera:

$$P_t = P(t - 1) + N(t - 1, t) - D(t - 1, t) + In(t - 1, t) - Em(t - 1, t)$$

donde:

P_t = Población al tiempo t

$N(t-1, t)$ = Nacimientos registrados entre $t-1$ y t

$D(t-1, t)$ = Defunciones registradas entre $t-1$ y t

$In(t-1, t)$ = Inmigrantes estimados entre $t-1$ y t

$Em(t-1, t)$ = Emigrantes estimados entre $t-1$ y t

Vale la pena mencionar que las fechas de referencia utilizadas por el INEGI para la información en cuestión son, para el CENSO el 14 de Febrero de 2000 y para el CONTEO se utilizó el 17 de Octubre de 2005.

La información contenida en ambas estadísticas incluye una estimación de la población de personas que corresponden a aquellas viviendas sin información de ocupantes. La estimación está distribuida en los renglones de no especificado (NE) de edad por lo que se decidió suavizar esta información mediante el prorrateo de "No especificados" ya que antes de iniciar el cálculo de cualquier

²⁰ Bogue, D.J., Principles of Demography, New York: John Wiley & Sons, 1969.

índice de calidad de la información, se debe realizar el prorrateo de aquellos casos que no declararon su edad.

Sea: u el grupo de edad o la edad y P_u la población de esa edad. Para prorratear los no especificados se siguen los siguientes pasos:

1. Se suman todos los grupos de edad sin incluir a los NE
2. Se calcula el porcentaje que representa la población de cada grupo de edad, es decir:

$$PP_u = \frac{P_u}{\sum_u P_u}$$

3. Se distribuyen los NE usando estos porcentajes $NP_u = P_u + NE * PP_u$ donde NP_u = Población del grupo u o edad u incluyendo NE.
4. Finalmente se redondean las cifras y se ajusta el último grupo de edad o la última edad para garantizar que:

$$\sum_u NP_u = \sum_u P_u + NE$$

Este procedimiento se realiza para hombres y para mujeres y se suman para poder obtener el total de ambos sexos.

Proyección de la población con base a la tasa de crecimiento poblacional. Como ya se mencionó anteriormente son muchas las variables que afectan a la dinámica poblacional, entre ellas está la natalidad, la mortalidad, la migración y la emigración entre otras. Para efectos de este estudio se utilizará el método tradicional de proyección, que consiste en utilizar la información censal y obtener una tasa de crecimiento.

Para proyectar la información de la población entre dos fechas se requiere:

1. Para cada grupo de edad se calcula la tasa efectiva de crecimiento haciendo:

$$t_u = \frac{P_u^{TF}}{P_u^{TI}} - 1$$

Esta tasa es efectiva para todo el período transcurrido entre el tiempo inicial (T I) y el tiempo final (T F).

2. Si lo que buscamos es proyectar la información a una fecha particular, digamos (T M), Necesitamos contar que fracción representan los días transcurridos entre T I y T M (DT = T M - T I) del total de tiempo entre T I y T F (T F - T I)

$$Fracción = \frac{DT}{TF - TI}$$

Entonces la población proyectada a TM sería:

$$P_u^{TM} = P_u^{TI} (1 + t_u)^{fracción}$$

ANEXO 2

Tablas abreviadas de mortalidad por regiones socioeconómicas.

República Mexicana

EDAD	Hombres					
	nq_x	nl_x	d_x	L_x	T_x	e_x
0 años	0.03437	1,000,000	34,368	982,816	70,475,996	70.48
1-4 años	0.00307	965,632	2,965	3,856,596	69,493,180	71.97
5 - 9 años	0.00152	962,666	1,459	4,809,685	65,636,584	68.18
10 - 14 años	0.00185	961,208	1,778	4,801,592	60,826,899	63.28
15 - 19 años	0.00532	959,429	5,103	4,784,389	56,025,306	58.39
20 - 24 años	0.00841	954,326	8,022	4,751,578	51,240,917	53.69
25 - 29 años	0.01048	946,305	9,917	4,706,731	46,489,339	49.13
30 - 34 años	0.01182	936,388	11,070	4,654,264	41,782,608	44.62
35 - 39 años	0.01454	925,318	13,450	4,592,964	37,128,343	40.12
40 - 44 años	0.01872	911,868	17,072	4,516,661	32,535,379	35.68
45 - 49 años	0.02649	894,796	23,702	4,414,727	28,018,718	31.31
50 - 54 años	0.03731	871,094	32,501	4,274,220	23,603,991	27.10
55 - 59 años	0.05446	838,594	45,667	4,078,800	19,329,771	23.05
60 - 64 años	0.07362	792,926	58,379	3,818,684	15,250,971	19.23
65 - 69 años	0.11027	734,547	81,000	3,470,237	11,432,287	15.56
70 - 74 años	0.15891	653,547	103,854	3,008,103	7,962,051	12.18
75 - 79 años	0.22477	549,694	123,556	2,439,578	4,953,948	9.01
80 - 84 años	0.31993	426,137	136,332	1,789,857	2,514,370	5.90
85 + años	1.00000	289,805	289,805	724,513	724,513	2.50

EDAD	Mujeres					
	nq_x	nl_x	d_x	L_x	T_x	e_x
0 años	0.02744	1,000,000	27,443	986,278	75,151,056	75.15
1-4 años	0.00263	972,557	2,557	3,885,114	74,164,777	76.26
5 - 9 años	0.00128	970,000	1,241	4,846,899	70,279,663	72.45
10 - 14 años	0.00129	968,759	1,252	4,840,666	65,432,764	67.54
15 - 19 años	0.00231	967,507	2,237	4,831,944	60,592,097	62.63
20 - 24 años	0.00271	965,271	2,614	4,819,818	55,760,153	57.77
25 - 29 años	0.00327	962,657	3,146	4,805,419	50,940,335	52.92
30 - 34 años	0.00394	959,511	3,776	4,788,114	46,134,916	48.08
35 - 39 años	0.00547	955,735	5,224	4,765,615	41,346,801	43.26
40 - 44 años	0.00870	950,511	8,269	4,731,882	36,581,186	38.49
45 - 49 años	0.01364	942,242	12,853	4,679,077	31,849,304	33.80
50 - 54 años	0.02223	929,389	20,661	4,595,293	27,170,228	29.23
55 - 59 años	0.03680	908,728	33,444	4,460,030	22,574,935	24.84
60 - 64 años	0.05320	875,284	46,566	4,260,005	18,114,905	20.70
65 - 69 años	0.08262	828,718	68,468	3,972,422	13,854,900	16.72
70 - 74 años	0.12436	760,250	94,546	3,564,886	9,882,478	13.00
75 - 79 años	0.18999	665,704	126,478	3,012,325	6,317,592	9.49
80 - 84 años	0.27407	539,226	147,786	2,326,666	3,305,267	6.13
85 + años	1.00000	391,440	391,440	978,601	978,601	2.50

Región 7

Hombres						
EDAD	nqx	Lx	Dx	Lx	Tx	Ex
0 años	0.02855	1,000,000	28,550	985,725	71,033,153	71.03
1-4 años	0.00261	971,450	2,540	3,880,719	70,047,428	72.11
5 - 9 años	0.00125	968,910	1,207	4,841,533	66,166,709	68.29
10 - 14 años	0.00126	967,703	1,219	4,835,469	61,325,176	63.37
15 - 19 años	0.00427	966,484	4,126	4,822,106	56,489,708	58.45
20 - 24 años	0.00688	962,358	6,623	4,795,233	51,667,602	53.69
25 - 29 años	0.00862	955,735	8,240	4,758,076	46,872,369	49.04
30 - 34 años	0.01012	947,495	9,589	4,713,504	42,114,293	44.45
35 - 39 años	0.01345	937,906	12,613	4,658,000	37,400,789	39.88
40 - 44 años	0.01836	925,294	16,989	4,583,997	32,742,789	35.39
45 - 49 años	0.02623	908,305	23,824	4,481,966	28,158,792	31.00
50 - 54 años	0.03675	884,481	32,501	4,341,153	23,676,826	26.77
55 - 59 años	0.05364	851,980	45,698	4,145,655	19,335,673	22.69
60 - 64 años	0.08052	806,282	64,920	3,869,108	15,190,019	18.84
65 - 69 años	0.11837	741,361	87,754	3,487,421	11,320,911	15.27
70 - 74 años	0.16078	653,607	105,088	3,005,314	7,833,490	11.99
75 - 79 años	0.24641	548,519	135,158	2,404,698	4,828,176	8.80
80 - 84 años	0.32743	413,361	135,345	1,728,439	2,423,477	5.86
85 + años	1.00000	278,015	278,015	695,038	695,038	2.50

Mujeres						
EDAD	nqx	Lx	Dx	Lx	Tx	ex
0 años	0.02311	1,000,000	23,112	988,444	75,621,827	75.62
1-4 años	0.00235	976,888	2,298	3,902,956	74,633,382	76.40
5 - 9 años	0.00101	974,590	980	4,870,500	70,730,426	72.57
10 - 14 años	0.00103	973,610	1,004	4,865,541	65,859,926	67.65
15 - 19 años	0.00203	972,606	1,972	4,858,101	60,994,385	62.71
20 - 24 años	0.00239	970,634	2,321	4,847,368	56,136,283	57.83
25 - 29 años	0.00316	968,313	3,057	4,833,922	51,288,916	52.97
30 - 34 años	0.00317	965,256	3,059	4,818,632	46,454,993	48.13
35 - 39 años	0.00469	962,197	4,516	4,799,693	41,636,361	43.27
40 - 44 años	0.00808	957,680	7,738	4,769,059	36,836,668	38.46
45 - 49 años	0.01262	949,943	11,989	4,719,742	32,067,609	33.76
50 - 54 años	0.02171	937,954	20,364	4,638,859	27,347,867	29.16
55 - 59 años	0.03692	917,590	33,879	4,503,251	22,709,009	24.75
60 - 64 años	0.05495	883,711	48,564	4,297,145	18,205,757	20.60
65 - 69 años	0.08640	835,147	72,158	3,995,341	13,908,613	16.65
70 - 74 años	0.12103	762,989	92,347	3,584,080	9,913,272	12.99
75 - 79 años	0.19405	670,643	130,139	3,027,866	6,329,192	9.44
80 - 84 años	0.27843	540,504	150,491	2,326,293	3,301,326	6.11
85 + años	1.00000	390,013	390,013	975,033	975,033	2.50

Región 6

Hombres						
EDAD	nqx	Lx	dx	Lx	Tx	ex
0 años	0.02762	1,000,000	27,622	986,189	71,192,289	71.19
1-4 años	0.00252	972,378	2,449	3,884,616	70,206,100	72.20
5 - 9 años	0.00160	969,930	1,554	4,845,764	66,321,484	68.38
10 - 14 años	0.00169	968,376	1,635	4,837,792	61,475,720	63.48
15 - 19 años	0.00505	966,741	4,882	4,821,499	56,637,928	58.59
20 - 24 años	0.00715	961,859	6,876	4,792,103	51,816,429	53.87
25 - 29 años	0.00875	954,983	8,360	4,754,014	47,024,325	49.24
30 - 34 años	0.00964	946,623	9,121	4,710,312	42,270,311	44.65
35 - 39 años	0.01236	937,502	11,587	4,658,542	37,559,999	40.06
40 - 44 años	0.01594	925,915	14,758	4,592,681	32,901,458	35.53
45 - 49 años	0.02351	911,157	21,422	4,502,232	28,308,777	31.07
50 - 54 años	0.03507	889,735	31,207	4,370,658	23,806,545	26.76
55 - 59 años	0.05378	858,528	46,169	4,177,217	19,435,886	22.64
60 - 64 años	0.07821	812,359	63,531	3,902,967	15,258,669	18.78
65 - 69 años	0.11687	748,828	87,515	3,525,350	11,355,702	15.16
70 - 74 años	0.17235	661,312	113,980	3,021,612	7,830,352	11.84
75 - 79 años	0.24116	547,332	131,993	2,406,681	4,808,740	8.79
80 - 84 años	0.34333	415,340	142,598	1,720,204	2,402,059	5.78
85 + años	1.00000	272,742	272,742	681,855	681,855	2.50

Mujeres						
EDAD	nqx	Lx	dx	Lx	Tx	ex
0 años	0.02198	1,000,000	21,979	989,010	75,550,174	75.55
1-4 años	0.00201	978,021	1,969	3,908,145	74,561,163	76.24
5 - 9 años	0.00111	976,052	1,080	4,877,558	70,653,018	72.39
10 - 14 años	0.00129	974,972	1,257	4,871,715	65,775,460	67.46
15 - 19 años	0.00209	973,714	2,040	4,863,472	60,903,745	62.55
20 - 24 años	0.00233	971,674	2,268	4,852,702	56,040,273	57.67
25 - 29 años	0.00275	969,406	2,667	4,840,365	51,187,571	52.80
30 - 34 años	0.00332	966,740	3,214	4,825,664	46,347,206	47.94
35 - 39 años	0.00503	963,526	4,845	4,805,518	41,521,543	43.09
40 - 44 años	0.00764	958,681	7,324	4,775,095	36,716,025	38.30
45 - 49 años	0.01266	951,357	12,047	4,726,667	31,940,930	33.57
50 - 54 años	0.02206	939,310	20,725	4,644,738	27,214,262	28.97
55 - 59 años	0.03783	918,585	34,752	4,506,046	22,569,524	24.57
60 - 64 años	0.05697	883,833	50,356	4,293,276	18,063,478	20.44
65 - 69 años	0.08559	833,477	71,341	3,989,032	13,770,203	16.52
70 - 74 años	0.12982	762,136	98,938	3,563,335	9,781,170	12.83
75 - 79 años	0.19778	663,198	131,165	2,988,078	6,217,835	9.38
80 - 84 años	0.28588	532,033	152,098	2,279,920	3,229,757	6.07
85 + años	1.00000	379,935	379,935	949,837	949,837	2.50

Región 5

Hombres						
EDAD	nqx	Lx	dx	Lx	Tx	Ex
0 años	0.03155	1,000,000	31,553	984,223	69,547,941	69.55
1-4 años	0.00287	968,447	2,779	3,868,229	68,563,717	70.80
5 - 9 años	0.00139	965,668	1,346	4,824,973	64,695,488	67.00
10 - 14 años	0.00177	964,322	1,703	4,817,351	59,870,515	62.09
15 - 19 años	0.00534	962,619	5,136	4,800,254	55,053,164	57.19
20 - 24 años	0.00931	957,483	8,917	4,765,122	50,252,910	52.48
25 - 29 años	0.01150	948,566	10,911	4,715,552	45,487,788	47.95
30 - 34 años	0.01255	937,655	11,770	4,658,849	40,772,236	43.48
35 - 39 años	0.01549	925,885	14,345	4,593,561	36,113,387	39.00
40 - 44 años	0.02009	911,540	18,310	4,511,923	31,519,826	34.58
45 - 49 años	0.02925	893,229	26,125	4,400,833	27,007,903	30.24
50 - 54 años	0.04212	867,104	36,524	4,244,209	22,607,070	26.07
55 - 59 años	0.06058	830,580	50,318	4,027,104	18,362,861	22.11
60 - 64 años	0.08467	780,262	66,064	3,736,151	14,335,757	18.37
65 - 69 años	0.12300	714,198	87,847	3,351,376	10,599,606	14.84
70 - 74 años	0.18328	626,352	114,799	2,844,762	7,248,230	11.57
75 - 79 años	0.25784	511,553	131,899	2,228,019	4,403,468	8.61
80 - 84 años	0.35398	379,655	134,392	1,562,293	2,175,449	5.73
85 + años	1.00000	245,263	245,263	613,156	613,156	2.50

Mujeres						
EDAD	nqx	Lx	dx	Lx	Tx	Ex
0 años	0.02283	1,000,000	22,834	988,583	75,224,240	75.22
1-4 años	0.00218	977,166	2,135	3,904,394	74,235,657	75.97
5 - 9 años	0.00115	975,031	1,124	4,872,345	70,331,263	72.13
10 - 14 años	0.00118	973,907	1,149	4,866,663	65,458,917	67.21
15 - 19 años	0.00276	972,758	2,681	4,857,088	60,592,254	62.29
20 - 24 años	0.00267	970,077	2,589	4,843,911	55,735,166	57.45
25 - 29 años	0.00340	967,488	3,286	4,829,224	50,891,255	52.60
30 - 34 años	0.00369	964,202	3,558	4,812,114	46,062,031	47.77
35 - 39 años	0.00573	960,644	5,507	4,789,450	41,249,917	42.94
40 - 44 años	0.00902	955,136	8,616	4,754,142	36,460,467	38.17
45 - 49 años	0.01473	946,520	13,939	4,697,754	31,706,326	33.50
50 - 54 años	0.02372	932,581	22,120	4,607,608	27,008,571	28.96
55 - 59 años	0.03892	910,462	35,438	4,463,715	22,400,963	24.60
60 - 64 años	0.05386	875,024	47,128	4,257,300	17,937,248	20.50
65 - 69 años	0.08302	827,896	68,729	3,967,658	13,679,947	16.52
70 - 74 años	0.13263	759,167	100,685	3,544,124	9,712,290	12.79
75 - 79 años	0.19633	658,482	129,279	2,969,215	6,168,166	9.37
80 - 84 años	0.29103	529,204	154,015	2,260,980	3,198,951	6.04
85 + años	1.00000	375,188	375,188	937,971	937,971	2.50

Región 4

Hombres						
EDAD	nqx	Lx	dx	Lx	Tx	Ex
0 años	0.03278	1,000,000	32,779	983,611	70,959,384	70.96
1-4 años	0.00309	967,221	2,984	3,862,918	69,975,773	72.35
5 - 9 años	0.00139	964,237	1,340	4,817,836	66,112,855	68.56
10 - 14 años	0.00159	962,897	1,531	4,810,659	61,295,019	63.66
15 - 19 años	0.00530	961,366	5,099	4,794,085	56,484,360	58.75
20 - 24 años	0.00746	956,267	7,133	4,763,504	51,690,276	54.05
25 - 29 años	0.00964	949,134	9,147	4,722,805	46,926,771	49.44
30 - 34 años	0.01101	939,987	10,349	4,674,065	42,203,967	44.90
35 - 39 años	0.01299	929,639	12,072	4,618,012	37,529,902	40.37
40 - 44 años	0.01736	917,566	15,928	4,548,013	32,911,889	35.87
45 - 49 años	0.02441	901,639	22,013	4,453,161	28,363,877	31.46
50 - 54 años	0.03619	879,626	31,830	4,318,554	23,910,716	27.18
55 - 59 años	0.05229	847,796	44,328	4,128,161	19,592,161	23.11
60 - 64 años	0.07188	803,468	57,753	3,872,959	15,464,001	19.25
65 - 69 años	0.10969	745,715	81,798	3,524,081	11,591,042	15.54
70 - 74 años	0.15829	663,917	105,091	3,056,860	8,066,961	12.15
75 - 79 años	0.22722	558,827	126,978	2,476,687	5,010,101	8.97
80 - 84 años	0.32671	431,848	141,090	1,806,517	2,533,413	5.87
85 + años	1.00000	290,758	290,758	726,896	726,896	2.50

Mujeres						
EDAD	nqx	Lx	dx	Lx	Tx	ex
0 años	0.02621	1,000,000	26,212	986,894	75,421,566	75.42
1-4 años	0.00257	973,788	2,499	3,890,154	74,434,672	76.44
5 - 9 años	0.00120	971,289	1,166	4,853,528	70,544,518	72.63
10 - 14 años	0.00123	970,123	1,198	4,847,618	65,690,990	67.71
15 - 19 años	0.00206	968,925	1,999	4,839,626	60,843,372	62.79
20 - 24 años	0.00260	966,926	2,518	4,828,336	56,003,745	57.92
25 - 29 años	0.00296	964,408	2,852	4,814,912	51,175,410	53.06
30 - 34 años	0.00378	961,556	3,632	4,798,703	46,360,498	48.21
35 - 39 años	0.00513	957,925	4,919	4,777,327	41,561,795	43.39
40 - 44 años	0.00843	953,006	8,035	4,744,943	36,784,468	38.60
45 - 49 años	0.01282	944,971	12,118	4,694,561	32,039,525	33.91
50 - 54 años	0.02090	932,853	19,499	4,615,519	27,344,964	29.31
55 - 59 años	0.03494	913,354	31,912	4,486,991	22,729,445	24.89
60 - 64 años	0.05321	881,442	46,902	4,289,956	18,242,453	20.70
65 - 69 años	0.08074	834,540	67,378	4,004,256	13,952,497	16.72
70 - 74 años	0.12539	767,162	96,192	3,595,330	9,948,241	12.97
75 - 79 años	0.18842	670,970	126,421	3,038,797	6,352,911	9.47
80 - 84 años	0.28280	544,549	154,000	2,337,743	3,314,114	6.09
85 + años	1.00000	390,548	390,548	976,371	976,371	2.50

Región 3

Hombres						
EDAD	nqx	Lx	dx	Lx	Tx	ex
0 años	0.03699	1,000,000	36,991	981,504	70,483,441	70.48
1-4 años	0.00298	963,009	2,874	3,846,287	69,501,937	72.17
5 - 9 años	0.00170	960,135	1,631	4,796,597	65,655,650	68.38
10 - 14 años	0.00238	958,504	2,277	4,786,827	60,859,053	63.49
15 - 19 años	0.00698	956,227	6,673	4,764,453	56,072,226	58.64
20 - 24 años	0.01056	949,554	10,029	4,722,699	51,307,773	54.03
25 - 29 años	0.01101	939,525	10,345	4,671,763	46,585,074	49.58
30 - 34 años	0.01345	929,180	12,497	4,614,658	41,913,311	45.11
35 - 39 años	0.01473	916,683	13,506	4,549,651	37,298,653	40.69
40 - 44 años	0.01929	903,177	17,422	4,472,332	32,749,002	36.26
45 - 49 años	0.02484	885,756	22,001	4,373,776	28,276,669	31.92
50 - 54 años	0.03409	863,755	29,446	4,245,159	23,902,893	27.67
55 - 59 años	0.05115	834,309	42,678	4,064,850	19,657,734	23.56
60 - 64 años	0.06900	791,631	54,619	3,821,607	15,592,885	19.70
65 - 69 años	0.09947	737,012	73,311	3,501,783	11,771,278	15.97
70 - 74 años	0.14833	663,701	98,447	3,072,389	8,269,495	12.46
75 - 79 años	0.20807	565,254	117,612	2,532,242	5,197,106	9.19
80 - 84 años	0.30938	447,642	138,491	1,891,985	2,664,864	5.95
85 + años	1.00000	309,152	309,152	772,879	772,879	2.50

Mujeres						
EDAD	nqx	Lx	dx	Lx	Tx	ex
0 años	0.03019	1,000,000	30,195	984,903	75,256,951	75.26
1-4 años	0.00243	969,805	2,352	3,874,517	74,272,048	76.58
5 - 9 años	0.00130	967,453	1,262	4,834,111	70,397,531	72.77
10 - 14 años	0.00151	966,191	1,462	4,827,301	65,563,421	67.86
15 - 19 años	0.00225	964,729	2,171	4,818,221	60,736,120	62.96
20 - 24 años	0.00250	962,559	2,409	4,806,772	55,917,899	58.09
25 - 29 años	0.00318	960,150	3,057	4,793,107	51,111,127	53.23
30 - 34 años	0.00396	957,093	3,792	4,775,986	46,318,020	48.39
35 - 39 años	0.00533	953,301	5,078	4,753,811	41,542,035	43.58
40 - 44 años	0.00825	948,223	7,827	4,721,550	36,788,223	38.80
45 - 49 años	0.01368	940,397	12,864	4,669,824	32,066,673	34.10
50 - 54 años	0.02124	927,533	19,705	4,588,403	27,396,849	29.54
55 - 59 años	0.03511	907,828	31,871	4,459,464	22,808,446	25.12
60 - 64 años	0.05284	875,957	46,286	4,264,071	18,348,982	20.95
65 - 69 años	0.07830	829,671	64,966	3,985,940	14,084,911	16.98
70 - 74 años	0.11783	764,705	90,107	3,598,257	10,098,971	13.21
75 - 79 años	0.18382	674,598	124,005	3,062,976	6,500,714	9.64
80 - 84 años	0.25126	550,592	138,341	2,407,110	3,437,738	6.24
85 + años	1.00000	412,251	412,251	1,030,629	1,030,629	2.50

Región 2

Hombres						
EDAD	nqx	Lx	Dx	Lx	Tx	Ex
0 años	0.03779	1,000,000	37,787	981,106	70,161,011	70.16
1-4 años	0.00322	962,213	3,098	3,842,655	69,179,905	71.90
5 - 9 años	0.00149	959,115	1,430	4,791,999	65,337,250	68.12
10 - 14 años	0.00208	957,685	1,994	4,783,438	60,545,252	63.22
15 - 19 años	0.00480	955,691	4,585	4,766,992	55,761,814	58.35
20 - 24 años	0.00851	951,106	8,095	4,735,294	50,994,822	53.62
25 - 29 años	0.01103	943,012	10,402	4,689,053	46,259,528	49.06
30 - 34 años	0.01212	932,610	11,302	4,634,793	41,570,475	44.57
35 - 39 años	0.01549	921,307	14,268	4,570,867	36,935,682	40.09
40 - 44 años	0.01967	907,040	17,838	4,490,602	32,364,814	35.68
45 - 49 años	0.02833	889,201	25,193	4,383,022	27,874,213	31.35
50 - 54 años	0.03984	864,008	34,418	4,233,992	23,491,191	27.19
55 - 59 años	0.05626	829,589	46,669	4,031,273	19,257,198	23.21
60 - 64 años	0.07130	782,920	55,822	3,775,045	15,225,926	19.45
65 - 69 años	0.10821	727,098	78,680	3,438,790	11,450,881	15.75
70 - 74 años	0.15384	648,418	99,755	2,992,700	8,012,090	12.36
75 - 79 años	0.21320	548,662	116,976	2,450,873	5,019,390	9.15
80 - 84 años	0.31001	431,687	133,827	1,823,867	2,568,517	5.95
85 + años	1.00000	297,860	297,860	744,650	744,650	2.50

Mujeres						
EDAD	nqx	Lx	Dx	Lx	Tx	Ex
0 años	0.03100	1,000,000	31,003	984,499	74,743,956	74.74
1-4 años	0.00286	968,997	2,776	3,870,436	73,759,458	76.12
5 - 9 años	0.00150	966,221	1,445	4,827,493	69,889,022	72.33
10 - 14 años	0.00132	964,776	1,271	4,820,703	65,061,529	67.44
15 - 19 años	0.00239	963,505	2,298	4,811,780	60,240,826	62.52
20 - 24 años	0.00278	961,207	2,669	4,799,363	55,429,046	57.67
25 - 29 años	0.00356	958,538	3,412	4,784,161	50,629,683	52.82
30 - 34 años	0.00416	955,126	3,970	4,765,705	45,845,522	48.00
35 - 39 años	0.00555	951,156	5,275	4,742,591	41,079,817	43.19
40 - 44 años	0.00941	945,881	8,901	4,707,151	36,337,225	38.42
45 - 49 años	0.01424	936,980	13,338	4,651,553	31,630,075	33.76
50 - 54 años	0.02277	923,642	21,035	4,565,620	26,978,522	29.21
55 - 59 años	0.03765	902,606	33,987	4,428,064	22,412,902	24.83
60 - 64 años	0.05202	868,619	45,182	4,230,141	17,984,838	20.71
65 - 69 años	0.08232	823,437	67,786	3,947,719	13,754,697	16.70
70 - 74 años	0.12455	755,651	94,115	3,542,967	9,806,978	12.98
75 - 79 años	0.19033	661,536	125,908	2,992,909	6,264,012	9.47
80 - 84 años	0.27859	535,628	149,221	2,305,086	3,271,102	6.11
85 + años	1.00000	386,407	386,407	966,016	966,016	2.50

Región 1

EDAD	Hombres					
	Nqx	Lx	Dx	Lx	Tx	Ex
0 años	0.04060	1,000,000	40,595	979,702	69,781,963	69.78
1-4 años	0.00403	959,405	3,869	3,829,881	68,802,261	71.71
5 - 9 años	0.00175	955,536	1,675	4,773,491	64,972,380	68.00
10 - 14 años	0.00195	953,861	1,864	4,764,644	60,198,889	63.11
15 - 19 años	0.00537	951,997	5,109	4,747,212	55,434,245	58.23
20 - 24 años	0.01022	946,888	9,675	4,710,251	50,687,033	53.53
25 - 29 años	0.01416	937,213	13,267	4,652,897	45,976,782	49.06
30 - 34 años	0.01604	923,946	14,824	4,582,669	41,323,886	44.73
35 - 39 años	0.02006	909,122	18,238	4,500,013	36,741,217	40.41
40 - 44 años	0.02262	890,884	20,150	4,404,043	32,241,204	36.19
45 - 49 años	0.03124	870,734	27,204	4,285,659	27,837,161	31.97
50 - 54 años	0.03675	843,530	31,002	4,140,145	23,551,502	27.92
55 - 59 años	0.05456	812,528	44,329	3,951,818	19,411,357	23.89
60 - 64 años	0.06275	768,199	48,207	3,720,480	15,459,539	20.12
65 - 69 años	0.09797	719,993	70,539	3,423,616	11,739,059	16.30
70 - 74 años	0.13874	649,454	90,104	3,022,009	8,315,443	12.80
75 - 79 años	0.19425	559,350	108,656	2,525,109	5,293,434	9.46
80 - 84 años	0.27153	450,694	122,375	1,947,530	2,768,326	6.14
85 + años	1.00000	328,318	328,318	820,796	820,796	2.50

EDAD	Mujeres					
	nqx	Lx	Dx	Lx	Tx	ex
0 años	0.03284	1,000,000	32,836	983,582	74,394,584	74.39
1-4 años	0.00394	967,164	3,812	3,861,035	73,411,002	75.90
5 - 9 años	0.00152	963,353	1,463	4,813,106	69,549,968	72.20
10 - 14 años	0.00136	961,889	1,312	4,806,168	64,736,862	67.30
15 - 19 años	0.00276	960,578	2,649	4,796,266	59,930,694	62.39
20 - 24 años	0.00385	957,929	3,685	4,780,430	55,134,429	57.56
25 - 29 años	0.00425	954,243	4,057	4,761,075	50,353,999	52.77
30 - 34 años	0.00595	950,187	5,656	4,736,793	45,592,923	47.98
35 - 39 años	0.00748	944,531	7,066	4,704,988	40,856,130	43.26
40 - 44 años	0.01040	937,464	9,747	4,662,954	36,151,143	38.56
45 - 49 años	0.01574	927,717	14,599	4,602,088	31,488,189	33.94
50 - 54 años	0.02478	913,118	22,623	4,509,034	26,886,101	29.44
55 - 59 años	0.03741	890,496	33,315	4,369,190	22,377,067	25.13
60 - 64 años	0.04846	857,181	41,537	4,182,060	18,007,876	21.01
65 - 69 años	0.08400	815,643	68,511	3,906,937	13,825,817	16.95
70 - 74 años	0.11854	747,132	88,564	3,514,250	9,918,879	13.28
75 - 79 años	0.17903	658,568	117,900	2,998,090	6,404,630	9.73
80 - 84 años	0.23988	540,668	129,694	2,379,105	3,406,540	6.30
85 + años	1.00000	410,974	410,974	1,027,435	1,027,435	2.50

Caracterización de los depósitos a la vista de la banca comercial Mexicana y administración de su riesgo de liquidez a partir de un Modelo de Tasas de Interés de Mercado

Leticia Sigrid González Méndez
Subdirector de Administración de Riesgo de Liquidez, BBVA Bancomer
Tel. 56 21 93 65; ls.gonzalez@bbva.bancomer.com

Mercedes Gregorio Domínguez
Departamento de Actuaría, ITAM
mercedes@itam.mx

Resumen

Los depósitos a la vista, una de las fuentes de financiamiento más importantes de la banca, conllevan de forma natural riesgo de liquidez por la posibilidad que ofrecen a sus titulares de un retiro total o parcial. En este trabajo, se prueba un marco cuantitativo predefinido para caracterizar el volumen y administrar el riesgo de liquidez de las cuentas de Cheques y Ahorro en Pesos de la banca múltiple.

I. Introducción

El negocio típico de un banco comercial es aceptar depósitos y otorgar créditos, siendo su ingreso el diferencial entre lo que recibe por los créditos otorgados y el costo de los depósitos recibidos. Los depósitos implican una fuente de recursos para un banco, pero a su vez, también conllevan una obligación contractual de devolver dichos recursos a sus propietarios. En particular, existe un tipo de

depósitos que pueden ser retirados, total o parcialmente, en cualquier momento por sus titulares, y se conocen como depósitos con vencimiento indeterminado.

En la banca comercial de México, un tipo de depósito con vencimiento indeterminado que conforma una porción muy importante de la captación del total de recursos de los bancos son los depósitos a la vista, en sus modalidades de cuentas de Cheques y cuentas de Ahorro (en marzo de 2009, el 34% de la captación total de recursos de la banca múltiple mexicana correspondía a estos productos). La incertidumbre en el comportamiento de estos depósitos abre riesgos para el negocio bancario, por lo cual requieren de un cuidadoso estudio para conocer la manera más adecuada y eficiente de manejarlos.

El objetivo general de este trabajo es aplicar el marco cuantitativo de Kalkbrener y Willing [9] para la caracterización de los depósitos a la vista en pesos del Sistema Bancario Mexicano (Chequeras con intereses, Chequeras sin intereses y Ahorro) y administración de su riesgo de liquidez.

Los objetivos particulares de este trabajo son:

- Caracterizar las tasas de interés de mercado gubernamentales, mediante la estructura de tasas de interés de Heath-Jarrow-Morton (realizar la calibración mediante un Análisis de Componentes Principales). Una vez caracterizadas, simular 50 escenarios de tasas.
- Caracterizar los volúmenes de los depósitos a la vista, considerando su función de tendencia y el proceso Ornstein-Uhlenbeck (el cual tiene tres

argumentos estocásticos, de los cuales dos provienen del modelo de tasas de interés de mercado). Una vez caracterizadas, simular 50 escenarios de volúmenes, tomando en cuenta las simulaciones de los procesos estocásticos de las tasas de interés de mercado.

- Para administrar el riesgo de liquidez de los depósitos a la vista, estimar su Estructura de Plazos de Liquidez con base en la estructura definida por Kalkbrener et al. [9] con un nivel de significancia de 1%.

Cabe señalar que el trabajo de forma intencional no se realiza en los depósitos a la vista de algún o algunos bancos en particular, con el objeto de no generar alguna percepción del estado de liquidez de la entidad analizada (i.e. por razones de confidencialidad).

II. Administración del Riesgo de Liquidez de los Depósitos a la Vista

La liquidez se ha definido de diversas formas; por mencionar algunas (Uyemura y van Deventer [12]):

“ La liquidez es tener acceso al suficiente efectivo para mantener el negocio normal en cualquier ocasión ”.

“ La liquidez es la habilidad de conseguir (o pedir prestado) fondos a un precio razonable en cualquier ocasión “.

“ La liquidez da tiempo para resolver los problemas ”.

De acuerdo con el Comité de Supervisión Bancaria de Basilea [4] del Banco de Pagos Internacionales, la liquidez es la capacidad de un banco para financiar aumentos en su volumen de activos y para hacer frente a sus obligaciones de pago al vencer éstas, sin incurrir en pérdidas inaceptables. El papel crucial que desempeñan los bancos en el proceso de transformación de vencimientos, captando depósitos a corto plazo y concediendo créditos a largo plazo, les hace intrínsecamente vulnerables al riesgo de liquidez, tanto al propio de cada institución como al que afecta al conjunto del mercado.

La liquidez se trata de un asunto de importancia sistémica, dado que, eventualmente, la mayor parte de las quiebras bancarias han sido ocasionadas o han contribuido a una crisis sistémica de liquidez (Matz y Neu [10]).

La crisis financiera iniciada a mediados de 2007 hizo evidente la importancia de la liquidez en el funcionamiento de los mercados financieros y del sector bancario. Antes de ella, los mercados de activos operaban de manera dinámica y pujante, con una gran disponibilidad de financiación a un costo reducido; con la crisis, las reacciones en los mercados revelaron la rapidez con que la liquidez puede agotarse y puso de manifiesto que su escasez puede prolongarse durante bastante tiempo. El sistema bancario padeció fuertes tensiones, que exigieron la intervención de los bancos centrales en apoyo tanto del funcionamiento de los mercados monetarios como de determinadas instituciones, en algunos casos.

Antes de la crisis actual, el principal documento que delineaba las prácticas a considerar en la administración del riesgo de liquidez era “Sound Practices for Managing Liquidity in Banking Organisations” del Comité de Supervisión Bancaria de Basilea, publicado en febrero de 2000 [2]. Más tarde, con el acuerdo de capital de Basilea II²¹ en 2006 [3], la administración del riesgo de liquidez se integra al Pilar II, aunque con un enfoque secundario. Ahora, principalmente tras las experiencias y lecciones de la crisis financiera actual, en varias partes del mundo se han publicado documentos que resaltan la importancia de la administración del riesgo de liquidez para los bancos.

De esta forma, el Comité de Supervisión Bancaria de Basilea realiza una revisión de fondo del documento de liquidez de 2000 [2], publicando “Principles for Sound Liquidity Risk Management and Supervision” en septiembre de 2008 [4]. Uno de los principios señala: “El banco deberá contar con un adecuado proceso de identificación, medición, vigilancia y control del riesgo de liquidez. Este proceso deberá incluir un marco robusto que ofrezca una proyección completa de los flujos de caja resultantes de los activos, pasivos y partidas fuera de balance para una serie de horizontes temporales relevantes.” En particular, al estimar los flujos de caja procedentes de los pasivos del banco, el documento indica que se debe evaluar la permanencia (*stickiness*) de las fuentes de financiación.

²¹ El marco de Basilea II trata de alinear los requerimientos de adecuación de capital con todos los riesgos materiales subyacentes de los bancos internacionalmente activos, incluyendo al riesgo de liquidez. En particular, Basilea II señala que los bancos deben contar con sistemas adecuados para medir, monitorear y medir el riesgo de liquidez (Basel Committee on Banking Supervision [3]).

El principal problema en la administración de riesgos de los pasivos con vencimiento indeterminado es la asignación de un perfil de vencimiento; esto es, preveer la fecha en que serán retirados de las cuentas de depósito y los importes de los retiros con el objeto de contar con el suficiente efectivo en el momento preciso para satisfacer dichos retiros. Algunas formas comunes que suelen elegir los bancos para esto son (Choudhry [5]):

- Suponer que todos los depósitos vencen en una fecha futura determinada con base en el horizonte de tiempo preferido por el banco.
- Suponer una tasa de retiro de los depósitos, ya sea anual, mensual, etc.
- Dividir el saldo de los depósitos en estables/permanentes e inestables. El importe permanente se determina con base en un análisis de la volatilidad total del balance en observaciones históricas. El excedente sobre el saldo estable se maneja como una deuda de muy corto plazo.
- Realizar proyecciones basadas en variables observadas correlacionadas con el volumen de los depósitos, como el crecimiento económico.

III. Marco Metodológico

III.1 Caracterización de las tasas de interés de mercado

Un modelo de tasas de interés viable debe excluir posibilidades de arbitraje (James y Webber [8]). Cuando no hay oportunidades de arbitraje, entonces existen las probabilidades neutrales al riesgo (Neftci [11]); a su vez, bajo las

probabilidades neutrales al riesgo, las esperanzas de los flujos futuros de los activos riesgosos, traídos a valor presente con la tasa libre de riesgo, son martingalas. Una martingala es una variable aleatoria, cuyo mejor pronóstico de valores futuros no observados es la observación más reciente (Neftci [11]).

Sea $(\Omega, \mathcal{F}, \mathcal{P})$ un espacio de probabilidad, y considérese el tiempo continuo, denotándolo mediante t , se define $W_t = (W_t^1, W_t^2, W_t^3)$ como un movimiento browniano tridimensional, donde \mathcal{F}_t es su filtración. Los dos primeros componentes (W_t^1, W_t^2) son necesarios para definir a las tasas de interés de mercado y se denotan \bar{W}_t ; el tercer componente W_t^3 se usa para definir al volumen de los depósitos. P^- es la distribución de probabilidad de \bar{W}_t , y F_t^- es la filtración generada por \bar{W}_t .

Se desea convertir en martingalas a las tasas de mercado; para esto, se necesita una Medida de Martingala Equivalente (EMM) a P^- , la cual se denominará \bar{Q} ²².

III.2 Modelo para la Estructura de tasas de interés de Heath-Jarrow-Morton

Sea $[0, \tau]$, el horizonte de tiempo para toda $0 \leq t \leq T \leq \tau$, con τ fija, se define:

²² La EMM \bar{Q} es una distribución de probabilidad que concuerda con P^- en relación a los eventos que son imposibles.

$f_t(T)$ como la tasa forward instantánea, es decir, la tasa al tiempo t de una inversión instantánea que empieza en el tiempo T y termina después de un período infinitesimal dt ;

$B_t(T)$ como el precio al tiempo t de un bono cupón cero con vencimiento en T , el cual en términos de tasas forward instantáneas se especifica así:

$$B_t(T) = \exp\left(-\int_t^T f_t(u)du\right); \quad (1)$$

r_t como la tasa corta o la tasa de interés teórica obtenida por una inversión instantánea al tiempo t en el intervalo de tiempo dt , cumpliéndose que:

$$r_t = f_t(t); \quad (2)$$

q_t como el factor de valor presente al tiempo t , es decir, el valor de mercado al tiempo t de una inversión de \$1 a la tasa corta, renovada continuamente, el cual en términos de tasas cortas se especifica así:

$$q_t = \exp\left(\int_0^t r_u du\right), \text{ siendo estrictamente positivo.} \quad (3)$$

Para caracterizar a las tasas de interés de mercado, se considera la aproximación de Heath, Jarrow y Morton (HJM), la cual modela la dinámica de las tasas forward instantáneas bajo P^- de la siguiente manera (James et al. [8]):

$$f_t(T) = f_0(T) + \int_0^t \alpha_u(T) du + \int_0^t \sigma_u(T) d\bar{W}_u, \quad (4)$$

donde $f_0(T)$ es la tasa forward instantánea inicial no aleatoria.

En términos de ecuaciones diferenciales, el modelo HJM se especifica así:

$$df_t(T) = \alpha_t(T) dt + \sigma_t(T) d\bar{W}_t. \quad (5)$$

En las ecuaciones (4) y (5), α y σ conforman los parámetros de cambio y difusión al tiempo t para una inversión instantánea en T , respectivamente. El primero conforma un componente no aleatorio y determinístico; el segundo corresponde a la parte aleatoria y puede ser negativo (Hull [7]).

Como \bar{W}_t es bidimensional, (5) se puede expresar también así:

$$df_t(T) = \sum_{i=1}^2 \alpha_t^i(T) dt + \sum_{i=1}^2 \sigma_t^i(T) dW_t^i, \quad (6)$$

en donde W_t^1 se interpreta como el proceso que determina los cambios en la pendiente de la curva de la estructura de tasas de interés, y W_t^2 como el proceso que determina la concavidad/convexidad de la curva (Haugh [6]).

Teorema: Si no existen posibilidades de arbitraje, es decir, si existe una EMM \bar{Q} , y si la dinámica de las tasas forward instantáneas está dada por (5) entonces:

$$\begin{aligned}\sum_{i=1}^d \alpha_i^i(T) &= \sum_{i=1}^d \sigma_i^i(T) \int_t^T \sigma_i^i(u) du - \sum_{i=1}^d \sigma_i^i(T) \lambda_i \\ &= \sum_{i=1}^d \sigma_i^i(T) \left(\int_t^T \sigma_i^i(u) du - \lambda_i \right)\end{aligned}$$

con $d \geq 1$. (7)

El punto central del teorema anterior es reconocer la relación explícita existente entre los parámetros de cambio y difusión en la dinámica de las tasas forward instantáneas (Haugh [6]).

λ_1 y λ_2 son los precios de mercado del riesgo asociados con W_t^1 y W_t^2 , respectivamente; estos dos parámetros especifican qué tanto difiere la medida del mundo real P^- de la EMM \bar{Q} .

Si (7) se sustituye en (6), se tiene que:

$$df_t(T) = \sum_{i=1}^2 \left(\sigma_i^i(T) \left(\int_t^T \sigma_i^i(u) du - \lambda_i \right) \right) dt + \sum_{i=1}^2 \sigma_i^i(T) dW_t^i. \quad (8)$$

III.3 Calibración de los parámetros de difusión

Se considera $R_j = (R_{j0}, \dots, R_{jm})$, con $j=1, \dots, n$, como $m+1$ muestras de n tasas spot continuas con plazos $\delta_1, \dots, \delta_n$.

Las tasas de interés de mercado se modelaron a partir de las tasas gubernamentales de los siguientes plazos e instrumentos:

- *1 Día*: Tasa de fondeo gubernamental (tasa promedio ponderada).
- *30 Días ó 1 Mes*: Cetes 28 días.
- *90 Días ó 3 Meses*: Cetes 91 días.
- *180 Días ó 6 Meses*: Cetes 182 días.
- *360 Días ó 1 Año*: Cetes 364 días.
- *1260 Días ó 3.5 Años*: Bonos 3 años.
- *1800 Días ó 5 Años*: Bonos 5 años.
- *3600 Días ó 10 Años*: Bonos 10 años.

Se supone que $\sigma_t(T)$ es una función determinística, la cual sólo depende del plazo al vencimiento $T - t$. Para estimar $\sigma_t(T)$, se realiza un análisis de componentes principales a las muestras.

Sea $R_t(T)$, la tasa spot continua al tiempo t con vencimiento en T , se define:

$$R_t(T) = -\frac{\ln(B_t(T))}{T-t} = \frac{1}{T-t} \int_t^T f_t(u) du. \quad (9)$$

De acuerdo con Kalkbrener et al. [9], en un modelo HJM de dos factores, la dinámica de la tasa spot puede representarse así:

$$dR_t(T) = -\left(\frac{1}{T-t}\right)\left(r_t - R_t(T) - \sum_{i=1}^2\left(a_t^i(T)\lambda_i + \left(\frac{1}{2}\right)a_t^i(T)^2\right)\right)dt - \sum_{i=1}^2\frac{a_t^i(T)}{T-t}dW_t^i, \quad (10)$$

donde

$$a_t^i(T) = -\int_t^T \sigma_t^i(u)du, \quad \text{con } i = 1, 2. \quad (11)$$

Al discretizar (10) con pasos de tamaño Δ , se obtiene:

$$R_{t+\Delta}(T) = R_t(T) - (r_t - R_t(T))\frac{\Delta}{T-t} + \sum_{i=1}^2\left(a_t^i(T)\lambda_i + \left(\frac{1}{2}\right)a_t^i(T)^2\right)\frac{\Delta}{T-t} - \sum_{i=1}^2\frac{a_t^i(T)}{T-t}\Delta W_t^i, \quad (12)$$

donde ΔW_t^i es una variable normalmente distribuida con media cero y varianza proporcional a Δ .

Dado que $a_t^1(T)$ y $a_t^2(T)$ sólo dependen del plazo al vencimiento $T-t$ pero no de t , para cualquier t y plazos δ_j con $j=1, \dots, n$, se define:

$$\tilde{\alpha}^i(\delta_j) = \left(a_t^i(t+\delta_j)\lambda_i + \left(\frac{1}{2}\right)a_t^i(t+\delta_j)^2\right)\frac{\Delta}{\delta_j}, \quad (13)$$

$$\tilde{\sigma}^i(\delta_j) = -\frac{a_t^i(t+\delta_j)}{\delta_j}. \quad (14)$$

A partir de la discretización en (12), se define:

$$A_t(\delta_j) = \sum_{i=1}^2 \tilde{\alpha}^i(\delta_j) + \sum_{i=1}^2 \tilde{\sigma}^i(\delta_j) \Delta W_t^i, \quad (15)$$

donde

$$A_t(\delta_j) = R_{t+\Delta}(t + \delta_j) - R_t(t + \delta_j) + (r_t - R_t(t + \delta_j)) \frac{\Delta}{\delta_j}. \quad (16)$$

Al transformar las muestras de tasas spot continuas R_{j0}, \dots, R_{jm} , $j=1, \dots, n$, en muestras de las tasas ajustadas $A_t(\delta_j)$, se considera la tasa spot con el plazo más corto δ_1 como una aproximación de la tasa corta (en el caso de este trabajo, se trata de la tasa de fondeo gubernamental).

Dado que no se dispone de información para las tasas con plazo $\delta_j - \Delta$, se realiza una interpolación lineal para aproximar $R_{t+\Delta}(t + \delta_j)$, para $j = 2, \dots, n$. De esta forma:

$$A_t(\delta_j) \approx \frac{(\delta_j - \delta_{j-1} - \Delta)R_{t+\Delta}(t + \Delta + \delta_j) + \Delta R_{t+\Delta}(t + \Delta + \delta_{j-1})}{\delta_j - \delta_{j-1}} - R_t(t + \delta_j) + (R_t(t + \delta_1) - R_t(t + \delta_j)) \frac{\Delta}{\delta_j}. \quad (17)$$

Esta aproximación se usa para ajustar y centrar las tasas spot, con lo cual se obtienen m muestras $s_j = (s_{j0}, \dots, s_{jm-1})$, con $j=2, \dots, n$, para los plazos $\delta_2, \dots, \delta_n$. Con dichas muestras, se calculan los componentes principales e_k en cada instante t :

$$e_k = \sum_{j=1}^{n-1} p_{kj+1} s_{j+1}, \quad k = 1, \dots, n-1 \quad (18)$$

con vectores de coeficientes $p_j = (p_{k2}, \dots, p_{kn})$.

Los componentes principales e_1, \dots, e_{n-1} no están correlacionados y tienen media 0 y varianzas $\beta_1 > \dots > \beta_{n-1}$. Con ellos, s_j se puede descomponer así:

$$s_{j+1} = \sum_{k=1}^{n-1} p_{kj+1} e_k, \quad j = 1, \dots, n-1. \quad (19)$$

En el caso del modelo de tasas de interés de mercado de este trabajo, sólo se consideran los primeros dos componentes principales:

$$s_{j+1} \approx p_{1j+1} e_1 + p_{2j+1} e_2, \quad j = 1, \dots, n-1. \quad (20)$$

Se identifica la variable normalmente distribuida ΔW^i con el componente principal e_k , y se escogen los parámetros de difusión $\tilde{\sigma}^i(\delta_{j+1})$ tal que las varianzas coincidan lo más que se pueda.

Dado que

$$\text{desv. estándar}(\tilde{\sigma}^i(\delta_{j+1}) \Delta W^i) = \tilde{\sigma}^i(\delta_{j+1}) \sqrt{\Delta}, \quad \text{desv. estándar}(p_{ij+1} e_i) = p_{ij+1} \sqrt{\beta_i}, \quad (21)$$

cada $\tilde{\sigma}^i(\delta_{j+1})$ para cada plazo y componente, se obtiene al minimizar:

$$\sum_{j=1}^{n-1} (\tilde{\sigma}^i(\delta_{j+1}) - p_{ij+1} \sqrt{\beta_i / \Delta})^2, \quad i = 1, 2. \quad (22)$$

$\tilde{\sigma}^j(\delta_{j+1})$ es equivalente a $\sigma^j(\delta_{j+1})$, con $j=1, \dots, n-1$. Asimismo $a_t^i(t+\delta_{j+1})$ ya se puede obtener. Dado lo anterior, se tiene lo necesario para estimar $R_{t+\Delta}(t+\delta_j)$, es decir, las tasas spot con plazo $\delta_j-\Delta$, considerando:

$$R_{t+\Delta}(t+\delta_j) = R_t(t+\delta_j) + dR_t(t+\delta_j), \quad (23)$$

con r_0 y $R_0(t+\delta_j)$ conocidas.

Las tasas $R_{t+\Delta}(t+\Delta+\delta_j)$, que serán requeridas subsecuentemente por (23), se calculan mediante las tasas estimadas $R_{t+\Delta}(t+\delta_j)$ y $R_{t+\Delta}(t+\delta_{j+1})$; para hacer esto, en términos de tasas spot compuestas mensuales, se interpolan linealmente dichas tasas para obtener la tasa spot compuesta mensual equivalente para $R_{t+\Delta}(t+\Delta+\delta_j)$. Cabe señalar que la tasa del mayor plazo se mantiene fija en la estimación.

Por último, para simular las tasas de interés de mercado, lo que se hace es realizar una simulación de Montecarlo de los procesos W_t^1 y W_t^2 , variables normalmente distribuidas cuyos parámetros se determinan de las medias y desviación estándar de los componentes principales e_1 y e_2 .

III.4 Caracterización de los volúmenes de depósito

El volumen de depósitos V_t es un proceso estocástico adaptado a la filtración $\mathcal{F}_{0 \leq t \leq \tau}$. Se considera un modelo normalmente distribuido, en el cual V_t se define como la suma de una función determinística lineal $f_t = a + bt$ y de un proceso Ornstein-Uhlenbeck X_t , es decir,

$$V_t = f_t + X_t, \tag{24}$$

donde $dX_t = \alpha_V X_t dt + \sigma_V d\tilde{W}_t$

$$(25)$$

con constantes $\sigma_V > 0$ y α_V , y el movimiento browniano \tilde{W}_t bajo \mathcal{F} (Kalkbrener et al. [9]). Cabe señalar que f_t sólo depende del tiempo.

La función f_t especifica la tendencia del volumen V_t , y X_t es un proceso Ornstein-Uhlenbeck que tiene reversión a la media (Neftci [11]).

Se utiliza una regresión lineal para determinar los parámetros de la función f_t , considerando a Δ como la longitud del intervalo de tiempo entre dos observaciones u_i y u_{i+1} .

Para estimar α_V y σ_V , se quita la tendencia a los volúmenes históricos restando f_t ; α_V y σ_V se escogen tal que coincidan con los dos primeros momentos de la serie de volúmenes sin tendencia.

En términos discretos, el proceso Ornstein-Uhlenbeck X_t es:

$$X_{t+\Delta} = X_t + \alpha_V X_t \Delta + \sigma_V \Delta \tilde{W}_t. \tag{26}$$

Así, usando el ajuste

$$\frac{(X_{t+\Delta} - X_t - \alpha_V X_t \Delta)}{\sigma_V} \tag{27}$$

se transforman los volúmenes de depósito en muestras sin tendencia $u = (u_1, \dots, u_m)$ de la variable $\Delta \tilde{W}$.

Se permiten correlaciones constantes c_1, c_2 entre \tilde{W} y los dos movimientos brownianos W^1, W^2 del modelo HJM; es por esto, que se establece:

$$\tilde{W}(t) = c_1 W_t^1 + c_2 W_t^2 + \sqrt{1 - c_1^2 - c_2^2} W_t^3. \quad (28)$$

Para estimar los parámetros de correlación c_1 y c_2 , sean e_1, e_2 , los primeros dos componentes principales definidos para el modelo de tasas de interés de mercado, dado que se ha identificado e_1 (respectivamente e_2) con ΔW^1 (respectivamente ΔW^2), se obtiene c_1 (respectivamente c_2) como la correlación entre u y e_1 (respectivamente e_2).

Por último, para simular los volúmenes de depósito, se consideran las simulaciones de las tasas de interés de mercado, y se realiza una simulación de Montecarlo del proceso W_t^3 . Cabe señalar que si el volumen estimado se hace negativo se establece un piso de V_t en 0.

III.5 Administración del riesgo de liquidez de los depósitos a la vista

Para la administración del riesgo de liquidez, Kalkbrener et al. [9] consideran la **Estructura de Plazos de Liquidez** (*Term Structure of Liquidity*), que es el importe

de depósitos a la vista disponible para invertir en cierto horizonte de tiempo $(0,t)$, con una probabilidad p establecida. Esta estructura proporciona la probabilidad de que el volumen caiga por debajo de determinados niveles en intervalos de tiempo específicos, ayudando a calcular restricciones de liquidez.

Sea V_u un proceso estocástico que especifica el volumen de los depósitos y sea $0 \leq t \leq \tau$, se define el proceso de minimización:

$$M_t := \min_{0 \leq u \leq t} V_u. \quad (29)$$

En cada escenario, el proceso estocástico M_t especifica el volumen mínimo en $[0,t]$, es decir, la cantidad exacta disponible para invertir a lo largo del período $[0,t]$ con probabilidad p .

Para $0 < p < 1$, la $TSL(t,p)$ se denota como el p -cuantil de M_t , donde p es la probabilidad de que el volumen caiga debajo del nivel $TSL(t,p)$ en el intervalo de tiempo $[0,t]$.

La $TSL(t,p)$ se construye mediante la simulación del proceso del volumen V_u .

Dados

$$\begin{aligned} M_t &= \min_{0 \leq s \leq t} V_s, \\ V_t &= a + bt + X_t, \\ dX_t &= \mu_V X_t dt + \sigma_V d\tilde{W}_t, \end{aligned} \quad (30)$$

bajo \mathcal{F} . Para cada escenario de V_t se construye el escenario mínimo correspondiente de M_t . El cuantil $TSL(t,p)$ se determina al ordenar los valores de los escenarios mínimos en t .

IV. Resultados

Las series históricas utilizadas para calibrar los modelos comprendieron 5 años y 1 mes de datos mensuales: de marzo de 2004 a marzo de 2009²³. Asimismo para evaluar la calidad de los pronósticos de los modelos ajustados, se utilizaron series históricas de las variables comentadas de abril a octubre de 2009.

IV.1 Caracterización de las tasas de interés de mercado

Tras realizar el análisis de componentes principales, se obtuvieron 7 componentes, de los cuales los dos primeros explicaron el 74.6% de la varianza de la muestra; dado esto, se consideró suficiente quedarse sólo con estos 2 primeros componentes para modelar las tasas de interés de mercado.

	Varianza de cada Componente Principal						
	1	2	3	4	5	6	7
Desv. Estándar	0.6006%	0.4007%	0.2679%	0.2154%	0.1783%	0.1415%	0.0864%
Varianza	0.0036%	0.0016%	0.0007%	0.0005%	0.0003%	0.0002%	0.0001%
Proporción	51.6%	23.0%	10.3%	6.6%	4.5%	2.9%	1.1%
Proporción Acum.	51.6%	74.6%	84.9%	91.5%	96.1%	98.9%	100%

Plazo	Vectores de Coeficientes de cada Componente Principal por Plazo						
	1	2	3	4	5	6	7
1 Mes	0.3099386	-0.8051845	0.4136939	0.0568916	-0.2666811	-0.0423097	0.0912567
3 Meses	0.3502004	-0.2427266	-0.4904109	-0.0651530	0.1027805	0.0375541	-0.7494809
6 Meses	0.3917950	-0.1093137	-0.5267002	-0.2677295	0.1658612	-0.3264461	0.5927715
1 Año	0.5359768	0.1909499	-0.0186802	0.4812824	0.0897466	0.6285766	0.2027863
3.5 Años	0.3776814	0.4180462	0.0332878	0.0387559	-0.7450175	-0.3380660	-0.1031722
5 Años	0.3435402	0.1938746	0.4415933	0.1908032	0.5681627	-0.5156770	-0.1557263
10 Años	0.2825956	0.1788867	0.3383485	-0.8070308	0.0699509	0.3389876	-0.0505481

²³ Dado que en febrero de 2005, se observó un cambio muy drástico en los volúmenes del Ahorro, se consideró más conveniente tomar datos históricos para este producto a partir de dicha fecha.

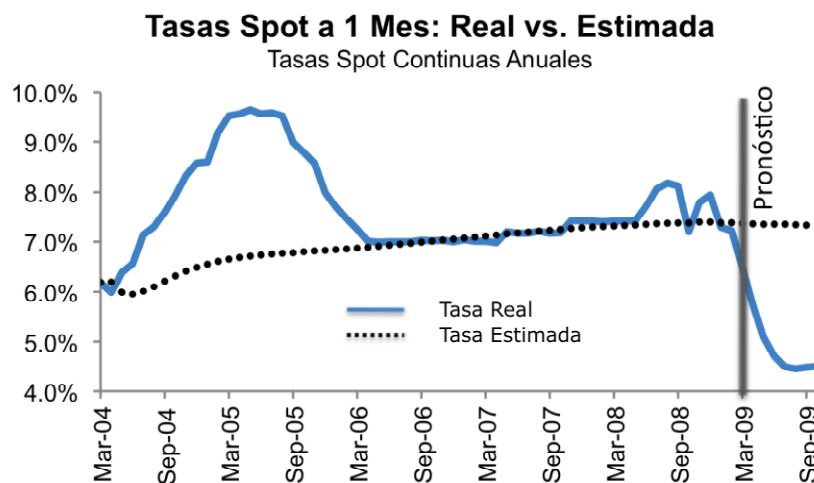
Los parámetros de difusión $\sigma^i(\delta_{j+1})$, con $i=1,2$, y $j=1, \dots, 7$, resultantes del análisis de componentes principales (para los componentes 1 y 2, y para cada uno de los plazos) fueron:

	Plazo						
	1 Mes	3 Meses	6 Meses	1 Año	3.5 Años	5 Años	10 Años
Componente 1	0.64487%	0.72864%	0.81518%	1.11517%	0.78582%	0.71478%	0.58798%
Componente 2	-1.11763%	-0.33691%	-0.15173%	0.26505%	0.58027%	0.26911%	0.24830%

En cuanto a los precios del riesgo λ_1 y λ_2 por componente, se obtuvo lo siguiente:

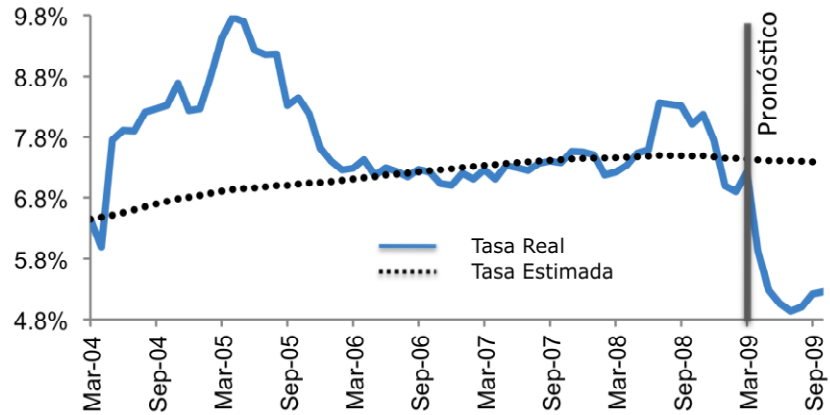
Precios del Riesgo	
Componente 1	27.047%
Componente 2	-99.241%

Como se podrá observar en las siguientes gráficas, el ajuste del modelo no es bueno en los primeros dos años, así como en el último año y en el pronóstico de todas las tasas (excepto la de 5 años); a continuación, se muestra la estimación de las tasas de interés de mercado y las tasas observadas para los plazos de 1 mes, 1 y 5 años, así como el pronóstico utilizando el modelo ajustado:



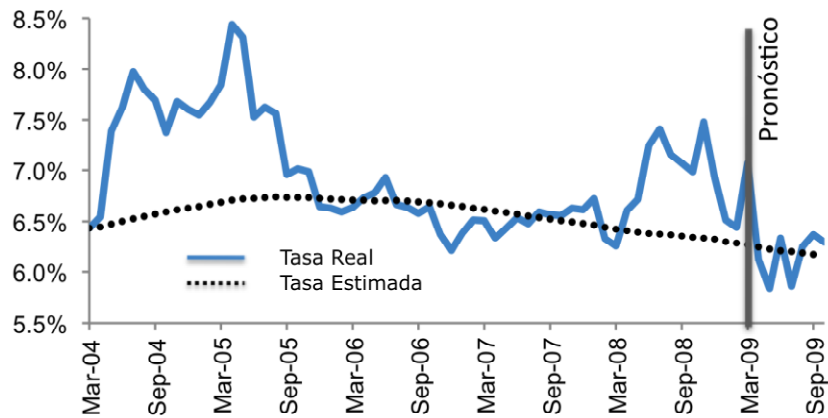
Tasas Spot a 1 Año: Real vs. Estimada

Tasas Spot Continuas Anuales

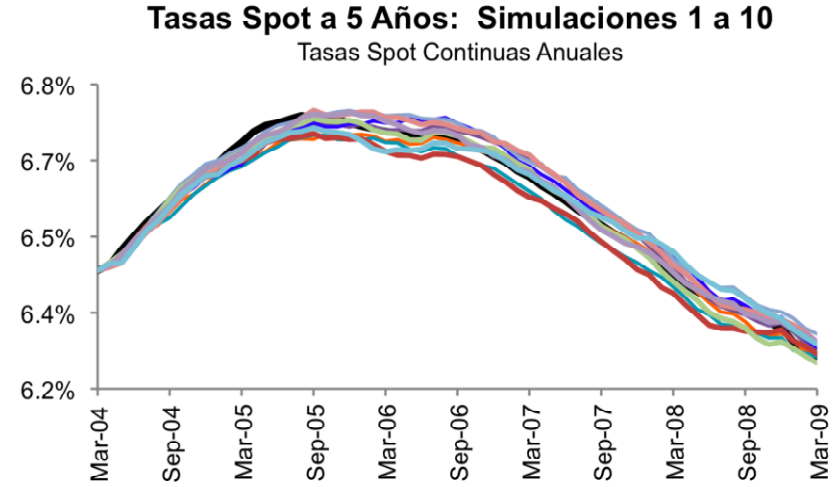
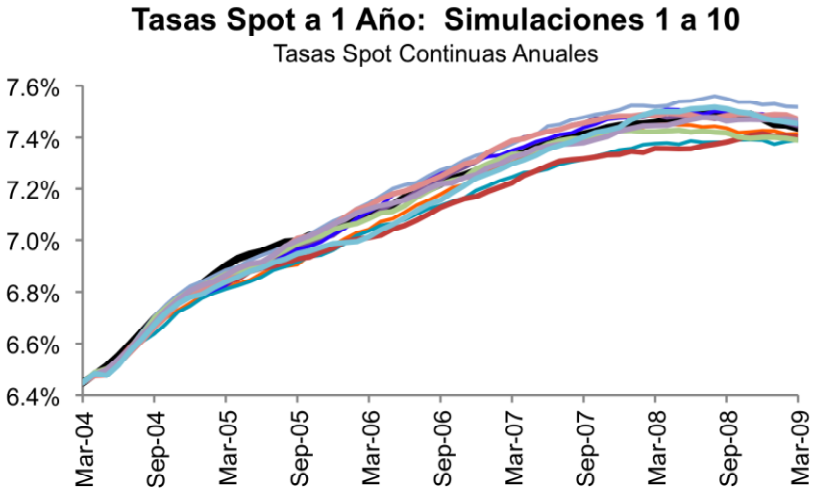
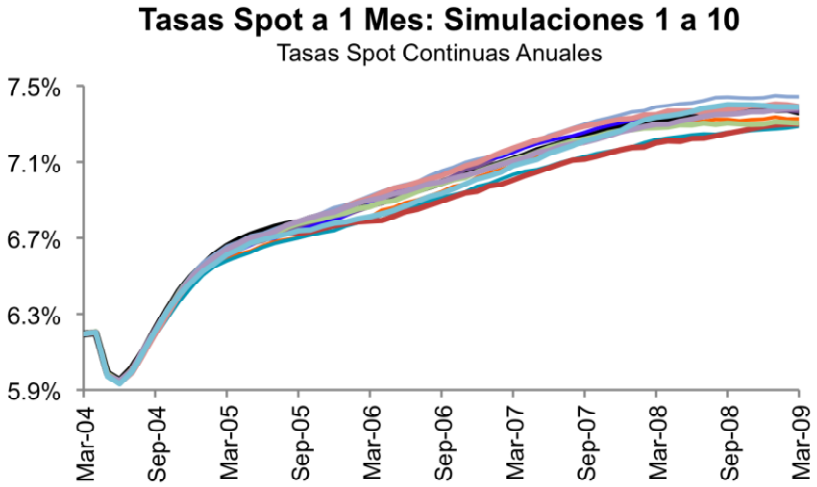


Tasas Spot a 5 Años: Real vs. Estimada

Tasas Spot Continuas Anuales

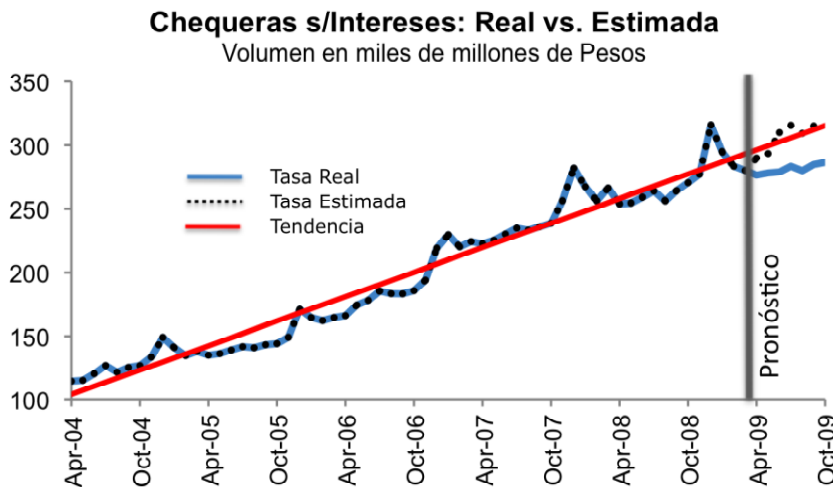
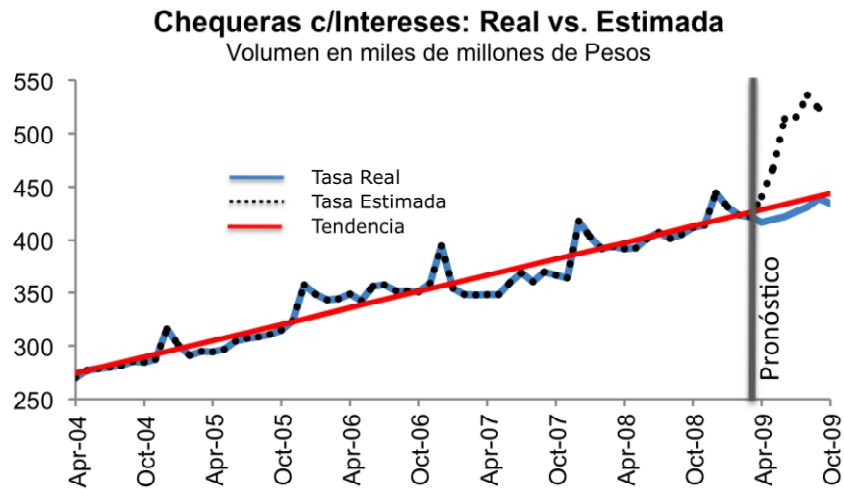


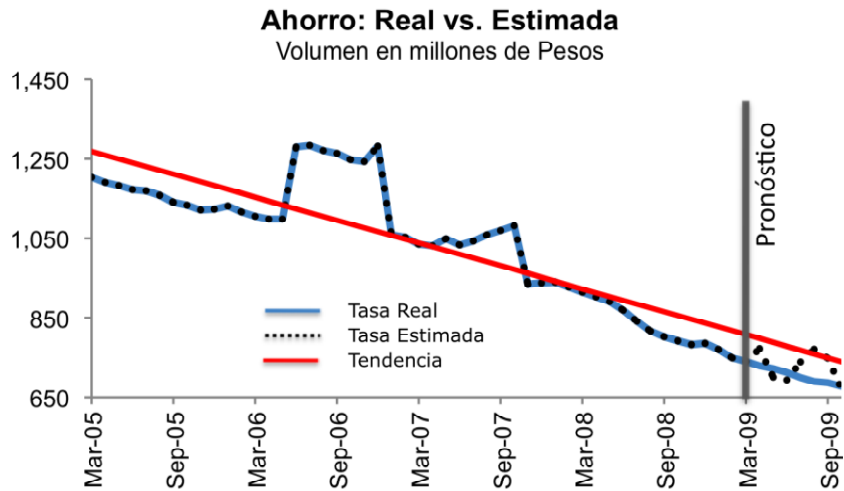
Las 10 primeras simulaciones de las tasas de interés de mercado, junto con la tasa base estimada, para los mismos plazos presentados en las gráficas anteriores, se ven como sigue:



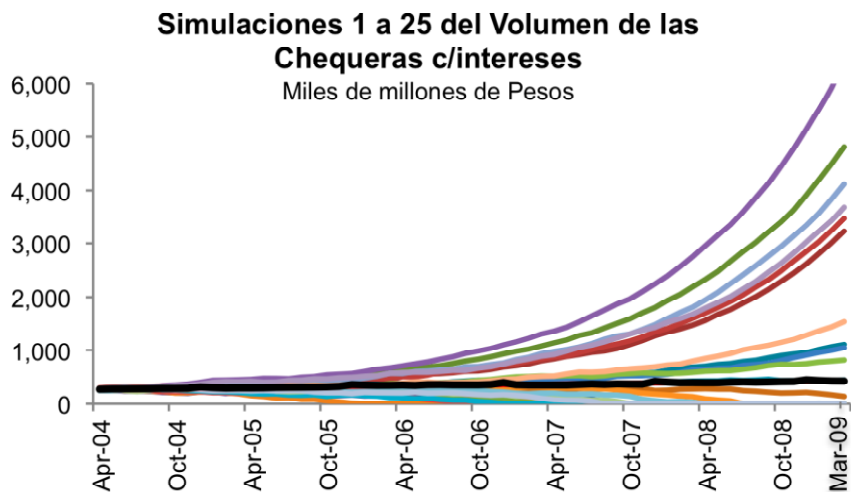
IV.2 Caracterización de los volúmenes de depósito

Al obtener la tendencia de los tres tipos de depósitos a la vista analizados (volumen) mediante un análisis de regresión, se logró un ajuste con un nivel de significancia elevado (0.1%). Después, se procedió a calcular ΔW^3 para cada uno, el cual junto con ΔW^1 y ΔW^2 de las tasas de interés de mercado, permitieron modelar el volumen de los depósitos. El pronóstico no es bueno principalmente en el caso de Chequeras con Intereses. A continuación, se presentan los volúmenes reales en relación a los estimados y pronosticados, así como la tendencia:



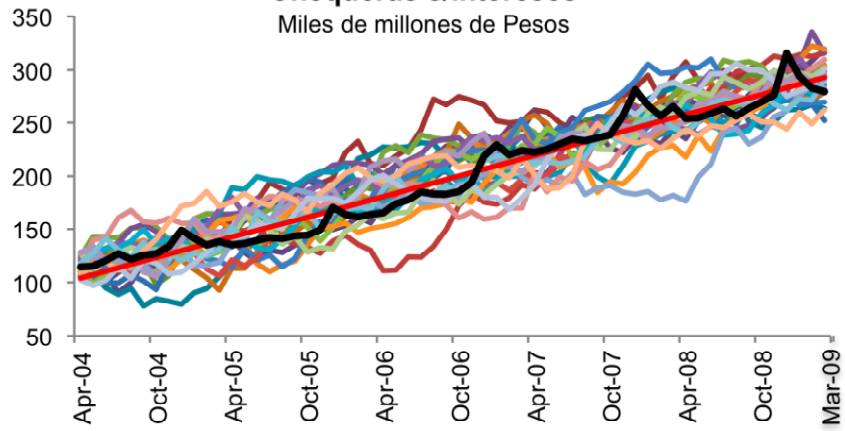


Al simular ΔW^3 , así como considerar las simulaciones de ΔW^1 y ΔW^2 realizadas para las tasas de interés de mercado, los volúmenes estimados se ven como se muestra a continuación (con el objeto que se distingan mejor las simulaciones entre sí, sólo se presentan las primeras 25; la línea negra es el nivel observado):



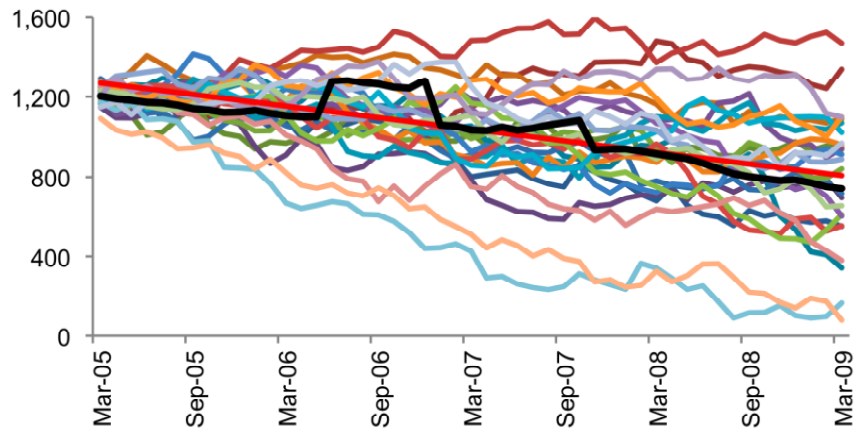
Simulaciones 1 a 25 del Volumen de las Chequeras s/intereses

Miles de millones de Pesos



Simulaciones 1 a 25 del Volumen del Ahorro

Millones de Pesos



IV.3 Administración del riesgo de liquidez de los depósitos a la vista

Una vez modelados los volúmenes de los depósitos a la vista, se procedió a estimar la Estructura de Plazos de Liquidez, con un nivel de significancia de 1%.

Chequeras con Intereses (Miles de millones de MXP)

Plazo de Inversión	1 Mes	3 Meses	6 Meses	9 Meses	1 Año	1.5 Años	2 Años	3 Años	4 Años	5 Años	Total a Invertir
Monto para Invertir	26	8	40	44	127	0	0	0	0	0	245

Chequeras sin Intereses (Miles de millones de MXP)

Plazo de Inversión	1 Mes	3 Meses	6 Meses	9 Meses	1 Año	1.5 Años	2 Años	3 Años	4 Años	5 Años	Total a Invertir
Monto para Invertir	6	9	4	0	0	0	0	0	0	80	100

Ahorro (Millones de MXP)

Plazo de Inversión	1 Mes	3 Meses	6 Meses	9 Meses	1 Año	1.5 Años	2 Años	3 Años	4 Años	Total a Invertir
Monto para Invertir	83	81	78	60	149	161	254	141	98	1,105

V. Conclusiones

Se puede afirmar que el marco cuantitativo de Kalkbrener et al. [9] para la caracterización de los depósitos con vencimiento indeterminado proporciona una buena alternativa para modelar el volumen de los depósitos a la vista de la banca mexicana, en específico cuentas de Cheques, con y sin Intereses, y cuentas de Ahorro; esto se mantiene aún cuando el ajuste del modelo de tasas de interés de mercado fue deficiente y aunque los pronósticos de los volúmenes no arrojaron buenos resultados (estos podrían mejorar con una actualización de la tendencia).

En cuanto a la Estructura de Plazos de Liquidez obtenida para la administración del riesgo de liquidez de los depósitos a la vista, se puede observar que la permanencia de las Chequeras con Intereses es baja; esto se afirma debido a que el monto estimado vence en un año. En cambio, en las Chequeras sin Intereses, se puede decir que se trata de un producto con una alta permanencia, dado que el 80% de la Estructura de Plazos vence al menos en 5 años. Por último, en el caso

del Ahorro, también se obtuvo una buena permanencia, dado que la mayor parte de la Estructura de Plazos vence entre 1 y 3 años.

Un problema que se podría señalar a la Estructura de Plazos de Liquidez es que el plazo máximo depende del tamaño de la muestra disponible; es por esto, que en este trabajo ésta se pudo obtener sólo hasta 5 años. No obstante, dados los resultados de permanencia obtenidos y por tomar una postura conservadora, se considera que no fue necesario contar con una historia más amplia.

Es posible que existan otras alternativas de modelos que arrojen resultados más precisos que los obtenidos con este marco; no obstante, habría que evaluar antes de usarlos qué tanta complejidad adicional conlleva su calibración, dado que la estructura simple del marco de Kalkbrener et al. [9] es una de sus fortalezas.

En la práctica, habría que hacer backtestings periódicos de la Estructura de Plazos de Liquidez, con el objeto de validar que tan adecuada continúa siendo la estructura, posiblemente volviéndola a calibrar con historia reciente; el principal tema de preocupación sería que la tendencia del volumen cambiara muy drástico.

Dado lo anterior, se puede decir que los objetivos primordiales de este trabajo se pudieron satisfacer. Es por esto que se puede sugerir que un banco que no cuente con un modelo establecido para caracterizar el volumen y administrar el riesgo de liquidez de sus depósitos a la vista (si es que conforman una fuente relevante de recursos para el banco), podría considerar el marco de Kalkbrener et al. [9] como

una alternativa para ello. Más aún, se podrían realizar análisis más detallados como por tipo de clientes en cada tipo de depósito a la vista, si se cuenta con la información necesaria. Asimismo los reguladores bancarios podrían usar este marco para evaluar la Estructura de Plazos de Liquidez de cada banco del sistema, con el objeto de detectar posibles debilidades en alguna institución.

En futuros análisis, otra alternativa a explorar con el marco de Kalkbrener et al. [9] es modelar otro tipo de productos, como por ejemplo las tarjetas de crédito.

Por último, no hay que perder de vista que el marco comentado se basa en el supuesto de condiciones normales de mercado, es decir, en el caso de escenarios de estrés habría que manejarlo con prudencia. En condiciones normales, los modelos son calibrados mediante series históricas y cierta volatilidad histórica; por su parte, en los análisis de estrés, los cambios en los factores económicos y en las reacciones de la gente difieren de lo visto en el pasado (Bardenhewer [1]).

VI. Bibliografía

[1] Bardenhewer, M. M. (2007). "Modeling Non-Maturing Products". En L. Matz y P. Neu (Eds.) *Liquidity Risk Measurement and Management* (pp. 220-256), John Wiley & Sons, Singapur.

[2] Basel Committee on Banking Supervision. (2000). *Sound Practices for Managing Liquidity in Banking Organisations*, Bank for International Settlements, Suiza.

- [3] Basel Committee on Banking Supervision. (2006). *International Convergence of Capital Measurement and Capital Standards*, Bank for International Settlements, Suiza.
- [4] Basel Committee on Banking Supervision. (2008). *Principles for Sound Liquidity Risk Management and Supervision*, Bank for International Settlements, Suiza.
- [5] Choudhry, M. (2007). *Bank Asset and Liability Management: Strategy, Trading, Analysis*, John Wiley & Sons, Singapur.
- [6] Haugh, M. (2005). "The Heath-Jarrow-Morton Framework", *Term Structure Models*, IEOR, 1-9.
- [7] Hull, J. C. (2002). *Options, Futures and Other Derivatives*, Quinta Edición, Prentice Hall, EE.UU.
- [8] James, J. y Webber, N. (2000). *Interest Rate Modelling*, John Wiley & Sons, Inglaterra.
- [9] Kalkbrener, M. y Willing, J. (2004). "Risk management of non-maturing liabilities", *Journal of Banking & Finance*, 28, 1547-1568.
- [10] Matz, L. y Neu, P. (Eds.) (2007). *Liquidity Risk Measurement and Management*, John Wiley & Sons, Singapur.
- [11] Neftci, S. N. (2000). *An Introduction to the Mathematics of Financial Derivatives*, Segunda Edición, Academic Press, EE.UU.
- [12] Uyemura, D. G. y van Deventer, D. R. (1993). *Financial risk management in banking: The theory & application of asset & liability management*, Bankline, EE.UU.