

ANÁLISIS ESPECTRAL Y REDES EN MATRICES DE CORRELACIÓN FINANCIERA

Linda Margarita Medina Herrera, Tecnológico de Monterrey
Ernesto Armando Pacheco Velázquez, Tecnológico de Monterrey

RESUMEN

En este artículo se usa la teoría de matrices aleatorias y árboles de expansión mínima para analizar la matriz de correlación C de los rendimientos de las 35 principales acciones cotizadas en la bolsa mexicana de valores. Se encuentra que el espectro de la matriz C tiene una estructura de banda aleatoria, los componentes de los vectores propios fuera de esta banda muestran importantes similitudes con las medidas topológicas del árbol de expansión mínima. Se obtiene que la empresa de telecomunicaciones AMX es el vértice central del árbol y la principal componente del vector propio más grande. Las medidas del árbol permiten concluir que sólo se forman dos conglomerados por sector económico: construcción y telecomunicaciones, los cuales presentan una gran influencia sobre el resto de las acciones.

PALABRAS CLAVES: Teoría de Matrices Aleatorias, Redes Financieras, Econofísica, Árbol de Expansión Mínima

SPECTRAL ANALYSIS AND NETWORKS IN FINANCIAL CORRELATION MATRICES

ABSTRACT

This paper uses the theory of random matrices and minimal spanning trees to analyze the correlation matrix C of returns of the top 35 stocks traded on the Mexican Stock Market. The results show the spectrum of the matrix C has a random band structure. The eigenvector components outside this band show significant similarities to the topological measures of the minimum spanning tree. The telecommunications company AMX is the central vertex of the tree and the main component of the largest eigenvector. The tree measures support the conclusion that only two clusters are formed by economic sectors: construction & telecommunications, which have a great influence on other stocks.

JEL: C16, C65, G11, C02, C22, C38, C45, C61, C8, D85

KEYWORDS: Random Matrix Theory, Stock Market Network, Econophysics, Minimum Spanning Tree

INTRODUCCIÓN

Uno de los problemas más frecuentes a los que nos enfrentamos al usar matrices de correlación financieras es que están “llenas de ruido” debido, entre otras cosas, a la finitud de las series de tiempo, a las condiciones cambiantes del mercado y al hecho de que dos activos que están correlacionados hoy pueden no estarlo mañana. Se han propuesto muchos modelos económicos y físicos para explicar la correlación de activos financieros, que es un sistema complejo con muchas relaciones difíciles de identificar y cuantificar. Las matrices de correlación financiera tienen un papel importante en administración de riesgos y construcción de portafolios.

De alguna manera, el problema de interpretar correlaciones entre los cambios en los rendimientos de activos financieros se asemeja a las dificultades experimentadas por los físicos en los años 50 cuando

trataban de explicar los cambios en los niveles de energía de un núcleo complejo. En este contexto el físico alemán Wigner (1951) y un poco más adelante Dyson y Mehta desarrollaron la Teoría de Matrices Aleatorias. Una serie de resultados importantes fueron establecidos y probados para el espectro de un núcleo complejo. Estos resultados forman un conjunto de propiedades universales de las matrices aleatorias. La parte de la matriz que no satisface las propiedades universales es aquella que puede ser considerada como no aleatoria y que por lo tanto contiene información sobre el sistema y la naturaleza de sus interacciones. Otra manera de estudiar la correlación entre activos financieros es construyendo un árbol de expansión mínima. El árbol de expansión mínima es una red cuyos vértices son los activos financieros y la distancia entre los vértices se construye a partir de los coeficientes de correlación. El árbol nos permite identificar el activo con mayor número de vértices conectados, el centro de masa del árbol, que de alguna manera representa el activo que más influencia tiene sobre el sistema. El árbol de expansión mínima también nos permite identificar conglomerados por sector económico y construir una medida de riesgo que puede ser usada en la construcción de portafolios.

La Teoría de matrices aleatorias y los árboles de expansión mínima son una buena combinación que no ha sido utilizada con frecuencia ya que la primera permite limpiar la matriz de correlación del ruido y una vez hecho esto, el árbol de expansión puede ser usado con la parte de la matriz limpia de ruido para realizar una taxonomía más confiable de las empresas y construir medidas topológicas mucho más adecuadas. Este artículo tiene dos objetivos fundamentales. El primero es aplicar la teoría de matrices aleatorias para estudiar la matriz de correlación C de los rendimientos diarios de las principales empresas que cotizan en la Bolsa Mexicana de Valores del 3 de enero de 2005 al 6 de septiembre de 2012.

Probamos las propiedades universales tales como distribuciones y espaciamiento de los valores propios de la matriz C , mostramos que C comparte propiedades universales con el conjunto de matrices aleatorias ortogonales gaussianas y adicionalmente cuantificamos el número de empresas que contribuyen significativamente al vector propio más grande, que representa la influencia de todo el mercado, usando el concepto de cociente de participación inverso. El segundo objetivo es construir un árbol de expansión mínima a partir de la matriz C , calcular las medidas topológicas sobre el árbol que permitan identificar el vértice central y realizar una taxonomía de las empresas, identificando conglomerados económicos y una medida de riesgo. Este artículo tiene varias contribuciones importantes a la literatura sobre el tema: da evidencia empírica de la potencia de la teoría de matrices aleatorias para obtener información real de la matriz de correlación en un mercado financiero emergente.

Compara los resultados obtenidos usando Teoría de Matrices aleatorias y Árboles de expansión mínima, encontrando una fuerte coincidencia en la taxonomía de las empresas pero interesantes diferencias que pueden ser aprovechadas en la construcción de índices y administración de riesgo. El árbol no se construye de la manera tradicional, con la distancia euclídea, se usa la distancia chi-cuadrada que separa más claramente los conglomerados y que puede ser usada para minimizar el riesgo en la construcción de un portafolio. El artículo está organizado de la siguiente forma: la sección dos presenta una revisión de la literatura sobre Teoría de Matrices Aleatorias y Árboles de expansión mínima en el contexto económico-financiero. La sección dos presenta la Teoría de Matrices Aleatorias y los Árboles de expansión mínima. La sección tres los datos y la metodología, la sección cuatro los resultados empíricos y la sección final presenta las conclusiones.

REVISION DE LA LITERATURA

Las propiedades de las matrices aleatorias es un problema con una historia interesante que tiene sus orígenes en los años 50, cuando Wigner, Dyson y Mehta estudiaban los niveles de energía de un núcleo complejo. Los modelos existentes no explicaban adecuadamente el comportamiento de los datos empíricos. Asumiendo un comportamiento aleatorio, construyeron matrices que mostraban las conexiones entre los diferentes niveles de energía. Las propiedades espectrales de estas matrices aleatorias les

permitieron encontrar predicciones significativas sobre cómo los niveles de energía se relacionaban unos con otros. La Teoría de Matrices Aleatorias establece un conjunto de propiedades universales aplicables en muchas áreas: física cuántica, nanotecnología, ecología, lingüística y muchas otras donde urge encontrar una conexión entre una cantidad importante de información aparentemente no correlacionada. Esta teoría también ha sido exitosamente aplicada en una colección de trabajos sobre matrices de correlación financiera, análisis de riesgo y optimización de portafolios.

Laloux et al (1999) encuentran un ajuste importante entre las predicciones teóricas y los datos empíricos de las series de tiempo de diferentes activos del S&P500 y discuten cómo esta idea puede ser implementada en la administración de riesgos. Laloux et al (2000) muestran que las matrices de correlación financieras tienen una cantidad importante de ruido que puede ser detectado y eliminado, estudiando el espectro de la matriz de correlación. Guhr (2003) presenta un método alternativo para detectar y eliminar el ruido mediante la potencia de una matriz.

Plerou (1999) muestra que la mayoría de los valores propios de la matriz de correlación de rendimientos de activos financieros coinciden sorprendentemente bien con las predicciones universales de la Teoría de las matrices aleatorias. Maslow (2001) usa la Teoría de matrices aleatorias para medir la globalización usando índices financieros mundiales, mostrando que hay fuertes interacciones entre economías individuales. La teoría de matrices aleatorias también ha sido utilizada en la optimización de portafolios donde los movimientos relacionados entre los activos financieros se miden a partir de la matriz de correlación: Laloux et al (2000) Bouchaud (2007), Potters et al (2000) (2007), Tola et al (2008), Eom et al. (2009). Rosenow (2002) encuentra que la teoría de matrices aleatorias hace una estimación de C que supera la estimación estándar en términos de una inversión que conlleva un nivel mínimo de riesgo.

Dali et al (2010) estudia el efecto de tres filtros de la Teoría de Matrices aleatorias en el riesgo de un portafolio, usando bootstrap y pruebas fuera de muestra. Otros mercados financieros fuera de Estados Unidos han sido analizados usando Teoría de Matrices Aleatorias: Tokio (Utsugi et al 2003), India (Kulkarni 2007), México (Medina- Mansilla 2007), UK (Livan-Rebechi 2012).

Sandoval (2010) estudia los mercados financieros en tiempos de crisis, demostrando que la alta volatilidad de los mercados está directamente vinculada con fuertes correlaciones entre ellos. Bouchaud (2007) presenta una breve pero interesante revisión de la Teoría de Matrices Aleatorias en finanzas y algunos problemas abiertos relacionados. En cuanto a los árboles de expansión mínima, los trabajos fundamentales en el contexto financiero aparecen en los años ochenta. Mantegna y Bonano introducen el concepto de redes en el entorno del mercado financiero como un método para encontrar los arreglos jerárquicos de los activos financieros a través del estudio de las agrupaciones de empresas, véase Mantegna y Stanley (2000), Bonano et al (2000). El concepto de árbol de expansión ha sido utilizado en la construcción de portafolios, índices financieros y volatilidad (Bonano et al 2003). Onnela, Chakraborti, Kaski y Kertész (2002) analizan árboles dinámicos, esto es, con ventanas en el tiempo, los mismos autores (2003) usan el concepto de vértice central, escogiendo el nodo más fuertemente conectado del árbol y definen una medida importante, el “promedio del nivel de ocupación” que durante las caídas del mercado aparece con un valor muy bajo. Los artículos de Eom et al (2009) y Medina, L. Mansilla, R. (2007) muestran evidencia empírica de la importancia de los árboles de expansión en el análisis de los movimientos de los precios en mercados financieros.

Las propiedades de las matrices aleatorias combinadas con el poder de los árboles de expansión mínima han hecho posible investigar los movimientos de los mercados de valores más importantes: Bonano et al. (2003) en Estados Unidos, Jung et al (2006) en Corea, Medina (2007) en México, Eom et al (2009) en Japón, y Tabak et al (2010) Brasil. Los árboles de expansión mínima también han sido usados para analizar los índices financieros globales: Tumminello (2010) y Medina (2012) y otros activos financieros; Miccichè et al (2003), Cukur et al (2007).

Teoría de Matrices Aleatorias

El conjunto más simple de matrices aleatorias es el conjunto de matrices simétricas R de tamaño N^2 , cuyos elementos son variables aleatorias independientes e idénticamente distribuidas. Este conjunto de matrices es conocido como matrices aleatorias Wishart de la Teoría de Matrices Aleatorias (Mehta, 1991). En notación matricial R se puede escribir como $R = \frac{1}{N}AA^T$, donde A es una matriz de tamaño $N \times T$ y A^T es la matriz transpuesta de A . Un resultado interesante se obtiene cuando la matriz R tiene un tamaño muy grande (esto es, $N \rightarrow \infty$), en este caso la distribución de sus valores propios tiene propiedades universales, que son independientes de la distribución de los elementos de la matriz.

Para el caso $N \neq T$ existe una fórmula [Mehta] para la densidad $\rho(\lambda_i)$ de los valores propios λ_i de R . En el límite $N, T \rightarrow \infty$, con un cociente fijo $Q = \frac{T}{N} \geq 1$, se tiene:

$$\rho(\lambda_i) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_- - \lambda_i)(\lambda_i - \lambda_+)}}{\lambda}, \quad \lambda_i \in [\lambda_-, \lambda_+] \quad (1)$$

Donde λ_- , el valor propio más pequeño y λ_+ , el valor propio más grande satisfacen

$$\lambda_{\pm} = \sigma^2(1 \pm \sqrt{1/Q})^2 \quad (2)$$

σ^2/T es la varianza de los elementos de A , o de manera equivalente σ^2 es el promedio de los valores propios de R .

Los valores λ_{\pm} dan una medida estrictamente cuantitativa para decidir cuándo un valor propio particular de una matriz de correlación empírica C refleja una señal de correlación real presente en los datos, o si es simplemente un efecto de ruido causado en particular por la finitud T de la serie de datos. En principio, cualquier valor propio que sea significativamente mayor que λ_+ debe ser tratado como una señal. Para aceptar o rechazar la hipótesis de aleatoriedad de una matriz C no es suficiente comparar la distribución de los valores propios de C con $\rho(\lambda_i)$. Hay matrices aleatorias que tienen unas diferencias drásticas en la distribución de los valores propios que comparten una estructura similar de correlación en sus valores propios debida únicamente a la simetría de las matrices. Por otra parte, matrices con la misma distribución de sus valores propios tienen diferencias drásticas en la correlación de los valores propios.

Así pues, una prueba de aleatoriedad de C debe involucrar una investigación de la correlación en los valores propios λ_{α} (Plerou et al 1999, 2000, 2002). Las matrices reales, cuadradas, aleatorias y simétricas cuyos elementos fuera de la diagonal son independientes e idénticamente distribuidos, presentan las propiedades universales del conjunto de las matrices cuyos elementos están distribuidos de acuerdo a una medida de probabilidad Gaussiana llamado el conjunto gaussiano ortogonal o simplemente GOE (de sus siglas en inglés Gaussian orthogonal ensemble) [Mehta]. Las pruebas estadísticas para los valores propios de C que se usarán, son precisamente las pruebas universales de correlación de los valores propios de matrices aleatorias reales simétricas. Para garantizar que los valores propios tengan un valor uniforme a través del espectro, se realiza una transformación $\xi_k = f(\lambda_k)$ que convierte los valores propios λ_k de C en nuevas variables llamadas “valores propios desplegados” [Mehta]. Las propiedades universales de las matrices GOE son:

1. La distribución de las diferencias de los valores propios de acuerdo al vecino más cercano $s = \xi_{k+1} - \xi_k$ está dada por la ecuación:

$$P_{GOE}(s) = \frac{\pi s}{2} \exp\left(-\frac{\pi}{4}s^2\right) \quad (3)$$

A menudo se refieren a esta distribución como “la conjetura de Wigner”.

2. La distribución de las diferencias de los valores propios de acuerdo al siguiente vecino más cercano $u = \xi_{k+2} - \xi_k$ es

$$P_{GOE}(u) = \frac{2^{18}}{3^6\pi^3} u^4 \exp\left(-\frac{64}{9\pi} u^2\right) \quad (4)$$

3. Correlaciones entre valores propios de rango largo. Para probar correlaciones entre pares de valores propios en rangos largos, se usará el estadístico conocido como “varianza número”:

$$\Sigma^2 = \langle [n(\xi, l) - l]^2 \rangle \quad (5)$$

que está definido como la varianza del número de valores propios desplegados en intervalos de longitud l alrededor de cada ξ_i , donde $n(\xi, l)$ es el número de valores propios desplegados en el intervalo $\left[\xi - \frac{l}{2}, \xi + \frac{l}{2}\right]$ y $\langle \dots \rangle_\xi$ es la media sobre todos los ξ .

Si los valores propios no están correlacionados, $\Sigma^2 \approx l$. En el extremo opuesto, cuando el espectro de los valores propios es rígido, Σ^2 es una constante. En las matrices aleatorias, para valores largos de l , la varianza número Σ^2 tiene un comportamiento intermedio $\Sigma^2 = \ln l$. Es importante notar que estas tres propiedades son independientes entre si.

Una vez probado que los valores propios en el intervalo $[\lambda_-, \lambda_+]$ satisfacen las propiedades universales, se podrán obtener importantes conclusiones del estudio de las componentes de los vectores propios que quedan fuera del intervalo. Los componentes del vector propio u^k de una matriz aleatoria de correlación R están distribuidos normalmente con media cero y varianza uno. Para cuantificar el número de componentes que participan significativamente en cada vector propio, se usará el cociente inverso de participación, una noción aplicada a menudo en la teoría de localización. El Cociente Inverso de Participación de un vector propio normalizado está definido como

$$I^k = \sum_{i=1}^N (u_i^k)^4 \quad (6)$$

Donde N es el número de activos y por lo tanto el número de componentes. La i -ésima componente u_i^k del vector u^k corresponde a la contribución de la i -ésima serie de tiempo a tal vector propio. Es importante notar que si todas las componentes son idénticas y $u_i^k = \frac{1}{\sqrt{N}}$ se tiene que $I^k = \frac{1}{N}$; si sólo hay una componente diferente de cero $u_i^k = 1$, $I^k = 1$. De esta manera, el Cociente Inverso de Participación es el recíproco del número de componentes del vector que contribuyen significativamente al mismo.

Arboles de Expansión Mínima

El problema del árbol de expansión mínima es un problema común de optimización combinatoria. Fue formulado inicialmente por Boruvka en 1926 con la intención de resolver un problema de electrificación del sur de Moravia, donde el proporcionó una solución para hallar la distribución más económica a través de una red de una línea de energía. Un árbol de expansión es una gráfica de N objetos o vértices unidos por $N-1$ arcos que permiten ir de un vértice a cualquier otro. Si cada arco representa una distancia o costo, la suma de los pesos de todos los lados de un árbol, será el peso total del árbol. Un árbol de

expansión mínima es un árbol de expansión cuyo peso total es el mínimo posible entre todos los árboles de expansión con los mismos vértices. Desde hace varias décadas, se ha tratado de encontrar algoritmos eficientes para el problema del árbol de expansión mínima. El algoritmo de Kruskal es un algoritmo muy efectivo y es el que se usará en este artículo. El árbol de expansión mínima que mostramos aquí, tendrá por vértices las principales empresas que cotizan en la BMV y los arcos son las distancias entre las empresas obtenidas a partir de la distancia chi-cuadrado. La distancia chi-cuadrado entre los logaritmos de los rendimientos estandarizados de dos empresas Z_i y Z_j con pesos w_i , $i = 1, 2, 3, \dots, T$ es

$$d_\chi(Z_i, Z_j) = \sum_{k=1}^T w_k (Z_{ik} - Z_{jk})^2 \quad (7)$$

donde $w_i = \frac{1}{R_i}$ donde R_i es el rango de los precios de las acciones del día i con $i = 1, 2, \dots, T$.

Para realizar una taxonomía de los vértices, es necesario seleccionar entre todos ellos el vértice que parece estar en el centro del árbol y que de alguna manera al moverse, mueve todo el árbol. El vértice central se puede seleccionar por medio de uno o varios de los siguientes criterios:

El grado del vértice: El grado del vértice i es el número de vértices que están conectados a él. El vértice central puede ser el vértice que tiene más nodos conectados, esto es, el vértice con mayor número de vecinos.

El peso del vértice: El peso del vértice i es la suma de los coeficientes de correlación de los vértices que están conectados a él. El vértice que tiene peso máximo puede ser elegido como vértice central.

El nodo central: Es el nodo que obtiene el menor promedio de nodos recorridos sobre el árbol para ir del vértice i al vértice j .

El centro de masa: Sea $niv(v_{ij})$ la suma de los arcos que hay que pasar sobre el árbol para ir del vértice i al vértice j . La forma en que se extienden los nodos en el árbol se puede caracterizar con el “promedio de ocupación”. El promedio de ocupación del vértice v_i se define como:

$$l(v_i) = \frac{1}{N} \sum_{j=1}^n niv(v_{ij}) \quad (8)$$

El vértice que produce el valor más bajo del promedio de ocupación, es el centro de masa (Barthélemy et al, 2005). Una vez seleccionado el vértice central v_c el cálculo del promedio de ocupación indica que tanto se extienden las ramas del árbol. Un valor alto de $l(v_c)$ refleja una estructura de mercado muy fina, mientras que en el otro extremo valores bajos se asocian con crisis en el mercado (las ramas del árbol se contraen).

DATOS Y METODOLOGÍA

Los datos usados para este estudio corresponden a las series de tiempo de los rendimientos diarios de 35 empresas que cotizan en la Bolsa Mexicana de Valores del 3 de enero de 2005 al 6 de septiembre de 2012. Para la elección de las empresas y la longitud de las series se tuvo en cuenta la bursatilidad, capitalización y mantenimiento de las mismas. La longitud final de las series es de 1936. Las empresas elegidas tienen la mayor bursatilidad de cada sector y juntas representan más del 83% de participación en el índice de precios y cotizaciones de la Bolsa Mexicana de Valores (IPC). Todas las acciones incluidas han permanecido activas en el periodo seleccionado para el estudio. La Tabla 2 muestra las empresas incluidas en el estudio, su código de bolsa y porcentaje de participación en el Índice de Precios y Cotizaciones (IPC) de la Bolsa Mexicana de Valores. La matriz de correlación que se analiza se construye de la siguiente forma: si se tienen N activos con precio $P_i(t)$ para el activo i en el tiempo t , con $t = 0, 1,$

2, ..., T, el logaritmo de los rendimientos de los activos $S_i(t)$ es $S_i(t) = \ln P_i(t) - \ln P_i(t-1)$. El rendimiento normalizado se define como

$$g_i(t) = \frac{S_i(t) - \langle S_i(t) \rangle}{\sigma_i} \quad (9)$$

Donde σ_i es la desviación estándar de S_i y $\langle . \rangle$ la media. Los elementos de la matriz de correlación empírica C son

$$C_{ij} = \frac{1}{T} \sum_{t=1}^T g_i(t) g_j(t) \quad (10)$$

Tabla 1: Empresas Incluidas en el Estudio

Empresa	Código De Bolsa	Ponderación Ipc	Sector
arca continental	ac	1.70	productos de consumo frecuente
grupo alfa	alfaa	3.32	industrial
controladora alsea	alsea	0.25	servicios y bienes de consumo no básico
américa móvil	amxl	25.54	servicio de telecomunicaciones
consorcio ara	ara	0.00	industrial
grupo aeroportuario del sureste	asurb	0.66	industrial
compañía minera autlán	autlanb	0.00	productos de consumo frecuente
tv azteca	aztecacpo	0.37	servicio de telecomunicaciones
grupo bimbo	bimboa	1.84	productos de consumo frecuente
cemex	cemexcpo	3.48	materiales
controladora comercial mexicana	comerciubc	0.00	productos de consumo frecuente
grupo elektra	elektra	2.04	servicios y bienes de consumo no básico
fomento económico mexicano	femsaubd	8.30	productos de consumo frecuente
grupo carso	gcarsoa5	0.00	otros
corporación geo	geob	0.30	industrial
grupo financiero inbursa	gfinburo	2.74	servicios financieros
grupo financiero banorte	gfnorteo	6.04	servicios financieros
grupo México	gmexicob	6.61	materiales
grupo modelo	gmodeloc	2.88	productos de consumo frecuente
grupo maseca	grumab	0.25	productos de consumo frecuente
desarrolladora homex	homex	0.27	industrial
empresas ica	ica	0.53	industrial
industrias ch	ichb	0.51	materiales
kimberly-clark	kimbera	1.67	productos de consumo frecuente
coca cola femsa	kofl	2.61	productos de consumo frecuente
mexichem	mexchem	1.88	materiales
industrias peñoles	pe&oles	2.78	materiales
sare holding	sareb	0.00	industrial
grupo simec	simecb	0.00	materiales
organización soriana	sorianab	0.00	productos de consumo frecuente
teléfonos de México	telmexl	0.00	servicio de telecomunicaciones
grupo televisa	tlevisacpo	6.76	servicio de telecomunicaciones
urbi desarrollos urbanos	urbi	0.17	industrial
vitro	vitroa	0.00	materiales

Esta tabla muestra las 35 empresas incluidas en el estudio, su código de bolsa (Bolsa Mexicana de Valores), el peso en el Índice de Precios y Cotizaciones del segundo semestre del 2012 y el sector económico perteneciente. Son las empresas más importantes de cada sector económico y sus series de tiempo tienen pocos o no tienen datos perdidos. Juntas aportan más del 83% de índice IPC.

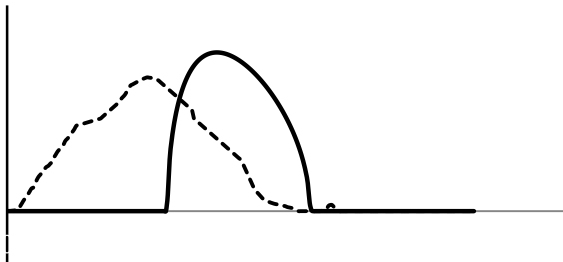
Se comparará las distribuciones empíricas de los valores propios y las estadísticas de los vectores propios de la matriz C construida a partir de estos datos con sus predicciones teóricas, asumiendo que la matriz de correlación es puramente aleatoria. Encontraremos los valores propios en el intervalo $[\lambda_-, \lambda_+]$ y una vez desplegados realizaremos las pruebas de las propiedades universales analizando las distribuciones P_{GOE} y P_{GSE} . El cálculo del estadístico varianza número y su posterior análisis permitirá determinar si el espectro satisface la tercera propiedad universal. Se analizarán los valores propios que caigan por encima y por debajo del intervalo $[\lambda_-, \lambda_+]$ y mediante el cociente inverso de participación se encontrará el número de empresas que aportan de manera significativa a estos vectores. Con las 35 empresas actuando como nodos, calculamos los arcos entre los nodos usando las distancias chi-cuadrada entre los logaritmos de los rendimientos estandarizados. Usaremos el algoritmo de Kruskal para encontrar un árbol de expansión mínima y el software pajek para graficarlo. Se hará el cálculo de las cuatro medidas topológicas en el

árbol para cada una de las 35 empresas con el fin de establecer el vértice central y los conglomerados por sector económico. Finalmente se compararán los resultados obtenidos por los dos métodos.

RESULTADOS EMPIRICOS

El cociente $Q = T / N = 1936/35 > 1$ garantiza que la matriz de correlación es definida positiva y permite la aplicación de la Teoría de Matrices Aleatorias. Obtenemos que $\lambda_- = 0.9064$ y $\lambda_+ = 1.5572$. Hay un valor propio por encima de λ_+ , lo cual es consistente con la mayoría de los estudios empíricos que encuentra que menos del 2% de los valores propios están por encima de λ_+ . La teoría asegura que si la matriz satisface las propiedades universales de la TMA, se podrá obtener información sustancial del vector propio correspondiente a este valor propio. Hay 28 valores propios por debajo de λ_- , que de acuerdo a la literatura, son más susceptibles al ruido. Los valores entre $[\lambda_-, \lambda_+]$ representan la parte de la matriz sin información alguna. El valor propio más grande 11.73 es más de siete veces mayor que λ_+ , se puede entonces atribuirle información real. La figura 1 muestra la distribución de los valores propios empírica (la de trazo firme) y teórica (con pequeñas oscilaciones). El pequeño salto que se encuentra por encima de λ_+ es el único valor propio fuera del intervalo $[\lambda_-, \lambda_+]$. Por debajo de λ_+ se encuentran 28 valores propios, sus vectores propios correspondientes están localizados, esto es, representan la influencia de sólo algunas empresas.

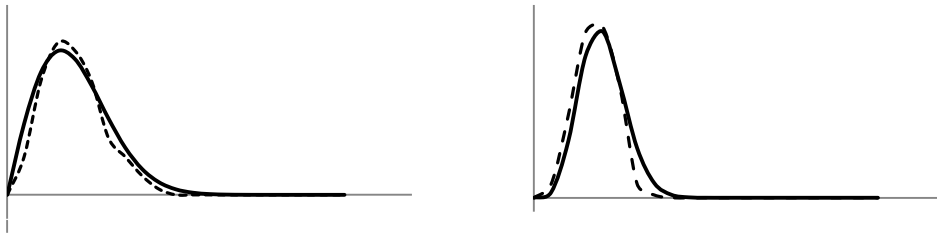
Figura 1: Distribución Teórica y Empírica de los Valores Propios de la Matriz C



La Figura 1. Muestra la distribución teórica (trazo firme) vs. La distribución empírica (punteada) de los valores propios de la matriz C . Sólo un valor propio se encuentra por encima del intervalo $[\lambda_-, \lambda_+]$, a su correspondiente vector propio se le atribuye información real (es el chipote pequeño a la derecha). Los valores propios que se encuentran en la intersección de las dos curvas no tienen información real, y los que se encuentran por debajo de la distribución empírica son valores propios cuyos vectores propios están localizados (pocas empresas contribuyen a ellos).

Es indispensable recordar que la información que se puede obtener del espectro sólo será válida si se prueba que los valores propios que pertenecen al intervalo $[\lambda_-, \lambda_+]$ satisfacen las propiedades universales de la Teoría de Matrices Aleatorias. Después de transformar los valores propios en $\xi_i = f(\lambda_i)$ con el desplegado, se obtienen los siguientes resultados para la matriz C : La hipótesis de la igualdad de las distribuciones es confirmada por el resultado de la prueba de bondad de ajuste de Kolmogorov-Smirnov. Permittiéndonos concluir que el espectro de C satisface las dos primeras propiedades universales de la TMA. La Figura 2a muestra la distribuciones P_{GOE} teórica vs. P_{GOE} empírica con un buen ajuste. La Figura 2b compara las distribuciones P_{GSE} empírica y teórica, obteniendo también un buen ajuste.

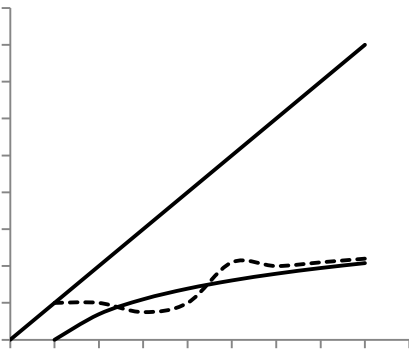
Figura 2 :a, 2b. Distribuciones P_{GOE} y P_{GSE}



La Figura 2a. (a la derecha) muestra la distribución P_{GOE} teórica vs. P_{GOE} empírica, revelando un buen ajuste. La Figura 2b. (a la izquierda) muestra las distribución P_{GSE} empírica vs. P_{GSE} teórica. La igualdad de las distribuciones es confirmada por la prueba de bondad de ajuste de Kolmogorov-Smirnov. Este ajuste en las distribuciones teóricas y empíricas permite concluir que la matriz de correlación C satisface las propiedades universales de las matrices aleatorias y que se podrá separar el ruido de la información real de la matriz C .

Para analizar las correlaciones entre valores propios de rango largo, obtenemos los valores del estadístico varianza número (Σ^2). La Figura 3 muestra el comportamiento de Σ^2 . A medida que l aumenta, Σ^2 se acerca a la función $\ln l$, que es el comportamiento esperado en el espectro de las matrices aleatorias.

Figura 3: El Estadístico Varianza Número



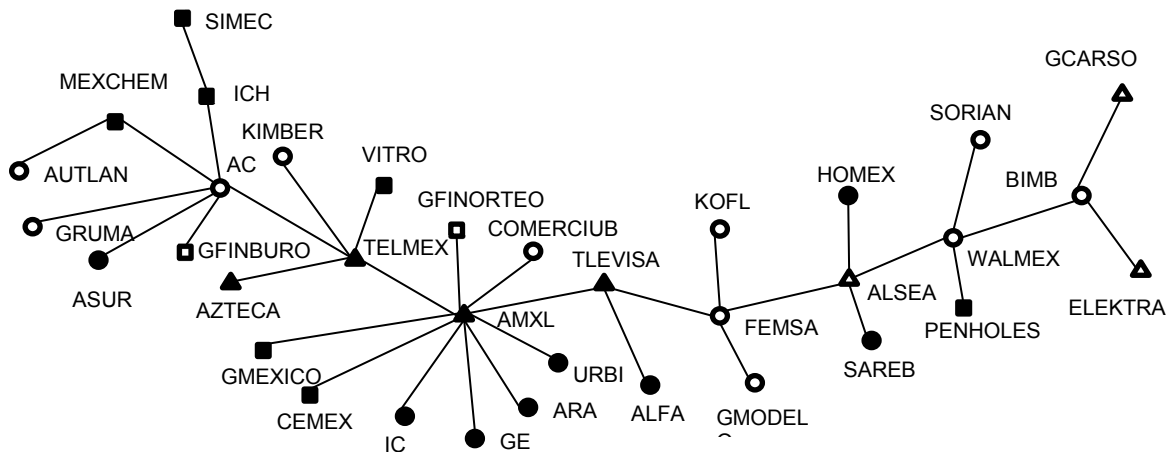
Esta Figura muestra el estadístico varianza número Σ^2 en el intervalo $[\lambda_-, \lambda_+]$. A medida que l crece, Σ^2 se aproxima a la función $f(l) = \ln l$. Si los valores propios no estuvieran correlacionados, la varianza número se acercaría a la recta $y=x$. En las matrices aleatorias, para valores largos de l , la varianza número tiene un comportamiento intermedio (función punteada) que se aproxima a $y = \ln x$.

El primer resultado interesante es que todas las componentes del vector propio asociado al mayor valor propio son positivas, lo que significa que no hay acciones con correlación inversa. Puesto que los vectores propios correspondientes a valores propios diferentes son ortogonales entre ellos, otros vectores propios contienen componentes negativas. La componente más grande del vector propio u^{35} (el vector propio asociado al único valor propio por encima de λ_+) corresponde a la empresa AMX. Esta es una de las empresas más importantes en el sector de telecomunicaciones, líder de servicios inalámbricos en América Latina y la cuarta más grande del mundo en términos de suscriptores proporcionales. La empresa AMX es una empresa con alta bursatilidad, con una participación del 23.79% en índice IPC, la más alta del índice. Las siguientes más altas componentes corresponden a empresas del sector de construcción: ICA, GEO y CEMEX. Las componentes más pequeñas corresponden a la embotelladora de Coca-Cola, AC y a la empresa GRUMA, líder en la producción de harina para tortillas, ambas del sector de productos de consumo frecuente. En cuanto a los vectores propios correspondientes a los valores propios más pequeños, se encuentra que están localizados, esto es, dan información de un solo sector. La mayor componente de u^1 corresponde a GEO y las menores, con signo contrario corresponden a ICA y URBI,

todas del sector de construcción. La mayor componente de u^2 es AMX siendo prácticamente la única componente significativa de este vector. El cociente inverso de participación (CIP) ayuda a cuantificar el número de componentes que participan significativamente en cada vector propio.

I^{35} tiene un valor de 0.032, por lo tanto, el vector propio u^{35} contiene $\frac{1}{I^{35}} = 30.74$ participantes significativos, esto es aproximadamente 30 empresas que son precisamente aquellas con mayor capitalización en el mercado. Los resultados obtenidos con los métodos de la Teoría de Matrices Aleatorias son consistentes con los obtenidos usando árboles de expansión mínima. La Figura 4 muestra el árbol de expansión mínima de las 35 empresas incluidas en el estudio. Este árbol fue construido con la distancia chi-cuadrada que permite distinguir los conglomerados existentes de una manera más clara (Medina, Pacheco, 2011) El árbol presenta dos conglomerados económicos, el de construcción con las empresas CEMEX, ICA, GEO, ARA, URBI y el de telecomunicaciones con AMX, TELMEX, TLVISA, AZTECA.

Figura 4: Árbol de Expansión Mínima en la Bolsa Mexicana de Valores



Esta figura muestra el árbol de expansión mínima de las 35 empresas más importantes de la Bolsa Mexicana de Valores. La empresa AMX es el vértice central. Dos conglomerados por sector económico son visibles en el árbol: el conglomerado de telecomunicaciones con triángulos negros (AMX, TELMEX, TLEVISA; AZTECA) y el conglomerado de construcción con círculos negros (ICA, GEO, ARA, URBI, ALFA). Los dos conglomerados, cercanos al vértice central, dominan las correlaciones del árbol, siendo los más influyentes en el mercado de valores mexicano.

Las empresas con mayores componentes en el vector propio más grande se encuentran precisamente en el centro del árbol y de manera sorprendente la empresa AMX, con el mayor peso en el vector propio más grande, resulta ser de manera indudable el vértice central del árbol. El hecho de que AMX sea el vértice central implica que el cambio en sus rendimientos está altamente correlacionado con el mayor número de empresas de la muestra, en términos de riesgo y diversificación conviene seleccionar acciones dentro y fuera de los vecinos de AMX. De acuerdo a la literatura las acciones cercanas al vértice central (los vecinos de AMX) son las menos riesgosas. Siendo AMX el vértice central del árbol, se considera como la empresa más importante de la muestra, afirmación que coincide con la apreciación del mercado al considerarla una de las mejores opciones para invertir en la bolsa y el hecho de ser la emisora de mayor peso en el IPC y la proveedora más importante de contenidos de habla hispana en el mundo.

La influencia del vértice central AMX en la Bolsa Mexicana de Valores puede ser positiva o negativa, recientemente la Bolsa Mexicana de Valores cerró su primera sesión de la semana con un retroceso de 0.27 por ciento, arrastrada por América Móvil (AMX), el principal indicador de la Bolsa Mexicana de Valores también fue influenciado negativamente por la caída preliminar de 1.13 por ciento de las acciones serie L de AMX. En este sentido el vértice central puede ser un agujero negro.

Las empresas en los extremos del árbol tienen componentes pequeñas en el vector u^{35} . Los vectores propios más pequeños están dominados por una o dos empresas cada uno, estas empresas son precisamente aquellas que obtuvieron mejores medidas en el árbol. La Tabla 2 muestra las empresas que obtuvieron los mejores resultados en las medidas que permiten escoger el vértice central. AMX es seleccionada como vértice central puesto que obtiene los mejores resultados en todas las medidas, de hecho, es la única que obtiene el primer lugar en todo.

Tabla 2: Medidas en el Árbol

	Centro De Masa	Peso	Grado	Nodos Promedio
amxl	0.50	2.00	10.00	2.74
tlevisapo	0.53	0.52	3.00	2.94
telmexl	0.55	0.94	5.00	3.06
femsaubd	0.58	0.70	4.00	3.26
ac	0.65	1.20	6.00	3.62
alsea	0.67	0.80	4.00	3.76
ica	0.68	0.19	1.00	3.71
ara	0.69	0.20	1.00	3.71
geob	0.70	0.21	1.00	3.71
walmexv	0.79	0.80	4.00	4.44

Esta tabla muestra los valores de las medidas topológicas y las empresas que obtuvieron los mejores. Se busca la empresa con centro de masa y nodos promedios mas bajos (el centro del árbol), la de mayor número de vecinos conectados (grado) y la de mayor peso. No siempre es posible hallar un nodo que sea el mejor en las cuatro medidas al mismo tiempo, en este estudio la empresa AMX lo logra. AMX es sin duda el vértice central del árbol. En esta lista aparecen las principales empresas del conglomerado de Telecomunicaciones (AMX, TLVISA, TELMEX) y las del conglomerado de construcción (ICA, ARA, GEO) mostrando la predominancia de estos dos conglomerados en el mercado de valores mexicano.

Las medidas obtenidas para la selección del vértice central y las componentes del vector propio más grande son consistentes con los pesos asignados a las empresas en el índice IPC, lo cual nos indica que un índice puede ser creado a partir de éstas medidas.

CONCLUSIONES

El espectro de la matriz de correlación de las 35 principales empresas de la Bolsa Mexicana de Valores muestra un comportamiento de banda aleatoria. Los valores propios en el intervalo $[\lambda_-, \lambda_+]$ satisfacen las propiedades universales de las matrices aleatorias, permitiéndonos separar los valores propios con información real de aquellos sin información o con información viciada. Solamente un valor propio se encuentra encima del intervalo $[\lambda_-, \lambda_+]$, su correspondiente vector propio muestra que la empresa de telefonía AMX tiene la componente más alta del vector mostrando su liderazgo en la bolsa mexicana de valores. El cociente inverso de participación indica que son 30 empresas las que aportan significativamente a este vector, las cuales coinciden con las 30 empresas de mayor peso en el índice IPC. Las empresas de construcción y telecomunicaciones se mantienen cercanas en peso a su conglomerado económico. La evidencia muestra que la Bolsa Mexicana de Valores forma conglomerados en los sectores de construcción y telecomunicaciones, con una gran influencia sobre las demás empresas; no hay evidencia de otros conglomerados. La empresa AMX resulta ser el vértice central, siendo la empresa con mayor peso, mayor número de vecinos y el centro de masa del árbol. Este resultado es consistente con el índice IPC que otorga a AMX el mayor peso del índice, casi 24%, el doble de la segunda empresa y el triple de la tercera empresa con pesos mas altos en el IPC.

Las empresas con mayores componentes en el vector propio más grande se encuentran precisamente en el centro del árbol y en las ramas exteriores aparecen las empresas con componentes más pequeños.

Al menos tres resultados importantes se obtienen de este estudio: 1) la matriz de correlación C , construida a partir de los rendimientos de las 35 empresas más importantes de la Bolsa Mexicana de Valores tiene una banda aleatoria que satisface las propiedades universales de las matrices aleatorias, lo cual permite

separar la parte de la matriz de correlación contaminada o con ruido de aquella que contiene información real. Analizando la información real por medio del cociente de participación inversa, se localizan las empresas que contribuyen de manera significativa a los cambios en la Bolsa Mexicana de Valores, siendo la de mayor contribución AMX. 2) el árbol de expansión mínima muestra una taxonomía de estas 35 empresas, seleccionando a la empresa AMX como vértice central; de acuerdo a la literatura, las acciones del portafolio de riesgo mínimo están localizadas lo más lejos posible del vértice central, y a medida que se mueve hacia portafolios con altos rendimientos esperados, las acciones incluidas en ese portafolio estarán localizadas cerca del vértice central. 3) Los resultados de la Teoría de Matrices Aleatoria y los del árbol de expansión mínima coinciden de manera importante en la matriz C , aumentando la validez del estudio. El poder de las medidas topológicas en el árbol y las componentes de los vectores propios significativos pueden ser explotados en investigaciones posteriores para la creación de índices, analizando matrices de correlación de variables pertenecientes a índices de precios al consumidor, índices de pobreza, etc. La metodología expuesta en este artículo tiene aplicaciones potenciales en otros campos donde el estudio de la verdadera correlación sea útil. En la Bolsa Mexicana de Valores esta metodología puede ser muy ventajosa al construir árboles dinámicos con precios intradía que muestren el cambio en la estructura del árbol a medida que se mueven los precios de las acciones.

BIBLIOGRAFIA

Barthélemy, M., Barrat, A., Pastor-Satorras, R., Vespignani, A. (2005) "Characterization and modeling of weighted networks", *Physica A* 346 34–43.

Bonanno, G. Caldarelli, G. Lillo, F. Mantegna, R. (2000) "Taxonomy of stock market indices", *Physical Review E* 62 R7615.

Bonanno, G. Caldarelli, G. Lillo, F. Mantegna, R. (2003) "Topology of correlation-based minimal spanning trees in real and model markets" *Physical Review E* 68 046130.

Bouchaud, L. Laloux, M.A. Miceli, and M. Potters, (2007) Large dimension forecasting models and random singular value spectra, *European Physical Journal B* 2 201-207.

Cukur, S., Eryigit, M., Eryigit, R., (2007) "Cross correlations in an emerging market financial data" *Physica A: Statistical Mechanics and its Applications* 376 555–564.

Dali, J. Crane, M. Ruskin, H. (2010) Random matrix theory filters and currency portfolio Optimization. *Journal of Physics: Conference Series* 221

Eom, C. Oh, G. Woo-Sung Jung, H. Jeong, S. Kim. (2009) "Topological properties of stock networks based on minimal spanning tree and random matrix theory in financial time series" *Physica A: Statistical Mechanics and its Applications* 388 900–906 .

Guhr, T. Kalber. (2003) A new method to estimate the noise in financial correlation matrices. *Phys. A: Math. Gen.* 36 3009–3032

Jung, W., Chae, S., Yang, J., Moon, H., (2006) "Characteristics of the Korean stock market correlations," *Physica A: Statistical Mechanics and its Applications* 361 263–271.

- Eom, C. Woo-Sung, J. Jeong, H. Kim, S. (2009) Topological properties of stock networks based on minimal spanning tree and random matrix theory in financial time series, *Physica A: Statistical Mechanics and its Applications* 388 900–906.
- Kulkarni, V. Deo, N. Correlation and volatility of an Indian stock market: a random matrix approach, *Eur. Phys. J. B* 60 (2007) 101-109.
- Kruskal, J., (1956), “On the Shortest Spanning Subtree and the Traveling Salesman Problem,” *Proceedings of the American Mathematical Society* 7, 48-50.
- Laloux, L. Cizeau, P. Bouchaud, Jean-Phillipe. Potters, Marc. (1999) *Random Matrix Theory and Financial Correlations. Mathematical Models and methods in applied sciences.* World Publishing company.
- Laloux, P. Cizeau, J-P Bouchaud, and M. Potters, (1999) Noise dressing of financial correlation matrices, *Phys. Rev. Letters* 83 1467-1470.
- Laloux, P. Cizeau, M. Potters, and J-P Bouchaud, (2000) *Random matrix theory and financial correlations, Mathematical Models and Methods in Applied Sciences*
- Livan, G. Rebecchi, L. (2012) Asymmetric correlation matrices: an analysis of financial data. arXiv:1201.6535v2 [q-fin.ST]
- Maslov S. (2001) Measures of globalization based on cross-correlations of world financial indices, *Physica A*, 301, 397-406.
- Mantegna, R., Stanley, H., (2000) “An Introduction to Econophysics: Correlations and Complexity in Finance”, Cambridge University Press, Cambridge.
- Medina, L., Mansilla, R., Un árbol de expansión mínima en la Bolsa Mexicana de Valores. (2007) *Revista de Administración, Finanzas y Economía (Journal of Management, Finance and Economics)* vol. 1, issue 2, pages 116-124
- Medina, L., M., Pacheco, E., (2012) “Caracterización y modelado de redes: índices financieros mundiales”. *Global Conference on Business and Finance Proceedings*. Vol 7. No 1. 774-779
- Medina, L., M., Pacheco, E., (2012) Distances and Networks: The case of Mexico. *Accounting and Taxation*. pp: 39-48. May . 2012.
- Mehta, Madan. (1991) *Random Matrices*. 2nd edition. Academic Press.
- Miccichè, S., Bonanno, G., Mantegna, R. (2003), “Degree stability of a minimum spanning tree of price return and volatility”. *Physica A-Statistical Mechanics and its Applications*, 324 (2003), pp. 66–73
- Onnela, J., Chakraborti, A., Kaski, K., Kertesz (2002) Dynamic asset trees and portfolio analysis. *The European Physical Journal B* 30, 285-288
- Onnela, J., Chakraborti, A., Kaski, K., Kertesz, J., Kanto, A., (2003) “Dynamics of market correlations: Taxonomy and portfolio analysis”, *Physical Review E* 68 056110.

Plerou, V. Gopikrishnan, P. Rosenow, B. Amaral, L. Stanley, H.E. (1999) Universal and nonuniversal properties and cross correlations in financial time series. *Phys. Rev. Lett.*, 83, 1471 -1474.

Plerou, P. Gopikrishman, B. Rosenow, L. A.N. Amaral, T. Guhr, and H. E. Stanley, (2002) Random matrix approach to cross-correlations in financial data, *Phys. Rev. E* 65 066126.

Plerou, P. Gopikrishman, B. Rosenow, L. A.N. Amaral, and H. E. Stanley, (2000). A random matrix theory approach to financial cross-correlations, *Physica A* 287 (2000) 374-382.

Potters, M. J-P Bouchaud, and L. Laloux, (2005). Financial applications of random matrix theory: old laces and new pieces, *Proceedings of the Cracow conference on “Applications of random matrix theory to economy and other complex systems”*

Rosenow, B. Plerou, Gopikrishnan, P. Amaral, L. Stanley, H. (2002) Application of random matrix theory to study cross-correlations of stock prices, *Int. J. of Theoret. and Appl. Finance* 3 399-403.

Sandoval, L. (2010) Correlation of financial markets in times of crisis. arxiv.org/abs/1102.1339

Tabak, B., Serra, T., Cajueiro, D., (2010) “Topological properties of stock market networks: The case of Brazil,” *Physica A* 389 3240-3249

Tola, V., Lillo, F., Gallegati, M., Mantegna, R., (2008) “Cluster analysis for portfolio optimization,” *Journal of Economic Dynamics & Control* 32, 235–258.

Tumminello, M., Mantegna, R., (2010) “Correlation, hierarchies, and networks in financial markets”, *Journal of Economic Behavior and Organization*, (doi:10.1016/j.jebo.2010.01.004).

Utsugi, A. Kazusumi, Ino. Oshikawa, M. (2003) Random Matrix Theory Analysis of Cross Correlations in Financial Markets. arXiv:cond-mat/0312643

Wigner, E. (1951) On the statistical distribution of the widths and spacings of nuclear resonance levels. *Proc. Cambridge Philos. Soc.* 47

BIOGRAFIA

Linda Medina Herrera es Doctora en Administración con Especialidad en Finanzas por el Tecnológico de Monterrey y la Universidad de Austin en Texas, tiene una maestría en Matemáticas por la Universidad de los Andes. Pertenece a la cátedra de investigación en Riesgos del Tecnológico de Monterrey. Tecnológico de Monterrey, CCM. Del puente 222 Col. Ejidos de Huipulco Tlalpan, 14380, México, D.F. México. Teléfono (55) 54832190. Fax +52(55) 54832163. linda.medina@itesm.mx

Ernesto Pacheco Velázquez es Doctor en Administración con Especialidad en Administración de Operaciones por el Tecnológico de Monterrey y la Universidad de Austin en Texas, tiene una maestría en Ingeniería Industrial por el Tec de Monterrey. Tecnológico de Monterrey CCM. Del puente 222 Col. Ejidos de Huipulco Tlalpan, 14380, México, D.F. México. Teléfono (55) 54832190. Fax +52(55) 54832163. epacheco@itesm.mx