

## Capítulo 3

### 3.1 Resumen

Aunque la sabiduría popular pregona que “El dinero no se da en los árboles”, en este capítulo se quiere mostrar que el concepto de árbol (grafo) tiene aplicaciones potenciales en el análisis de mercados financieros, en particular, en la Bolsa Mexicana de Valores (BMV).

Se analizará la correlación de los rendimientos de las principales acciones que cotizan en la BMV, usando árboles de expansión mínima, un novedoso método que en el contexto financiero fue desarrollado por Mantenga (1998) y que ha tenido aplicaciones importantes en los últimos años. Los árboles nos permiten visualizar de una manera muy especial la estructura de correlación del mercado.

En este capítulo se presenta la construcción de los árboles de expansión mínima generados a partir de las matrices  $C$ ,  $C^{(q)}$  y  $C_f$ . Se muestra las diferencias y semejanzas de los mismos, a través del estudio de conglomerados.

Usando el concepto de vértice central, el nodo que más conexiones tiene, y promedio de ocupación, se busca aquellas acciones que hacen parte de un portafolio con riesgo mínimo y también aquellas pertenecientes al portafolio con altos rendimientos esperados.

### 3.2 Construcción y análisis de los árboles de expansión mínima.

Se construyen los árboles de expansión mínima generados a partir de las matrices  $C$ ,  $C^{(q)}$  y  $C_f$ .

Las componentes  $d_{ij}$  de las matrices de distancias darán el peso del arco que une el nodo  $i$  con el nodo  $j$ .

La componente  $d_{ij}$  es:

$$d_{ij} = \sqrt{2(1 - \rho_{ij})}$$

Tomando  $\rho_{ij}$  de las matrices  $C$ ,  $C^{(q)}$  y  $C_f$ , se construye sus correspondientes matrices de distancia  $D$ ,  $D^{(q)}$  y  $D_f$ .

El objetivo es encontrar arreglos jerárquicos de acciones a través del estudio de conglomerados de compañías e investigar si es posible determinar cuándo el movimiento de los precios de una acción precede al movimiento del precio de otra acción.

La Figura 2.26 muestra los colores de los vértices y las empresas de acuerdo al sector económico.

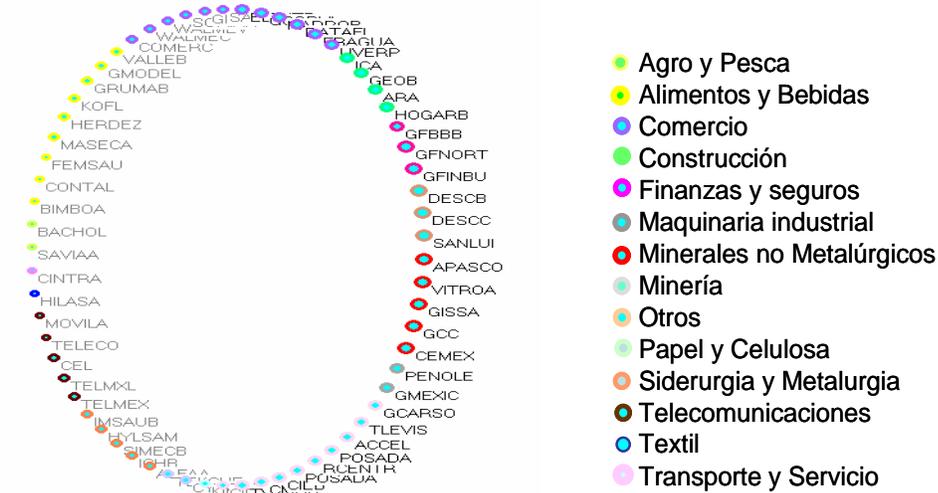


Figura 2.26 Sectores Económicos

Las Figuras 2.27, 2.28 y 2.29 muestran los árboles de expansión mínima de  $C$ ,  $C^{(q)}$  y  $C_f$ . Los árboles fueron hechos con el software Pajek y el camino mínimo con Windqsb.

Lo que los árboles tienen en común:

1. El conglomerado (cluster) más grande está liderado por Cemex, empresa líder en la producción y comercialización de cemento, concreto y productos relacionados. Todo el sector de minerales no metalúrgicos se mueve alrededor de Cemex: Apasco, Vitro, Gissa y GCC (cementos Chihuahua). En los árboles  $C$  y  $C^{(q)}$  las empresas de construcción: Geo, Ara, Hogar e incluso ICA, la

empresa de ingeniería, procuración y construcción más grande de México, también permanecen cerca de Cemex.

2. El segundo conglomerado es el de las empresas de telecomunicaciones, dominado por Telmex, en este conglomerado se encuentran Iusacell, Telecom Carso y Movil Acces.
3. Las acciones de Wal Mart de México aparecen en las ramas exteriores del árbol, mostrando cierta independencia del resto del mercado. De la misma forma aparece Peñoles, el mayor productor de plata y oro afinado, Gmodelo, empresa líder en elaboración, distribución y venta de cerveza y sorprendentemente Cintra, la empresa que reúne las más importantes líneas aéreas mexicanas y cuyo mayor accionista es el gobierno mexicano. En las ramas exteriores de los árboles  $C$  y  $C^{(q)}$  se encuentra al Grupo Posadas, la operadora hotelera más importante en México.
4. Las acciones de los bancos Bancomer (GFBBB) y Banorte (GFNorte) aparecen unidas en todos los árboles, no así Inbursa, que siempre aparece en otra rama.
5. Las acciones de las televisoras Tvazteca y Televisa se encuentran muy cerca. Del nodo de Televisa se desprende no solamente Tvazteca sino también las acciones de Kofl (Coca-cola) y Cel (Iusacell).



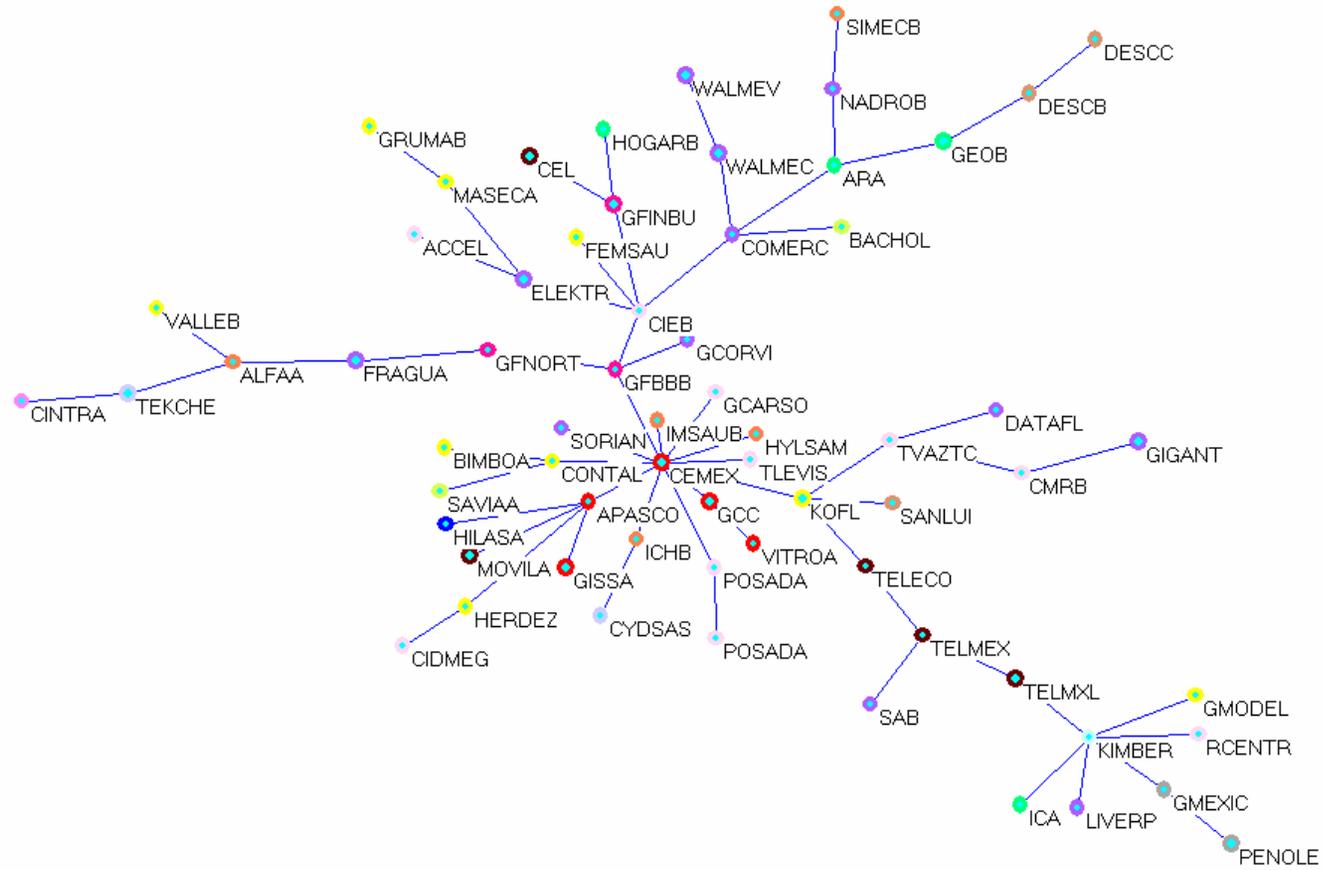


Figura 2.28 Árbol de expansión mínima de  $C_f$

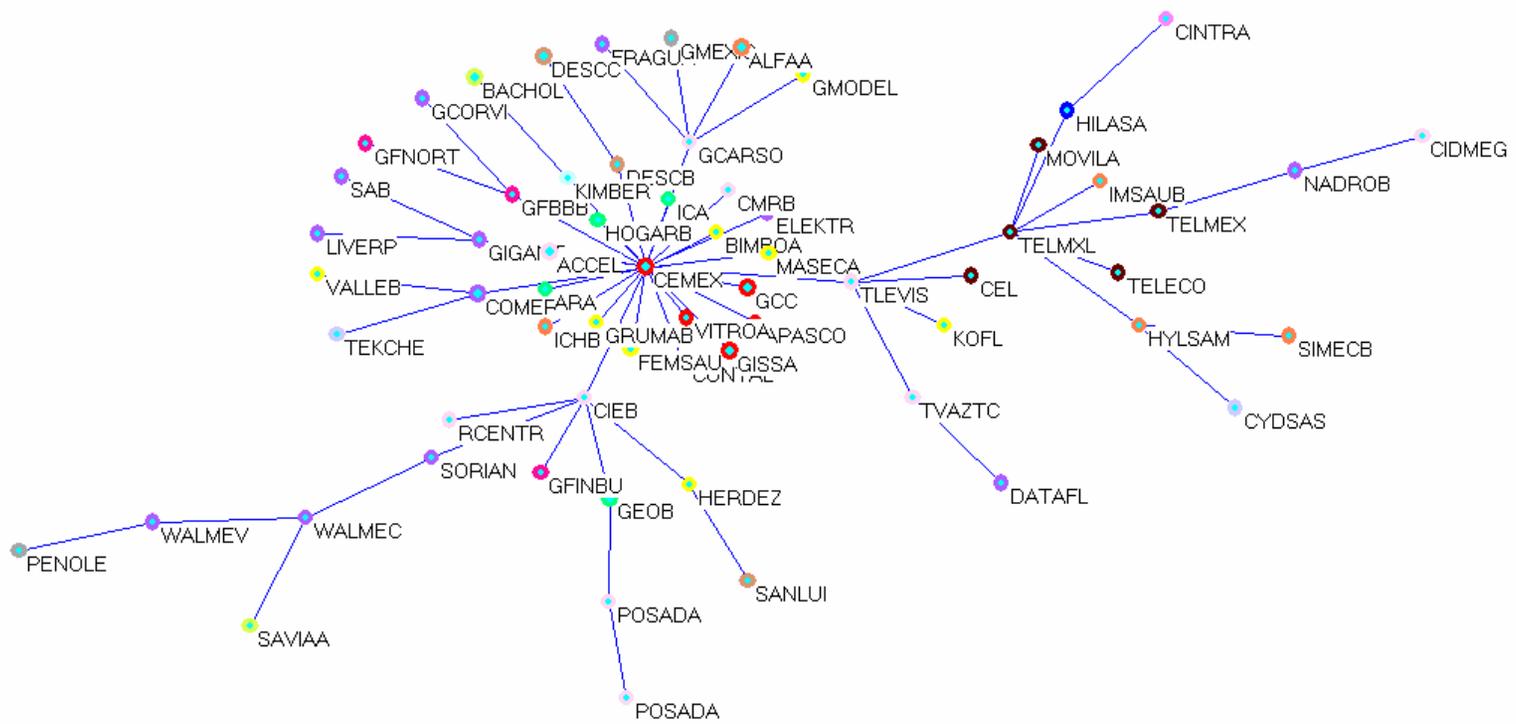


Figura 2.29 Árbol de expansión mínima de C<sup>q</sup>

6. El nodo correspondiente a CIEB (Corporacion Interamericana de Entretenimiento SA de CV) es el origen de una rama del árbol que tiene influencia sobre acciones de diferentes sectores económicos como Soriana (comercio) , GFImbuero (finanzas y seguros) y Herdez (alimentos y bebidas) , entre otros.

Los árboles  $C$  y  $C^q$  contienen prácticamente la misma información, esto es, presentan los mismos conglomerados y las acciones en las ramas exteriores coinciden. La ventaja del árbol  $C^q$  es que muestra en forma más clara los conglomerados.

El árbol  $C_f$  distribuye las acciones de una manera muy diferente a  $C$  y  $C^q$  pero aún así, los conglomerados liderados por Cemex y Telmex se conservan, y las empresas como Peñoles y Walmart continúan en las ramas externas del árbol. Con este árbol se gana un nuevo conglomerado “las empresas de Comercio”, un poco esparcido pero todo en la parte superior del árbol: Comercial mexicana, Walmart, Gigante, Elektra, Nadro (farmacéuticos, higiene y belleza), Liverpool y Soriana.

En árboles de acciones altamente capitalizadas en mercados financieros de Estados Unidos [5] y [40], los conglomerados formados por sectores

económicos son muy claros, no así en los árboles del mercado mexicano, donde los únicos conglomerados por sector económico son: el de minerales no metalúrgicos y el de telecomunicaciones.

### 3.3 Análisis de portafolios

Se aplica los conceptos de promedio de ocupación, y vértice central al problema de optimización de un portafolio.

Para comenzar, se analizará el árbol de expansión mínima construido a partir de C.

El criterio del grado del vértice nos deja a Cemex con 23 vecinos, seguido por Telmex con 7 y Cieb con 5.

El criterio del peso del vértice nos da como resultado: en primer lugar Cemex con 25.0354 y segundo lugar Telmex con 5.3976.

Por último, el promedio de ocupación mas bajo lo obtuvo Cemex, quedando de manera indiscutible como vértice central.

Es muy interesante comparar los resultados de Cemex con el que se podría esperar que fuera el vértice central: Telmex.

Empresa	Grado del vértice (Número de vecinos)	Peso del vértice (Suma de corr)	%IPC	# Emp
Cemex	23	25.0354	25%	18
Telmex	7	5.3976	5%	1

La tabla muestra que de los 23 vecinos de Cemex, 18 pertenecen al IPC para un total del 25% de participación en el IPC sin contar la participación de Cemex. De los 7 vecinos de Telmex solo uno pertenece al IPC con 5% de participación.

La Figura 2.30 muestra las empresas alrededor de Cemex. Entre las cuales encontramos a Femsa, Televisa, Gfbb, Gcarso, Bimbo y los conglomerados de construcción (en verde) y de minerales no metalúrgicos (en rojo)

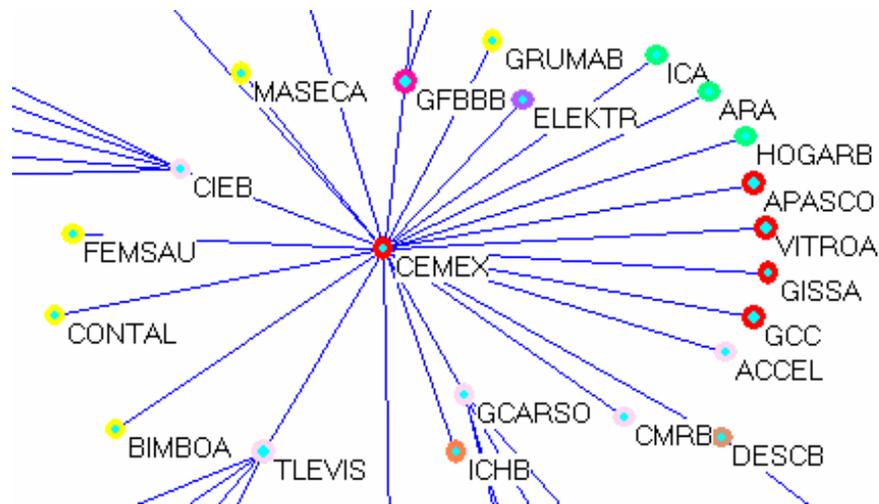


Figura 2.30 Cemex: el vértice central en el árbol C

La Figura 2.31 muestra las empresas alrededor de Telmex que son prácticamente el conglomerado de telecomunicaciones.

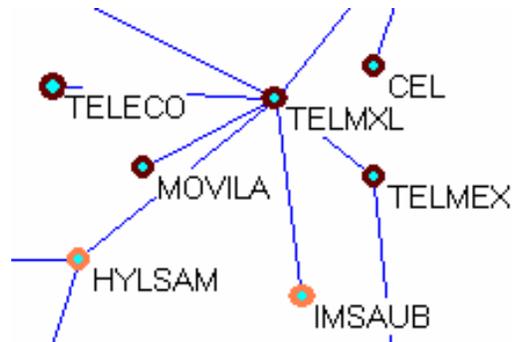


Figura 2.31 Los vecinos de Telmex

La existencia de un centro significativo en el árbol no es un asunto trivial y menos el hecho de que éste coincida con el centro de masa.

La Figura 2.32 confirma, entre otras cosas, que Cemex es el vértice central y/o centro de masa. Es importante recordar que el vértice central  $v_c = Cemex$  es considerado la raíz del árbol. Se usará como punto de referencia, contra el cual la posición de los demás vértices es relativa.

Partiendo del vértice central se construye el nivel  $niv(v_i)$  del vértice  $v_i$  en relación con el vértice central, cuyo nivel se toma como cero.

El  $niv(v_i)$  es la distancia en el árbol de  $v_i$  al vértice central. Por ejemplo:

$$niv(46) = niv(Posadas) = 0.93 + 1.05 + 1.19 + 1.22 = 4.39$$

La Figura 2.30 muestra las distancias entre los nodos del árbol de expansión mínima. Con ellas se construye el  $niv(v_i)$ .

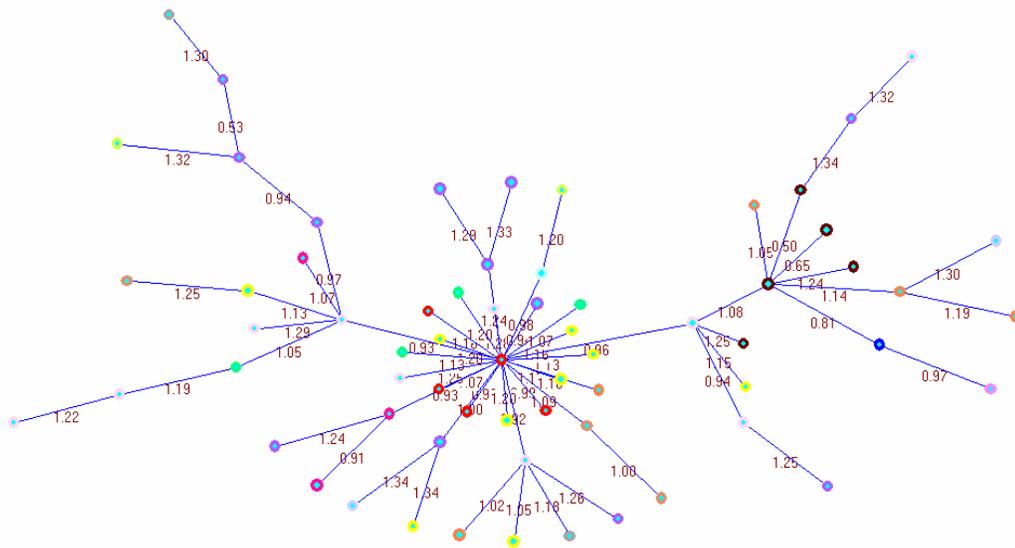


Figura 2.30 Distancias medidas desde Cemex

La Tabla 2.31 muestra el  $niv(v_i)$  en el árbol C con  $1 \leq i \leq 65$ .

$i$	Empresa	$niv(v_i)$	$i$	Empresa	$niv(v_i)$
50	CIDMEGA	5.2018	60	TELMEXL	2.0404
39	PE&OLES	4.6751	30	GFINBURO	1.9992
52	CYDSASA	4.4820	25	GEO	1.9819
46	POSADAS	4.3912	10	GMODELO	1.9652
56	SIMEC	4.3767	54	ALFA	1.9423
1	SAVIAA	4.1691	15	SORIANA	1.9026
20	NADRO	3.8837	49	TVAZTCA	1.9015
65	CINTRA	3.8221	29	GFNORTEO	1.8439
14	WALMEXV	3.3790	43	ACCELSA	1.2801
33	SANLUIS	3.3083	48	CMR	1.2535
63	MOVILAB	3.2851	16	GIGANTE	1.2445
57	HYLSAMX	3.1848	27	HOGAR	1.2026
44	POSADAS	3.1740	6	MASECA	1.2023
21	DATAFLX	3.1492	9	GRUMA	1.1979
58	IMSA	3.0860	4	CONTAL	1.1646
64	HILASALA	2.8471	36	GISSA	1.1642
13	WALMEXC	2.8467	5	FEMSA	1.1487
62	TELECOM	2.6924	3	BIMBO	1.1346
17	SAB	2.5746	24	ICA	1.1269
59	TELMEXA	2.5450	55	ICHB	1.1046
23	LIVEPOL1	2.5395	31	DES	1.0948
11	VALLEB	2.3465	26	ARA	1.0683
53	TEKCHEM	2.3446	35	VITRO	1.0667
45	RCENTRO	2.2156	12	COMERCI	1.0040
61	CEL	2.2103	37	GCC	0.9888
2	BACHOCO	2.1806	51	KIMBER	0.9796
22	FRAGUA	2.1769	42	TLEVISA	0.9644
19	GCORVI	2.1680	18	ELEKTRA	0.9383
8	KOFL	2.1188	34	APASCO	0.9330
40	GMEXICO	2.0988	28	GFBBB	0.9295
32	DESCC	2.0931	47	CIEB	0.9281
7	HERDEZ	2.0557	41	GCARSO	0.9185

Tabla 2.31  $niv(v_i)$  en C

Como se ha visto, las acciones del portafolio de riesgo mínimo están localizadas lo más lejos posible del vértice central. En el árbol C estas resultan ser: Las hoteleras Cid Mega y Posadas, la minera Peñoles, Walmart de comercio y Cintra de transporte aéreo. A medida que se mueve hacia portafolios con altos rendimientos esperados, las acciones incluidas en ese portafolio están localizadas cerca del vértice central: Cemex, Ica, Apasco, Maseca, Gfbbb y Femsa entre las 23 que rodean el vértice Cemex.

El “promedio de ocupación”  $l(Cemex)$  es:

$$l(Cemex) = \frac{1}{65} \sum_{i=1}^{65} niv(v_i) = 2.14$$

En las bolsas financieras americanas el promedio de ocupación oscila entre 3 alcanzado en 1986 y 9.5 en 1994, manteniéndose la mayoría del tiempo por encima de 4.

El valor 2.14 nos muestra que aunque el mercado mexicano no está en crisis, es aún un mercado en desarrollo, uno donde el comportamiento del sistema es todavía muy homogéneo.

Los resultados del árbol  $C^q$  son similares a  $C$ . A continuación se presentarán los resultados del árbol  $C_f$ .

El vértice central en este árbol también es Cemex, obteniendo los siguientes resultados:

El criterio del grado del vértice nos deja a Cemex con 11 vecinos, seguido por Cieb y Apasco, ambos con 4.

El criterio del peso del vértice nos da como resultado: en primer lugar Cemex con 11.47, segundo lugar Apasco con 4.59 y tercero Cieb con 4.01.

El promedio de ocupación más bajo lo obtuvo Cemex y es:

$$l(Cemex) = \frac{1}{65} \sum_{i=1}^{65} niv(v_i) = 3.23$$

El promedio de ocupación es más alto que el del árbol  $C$  pero aún muestra una diferencia importante con los mercados más desarrollados.

Las acciones cercanas al vértice central son: Apasco, Gfbbb, Kofl, Televisa y Soriana, entre otras.

Tanto el vértice central como las principales empresas que lo rodean coinciden en los tres árboles, siendo el árbol  $C_f$  más selectivo.

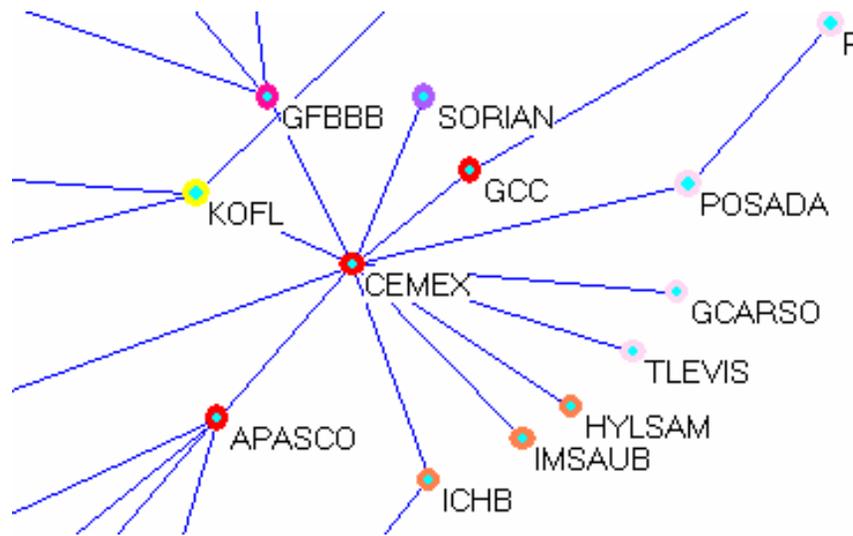


Figura 2.32 Cemex: el vértice central en el árbol  $C_f$

La Figura 2.33 muestra los vecinos de Cie (Corporación interamericana de entretenimiento)

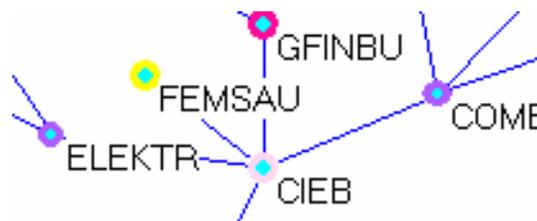


Figura 2.33 Vecinos de Cieb

La Tabla 2.33 muestra el  $niv(v_i)$  en el árbol  $C_f$  con  $1 \leq i \leq 65$ .

$i$	Empresa	$niv(v_i)$
39	PE&OLES	7.1476
32	DESCC	6.9534
65	CINTRA	6.2568
9	GRUMA	6.2054
56	SIMEC	6.1536
40	GMEXICO	5.8594
31	DESC	5.8157
45	RCENTRO	5.7628
23	LIVEPOL	5.6946
24	ICA	5.6899
10	GMODELO	5.4946
43	ACCELSA	5.1581
11	VALLE	5.001
6	MASECA	4.9644
53	TEKCHEM	4.9176
20	NADRO	4.8602
51	KIMBER	4.5988
25	GEO	4.5776
16	GIGANTE	4.2328
17	SAB	4.1227
2	BACHOCO	3.9974
14	WALMEXV	3.9946
18	ELEKTRA	3.8295
27	HOGAR	3.8173
60	TELMEXL	3.8146
61	CEL	3.6852
54	ALFA	3.6837
13	WALMEXC	3.6689
26	ARA	3.6377
21	DATAFLX	3.4114
48	CMR	3.2073
5	FEMSA	2.9895
50	CIDMEGA	2.9097
59	TELMEXA	2.8256
12	COMERCI	2.7093
30	GFINBURO	2.584
1	SAVIA	2.5836
22	FRAGUA	2.5149
3	BIMBO	2.3188
52	CYDSASA	2.2656
46	POSADASA	2.2293
35	VITRO	2.1656
19	GCORVI	2.1366
49	TVAZTCA	2.0764
64	HILASAL	2.0443
33	SANLUIS	2.009
36	GISSA	1.9805
63	MOVILA	1.9573
62	TELECOM	1.8475
7	HERDEZ	1.776
47	CIEB	1.7027
29	GFNORTEO	1.4345
58	IMSA	1.2467
4	CONTAL	1.2291
15	SORIANA	1.1965
57	HYLSAMX	1.184
44	POSADASL	1.182
41	GCARSO	1.0417
42	TLEVISA	1.0292
55	ICHB	1.0115
37	GCC	0.9763
8	KOFL	0.9156
28	GFBBB	0.8914
34	APASCO	0.7919

Tabla 2.31  $niv(v_i)$  en  $C_f$

En las ramas exteriores coinciden en todos los árboles: Peñoles, Simec, y Cintra. Alrededor de Cemex coinciden: Apasco, Gfbbb, Televisa, Gcc, Ichb y Gcarso.

Es importante observar que el promedio de ocupación fluctúa con el tiempo, se pueden construir árboles dinámicos y compararlos. En los mercados americanos las crisis del mercado muestran cambios importantes en la estructura de los árboles. Además, nuestros estudios de la estructura libre de escalas del árbol de expansión mínima muestran que esta gráfica no es únicamente jerárquica en el sentido de un árbol, sino que posee nodos altamente conectados, desde los cuales se construye la estructura jerárquica.

Finalmente, el árbol de activos se puede ver como una poderosa herramienta gráfica, pues aunque parece estar fuertemente reducido, contiene información esencial del mercado y se puede usar para añadir un juicio subjetivo al problema de optimización de un portafolio.

## **INTRODUCCIÓN**

Las matrices de correlación empíricas son de gran importancia en la administración de riesgos y la asignación de activos. La probabilidad de grandes pérdidas para cierto portafolio está sujeta a los movimientos correlacionados de sus componentes. El estudio de matrices de correlación tiene una larga historia en finanzas y es una de las piedras angulares en la teoría de optimización de portafolios de Markovitz.

Uno de los mayores problemas en la determinación empírica de una matriz de correlación es la finitud de las series de tiempo de los activos que la componen. Si la longitud de la serie no es muy larga, uno puede esperar que la determinación de la covarianza esté viciada, resultando de ello una matriz de correlación cuya estructura depende en gran medida del ruido. Otra de las dificultades para analizar el significado y la importancia de la matriz de correlación empírica se debe a que las condiciones del mercado cambian con el tiempo y la correlación que existe entre un par de activos puede no ser permanente.

Todo lo anterior, sin duda alguna incide en la elaboración de portafolios eficientes. Muchos estudios se han realizado para detectar y eliminar el ruido de las matrices financieras, algunos de ellos provenientes de otras disciplinas. En este trabajo se presentan tres métodos para obtener la información real de una matriz de correlación empírica. Los dos primeros

se aplicarán por primera vez en el mercado mexicano y el tercero, aunque ya conocido, da pie para analizar de una nueva manera la estructura de correlación de las principales acciones de la Bolsa Mexicana de Valores (BMV). Los métodos son:

1. La Teoría de Matrices Aleatorias. La cual analiza el espectro de la matriz de correlación distinguiendo los valores y vectores propios que contienen información real de aquellos que tienen información inútil e inestable en el tiempo.
2. La matriz Potencia. Este método puede ser interpretado como una prolongación efectiva de la series de tiempo. El método consiste en aplicar una función potencia a la matriz de correlación  $C$  convirtiéndola en una matriz  $C^p$  con la misma información de  $C$  pero quitándole eficientemente el ruido.
3. La serie de Fourier. Este es uno de los métodos que se usa en la teoría de comunicaciones, para diferenciar las señales del ruido.

Además de obtener información real de la matriz de correlación de la BMV, se usa la noción de árbol de expansión mínima, proveniente de la teoría de grafos, como un método para encontrar arreglos jerárquicos de acciones a través del estudio de conglomerados de compañías. Como veremos a lo largo de este trabajo, esto puede ser de gran utilidad para la

construcción de portafolios eficientes, con una perspectiva diferente a la elaborada por Markowitz.

En este trabajo se presenta, la topología de correlación y tres árboles de expansión mínima que reflejan los movimientos de las principales acciones de la BMV. Los árboles de expansión mínima han sido usados exitosamente como herramienta en la elección de un portafolio, puesto que los activos que minimizan el riesgo se encuentran en las ramas exteriores del árbol.

Se analiza la posición relativa de las acciones de la BMV en el árbol de expansión mínima y se podrá observar que la longitud y el centro del árbol (vértice central) ofrecen información importante sobre el mercado mexicano y sobre las acciones de la BMV elegibles para un portafolio con riesgo mínimo y/o con altos rendimientos esperados.

En el Capítulo I se presentan las herramientas teóricas que serán utilizadas en el análisis de la matriz de correlación empírica. En el Capítulo II se construyen y analizan la matriz de correlación  $C$ , la matriz potencia  $C^p$  y la matriz con las series de Fourier  $C_f$ . El Capítulo III muestra la construcción y análisis de los árboles de expansión mínima con su aplicación a la optimización de portafolios. Finalmente se enuncian las conclusiones del trabajo.

# CAPITULO 1

## 1.1 Resumen

En la minimización del riesgo de un portafolio es muy importante considerar cómo los rendimientos de diferentes compañías están correlacionados con otros. Por otro lado, es bien conocido que las matrices de correlación financieras poseen ruido aleatorio que en ocasiones perturba de manera significativa a la señal original.

En la primera parte de este capítulo se presentan tres métodos diferentes para estimar el ruido y obtener información real de la matriz de correlación de acciones de la Bolsa Mexicana de Valores.

El primer método incluido en este estudio, corresponde a una aplicación de la Teoría de Matrices Aleatorias en Física Atómica, desarrollada originalmente por Wigner [71] y llevado a su plenitud por Mehta [44], permite distinguir los valores y vectores propios de la matriz de correlación que contienen información real de aquellos que tienen información inútil e inestable en el tiempo.

El segundo, es un método nuevo desarrollado por Guhr y Kalber (2002) [26] basado en una transformación de la matriz  $C$  en una matriz potencia  $C^q$  que suprime el ruido de manera eficiente.

El último es un método muy conocido, el que se usa en la teoría de comunicaciones, justamente para diferenciar las señales del ruido: la transformada de Fourier y el espectro de potencias.

La segunda parte del capítulo presenta una forma de investigar la correlación y la eficiencia de un portafolio, usando “árboles de expansión mínima”. Este concepto, en el ambiente de mercados financieros, fue introducido por Mantenga, Bonano y otros (2002) [5] como un método para encontrar arreglos jerárquicos de acciones a través del estudio de conglomerados de compañías.

Cada uno de estos temas viene acompañado de las definiciones pertinentes y de una breve historia de su desarrollo, mencionando los principales autores que han aportado a los mismos.

## **1.2 Matrices de Correlación**

Aunque la idea de correlación había sido usada con anterioridad, no fue sino hasta 1885 que Sir Francis Galton la definió formalmente como un tópico estadístico. Una década después, Karl Pearson desarrolló el índice que todavía se usa para medir la correlación, el coeficiente de correlación de Pearson “ $\rho$ ”. En 1923 Pearson menciona el trabajo de Gauss sobre la

distribución normal de  $n$  variables correlacionadas. Sin embargo Gauss no tenía ningún interés en la correlación como una noción conceptualmente diferente, El la interpretaba como uno de los varios parámetros de sus ecuaciones de distribución. En 1868 el primo de Galton, Charles Darwin, usó el concepto de correlación mencionando que “todas las partes de una organización están, hasta cierto punto, conectadas o correlacionadas entre si”. Durante el mismo periodo, muchos filósofos contribuyeron al concepto de correlación. John Stuart Mill (1843) en sus “Cinco cánones de la investigación experimental”, afirmaba que “Cualquier fenómeno que varía de cierta manera cuando otro fenómeno varía de alguna manera particular, es una causa o un efecto de tal fenómeno, o está relacionado con el por algún hecho o causa”.

La fórmula para medir la correlación tal y como hoy se conoce, fue desarrollada por Pearson en 1895. Hoy, más de un siglo después, no se aprecia el hecho de que antes de Galton y Pearson, la única forma de establecer una relación entre variables era obtener una conexión de causa y efecto. En la actualidad el coeficiente de correlación constituye una de las principales metodologías estadísticas para experimentos en muchas disciplinas.

El coeficiente de correlación de Pearson es un índice sin dimensión, invariante bajo transformaciones lineales de cualquiera de las variables.

La fórmula de Pearson establece que un estimador muestral para la correlación  $\rho$  es:

$$\rho = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

En el numerador aparece la suma acumulada de los productos de las variables centradas por su media. El denominador ajusta las escalas de tal forma que desaparecen las unidades. Esta ecuación describe a  $\rho$  como la suma centrada y estandarizada del producto de dos variables.

Debido a la desigualdad de Cauchy–Schwarz, el coeficiente de correlación satisface

$$-1 \leq \rho \leq 1$$

tomando tres valores especiales

$$\rho = \begin{cases} 1 & \text{Si las variables est\u00e1n completamente correlacionadas} \\ 0 & \text{Si no hay correlaci\u00f3n entre las variables} \\ -1 & \text{Si las variables tienen correlaci\u00f3n completamente inversa} \end{cases}$$

En los mercados financieros, muchas acciones son operadas simult\u00e1neamente. Una forma de detectar diferencias y similitudes en la evoluci\u00f3n en el tiempo de un par de acciones  $i$  y  $j$  es estudiar el

coeficiente de correlación  $\rho_{ij}$  entre los logaritmos de los rendimientos de los activos.

Se tienen N acciones con precio  $P_i(t)$  para el activo  $i$  en el tiempo  $t$ , con  $t = 0, 1, \dots, T$ . Sea  $S_i(t)$  el logaritmo de los rendimientos de los activos

$$S_i(t) = \ln P_i(t) - \ln P_i(t-1) \quad (1.1)$$

De acuerdo a la fórmula de Pearson el coeficiente de correlación  $\rho_{ij}$  se calcula así:

$$\rho_{ij} = \frac{\sum_{t=1}^T (S_i(t) - \bar{S}_i(t))(S_j(t) - \bar{S}_j(t))}{\sqrt{\sum_{t=1}^T (S_i(t) - \bar{S}_i(t))^2 \sum_{t=1}^T (S_j(t) - \bar{S}_j(t))^2}}$$

Con los números  $\rho_{ij}$  se forma una matriz, la “matriz de correlación” que es la matriz cuyas entradas son precisamente  $\rho_{ij}$ :

$$C = \begin{bmatrix} \rho_{11} & \rho_{12} & \rho_{13} & \cdots & \rho_{1N} \\ \rho_{21} & \rho_{22} & \rho_{23} & \cdots & \rho_{2N} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{N1} & \rho_{N2} & \cdots & & \rho_{NN} \end{bmatrix}$$

Puesto que  $\rho_{ij} = 1$  si  $i = j$  y  $\rho_{ij} = \rho_{ji}$  la matriz de correlación es una matriz cuadrada  $N \times N$ , simétrica, en cuya diagonal sólo hay unos. La matriz tiene a lo más  $\frac{N(N-1)}{2}$  elementos diferentes.

Las matrices de correlación empíricas son de gran importancia en la administración de riesgos y en la asignación de activos.

La probabilidad de grandes pérdidas para cierto portafolio está controlada por movimientos correlacionados de sus diferentes componentes, por ejemplo, una posición que es simultáneamente larga en acciones y corta en bonos puede ser riesgosa porque acciones y bonos usualmente se mueven en direcciones opuestas en periodos de crisis.

El estudio de matrices de correlación tiene una historia larga en finanzas y es una piedra angular de la teoría de optimización de portafolios de Markowitz.

Se han hecho muchos intentos para resolver el problema de la optimización de un portafolio, desde el clásico estudio de Markowitz hasta formas mas sofisticadas, con tratamientos estocásticos. En todos los intentos de resolver el problema, el coeficiente de correlación entre los precios de los activos juega un papel importante para cuantificar el grado de sincronización de una acción  $i$  con una acción  $j$ .

En éste trabajo, el coeficiente  $\rho_{ij}$  permitirá definir una distancia relativa entre las acciones con las cuales se construye y proveerá de un método para extraer información económica guardada en las series de tiempo de los precios de las acciones. Se utilizarán solo aquellas que cotizan en la Bolsa Mexicana de Valores.

Las propiedades estadísticas de la matriz de correlación de rendimientos han sido investigadas en diferentes campos como economía, finanzas y física. En economía, se ha usado principalmente para determinar el número de  $k$  factores presentes en los mercados financieros usando la teoría de fijación de precio del arbitraje (APT) originalmente desarrollada por Ross [1976].

### 1.2.1 Valores y vectores propios

Si  $C$  es una matriz  $N \times N$ , entonces un vector  $\mathbf{x}$  diferente de cero en  $\mathbf{R}^n$  se denomina vector propio de  $C$  si  $C\mathbf{x}$  es un múltiplo escalar de  $\mathbf{x}$ ; es decir,

$$C\mathbf{x} = \lambda\mathbf{x}$$

para algún escalar  $\lambda \neq 0$ . El escalar  $\lambda$  se denomina valor propio de  $C$ , y se dice que  $\mathbf{x}$  es un vector propio correspondiente a  $\lambda$ .

Los valores propios de la matriz de correlación también han sido tema de estudio en varias disciplinas. Stanley y otros [57] utilizan los vectores propios correspondientes a los valores propios más grandes para hacer una partición del conjunto de todas las acciones en subconjuntos, encontrando esta partición similar a los sectores económicos convencionalmente identificados.

Se requiere bastante trabajo para determinar empíricamente una matriz de correlación, de manera confiable. Si se tiene un conjunto de  $N$  acciones diferentes, las  $\frac{N(N-1)}{2}$  entradas de la matriz de correlación, deben ser

determinadas con  $N$  series de tiempo de longitud  $T$ ; si  $T$  no es muy largo comparado con  $N$ , uno puede esperar que la determinación de la covarianza este llena de ruido, resultando de ello una matriz de correlación cuya estructura depende en gran medida del ruido, por lo tanto hay que tener cuidado al utilizar este tipo de matriz de correlación en aplicaciones.

Otra de las dificultades para analizar el significado y la importancia de la matriz de correlación empírica se debe a que las condiciones del mercado cambian con el tiempo y la correlación que existe entre un par de activos puede no ser permanente.

Muchos estudios se han realizado para detectar y eliminar el ruido producido por la finitud de las series de tiempos de los rendimientos. Las

estimaciones también pueden verse afectadas por el sentido variable de las correlaciones cruzadas, que contribuyen con factores aleatorios a la matriz de correlación. En general es un problema difícil estimar de  $C$  las correlaciones cruzadas que no son producto de la aleatoriedad. Así que es importante utilizar métodos que permitan distinguir “señales” de “ruidos”. En este trabajo se usan tres métodos diferentes para detectar y eliminar el ruido:

1. La teoría de matrices aleatorias
2. La matriz potencia
3. La transformada de Fourier

Los resultados serán comparados, discutiendo debilidades y fortalezas para obtener mejores conclusiones.

El primer método incluido en nuestro estudio, que corresponde a una aplicación de la Teoría de Matrices Aleatorias desarrollada por Mehta [44], permite distinguir los valores y vectores propios de la matriz de correlación que contienen información real de aquellos que tienen información inútil e inestable en el tiempo. El método consiste en comparar las propiedades de una matriz de correlación  $C$  contra la “hipótesis nula” de una matriz de correlación aleatoria, esto es una matriz de correlación construida con

series de tiempo de activos estrictamente independientes. Si las propiedades de  $C$  se asemejan a aquellas de la matriz de correlación aleatoria, entonces se sigue que el contenido de la matriz  $C$  es aleatorio. Recíprocamente, si las propiedades de las dos matrices difieren, la correlación es genuina. Una de las contribuciones valiosas de la Teoría de Matrices Aleatorias es que permite separar el contenido de  $C$  en dos grupos: la parte de  $C$  que tiene propiedades semejantes a las de la matriz aleatoria (“ruido”) y la parte de  $C$  que difiere de la matriz aleatoria (“información”).

El segundo, es un método nuevo desarrollado por Guhr y Kalber (2002) [26], basado en trabajar con una nueva matriz  $C^{(q)}$  que contiene la información relevante de la matriz  $C$ , pero sin ruido y que en cierta forma es equivalente a alargar la serie de tiempo de los rendimientos, esto es hacer  $T \rightarrow \infty$ .

El último es un método muy conocido, que se usa en la teoría de comunicaciones, justamente para diferenciar las señales del ruido: la transformada de Fourier y su correspondiente espectro de potencias. Se construirá la matriz  $C_f$  formada por las aproximaciones de las  $N$  series usando la transformada de Fourier.

### 1.3 Teoría de Matrices Aleatorias

El estudio de las propiedades estadísticas de las matrices aleatorias tiene sus orígenes históricos en la física nuclear [71]. El problema data de los años cincuenta; los modelos existentes no explicaban de manera adecuada los niveles de energía de un núcleo complejo. La Teoría de Matrices Aleatorias (TMA) desarrollada en ese contexto fue presentada por Wigner, Dyson y Mehta [46]. Ellos establecían una serie de importantes predicciones, que fueron probadas exitosamente para el espectro de núcleos complejos, bajo el supuesto de un Hamiltoniano aleatorio dado por una matriz simétrica  $\mathbf{R}$  con elementos aleatorios independientes  $R_{ij}$ .

Las predicciones de la TMA representan un promedio de todas las posibles interacciones. Las predicciones que difieren de las *predicciones universales* de la TMA permiten identificar sistemas específicos, propiedades no aleatorias del sistema bajo consideración y dan algunas pistas sobre las interacciones de los elementos en cuestión. Stanley, Amaral y otros (1999) [58] usan la metodología de TMA para analizar la correlación cruzada entre los precios de las acciones de New York Stock Exchange, American Stock Exchange y de Nasdaq. Ellos probaron las estadísticas de los valores propios de la matriz de correlación empírica contra la hipótesis nula de una matriz de correlación aleatoria, lo que les

permitió distinguir la correlación genuina de la correlación aparente presente en las matrices aleatorias, además mostraron que en su caso los valores propios más altos y sus correspondientes vectores propios representan la influencia de todo el mercado sobre las acciones y usando el resto de los valores propios, esto es, los que difieren de la matriz aleatoria, realizan una partición del conjunto de todas las acciones estudiadas en subconjuntos diferentes que corresponden de manera idéntica a los sectores económicos convencionalmente identificados. Potters, Cizeau y Laloux (2002) [32] muestran que los resultados de la TMA son de gran interés para entender la estructura estadística de las matrices de correlación empíricas. Usando series de tiempo de diferentes acciones de S&P500, encuentran que hay una similitud importante entre las predicciones teóricas (basadas en que la matriz de correlación es aleatoria) y los datos empíricos. En este mismo artículo Potters, Cizeau y Laloux muestran que este método puede ser implementado en la administración de riesgos. Maslov (2001) [45] utiliza la TMA para medir la globalización usando la matriz de correlación de los índices de mercado de 37 países, mostrando que hay fuertes interacciones entre las economías individuales. Burda, Gorlich, Jarosz y Jurkiewicz (2004) [9] determinan, usando la TMA, una relación exacta entre el espectro de los valores propios de la matriz de correlación y sus estimaciones.

### 1.3.1 Matrices Aleatorias

El conjunto más simple de matrices aleatorias es el conjunto de matrices simétricas  $\mathbf{R}$  de tamaño  $N \times N$ , cuyos elementos son variables aleatorias independientes e idénticamente distribuidas. Este conjunto de matrices es conocido como matrices aleatorias Wishart o el conjunto Laguerre de la TMA [44].

Un resultado interesante se obtiene cuando la matriz  $\mathbf{R}$  tiene un tamaño muy grande ( $N \rightarrow \infty$ ), en este caso la distribución de sus valores propios tiene propiedades universales, que son independientes de la distribución de los elementos de la matriz.

Sean  $\mathbf{R}$  una matriz simétrica cuadrada de tamaño  $N \times N$  y  $\lambda_\alpha$  con  $\alpha = 1, \dots, N$  los valores propios de  $\mathbf{R}$ .

La densidad de los valores propios se define como

$$\rho(\lambda) = \frac{1}{N} \sum_{\alpha=1}^N \delta(\lambda - \lambda_\alpha)$$

Donde  $\delta$  es la función de Dirac.

A menudo se encuentran matrices  $\mathbf{R}$ , que tienen la propiedad especial de ser positivas definidas, esto es, que todos sus valores propios son

positivos. En notación matricial  $\mathbf{R}$  se puede escribir como  $\mathbf{R} = \frac{1}{N} \mathbf{A} \mathbf{A}^T$ ,

donde  $\mathbf{A}^T$  es la matriz transpuesta de  $\mathbf{A}$ . En general  $\mathbf{A}$  es una matriz de tamaño  $N \times T$ ,  $\mathbf{A}^T$  es de  $T \times N$  y así  $\mathbf{R}$  es  $N \times N$ .

Para el caso  $N \neq T$  existe una formula [46] para la densidad  $\rho(\lambda)$  de los valores propios de  $\mathbf{C}$ . En el límite  $N, T \rightarrow \infty$ , con un cociente fijo

$Q = \frac{T}{N} \geq 1$ , se tiene:

$$\rho(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_- - \lambda)(\lambda - \lambda_+)}}{\lambda}$$

$$\lambda_{\pm}^{\pm} = \sigma^2 \left( 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}} \right) \quad (1.2)$$

con  $\lambda \in [\lambda_-, \lambda_+]$  y donde  $\frac{\sigma^2}{T}$  es la varianza de los elementos de  $\mathbf{A}$ , o de manera equivalente  $\sigma^2$  es el promedio de los valores propios de  $\mathbf{R}$ .

De la ecuación 1.2 se puede inferir que:

- El extremo inferior del intervalo,  $\lambda_-$ , es positivo (excepto para  $Q=1$  donde es cero); de aquí que no hay valores propios entre 0 y  $\lambda_-$ . Cerca de éste extremo, la densidad de los valores propios

muestra un máximo puntiagudo, excepto en el límite  $Q = 1$  donde diverge como  $\frac{1}{\sqrt{\lambda}}$ .

- o La densidad de los valores propios también desaparece alrededor del extremo superior  $\lambda_+$ .

Note que estos resultados son válidos cuando  $N \rightarrow \infty$ . Para  $N$  finito, las singularidades que se presentan en ambos extremos son suaves, los extremos se vuelven imprecisos, con una probabilidad muy pequeña de encontrar valores propios arriba de  $\lambda_+$  y debajo de  $\lambda_-$  quien tiende a 0 cuando  $N$  es muy grande.

La ecuación (1.2) da una medida estrictamente cuantitativa para decidir cuándo un valor propio particular de una matriz de correlación empírica  $\mathbf{C}$  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  $\lambda_+$  debe ser tratado como una señal.

### **1.3.2 Propiedades universales de las matrices aleatorias.**

La idea es comparar las propiedades de una matriz de correlación empírica  $\mathbf{C}$ , con las de una matriz de correlación aleatoria  $\mathbf{R}$ .

La matriz de correlación **C** se puede escribir como

$$C = \frac{1}{N} HH^T$$

donde H es una matriz de tamaño  $N \times T$ , como ya se ha mencionado. En el Capítulo II la matriz C será la matriz de correlación de N acciones diferentes cuyos precios de cierre diarios forman series de tiempo de longitud T.

La matriz **R** viene dada por la expresión:

$$R = \frac{1}{N} AA^T$$

Donde A es una matriz de tamaño  $N \times T$  que contiene N series de tiempo de longitud T cuyos elementos son variables aleatorias mutuamente no correlacionados. Por construcción **R** es una matriz de Wishart y por lo tanto una matriz aleatoria.

Para aceptar o rechazar la hipótesis de aleatoriedad de la matriz **C** no es suficiente comparar la distribución de los valores propios de **C** con  $\rho(\lambda)$ .

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  $\mathbf{C}$  debe involucrar una investigación de la correlación en los valores propios  $\lambda_\alpha$ .

Puesto que por definición  $\mathbf{C}$  es una matriz real simétrica, las pruebas estadísticas para los valores propios de  $\mathbf{C}$  que se deben usar, son precisamente las pruebas universales de correlación de los valores propios de matrices aleatorias reales simétricas.

Sea  $\mathbf{S}$  una matriz real, aleatoria y simétrica de tamaño  $M \times M$  cuyos elementos fuera de la diagonal  $\mathbf{S}_{ij}$  con  $i < j$  son independientes e idénticamente distribuidos. Se ha conjeturado, basado en evidencias analíticas y numéricas extensivas [58] que en el límite  $M \rightarrow \infty$ , sin considerar la distribución de los elementos  $\mathbf{S}_{ij}$ , esta clase de matrices presenta las propiedades universales (funciones de correlación de los valores propios) 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) [46].

Formalmente, GOE está definido es el espacio de las matrices reales simétricas que satisfacen dos condiciones:

- El conjunto es invariante bajo transformaciones ortogonales, esto es, si  $Z$  es una matriz GOE, la transformación  $Z \rightarrow Z^* \equiv W^T Z W$

donde  $W$  es cualquier matriz ortogonal real ( $WW^T = I$ ), deja la probabilidad conjunta  $P(Z)dZ$  de los elementos  $Z_{ij}$  sin cambios:

$$P(Z^*)dZ^* = P(Z)dZ.$$

- Los elementos  $\{Z_{ij} : i \leq j\}$  son estadísticamente independientes.

Por definición, las matrices de correlación aleatorias  $\mathbf{R}$  que interesan en este trabajo, no son estrictamente matrices GOE, pero pertenecen de alguna manera a un conjunto especial llamado “chiral” GOE [25]. Este conjunto se puede describir de la siguiente manera: se define una matriz  $B$

$$B = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$$

Los valores propios  $\gamma$  de  $B$  están dados por  $\det(\gamma^2 I - AA^T) = 0$  de manera similar, los valores propios  $\lambda$  de  $R$  están dados por  $\det(\lambda I - AA^T) = 0$ . Así que todos los valores propios diferentes de cero de  $B$ , vienen en pares, esto es, para cada valor propio  $\lambda$  de  $R$ ,  $\gamma = \pm\sqrt{\lambda}$  son los valores propios de  $B$ . Puesto que los valores propios vienen en pares, los espectros de los valores propios tanto de  $B$  como de  $R$  tienen propiedades especiales en una vecindad de cero que son diferentes a las propiedades estándares de GOE. Debido a que estas propiedades decaen rápidamente a medida

que se aleja de cero, la correlación de los valores propios de  $R$  en el grueso del espectro todavía es consistente con las de GOE. Así que el propósito en esta parte debe ser probar que el grueso del espectro de los valores propios de la matriz de correlación empírica  $C$  posee las propiedades universales conocidas de las matrices GOE.

Las tres propiedades universales que poseen las matrices GOE son:

- Distribución de las diferencias de los valores propios de acuerdo al vecino más cercano.
- Distribución de las diferencias de los valores propios de acuerdo al siguiente vecino más cercano.
- Correlaciones entre valores propios de rango largo.

Antes de analizar las propiedades mencionadas es importantes resaltar que estas se cumplen si las diferencias de valores propios adyacentes (ordenados de menor a mayor) se expresan en unidades del promedio de las diferencias de los valores propios. De manera más general, el promedio de las diferencias de los valores propios cambia de una parte del espectro de los valores propios al siguiente. Así, para asegurar que las diferencias de los valores propios tenga un valor promedio uniforme a través del espectro, es importante hallar una transformación de

“desplegado” que convierte los valores propios  $\lambda_i$  en nuevas variables llamadas “valores propios desplegados”  $\xi_i$ , [46]. El desplegado asegura que las distancias entre los valores propios estén expresadas en unidades del promedio del espaciamiento local, facilitando así la comparación con resultados teóricos.

La transformación se encuentra de la siguiente manera:

Primero se define la función de distribución acumulativa de los valores propios, que cuenta el número de valores propios en el intervalo  $\lambda_i \leq \lambda$ ,

$$F(\lambda) = N \int_{-\infty}^{\lambda} P(x) dx$$

Donde  $P(x)$  es la densidad de probabilidad de los valores propios y  $N$  el número total de valores propios. La función  $F(\lambda)$  se puede descomponer en el promedio y la parte que fluctúa,

$$F(\lambda) = F_{prom}(\lambda) + F_{fluc}(\lambda)$$

Puesto que  $P_{fluc} \equiv \frac{dF_{fluc}(\lambda)}{d\lambda} = 0$  en promedio,

$$\rho(\lambda_R) = \frac{dF_{prom}(\lambda)}{d\lambda}$$

Es la densidad de los valores propios promediada. La falta de dimensión, de los valores propios desplegados está dada por

$$\xi_i \equiv F_{prom}(\lambda_i) \quad (1.3)$$

Para encontrar  $F_{prom}(\lambda)$  se puede sustituir la función de distribución acumulada

$$F(\lambda) = N \int_{-\infty}^{\lambda} P(x) dx$$

Por la expresión dada por

$$F_{rm}(\lambda) = N \int_{-\infty}^{\lambda} P_{rm}(x) dx$$

Donde

$$P_{rm}(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} .$$

El ajuste se realiza tomando  $\lambda_+, \lambda_-$  y  $N$  como parámetros libres. La función así obtenida es una estimación para  $F_{prom}(\lambda)$ .

### 1.3.2.1 Distribución de las diferencias de los valores propios de acuerdo al vecino más cercano.

Para matrices GOE, la distribución de las diferencias de los valores propios de acuerdo al vecino mas cercano  $s \equiv \xi_{k+1} - \xi_k$  está dada por la ecuación:

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

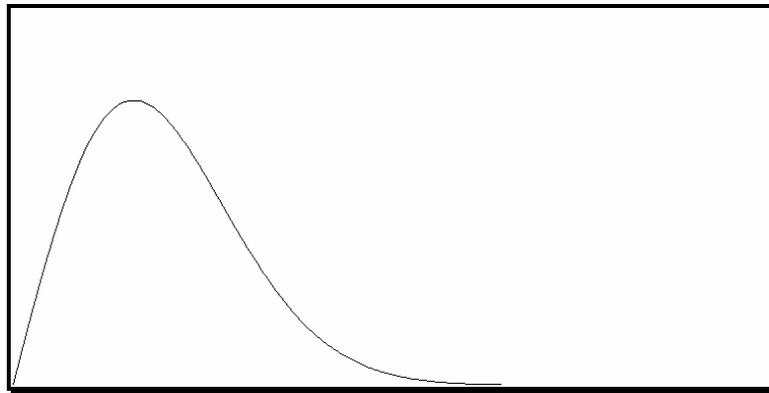


Figura 1.1 Distribución  $P_{GOE}(s)$

A menudo se refieren a esta distribución como “la conjetura de Wigner”. Vale la pena agregar, que esta distribución se encuentra también en las diferencias consecutivas de los ceros de la función zeta de Riemann. Esta

distribución permanece invariante después de aplicarle procedimientos de “desplegado” a los valores propios debido al decaimiento exponencial de  $P_{GOE}(s)$  para valores grandes de  $s$ .

Para probar la aleatoriedad de  $C$  se comparará la distribución de las diferencias de los valores propios “desplegados” de  $C$ ,  $\rho_m(s)$  con la distribución  $P_{GOE}(s)$ . Adicionalmente, para ver el grado de coincidencia de  $\rho_m(s)$  a los resultados de la TMA se puede utilizar la prueba no paramétrica de Kolmogorov-Smirnov [64] para ver que  $P_{GOE}(s)$  es una descripción correcta para  $\rho_m(s)$ .

#### **1.3.2.1.1 La prueba de Bondad de Ajuste de Kolmogorov-Smirnov.**

Suponga que  $Y_1, \dots, Y_n$  representan datos muestrales de una distribución continua, y que se desea probar la hipótesis nula  $H_0$  que afirma que  $F$  es la distribución poblacional, donde  $F$  es una función de distribución continua especificada. Una manera de probar  $H_0$ , una vez observadas  $Y_1, \dots, Y_n$  es la siguiente:

Sea  $F_e$  la distribución empírica definida mediante

$$F_e(x) = \frac{\#i: Y_i \leq x}{n}$$

Es decir,  $F_e(x)$  es la proporción de valores observados que son menores o iguales a  $x$ . Como  $F_e(x)$  es un estimador natural de la probabilidad de que una observación sea menor o igual a  $x$ , entonces si la hipótesis nula de que  $F$  es la distribución subyacente correcta, deberá parecerse a  $F(x)$ . Como esto vale para toda  $x$ , una cantidad natural en la cual basar una prueba  $H_0$  es la cantidad de prueba

$$D = \underset{x}{\text{Máximo}} |F_e(x) - F(x)|$$

Donde se toma el máximo sobre todo los valores de  $x$  desde  $-\infty$  hasta  $+\infty$ . A la cantidad  $D$  se le llama *estadístico de prueba Kolmogorov-Smirnov*.

Para calcular el valor de  $D$  de un conjunto de datos dado  $Y_j = y_j, j=1, \dots, n$ , se indicará con  $y_{(1)}, y_{(2)}, \dots, y_{(n)}$  los valores de las  $y_j$  en orden creciente. Es decir,  $y_{(j)} = j$  el valor mínimo de  $y_1, \dots, y_n$ .

Puede mostrarse que:

$$\begin{aligned} D &= \underset{x}{\text{Máximo}} |F_e(x) - F(x)| \\ &= \text{Máximo} \left\{ \text{Máximo} \{F_e(x) - F(x)\}, \text{Máximo} \{F(x) - F_e(x)\} \right\} \\ &= \text{Máximo} \left\{ \frac{j}{n} - F(y_{(j)}), F(y_{(j)}) - \frac{j-1}{n}, j = 1, \dots, n \right\} \end{aligned}$$

Es posible obtener una prueba de significancia  $\alpha$  considerando la cantidad  $D^*$  definida mediante

$$D^* = (\sqrt{n} + 0.12 + 0.11/\sqrt{n})D$$

Si se toma  $d_\alpha^*$  de manera que

$$P_F \{D^* \geq d_\alpha^*\} = \alpha$$

Entonces las siguientes son aproximaciones exactas de  $d_\alpha^*$  para diversos valores:

$$d_{.1}^* = 1.224, \quad d_{.05}^* = 1.358 \quad d_{.025}^* = 1.480, \quad d_{.01}^* = 1.626$$

La prueba de nivel  $\alpha$  rechazará la hipótesis nula de que  $F$  es la distribución del valor observado de  $D^*$  es por lo menos tan grande como  $d_\alpha^*$ .

### 1.3.2.2 Distribución de las diferencias de los valores propios de acuerdo al siguiente vecino más cercano.

Una segunda prueba, independiente de la anterior para GOE es la distribución  $\rho_{nm}(s')$  de las diferencias de los valores propios “desplegados” de acuerdo al siguiente vecino más cercano  $s' = \xi_{k+2} - \xi_k$ .

Para matrices GOE esta distribución viene dada en [46]

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

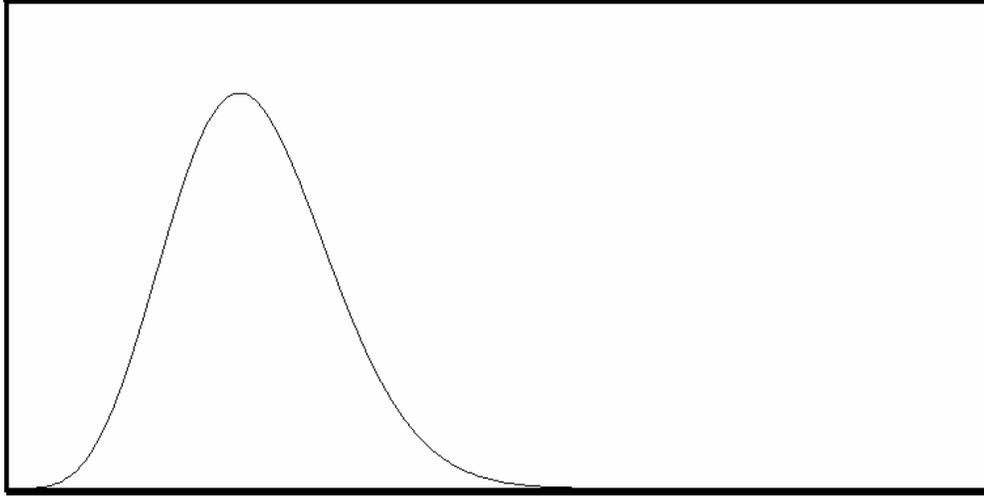


Figura 1.2 Distribución  $P_{GSE}(s)$

Las siglas GSE provienen del inglés Gaussian symplectic ensemble.

### 1.3.2.3 Correlaciones entre valores propios en un rango largo.

Para probar correlaciones entre pares de valores propios en rangos largos, se usará el estadístico  $\Sigma^2$  conocido como “varianza número”, que está definido como la varianza del número de valores propios desplegados en intervalos de longitud  $l$  alrededor de cada  $\xi_i$

$$\Sigma^2 \equiv \left\langle [n(\xi, l) - l]^2 \right\rangle_{\xi},$$

Donde  $n(\xi, l)$  es el número de valores propios desplegados en el intervalo  $[\xi - 1/2, \xi + 1/2]$  y  $\langle \dots \rangle_{\xi}$  es el promedio sobre todos los  $\xi$ . Si los valores

propios no están correlacionados,  $\Sigma^2 \approx l$ . En el extremo opuesto, cuando el espectro de los valores propios es rígido,  $\Sigma^2$  es una constante. De una manera más general, la varianza número  $\Sigma^2$  se puede expresar como

$$\Sigma^2(l) = l - 2 \int_0^l (l-x)Y(x)dx,$$

Donde  $Y(x)$ , conocida como la función de conglomerado de segundo nivel, está relacionada con la función de correlación de dos puntos [46].

Para el caso de GOE,  $Y(x)$  está dada explícitamente por

$$Y(x) \equiv s^2(x) + \frac{ds}{dx} \int_x^\infty s(t)dt,$$

Donde

$$s(x) = \frac{\text{sen}(\pi x)}{\pi x},$$

Para valores largos de  $l$ , la varianza número  $\Sigma^2$  para GOE tiene un comportamiento intermedio

$$\Sigma^2 = \ln l$$

### 1.3.3 Estadísticas de los vectores propios.

Después de analizar el comportamiento de los valores propios, las conclusiones más importantes se obtendrán del estudio de las componentes de los vectores propios.

Los componentes  $\{u_l^k; l=1, \dots, N\}$  del vector propio  $u^k$  de una matriz aleatoria de correlación R están distribuidos normalmente con media cero y varianza uno:

$$\rho(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

Para cuantificar el número de componentes que participan significativamente en cada vector propio, se usará el cociente inverso de participación (CIP), una noción aplicada a menudo en la teoría de localización [8].

El CIP de un vector propio normalizado  $u^k$  está definido como

$$I^k = \sum_{l=1}^N (u_l^k)^4$$

Donde N es el número de series de tiempo (empresas en nuestro caso) 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.

El significado de  $I^k$  se puede ilustrar observando los siguientes casos:

- 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  $u_1^k = 1$  ,  $I^k = 1$

De esta manera, el CIP es el recíproco del número de componentes del vector que contribuyen significativamente.

#### 1.4 La Matriz Potencia

Además de usar el espectro de los valores propios y la teoría de matrices aleatorias, se utilizará un método desarrollado por Guhr y Kälber [26] para estimar el ruido en la matriz de correlación. El método consiste en aplicar una función potencia a la matriz de correlación  $C$  convirtiéndola en una matriz  $C^p$  con la misma información de  $C$  pero quitándole eficientemente el ruido. Guhr y Kälber aplicaron este método a diferentes matrices de

correlación empíricas, permitiéndoles distinguir diferentes estructuras de correlación.

Considérese un mercado con  $K$  compañías,  $k = 1, 2, \dots, K$ , y  $B$  sectores industriales,  $b = 1, 2, \dots, B$ . Se asume que las compañías pertenecientes al mismo sector industrial están correlacionadas. Se ordenan las compañías de forma tal que los índices  $k$  de las compañías del mismo sector están seguidos. El índice  $b$  de los sectores económicos se puede ver como una función del índice  $k$  de la compañía, esto es  $b = b(k)$ . Para las  $k$  compañías que no pertenecen a ningún sector industrial, se tiene que  $b(k) = 0$ . Sea  $k_b$  el tamaño del sector industrial  $b$ . Obviamente se tiene que

$$\sum_{b=1}^B k_b + k = K$$

Se asumirá que  $k_b > 1$  y  $k \geq 0$ .

La serie de tiempo normalizada  $M_k(t)$ ,  $k = 1, 2, \dots, K$  de los rendimientos de las  $K$  compañías se pueden modelar como la suma de dos contribuciones puramente aleatorias: la primera modela la correlación entre los elementos de un sector dado siendo así común para ese sector ( $p_{b(k)}$ ), incluye además un número aleatorio  $\eta_b(t)$ , la segunda es específicamente para la compañía e incluye un número aleatorio  $\varepsilon_k(t)$ ,

$$M_k(t) = \sqrt{\frac{p_{b(k)}}{1+p_{b(k)}}} \eta_{b(k)}(t) + \frac{1}{\sqrt{1+p_{b(k)}}} \varepsilon_{k(t)}$$

Las contribuciones se pesan con el parámetro  $p_{b(k)}$ , común a todas las compañías del sector  $b$ , se toma  $p_{b(k)} \geq 0$ . Se asume que las variables  $\eta_b(t)$  y  $\varepsilon_k(t)$  no están correlacionadas y tienen distribución normal con media cero. Si  $k$  no pertenece a ningún sector, es decir para  $b=0$ ,  $p_{b(k)}=0$ . Aquí, se usa pasos discretos y normalizamos las unidades de tiempo tal que  $dt=1$ . Las series de tiempo  $M_k(t)$  son de longitud  $T$ ,  $t=1,2,\dots,T$ .

Si la serie de tiempo es infinitamente larga, esto es  $T \rightarrow \infty$ , se tiene que

$$[\mu_k]_{\infty} = 0 \quad y \quad [\sigma_k^2]_{\infty} = 1$$

Esto es porque  $\langle \eta_b(t) \rangle = 0$ ,  $\langle \varepsilon_k(t) \rangle_{\infty} = 0$  y  $\langle \eta_{b(k)}(t) \eta_{b(l)}(t) \rangle_{\infty} = \delta_{b(k)b(l)}$ ,

$\langle \eta_{b(k)}(t) \varepsilon_l(t) \rangle_{\infty} = 0$ ,  $\langle \varepsilon_k(t) \varepsilon_l(t) \rangle_{\infty} = \delta_{kl}$ .  $\langle \rangle$  es el promedio.

El coeficiente de correlación entre las compañías  $k$  y  $l$  es el promedio sobre el producto de las dos series normalizadas,

$$C_{kl}(T) = \frac{1}{T} \sum_{t=1}^T M_k(t)M_l(t) = \langle M_k(t)M_l(t) \rangle_T$$

Si se ve los números  $M_k(t)$  como las entradas de una matriz rectangular M de tamaño  $K \times T$ , se tiene

$$C(T) = \frac{1}{T} MM^t = \langle M(t)M^t(t) \rangle_T$$

la matriz de correlación C es de tamaño  $K \times K$ . Puesto que estos promedios dependen de la longitud T de la series de tiempo, se añade el argumento T a la matriz de correlación. Para series infinitamente largas,  $T \rightarrow \infty$  se tiene que

$$C_{kl}(\infty) = \lim_{T \rightarrow \infty} C_{kl}(T) = \frac{1}{1 + p_{b(k)}} (p_{b(k)} \delta_{b(k)b(l)} + \delta_{kl})$$

Así, la matriz  $C(\infty)$  consiste de B bloques cuadrados en la diagonal, de dimensiones  $k_b \times k_b$ . las entradas fuera de la diagonal son  $\frac{p_b}{1 + p_b}$  para el sector b, y una submatriz de  $k \times k$  unidades para las compañías que no están en ningún sector. Las entradas de la diagonal son todas iguales a

uno. Las demás entradas son cero: los coeficientes de correlación entre compañías que no están en ningún sector, las compañías que pertenecen a diferentes sectores y las que pertenecen a un sector con las de ningún sector.

Noh [52] mostró que la elección de

$$p_b = 1 - \frac{1}{k_b}$$

permite que la densidad espectral concuerde con los datos.

#### 1.4.1 Identificación del ruido

La correlación verdadera que se esconde detrás del ruido llega a ser visible en la densidad espectral si la serie es suficientemente larga. Guhr y Kälber [26] encontraron un procedimiento equivalente, en algún sentido, a la prolongación de la serie de tiempo, que les permite identificar y cuantificar el ruido de una matriz de correlación dada. Todo esto mediante una función potencia descrita a continuación.

La idea es transformar la matriz  $C(T)$  en la matriz  $C^{(q)}(T)$ , donde  $q$  es un número positivo y los elementos de  $C^{(q)}(T)$  se calculan de acuerdo a la siguiente definición

$$C_{kl}^{(q)}(T) = \text{sig}(C_{kl}(T)) |C_{kl}(T)|^q$$

De esta manera la función potencia preserva el signo de los elementos de la matriz  $C_{kl}(T)$  y eleva el módulo a la  $q$ -ésima potencia. La matriz  $C^{(q)}(T)$  es la matriz de la potencia de los elementos de  $C(T)$ , no es la matriz potencia  $[C(T)]^q$ .

Esto es importante porque el espectro de  $C(T)$  y el de  $[C(T)]^q$  se relacionan de una manera sencilla. Si  $\lambda_k$ ,  $k = 1, 2, 3, \dots, K$  son los valores propios de  $C(T)$ ,  $\lambda_k^q$ ,  $k = 1, 2, 3, \dots, K$  son los valores propios de  $[C(T)]^q$ . El espectro de  $C^{(q)}(T)$  es mucho más complicado y depende de los valores y vectores propios de  $C(T)$ . Sin embargo, dependiendo del valor escogido para  $q$ , se puede hacer que algunos elementos de  $C(T)$  ya no aparezcan en  $C^{(q)}(T)$ .

La función potencia transforma la densidad original desde  $\rho_{T_0}^{(1)}(\lambda) = \rho_{T_0}(\lambda)$  para  $q = 1$  hasta  $\rho_{T_0}^{(q)}(\lambda)$ , que para valores intermedios de  $q$  muestra dos picos claramente separados. Guhr y Kälber encontraron que la mejor separación se obtenía para  $q = 1.5$  y que para valores mayores de  $q = 2$ , los dos picos se unían de nuevo, perdiéndose la separación. Además, para  $1.25 \leq q \leq 2$  los dos picos tienen un significado importante: el pico de la izquierda corresponde a las correlaciones reales, mientras que el pico

de la derecha es debido al ruido alrededor de la autocorrelación trivial cuando  $k = l$ .

Para entender el efecto de la función potencia, se distinguen tres casos diferentes al considerar  $|C_{kl}(T)|^q$ . Primero, se aplica la potencia a los elementos de la diagonal  $C_{kk}(T)$ . Si T es suficientemente grande, una gran mayoría de los elementos de la matriz son positivos, de tal forma que para simplificar la discusión, se puede ignorar el signo del valor absoluto y se obtiene

$$(C_{kk}(T))^q = 1 + \frac{q}{1 + p_{b(k)}} \left( \sqrt{2} p_{b(k)} a_{b(k)b(k)} + \sqrt{2} a_{kk} + \sqrt{p_{b(k)}} (a_{b(k)k} + a_{kb(k)}) \right) \frac{1}{\sqrt{T}}$$

Segundo, se aplica la potencia a los elementos fuera de la diagonal  $C_{kl}(T)$  en los bloques de los sectores industriales donde  $k \neq l$  pero  $b(k) = b(l)$ . Por la misma razón del caso anterior se puede ignorar el signo del valor absoluto, y se tiene que

$$(C_{kl}(T))^q = \left( \frac{p_{b(k)}}{1 + p_{b(k)}} \right)^q + \left( \frac{q(p_{b(k)})^{q-1}}{(1 + p_{b(k)})^q} \right) \left( \sqrt{2} p_{b(k)} a_{b(k)b(l)} + \sqrt{2} a_{kl} + \sqrt{p_{b(k)}} (a_{b(k)l} + a_{kb(l)}) \right) \frac{1}{\sqrt{T}}$$

Tercero, se aplica la potencia a los elementos de  $C_{kl}(T)$  fuera de los bloques, donde  $k \neq l$  y  $b(k) \neq b(l)$ . Puesto que todos los  $\delta$  de Kronecker en la última ecuación son cero en este caso, se obtiene que

$$|C_{kl}(T)|^q = \frac{1}{\left[(1 + p_{b(k)})(1 + p_{b(l)})\right]^{q/2}} \left| \sqrt{p_{b(k)}} \sqrt{p_{b(l)}} a_{b(k)b(l)} + a_{kl} + \sqrt{p_{b(k)}} a_{b(k)l} + \sqrt{p_{b(l)}} a_{kb(l)} \right|^q \frac{1}{T^{q/2}}$$

en este tercer caso, los elementos de la matriz  $C_{kl}(T)$  pueden ser positivos o negativos con igual probabilidad.

En los primeros dos casos, la potencia  $(C_{kl}(T))^q$  contiene  $T$  términos independientes mas un término del orden  $1/\sqrt{T}$ . En el tercer caso, no hay un término independiente con  $T$ , y el orden de toda la expresión es  $1/T^{q/2}$ .

Así que para  $q > 1$ ,  $|C_{kl}(T)|^q$  se desvanece más rápido en el tercer caso que en los dos primeros. Como el tercer caso comprende todos los elementos fuera de los bloques, se encuentra que, para  $q > 1$  la matriz potencia  $C^{(q)}(T)$  tiene bloques en la diagonal de orden  $1/\sqrt{T}$ . Esto explica por qué la función potencia tiene un efecto comparable al de la prolongación de la serie de tiempo. A primera vista, uno podría esperar que el efecto sea mayor a medida que  $q$  es más grande. Sin embargo, no

es esto lo que ocurre, porque los  $T$  términos independientes son diferentes debido a los elementos que están en cada bloque de la matriz, en el primer caso son unos, pero en el segundo caso son números menores que uno que al elevarse a una potencia cada vez mayor, se van volviendo mas pequeños. De este modo los unos de la diagonal dominan cada vez más, haciendo que los valores propios tiendan a uno.

Para series infinitamente largas  $T \rightarrow \infty$ , la matriz potencia  $C_q(\infty)$  es de manera trivial una diagonal de bloques.

### **1.5 La matriz $C$ usando la Transformada de Fourier**

La transformada de Fourier se está convirtiendo en una herramienta cada vez más popular e importante en la economía financiera porque permite obtener el valor de un activo en tiempo real teniendo en cuenta una estructura realista para los rendimientos del mismo, considerando exceso de kurtosis, volatilidad estocástica y los efectos de apalancamiento. Además de una herramienta para predecir el precio de un activo, la transformada de Fourier también se utiliza para aproximar los datos perdidos de una serie de tiempo y para suavizarlas de tal modo que se reduzca el ruido.

El análisis de Fourier esta basado en la descomposición de una función arbitraria  $s(t)$  posiblemente periódica, en simples formas de ondas, esto

es, en una suma de funciones senos y cosenos con diferentes frecuencias.

La Figura 1.1 muestra la descomposición en funciones sinusoidales de las gráficas en negro.

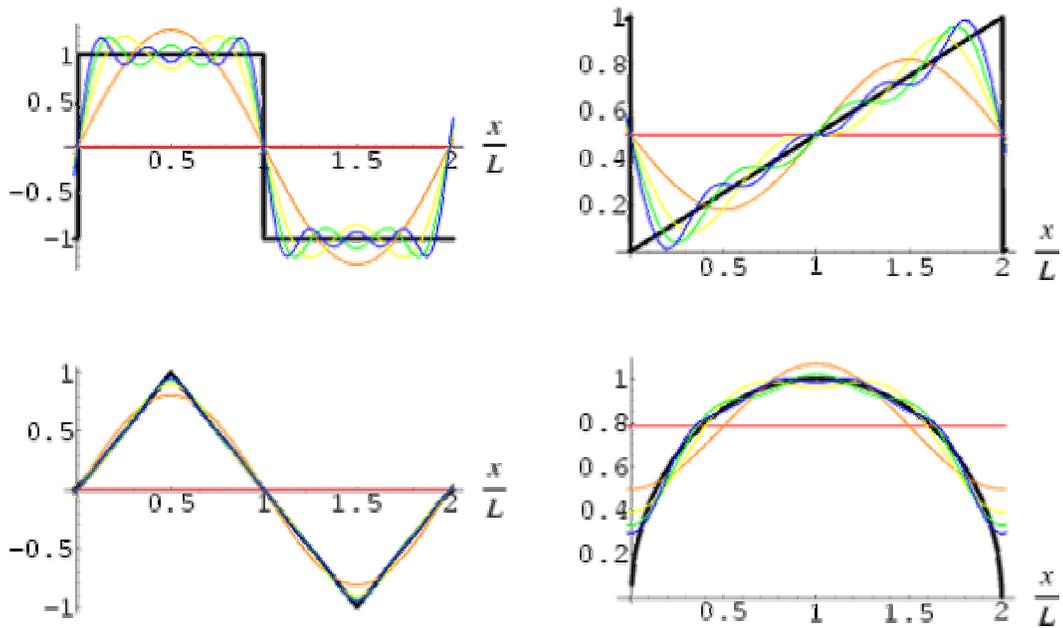


Figura 1.3 Aproximaciones con series de Fourier

La idea de la expansión de una función como una serie trigonométrica fue usada por el matemático francés Jean Baptiste Joseph Fourier alrededor de 1804, en ese entonces fue puesta en duda por matemáticos famosos

como Lagrange y Laplace, pero sirvió de base para trabajos posteriores en series trigonométricas y teoría de funciones de variable real.

Para vectores de datos, el análisis de Fourier se realiza usando la Transformada Discreta de Fourier (TDF). La Transformada Discreta de Fourier era conocida en teoría desde hace muchos años, pero solamente con la llegada de la computadora fue llevada a la práctica. Probablemente la primera persona que concibió la TDF fue Karl Friedrich Gauss, el famoso matemático alemán del siglo XIX, aunque por cierto él no tenía una computadora en la que pudiera implementarla.

W. Giles Mellon (1964) en [48] explica el uso de la serie de Fourier como un instrumento para la predicción financiera. Michio Hatanaka (1972) [69] muestra cómo recuperar la pérdida de grados de libertad causados por la elección de una ventana de datos que no contenga frecuencias altas.

Peijie Wang (2004) [69] utiliza la transformada de Fourier para identificar patrones típicos y analizar el comportamiento de datos económicos y financieros. Bogdan Negrea (2002) en [50] propone un modelo de valuación estocástico para acciones usando la transformada de Fourier. Ales Cerny (2004) en [15] explica la labor de la transformada de Fourier en el modelo binomial de valuación de opciones.

En este trabajo se usará la TDF en dos sentidos: uno para completar las series de tiempo cuya fecha de inicio difiere en menos de 60 días de la fecha escogida y la segunda, la más importante, para limpiar cada serie de tiempo de tal manera que la matriz  $C_f$  obtenida a partir de las nuevas series tenga menos ruido.

Estrictamente hablando, este método puede ser aplicado únicamente a series de tiempo estacionarias, esto es, cuyas características no cambian con el tiempo. Sin embargo, las investigaciones muestran que en aplicaciones prácticas el problema se puede manejar satisfactoriamente.

Sea  $f(t)$  ( $-\infty < t < \infty$ ) una serie de tiempo discreta, entonces su transformada de Fourier se define como

$$F(\omega) = \sum_{t=-\infty}^{\infty} f(t)e^{-j\omega t}$$

$F(\omega)$  se conoce también como la función de densidad espectral de  $f(t)$ , existe una transformada inversa de Fourier, que es continua, de modo que

$$f(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\omega)e^{i\omega t} d\omega$$

En la práctica, para una serie de tiempo discreta con N observaciones, la transformada de Fourier se convierte en la transformada discreta de Fourier (TDF) que también tiene su inversa, la fórmula de la TDF es:

$$F(k) = \frac{1}{N} \sum_{n=0}^{N-1} f(n) e^{-jn \frac{2k\pi}{N}}$$

y la de la inversa (ITDF):

$$f(n) = \sum_{k=0}^{N-1} F(k) e^{jk \frac{2n\pi}{N}}$$

con  $F(k) = F(k\Delta\omega) = F\left(\frac{2k\pi}{N}\right)$  pues  $\Delta\omega = \frac{2\pi}{N}$ .

Si  $f(n)$  es real, se puede escribir la última expresión en términos de funciones seno y coseno con coeficientes reales

$$f(n) = \frac{1}{N} \sum_{k=1}^N a(k) \cos\left(\frac{2\pi(k-1)(n-1)}{N}\right) + b(k) \sin\left(\frac{2\pi(k-1)(n-1)}{N}\right)$$

Donde

$$a(k) = \text{real}[F(k)], \quad b(k) = -\text{imag}[F(k)]$$

Empíricamente  $F(k)$  tiene que ser filtrada y estimada. La manera más sencilla de filtrar es hacer pasar a  $F(k)$  a través de una ventana rectangular o periodograma y truncar la señal en un punto  $M$  convenientemente, esto es

$$\overline{F(k)} = F(k) * W$$

donde  $W$  es un vector cuyas componentes son

$$W(i) = \begin{cases} 1 & \text{si } i \leq M \\ 0 & \text{si } i > M \end{cases}$$

La nueva ITDF es

$$f(n) = \sum_{k=0}^{N-1} \overline{F(k)} e^{jk \frac{2n\pi}{N}}$$

La cual reconstruye la señal sin frecuencias altas (sin ruido).

El error cometido al ajustar la serie de tiempo con polinomios trigonométricos esta distribuido normalmente con media cero.

## **1.6 El Árbol de expansión mínima.**

Recientemente se ha venido prestando mucha atención al estudio de las propiedades topológicas de las redes. En particular, se ha mostrado que muchos sistemas naturales y sociales presentan propiedades estadísticas inesperadas de relaciones que conectan diferentes elementos del sistema y que no pueden ser descritas con graficas aleatorias. Investigaciones recientes muestran propiedades de redes que describen sistemas físicos y sociales como el sistema de World Wide Web , Internet y redes sociales [59].

Un árbol de expansión mínima es una red con características especiales.

La noción de árbol proveniente de la teoría de grafos, tiene aplicaciones potenciales en el análisis de mercados financieros.

El concepto de árbol de expansión mínima, en el ambiente de mercados financieros, fue introducido por Mantenga, Bonano y otros (1989) [39] como un método para encontrar arreglos jerárquicos de acciones a través del estudio de conglomerados de compañías, usando las correlaciones de los rendimientos de las acciones.

En sus últimos artículos lo emplea para investigar no sólo las acciones de un portafolio, sino también índices financieros y volatilidad [42].

Con una métrica apropiada, basada en la matriz de correlación, se define un grafo totalmente conectado donde los nodos son compañías o acciones y las distancias entre ellas son obtenidas de sus correspondientes coeficientes de correlación.

El árbol de expansión mínima se genera de un grafo seleccionando las correlaciones más importantes y se usa para identificar conglomerados de compañías.

Onnela, Chakraborti, Kaski y Kertész (2002) [53] usan árboles dinámicos, esto es, con ventanas en el tiempo, para mostrar que los activos del portafolio óptimo de Markowitz están prácticamente todo el tiempo en las ramas externas del árbol. A finales del 2002 Mantegna, Bonano y otros [5] compararon los árboles de expansión mínima dinámicos obtenidos con los rendimientos de acciones con los obtenidos con las series de tiempo de la volatilidad, concluyendo que la estabilidad de los árboles de expansión mínima de la volatilidad es menor que los de los rendimientos de los activos. En Onnela, Chakraborti, Kaski y Kertész y otros (2003) [54] se usa 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. A principios del 2004 Mantegna, Bonano y otros [42] investigan portafolios, índices financieros y volatilidad por medio de redes. Los artículos anteriores muestran la manera tan acertada en que los árboles de expansión mínima reflejan la taxonomía económica de los activos.

**Un árbol de expansión** (o árbol recubridor) es una gráfica de N objetos (vértices o nodos) unidos por N-1 arcos que permiten ir de un vértice a cualquier otro vértice. Para N objetos existen  $N^{N-2}$  árboles recubridores diferentes.

La Figura 1.4 es un ejemplo de los 16 árboles de expansión de 4 vértices.

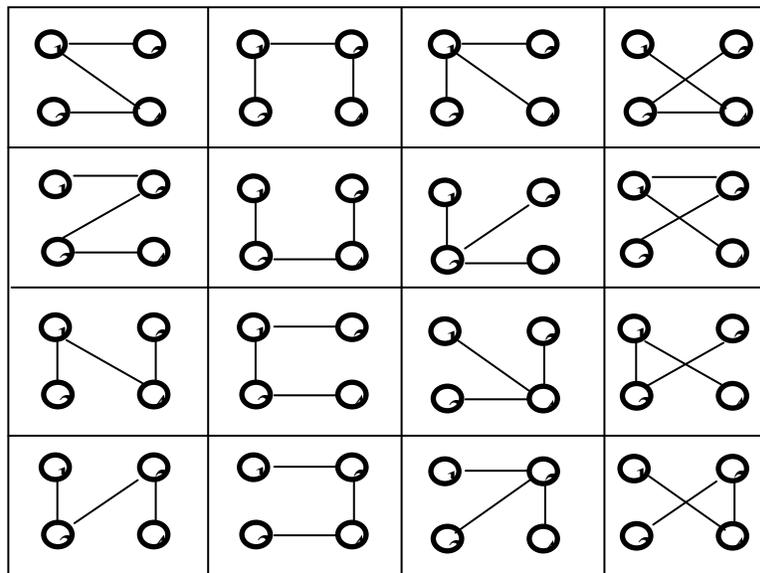


Figura 1.4 Árboles de expansión de 4 vértices

Si cada arco representa una distancia o costo, o en general si a cada arco se le asocia un peso (número real), 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.

En el caso de la tabla anterior, si los arcos representan las distancias entre los vértices de un cuadrado, se observa que los árboles de la segunda columna son de expansión mínima. Este ejemplo sugiere que la solución del problema de minimización de la distancia total no es única.

El problema del árbol de expansión mínima fue originalmente establecido por Boruvka en 1926 mientras consideraba la electrificación rural del sur de Moravia en Checoslovaquia.

Posiblemente el más famoso de los algoritmos para solucionar este problema se debe a Kruskal. Los pasos del algoritmo son:

Paso 1      Arbitrariamente seleccione un nodo.

- Paso 2 Conecte este nodo al nodo más cercano en términos del criterio que esté usando (tiempo, costo, distancia...) esto es, conéctelo al nodo que minimice el peso total.
- Paso 3 Considere todos los nodos que están conectados, encuentre y conecte el nodo más cercano que no esté conectado.
- Paso 4 Repita el paso tres hasta que todos los nodos estén conectados.
- Paso 5 Si en cualquiera de los pasos anteriores se encuentra en una encrucijada porque hay varios nodos igualmente cercanos a los conectados, seleccione uno arbitrariamente y continúe. Esto sugerirá que puede haber más de una solución óptima.

Para conectar los nodos es necesario definir un criterio o distancia entre ellos. En la siguiente sección se definirá una distancia a partir de las matrices de correlación  $C$ ,  $C^{(q)}$  y  $C_f$ . La distancia permitirá jerarquizar las acciones y exponerlas en árboles de expansión mínima.

### 1.6.1 Espacios ultramétricos

Un espacio ultramétrico es un espacio métrico donde la distancia  $\hat{d}_{ij}$  entre los objetos satisface las siguientes condiciones:

$$\text{i) } \hat{d}_{ij} = 0 \text{ si y sólo si } i = j$$

$$\text{ii) } \hat{d}_{ij} = \hat{d}_{ji}$$

$$\text{iii) } \hat{d}_{ij} \leq \max\{\hat{d}_{ik}, \hat{d}_{kj}\}$$

La tercera condición es más fuerte que la desigualdad triangular y es conocida como una desigualdad ultramétrica.

El concepto de ultrametricidad está íntimamente ligado al concepto de jerarquía, de ésta manera los espacios ultramétricos proporcionan un camino natural para describir jerárquicamente estructuras de sistemas complejos.

Partiendo de un espacio de N objetos con una distancia métrica asociada, se pueden obtener varios espacios ultramétricos haciendo particiones sobre el conjunto de los N objetos. Uno de éstos espacios con estructura ultramétrica es de particular importancia, debido a su simplicidad y a las

propiedades importantes que posee, éste espacio se denomina el “espacio ultramétrico subdominante”.

En un espacio métrico donde N objetos se pueden unir, el espacio ultramétrico subdominante se puede obtener determinando el árbol de expansión mínima que conecta los N objetos.

El siguiente método permite determinar la distancia entre las acciones  $i$  y  $j$  :

Sea  $S_i$  el logaritmo de los rendimientos de los activos como en la ecuación (1.1) y

$$\tilde{S}_i = \frac{S_i - \langle S_i \rangle}{\sqrt{S_i - \langle S_i \rangle}}$$

$\tilde{S}_i$  es simplemente  $S_i$  menos su media y dividido sobre su desviación estándar. Para cada activo  $i$  se forma el vector columna de rendimientos  $\vec{S}_i$  como en (1.2).

La distancia euclidea entre los vectores  $\vec{S}_i$  y  $\vec{S}_j$  se obtiene de la relación pitagórica:

$$d_{ij}^2 = \left\| \vec{S}_i - \vec{S}_j \right\|^2 = \sum_{k=1}^T \left( \tilde{S}_{ik} - \tilde{S}_{jk} \right)^2$$

Puesto que  $\tilde{S}_i$  por construcción es un vector unitario, se tiene que

$$\sum_{k=1}^T \tilde{S}_{ik}^2 = 1$$

y

$$d_{ij}^2 = \sum_{k=1}^T \left( \tilde{S}_{ik}^2 + \tilde{S}_{jk}^2 - 2\tilde{S}_{ik}\tilde{S}_{jk} \right) = 2 - 2 \sum_{k=1}^T \tilde{S}_{ik}\tilde{S}_{jk} = 2 - 2\rho_{ij}$$

Donde  $\rho_{ij}$  es el coeficiente de correlación entre el activo  $i$  y el activo  $j$ .

De aquí

$$d_{ij} = \sqrt{2(1 - \rho_{ij})}$$

Puesto que  $d_{ij}$  está definida a partir de la distancia euclidea entre dos vectores,  $d_{ij}$  satisface las tres propiedades de distancia.

Debido a que  $-1 \leq \rho_{ij} \leq 1$ , se tiene que  $0 \leq d_{ij} \leq 2$ . En particular:

$$\rho_{ij} = -1 \mapsto d_{ij} = 2$$

$$\rho_{ij} = 0 \mapsto d_{ij} = \sqrt{2}$$

$$\rho_{ij} = \frac{1}{2} \mapsto d_{ij} = 1$$

$$\rho_{ij} = 1 \mapsto d_{ij} = 0$$

Puesto que  $\rho_{ij} = 1$  si y sólo si  $i = j$ , y que  $\rho_{ij} = \rho_{ji}$  se puede verificar fácilmente que

i)  $d_{ij} = 0$  si y sólo si  $i = j$

ii)  $d_{ij} = d_{ji}$

En términos de acciones estas propiedades dicen que dos acciones diferentes tienen una distancia diferente de cero entre ellas, y que la distancia de la acción  $i$  a la  $j$  es la misma que de la acción  $j$  a la  $i$ .

Usando la distancia  $d_{ij}$  y el algoritmo de Kruskal para obtener el árbol de expansión mínima, queda definida una nueva distancia  $\hat{d}_{ij}$  entre las acciones, que satisface i) y ii), pero lo más importante es que satisface

$$\hat{d}_{ij} \leq \max\{\hat{d}_{ik}, \hat{d}_{kj}\}$$

Corroborando el hecho de que el árbol de expansión mínima así obtenido, es un espacio ultramétrico subdominante.

El siguiente ejemplo muestra la construcción de un árbol de expansión mínima de 5 nodos a partir de una matriz distancia  $(d_{ij})$  y la matriz distancia  $(\hat{d}_{ij})$  obtenida del árbol.

Se considera la siguiente matriz de distancia  $(d_{ij})$

$d_{ij}$	A	B	C	D	E
A	0	0.2	0.5	0.8	1
B	0.2	0	1.2	1.5	1.7
C	0.5	1.2	0	1.8	2
D	0.8	1.5	1.8	0	0.3
E	1	1.7	2	0.3	0

Se escoge arbitrariamente un nodo, por ejemplo el nodo C y se conecta al nodo más cercano que resulta ser el nodo A, que está a 0.5 de C, como muestra la Figura 1.5.

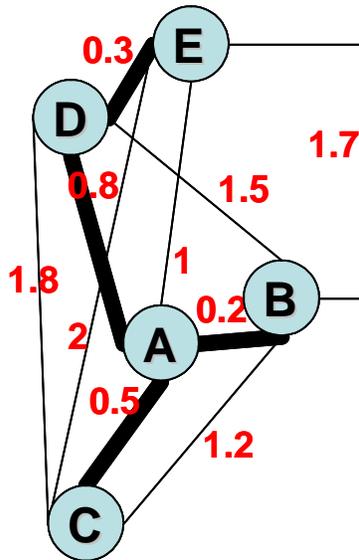


Figura 1.5 Árbol de expansión mínima

Considerando los nodos C y A se encuentra y conecta el nodo más cercano que no está conectado, resultando ser el B, a una distancia de 0.2 de A.

Teniendo en cuenta los nodos C, A y B, el siguiente nodo con distancia más corta es el D, a 0.8 de A. Por último conectamos el nodo E, a 0.2 de D.

El camino marcado resulta ser un árbol de expansión mínima y se considerará el único camino que conecta los cinco nodos. Así que de ahora en adelante la distancia  $\hat{d}_{ij}$  entre dos nodos será la distancia que se recorre en el árbol para llegar de un nodo a otro.

Por ejemplo:

$$\begin{aligned}\hat{d}_{CE} &= d_{CA} + d_{AD} + d_{DE} \\ &= 0.5 + 0.8 + 0.3 = 1.6\end{aligned}$$

Las distancias en el árbol forman la siguiente matriz de distancias  $\hat{d}_{ij}$

$\hat{d}_{ij}$	A	B	C	D	E
A	0	0.2	0.5	0.8	1.1
B	0.2	0	0.7	1	1.3
C	0.5	0.7	0	1.3	1.6
D	0.8	1	1.3	0	0.3
E	1.1	1.3	1.6	0.3	0

Que satisface (como se puede verificar en la tabla)

$$\hat{d}_{ij} \leq \max\{\hat{d}_{ik}, \hat{d}_{kj}\}$$

el cumplimiento de la desigualdad es consecuencia directa del algoritmo de Kruskal, en el árbol significa simplemente que la distancia entre dos nodos es un pedazo de la autopista que une todos los nodos.

### 1.6.2 El Promedio de ocupación y el vértice central.

Es importante caracterizar la forma en que se extienden los nodos en el árbol. Para ello, se define la cantidad “promedio de ocupación”  $l(v_c)$  así:

$$l(v_c) = \frac{1}{N} \sum_{i=1}^N niv(v_i)$$

donde  $niv(v_i)$  es el nivel del vértice  $v_i$ . Los niveles, (no confundirlos con la distancia  $d_{ij}$  entre los nodos) se miden con relación a un vértice especial, llamado el vértice central  $v_c$ , cuyo nivel se toma como cero. De esta forma, el promedio de ocupación indica el centro de masa 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.

El vértice central  $v_c$  es considerado como el padre de todos los vértices del árbol o también como la raíz del mismo. Este se usa como punto de referencia en el árbol, contra el cual la posición de los demás vértices es relativa.

Hay un poco de arbitrariedad en la elección del vértice central, sin embargo los siguientes criterios pueden ayudar a escoger al mejor candidato:

- Es el vértice que tiene más nodos conectados, esto es, el vértice con mayor número de vecinos. El número de vecinos se conoce como el grado del vértice.
- Es aquel cuya suma de los coeficientes de correlación de los vértices vecinos es máxima. Este criterio se conoce como el peso del vértice.
- Es el vértice que produce el valor más bajo del promedio de ocupación, esto es, el centro de masa.

Intuitivamente hablando, es muy probable que los tres criterios coincidan. Un vértice con un grado de vértice alto, el vértice central en particular, carga mucho peso alrededor de él (los nodos vecinos), quienes a su vez pueden estar altamente conectados con otros y así sucesivamente.

### **1.6.3 Análisis de portafolios**

Ahora, se aplicará los conceptos y las medidas discutidas al problema de optimización de un portafolio. Esto es una herramienta cuantitativa importante, que permite visualizar el problema y aporta argumentos significativos al proceso de decisión.

Sea  $P$  un portafolio de Markowitz con pesos de los activos  $w_1, w_2, \dots, w_N$ . En el esquema clásico de optimización del portafolio de Markowitz, los activos financieros se caracterizan por su riesgo y rendimiento promedio, donde el riesgo asociado con un activo se mide con la desviación estándar. Usualmente se realiza la optimización de Markowitz usando datos históricos. La idea es optimizar los pesos de los activos de tal forma que el riesgo del portafolio sea minimizado para un rendimiento del portafolio  $r_P$ . Sin embargo, en el marco de árboles de activos, la tarea es determinar cómo el activo está localizado con respecto al vértice central. Sean  $r_m$  y  $r_M$  los rendimientos mínimo y máximo de un portafolio, respectivamente. El rendimiento esperado varía entre estos dos extremos, y se puede expresar como

$$r_{P,\theta} = (1-\theta)r_m + \theta r_M$$

$$0 \leq \theta \leq 1$$

Se define  $l_p$ , el “promedio ponderado del portafolio”, de la siguiente manera:

$$l_p(\theta) = \sum_{i \in P} w_i \text{niv}(v_i)$$

Donde  $\sum_{i=1}^N w_i = 1$  y más aún, como punto de partida, se puede considerar

$w_i \geq 0$  para todo  $i$ , lo que es equivalente a asumir que no hay ventas en corto. El propósito de esta restricción es evitar que  $l_p$  tome valores negativos, lo cual no tendría una explicación significativa en nuestro marco de árboles con vértice central. Sin embargo, es importante recalcar, que en la práctica, la medida del promedio ponderado del portafolio nunca asume valores.

Puesto que los activos que minimizan el riesgo de un portafolio se encuentran en las ramas exteriores del árbol, se espera que árboles largos (con  $l$  grande) tengan mayor potencial de diversificación, esto es, la oportunidad del mercado financiero para eliminar un riesgo específico del portafolio de riesgo mínimo ( $\theta = 0$ ).

A medida que se incrementa  $\theta$  hasta llegar a la unidad, el riesgo del portafolio en función del tiempo empieza prontamente a comportarse muy diferente del promedio de ocupación  $l$ . Consecuentemente, ya no es útil para describir la diversificación potencial del mercado. Sin embargo, emerge otro resultado interesante: el promedio ponderado del portafolio  $l_p(\theta)$  decrece cuando  $\theta$  aumenta su valor. Esto significa que de todos los posibles portafolios de Markovitz, 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.

## CAPITULO 2

### 2.1 Resumen

El estudio de las matrices de correlación tiene una larga historia en finanzas y es una de las piedras angulares de la teoría de optimización de portafolios de Markovitz. Sin embargo, determinar de una manera confiable una matriz de correlación empírica es un trabajo difícil.

La primera parte del capítulo presenta la forma como han sido seleccionadas las empresas para el estudio y la longitud de las series de tiempo.

Debido a la finitud de las series y a los datos perdidos, la estructura de la matriz está dominada por el ruido.

El objetivo de este capítulo es encontrar la información real contenida en la matriz  $C$ , para ello, en la segunda parte del capítulo, se muestra cómo limpiar la matriz de correlación  $C$  usando tres formas diferentes:

- *La teoría de matrices aleatorias.* Aquí se prueba que la matriz  $C$  satisface las propiedades universales de las matrices aleatorias y que por lo tanto su espectro se puede separar en: los valores propios con información real, los que no tienen información y los que tienen información viciada. Se presenta, pues, un análisis espectral, que permite analizar el comportamiento de las principales

acciones que cotizan en la BMV. Más aún, se usará el cociente de participación inversa para analizar los vectores propios.

- *La matriz potencia.* Se expondrá la potencia “ $p$ ” a la cual se debe elevar los componentes de la matriz  $C$ , para que el espectro se separe en dos: la parte de la matriz con ruido y la parte con información real. La potencia suprime de la matriz de correlación aquellos elementos que se asocian con el ruido, simulando un efecto de prolongación de las series de tiempo, esto es, los elementos fuera de los rectángulos de la diagonal de la matriz de correlación  $C$ , se hacen ceros.
- *La transformada de Fourier.* Se encontrará para cada serie una buena aproximación con la serie de Fourier, truncando la señal en un punto conveniente para eliminar el ruido. La nueva matriz construida a partir de las series de Fourier es sometida a la Teoría de Matrices Aleatorias para comprobar si efectivamente el nuevo espectro contiene información adicional.

Las tres matrices  $C$ ,  $C^p$  y  $C_f$  serán consideradas en el siguiente capítulo, en la construcción de los árboles de expansión mínima.

## **2.2 Las series de tiempo**

La Bolsa Mexicana de Valores (BMV) tiene aproximadamente 158 empresas emisoras, si se es riguroso en la elección de las series para nuestro estudio, exigiendo bursatilidad (alta y media), valor de capitalización y condiciones de mantenimiento, prácticamente quedarían las 33 empresas que en la actualidad conforman el Índice de Precios y Cotizaciones (IPC). Una de las ventajas de los métodos que se utilizarán para encontrar la correlación real entre las series de tiempo, es que permiten ser un poco más flexibles en la elección de las empresas que se incluyen en el estudio.

Después de analizar minuciosamente la longitud de las series, el número e importancia (bursatilidad, capitalización y mantenimiento) de las mismas, se escogió el periodo comprendido entre el 19/09/97 y el 06/02/04 y un total de 65 empresas, 32 de las cuales han sido tomadas en cuenta en los últimos dos años como parte de la muestra para construir el Índice de Precios y Cotizaciones (IPC).

Después de quitar los días donde la bolsa cerró y algunos donde más del 70% de las empresas no cotizaron, las series quedaron con una longitud T de 1598.

Hay que notar que en la BMV no todas las acciones son comercializadas diariamente, algunas de las acciones que no se comercializan a menudo ocasionalmente pueden realizar transacciones en volúmenes grandes. Esta realidad exagera el problema de estimar la correlación de una manera confiable.

Algunas de las empresas consideradas cotizan semanalmente o cada 30 días, por ello el número de datos originales es significativamente más bajo que el de las demás, es posible que la inclusión de las mismas aumente la autocorrelación espuria, pero el problema será detectado y eliminado en gran medida con los métodos que se usarán.

Después de comparar los resultados obtenidos utilizando diferentes métodos para completar los valores perdidos (los métodos propuestos por SPSS y Matlab), se ha decidido usar la interpolación lineal. Se utilizan para la interpolación el último valor válido antes del valor perdido y el primer valor válido después del valor perdido.

Los datos fueron extraídos de ECONOMATICA y REUTERS.

La Tabla 2.1 muestra las empresas seleccionadas, el sector al que pertenecen, el código de bolsa (que en adelante se usarán para nombrar las acciones), el porcentaje de participación en el IPC y el número de datos originales de la serie.

	<b>Nombre</b>	<b>Sector</b>	<b>Codigo de Bolsa</b>	<b>% IPC</b>	<b># de Datos</b>
1	Savia A	Agro & Pesca	SAVIAA	0.06	1587
2	Bachoco Industrias UBL	Agro & Pesca	BACHOCOUBL		1181
3	Bimbo Gpo A	Alimentos y Beb	BIMBOA	2.22	1594
4	Continental Grupo	Alimentos y Beb	CONTAL	1.32	1549
5	Fomento Econ Mex UBD	Alimentos y Beb	FEMSAUBD	2.47	1598
6	Maseca GI B	Alimentos y Beb	MASECAB		1428
7	Herdez S.A.	Alimentos y Beb	HERDEZ		1003
8	KOF CocaCola L	Alimentos y Beb	KOFL		1275
9	GMaseca (Gruma) B	Alimentos y Beb	GRUMAB		1389
10	GModelo C	Alimentos y Beb	GMODELOC	1.55	1598
11	Valle Jugos del B	Alimentos y Beb	VALLEB		952
12	Comercial Mexicana UBC	Comercio	COMERCIUBC	0.35	1598
13	Wal Mart de Mexico C	Comercio	WALMEXC	1.43	1598
14	Wal Mart de Mexico V	Comercio	WALMEXV	11.28	1598
15	Soriana Organizacio B	Comercio	SORIANAB	1.43	1598
16	Gigante Gpo	Comercio	GIGANTE		982
17	Saba Casa Grupo	Comercio	SAP		809
18	Elektra Gpo	Comercio	ELEKTRA	1.05	1596
19	GCorvi UBL	Comercio	GCORVIUBL		1058
20	Nadro S.A. B	Comercio	NADROB		995
21	Dataflux B	Comercio	DATAFLXB		1441

22	Fragua Corporativo B	Comercio	FRAGUAB		737
23	Liverpool Puerto de 1	Comercio	LIVEPOL1		677
24	ICA Soc Controlad	Construcción	ICA	0.49	1596
25	Geo Corporacion B	Construcción	GEOB	0.57	1589
26	Ara Consorcio	Construcción	ARA	0.9	1574
27	Hogar Consorcio B	Construcción	HOGARB		1310
28	GFBBVA Bancomer B	Finanzas y Seguros	GFBBB	4.18	1598
29	GFBanorte O	Finanzas y Seguros	GFNORTEO	1.66	1594
30	GFInbursa O	Finanzas y Seguros	GFINBURO	3.31	1571
31	Desc Soc Fom Ind B	Maquinaria Indust	DESCB	0.16	1587
32	Desc Soc Fom Ind C	Maquinaria Indust	DESCC		1158
33	San Luis Corp CPO	Maquinaria Indust	SANLUISCPO		1088
34	Apasco S.A.	Minerales no Met	APASCO	2.21	1595
35	Vitro A	Minerales no Met	VITROA	0.32	1573
36	GInd Saltillo	Minerales no Met	GISSA	0.46	1517
37	Cementos Chihuahua	Minerales no Met	GCC		1357
38	Cemex S.A. CPO	Minerales no Met	CEMEXCPO	9.12	1596
39	Penoles Industrias	Minería	PE&OLES	1.59	1544
40	GMexico B	Minería	GMEXICOB	2.35	1584
41	GCarso A1	Otros	GCARSOA1	3.04	1598
42	Televisa Gpo CPO	Otros	TLEVISACPO	4.56	1595
43	Accel S.A. B	Otros	ACCELSAB		942
44	Posadas Gpo L	Otros	POSADASL		1154

45	Radio Centro A	Otros	RCENTROA		646
46	Posadas Gpo A	Otros	POSADASA		1228
47	Corp Interam de Ent B	Otros	CIEB	0.61	1596
48	Corp Mex Restaurant B	Otros	CMRB		784
49	TV Azteca CPO	Otros	TVAZTCACPO	0.99	1597
50	Cid Mega Resort	Otros	CIDMEGA		732
51	Kimberly Clark Mex A	Papel y Celulosa	KIMBERA	1.52	1598
52	Cydsa S.A. A	Química	CYDSASAA		853
53	Tekchem S.A. A	Química	TEKCHEMA		802
54	Alfa S.A. A	Siderur & Metalur	ALFAA	2.05	1598
55	Industrias CH B	Siderur & Metalur	ICHB		1023
56	Simec Grupo B	Siderur & Metalur	SIMECB		1007
57	Hylsamex B	Siderur & Metalur	HYLSAMXB		1487
58	Imsa Gpo UBC	Siderur & Metalur	IMSAUBC		1439
59	Telefs de Mex A	Telecomunicación	TELMEXA		1466
60	Telefs de Mex L	Telecomunicación	TELMEXL	12.44	1598
61	Iusacell Gpo	Telecomunicación	CEL	0.12	1096
62	Telecom Carso Globa A1	Telecomunicación	TELECOMA1	5.02	1598
63	Movil Acces B	Telecomunicación	MOVILAB		984
64	Hilasal Mexicana A	Textil	HILASALA		1208
65	Cintra S.A. A	Transporte Servic	CINTRAA		760

Tabla 2.1 Datos de las empresas incluidas en el estudio

Dentro de las 65 empresas seleccionadas para el estudio se encuentran representados todos los sectores económicos, las empresas elegidas tienen la mayor bursatilidad de cada sector y juntas representan más del 85% de participación en el IPC y el 100% del índice México (INMEX). Todas las acciones incluidas han permanecido activas en el periodo seleccionado para el estudio.

La Tabla 2.2 muestra el número de empresas que cotizan en la BMV por sector económico y las que han sido incluidas en el estudio.

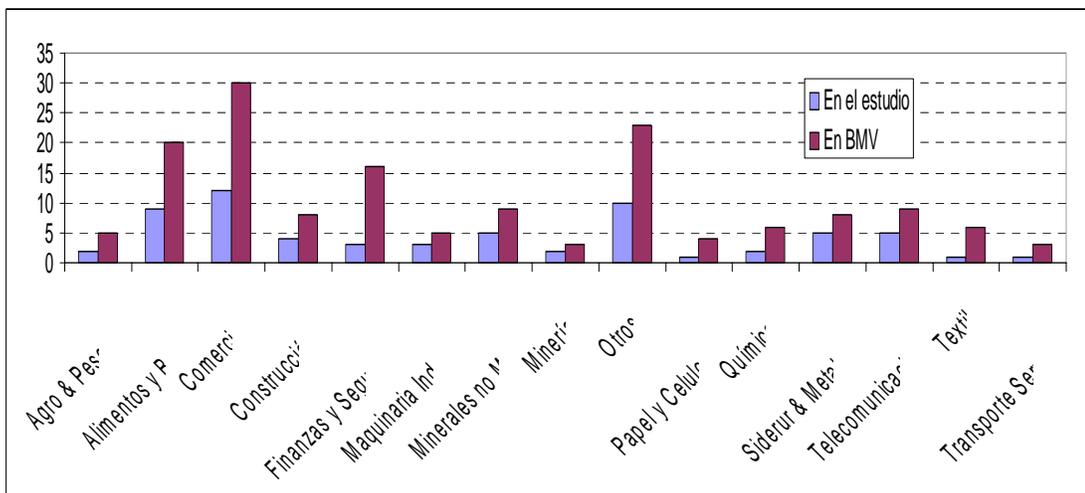


Tabla 2.2 Empresas por sector económico

### 2.3 Construcción de la matriz de correlación empírica $C_{ij}$

En este trabajo se tiene, pues,  $N=65$  acciones con precio  $P_i(t)$  para la acción  $i$  en el tiempo  $t$ , con  $t=0,1,\dots,1598$ . Después de considerar el logaritmo de los rendimientos de los activos  $S_i(t) = \ln P_i(t) - \ln P_i(t-1)$  y estandarizar cada una de las series, se calcula la matriz de correlación estandarizada  $C_{ij}$  que tiene un tamaño de  $65 \times 65$ .

El valor  $Q = \frac{T}{N} = \frac{1597}{65} = 24.57 > 1$  garantiza que la matriz de correlación

es definida positiva y permitirá la aplicación de la TMA. Los valores de  $\lambda_{\pm}$

se calculan con la fórmula  $\lambda_{\pm} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}$  de tal modo que

$\lambda_+ = 1.4441$
$\lambda_- = 0.63720$

La Figura 2.1 muestra el histograma de la distribución de los valores propios de  $C$ .

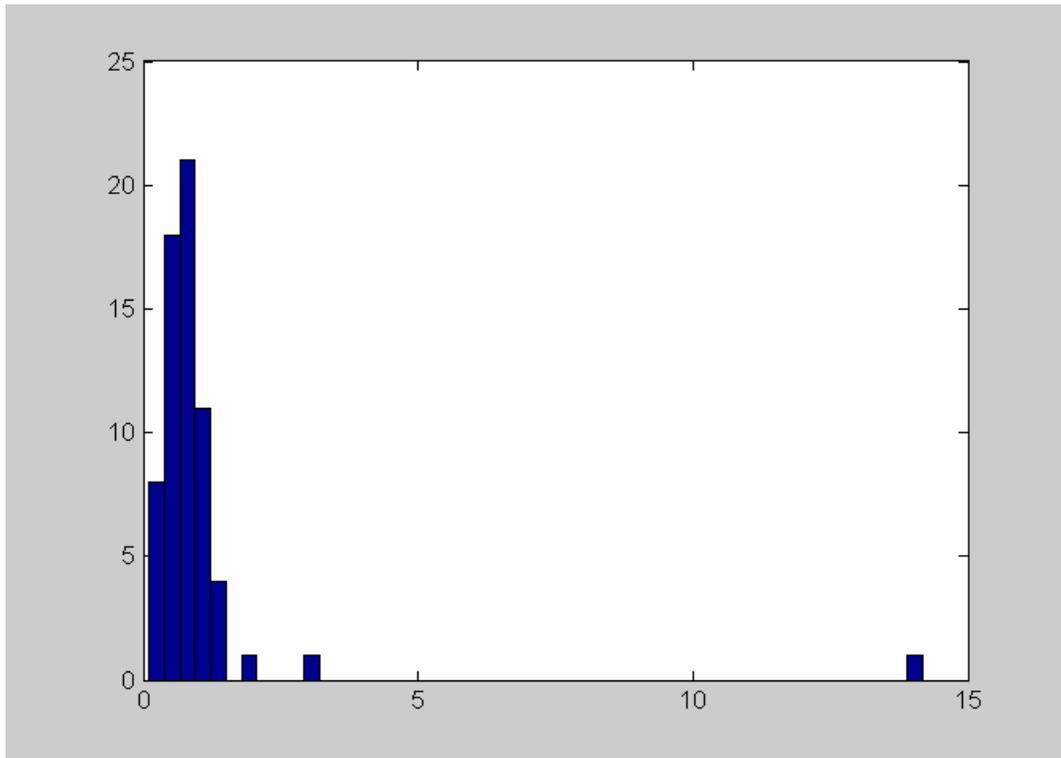


Figura 2.1 Distribución de los valores propios de C

En el histograma se puede observar que tres valores propios se alejan visiblemente del grueso del espectro. Estos tres valores propios son mayores que  $\lambda_+$ . La teoría asegura que si la matriz satisface las propiedades universales de la TMA, se podrá obtener información sustancial de estos tres valores propios.

Hay también una cantidad importante de valores propios por debajo de  $\lambda_-$ , 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.

La Figura 2.2 muestra, en azul, la distribución de los valores propios de  $C_{ij}$  y en rojo la distribución dada por

$$P_{rm}(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda}$$

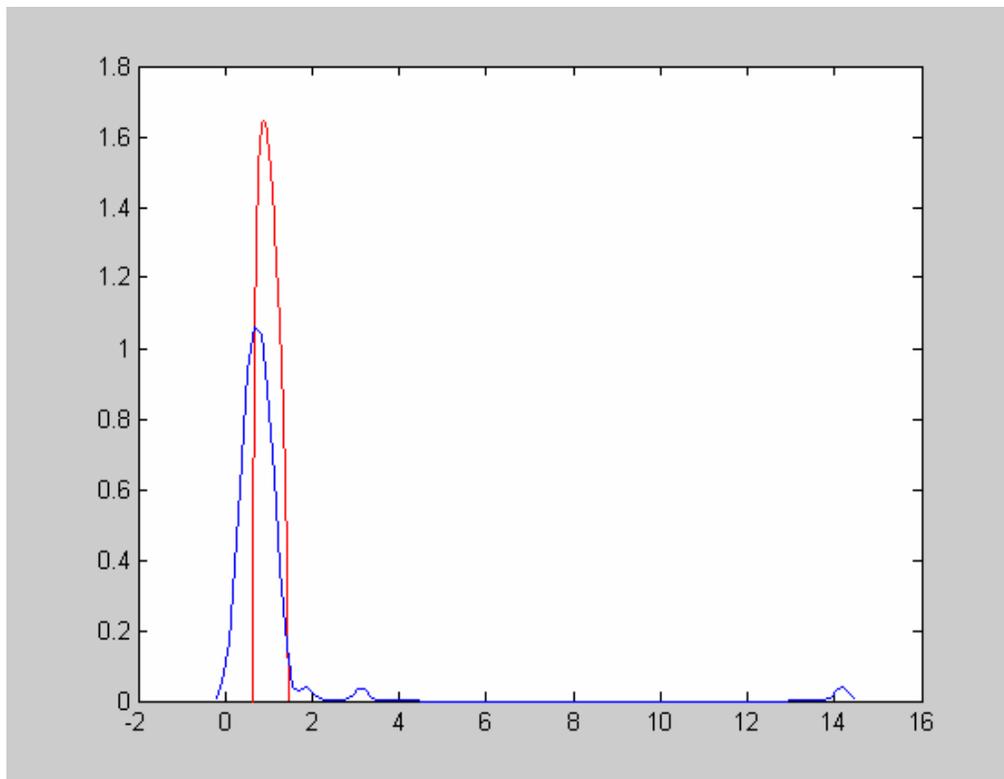


Figura 2.2 Distribución de los valores propios de  $C$  y  $P_{rm}(\lambda)$

Las diferencias entre las dos curvas de la Fig. 2.2 se deben precisamente a la existencia de valores propios por encima de  $\lambda_+$  y por debajo de  $\lambda_-$ .

Si la matriz C fuera completamente aleatoria, las dos curvas coincidirían y no se tendría información relevante.

En la intersección de las dos curvas se encuentran los valores propios del intervalo  $[\lambda_-, \lambda_+]$ . Los valores propios por encima de  $\lambda_+$ , que contienen información real, están a la derecha de la curva roja, en azul. Y se puede ver que hay una cantidad importante de valores propios por debajo de  $\lambda_-$  en el intervalo  $(0, \lambda_-]$ .

24 valores propios están por debajo de  $\lambda_-$ , 38 en el intervalo  $(\lambda_-, \lambda_+)$  y 3 son mayores que  $\lambda_+$  (los tres rectángulos que se ven separados en la Fig. 2.1).

El valor propio más grande 14.166 es más de 9 veces mayor que  $\lambda_+$ , se puede entonces atribuirle información real. El segundo mas grande valor propio 3.13 es mas del doble de  $\lambda_+$ , también proporcionará información importante, el siguiente valor propio 1.8615 es apenas un poco mayor que  $\lambda_+$  sin embargo es posible hallar en el información importante.

### 2.3.1 Análisis de la Matriz $C_{ij}$ usando la Teoría de Matrices Aleatorias

Es indispensable recordar que la información que se puede obtener del espectro sólo será válida si se prueba que el espectro satisface las propiedades universales de la TMA:

1. Distribución de las diferencias de los valores propios de acuerdo al vecino más cercano.

Para asegurar que las diferencias de los valores propios tenga un valor promedio uniforme a través del espectro, se halla una transformación llamada “desplegado” que convierte los valores propios  $\lambda_i$  en nuevas variables llamadas “valores propios desplegados”  $\xi_i$

La transformación se encuentra mediante la integral

$$\xi_i = N \int_{\lambda_-}^{\lambda_i} \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} d\lambda$$

Tomando N de acuerdo al número de valores propios dentro del intervalo  $(\lambda_-, \lambda_+)$  y ajustando  $\sigma$  de tal manera que se consiga la misma distribución.

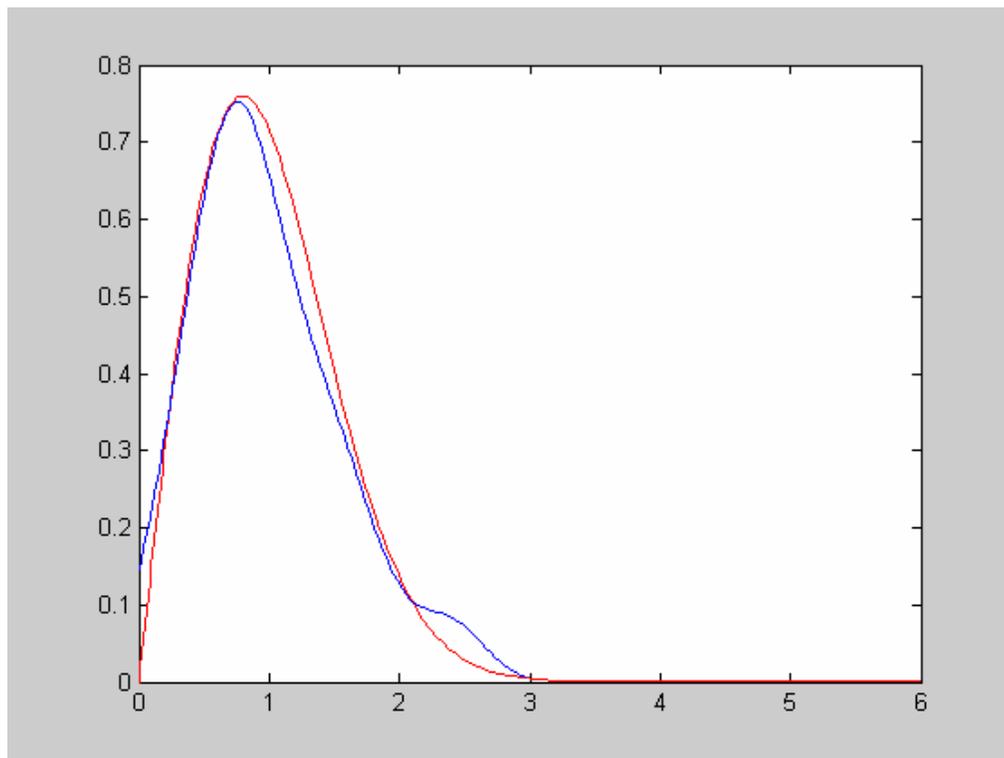
Las integrales se evaluaron usando el método de cuadratura de Simpson en Matlab.

Después de ordenar los  $\xi_i$  de menor a mayor, se considera

$$s_i = \xi_i - \xi_{i-1}$$

La Figura 2.3 muestra la distribución de  $s_i$  en azul y en rojo la de las matrices GOE dada por la fórmula

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



**Figura 2.3**  $P(s)$  y  $P_{GOE}(s)$

Gráficamente las dos curvas parecen tener la misma distribución, pero es necesario realizar una prueba de ajuste de las distribuciones. Se ha escogido la prueba de bondad de ajuste de Kolmogorov- Smirnov, cuyos resultados se presentan después de la Figura 2.4.

La Figura 2.4 muestra en rojo la distribución acumulada  $F_{GOE}$  y la distribución acumulada de  $s$  en azul.

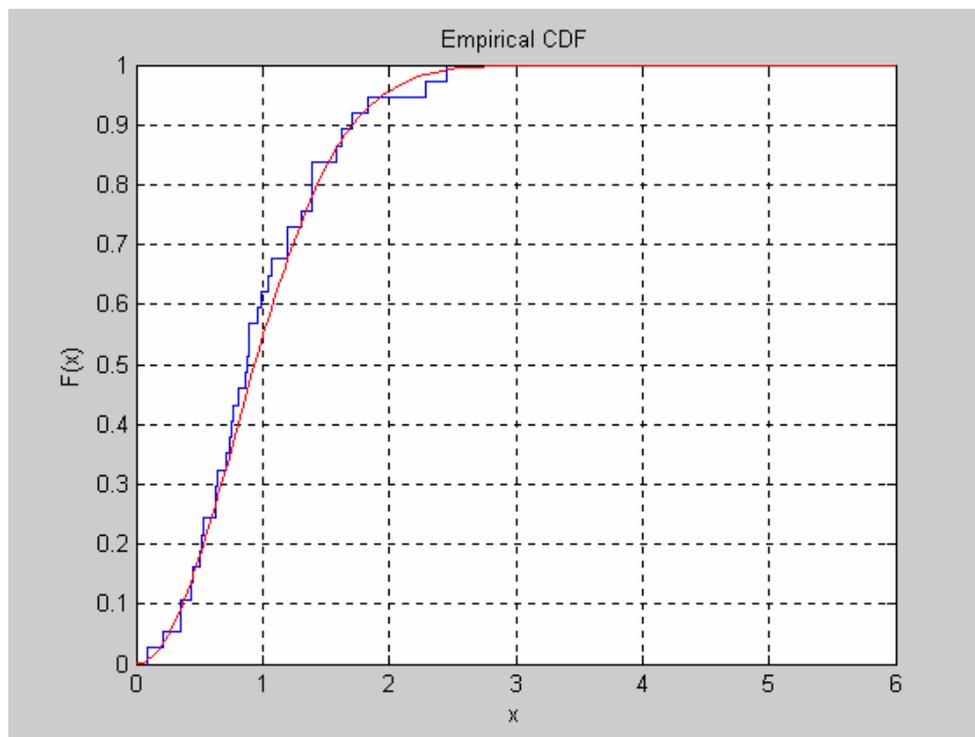


Figura 2.4  $F_{GOE}$  y  $F_s$

La grafica de las distribuciones acumuladas también muestra que las dos distribuciones se ajustan muy bien.

El resultado de la prueba de bondad de ajuste de Kolmogorov-Smirnov:

$$\begin{array}{l} D = \text{MAX}(0.0381860063200548, 0.105356234419861) \\ D^* = 0.655404953599772 \end{array}$$

permite concluir que la distribución de  $s$  es  $P_{GOE}(s)$  y de esta manera, que el espectro de  $C$  satisface la primera propiedad universal de la TMA.

2. Distribución de las diferencias de los valores propios de acuerdo al siguiente vecino más cercano.

En la segunda prueba, se consideran las diferencias de los valores propios “desplegados” de acuerdo al siguiente vecino más cercano

$$s' = \xi_{k+2} - \xi_k.$$

La Figura 2.5 muestra la distribución después de un reescalamiento de  $s' = \xi_{k+2} - \xi_k$  en azul y en rojo la de las matrices GOE dada por la fórmula

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

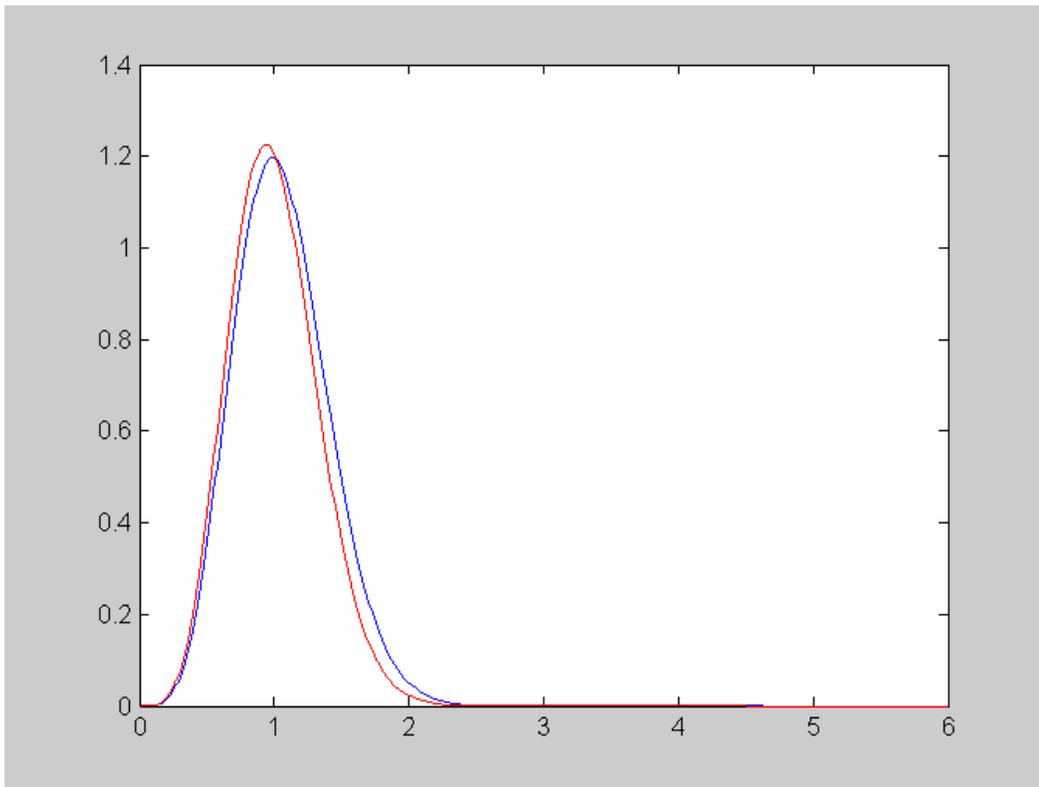


Figura 2.5  $P(s')$  y  $P_{GSE}(s')$

La gráfica muestra que las dos gráficas se ajustan muy bien.

Para confirmar la suposición se presenta el resultado de la prueba de bondad de ajuste de Kolmogorov-Smirnov :

**D=MAX(0,0.098889)**

**D\*=0.607014**

El cual, permite concluir que la distribución de  $s'$  es  $P_{GSE}(s')$  y de esta manera, que el espectro de C satisface la segunda propiedad universal de la TMA.

### 3. Correlaciones entre valores propios en un rango largo.

Para la tercera prueba, se usará el estadístico varianza número  $\Sigma^2$ , el cual se calcula empíricamente usando los valores propios desplegados en intervalos de longitud  $l$  alrededor de cada  $\xi_i$ , con la fórmula

$$\Sigma^2(l) \equiv \left\langle [n(\xi, l) - l]^2 \right\rangle_{\xi},$$

Donde  $n(\xi, l)$  es el número de valores propios desplegados en el intervalo  $[\xi - \frac{1}{2}, \xi + \frac{1}{2}]$  y  $\langle \dots \rangle_\xi$  es el promedio sobre todos los  $\xi$ .

La Figura 2.6 muestra las gráficas de las funciones  $f(l) = l$  en verde,  $\Sigma^2(l)$  en azul y  $g(l) = \ln(l)$  en rojo.

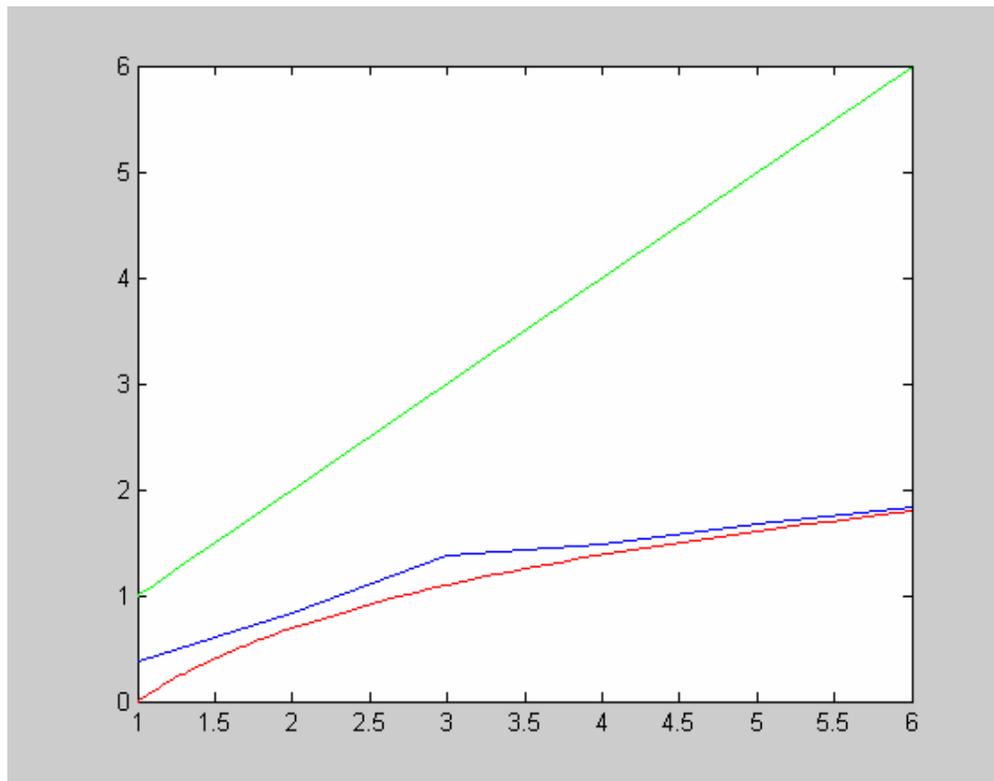


Figura 2.6  $\Sigma^2(l)$

Si los valores propios no estuvieran correlacionados  $\Sigma^2 \approx l$ , esto es, la función azul se acercaría a la verde, contrariamente de lo que sucede en la Figura. A medida que  $l$  crece  $\Sigma^2$  se separa de  $f(l) = l$ . Aunque para valores pequeños de  $l$  no es claro el acercamiento entre  $\Sigma^2$  en azul y  $\ln(l)$  en rojo, se nota que a medida que el valor de  $l$  aumenta,  $\Sigma^2$  se acerca a la función  $\ln(l)$ , esto es justamente lo que sucede en el espectro de las matrices GOE.

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. El vector propio asociado al mayor valor propio  $u^{65}$  tiene fuertes componentes en WalmexC , GFBBB, Cemex CPO, GcarsoA1, Televisa CPO, TV Azteca CPO y TelmexL. Las componentes mas pequeñas corresponden a Cid Mega Resort, y Sab.

Al analizar los otros dos vectores propios que se desvían de RMT  $u^{64}$  y  $u^{63}$  se encuentra que las componentes significativas pertenecen a

industrias similares o relacionadas. El vector propio  $u^{64}$  está dominado significativamente por empresas de comunicación: Telmex A y L, CEL y Telecom A1 con componentes negativas, la única componente positiva significativa es Apasco S.A. En el vector propio  $u^{63}$  dominado por empresas de comercio, se encuentra que las componentes negativas significativas corresponden a Walmart, Soriana y Elektra mientras que las positivas corresponden a Herdez y Gigante.

En cuanto a los vectores propios correspondientes a los valores propios más pequeños, se encuentra que contienen parejas de acciones como componentes significativas y son precisamente las que tienen las correlaciones más grandes  $C_{ij}$  de nuestra muestra. Por ejemplo, las dos mayores componentes de  $u^1$  corresponden a TelmexA y TelmexL con  $C_{ij} = 0.87291$ , la mayor correlación de la muestra. Las mayores componentes de  $u^2$  son WalmexC y WalmexV con  $C_{ij} = 0.85832$ , la segunda mayor correlación de la muestra. El valor propio  $u^3$  tiene a TelmexL y TelecomA1 como componentes principales con una correlación  $C_{ij} = 0.78744$  la tercera mayor de la muestra. En los tres vectores propios, el signo de las dos mayores componentes es contrario.

El cociente inverso de participación (CIP) ayuda a cuantificar el número de componentes que participan significativamente en cada vector propio.

La Figura 2.7 muestra el cociente inverso de participación (CIP)  $I^k$  como una función de los valores propios  $\lambda_k$  de la matriz C. La región en el rectángulo rojo muestra el intervalo  $[\lambda_+, \lambda_-]$

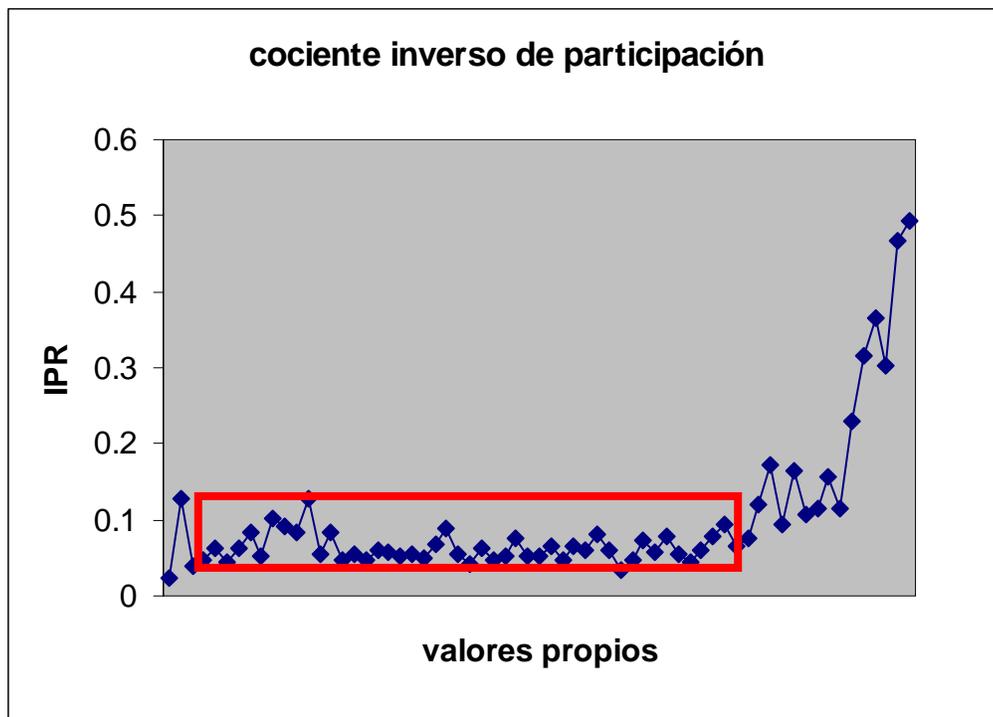


Figura 2.7

Se puede decir que prácticamente los valores en intervalo  $[\lambda_+, \lambda_-]$  permanecen en una banda o rango (el recuadro rojo), mientras que el CIP mínimo se obtiene en el valor propio más grande y los valores de CIP más altos, corresponden a los valores propios más pequeños.

$I^{65}$  tiene un valor de 0.024, si todas las componentes fueran iguales a

$\frac{1}{\sqrt{N}} = \frac{1}{\sqrt{65}}$ , el cociente sería 0.015 el cual representaría la influencia de

todo el mercado. El hecho de que algunas componentes dominen este vector da una idea de las empresas líderes en el mercado y justifica el hecho de que el cociente de participación inverso sea mayor que 0.015.

El vector propio  $u^{65}$  contiene aproximadamente  $1/I^{65} = 40$  participantes significativos, que son precisamente las acciones con mayor capitalización en el mercado.

En la grafica se observa que el cociente de participación inverso de los valores propios menores que  $\lambda_-$  son considerablemente mayores, de hecho son de tres a cinco veces el promedio del CIP, lo que sugiere que los vectores están localizados, esto es, solo algunas acciones contribuyen a ellos. Por ejemplo el vector propio  $u^1$  contiene  $1/I^1 = 2$  acciones que contribuyen a el: Telmex A y L.

Es interesante observar la influencia de la longitud de las series sobre la información obtenida. Para ello se analizan dos ventanas de la matriz  $C$ , la matriz  $C_1$  de tamaño  $400 \times 65$  tomando los 400 primeros datos y la matriz  $C_2$  de tamaño  $800 \times 65$  tomando los 800 primeros datos.

La Figura 2.8 muestra el histograma de la distribución de los valores propios de  $C_1$ .

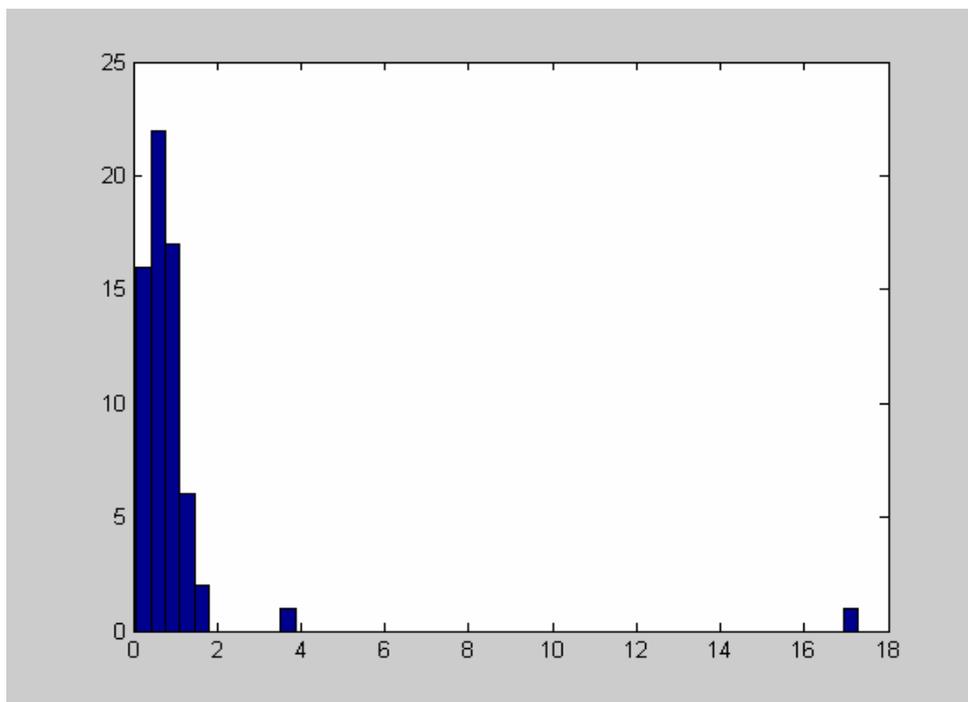


Figura 2.8 Valores propios de  $C_1$

Se puede observar dos valores propios mayores que  $\lambda_+$ , un valor propio menos, en ese rango, que en la matriz **C**. Al perder un valor propio, se pierde información.

La Figura 2.9 muestra, en azul, la distribución de los valores propios de  $C_1$  y en rojo la distribución de  $P_{rm}(\lambda)$ .

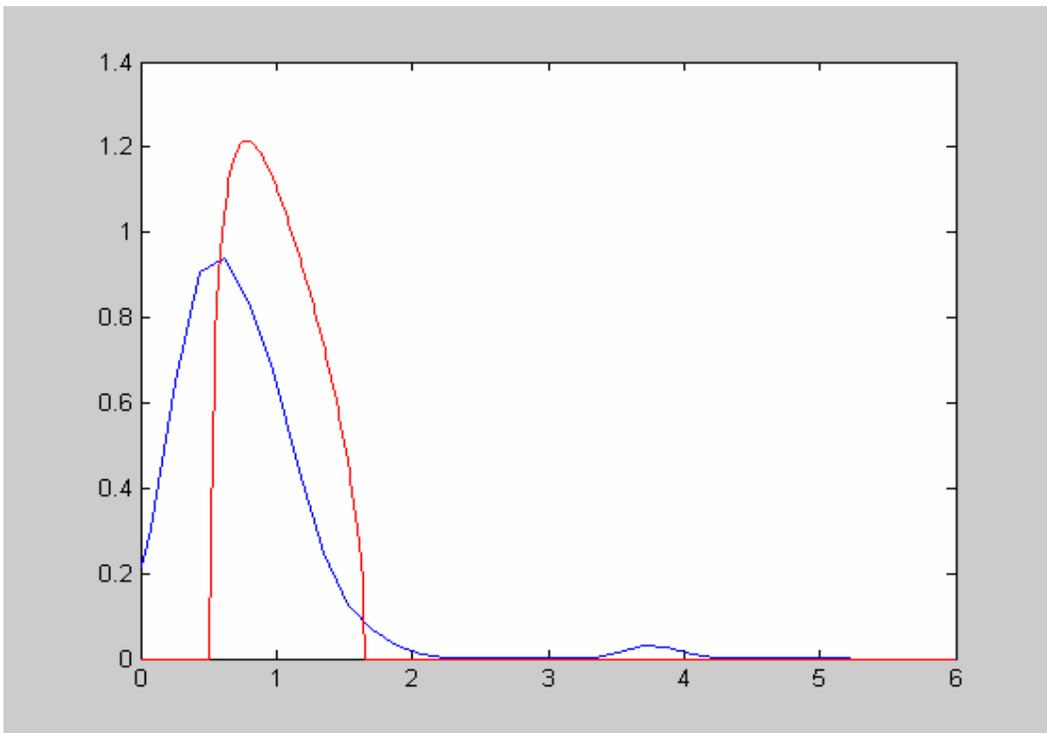


Figura 2.9 Valores propios de  $C_1$  y de  $P_{rm}(\lambda)$

Hay dos valores propios por encima de  $\lambda_+$ , los valores propios menores que  $\lambda_-$  (los más susceptibles al ruido) tienen una influencia tal que mueven la grafica a la izquierda y la achata. La matriz  $C_1$  satisface sólo la primera de las tres propiedades universales, así que la información que se obtiene de ella puede estar viciada, de hecho al comparar los resultados con los de la matriz  $C$ , se concluye que así es.

La Figura 2.10 muestra la distribución acumulada de  $s_i$  de  $C_1$  en azul y en rojo la de las matrices GOE.

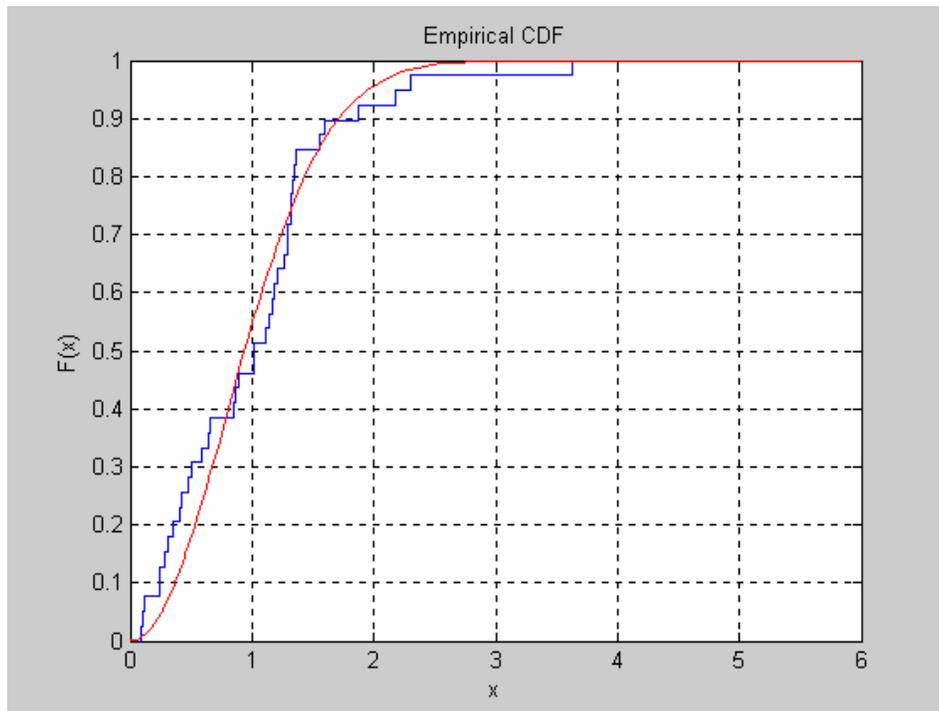


Figura 2.10

La Figura 2.11 muestra en azul la distribución de  $P(s_i \hat{)} de  $C_1$  y en rojo la de las matrices  $P_{GSE}$$

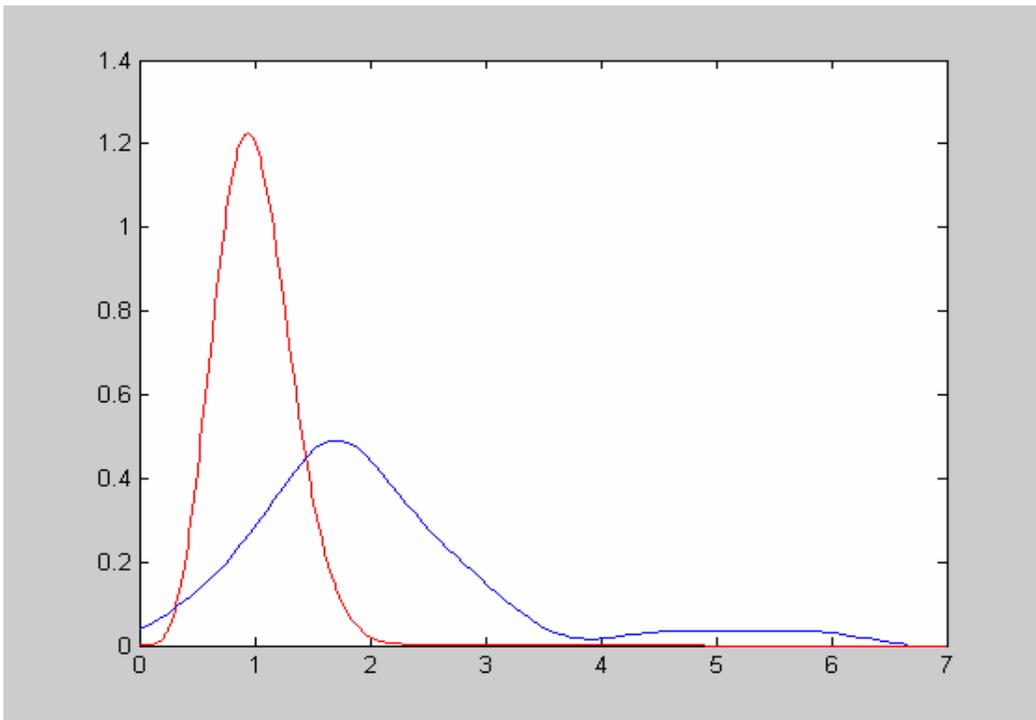


Figura 2.11  $P(s_i \hat{)} y  $P_{GSE}$$

Gráficamente se observa que  $P(s_i \hat{)} y  $P_{GSE}$$  no tienen la misma distribución, el mismo resultado se obtiene, usando Kolmogorov – Smirnov.

Al igual que en C, todas las componentes del vector propio asociado al mayor valor propio son positivas, lo que significa que no hay acciones con correlación inversa. El vector propio asociado al mayor valor propio  $u^{65}$  de  $C_1$  tiene fuertes componentes en Walmex C , GFBBB y Cemex CPO las demás empresas que aparecían en  $u^{65}$  de C ya no son significativas.  $u^{64}$  tiene como componentes significativas Telmex A y L.

Además de la información que se perdió de  $u^{65}$  y de  $u^{64}$  se ha perdido totalmente el tercer valor propio, que ahora pertenece al intervalo  $[\lambda_-, \lambda_+]$ .

La Figura 2.12 muestra la distribución de los valores propios de  $C_2$ .

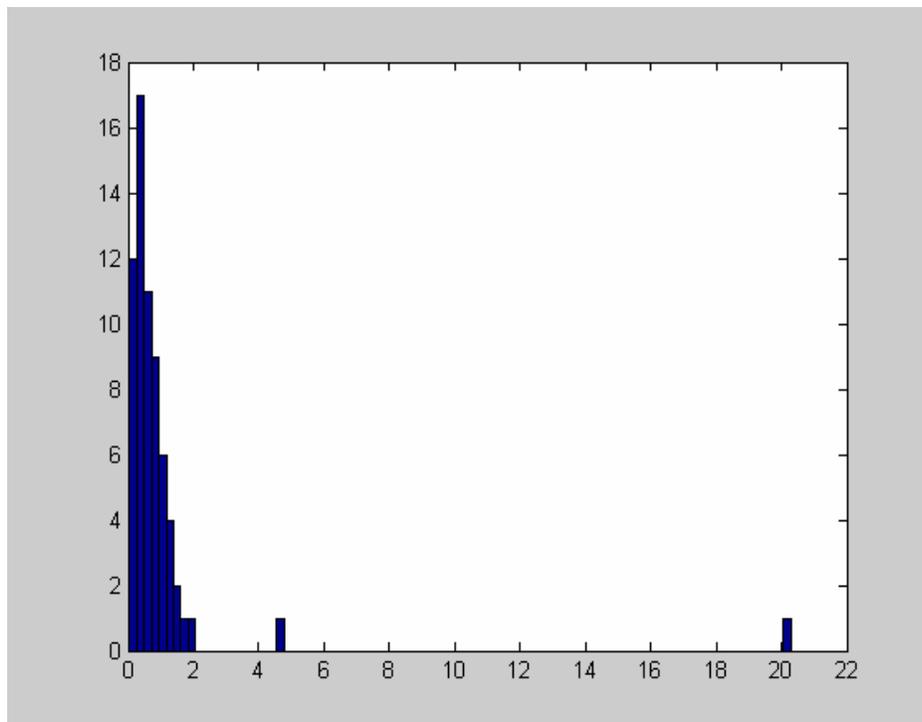


Figura 2.12

Se tiene dos valores propios por encima de  $\lambda_+$ , y un tercero a punto de escapar del grueso del espectro, sólo un poco mayor  $\lambda_+$ . Aún se tiene pérdida de información al comparar los resultados con los de la matriz **C**.

La Figura 2.13 muestra en azul la distribución de los valores propios de  $C_2$  y en rojo la distribución de  $P_{rm}(\lambda)$

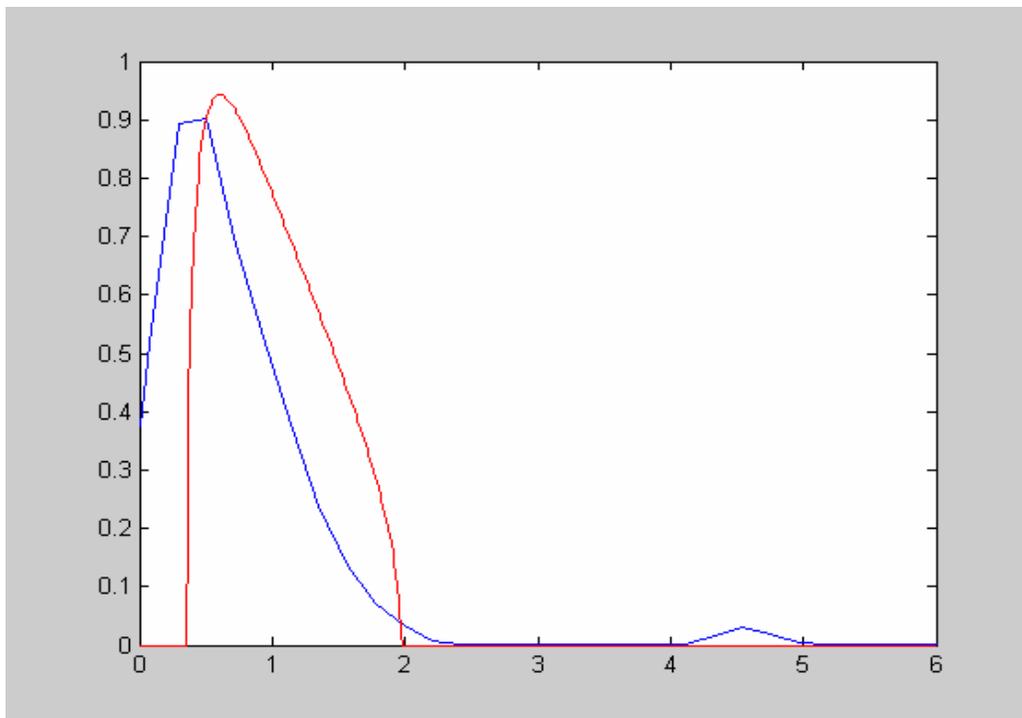


Figura 2.13 Distribución de los valores propios de  $C_2$

Aquí ya aparecen tres valores propios por encima de  $\lambda_+$ , los valores propios menores que  $\lambda_-$  mueven la grafica a la izquierda. Al igual que  $C_1$  La matriz  $C_2$  satisface sólo la primera de las tres propiedades universales.

La Figura 2.14 a la izquierda muestra la distribución acumulada de  $s_i$  de  $C_2$  en azul y en rojo la de las matrices GOE.

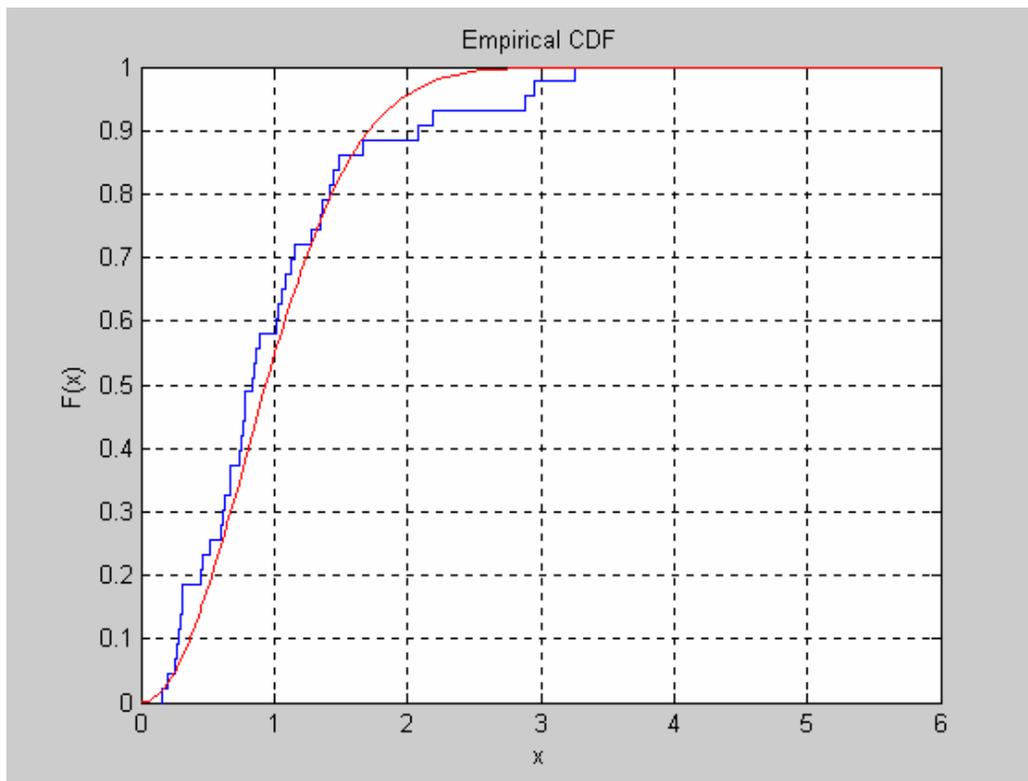


Figura 2.14

La Figura 2.15 muestra la distribución de  $P(s_i \hat{)}$  de  $C_2$  en azul y en rojo la de las matrices  $P_{GSE}$

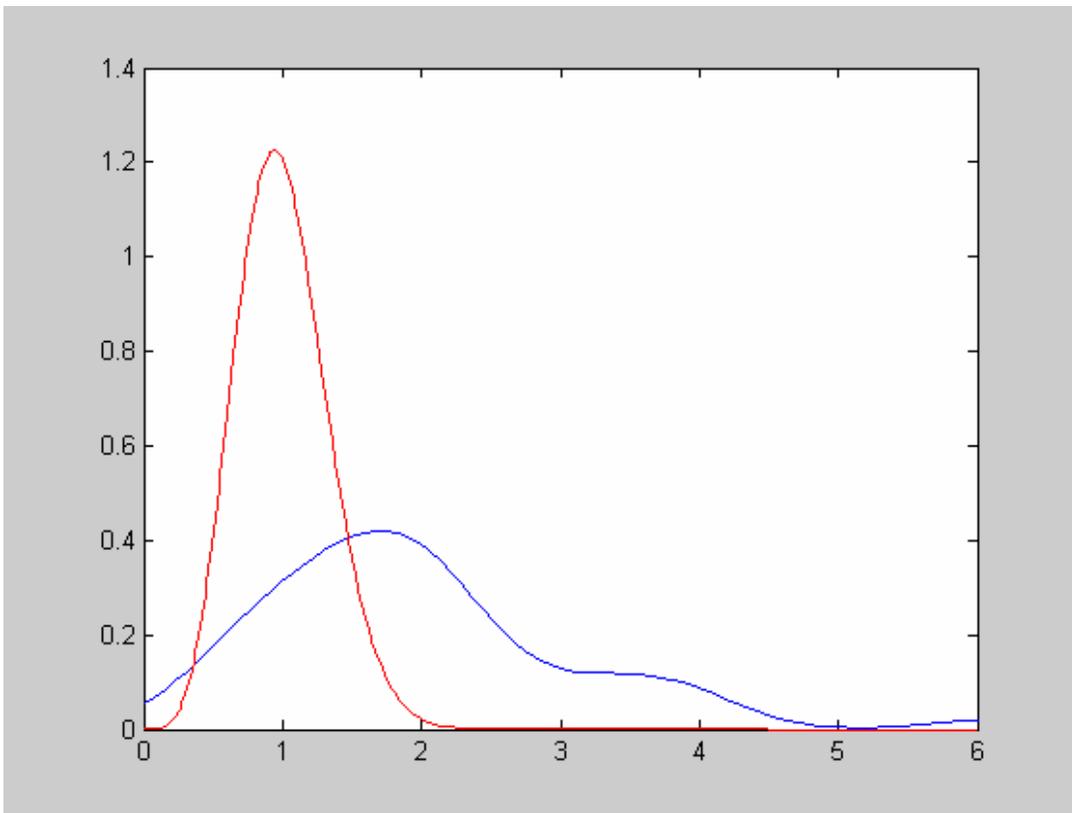


Figura 2.15

Claramente las distribuciones  $P(s_i \hat{)}$  de  $C_2$  y la de las matrices  $P_{GSE}$  son diferentes.

Al igual que en  $C$  y  $C_1$ , no hay acciones con correlación inversa.

El vector propio asociado al mayor valor propio  $u^{65}$  de  $C_2$  tiene fuertes componentes en WalmexC , GFBBB , Cemex CPO y TelmexL las demás empresas que aparecían en  $u^{65}$  de  $C$  ya no son significativas.  $u^{64}$  tiene como componentes significativas TelmexA y L y TelecomA1. En el vector propio  $u^{63}$  de  $C_2$  tiene las mismas componentes significativas que  $u^{63}$  de  $C$  : Walmart C y V.

A pesar de la información perdida, se encuentra que las empresas más significativas de cada vector permanecen iguales en todas las matrices.

## **2.4 Construcción de la matriz potencia $C^{(q)}$**

En esta sección se encuentra el valor de la potencia  $q$ , que de acuerdo con Guhr y Kalber [26] permitirá identificar y estimar el ruido de tal forma que la matriz  $C^{(q)}$  no posea aquellos elementos que están asociados al ruido, de alguna manera, este procedimiento es equivalente a ampliar la longitud de las series de tiempo.

Los elementos de  $C^{(q)}(T)$  se calculan de acuerdo a la formula

$$C_{kl}^{(q)}(T) = \text{sig}(C_{kl}(T)) |C_{kl}(T)|^q$$

Esto es, se eleva cada componente de  $C$  a la  $q$  y se preserva sus signos. Lo importante ahora es encontrar la potencia  $q$  cuyo espectro separe convenientemente el ruido de la información.

Usando Matlab se obtiene las matrices  $C^{(q)}$  para las potencias  $q = 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5$ .

Observando las gráficas de las densidades espectrales de cada una de las matrices anteriores, se puede identificar aquella que separe de una manera mas notable el ruido.

La Figura 2.16 muestra la densidad espectral  $\rho^{(q)}(\lambda)$  de la matriz potencia  $C^{(q)}$ .

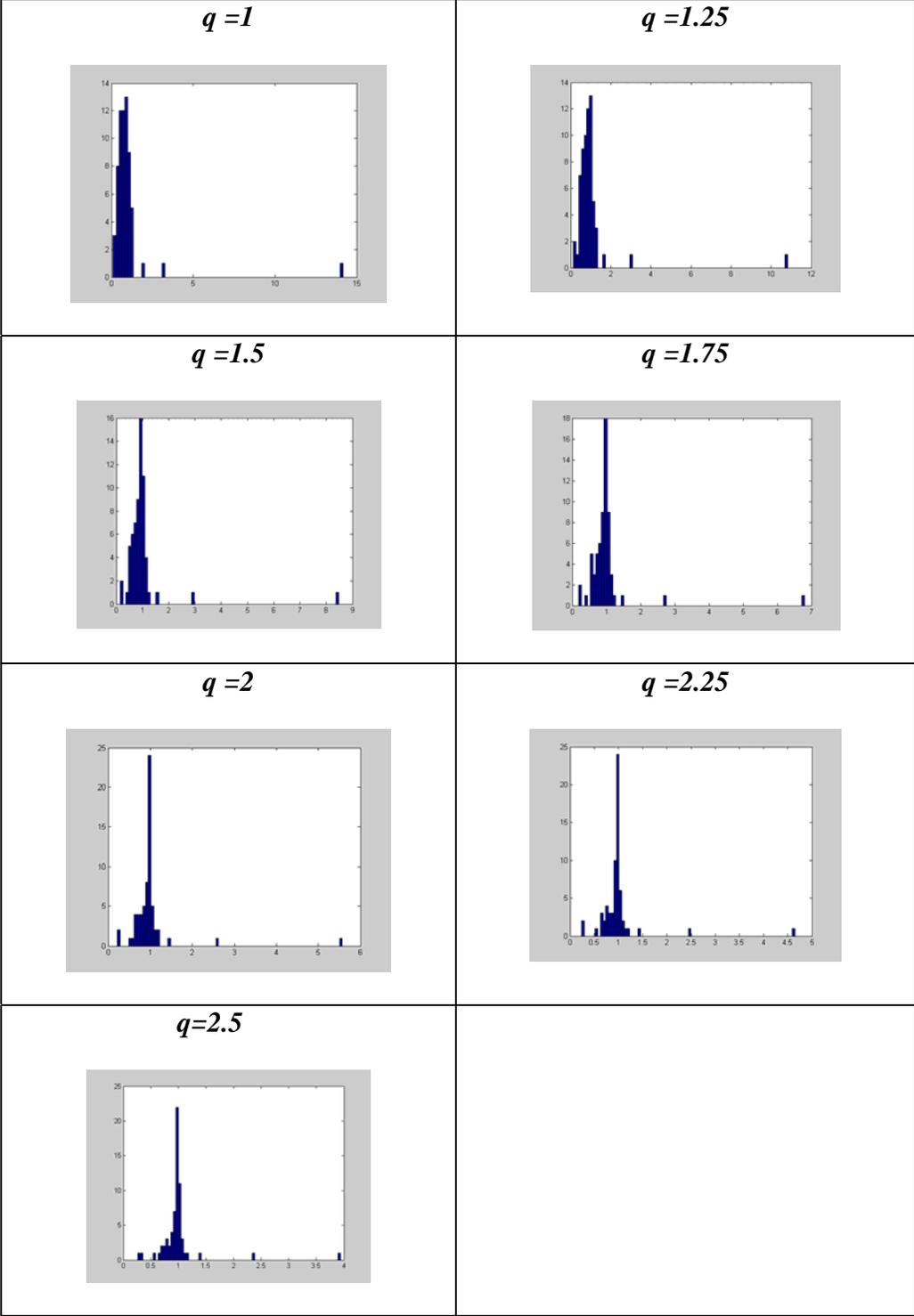


Figura 2.16

Lo mejor separación que se encuentra, corresponde a  $q = 2$ . Son tres pedazos en lugar de dos a causa del número de acciones incluidas en nuestro estudio y a la longitud de ellas.

Debido a que el número de valores propios de nuestro estudio es pequeño, no son claros los picos que se producen para los valores intermedios de  $q$ , sin embargo se puede notar que en  $q=1.25$  comienza una separación en tres partes que culmina en  $q=2$ , para valores mayores de 2 esta separación comienza lentamente a perderse.

Es interesante observar que  $\rho^{(2)}(\lambda)$  tiene una división similar a

$$[\lambda_1, \lambda_-) \cup [\lambda_-, \lambda_+] \cup (\lambda_+, \lambda_{64}]$$

que se obtiene de  $C$  usando la Teoría de Matrices Aleatorias.

Se escoge, pues,  $q=2$ . Se analizará la información que se obtiene de la matriz potencia  $C^{(2)}$ , usando el árbol de expansión mínima.

## 2.5 Construcción de la matriz $C_f$

La construcción de la matriz  $C_f$  comienza buscando la mejor aproximación con la serie de Fourier de cada una de las series de tiempo empíricas.

Usando Matlab se encuentra la TDF para cada serie, esto es

$$z_k = F(k) = \frac{1}{N} \sum_{n=0}^{N-1} f(n) e^{-jn \frac{2k\pi}{N}}$$

Para construir la ventana que permitirá filtrar la señal, se grafica el espectro de potencias

$$p = z_k * \text{conj}(z_k)$$

Se usa la grafica del espectro de potencia para encontrar empíricamente el valor de M para truncar la señal convenientemente, de ese modo se obtiene

$$\overline{F(k)} = F(k) * W$$

Donde  $W$  es un vector cuyas componentes son

$$W(i) = \begin{cases} 1 & \text{si } i \leq M \\ 0 & \text{si } i > M \end{cases}$$

La nueva ITDF es

$$f(n) = \sum_{k=0}^{N-1} F(k) e^{jk \frac{2n\pi}{N}}$$

La cual reconstruye la señal sin frecuencias altas (sin ruido).

El número  $M$  se encuentra empíricamente, ubicando el punto donde las frecuencias se hacen bajas. Para esto se usa el espectro de potencia y se selecciona el punto donde las frecuencias comienzan a ser bajas.

A continuación se expondrá el proceso paso a paso con la serie ICA. En el anexo se encuentran las gráficas de las demás series de tiempo y el programa de Matlab paso a paso para la construcción de las nuevas series.

La Figura 2.17 muestra el espectro de potencias de la serie ICA y el  $M$  escogido para filtrar la señal

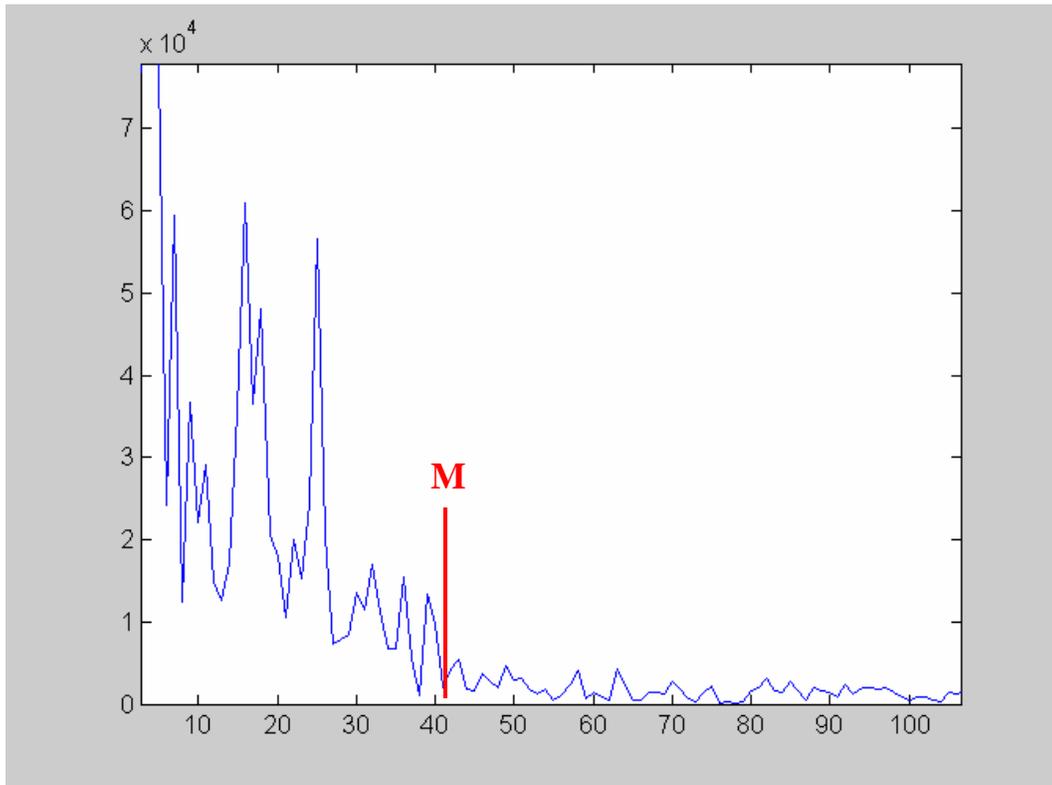


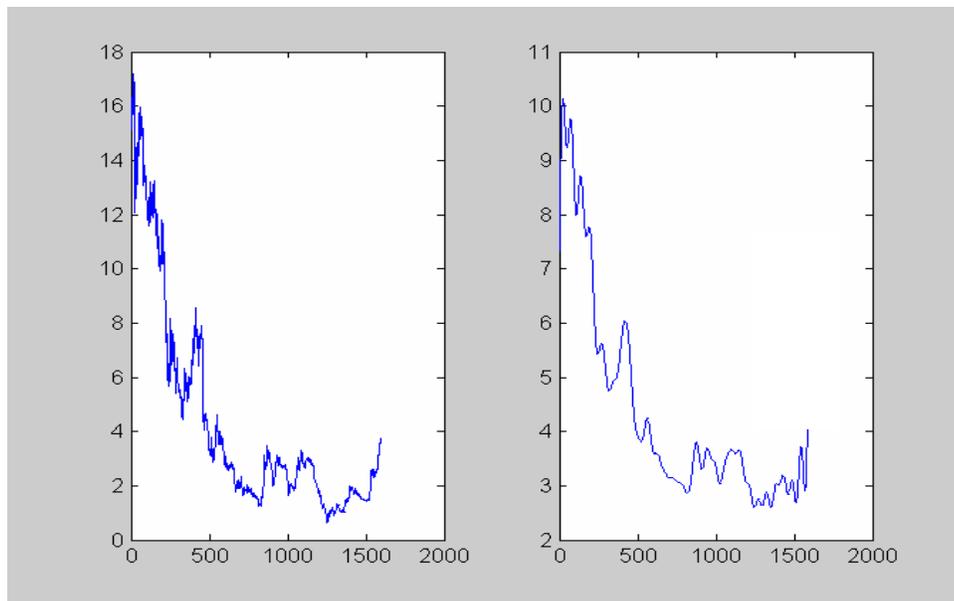
Figura 2.17 Espectro de potencias de la serie ICA

En la práctica, algunos usuarios de esta técnica suelen cortar la señal al terminar las frecuencias más altas, esto es  $M \approx 28$ . Aquí se ha decidido conservar algunas frecuencias bajas y tomar  $M = 42$ .

Es muy importante ver cómo la elección de  $M$  influye en la distribución de los errores cometidos al ajustar la señal. Este es otro aspecto importante a tener en cuenta en la elección de  $M$ .

Una vez encontrado  $M$  se puede pasar a la reconstrucción de la señal sin frecuencias altas.

La Figura 2.18 muestra a la derecha la señal original y a la izquierda su reconstrucción sin frecuencias altas.

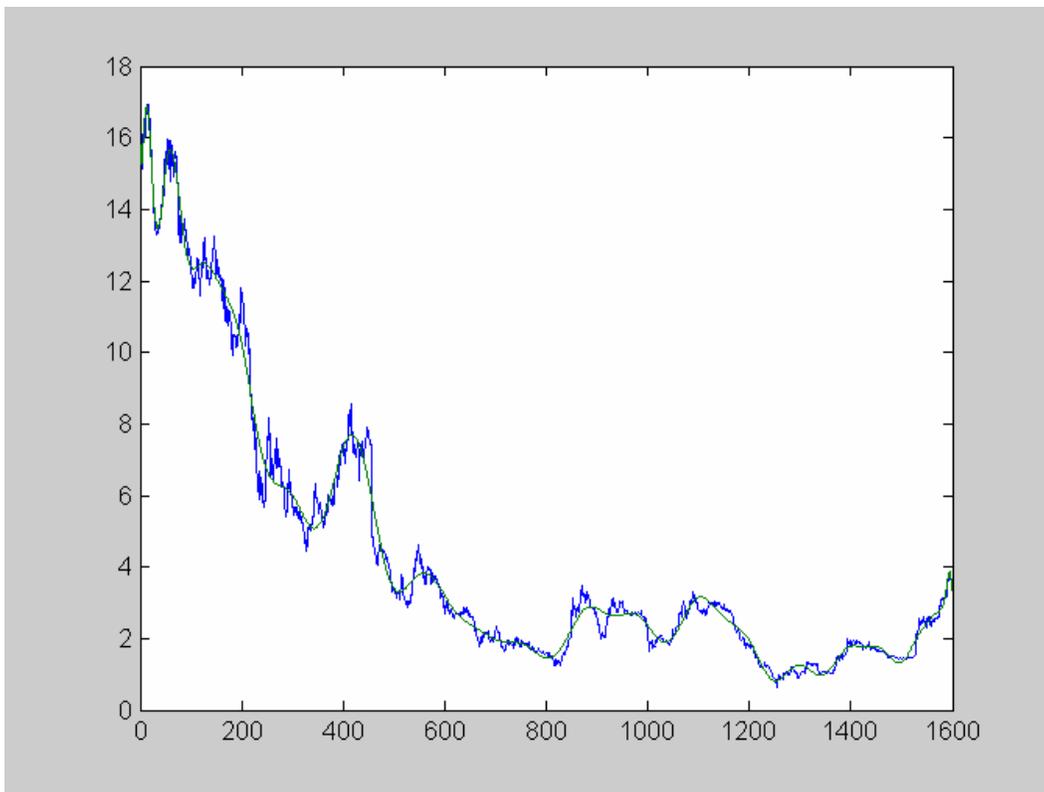


**Figura 2.18** Aproximación de la serie ICA con ITDF

La nueva serie ya no presenta cambios bruscos, pero ha conservado las propiedades importantes de la serie original y se espera que también

haya conservado la información importante, esto se verá al analizar la nueva matriz de correlación formada con las series de tiempo ajustadas.

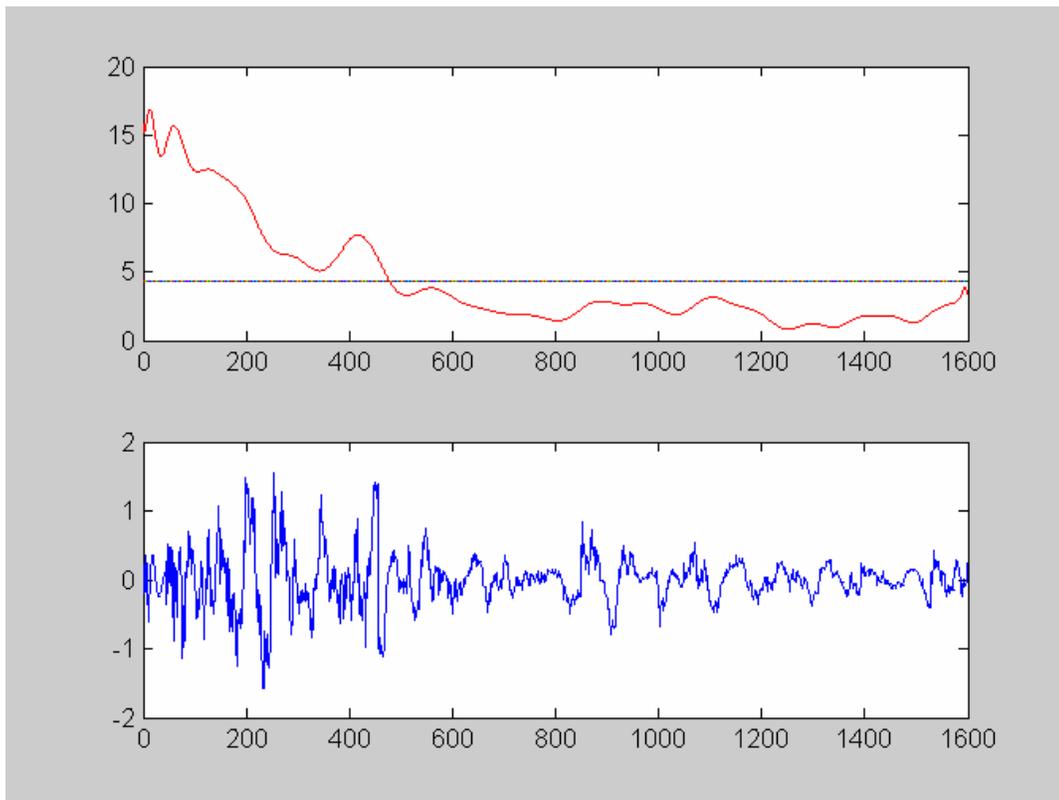
La Figura 2.19 muestra en azul la serie ICA y en verde el ajuste con polinomios trigonométricos, de la serie de tiempo de ICA.



**Figura 2.19** La serie ICA

Se observa que hay un buen ajuste entre las dos series, pero es importante prestar atención al error que se comete al hacer la aproximación.

La Figura 2.20 muestra en rojo la aproximación por serie de Fourier de la serie ICA y en azul los errores cometidos al ajustar la serie.

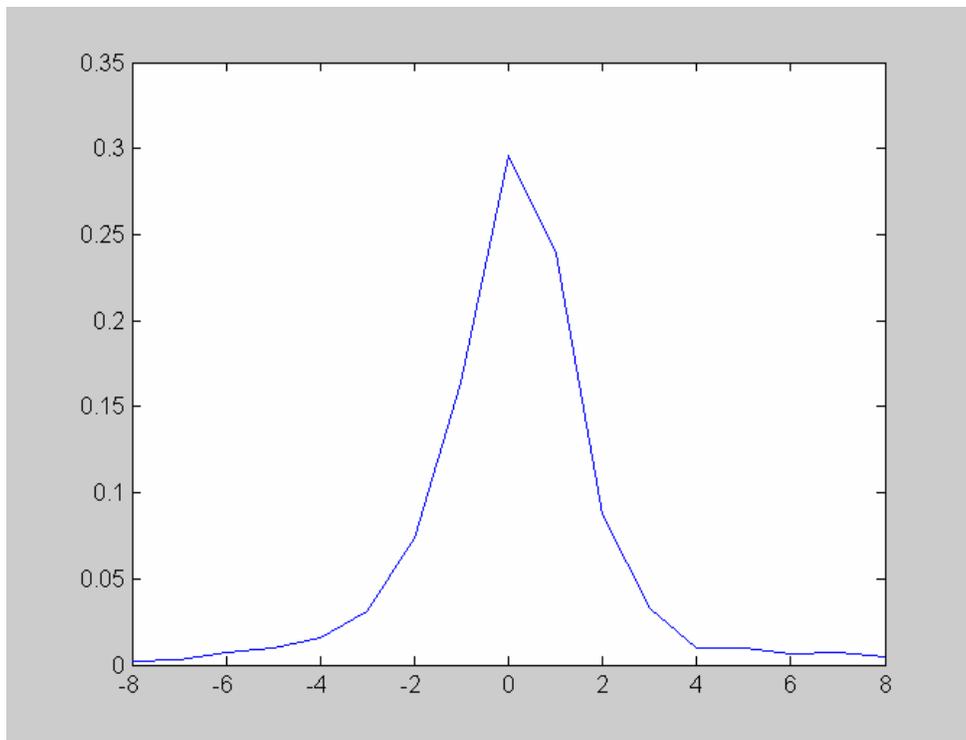


**Figura 2.20** La Serie aproximada de ICA y sus errores

En teoría el error cometido debe tener media cero y distribución normal.

Aunque en la práctica es difícil que los errores se distribuyan de tal forma, lo que se hace es escoger el valor de  $M$  que no solo separe las frecuencias altas de las bajas sino cuya distribución de los errores sea lo mas parecida a la deseada (se puede utilizar el estadístico Jarque- Bera con ese fin).

La Figura 2.21 muestra la distribución de los errores de la serie ICA.



**Figura 2.22 Distribución de los errores de ICA**

Una vez encontradas las aproximaciones trigonométricas de cada una de las series empíricas, se forma con ellas la matriz  $C_f$ .

La Figura 2.22 muestra la distribución de los valores propios de  $C_f$

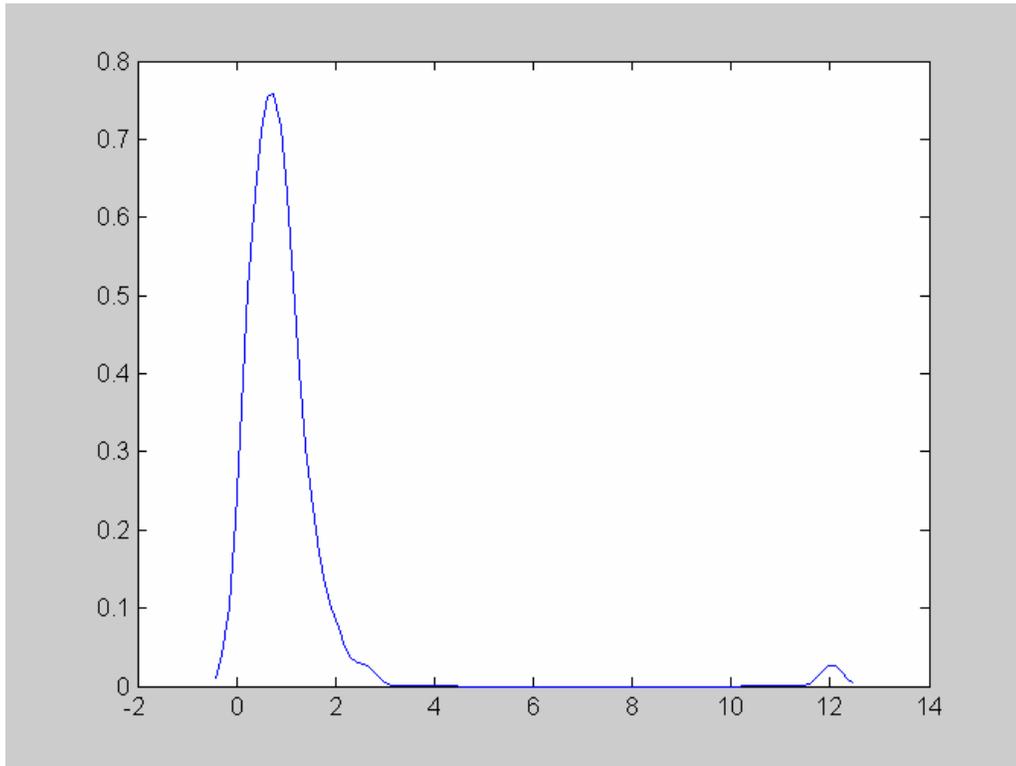


Figura 2.23 Distribución de los valores propios de  $C_f$

26 valores propios están por debajo de  $\lambda_-$ , 31 en el intervalo  $(\lambda_-, \lambda_+)$  y ahora se tienen 8 valores propios mayores que  $\lambda_+$  (5 más que en la matriz C), lo que significa que es posible que se puede obtener más información de la matriz  $C_f$ .

Después de desplegar los valores propios de  $C_f$ , se puede ver la distribución del vecino más cercano y la del siguiente vecino mas cercano.

La Figura 2.23 muestra, en azul, la distribución de los valores propios de  $C_{ij}$  y en rojo la distribución dada por  $P_{rm}(\lambda)$ .

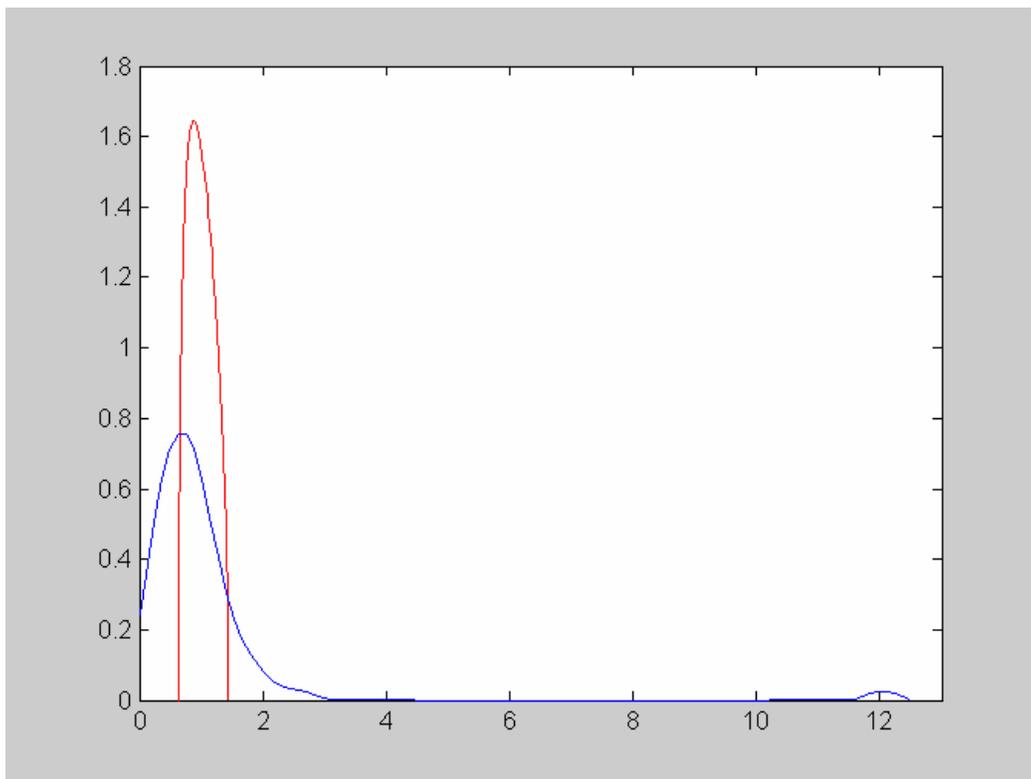


Figura 2.24 Distribución de los valores propios de  $C_f$  y  $P_{rm}(\lambda)$

Las diferencias entre las dos curvas de la Fig. 2.24 se deben precisamente a la existencia de valores propios por encima de  $\lambda_+$  (8 en total) y por debajo de  $\lambda_-$  (26).

La Figura 2.25 muestra en rojo la distribución  $F_{GOE}$  y la distribución de  $s$  en azul.

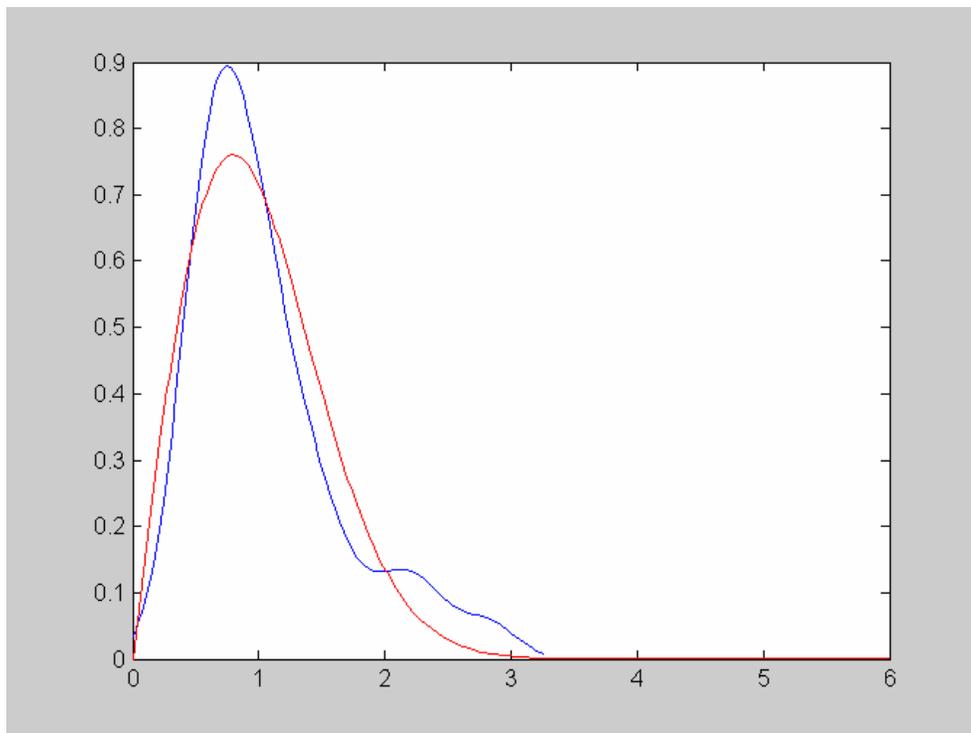


Figura 2.25  $F_{GOE}$  y  $F_s$

Puesto que los resultados de las pruebas no fueron suficientes para concluir que la matriz  $C_f$  satisface las propiedades universales de las matrices aleatorias, no se analizarán los valores propios.

La información de la matriz  $C_f$  se obtendrá a partir del árbol de expansión mínima.

**INSTITUTO TECNOLÓGICO Y DE ESTUDIOS  
SUPERIORES DE MONTERREY**

---

**CAMPUS CIUDAD DE MEXICO**



**UN NUEVO MÉTODO PARA LA CONSTRUCCIÓN DE PORTAFOLIOS.  
EL CASO MEXICANO.**

**DOCTORADO EN ADMINISTRACIÓN**

**TESIS PRESENTADA POR  
LINDA MARGARITA MEDINA HERRERA**

**JUNIO 2005**

**INSTITUTO TECNOLÓGICO Y DE ESTUDIOS  
SUPERIORES DE MONTERREY**

---

**CAMPUS CIUDAD DE MEXICO**



**UN NUEVO MÉTODO PARA LA CONSTRUCCIÓN DE PORTAFOLIOS.  
EL CASO MEXICANO.**

**DOCTORADO EN ADMINISTRACIÓN**

**TESIS PRESENTADA POR  
LINDA MARGARITA MEDINA HERRERA**

**ASESOR  
DR. RICARDO MANSILLA CORONA**

**JUNIO 2005**

## RESUMEN

En este trabajo se usan los métodos de la teoría de matrices aleatorias, la matriz potencia y las series de Fourier para analizar la matriz de correlación  $C$  de los rendimientos de los precios de las principales acciones de la Bolsa Mexicana de Valores. Además se utiliza la noción de árbol de expansión mínima, como un método para encontrar arreglos jerárquicos de acciones a través del estudio de conglomerados de compañías y como una herramienta importante en la elección de portafolios con riesgo mínimo y/o con altos rendimientos esperados.

Los resultados obtenidos muestran que la matriz de correlación,  $C$ , satisface las propiedades universales de las matrices aleatorias, permitiendo separar los valores y vectores propios que contienen información real de aquellos que tienen información inútil e inconsistente en el tiempo. Mas aún, se analiza los vectores propios de  $C$  a través del cociente de participación inverso en ambos lados del espectro.

Además de obtener información económica, a partir de los árboles de expansión mínima, se señalan las acciones elegibles para un portafolio óptimo y se da una medida del potencial de diversificación de la inversión en el mercado mexicano.

# ÍNDICE

<b>INTRODUCCIÓN .....</b>	<b>1</b>
<b>CAPITULO 1 .....</b>	<b>4</b>
<b>1.1. Resumen.....</b>	<b>4</b>
<b>1.2. Matrices de correlación .....</b>	<b>5</b>
1.2.1. Valores y vectores propios.....	10
<b>1.3. Teoría de Matrices Aleatorias.....</b>	<b>13</b>
1.3.1. Matrices aleatorias .....	15
1.3.2. Propiedades universales de las matrices aleatorias .....	18
1.3.2.1. Distribución de las diferencias de los valores propios de acuerdo al vecino más cercano. ....	25
1.3.2.1.1. La prueba de bondad de ajuste de Kolmogorov-Smirnov.....	26
1.3.2.2. Distribución de las diferencias de los valores propios de acuerdo al siguiente vecino más cercano.....	28
1.3.2.3. Correlaciones entre valores propios en un rango largo.....	29
1.3.3. Estadísticas de los valores propios.....	31
<b>1.4. La Matriz Potencia.....</b>	<b>32</b>
1.4.1. Identificación del ruido .....	36
<b>1.5. La Matriz C usando la transformada de Fourier .....</b>	<b>40</b>
<b>1.6. El árbol de expansión mínima.....</b>	<b>46</b>
1.6.1. Espacios ultramétricos .....	51
1.6.2. El promedio de ocupación y el vértice central.....	58
1.6.3. Análisis de portafolios.....	59
<b>CAPITULO 2 .....</b>	<b>63</b>
<b>2.1. Resumen.....</b>	<b>63</b>
<b>2.2. Las series de tiempo .....</b>	<b>65</b>
<b>2.3. Construcción de la matriz de correlación empírica <math>C_{ij}</math>.....</b>	<b>71</b>
2.3.1. Análisis de la matriz $C_{ij}$ usando la teoría de matrices aleatorias.....	75

2.4.	Construcción de la matriz potencia $C^{(q)}$ .....	94
2.5.	Construcción de la matriz $C_f$ .....	98
<b>CAPITULO 3 .....</b>		<b>108</b>
3.1.	Resumen.....	108
3.2.	Construcción y análisis de los árboles de expansión mínima.....	109
3.3.	Análisis de portafolios.....	116
<b>CONCLUSIONES.....</b>		<b>126</b>
<b>BIBLIOGRAFÍA.....</b>		<b>130</b>
<b>ANEXOS</b>		

## Conclusiones

Se ha encontrado que la mayoría de los valores propios en el espectro de la matriz de correlación  $C$  coinciden notablemente bien con las predicciones universales de la Teoría de Matrices Aleatorias. En particular, se ha encontrado que la matriz  $C$  satisface las propiedades universales del conjunto gaussiano ortogonal de matrices simétricas aleatorias, lo cual nos permite distinguir los valores y vectores propios de la matriz que contienen información real de aquellos que tienen información inútil e inestable en el tiempo.

La matriz  $C$  tiene 3 valores propios mayores que  $\lambda_+$ , los vectores propios asociados a estos valores son precisamente los que contienen información genuina. El vector propio  $u^{65}$ , nos da información general del mercado y nos muestra que no hay acciones con correlación inversa. Las empresas con mayores componentes en  $u^{65}$  (las que dan dirección al vector y por tanto dirigen el mercado) son: Cemex, TelmexL, Walmex, GFBB, Gcarso, Televisa, y TV Azteca. El segundo vector propio  $u^{64}$  está dominado significativamente por empresas de comunicación: Telmex, CEL y Telecom. El tercer vector propio esta sujeto a las principales empresas de comercio: Walmart, Soriana, Elektra y Gigante.

Es muy interesante resaltar que Cemex, la mayor componente del vector propio  $u^{65}$  es precisamente el vértice central del árbol de expansión mínima, se advierte aquí que dos métodos diferentes refieren a Cemex como un activo importante en el mercado.

El cociente inverso de participación soporta la idea de que algunas acciones dominan el mercado y más específicamente nos dice que el vector propio  $u^{65}$  contiene aproximadamente  $1/I^{65} = 40$  participantes significativos, que son precisamente las acciones con mayor capitalización en el mercado. ¿Será posible construir un índice de participación (tipo IPC) a partir del análisis de los vectores propios y del cociente de participación?

La longitud de las series (1598) ha permitido que la matriz C satisfaga las condiciones de la TMA. Pero se ha probado que para longitudes menores (400 y 800) las matrices correspondientes no satisfacen todas las propiedades universales de la TMA.

La matriz potencia ha resultado ser la más indicada para encontrar conglomerados por sectores económicos, mientras que la matriz de Fourier nos permite rescatar mayor información en general al tener 8 valores propios por encima de  $\lambda_+$  (5 mas que la matriz C) y resulta ser

más eficiente en separar las acciones para construir un portafolio eficiente usando el árbol de expansión mínima.

A través del estudio de los valores propios de la matriz  $C$ , se encuentra que se puede definir una métrica sobre el espacio de las compañías y que existe una distancia ultramétrica entre cada par de acciones.

Usando las distancias entre acciones y el algoritmo de Kruskal se construyen tres árboles de expansión mínima. Los árboles han mostrado ser un instrumento eficaz para visualizar la distribución de las acciones más importantes de la Bolsa Mexicana de Valores y para filtrar información económica y financiera de la matriz de correlación.

A diferencia de los mercados americanos, donde los conglomerados por sector económicos son claramente identificables, los árboles de la BMV sólo presentan dos: el de Minerales no Metalúrgicos liderado por Cemex y el de Telecomunicaciones liderado por Telmex.

Se analiza la longitud de los árboles como una manera de medir la diversificación del mercado (una de las formas de eliminar el riesgo sistemático) mediante “el promedio de ocupación” y se encuentra que aunque el mercado mexicano no está en crisis, es aún un mercado en desarrollo, donde el comportamiento del sistema es todavía muy homogéneo.

Cemex, el vértice central, coincide en todos los árboles con el centro de masa y por esto ha sido elegido como punto de referencia, contra el cual la posición de los demás vértices es relativa. De todos los posibles portafolios de Markovitz, las acciones del portafolio de riesgo mínimo están localizadas lo más lejos posible del vértice central, y a medida que se pasa hacia portafolios con altos rendimientos esperados, las acciones incluidas en ese portafolio estarán localizadas cerca del vértice central.

Es posible construir un programa que día a día muestre el árbol de expansión mínima de la BMV (árbol dinámico) y que calcule el vértice central y el promedio de ocupación diario. Teniendo estos datos de por lo menos un año, se podrán encontrar importantes conclusiones sobre el mercado mexicano.

De la misma forma se puede construir árboles con activos diferentes (no solo acciones), inclusive con los activos que componen un portafolio para analizar su eficiencia. Los árboles contruidos a partir de series de volatilidad también parecen ser interesantes.

## BIBLIOGRAFIA

- [1] Anderson, David. Sweeney, Dennis. Williams, Thomas. (2003) *An introduction to management science: quantitative approaches to decision making*. 10<sup>th</sup> ed. Cincinnati, Ohio: South-Western College Pub.
  
- [2] Bai, Z.D. (1999). *Methodologies in spectral analysis of large dimensional random matrices, a review*. Statistica Sinica 9. 611-677.
  
- [3] Batagelj, Vladimir. Mrvar, Andrej.(2005) *Pajek. Program for Analysis and Visualization of Large Networks. Reference Manual*.  
<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>
  
- [4] Bloomfield, Peter. (1976) *Fourier Analysis of Time Series: An Introduction*, *Wiley Series in Probability Mathematical*. John Wiley, United States of America,
  
- [5] Bonanno, Giovanni. Caldarelli, G. Lillo, F. Mantenga, R.N. (2002) *Topology of correlation based minimal spanning trees in real and model markets*. e-print <http://xxx.lanl.gov/cond-mat/0211546>.

- [6] Bonanno, Giovanni. Lillo, F. Mantegna, R.N. (2001) *High-frequency Cross-correlation in a Set of Stocks*, *Quantitative Finance*, 1, 96-104, e-print <http://xxx.lanl.gov/cond-mat/0009350>
- [7] Bouchaud, Jean-Phillipe. Potters, Marc. (2000) *Theory of Financial Risks -from statistical physics to risk management-*. Cambridge University Press, Cambridge.
- [8] Brody, T.A. Flores, J. French, J.P. Mello, P. Pandey, A.Wong, S. (1981) *Random- matrix physics spectrum and strength fluctuations*. *Rev.Mod. Phys.*, 53, No. 3, 385-480.
- [9] Burda, Zdzislaw. Jurkiewicz, Jerzy. (2004) *Signal and Noise in Financial Correlation Matrices v2* e-print <http://xxx.lanl.gov/cond-mat/0312496>
- [10] Burda, Zdzislaw. Jurkiewicz, Jerzy. Nowak, J. Papp, M.A. Zahed, G. I. (2003) *Levy Matrices and Financial Covariance*. e-print <http://xxx.lanl.gov/cond-mat/0103108>
- [11] Burda, Zdzislaw. Goerlich, A. Jarosz, A. Jurkiewicz, Jerzy. (2004) *Signal and Noise in Correlation Matrix*, e-print <http://xxx.lanl.gov/cond-mat/0305627>

- [12] Campbell, J. Lo, A. Mackinlay, A. (1997) *The econometrics of financial markets*. Princeton University Press.
- [13] Carr, P. Madan, D. B. (1999) *Option valuation using the fast Fourier transform*. Journal of Computational Finance 2, 61-73.
- [14] Carr, P. Wu, L. 1 (2004) *Time-changed Lévy processes and option pricing*. Journal of Financial Economics 71, 113-141.
- [15] Cerny, Ales. (2004) *Introduction to Fast Fourier Transform in Finance*. Working paper series. SSRN. Imperial College London.
- [16] Drozd, S. Kwapien, J. Grümmer, F. Ruf, F. Speth, J. (2001) *Quantifying dynamics of the financial correlations*. Physica A, 299, 144- , e-print <http://xxx.lanl.gov/cond mat/0102402>
- [17] Drozd, S. Kwapien, J. Speth, J. Wojcik, M. 314 (2002) *Identifying Complexity by Means of Matrices*. Physica A, 355-361, e-print <http://xxx.lanl.gov/cond-mat/0112271>
- [18] Edelman, A. (1988) *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix. Anal. Appl., No. 4, 543-560

- [19] Fama, E. French, K. (1992) *The cross section of expected share returns*, J. Finance, 47, 427-465.
- [20] Galluccio, S. Bouchaud, Jean-Phillipe. Potters, Marc. (1998) *Rational decisions, random matrices and spin glasses*. Physica A, 259, 449-456
- [21] Giada, L. Marsili, M. (2001), *Data clustering and noise undressing of correlation matrices*. Phys. Rev. E, 63, 061101-
- [22] Gopikrishnan, P. Rosenow, B. Plerou, V. Stanley, H. E. (2001) *Identifying Business Sectors from Stock Price Fluctuation*. Phys. Rev. E, 64, 035106
- [23] Grimaldi, Ralph. (1997) *Matemáticas discreta y combinatoria*. 3ª Edición. Pearson.
- [24] Gross, Jonathan. Yellen, J. (1999) *Graph theory and its applications*. Boca Raton, Fla. CRC Press.

- [25] Guhr, Thomas. Müller-Groeling, A. Weidenmüller, H. (1998) *Random matrix theories in quantum physics: common concepts*. Phys. Rep., 299, 190-
- [26] Guhr, Thomas. Kalber, Bernd. (1999) *A New Method to Estimate the Noise in Financial Correlation Matrices*. e-print <http://xxx.lanl.gov/cond-mat/0206577>
- [27] Gupta, A.K. Nagar, D. K. (2000) *Matriz variate distributions*. Chapman and Hall/CRG.
- [28] Hatanaka, Michio. (1972) *The Estimation of Spectra and Cross-Spectra on Short Time Series Data*. International Economic Review, Vol. 13, No. 3, 679-704.
- [29] Heston, S. (1993) *A closed-form solution for options with stochastic volatility with applications to bond and currency options*. Review of Financial Studies 6, 327—344.
- [30] Jhonson, Neil F. McDonald, Mark. Suleman, Omer. Williams, Stacy. Howison, Sam. (2005) *What shakes the FX tree? Understanding currency dominance, dependence and dynamics*. . e-print <http://xxx.lanl.gov/physics/0503014v1>

- [31] Kim, Dong-Hee. Jeong, Hawoong. (2005) *Systematic analysis of group identification in stock markets*. . e-print.  
<http://xxx.lanl.gov/physics/0503076v1>
- [32] Laloux, L. Cizeau, P. Bouchaud, Jean-Phillipe. Potters, Marc. (2002) *Noise Dressing of Financial Correlation Matrices*, *Phys. Rev. Lett.*, 83, 1467-1470.
- [33] 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.
- [34] Lee, W. (2000) *Theory and methodology of tactical asset allocation*. F.J.Fabozzi Associates, New Hope, Pennsylvania.
- [35] Koopmans, Lambert. (1995) *The spectral analysis of time series*. Probability and mathematical statistic. Volume 22. Academic Press.
- [36] Kullman, L. Kértész, J. Kaski, K. (1999) *Universal and non-universal properties of cross correlation in financial time series*. e-print <http://xxx.lanl.gov/cond-mat/9902283>.

- [37] Mansilla, Ricardo. (2003) *Una breve introducción a la econofísica*. Equipo Sirius.
- [38] Malevergne, Y. Soroette, D. (1999) *Hedging Extreme Co-Movements*. e-print <http://xxx.lanl.gov/cond-mat/0205636>
- [39] Mantegna, R. (1998) *Hierarchical Structure in Financial Markets*. Eur. Phys. J. B, 11, 193-197.
- [40] Mantegna, R. Stanley, H.E. (2000) *An Introduction to Econophysics - correlations and complexity in finance*. Cambridge University Press, Cambridge.
- [41] Mantenga, R. Bonanno, S. Lillo, G. (1999) *Degree stability of a minimum spanning tree of price return and volatility*. e-print <http://xxx.lanl.gov/cond-mat/0212338>
- [42] Mantegna, R. Et all. (2004) *Networks of equities in financial markets*. e-print <http://xxx.lanl.gov/cond-mat/041300v1>
- [43] Markowitz, H. (1959) *Portfolio Selection: Efficient Diversification of Investments*, Wiley, New York.

- [44] Marsili, M. (2002) *Dissecting financial markets: sectors and states*. *Quantitative Finance*, 2, 297-302.
- [45] Maslov S. (2001) Measures of globalization based on cross-correlations of world financial indices, *Physica A*, 301, 397-406.
- [46] Mehta, Madan. (1991) *Random Matrices*. 2<sup>nd</sup> edition. Academic Press.
- [47] Mehta, Madan. (1960) *On the statistical properties of the level spacing in nuclear spectra*, *Nucl. Phys.*, 18, 395-419.
- [48] Mellon, Giles W. (1964) *On the use of Time Series Analysis for Financial Prediction*. *The Journal of Finance*. Vol. 19. No. 2. Part 1: Papers and Proceedings of the Twenty-Second Annual Meeting of the American Finance Association. Boston, Massachusetts. 27-29, 170-185.
- [49] Merton, Robert C. (1994) *Influence of Mathematical Models in Finance on Practice: Past, Present and Future*. *Philosophical Transactions: Physical Sciences and Engineering*, Vol. 347. No. 1684, Mathematical Models in Finance 451-462.

- [50] Negrea, Bogdan. (2002) *Option Pricing with Stochastic Volatility: A closed-Form Solution Using the Fourier Transform*. National center for scientific research. SSRN. Working paper series.
- [51] Nicewander, Alan. Rodgers, Joseph. (1988) *Thirteen ways to look the correlation coefficient*. The American Statistian, Vol.42, No.1
- [52] Noh, J.D. (1999) *A model for correlations in stock markets*. e-print <http://xxx.lanl.gov/cond-mat/9912076v1>
- [53] Onnela, J.-P. Chakraborti, A. Kaski, K. (2002) *Dynamic asset trees and portfolio analysis*. e-print <http://xxx.lanl.gov/cond-mat/0208131>
- [54] Onnela, J.-P. Chakraborti, A. Kaski, K. (2003) *Dynamics of market correlations: Taxonomy and portfolio analysis*. e-print <http://xxx.lanl.gov/cond-mat/0302546v1>
- [55] Onnela, J.-P. Chakraborti, A. Kaski, K. (2003) *Asset trees and asset graphs in financial markets*. e-print <http://xxx.lanl.gov/cond-mat/0303579>

- [56] Pafka, S. Kondor, I. (2003) *Noisy covariance matrices and portfolio optimisation II*. Physica A, 319, 487-494, e-print <http://xxx.lanl.gov/cond-mat/0205119>
- [57] Pinsky, Mark. (2003) *Introducción al análisis de Fourier y las ondas*. Ed. Thomson.
- [58] 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.
- [59] Plerou, V. Gopikrishnan, P. Rosenow, B. Amaral, L. Guhr, Thomas. Stanley, H. E. (2001) *A Random Matrix Approach to Cross-Correlations in Financial Data*. e-print <http://xxx.lanl.gov/cond-mat/0108023v1>
- [60] Portillo Pérez de Viñaspre, Fabiola. Pérez Pascual, Pedro A. (2004) *Harmonic Analysis: The application of 'theoretical cycles' to the economic analysis. I Internacional Meeting on Economic Cycles. Quantitative Methodology of Economic Cycles*

- [61] Radicchi, Filippo. Castellano, Claudio. Cecconi, Federico. Loreto, Vittorio. Parisi, Domenico. (2003) *Defining and identifying communities in networks*. e-print <http://xxx.lanl.gov/cond-mat/0309488v1>
- [62] Rosenow, B. Plerou, V. Gopikrishnan, P. Stanley, H.E. (2002) *Portfolio optimization and the random magnet problem*. Europhys. Lett. A, 59, 500-506.
- [63] Rosenow, B. Plerou, V. Gopikrishnan, P. Stanley, H.E. (2003) *Dynamics of cross-correlations in the stock market*. Physica A, 324, 241-246
- [64] Ross, S. (2002) Probabilidad y estadística para ingenieros. Segunda edición. McGraw Hill.
- [65] Salvatore, M. Bonanno, G. Lillo, F. Mantenga, R. (1999) *Degree stability of a minimum spanning tree of price return and volatility*, e-print <http://xxx.lanl.gov/cond-mat/0212338>
- [66] Sengupta, A.M. Mitra, P.P. (1999) *Distribution of Singular Values for Some Random Matrices*. Phys.Rev. E, 60, 3389-

- [67] Tumminello, M. Aste, T. Di Matteo, T. Mantegna, R.N. (2005) *A new tool for filtering information in complex systems*. E-print <http://xxx.lanl.gov/cond-mat/0501335v1>
- [68] Vilela, R. Araújo T. Louca, F. (1999) *Reconstructing an economic space from a market metric*. e-print <http://xxx.lanl.gov/cond-mat/0211108>
- [69] Wang, Peijie. (2004) *Spectral Analysis of Economic Time Series Behaviour*. JEL No: 22.
- [70] West, Douglas B. (1996) *Introduction to graph theory*. Second edition. Prentice Hall.
- [71] Wilcox, Diane. Gebbie, Tim. (2004) *An Analysis of Cross-Correlations in South African Market Data*. e-print <http://xxx.lanl.gov/cond-mat/042389v1>
- [72] Yoon, Seong-Min. Kim, Kyungsik. (2005) *Financial Networks in the Korean Stock Exchange Market*. e-print <http://xxx.lanl.gov/physics/0503017v1>

## **Anexos**

**Total Variance Explained**

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	22.915	44.066	44.066	22.915	44.066	44.066
2	14.040	27.001	71.067	14.040	27.001	71.067
3	4.869	9.363	80.431	4.869	9.363	80.431
4	3.528	6.785	87.216	3.528	6.785	87.216
5	1.514	2.912	90.127			
6	1.228	2.362	92.490			
7	.613	1.179	93.669			
8	.558	1.073	94.742			
9	.429	.825	95.567			
10	.353	.679	96.245			
11	.297	.571	96.816			
12	.220	.423	97.239			
13	.152	.293	97.532			
14	.132	.254	97.786			
15	.112	.215	98.001			
16	.103	.198	98.198			
17	9.665E-02	.186	98.384			
18	8.320E-02	.160	98.544			
19	7.709E-02	.148	98.692			
20	6.493E-02	.125	98.817			
21	6.170E-02	.119	98.936			
22	5.847E-02	.112	99.048			
23	4.991E-02	9.599E-02	99.144			
24	4.391E-02	8.445E-02	99.229			
25	4.090E-02	7.865E-02	99.307			
26	3.851E-02	7.407E-02	99.382			
27	3.467E-02	6.667E-02	99.448			
28	3.077E-02	5.917E-02	99.507			
29	2.967E-02	5.706E-02	99.564			
30	2.790E-02	5.366E-02	99.618			
31	2.450E-02	4.712E-02	99.665			
32	2.233E-02	4.294E-02	99.708			
33	2.023E-02	3.891E-02	99.747			
34	1.961E-02	3.771E-02	99.785			
35	1.809E-02	3.479E-02	99.820			
36	1.567E-02	3.013E-02	99.850			
37	1.221E-02	2.348E-02	99.873			
38	1.183E-02	2.276E-02	99.896			
39	1.025E-02	1.971E-02	99.916			
40	9.923E-03	1.908E-02	99.935			
41	9.304E-03	1.789E-02	99.953			
42	7.863E-03	1.512E-02	99.968			
43	6.130E-03	1.179E-02	99.980			
44	3.811E-03	7.328E-03	99.987			
45	2.250E-03	4.327E-03	99.991			
46	1.071E-03	2.059E-03	99.993			
47	9.378E-04	1.803E-03	99.995			
48	8.374E-04	1.610E-03	99.997			
49	6.438E-04	1.238E-03	99.998			
50	4.866E-04	9.358E-04	99.999			
51	4.187E-04	8.052E-04	100.000			
52	1.823E-04	3.506E-04	100.000			

Extraction Method: Principal Component Analysis.

Este programa filtra la serie.

Despues del load hay que poner el path completo del file.

```
load c:\tesis\linda\series\ica.txt;
```

```
x=ica;
```

```
% Aqui se calcula la transformada de Fourier
```

```
z=fft(x);
```

```
%Aqui se multiplica por la conjugada para calcular el espectro de potencias.
```

```
p=z.*conj(z);
```

```
plot(p(10:911));
```

```
%Aqui se contruye la ventana para filtrar la señal. El 30 se puede
```

```
%cambiar, pero hay que restarle 30 a 1598, que es la longitud total de la serie.
```

```
w=[ones(1,30) zeros(1,1568)'];
```

```
% zn es la transformada de Fourier filtrada
```

```
zn=z.*w;
```

```
% Aqui se reconstruye la señal sin las frecuencias altas (ifft).
```

```
xn=ifft(zn);
```

```
figure;
```

```
subplot(1,2,1),plot(x);
```

```
subplot(1,2,2),plot(real(xn));
```

En este programa se calcula el ajuste con polinomios

```
% trigonométricos de la serie de tiempo
```

```
% a partir del análisis espectral hecho en filtrado.m.
```

```
load c:\tesis\linda\series\ica.txt;
```

```
x=ica;
```

```
y=mean(x);
```

```
zf=fft(y);
```

```
wp=[ones(40,1)' zeros(1558,1)'];
```

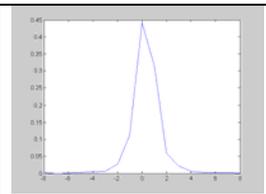
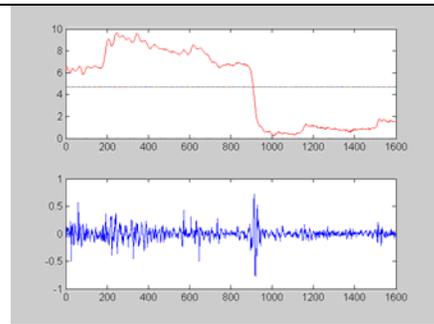
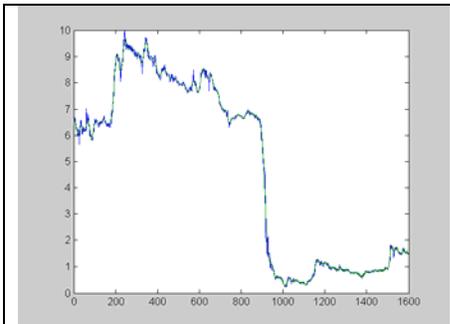
```
xr=real(ifft(zf.*wp));
```

```
t=1:1:1598;t=t';
```

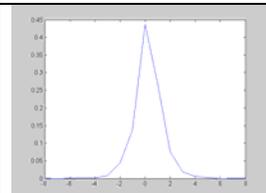
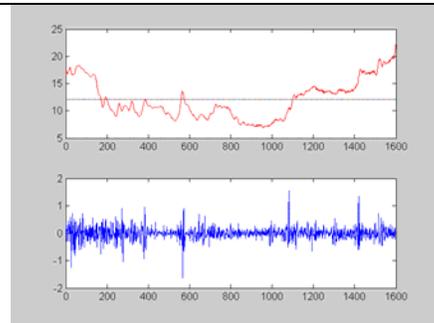
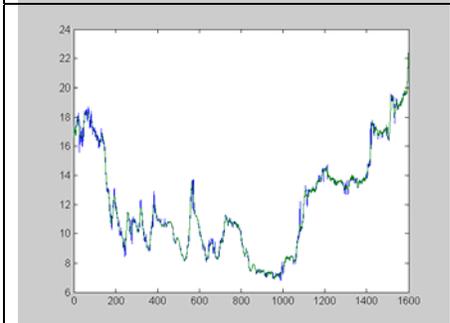
```
y1=x-mean(y).*ones(size(y));
```

```
w=pi/1557;
```

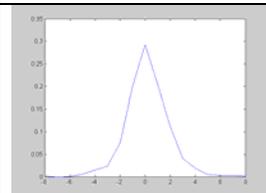
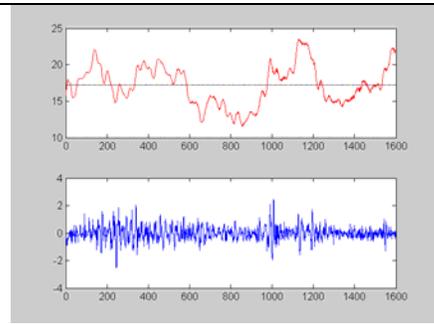
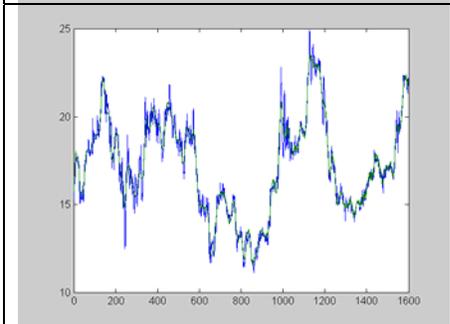
```
X=[cos((2*w).*t) sin((2*w).*t) cos((3*w).*t) sin((3*w).*t) cos((4*w).*t) sin((4*w).*t) cos((5*w).*t) sin((5*w).*t) cos((6*w).*t) sin((6*w).*t) cos((7*w).*t) sin((7*w).*t) cos((8*w).*t) sin((8*w).*t) cos((9*w).*t) sin((9*w).*t) cos((10*w).*t) sin((10*w).*t) cos((11*w).*t) sin((11*w).*t) cos((12*w).*t) sin((12*w).*t) cos((13*w).*t) sin((13*w).*t) cos((14*w).*t) sin((14*w).*t) cos((15*w).*t) sin((15*w).*t) cos((16*w).*t) sin((16*w).*t) cos((17*w).*t) sin((17*w).*t) cos((18*w).*t) sin((18*w).*t) cos((19*w).*t) sin((19*w).*t) cos((20*w).*t) sin((20*w).*t) cos((21*w).*t) sin((21*w).*t) cos((22*w).*t) sin((22*w).*t) cos((23*w).*t) sin((23*w).*t) cos((24*w).*t) sin((24*w).*t) cos((25*w).*t) sin((25*w).*t) cos((26*w).*t) sin((26*w).*t) cos((27*w).*t) sin((27*w).*t) cos((28*w).*t) sin((28*w).*t) cos((29*w).*t) sin((29*w).*t) cos((30*w).*t) sin((30*w).*t) cos((31*w).*t) sin((31*w).*t) cos((32*w).*t) sin((32*w).*t) cos((33*w).*t) sin((33*w).*t) cos((34*w).*t) sin((34*w).*t) cos((35*w).*t) sin((35*w).*t) cos((36*w).*t) sin((36*w).*t) cos((37*w).*t) sin((37*w).*t) cos((38*w).*t) sin((38*w).*t) cos((39*w).*t) sin((39*w).*t) cos((40*w).*t) sin((40*w).*t)]
```



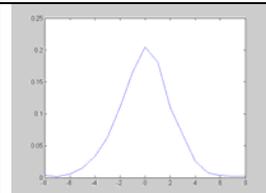
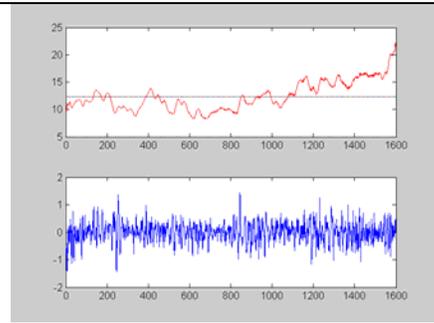
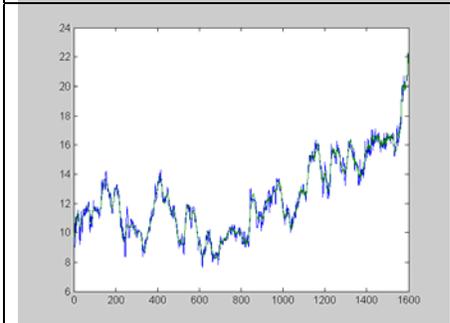
Saviaa



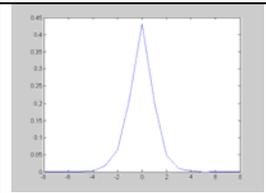
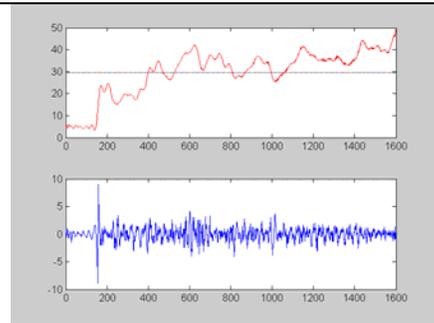
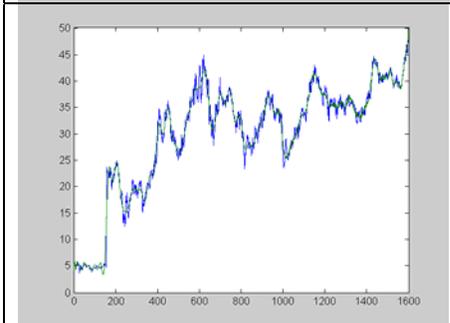
Bachoco



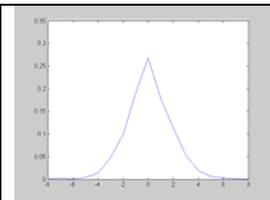
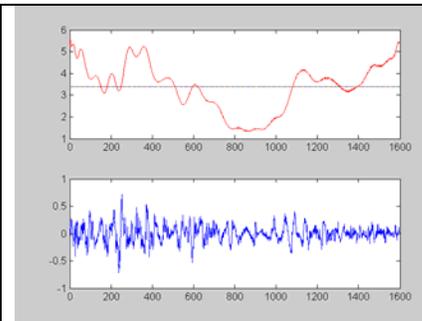
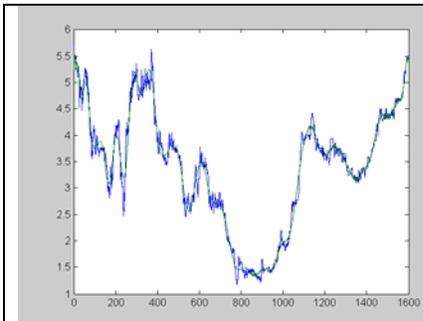
Bimboa



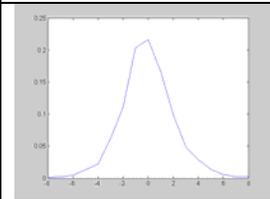
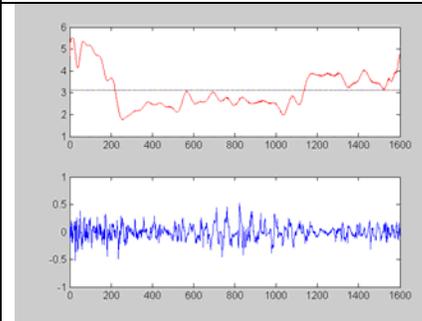
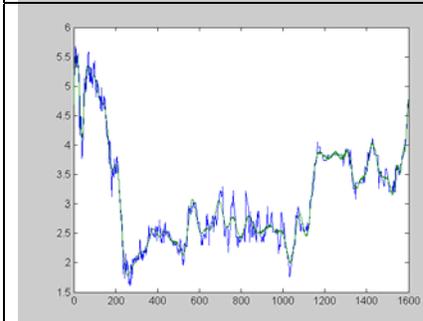
Contal



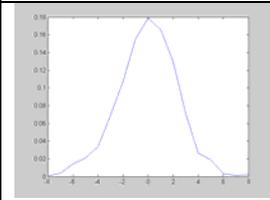
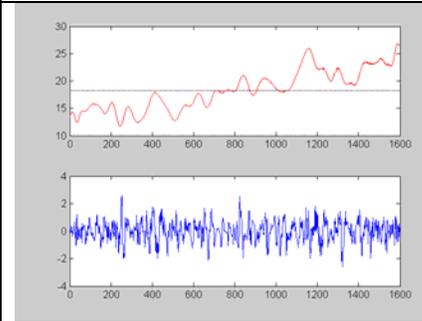
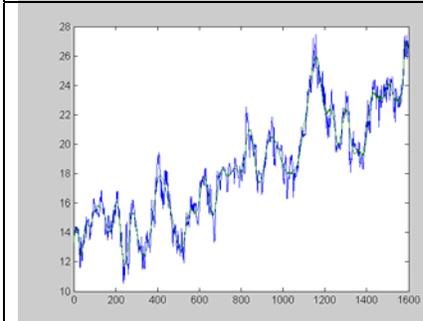
Femsa



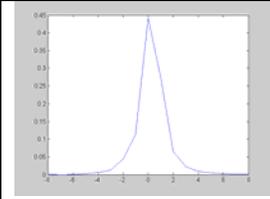
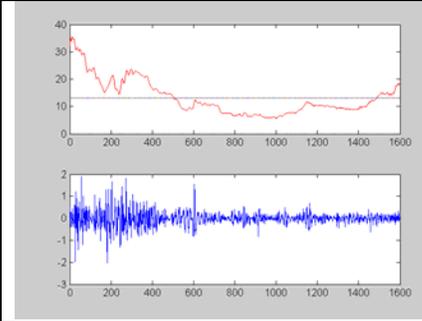
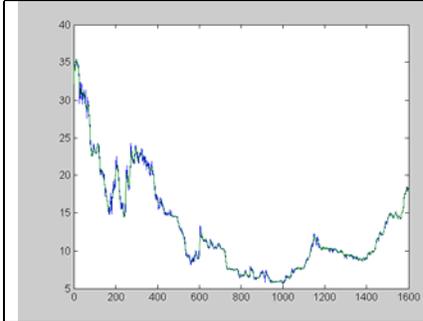
Maseca



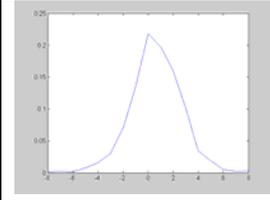
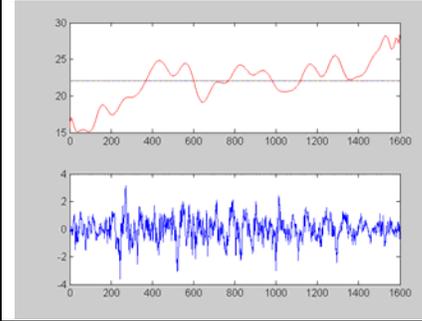
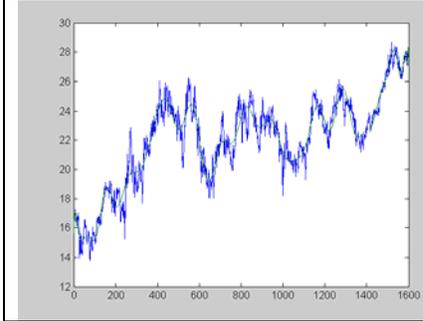
Herdez



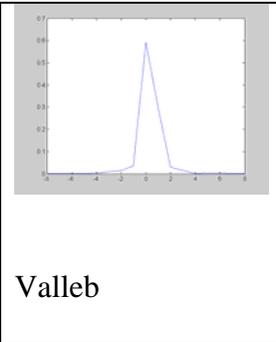
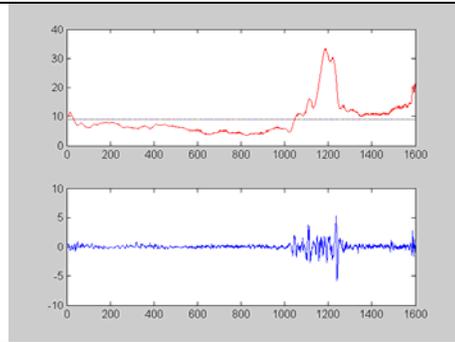
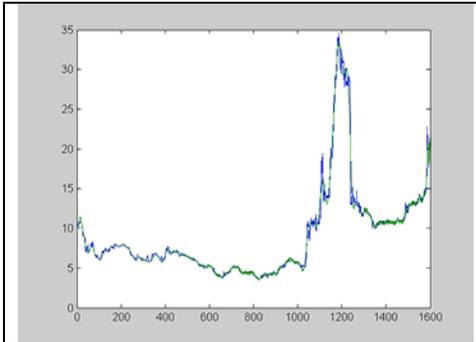
Kofl



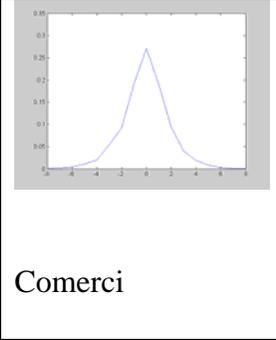
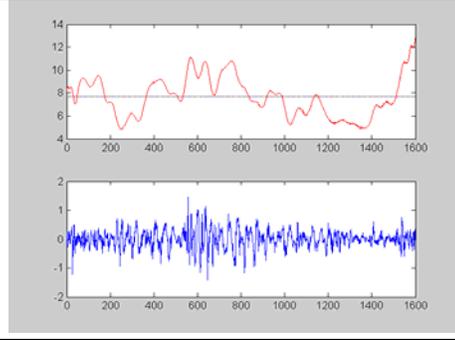
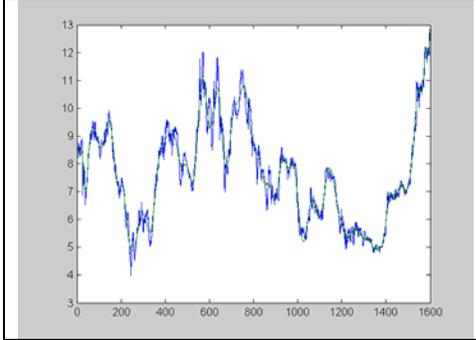
Gruma



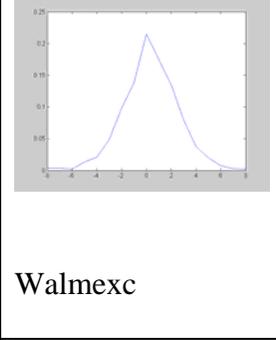
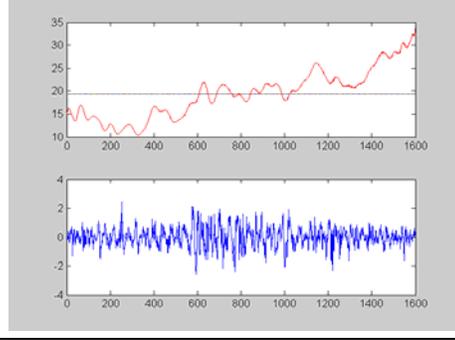
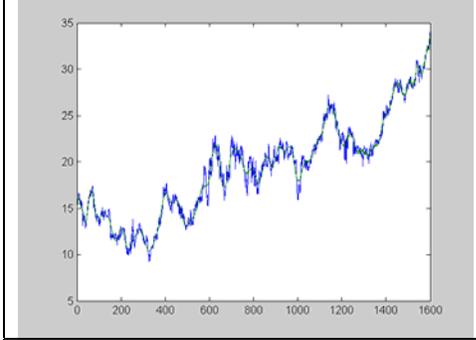
Gmodelo



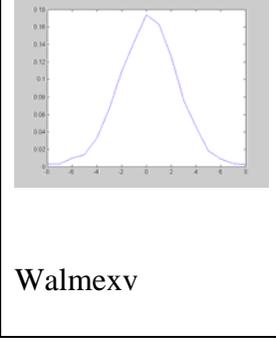
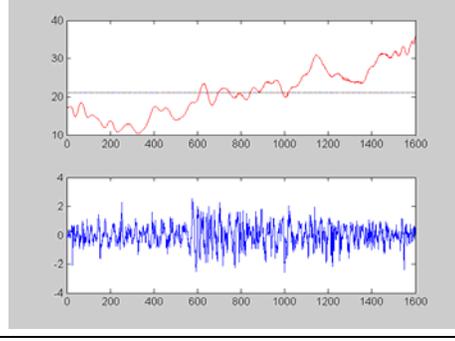
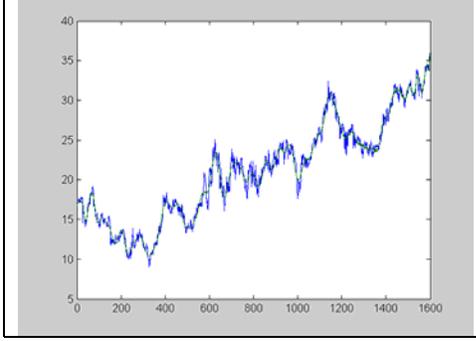
Valleb



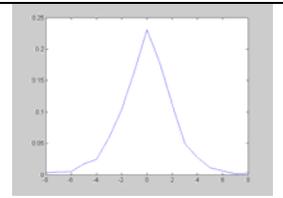
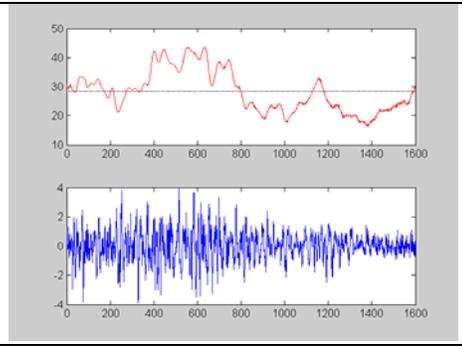
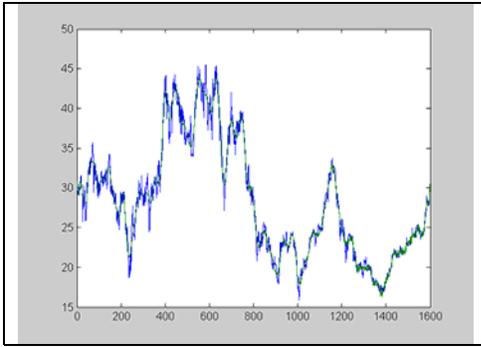
Comerci



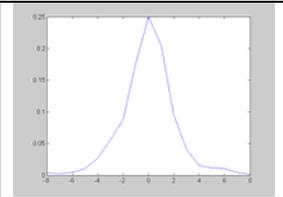
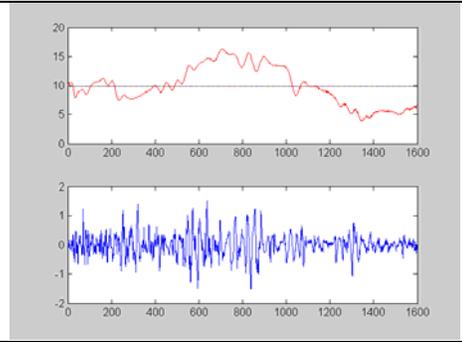
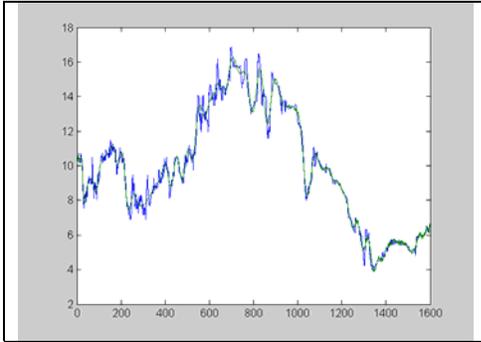
Walmexc



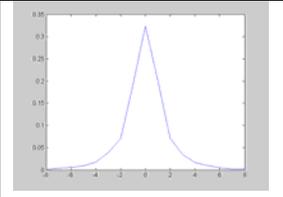
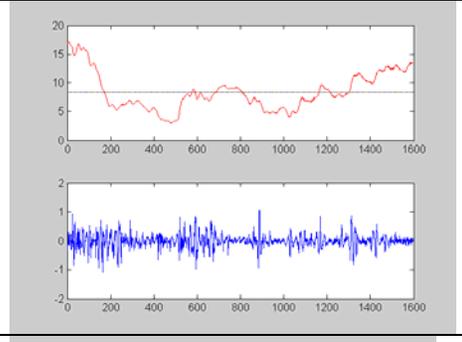
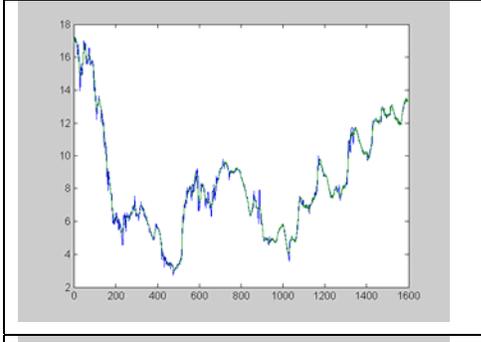
Walmexv



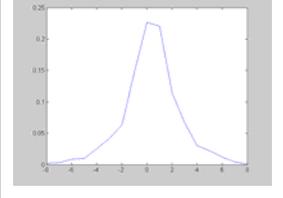
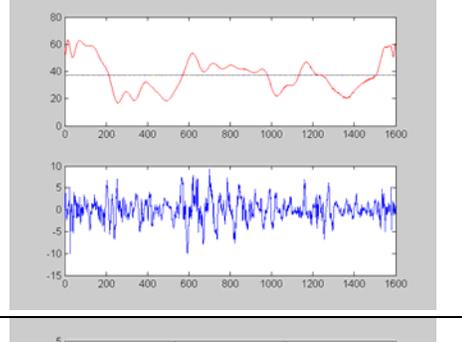
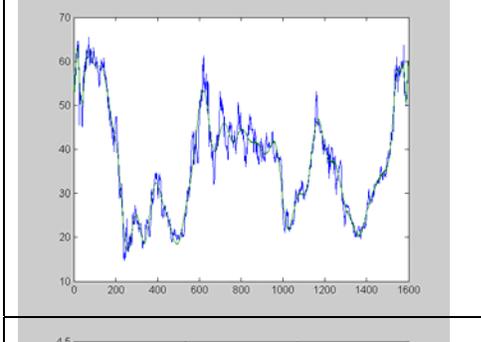
Soriana



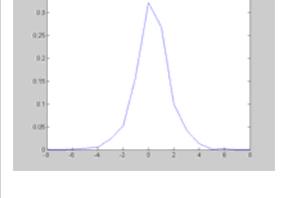
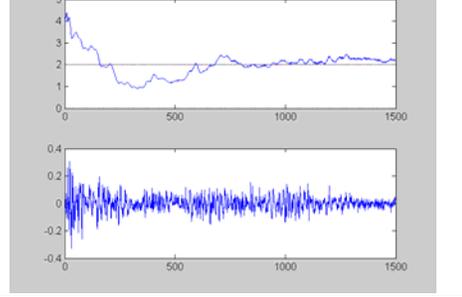
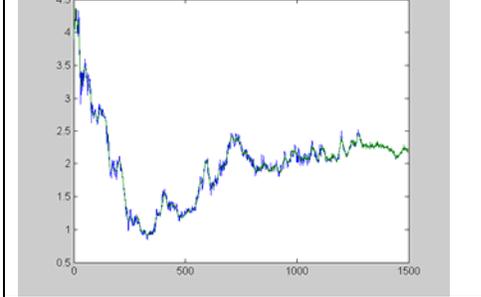
Gigante



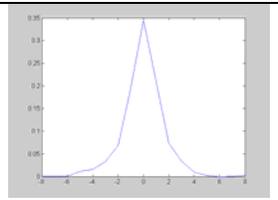
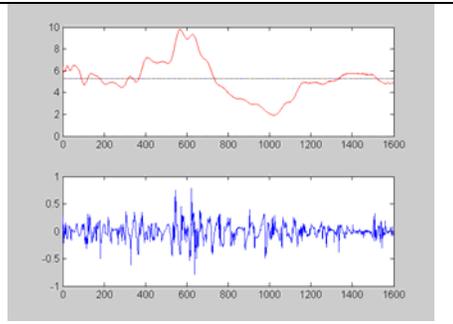
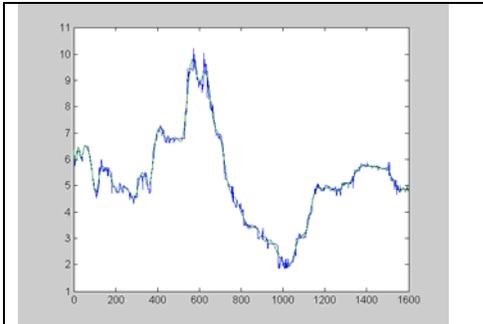
sab



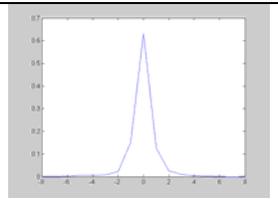
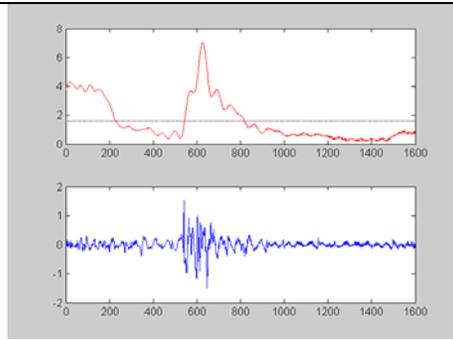
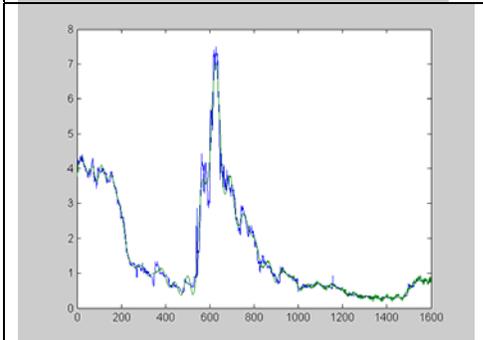
Electra



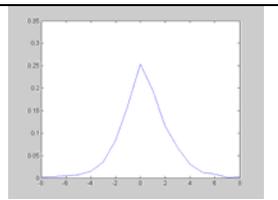
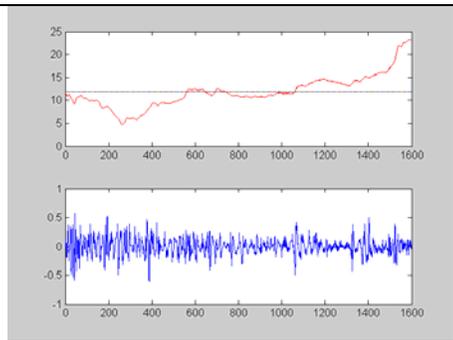
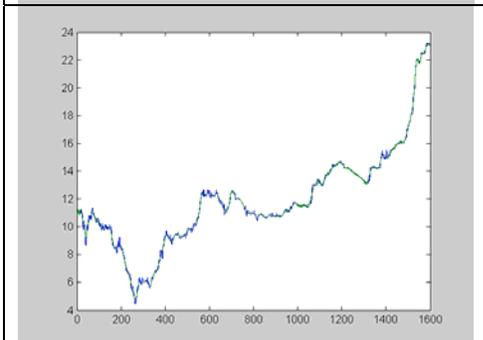
Gcorvi



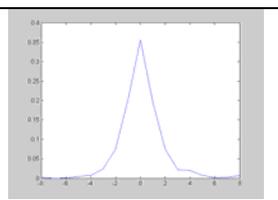
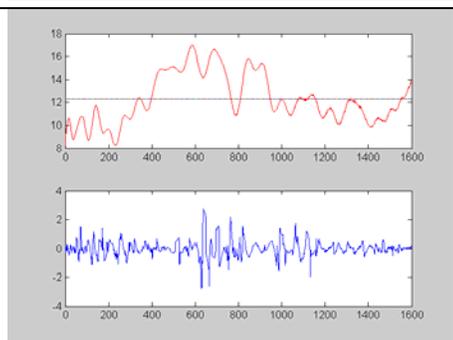
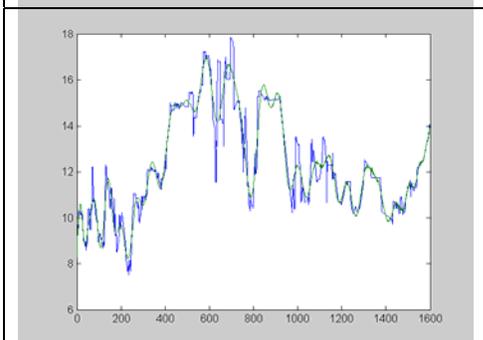
Nadrob



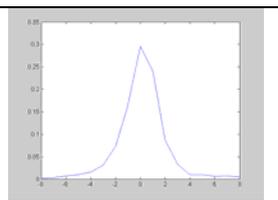
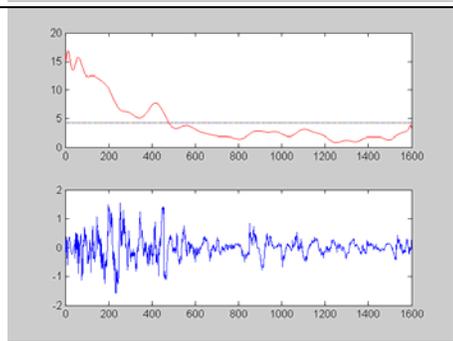
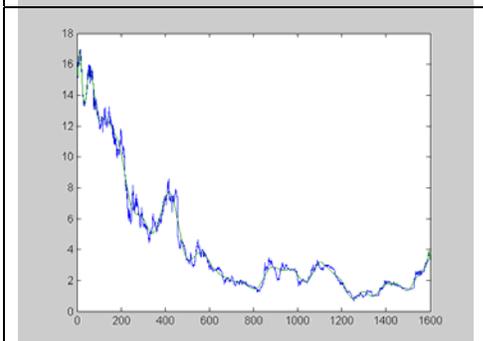
dataflx



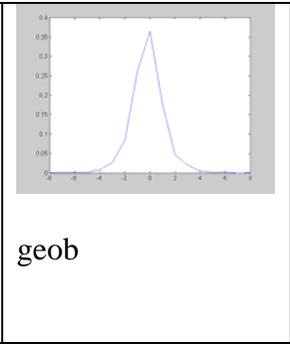
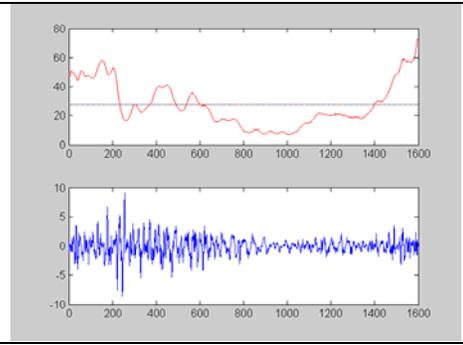
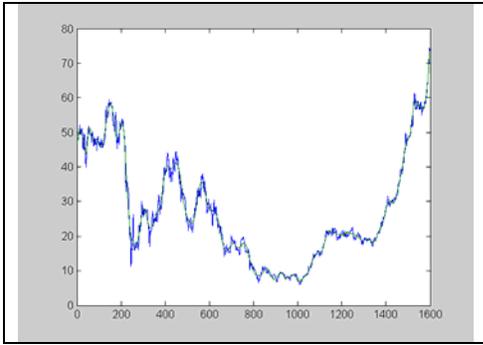
fragua



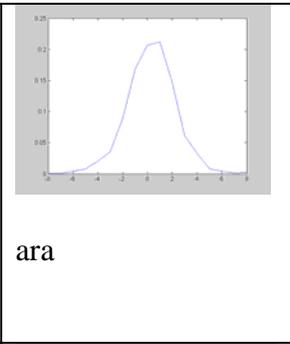
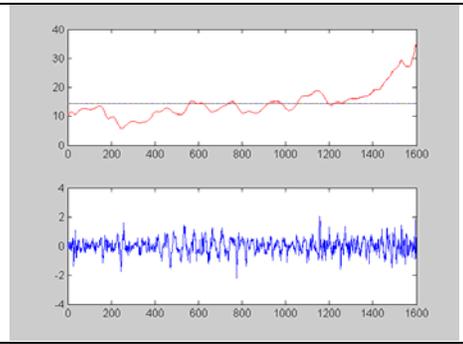
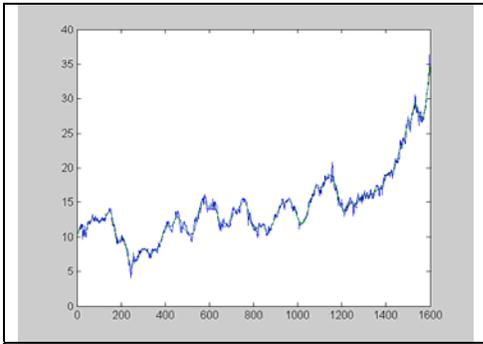
liverpol



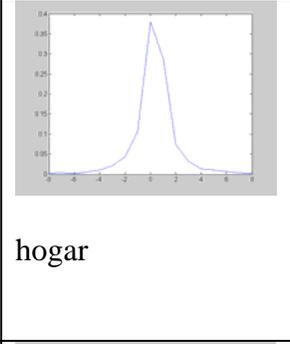
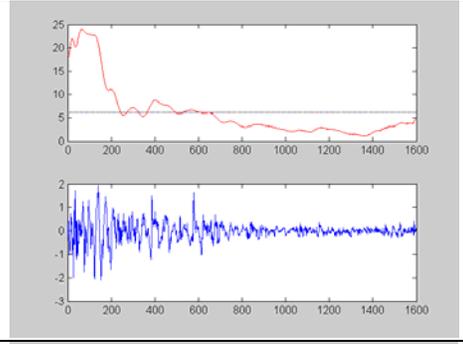
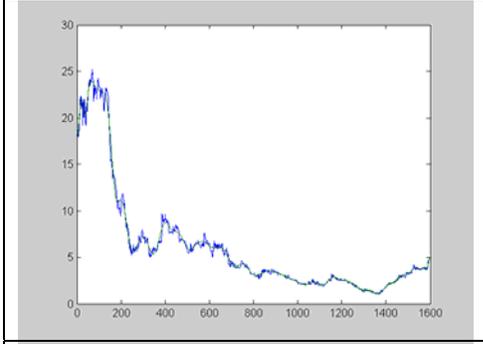
ICA



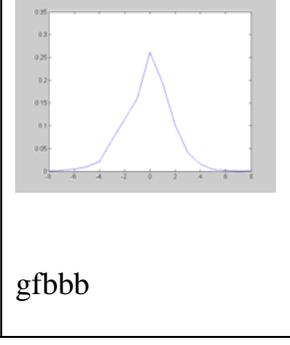
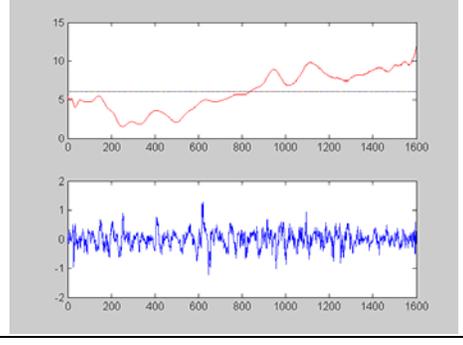
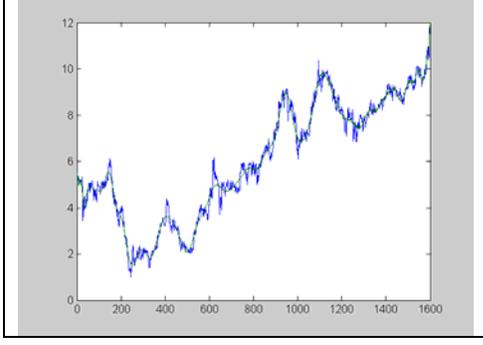
geob



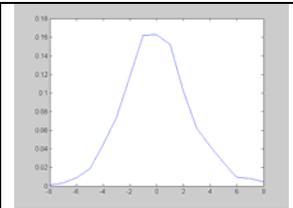
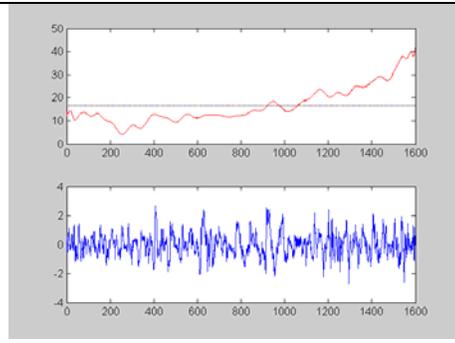
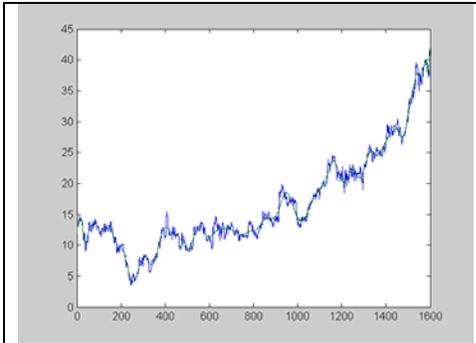
ara



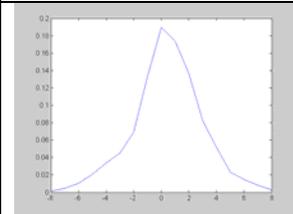
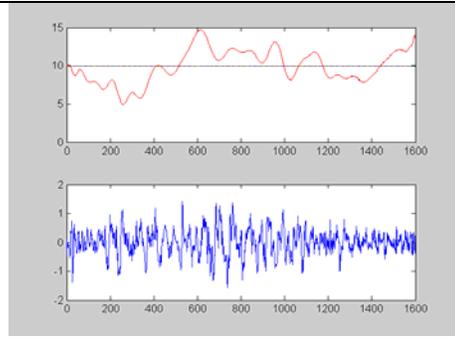
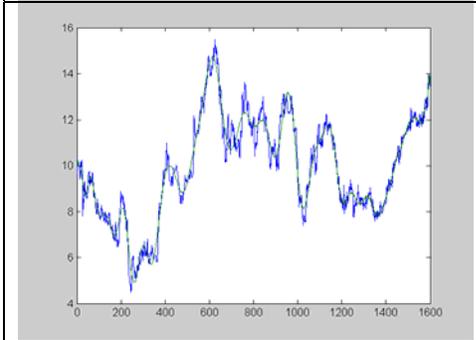
hogar



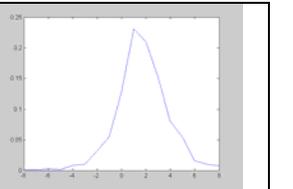
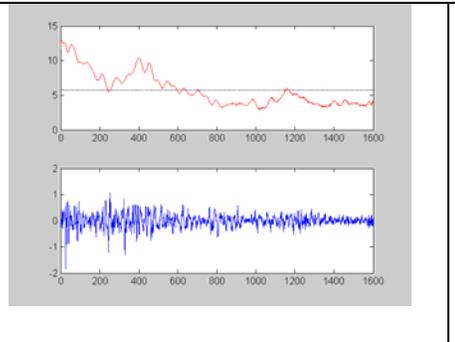
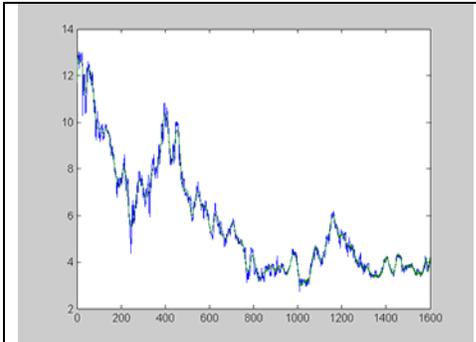
gfbbb



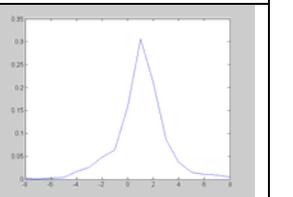
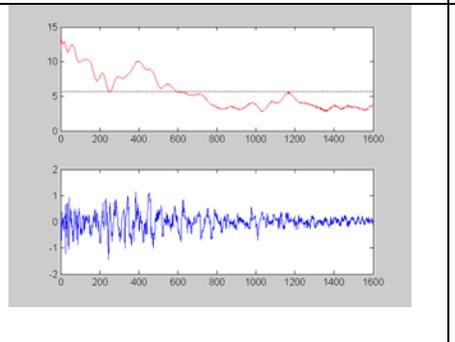
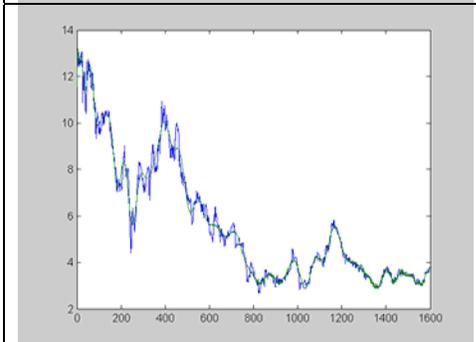
**gfnorte**



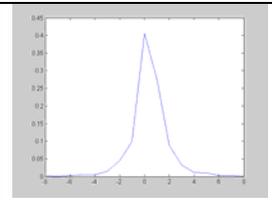
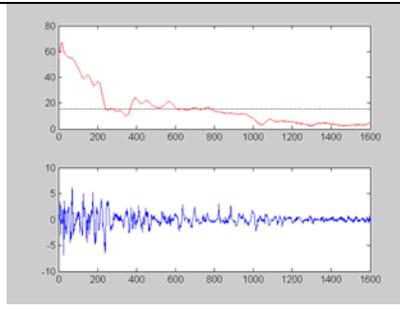
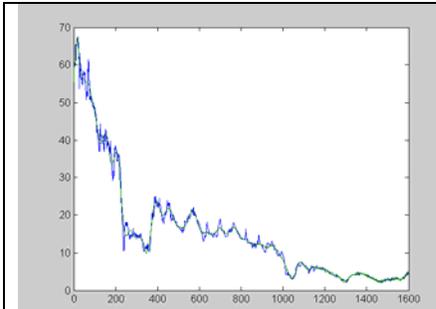
**gfimburo**



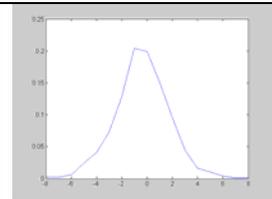
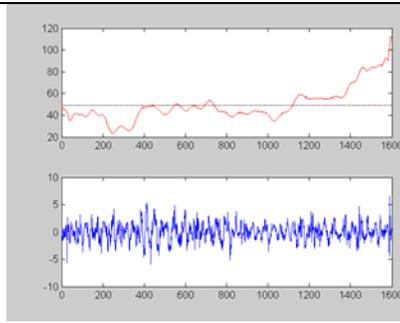
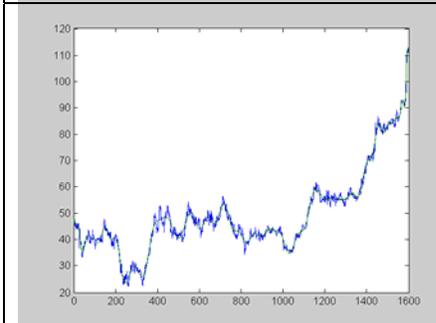
**desb**



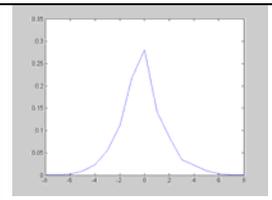
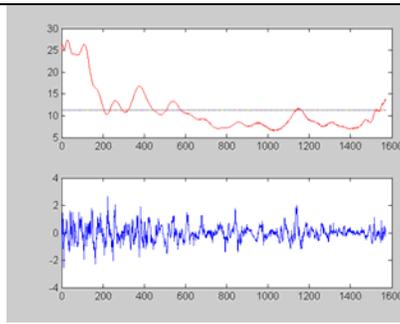
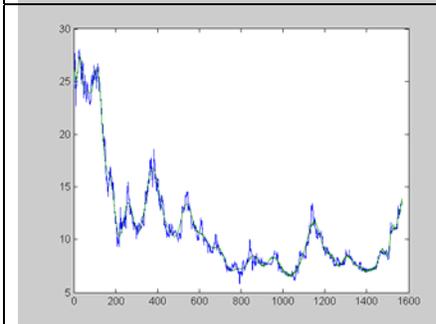
**desc**



sanluis

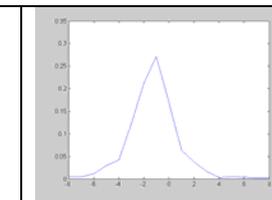
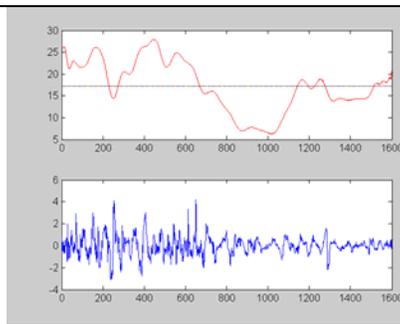
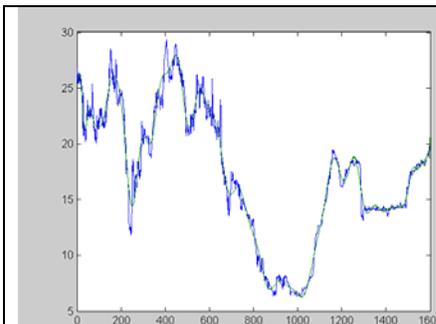


apasco

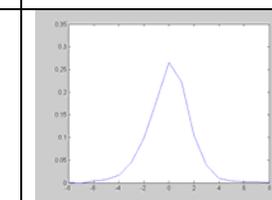
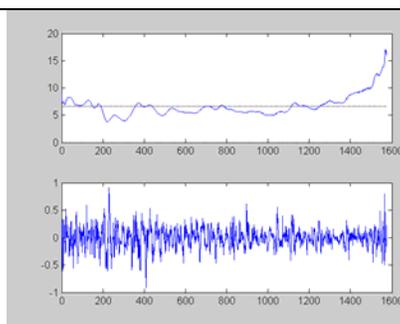
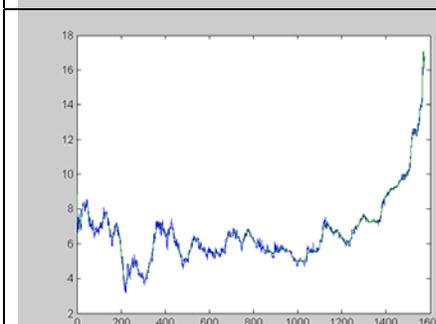


vitro

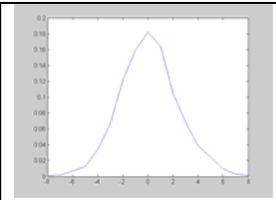
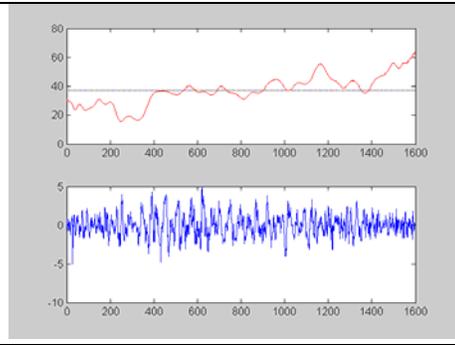
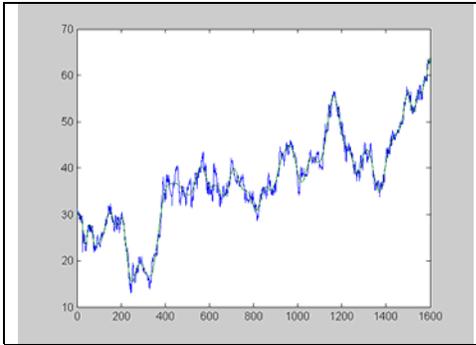
X



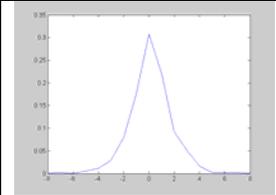
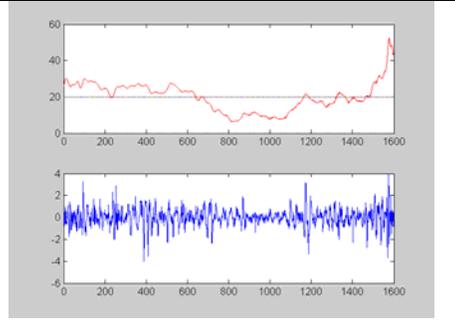
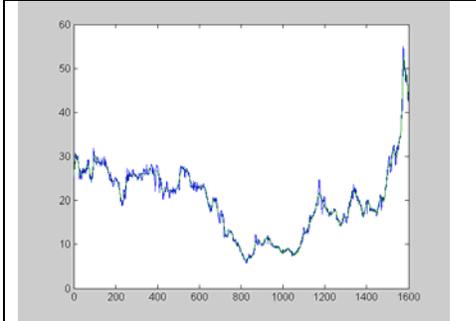
gissa



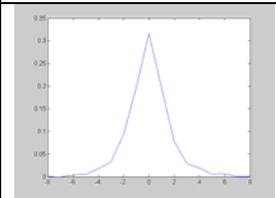
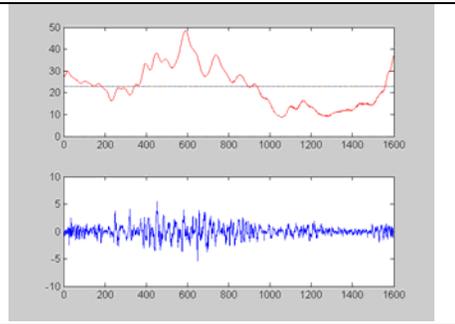
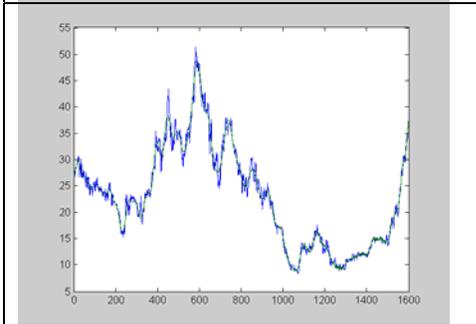
gcc



cemex

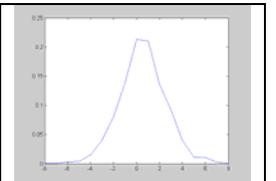
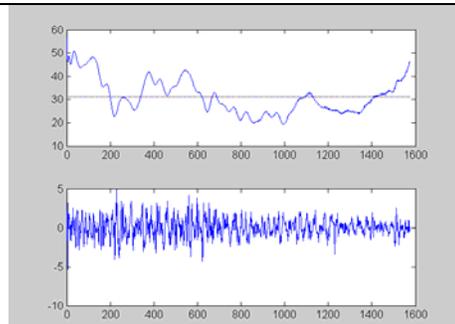
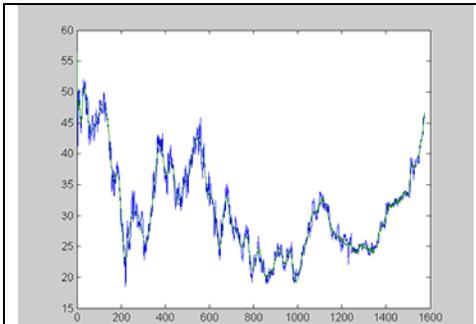


penoles

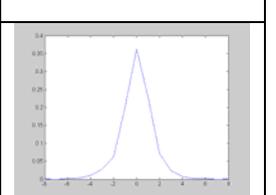
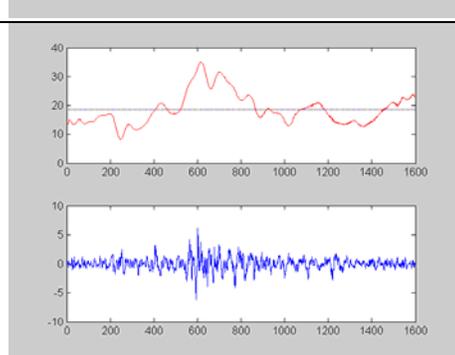
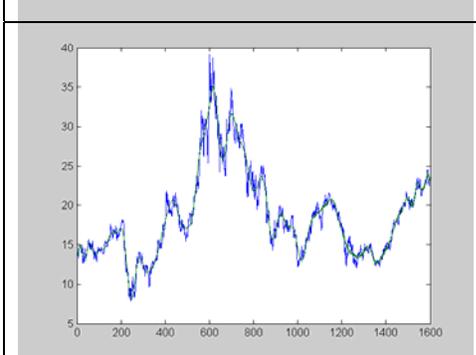


gmexico

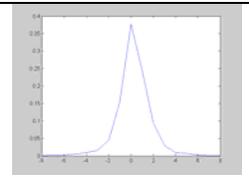
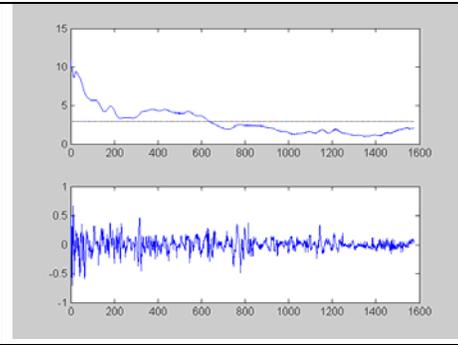
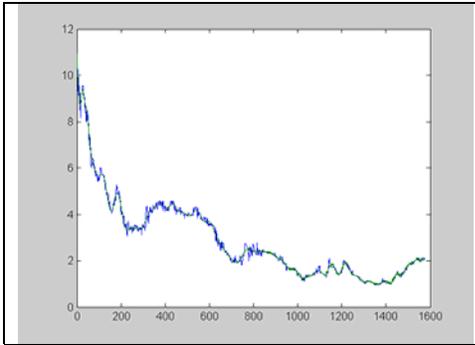
M



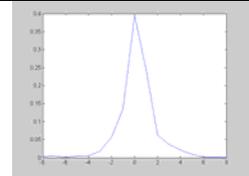
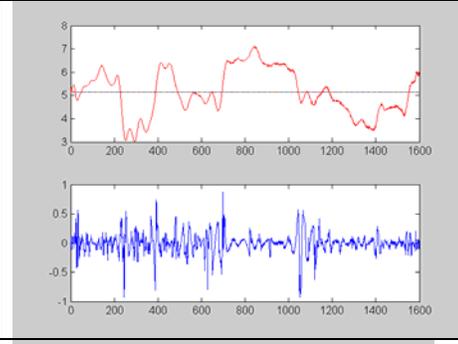
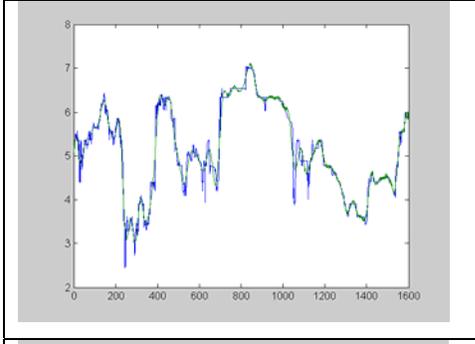
gcarso



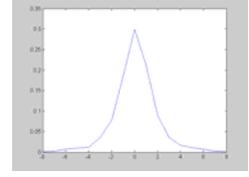
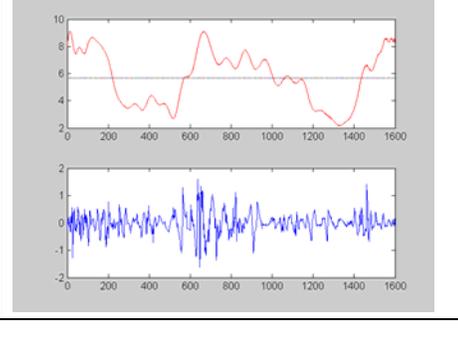
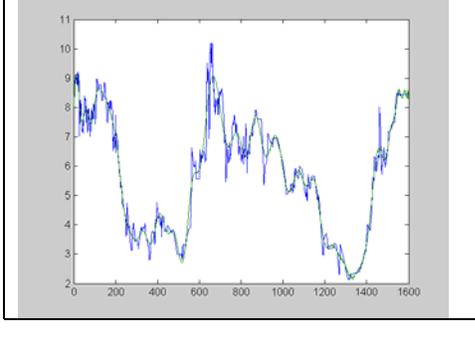
tlevisa



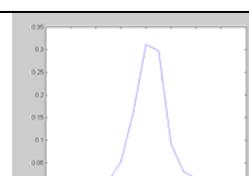
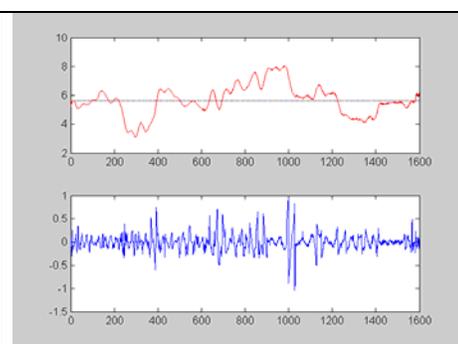
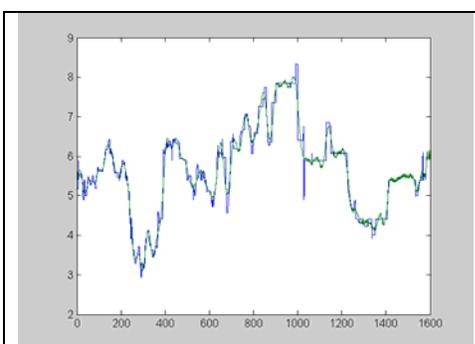
accel



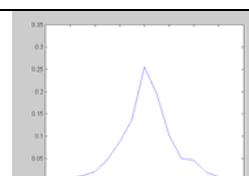
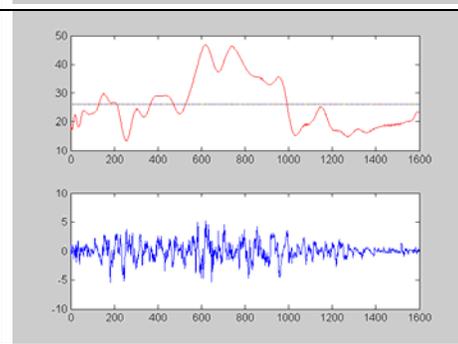
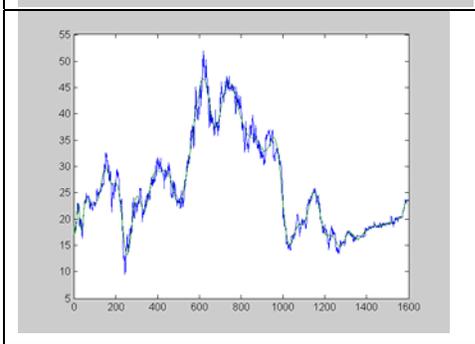
posadasl



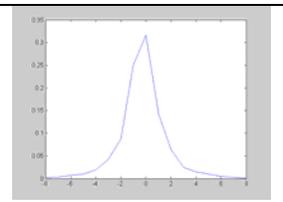
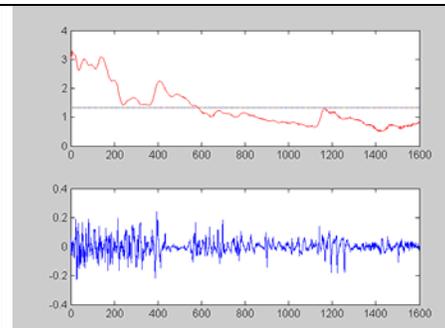
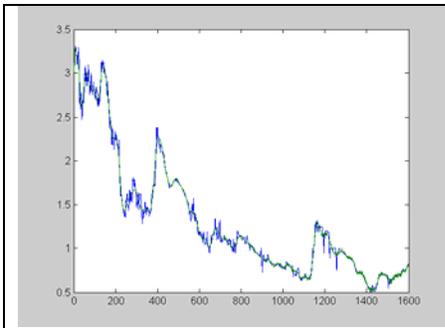
rcentro



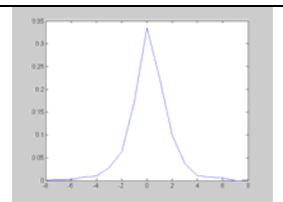
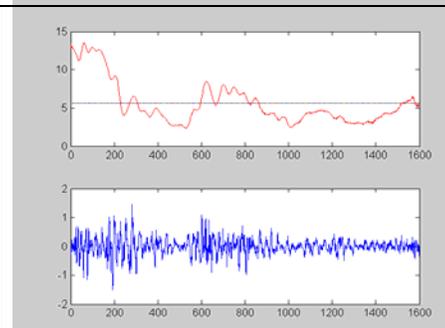
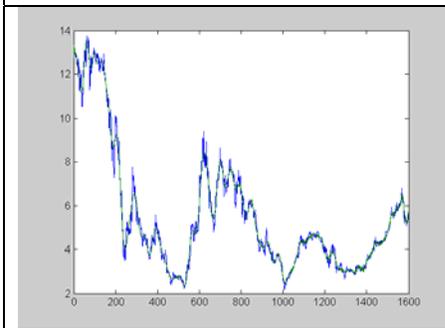
Posadasa



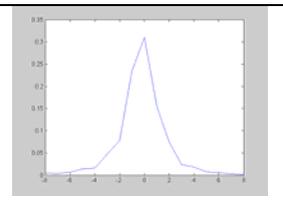
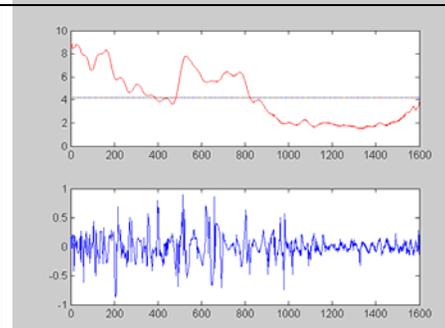
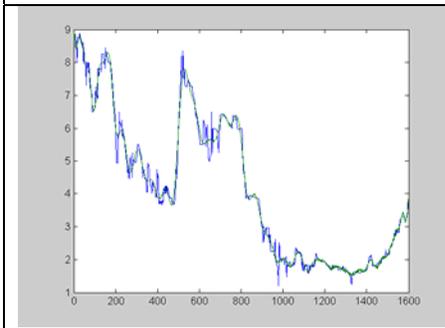
cieB



cmrb

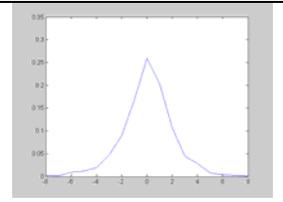
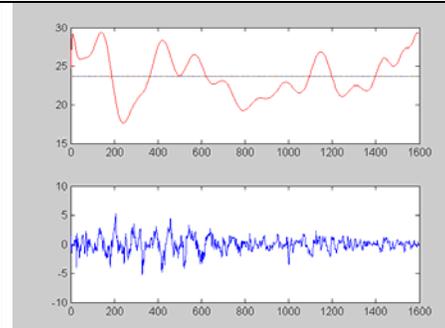
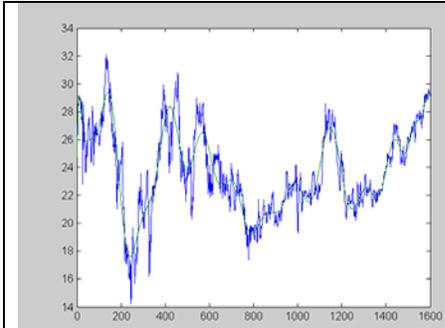


tvaztca

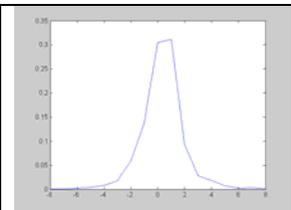
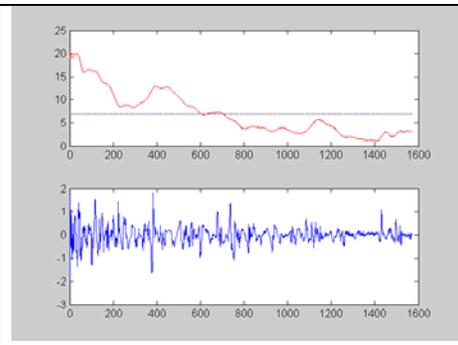
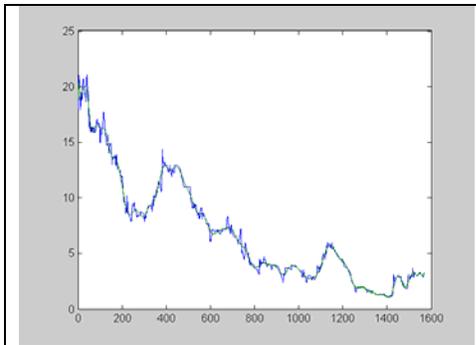


cidmega

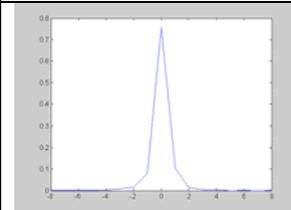
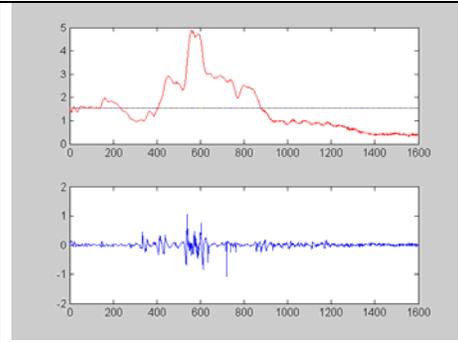
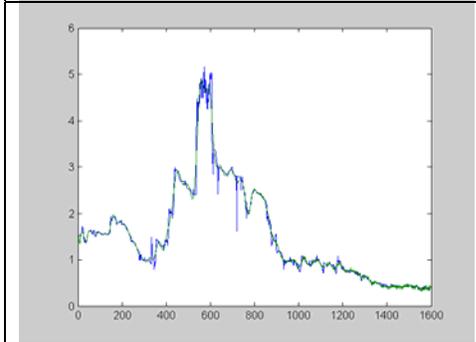
T



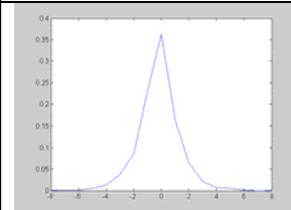
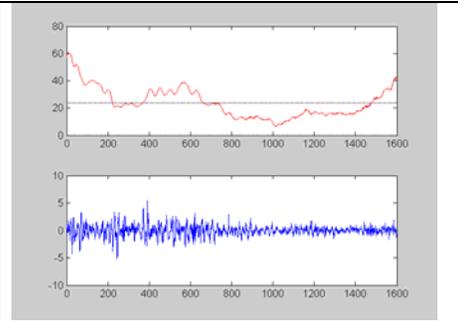
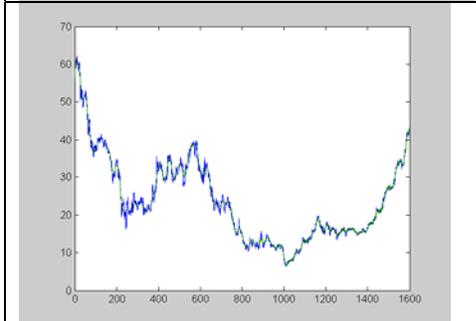
kimber



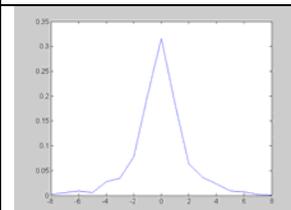
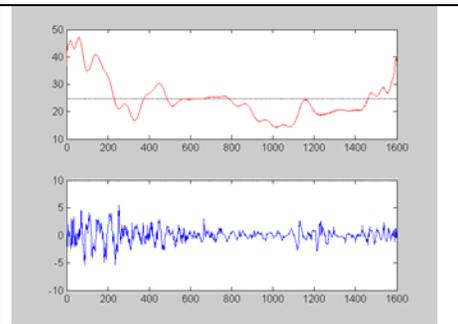
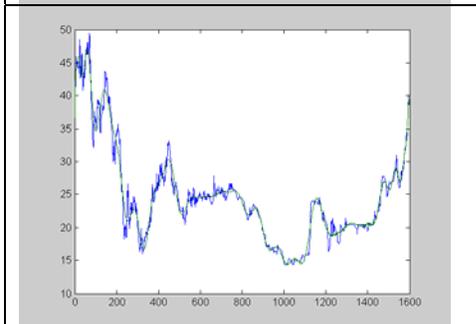
cydsa



tekchem

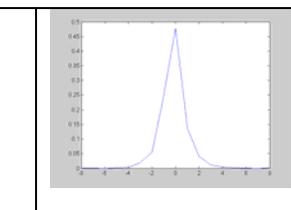
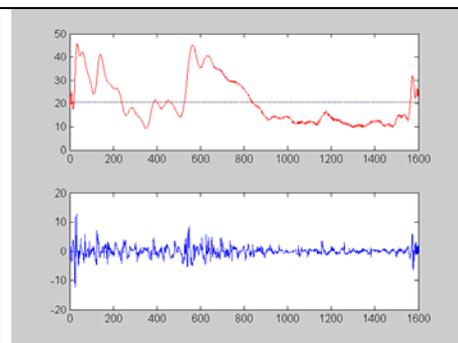
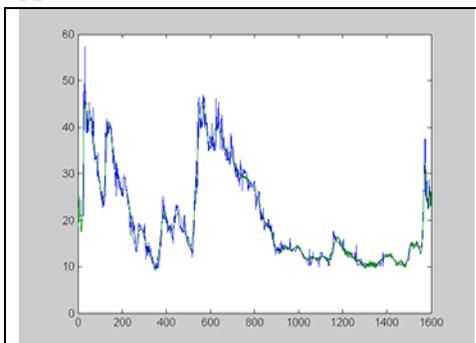


alfa

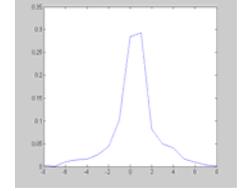
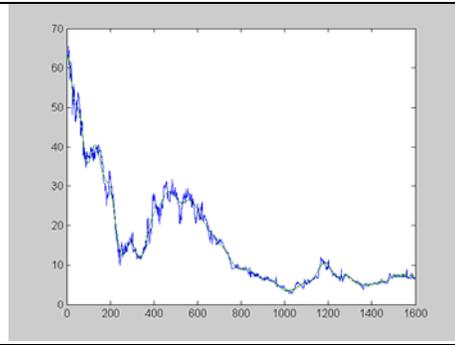
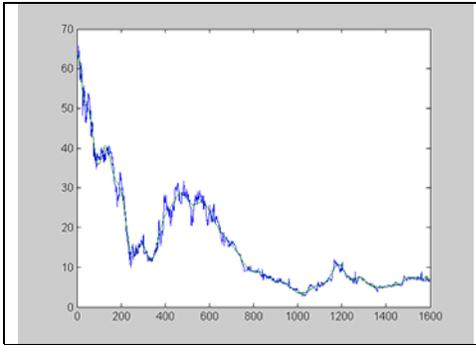


ichb

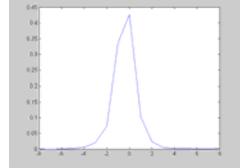
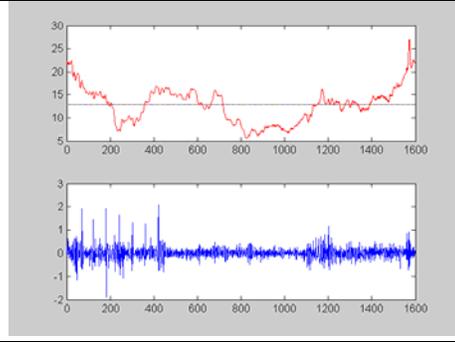
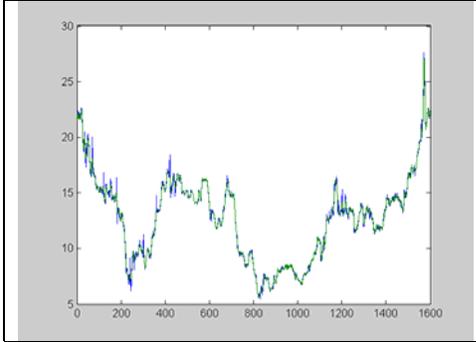
A



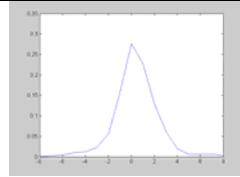
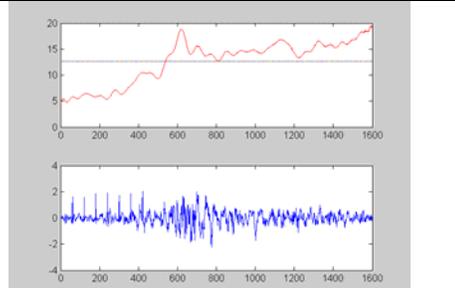
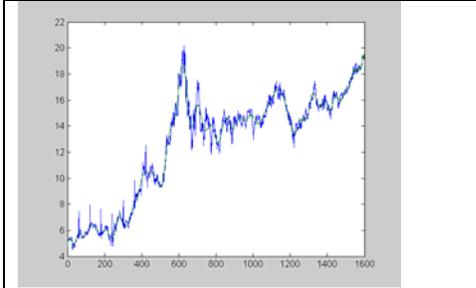
simecb



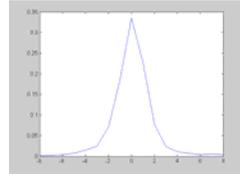
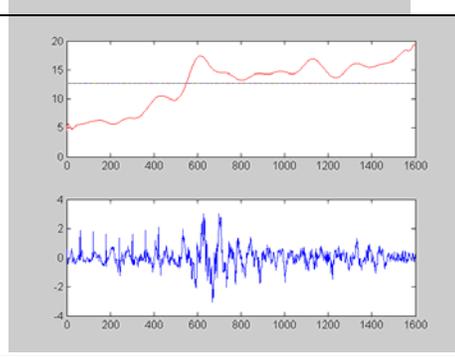
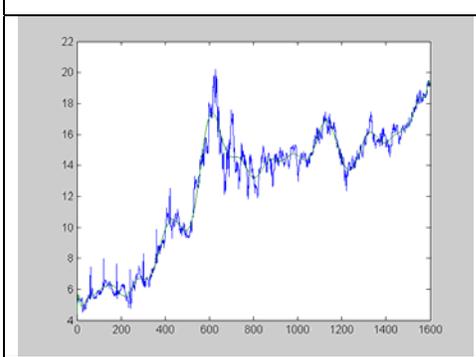
hylsa



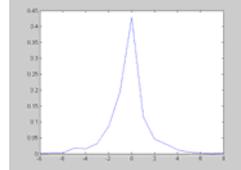
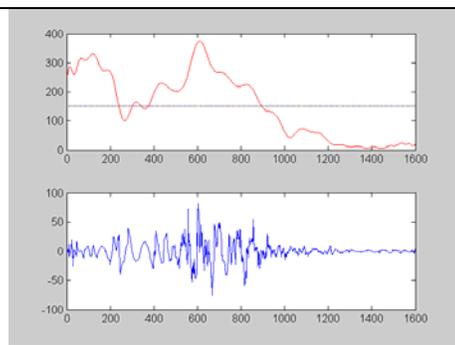
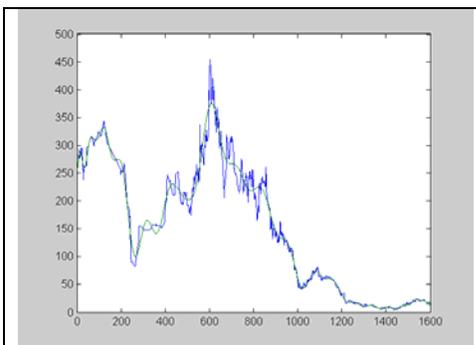
imsa



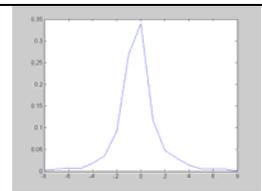
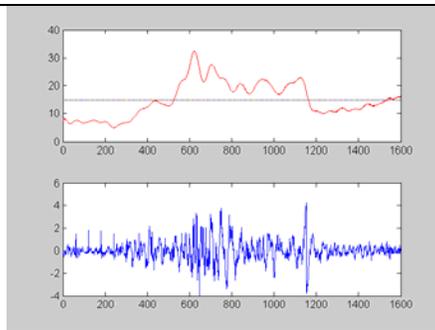
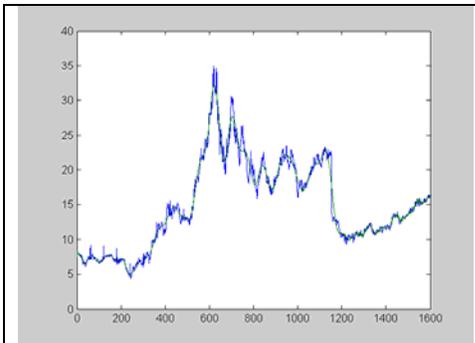
telmexa



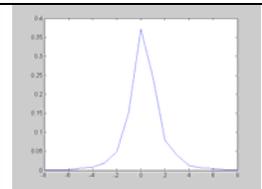
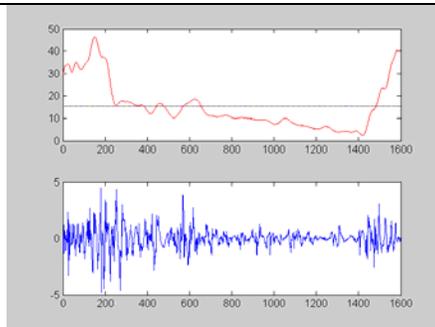
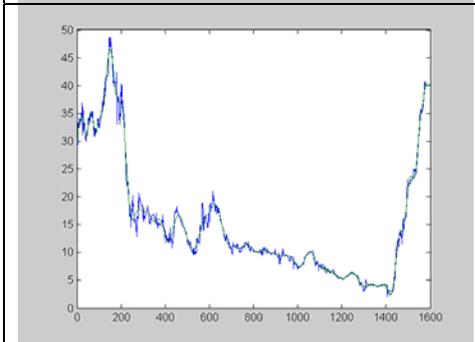
telmexl



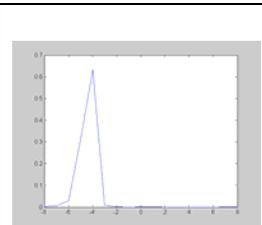
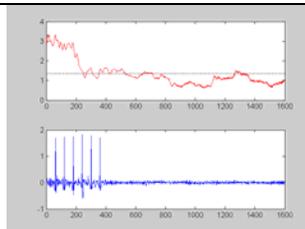
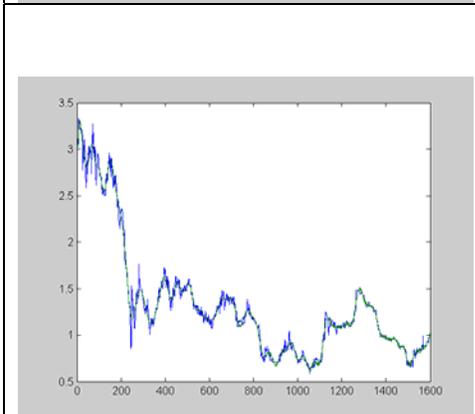
cel



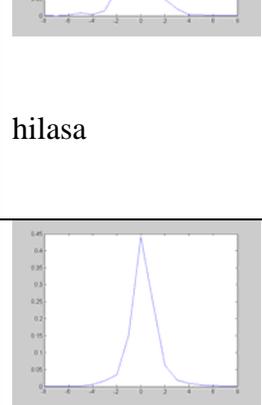
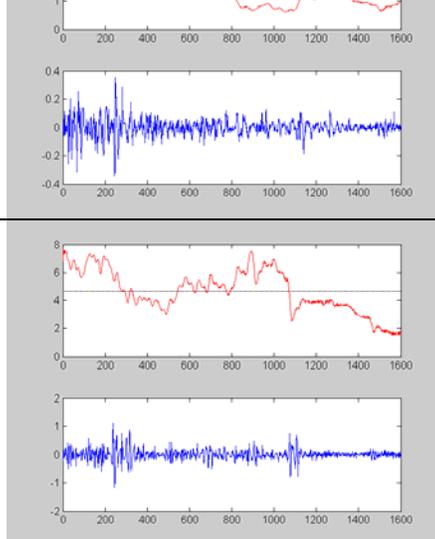
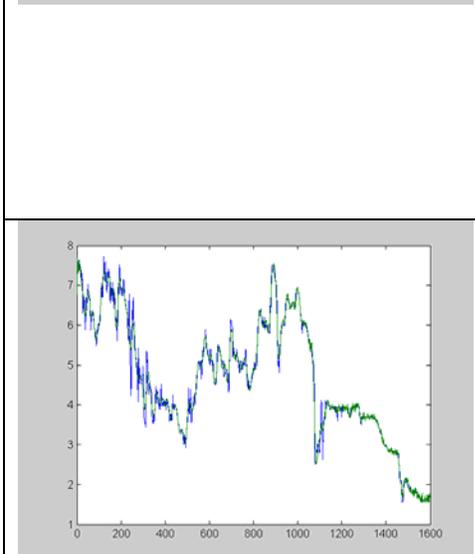
telecom



movil



hilasa



cintra

Solution for Minimal Spanning Tree Problem disC

	From Node	Connect To	Distance/Cost		From Node	Connect To	Distance/Cost
1	Node51	Node2	1.201	33	Node38	Node34	0.933
2	Node38	Node3	1.1346	34	Node38	Node35	1.0667
3	Node38	Node4	1.1646	35	Node38	Node36	1.1642
4	Node38	Node5	1.1487	36	Node38	Node37	0.9888
5	Node38	Node6	1.2023	37	Node47	Node38	0.9281
6	Node47	Node7	1.1276	38	Node14	Node39	1.2961
7	Node42	Node8	1.1544	39	Node41	Node40	1.1803
8	Node38	Node9	1.1979	40	Node38	Node41	0.9185
9	Node41	Node10	1.0467	41	Node38	Node42	0.9644
10	Node12	Node11	1.3425	42	Node38	Node43	1.2801
11	Node38	Node12	1.004	43	Node25	Node44	1.1921
12	Node1	Node13	1.3224	44	Node47	Node45	1.2875
13	Node13	Node14	0.5323	45	Node44	Node46	1.2172
14	Node13	Node15	0.9441	46	Node15	Node47	0.9745
15	Node38	Node16	1.2445	47	Node38	Node48	1.2535
16	Node16	Node17	1.3301	48	Node42	Node49	0.9371
17	Node38	Node18	0.9383	49	Node20	Node50	1.3181
18	Node28	Node19	1.2385	50	Node38	Node51	0.9796
19	Node59	Node20	1.3391	51	Node57	Node52	1.2972
20	Node49	Node21	1.2477	52	Node12	Node53	1.3406
21	Node41	Node22	1.2584	53	Node41	Node54	1.0238
22	Node16	Node23	1.295	54	Node38	Node55	1.1046
23	Node38	Node24	1.1269	55	Node57	Node56	1.1919
24	Node47	Node25	1.0538	56	Node60	Node57	1.1444
25	Node38	Node26	1.0683	57	Node60	Node58	1.0456
26	Node38	Node27	1.2026	58	Node60	Node59	0.5042
27	Node38	Node28	0.9295	59	Node42	Node60	1.076
28	Node28	Node29	0.9144	60	Node42	Node61	1.2459
29	Node47	Node30	1.0711	61	Node60	Node62	0.652
30	Node38	Node31	1.0948	62	Node60	Node63	1.2447
31	Node31	Node32	0.9983	63	Node60	Node64	0.8067
32	Node7	Node33	1.2526	64	Node64	Node65	0.975
	Total	Minimal	Connected	Distance	or Cost	=	70.66

Solution for Minimal Spanning Tree Problem potenciaC

	From Node	Connect To	Distance/Cost		From Node	Connect To	Distance/Cost
1	Node51	Node2	1.306	33	Node38	Node34	1.0729
2	Node38	Node3	1.2548	34	Node38	Node35	1.1975
3	Node38	Node4	1.2786	35	Node38	Node36	1.2783
4	Node38	Node5	1.2661	36	Node38	Node37	1.1266
5	Node38	Node6	1.307	37	Node47	Node38	1.0681
6	Node47	Node7	1.2491	38	Node14	Node39	1.3682
7	Node42	Node8	1.2706	39	Node41	Node40	1.2906
8	Node38	Node9	1.3037	40	Node38	Node41	1.0586
9	Node41	Node10	1.1798	41	Node38	Node42	1.1034
10	Node12	Node11	1.392	42	Node38	Node43	1.3588
11	Node38	Node12	1.1408	43	Node25	Node44	1.2994
12	Node1	Node13	1.3824	44	Node47	Node45	1.3632
13	Node13	Node14	0.64	45	Node44	Node46	1.3176
14	Node13	Node15	1.0838	46	Node15	Node47	1.113
15	Node38	Node16	1.3363	47	Node38	Node48	1.3422
16	Node16	Node17	1.3862	48	Node42	Node49	1.077
17	Node38	Node18	1.0781	49	Node20	Node50	1.3801
18	Node28	Node19	1.3323	50	Node38	Node51	1.1178
19	Node59	Node20	1.3905	51	Node57	Node52	1.3688
20	Node49	Node21	1.3384	52	Node12	Node53	1.3912
21	Node41	Node22	1.3454	53	Node41	Node54	1.159
22	Node16	Node23	1.3675	54	Node38	Node55	1.23
23	Node38	Node24	1.2486	55	Node57	Node56	1.2993
24	Node47	Node25	1.1861	56	Node60	Node57	1.2627
25	Node38	Node26	1.1988	57	Node60	Node58	1.1788
26	Node38	Node27	1.3072	58	Node60	Node59	0.6074
27	Node38	Node28	1.0695	59	Node42	Node60	1.2056
28	Node28	Node29	1.0546	60	Node42	Node61	1.3372
29	Node47	Node30	1.2013	61	Node60	Node62	0.7762
30	Node38	Node31	1.2217	62	Node60	Node63	1.3364
31	Node31	Node32	1.1355	63	Node60	Node64	0.9444
32	Node7	Node33	1.3416	64	Node64	Node65	1.1135
	Total	Minimal	Connected	Distance	or Cost	=	77.74

Solution for Minimal Spanning Tree Problem  
fourierC

	From Node	Connect To	Distance/Cost		From Node	Connect To	Distance/Cost
1	Node12	Node2	1.2881	33	Node38	Node34	0.7919
2	Node4	Node3	1.0897	34	Node37	Node35	1.1893
3	Node1	Node4	1.3545	35	Node34	Node36	1.1886
4	Node47	Node5	1.2868	36	Node38	Node37	0.9763
5	Node18	Node6	1.1349	37	Node4	Node38	1.2291
6	Node34	Node7	0.9841	38	Node40	Node39	1.2882
7	Node38	Node8	0.9156	39	Node51	Node40	1.2606
8	Node6	Node9	1.2410	40	Node38	Node41	1.0417
9	Node51	Node10	0.8958	41	Node38	Node42	1.0292
10	Node54	Node11	1.3173	42	Node18	Node43	1.3286
11	Node47	Node12	1.0066	43	Node38	Node44	1.1820
12	Node12	Node13	0.9596	44	Node51	Node45	1.1640
13	Node13	Node14	0.3257	45	Node44	Node46	1.0473
14	Node38	Node15	1.1965	46	Node28	Node47	0.8113
15	Node48	Node16	1.0255	47	Node49	Node48	1.1309
16	Node59	Node17	1.2971	48	Node8	Node49	1.1608
17	Node47	Node18	0.8439	49	Node7	Node50	1.1337
18	Node28	Node19	1.2452	50	Node60	Node51	0.7842
19	Node26	Node20	1.2225	51	Node55	Node52	1.2541
20	Node49	Node21	1.3350	52	Node54	Node53	1.2339
21	Node29	Node22	1.0804	53	Node22	Node54	1.1688
22	Node51	Node23	1.0958	54	Node38	Node55	1.0115
23	Node51	Node24	1.0911	55	Node20	Node56	1.2934
24	Node26	Node25	0.9399	56	Node38	Node57	1.1840
25	Node12	Node26	0.9284	57	Node38	Node58	1.2467
26	Node30	Node27	1.2333	58	Node62	Node59	0.9781
27	Node38	Node28	0.8914	59	Node59	Node60	0.9890
28	Node28	Node29	0.5431	60	Node30	Node61	1.1012
29	Node47	Node30	0.8813	61	Node8	Node62	0.9319
30	Node25	Node31	1.2381	62	Node34	Node63	1.1654
31	Node31	Node32	1.1377	63	Node34	Node64	1.2524
32	Node8	Node33	1.0934	64	Node53	Node65	1.3392

Total Minimal Connected Distance or Cost = 70.01

