



Universidad de Valladolid

**Facultad de Ciencias
Económicas y Empresariales
Trabajo de Fin de Grado**

Grado en Administración y Dirección de Empresas

Medidas de dependencia alternativas al coeficiente de correlación lineal de Pearson. Aplicación a los mercados financieros I

Presentado por:

Álvaro García Álvarez

Tutelado por:

Ana Pérez Espartero

RESUMEN

Este trabajo analiza diferentes alternativas al coeficiente de correlación lineal de Pearson para cuantificar el grado de dependencia entre la rentabilidad de distintos mercados financieros. En primer lugar, definimos dicho coeficiente, así como las alternativas presentadas, basadas en rangos y/o en modificaciones robustas del coeficiente de correlación de Pearson. En la aplicación empírica, contrastamos la ausencia de normalidad de las distribuciones marginales de los rendimientos de 4 índices bursátiles y comentamos que consecuencias tienen los resultados sobre el análisis de la dependencia. Por último, aplicamos todos los coeficientes descritos a dichos rendimientos y discutimos si las medidas alternativas proporcionan una medida más precisa de la dependencia entre los rendimientos de los índices bursátiles analizados.

Palabras clave: Rendimientos bursátiles, Correlación, Mercados financieros.

ABSTRACT

This paper analyzes different alternatives to Pearson's correlation coefficient to quantify the degree of dependence between the returns of different stock markets. First, we define such coefficient as well as the alternatives based on ranks and/or based on the robust modifications of Pearson's correlation coefficient. In the empirical application, we test the absence of normality of the marginal distributions of the returns of 4 stock market indices and we comment on what consequences the results have on the dependency analysis. Finally, we apply all the coefficients described to these returns and discuss whether the alternative measures provide a more accurate measure of the dependence between the returns of the analyzed stock indices or not.

Keywords: Stock returns, Correlation, Stock market.

ÍNDICE

1. INTRODUCCIÓN.....	5
2. METODOLOGÍA.....	7
2.1 Coeficiente de correlación lineal de Pearson	7
2.2 Coeficiente de asociación Tau de Kendall	10
2.3 Coeficiente de correlación por rangos de Spearman	12
2.4 Medidas robustas de correlación	14
2.5 Medidas de dependencia basadas en datos agrupados	17
3. APLICACIÓN EMPÍRICA	18
3.1 Descripción de los datos	18
3.2 Análisis de la distribución de los rendimientos bursátiles.....	20
3.2.1 Análisis descriptivo	21
3.2.2 Contrastes de normalidad.....	23
3.3 Medidas de dependencia entre los mercados financieros.....	27
3.4 Análisis de dependencia con datos agrupados por quintiles.....	32
4. CONCLUSIONES.....	39
5. BIBLIOGRAFÍA	41
6. ANEXO.....	42

ÍNDICE DE GRÁFICOS

Gráfico 1. Diagramas de dispersión de distintas variables junto a su correspondiente coeficiente de correlación lineal de Pearson.....	9
Gráfico 2. Histogramas de las rentabilidades de los cuatro índices bursátiles junto a una distribución Normal de misma media y desviación típica	27
Gráfico 3. Evolución temporal de los rendimientos diarios de los índices bursátiles.....	28

Gráfico 4. Diagramas de dispersión de las 6 posibles combinaciones de los 4 índices bursátiles.....	29
---	----

ÍNDICE DE TABLAS

Tabla 1. Distribución conjunta por quintiles de las variables X e Y tanto en frecuencia absoluta como en relativa (entre paréntesis).	18
Tabla 2. Estadísticos descriptivos básicos de los rendimientos de los índices bursátiles.	22
Tabla 3. Valores de los estadísticos del contraste de Normalidad para las rentabilidades de los distintos índices bursátiles junto a su respectivo p-valor (entre paréntesis).	26
Tabla 4. Valores de los distintos coeficientes de asociación y correlación para las distintas parejas de índices bursátiles.	30
Tabla 5. Distribución conjunta por quintiles del NASDAQ100 y el IBEX35 de frecuencias absolutas y relativas (entre paréntesis).....	34
Tabla 6. Distribución conjunta por quintiles del NASDAQ100 y el IEUROSTOXX50 de frecuencias absolutas y relativas (entre paréntesis).....	34
Tabla 7. Distribución conjunta por quintiles del NASDAQ100 y el NIKKEI225 de frecuencias absolutas y relativas (entre paréntesis).....	35
Tabla 8. Distribución conjunta por quintiles del IBEX35 y el EUROSTOXX50 de frecuencias absolutas y relativas (entre paréntesis).....	35
Tabla 9. Distribución conjunta por quintiles del IBEX35 y el NIKKEI225 de frecuencias absolutas y relativas (entre paréntesis).....	36
Tabla 10. Distribución conjunta por quintiles del EUROSTOXX50 y el NIKKEI225 de frecuencias absolutas y relativas (entre paréntesis).....	36

1. INTRODUCCIÓN

La medición de las dependencias existentes entre diversos mercados financieros mundiales es un tema de especial relevancia desde el punto de vista de los inversores debido a que estos buscarán reducir el riesgo de las inversiones que hacen. Para ello sería interesante conocer las relaciones existentes entre los distintos mercados para poder diversificar el riesgo de sus activos. Por ejemplo, invertir en mercados cuyos movimientos se comporten de manera opuesta ante un mismo evento de modo que uno de ellos aumente cuando otro disminuya, contribuiría a minimizar el riesgo.

Un error común a la hora de analizar las dependencias entre los mercados financieros es suponer que las rentabilidades de los activos que cotizan en esos mercados siguen distribuciones gaussianas y utilizar el coeficiente de correlación de Pearson como medida de la posible relación entre dichos mercados. Sin embargo, los datos sobre las fluctuaciones en los mercados financieros no siguen este tipo de distribución, sino que presentan diferentes patrones y características que los alejan de una distribución puramente gaussiana. Una de estas características es la presencia de “demasiados” datos en las colas de la distribución (“fat tails”) en comparación con lo que cabría esperar en una distribución gaussiana, es decir, la existencia de una cantidad inusualmente grande de datos correspondientes a grandes crecimientos o grandes pérdidas en estos mercados. Por ejemplo, caídas muy importantes en las bolsas como las derivadas de la quiebra de Lehman Brothers o la crisis más reciente derivada de la pandemia que ha supuesto pérdidas importantes. Además, se observan otras diferencias con respecto a la distribución normal como una ligera asimetría en la distribución, pero estas son de menor relevancia que las “fat tails” comentadas previamente (Cont, 2001; Aparicio y Estrada, 2001).

La importancia de comprobar si las rentabilidades de los activos financieros siguen o no una distribución gaussiana radica en el hecho de que el coeficiente de correlación lineal de Pearson, uno de los más utilizados para medir la relación entre dos variables, no es un buen indicador de dependencia cuando dichas variables no siguen distribuciones gaussianas (Embrechts, McNeil, y Straumann, 2002 y Drouet y Kotz, 2001).

Para poder realizar un análisis más adecuado de las relaciones no necesariamente lineales entre dos variables existen diferentes alternativas disponibles. La primera de ellas es la utilización de los rangos en lugar de valores concretos de las variables como ocurre con el coeficiente de correlación por rangos de Spearman. También existe el coeficiente Tau de Kendall cuyo análisis se basa en las probabilidades de concordancia entre las distintas observaciones de las variables. Adicionalmente, existen diferentes coeficientes de correlación “robustos” especialmente diseñados para evitar el efecto adverso de la presencia de datos atípicos en las rentabilidades de los mercados financieros que está relacionada, en cierta forma, con las “fat-tails” comentadas previamente.

El objetivo de este trabajo es doble. En primer lugar, se quiere contrastar la normalidad de las distribuciones de los rendimientos de los índices bursátiles de varios mercados financieros, y después, se pretende comparar los valores de distintos coeficientes de asociación alternativos con el de correlación lineal de Pearson en busca de posibles diferencias.

La principal conclusión que se extrae de este estudio es que los rendimientos de los índices bursátiles no siguen distribuciones Normales y que el coeficiente de correlación lineal de Pearson devuelve valores muy distintos a los otros coeficientes estudiados. Además, observamos que el mayor grado de dependencia se produce en las colas de la distribución, es decir, entre los valores extremadamente grandes y los extremadamente pequeños.

La estructura del trabajo es la siguiente. En el capítulo 2 comenzaré revisando el coeficiente de correlación lineal de Pearson y sus ventajas e inconvenientes; después, describiré algunos de los diferentes coeficientes alternativos que existen y realizaré un pequeño análisis de sus propiedades, ventajas e inconvenientes. En el capítulo 3, realizaré un análisis empírico de la relación entre la rentabilidad de los cuatro índices bursátiles representativos de cuatro mercados financieros (España, USA, Asia, Europa). En concreto, analizaré las rentabilidades diarias de los índices: IBEX35 (España), NASDAQ100 (USA), NIKKEI225 (Japón) y EUROSTOXX50 (Europa) en el periodo comprendido entre 2010 y 2020. En primer lugar, contrastaré la no gaussianidad de estos datos para constatar el pilar principal sobre el que se asienta la necesidad de utilizar otros

indicadores distintos del coeficiente de correlación lineal de Pearson. En segundo lugar, calcularé cada uno de los coeficientes previamente mencionados para los datos diarios del periodo 2010-2020 analizando la consistencia de los resultados obtenidos con los distintos indicadores utilizados. Por último, discutiré otra forma alternativa de visualizar la dependencia entre estos índices usando los datos por quintiles. Para cerrar el trabajo, en el capítulo 4, resumiré las principales conclusiones.

2. METODOLOGÍA

En esta sección vamos a describir una serie de coeficientes que permiten medir la dependencia entre dos variables para poder, después, aplicarlos a los datos de los índices bursátiles. Para ello analizaremos sus expresiones muestrales y describiremos sus principales propiedades. Partimos de T observaciones $(x_1, y_1), (x_2, y_2) \dots (x_T, y_T)$ de una variable aleatoria bidimensional (X, Y) donde X denotará, en nuestro caso, la rentabilidad diaria de un índice bursátil e Y la rentabilidad diaria de otro índice bursátil.

2.1 Coeficiente de correlación lineal de Pearson

En primer lugar, vamos a explicar los conceptos de dependencia y correlación ya que son un punto fundamental de este trabajo.

Cuando hablamos de dependencia nos referimos a lo opuesto a independencia y decimos que dos variables son independientes cuando el hecho de que una ocurra no altera en absoluto las probabilidades de que la otra ocurra. En este caso, la distribución conjunta de ambas será el producto de las distribuciones marginales de cada una de ellas. Si esto último no se cumple, entonces las variables son dependientes.

Ahora bien, la cuestión es cómo cuantificar el grado de dependencia entre las variables. A lo largo de la historia han sido muchos los autores que han intentado cuantificar la relación de dependencia mediante un coeficiente matemático. Entre

ellos merece especial atención Karl Pearson que, en 1896, describe, por primera vez, el coeficiente de correlación lineal de Pearson. El autor justifica su uso en la sencillez de su cálculo y su capacidad de predecir la dependencia lineal entre sucesos. Por tanto, este coeficiente permite cuantificar el grado de correlación lineal entre las variables; cuando este coeficiente toma el valor cero decimos que las variables están incorrelacionadas (Drouet y Kotz, 2001).

Partiendo de dos variables aleatorias X e Y, como las presentadas previamente, este coeficiente se define como:

$$r_{XY} = \frac{S_{XY}}{S_X * S_Y} \quad (1)$$

donde S_{XY} es la covarianza entre X e Y y S_X y S_Y son las desviaciones típicas de X e Y, respectivamente, definidas como:

$$S_{xy} = \frac{\sum_{i=1}^T (x_i - \bar{x})(y_i - \bar{y})}{T}, \quad (2)$$

$$S_X = \sqrt{\frac{\sum_{i=1}^T (x_i - \bar{x})^2}{T}}, \quad (3)$$

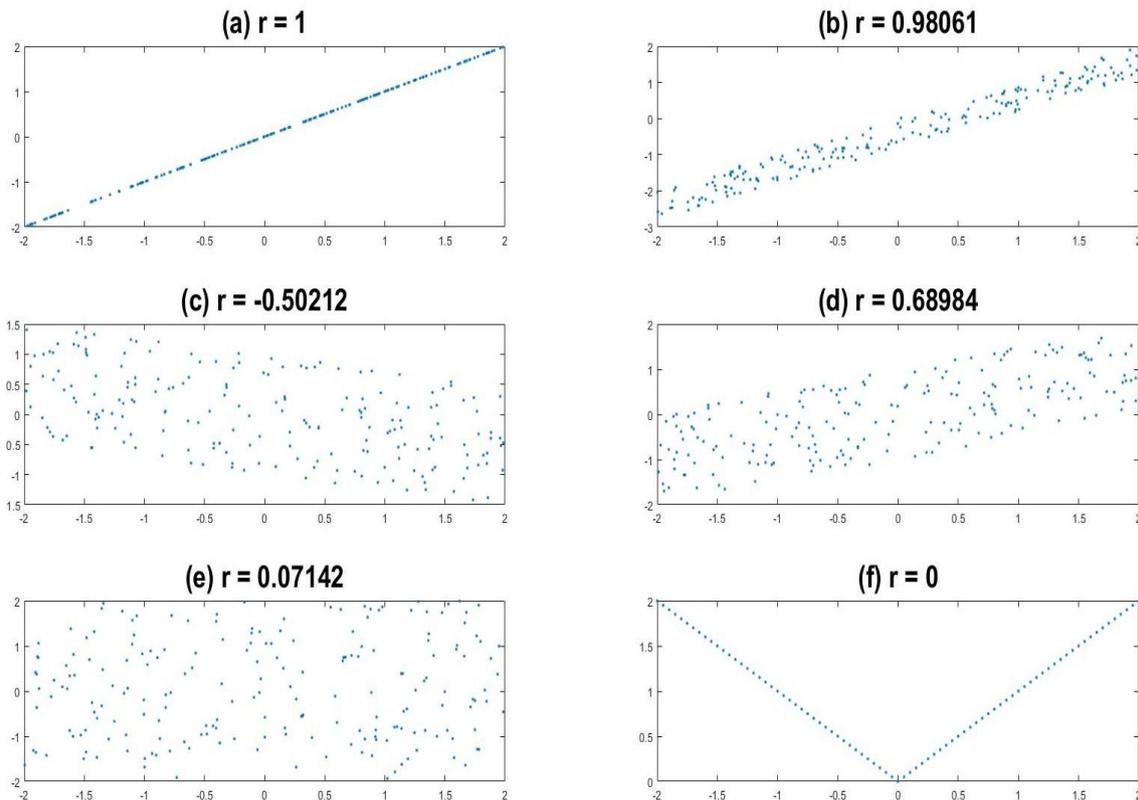
$$S_Y = \sqrt{\frac{\sum_{i=1}^T (y_i - \bar{y})^2}{T}}, \quad (4)$$

siendo \bar{x} e \bar{y} las medias muestrales de X e Y, respectivamente.

Este coeficiente puede tomar todos los valores entre -1 y 1. Si obtenemos un valor de 1, ambas variables tienen una relación lineal exacta positiva, es decir, cuando una aumenta la otra lo hace proporcionalmente y conforme a una relación lineal exacta. Por el contrario, un coeficiente de -1 muestra una correlación totalmente negativa donde el crecimiento de una de las variables provoca un descenso en la otra conforme a una relación lineal exacta de pendiente negativa (Peña, 2014). Si las variables son independientes entonces $r(x,y)$ es igual a 0, indicando ausencia de relación lineal entre las variables. Sin embargo, el hecho de que dos variables no estén correlacionadas no implica que sean independientes.

En el gráfico 1 observamos diferentes nubes de puntos obtenidas a partir de variables con distinto grado de dependencia acompañadas del coeficiente de correlación lineal de Pearson.

Gráfico 1. Diagramas de dispersión de distintas variables junto a su correspondiente coeficiente de correlación lineal de Pearson



En el ejemplo del gráfico 1 (a) podemos observar que la dependencia es lineal y exacta formando una línea perfecta como habíamos comentado en la interpretación del coeficiente, que en este caso toma el valor de 1. Si observamos los ejemplos (b) y (d) observamos que ambas nubes de puntos se sitúan en torno a la diagonal que parte de la esquina inferior izquierda. Cuanto mayor es la dispersión respecto de la diagonal, menor es el coeficiente. En el caso del gráfico 1(c) vemos como los puntos se sitúan alrededor de la diagonal opuesta dando como resultado un coeficiente negativo. En el gráfico 1 (e) ambas variables carecen de dependencia y el valor de correlación casi cero es correcto. No obstante, en el gráfico 1 (f) se muestra un ejemplo de la problemática comentada anteriormente, donde dos variables están funcionalmente relacionadas ($Y=|X|$) y son totalmente dependientes, pero el coeficiente de correlación lineal de Pearson devuelve un valor nulo. Esta problemática surge de la incapacidad de este coeficiente de recoger cualquier dependencia que no sea lineal.

Si la distribución conjunta de las dos variables sigue una distribución normal, el coeficiente de correlación lineal de Pearson es muy relevante, puesto que el concepto de independencia e incorrelación lineal se vuelve equivalente ya que, entre dichas variables, solo pueden existir relaciones lineales las cuales sí son capturadas por este coeficiente (Embrechts, McNeil, y Straumann, 2002 y Peña, 2004).

En el caso de los mercados financieros, el problema surge cuando las rentabilidades de los activos financieros no siguen distribuciones gaussianas por lo que pueden existir entre ellas relaciones no lineales que no se recogen en el coeficiente de correlación lineal de Pearson y pueden provocar errores importantes en el análisis de la dependencia entre ellas si se utiliza este coeficiente. Por ello, resulta de especial interés analizar cuáles son las posibles alternativas a este coeficiente, examinando las ventajas e inconvenientes que cada una presenta, y analizar la posible inconsistencia de resultados obtenidos al aplicar distintas medidas de dependencia entre las mismas variables (Embrechts, McNeil, y Straumann, 2002).

2.2 Coeficiente de asociación Tau de Kendall

El concepto de mayor relevancia para el análisis mediante este coeficiente es el de par concordante. Para dos pares de observaciones (x_i, y_i) y (x_j, y_j) decimos que estos pares son concordantes cuando $x_i < x_j$ e $y_i < y_j$ o $x_i > x_j$ e $y_i > y_j$, es decir, cuando $(x_i - x_j)(y_i - y_j) > 0$. En caso contrario diremos que los pares de observaciones no son concordantes. En nuestro caso no nos enfrentamos al tercer posible caso, que es cuando aparecen empates entre los valores ya que en los índices bursátiles es casi imposible encontrar dos valores iguales ya que los valores son muy precisos y están en constante actualización (Malevergne y Sornette, 2006).

Con el fin de facilitar el entendimiento de este concepto voy a incluir un breve ejemplo. Tenemos 3 pares de valores:

$$(x_1=1, y_1=1), (x_2=2, y_2=3) \text{ y } (x_3=6, y_3=0)$$

Si los analizamos vemos que los dos primeros pares formados son concordantes puesto que $x_1 < x_2$ e $y_1 < y_2$, por tanto $(x_1 - x_2)(y_1 - y_2) > 0$, es decir, el individuo 1 está “peor” posicionado con respecto a ambas variables. En el caso del primer y tercer par, el resultado es no concordantes puesto que $x_1 < x_3$ pero $y_1 > y_3$, por tanto, $(x_1 - x_3)(y_1 - y_3) < 0$, es decir, mientras el individuo 1 esta “mejor” posicionado en la variable Y, el 3 esta “mejor” posicionado en la variable X. Por último, los dos últimos pares son no concordantes puesto que arrojan el mismo resultado que el caso anterior, $(x_2 - x_3)(y_2 - y_3) < 0$.

Para poder trabajar con este coeficiente debemos calcular la probabilidad de que los pares sean concordantes y la probabilidad de que no lo sean. Una vez obtenidas ambas, el coeficiente de asociación Tau de Kendall, τ , resulta de la resta de la probabilidad de que sean concordantes menos la probabilidad de que no lo sean. En la práctica, para calcular este coeficiente, trabajaremos con el número de pares no concordantes y concordantes. En concreto, para su cálculo directo tenemos que:

$$\tau = \frac{(\text{número de pares concordantes}) - (\text{número de pares no concordantes})}{\binom{T}{2}} \quad (4)$$

donde $\binom{T}{2} = \frac{T(T-1)}{2}$ que denota todos los posibles pares que se pueden formar con T observaciones.

Al tratarse de una resta entre dos probabilidades los valores de τ oscilan entre -1 y 1 pudiendo tomar cualquier valor dentro de ese intervalo. En el caso de obtener un valor de 1 o muy próximo, tenemos dos variables que casi siempre son concordantes, por lo tanto, podemos hablar de una relación positiva entre las dos variables (cuando una toma valores pequeños, la otra también lo hace, y cuando una toma valores grandes, la otra también presenta valores grandes). Por el contrario, si obtenemos un valor de -1, todos los pares son no concordantes, por lo que se trata de una relación inversa o negativa donde una variable aumenta cuando la otra disminuye. Por último, si se obtiene un valor de 0, el coeficiente de asociación Tau de Kendall nos daría una ausencia de relación entre ambas variables (Liebetreau, 1983).

La interpretación de este índice en el ámbito de los mercados financieros sería la siguiente: el coeficiente de Kendall mide la probabilidad de que dos mercados

se muevan en la misma dirección, es decir, que aumenten o disminuyan a la vez. Por lo tanto, a los inversores que quieran diversificar el riesgo, les interesarán coeficientes de Kendall negativos o próximos a cero que indicarán, en el primer caso, que dos mercados se mueven en sentido opuesto, de modo que cuando en uno hay ganancias en el otro hay pérdidas y viceversa y, en el segundo, que no tienen relación ya que estos casos serían los mejores para la diversificación del riesgo.

A modo de ejemplo podríamos calcular el valor de este coeficiente a partir del ejemplo dado anteriormente, donde tenemos:

Número de pares concordantes: 1

Número de pares no concordantes: 2

Número de observaciones: 3

Tau de Kendall = -0.33333

Estaríamos ante un caso de dos variables que tienden a moverse de manera opuesta lo que implicaría que un aumento en una de ellas trae consigo una disminución en la otra en la mayoría de las ocasiones. Este ejemplo se ha realizado con solo 3 observaciones por lo que la interpretación del valor del índice no es tan clara como el que se obtiene al utilizar un gran número de observaciones como en el caso de los índices bursátiles que veremos más adelante.

2.3 Coeficiente de correlación por rangos de Spearman

El coeficiente de correlación de Spearman se basa en la utilización de rangos posicionales en lugar de los valores de cada observación. Para hallar el rango se parte de todas las observaciones de una variable y se asigna el valor 1 a aquella observación de menor valor, 2 a la siguiente más pequeña y así sucesivamente. Esto se realiza, también, con la otra variable cuya relación se quiere analizar.

Una vez hecho esto, se comparan los rangos de cada variable: si los rangos asignados a las dos variables fueran iguales, existiría una relación positiva perfecta, el individuo que es primero en una variable también lo es en la otra, el que es segundo es una variable también lo es en la otra, ... y el que es último en una variable también lo es en la otra. Para realizar esta comparación se obtiene la diferencia entre los rangos de la misma observación en cada variable, por ejemplo, a la observación i -ésima, (x_i, y_i) , se le asignan sus rangos, (R^x_i, R^y_i) , que son números del 1 a T , y a la diferencia de los rangos la denotaremos como $d_i = R^x_i - R^y_i$

Para facilitar la comprensión realizaremos un pequeño ejemplo con 6 valores de dos variables distintas.

Partimos de la siguiente muestra:

X	12	15	5	9	2	7
Y	10	4	6	2	8	14

A partir de ésta, calculamos tanto los rangos como la diferencia entre ellos

X	12	15	5	9	2	7
Rango de X	5	6	2	4	1	3
Y	10	4	6	2	8	14
Rango de Y	5	2	3	1	4	6
d	0	4	-1	3	-3	3

Ahora cabe de nuevo la posibilidad de obtener valores repetidos de X o de Y y habría que decidir qué rango asignar a las observaciones repetidas, lo cual complicaría el cálculo. Como hemos comentado, en nuestro caso esto no va a ocurrir por lo que podemos evitar comentar esta problemática.

En ausencia de empates la expresión del coeficiente de correlación de rangos de Spearman, r_s , que se obtiene es bastante sencilla:

$$r_s = 1 - \frac{6 \sum_{i=1}^T d_i^2}{T(T^2-1)}, \quad (5)$$

donde d_i es la diferencia previamente definida y T es el número de observaciones.

El valor de este coeficiente está siempre entre -1 y 1, como los dos anteriores, y la interpretación de los valores sería la siguiente:

- Un valor de 0 nos indica que no hay asociación entre las variables.
- Un valor positivo nos indica una asociación positiva de rangos entre las variables, es decir, los individuos mejor posicionados en una de las variables también tienden a estar mejor posicionados en la otra variable, siendo esta asociación perfecta cuando el valor es 1, es decir, cuando los rangos de los individuos en ambas variables son idénticos.
- Un valor negativo nos indica una asociación negativa de rangos entre las variables, es decir, los individuos mejor posicionados en una variable tienden a ser los peor posicionados en la otra, siendo esta asociación perfecta cuando el valor es -1, lo cual ocurre cuando los rangos de los individuos en ambas variables son completamente contrarios (el primero en una variable es el último en la otra, el segundo en una variable es el penúltimo en la otra y así sucesivamente) (Liebetreau, 1983).

Para el ejemplo previo, obtendríamos un coeficiente de correlación por rangos de Spearman de 0.790476 que indicaría una fuerte asociación positiva, por lo que cabría esperar que, a mayor valor de X en una observación, mayor valor de Y .

2.4 Medidas robustas de correlación

Uno de los problemas del coeficiente de correlación lineal de Pearson es que, al estar basado en promedios, es muy sensible a la presencia de datos atípicos, característica habitual en los mercados financieros debido a la presencia de valores extremos bien por fuertes caídas o fuertes rebotes del mercado. Una forma de “robustificar” este coeficiente es sustituir los promedios que aparecen en las ecuaciones (2), (3) y (4) por otras medidas de posición más robustas como

la mediana. Surgen así distintas medidas de correlación robustas, algunas de las cuales se describen a continuación.

En primer lugar, vamos a explicar la utilización de medianas en lugar de medias. La justificación es que la mediana se muestra mucho más robusta a la presencia de datos atípicos. Para ello podemos llevar a cabo un breve ejemplo. Tenemos una muestra de 5 observaciones de una variable que son:

$$x_1=1, x_2=1.2, x_3=1.1, x_4=1.15, x_5=0.9.$$

Para esta muestra, su media sería de 1.07 y su mediana sería 1.10, valores muy similares. Ahora supongamos que el dato x_4 toma el valor de 10, es decir, un dato atípico que difiere mucho del resto. Si realizamos el cálculo de media y mediana de nuevo, obtenemos 2.84 y 1.1 respectivamente. Como se puede observar, la media ha sufrido un cambio muy importante debido a la presencia del dato atípico extremadamente grande, que ha incrementado notablemente su valor haciéndolo no representativo, mientras que la mediana se ha mantenido robusta frente al mismo.

El primer coeficiente de correlación que utiliza la mediana en lugar de la media se denomina “estimador de la correlación a partir de la mediana” el cual denotaremos como r_{COMED} por su nombre en inglés *correlation median estimator* (Shevlyakov y Smirnov, 2011).

Su expresión es:

$$r_{COMED} = \frac{med((x-med(x))(y-med(y)))}{MAD(x)MAD(y)}, \quad (6)$$

donde $med(x)$ y $med(y)$ denotan la mediana de x e y , respectivamente, y MAD es una medida robusta de dispersión definida como

$$MAD(x) = med(|x - med(x)|). \quad (7)$$

El numerador en (6) surge de la utilización de las medianas en lugar de las medias en la ecuación (2) mientras que el denominador en (6), además de usar medianas en lugar de medias, utiliza el valor absoluto en lugar de elevar al cuadrado como se haría con una desviación típica normal en las ecuaciones (3) y (4). El objetivo de esto es el mismo, puesto que el valor absoluto corrige los posibles signos tanto negativos como positivos, para recoger solo la magnitud

de las desviaciones a la mediana. La diferencia es que, al evitar elevar al cuadrado, reducimos el impacto de los datos atípicos, los cuales, al elevarse su diferencia a la media al cuadrado, provocarían un mayor cambio en la varianza.

El segundo coeficiente basado en medidas de correlación robustas será el coeficiente de correlación cuadrática con signo, la r_Q de Blomquist. Este coeficiente se define como el promedio del signo de las desviaciones a la mediana de la muestra:

$$r_Q = T^{-1} \sum_{i=1}^T \text{sgn}(x_i - \text{med}(x))\text{sgn}(y_i - \text{med}(y)), \quad (8)$$

donde $\text{sgn}(x)$ vale 1 si $x > 0$, -1 si $x < 0$ y 0 en caso de que x sea igual a 0.

Por último, vamos a analizar otro coeficiente de correlación robusto obtenido a partir de la desviación absoluta a la mediana definida en (7).

Este coeficiente toma de partida el coeficiente de correlación lineal r_{XY} de Pearson definido en (1), a partir del cual se obtiene la siguiente identidad (Gnanadesikan y Kettenring, 1972):

$$r_{XY} = \frac{\text{VAR}(U) - \text{VAR}(V)}{\text{VAR}(U) + \text{VAR}(V)}, \quad (9)$$

donde $U = \left(\frac{x}{s_x} + \frac{y}{s_y}\right) / \sqrt{2}$ y $V = \left(\frac{x}{s_x} - \frac{y}{s_y}\right) / \sqrt{2}$ son dos variables tales que $\text{Cov}(U, V) = 0$ y $\text{Var}(U) = \text{Var}(V) = 1 + r_{XY}$. Sustituyendo en (9) las desviaciones típicas por su equivalente robusto, que es el MAD en (7) se obtiene un coeficiente que denotaremos como r_{MAD} , y tiene la siguiente expresión

$$r_{MAD} = \frac{\text{MAD}(u)^2 - \text{MAD}(v)^2}{\text{MAD}(u)^2 + \text{MAD}(v)^2}, \quad (10)$$

donde

$$u = \frac{x - \text{med}(x)}{\sqrt{2}\text{MAD}(x)} + \frac{y - \text{med}(y)}{\sqrt{2}\text{MAD}(y)}, \quad (11)$$

$$v = \frac{x - \text{med}(x)}{\sqrt{2}\text{MAD}(x)} - \frac{y - \text{med}(y)}{\sqrt{2}\text{MAD}(y)}. \quad (12)$$

Véase Shevlyakov y Smirnov (2011) para un desarrollo más detallado de este y los otros coeficientes anteriormente descritos.

2.5 Medidas de dependencia basadas en datos agrupados

En este último apartado de la metodología discutimos la implementación de otras formas de analizar la dependencia, basadas en datos agrupados.

En primer lugar, dividimos los datos de cada variable en grupos definidos por sus respectivos cuantiles, por ejemplo, en quintiles. Se obtiene el quintil al que pertenece cada observación tanto en X como en Y. Una vez hecho esto, se construye una tabla de doble entrada con una casilla para cada posible combinación de quintiles en cada variable. Por último, se calcula para cada casilla el número de observaciones (x_i, y_i) que están simultáneamente en la combinación de quintiles correspondiente a esa casilla (Hu, 2006).

Para facilitar su entendimiento vamos a realizar un pequeño ejemplo. Partimos de una muestra de dos variables de 10 observaciones cada una

X	1	2	3	4	5	6	7	8	9	10
Y	4	13	6	2	17	9	5	12	16	1

Calculamos los quintiles de cada variable, $(2,4,6,8,10)$ en el caso de X y $(2,5,9,13,17)$ en el caso de Y. Realizamos la división en quintiles y asignamos a cada observación el valor del quintil al que pertenece cada observación

X	1	1	2	2	3	3	4	4	5	5
Y	2	4	3	1	5	3	2	4	5	1

Creamos, por último, la tabla de frecuencias absolutas y relativas que recoge el número y proporción de observaciones (x_i, y_i) que están simultáneamente en un determinado quintil de X y en otro de Y; véase tabla 1.

En una situación de independencia entre las variables, cabría esperar una distribución uniforme por todo el cuadro con frecuencias relativas conjuntas de 0.04 en cada celda, mientras que, si las ordenaciones de los individuos en estas variables fueran iguales, es decir, si los que están en el primer quintil en una variable también están en el primer quintil de la otra y así sucesivamente, los

valores aparecerían únicamente distribuidos en la diagonal principal, con una frecuencia relativa de 0.2 en cada celda de dicha diagonal y fuera de la diagonal solo habría ceros.

Tabla 1 Distribución conjunta por quintiles de las variables X e Y tanto en frecuencia absoluta como en relativa (entre paréntesis).

X	Y					Total
	1 ^{er} quintil	2 ^o quintil	3 ^{er} quintil	4 ^o quintil	5 ^o quintil	
1 ^{er} quintil	0 (0)	1 (0.10)	0 (0)	1 (0.1)	0 (0)	2 (0.2)
2 ^o quintil	1 (0.1)	0 (0)	1 (0.1)	0 (0)	0 (0)	2 (0.2)
3 ^{er} quintil	0 (0)	0 (0)	1 (0.1)	0 (0)	1 (0.1)	2 (0.2)
4 ^o quintil	0 (0)	1 (0.1)	0 (0)	1 (0.1)	0 (0)	2 (0.2)
5 ^o quintil	1 (0.1)	0 (0)	0 (0)	0 (0)	1 (0.1)	2 (0.2)
Total	2 (0.2)	2 (0.2)	2 (0.2)	2 (0.2)	2 (0.2)	10 (1)

Cabe destacar que la realización de un ejemplo para este método con tan pocas observaciones solo busca explicar la ejecución de éste. Para poder analizar el resultado se requiere de un número mucho mayor de observaciones, como el que veremos en la aplicación.

3. APLICACIÓN EMPÍRICA

3.1 Descripción de los datos

En esta sección vamos a evaluar la dependencia entre distintos mercados financieros aplicando los diferentes coeficientes, previamente comentados, a series diarias de datos de la rentabilidad de los índices bursátiles de 4 mercados financieros en el periodo 2010-2020. Concretamente, los índices elegidos son los siguientes:

- IBEX35: Está formado por las 35 empresas con más liquidez que cotizan en las cuatro bolsas españolas pertenecientes al Sistema de Interconexión Bursátil Español. Además, es un índice ponderado, por lo que no todas las empresas tienen el mismo peso. He escogido este índice debido a que es el índice de referencia de España de modo que, al incluirlo, hago más cercanas todas las conclusiones de este trabajo a la economía financiera española y puedo compararlo con la situación de otros mercados.
- NASDAQ100: Este índice está formado por las 100 compañías más importantes del sector de la industria y la tecnología. Estas empresas son principalmente estadounidenses, pero también incluye algunas de otros países. Se trata de uno de los grandes índices bursátiles estadounidenses junto con el S&P500 y el Dow Jones. Es importante incluir uno de ellos ya que son los índices de referencia del mercado financiero más grande del mundo y tiene algunas de las empresas más influyentes. Además, el mercado financiero estadounidense ha sido el más utilizado como modelo de mercado en el análisis teórico y práctico de los mercados financieros.
- NIKKEI 225. Este índice se compone de los 225 valores más líquidos de los que cotizan en la Bolsa de Tokio. Se trata del mayor índice bursátil de Japón. He incluido este índice ya que me parecía muy importante incluir un mercado asiático, dado que utilizar solo mercados occidentales podría arrojar resultados engañosos sobre las dependencias entre ellos. La elección del mercado nipón se basa en que no solo es uno de los mayores de toda la región, sino que también es uno de los más accesibles en cuanto a datos, tanto actuales como pasados, facilitando la tarea y ofreciendo mayores garantías sobre ellos, especialmente cuando nos remontamos en el pasado.
- Euro Stoxx 50: Este índice está formado por las 50 empresas más grandes de los 19 supersectores en términos de capitalización de mercado en 11 de los países de Europa. Estos países son Alemania, Austria, Bélgica, España, Finlandia, Francia, Irlanda, Italia, Luxemburgo, Países Bajos y Portugal. He escogido este índice ya que representa la situación general de Europa. Esto tiene dos puntos positivos, el primero es que refleja mejor la situación del continente que el IBEX35 y en segundo lugar permite realizar un análisis de la dependencia de España con Europa la cual debería ser alta.

Todos estos datos se han obtenido de Yahoo Finance (www.finance.yahoo.com) un portal financiero que contiene series históricas de los precios de cierre de los principales índices bursátiles. En nuestro caso, los datos se han adquirido con una frecuencia diaria y con un total de once años completos, de 2010 a 2020.

A la hora de analizar la dependencia entre mercados, en lugar de utilizar los valores de cierre de los índices o la rentabilidad porcentual diaria (diferencia porcentual entre los valores de cierre de dos días consecutivos), vamos a utilizar los denominados “compound returns”, esto es, la diferencia en logaritmos de los precios de cierre, definidos de la siguiente forma. Si denotamos por P_t el valor de cierre del índice bursátil en el tiempo t , los rendimientos X_t que vamos a analizar se definen como

$$X_t = 100 * (Ln(P_t) - Ln(P_{t-1})) \quad (13)$$

Las ventajas de estos rendimientos frente a los rendimientos porcentuales habituales pueden verse en Törnqvist, Vartia y Vartia (1985). Se incluye una explicación breve al respecto en el Anexo.

Adicionalmente, es importante destacar que estamos realizando un análisis de relaciones contemporáneas comparando los precios de cierre de cada índice siempre en el mismo día, es decir, de X_t con Y_t para un mismo t . Se podría realizar un análisis dinámico en el que se compararan los valores de un día en un índice bursátil, X_t , con los de otro día de otro índice bursátil, Y_{t-h} , para varios retardos $h=1,2, \dots$. Esto es especialmente importante cuando trabajamos con mercados situados en husos horarios diferentes, ya que el impacto de un suceso puede repercutir en tiempo real en un mercado mientras que otro se encuentra ya cerrado y su efecto se contabilizará en el siguiente día. Además, los shocks que pueden afectar a un mercado pueden tener efectos retardados en el propio mercado y no solo efectos contemporáneos. No obstante, este estudio dinámico queda fuera del alcance de este trabajo, pero será tenido en cuenta a la hora de analizar los resultados de este.

3.2 Análisis de la distribución de los rendimientos bursátiles.

En esta sección vamos a analizar las principales propiedades estadísticas de los rendimientos bursátiles.

Comenzamos estudiando los estadísticos descriptivos básicos de las series que tenemos, concretamente las principales medidas de posición central, dispersión y forma (véase Peña, 2014), y luego contrastaremos el supuesto de normalidad para la distribución marginal. Denotaremos por T el número total de observaciones en una serie dada.

3.2.1 Análisis descriptivo

Los principales estadísticos descriptivos que vamos a analizar son los siguientes:

- Media. Medida básica del centro de una distribución que representa el punto de equilibrio, de modo que las diferencias positivas y negativas a la media suman cero. Su expresión es la siguiente:

$$\bar{x} = \frac{\sum_{i=1}^T x_i}{T} \quad (14)$$

- Mediana. Valor central de la muestra al ordenar los valores de menor a mayor, de modo que el 50% de los valores son menores o iguales que el valor mediano. Si la distribución es simétrica, la media y la mediana coinciden, mientras que un valor de la media menor que la mediana se considera un indicativo de asimetría a la izquierda y viceversa, un valor de la media mayor que la mediana indica asimetría a la derecha.
- Desviación típica. Medida de dispersión de los datos que cuantifica la representatividad de la media. Definida como:

$$s_x = \sqrt{\frac{\sum_{i=1}^T (x_i - \bar{x})^2}{T}} \quad (15)$$

- Coefficiente de asimetría. Se trata de una medida de forma con la siguiente expresión:

$$CA_x = \frac{\sum_{i=1}^T (x_i - \bar{x})^3}{T s_x^3} \quad (16)$$

Es una cantidad adimensional que vale cero si la distribución es simétrica respecto a la media. En caso de que el valor sea positivo indica una asimetría a la derecha mientras que valores negativos indicarán una asimetría a la izquierda.

- Coeficiente de curtosis. Es otra medida adimensional de forma que mide el grado de concentración de probabilidad tanto en las colas como alrededor de la media y lo compara con una distribución normal. Su expresión es la siguiente:

$$g_x = \frac{\sum_{i=1}^T (x_i - \bar{x})^4}{T s_x^4} \quad (17)$$

Para una distribución Normal el valor de este coeficiente es $g_x = 3$. Valores mayores que éste indicarán una distribución con colas más gruesas que en una Normal, es decir, más probabilidad de valores extremos que en una distribución Normal. Por el contrario, valores menores a 3 nos indicarán colas menos gruesas que la distribución Normal.

En la tabla 2 se muestran todos los valores previamente definidos junto con el valor máximo y mínimo obtenidos para los rendimientos de cada índice bursátil. El número de observaciones es $T=2828$ en todos los casos ya que las fechas en las que alguno de los índices bursátiles no estaba disponible han sido eliminadas de la muestra para homogenizar las series y poder calcular las correlaciones.

Tabla 2. Estadísticos descriptivos básicos de los rendimientos de los índices bursátiles.

	media	mediana	S	CA	g	Máx.	Mín.
NASDAQ100	0.073	0.126	1.312	-0.678	11.886	8.934	-13.149
IBEX35	-0.005	0.056	1.558	-0.651	12.845	13.483	-15.151
EUROSTOXX50	0.011	0.050	1.424	-0.515	9.917	9.846	-13.240
NIKKEI225	0.0393	0.0680	1.401	-0.3494	7.5507	7.731	-11.153

En primer lugar, observamos valores de media y mediana muy cercanos a cero en todos los casos, algo que cabría esperar como resultado del equilibrio entre las fluctuaciones positivas (subidas de precios) y negativas (bajadas de precios) dando lugar a una suma prácticamente igual a cero. No obstante, al comparar

las medias y medianas, vemos que la media es menor a la mediana en todos los casos, lo cual es un primer indicativo de no normalidad, ya que el valor central es mayor que el valor medio de los datos, algo propio de las distribuciones asimétricas a la izquierda.

La desviación típica presenta valores semejantes en todos los casos teniendo una menor volatilidad el NASDAQ y una mayor el IBEX mientras que los otros dos se posicionan en un terreno intermedio.

Los dos siguientes coeficientes son los más interesantes desde el punto de vista del análisis de la forma de la distribución. Vemos cómo se cumplen las singularidades que hemos comentado al inicio del trabajo, presentando todas las distribuciones una asimetría a la izquierda. Todos los valores del coeficiente de asimetría son negativos y semejantes, aunque podemos observar una menor asimetría en el EUROSTOXX50. La presencia de asimetría en las distribuciones, que ya habíamos apuntado por la posición de media y mediana, es de nuevo un indicio de la falta de normalidad.

Por último, el coeficiente de curtosis presenta siempre valores ampliamente mayores que 3, lo cual demuestra una forma con colas más grandes que la distribución Normal, es decir, los rendimientos bursátiles presentan más valores extremos (positivos o negativos) de los que cabría esperar en una distribución Normal.

Una vez analizados estos estadísticos, pasamos al estudio de la normalidad de la distribución en el siguiente epígrafe. Como hemos comentado previamente, éste es uno de los pilares de este trabajo puesto que la problemática relacionada con el uso del coeficiente de correlación lineal de Pearson es la no normalidad de las distribuciones en los índices bursátiles.

3.2.2 Contrastes de normalidad

Para contrastar la posible ausencia de Normalidad en los rendimientos financieros que parecen indicar los estadísticos descriptivos previamente analizados se han usado 3 contrastes de normalidad diferentes.

Estos contrastes siguen un modelo similar cuyo objetivo final es contrastar las siguientes hipótesis

$$H_0: X \rightarrow N(\mu, \sigma)$$

$$H_1: X \nrightarrow N(\mu, \sigma)$$

Kolmogorov-Smirnov-Lilliefors

El estadístico para este contraste es el siguiente:

$$D_T = \max(F_T(x_i) - F(x_i)) , \quad (18)$$

donde F_T es la función de distribución empírica obtenida a partir de la muestra y F es la distribución teórica bajo la hipótesis nula, en nuestro caso, la distribución normal. El estadístico D_T mide la diferencia máxima existente entre ambas y la compara con un valor crítico k_α de referencia: si la diferencia es mayor que k_α , rechazamos H_0 al nivel α . Habitualmente trabajaremos con valores de $\alpha = \{0.01, 0.05, 0.1\}$.

Para el propósito de este trabajo vamos a utilizar la modificación de Lilliefors. Esta modificación permite utilizar el test, previamente comentado, en el caso de utilizar una distribución Normal bajo H_0 donde no se especifiquen su media ni su desviación típica. Esto es posible a partir de la tabulación de los valores críticos realizada por Lilliefors para el estadístico D_T en caso de normalidad, estimando μ y σ mediante los correspondientes valores muestrales \bar{x} y S (Peña, 2014) quedando $F(x_i)$ definido como:

$$F(x_i) = \Phi\left(\frac{x_i - \bar{x}}{s}\right) , \quad (19)$$

donde Φ denota la función de distribución de una variable normal $N(0,1)$.

Anderson Darling

De nuevo estamos ante una modificación de la prueba de Kolmogorov-Smirnov que busca incrementar su sensibilidad en las colas de las distribuciones por lo

que es especialmente útil para nuestro estudio. Las T observaciones se ordenan de menor a mayor de modo que $x_1 < x_2 \dots < x_n$ y para la notación asumimos que x_i representa las observaciones ordenadas.

La expresión del estadístico de contraste es la siguiente:

$$A_N^2 = -T - \sum_{i=1}^T \frac{2i-1}{T} [\ln(F(x_i)) + \ln(1 - F(x_{T+1-i}))] \quad (20)$$

Como en el caso anterior, si bajo H_0 no se especifica ni μ ni σ , estos valores se estiman mediante sus valores muestrales \bar{x} y S y $F(x_i)$ queda definida de igual modo que en (19).

El estadístico obtenido se compara con el valor crítico escogido para un determinado nivel de significación α y se decide si la hipótesis puede o no ser rechazada para ese valor de significación.

Jarque-Bera

Este contraste se basa en comparar simultáneamente, si la distribución formada por los datos de estudio se asemeja a una distribución normal, tanto en simetría como en curtosis. La expresión del estadístico de contraste es la siguiente.

$$JB = N \left(\frac{CA^2}{6} + \frac{(g-3)^2}{24} \right), \quad (21)$$

donde CA es el coeficiente de asimetría definido en (16) y g es en coeficiente de curtosis definido en (17). El valor de este estadístico es siempre positivo o cero. Cuanto más se aleje de cero mayor indicativo de no normalidad. Al igual que en los contrastes anteriores, el valor de JB se compara con el valor crítico para un nivel de significación α y si es mayor que ese valor, se rechaza H_0 a nivel de significación α .

La tabla 3 recoge los resultados de estos 3 test de Normalidad, correspondientes a cada índice bursátil, con el p-valor entre paréntesis. El p-valor define el límite de significación a partir del cual se puede rechazar la hipótesis nula. Los resultados se han obtenido con MATLAB.

Tabla 3. Valores de los estadísticos del contraste de Normalidad para las rentabilidades de los distintos índices bursátiles junto a su respectivo p-valor (entre paréntesis).

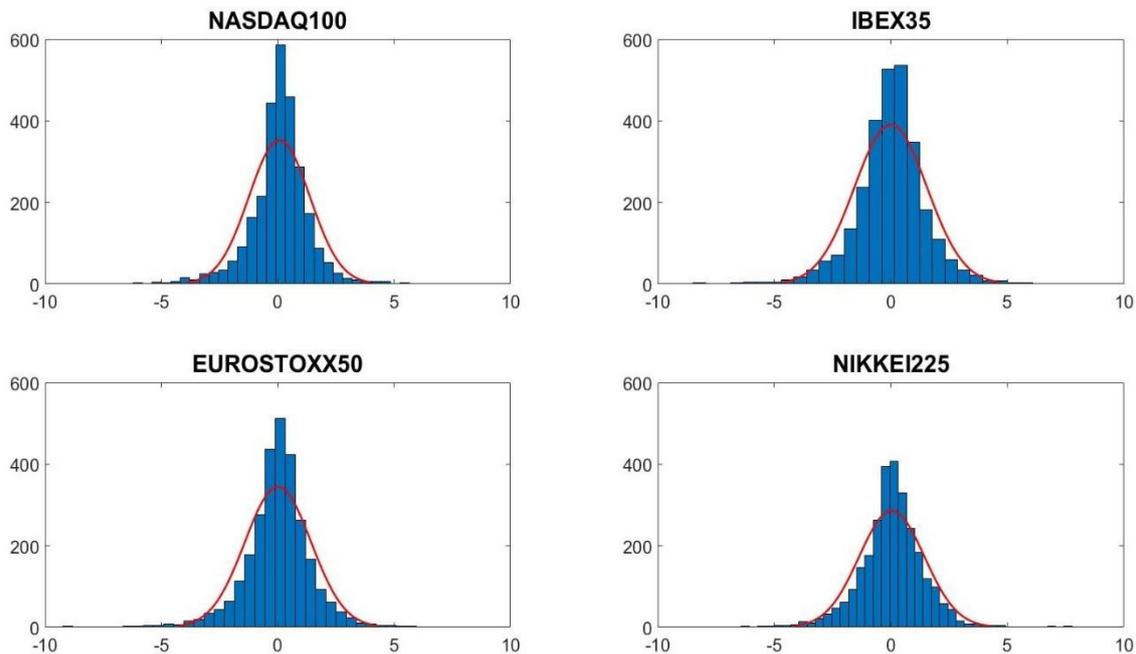
	Kolmogorov-Smirnov-Lilliefors	Anderson-Darling	Jarque-Bera
NASDAQ100	0.103 (<0.001)	53.428 (<0.0005)	9523.3 (<0.001)
IBEX35	0.0730 (<0.001)	Inf (<0.0005)	11623 (<0.001)
EUROSTOXX50	0.0777 (<0.001)	36.8116 (<0.0005)	5763.7 (<0.001)
NIKKEI225	0.0665 (<0.001)	23.2327 (<0.0005)	2497.8 (<0.001)

Como se puede observar, se rechaza la hipótesis nula a todos los niveles de significación usados habitualmente ya que los p-valores son muy pequeños. En consecuencia, podemos concluir que no podemos considerar ninguna de las series de rentabilidades de estos índices bursátiles como distribuidos normalmente y que, si lo hacemos, los resultados que de este supuesto se deriven deberían, al menos, interpretarse con cautela.

De manera gráfica podemos visualizar, también, esta no normalidad. Concretamente, el gráfico 2 representa, para las cuatro series analizadas, su histograma acompañado de una distribución normal con la misma media y desviación típica para observar las notorias diferencias que existen entre ambas.

En el gráfico 2 podemos apreciar esa ligera asimetría a la izquierda que obtuvimos con el coeficiente de asimetría y comparando la posición de media y mediana. Además, podemos ver con claridad como la presencia de valores extremos negativos y positivos, es más frecuente que en una distribución Normal: en las colas, el histograma siempre está por encima de la densidad teórica de la distribución Normal. Además, podemos observar la presencia de algunas rentabilidades extremas negativas en la cola izquierda.

Gráfico 2. Histogramas de las rentabilidades de los cuatro índices bursátiles junto a una distribución Normal de misma media y desviación típica



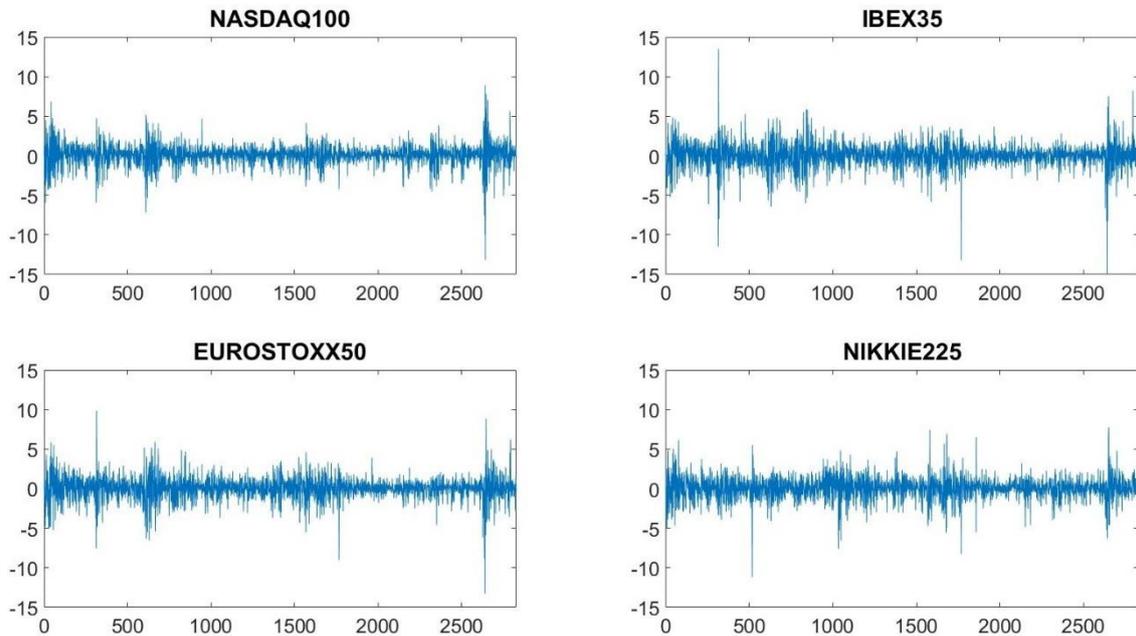
Una vez comprobado la falta de Normalidad de las rentabilidades de los índices bursátiles, se justifica el cálculo de los diferentes coeficientes de asociación y correlación presentados en la sección 2 para comprobar si estos difieren de manera notoria del coeficiente de correlación lineal de Pearson.

3.3 Medidas de dependencia entre los mercados financieros

En este apartado pasaremos al cálculo de todos los coeficientes de dependencia descritos en la sección 2, así como a su interpretación. Previamente, realizaremos una serie de gráficos sobre los rendimientos de los índices bursátiles.

En el gráfico 3 podemos observar la evolución temporal de los rendimientos diarios de los 4 índices bursátiles en el periodo 2010-2020.

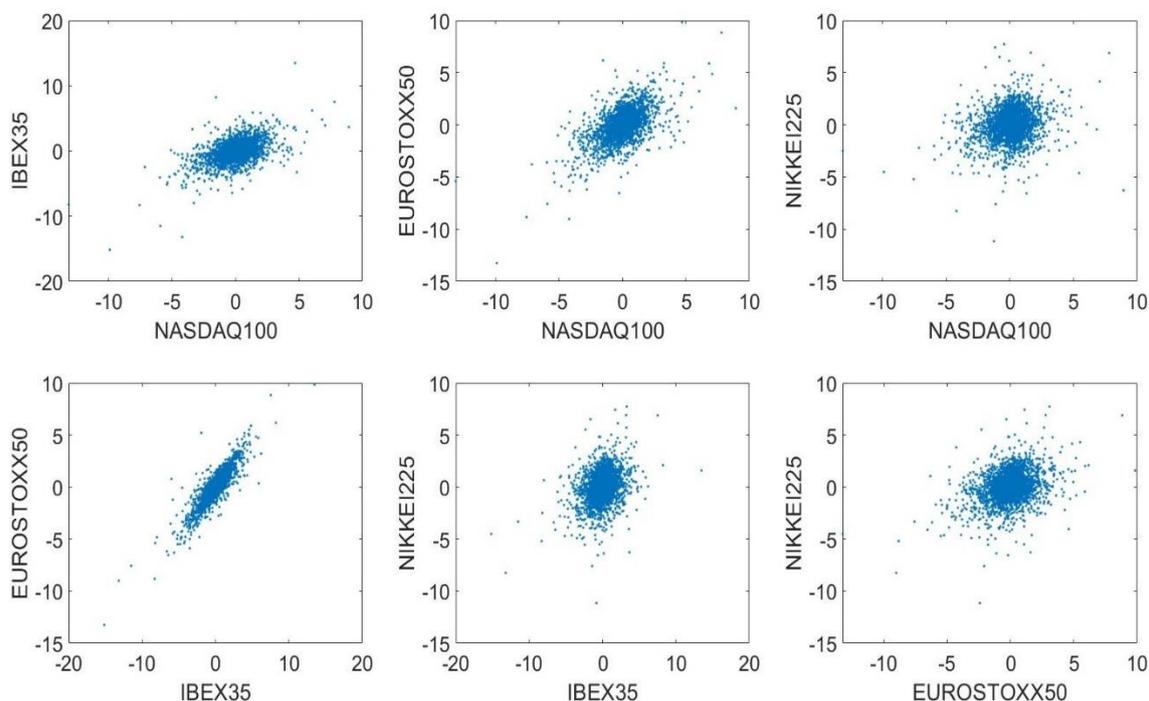
Gráfico 3. Evolución temporal de los rendimientos diarios de los índices bursátiles.



En el gráfico 3 se puede observar cómo los periodos de volatilidad suelen aparecer en los mismos instantes temporales en todos los gráficos. Por ejemplo, en la observación $t=2641$, correspondiente al 12 de marzo de 2020, fecha en la cual comenzaron las grandes restricciones con motivo del COVID-19, vemos una importante caída en la rentabilidad de todos los índices bursátiles a excepción del NIKKEI225 (en el caso del IBEX35 llegó hasta pérdidas del 15%). Otro caso destacado es la gran caída acompañada de una fuerte recuperación, especialmente del IBEX35 y el EUROSTOXX50, en $t=315$, que corresponde al 6 de mayo de 2010, conocido como el “2010 flash crash”, un evento que provocó una fuerte caída del mercado financiero acompañada de una gran recuperación poco después. A simple vista, parece que es el IBEX35 el que tiene una mayor volatilidad mientras que el NIKKEI225 tiene los mínimos y máximos menos extremos, como habíamos visto en la tabla 2. El NASDAQ100 es el que menor desviación típica presenta lo cual coincide con un diagrama más concentrado a excepción de algunos sucesos concretos.

En el gráfico 4 podemos ver la representación como nube de puntos de cada combinación posible de rentabilidades de los índices bursátiles, es decir, se representan los T valores (x_i, y_i) , con $i=1, \dots, T$ para cada par de índices.

Gráfico 4. Diagramas de dispersión de las 6 posibles combinaciones de los 4 índices bursátiles.



A la vista de los diagramas de dispersión podemos prever que la mayor correlación sea entre el IBEX35 y el EUROSTOXX50, ya que su diagrama de dispersión presenta valores muy alineados y bastante concentrados en torno a una recta, y las menores correlaciones serán las que incluyen al NIKKEI225, ya que su diagrama de dispersión frente a cualquiera de los otros índices es más redondeado y presenta mayor dispersión. Adicionalmente, observamos que la mayoría de los puntos se colocan en el cuadrante superior derecho e inferior izquierdo, lo cual indica una correlación positiva. Apenas nos encontramos puntos en las esquinas inferior derecha y superior izquierda que indicarían lo contrario, esto es, una correlación negativa.

A continuación, calcularemos los seis coeficientes de asociación presentados en la sección 2. Estos coeficientes, a diferencia de los calculados en los apartados anteriores, son correspondientes a pares de variables por lo que no se realizará por separado para cada índice bursátil sino sobre las 6 posibles combinaciones por pares entre ellos. La tabla 4 recoge los valores de estos coeficientes.

Tabla 4. Valores de los distintos coeficientes de asociación y correlación para las distintas parejas de índices bursátiles.

	Pearson	Kendall	Spearman	r_{COMED}	r_Q	r_{MAD}
NASDAQ100, IBEX35	0.518	0.300	0.425	0.228	0.264	0.446
NASDAQ100, EUROSTOXX50	0.576	0.344	0.480	0.288	0.314	0.534
NASDAQ100, NIKKEI225	0.202	0.109	0.160	0.043	0.079	0.172
IBEX35, EUROSTOXX50	0.889	0.696	0.864	0.792	0.701	0.889
IBEX35, NIKKEI225	0.302	0.172	0.252	0.128	0.162	0.263
EUROSTOXX50, NIKKEI225	0.329	0.183	0.267	0.127	0.178	0.284

A la vista de los resultados obtenidos en la tabla 4, podemos extraer las siguientes conclusiones:

- Todos los coeficientes son positivos. Esto implica que la relación de dependencia es positiva en todos los casos, es decir, los distintos mercados tienden a moverse en la misma dirección: cuando uno tiene beneficios el otro también, y viceversa, cuando uno tiene pérdidas el otro también. Este patrón concuerda con las formas de las nubes de puntos del gráfico 4.
- Se podrían realizar tres subgrupos en base a la similitud de los valores:
 - El coeficiente de correlación lineal de Pearson siempre toma un valor más grande que cualquiera de los otros coeficientes. Por lo tanto, parece sobreestimar la relación de dependencia existente.
 - El coeficiente de correlación de Spearman y r_{MAD} adquieren valores semejantes para todas las combinaciones estudiadas y ligeramente inferiores al coeficiente de correlación lineal de Pearson.
 - Los otros tres coeficientes devuelven valores semejantes entre sí y más pequeños que los otros coeficientes, aunque el orden, de menor a mayor, es r_{COMED} , r_Q y τ de Kendall. No obstante, este orden no se cumple en la combinación IBEX35 y EUROSTOXX35.
- Si ordenamos los diferentes pares de índices de menor a mayor en función de la relación entre ellos, obtenemos exactamente la misma ordenación con todos los coeficientes estudiados: la mayor dependencia se observa siempre entre los índices IBEX35 y EUROSTOXX50, mientras que la menor dependencia se aprecia siempre entre el índice bursátil estadounidense y el nipón. La única excepción sería el coeficiente r_{COMED} en las dos últimas parejas, donde aparecen intercambiadas, aunque con valores casi idénticos. Esto podría indicar que, a pesar de que los distintos coeficientes describan la dependencia de forma distinta, todos ellos pueden ser usados con el propósito de averiguar si dos mercados son más o menos dependientes entre sí que otros dos usados de referencia.
- Cuanto mayores son los coeficientes, por ejemplo, el coeficiente de correlación lineal de Pearson, para una combinación de rentabilidades de

índices bursátiles, menor es la diferencia existente entre los distintos coeficientes obtenidos para ella. Esto es un indicativo de que cuando dos mercados están estrechamente relacionados, su relación se aproxima más a una lineal que cuando no lo están. En nuestro caso, la relación entre IBEX35 y EUROSTOXX50 es la que mayores coeficientes tiene y en la que menor diferencia existe entre ellos. Si comparamos esto con lo obtenido en el gráfico 4, vemos que la relación entre estos dos índices es la que más se aproxima a una recta en el diagrama de dispersión. De igual manera, el caso contrario sería la relación entre el mercado nipón y el americano, que obtiene los coeficientes más pequeños y una gran diferenciación entre los diferentes coeficientes estudiados. Esto coincide con un diagrama de dispersión muy redondeado donde apenas se puede establecer una relación lineal.

De todo lo anterior se puede concluir que los índices bursátiles más dependientes entre sí son los europeos y que éstos se comportan de manera casi idéntica cuando se analizan en relación con otro índice no europeo. Esto no es algo que sorprenda debido a que ambos son índices europeos y sus movimientos deberían ser semejantes puesto que son muy próximos, no solo geográficamente, sino también cultural e institucionalmente. Entre los otros, el NIKKEI225 es el que menos depende contemporáneamente de los demás, en especial del estadounidense. Que el índice asiático sea el menos relacionado con el resto era un resultado esperable ya que son los más diferenciados en cultura y régimen institucional de los estudiados, aunque no se conociera con cuál de los demás índices bursátiles de los otros continentes tendría menor dependencia. Por último, el estadounidense es más dependiente de los europeos que del nipón.

3.4 Análisis de dependencia con datos agrupados por quintiles

Para finalizar vamos a realizar un análisis de los datos basado en sus quintiles conforme se ha explicado en la sección 2.5:

- Clasificamos todos los datos de cada índice bursátil en 5 grupos en base al quintil al que pertenecen.
- Tomamos cada uno de los posibles pares de índices y analizamos a qué dos quintiles pertenecen cada observación (x_i, y_i) de las dos variables.
- Sumamos cuántas veces aparece cada una de las combinaciones de quintiles posibles y anotamos el resultado en la casilla correspondiente de la tabla.

El resultado se muestra en las tablas 5-10 donde las filas equivalen a cada uno de los quintiles del primer índice bursátil y las columnas a los quintiles del segundo. En nuestro caso la frecuencia relativa, f_{ij} , independientes con $i = 1, 2, \dots, 5$ y $j = 1, 2, \dots, 5$, debería ser 0.04 (0.2×0.2) en todas las celdas si los dos índices analizados fueran independientes.

Adicionalmente, se ha calculado el valor del estadístico chi cuadrado para cada una de las tablas con el propósito de contrastar la independencia entre los dos pares de índices correspondientes.

El estadístico Chi-cuadrado compara los valores de la frecuencia observada, n_{ij} , de cada celda de la tabla con las frecuencias absolutas esperadas, e_{ij} , que aparecerían si las dos variables fueran independientes ($e_{ij} = n_{i.} \times n_{.j} / T$). La expresión del estadístico de contraste sería:

$$\chi^2 = \sum_{i=1}^5 \sum_{j=1}^5 \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad (22)$$

Una vez obtenido el estadístico de contraste, se compara con la distribución chi cuadrado teniendo en cuenta el número de grados de libertad. Los grados de libertad se obtienen como el número de filas menos 1 por el número de columnas menos uno. Luego, para nuestro caso, los valores del estadístico chi cuadrado deberían compararse con el percentil correspondiente de una variable aleatoria chi cuadrado con 16 grados de libertad. Para los tres grados de significación más usados, 0.1, 0.05 y 0.01, esos valores críticos serían de 23.5, 26.3 y 32, respectivamente.

Tabla 5. Distribución conjunta por quintiles del NASDAQ100 y el IBEX35 de frecuencias absolutas y relativas (entre paréntesis).

NASDAQ100	IBEX35					Total
	1 ^{er} quintil	2 ^o quintil	3 ^{er} quintil	4 ^o quintil	5 ^o quintil	
1 ^{er} quintil	261 (0.092)	146 (0.052)	55 (0.019)	56 (0.02)	48 (0.017)	566 (0.2)
2 ^o quintil	124 (0.044)	127 (0.045)	146 (0.052)	109 (0.039)	59 (0.021)	565 (0.2)
3 ^{er} quintil	73 (0.025)	141 (0.05)	135 (0.048)	127 (0.045)	90 (0.032)	566 (0.2)
4 ^o quintil	62 (0.022)	85 (0.030)	151 (0.053)	148 (0.052)	119 (0.042)	565 (0.2)
5 ^o quintil	46 (0.016)	66 (0.023)	79 (0.028)	125 (0.044)	250 (0.088)	566 (0.2)
Total	566 (0.2)	565 (0.2)	566 (0.2)	565 (0.2)	566 (0.2)	2828 (100%)

Nota: $\chi^2 = 659.641$.

Tabla 6. Distribución conjunta por quintiles del NASDAQ100 y e IEUROSTOXX50 de frecuencias absolutas y relativas (entre paréntesis).

NASDAQ100	EUROSTOXX50					Total
	1 ^{er} quintil	2 ^o quintil	3 ^{er} quintil	4 ^o quintil	5 ^o quintil	
1 ^{er} quintil	293 (0.104)	110 (0.04)	69 (0.024)	57 (0.020)	37 (0.013)	566 (0.2)
2 ^o quintil	115 (0.041)	154 (0.055)	136 (0.048)	92 (0.033)	68 (0.024)	565 (0.2)
3 ^{er} quintil	60 (0.021)	156 (0.055)	138 (0.049)	137 (0.048)	75 (0.027)	566 (0.2)
4 ^o quintil	56 (0.02)	93 (0.033)	143 (0.051)	158 (0.056)	115 (0.041)	565 (0.2)
5 ^o quintil	42 (0.015)	52 (0.018)	80 (0.028)	121 (0.043)	271 (0.1)	566 (0.2)
Total	566 (0.2)	565 (0.2)	566 (0.2)	565 (0.2)	566 (0.2)	2828 (100%)

Nota: $\chi^2 = 854.814$.

Tabla 7. Distribución conjunta por quintiles del NASDAQ100 y el NIKKEI225 de frecuencias absolutas y relativas (entre paréntesis).

NASDAQ100	NIKKEI225					Total
	1 ^{er} quintil	2 ^o quintil	3 ^{er} quintil	4 ^o quintil	5 ^o quintil	
1 ^{er} quintil	194 (0.069)	113 (0.04)	98 (0.035)	76 (0.027)	85 (0.03)	566 (0.2)
2 ^o quintil	93 (0.033)	131 (0.046)	114 (0.040)	134 (0.047)	93 (0.033)	565 (0.2)
3 ^{er} quintil	87 (0.031)	119 (0.042)	135 (0.048)	114 (0.040)	111 (0.039)	566 (0.2)
4 ^o quintil	89 (0.031)	106 (0.037)	122 (0.043)	122 (0.043)	126 (0.045)	565 (0.2)
5 ^o quintil	103 (0.036)	96 (0.034)	97 (0.034)	119 (0.042)	151 (0.0534)	566 (0.2)
Total	566 (0.2)	565 (0.2)	566 (0.2)	565 (0.2)	566 (0.2)	2828 (100%)

Nota: $\chi^2 = 130.716$.

Tabla 8. Distribución conjunta por quintiles del IBEX35 y el EUROSTOXX50 de frecuencias absolutas y relativas (entre paréntesis).

IBEX35	EUROSTOXX50					Total
	1 ^{er} quintil	2 ^o quintil	3 ^{er} quintil	4 ^o quintil	5 ^o quintil	
1 ^{er} quintil	423 (0.15)	110 (0.039)	22 (0.008)	7 (0.002)	4 (0.001)	566 (0.2)
2 ^o quintil	115 (0.041)	279 (0.099)	138 (0.049)	29 (0.010)	4 (0.001)	565 (0.2)
3 ^{er} quintil	16 (0.006)	143 (0.051)	241 (0.085)	148 (0.052)	18 (0.006)	566 (0.2)
4 ^o quintil	8 (0.003)	28 (0.01)	136 (0.048)	275 (0.097)	118 (0.042)	565 (0.2)
5 ^o quintil	4 (0.001)	5 (0.002)	29 (0.010)	106 (0.037)	422 (0.149)	566 (0.2)
Total	566 (0.2)	565 (0.2)	566 (0.2)	565 (0.2)	566 (0.2)	2828 (100%)

Nota: $\chi^2 = 3383.6$.

Tabla 9. Distribución conjunta por quintiles del IBEX35 y el NIKKEI225 de frecuencias absolutas y relativas (entre paréntesis).

IBEX35	NIKKEI225					Total
	1 ^{er} quintil	2 ^o quintil	3 ^{er} quintil	4 ^o quintil	5 ^o quintil	
1 ^{er} quintil	200 (0.071)	125 (0.044)	100 (0.035)	75 (0.027)	66 (0.023)	566 (0.2)
2 ^o quintil	115 (0.041)	133 (0.047)	118 (0.042)	126 (0.045)	73 (0.026)	565 (0.2)
3 ^{er} quintil	99 (0.035)	112 (0.04)	117 (0.041)	116 (0.041)	122 (0.043)	566 (0.2)
4 ^o quintil	76 (0.027)	102 (0.036)	131 (0.046)	126 (0.045)	130 (0.046)	565 (0.2)
5 ^o quintil	76 (0.027)	93 (0.033)	100 (0.035)	122 (0.043)	175 (0.062)	566 (0.2)
Total	566 (0.2)	565 (0.2)	566 (0.2)	565 (0.2)	566 (0.2)	2828 (100%)

Nota: $\chi^2 = 195.992$.

Tabla 10. Distribución conjunta por quintiles del EUROSTOXX50 y el NIKKEI225 de frecuencias absolutas y relativas (entre paréntesis).

EUROSTOXX50	NIKKEI225					Total
	1 ^{er} quintil	2 ^o quintil	3 ^{er} quintil	4 ^o quintil	5 ^o quintil	
1 ^{er} quintil	210 (0.074)	132 (0.047)	86 (0.030)	77 (0.027)	61 (0.022)	566 (0.2)
2 ^o quintil	120 (0.042)	121 (0.043)	127 (0.045)	110 (0.039)	87 (0.031)	565 (0.2)
3 ^{er} quintil	89 (0.031)	127 (0.045)	112 (0.04)	127 (0.045)	111 (0.039)	566 (0.2)
4 ^o quintil	67 (0.024)	97 (0.034)	135 (0.048)	131 (0.046)	135 (0.048)	565 (0.2)
5 ^o quintil	80 (0.028)	88 (0.031)	106 (0.037)	120 (0.042)	172 (0.061)	566 (0.2)
Total	566 (0.2)	565 (0.2)	566 (0.2)	565 (0.2)	566 (0.2)	2828 (100%)

Nota: $\chi^2 = 224.741$.

Como hemos comentado previamente, si los mercados financieros fueran independientes, veríamos una distribución uniforme a lo largo de la tabla con un

valor próximo al 0.04 de frecuencia relativa en cada celda. Sin embargo, como se puede observar, este no es el caso para nuestros datos en ninguna de las tablas analizadas. Esto es un claro indicativo de que existen relaciones de dependencia entre todos los mercados financieros analizados. Podemos observar, además, que los valores más grandes y las frecuencias relativas superiores a 0.04 se concentran en torno a la diagonal central de cada tabla, lo cual implica una relación de dependencia entre las variables positiva, puesto que indica que los dos mercados analizados en cada tabla se mueven en sintonía, cuando la rentabilidad de uno toma valores en un quintil concreto, la rentabilidad del otro tiende a estar en ese mismo quintil o en los quintiles contiguos. Por otra parte, las frecuencias más elevadas se observan en los extremos de la diagonal principal, es decir, siempre entre los dos quintiles extremos, el primero y el quinto. Eso significa que la dependencia entre los mercados se debe fundamentalmente a la dependencia entre los valores extremos.

Además, vemos como todos los valores del estadístico de contraste chi cuadrado están muy por encima de los valores críticos para los niveles de significación habituales, lo cual refuerza la conclusión de que los mercados financieros no son independientes, sino que están muy interrelacionados entre sí.

Si analizamos las tablas por separado podemos observar como la que contiene una distribución conjunta más uniforme sobre las 25 celdas, con frecuencias relativas próximas a 0.04, es la tabla 7 que recoge la relación entre los índices NASDAQ100 y NIKKEI225, lo cual concuerda con los resultados de la tabla 4, puesto que estos dos mercados presentan los menores coeficientes de correlación y asociación. El hecho de que aparezcan valores semejantes en todas las casillas indica que no existe una gran relación de dependencia entre las variables puesto que el hecho de que una esté en un quintil no parece indicar que la otra tenga que estar en otro concreto (lo cual indicaría una dependencia); véase también la nube de puntos correspondiente en el gráfico 4, en consonancia con este resultado.

Los índices IBEX35 y EUROSTOXX50 son los que más concentración muestran en la diagonal, con frecuencias muy elevadas en la diagonal principal indicando que ambos índices bursátiles se mueven juntos, es decir, aumentos o

disminuciones en uno, suelen ir acompañados de un cambio similar en el otro. Este resultado de nuevo es concordante con lo obtenido en la tabla 4 puesto que el IBEX35 y el EUROSTOXX50 son los dos índices que tenían mayor dependencia independientemente del coeficiente de correlación utilizado.

Si comparamos los valores de χ^2 vemos que también reflejan como más dependientes a IBEX35 y EUROSTOXX50, al devolver el valor más elevado, mientras que los menos dependientes de los estudiados serían NASDAQ100 y NIKKEI225, puesto que su valor χ^2 es el más pequeño.

A la vista de estos resultados, los de la tabla 4 y los del gráfico 4, podríamos concluir, a modo de resumen, que los índices NASDAQ100 y NIKKEI225 son los menos dependientes contemporáneamente mientras que IBEX35 y EUROSTOXX50 son los más dependientes. Esta última conclusión es algo que suponíamos verdadero puesto que el IBEX35 está relacionado con el mercado español y el EUROSTOXX50 con el mercado europeo, los cuales deberían estar en mayor sintonía que las otras combinaciones de mercados internacionales debido a su proximidad, tanto geográfica como cultural e institucional. Otro aspecto a tener en cuenta son los husos horarios ya que los índices con husos horarios más próximos son los más dependientes. Esto se debe a que un “shock” que cause grandes pérdidas o ganancias en un mercado en un determinado día, como, por ejemplo, la caída de Lehman Brothers, puede afectar en ese momento a dicho mercado mientras el otro ya ha cerrado y registrará el impacto al día siguiente. En este caso dicho suceso quedaría registrado como un aumento o disminución del precio de apertura del día siguiente. Al realizar un análisis contemporáneo perdemos la capacidad de analizar las relaciones dinámicas entre ambas variaciones del precio de cierre al tratarse de dos días distintos para los índices bursátiles. Este efecto se intensifica aún más si tenemos en cuenta que grandes caídas o subidas suelen ir acompañadas de un “efecto rebote” que provoca el efecto contrario en los días posteriores llegando a poder analizar una relación entre la bajada de uno y la subida del día siguiente en el otro mercado. Esto es especialmente importante entre los mercados japoneses y estadounidense ya que son los que tiene una mayor diferencia horaria.

Del resto de mercados analizados, podemos decir que se mueven entre los dos extremos anteriormente mostrados y que las conclusiones extraídas de las tablas de quintiles y la tabla 4 de coeficientes de correlación y asociación son similares.

4. CONCLUSIONES

En el presente trabajo hemos estudiado la capacidad del coeficiente de correlación lineal de Pearson para cuantificar las dependencias existentes entre las rentabilidades diarias de 4 índices bursátiles, concretamente, entre un europeo (EUROSTOXX50), un español (IBEX35), un estadounidense (NASDAQ100) y un japonés (NIKKEI225), a lo largo del periodo 2010-2020.

Para ello, en primer lugar, hemos analizado las limitaciones de dicho coeficiente definiendo qué características de los datos a analizar pueden causar que el valor de este coeficiente sea impreciso en la medida de la dependencia.

Después, hemos presentado y explicado algunas alternativas a este coeficiente disponibles en la literatura para tratar de mejorar la cuantificación de la dependencia entre los índices bursátiles permitiendo así recoger relaciones no necesariamente lineales entre variables no gaussianas. En concreto, hemos utilizado coeficientes de asociación basados en rangos, como el coeficiente de asociación de Spearman, y el τ de Kendall, basado en el concepto de la concordancia y algunos coeficientes de correlación robustos que tratan de minimizar el posible efecto pernicioso de los datos atípicos.

Posteriormente, se ha realizado un análisis de los datos recogidos. En primer lugar, se han analizado las características de las distribuciones de los rendimientos diarios de los índices bursátiles y se ha comprobado que no son gaussianas, lo que puede distorsionar la interpretabilidad del coeficiente de correlación lineal de Pearson. En segundo lugar, se han obtenido los valores de todos los coeficientes alternativos escogidos, así como el coeficiente de correlación lineal de Pearson, demostrando que éste muestra valores muy diferentes a los otros.

En concreto, hemos comprobado que el r_{COMED} , r_Q y T de Kendall devuelven valores similares entre sí y muy inferiores al coeficiente de correlación lineal de Pearson y que el r_{MAD} y el coeficiente de asociación por rangos de Spearman devuelven valores muy similares entre sí y ligeramente menores al de Pearson, aunque superiores a los otros tres.

Por consiguiente, podemos concluir que el coeficiente de correlación lineal de Pearson puede no ser un buen cuantificador del grado de dependencia existente entre los distintos índices bursátiles y sería recomendable usar uno de los otros propuestos como alternativa y comparar los resultados obtenidos.

A pesar de estas diferencias, cabe señalar que todos los coeficientes analizados indican que los dos mercados más relacionados son el IBEX35 y el EUROSTOXX50 y los dos menos relacionados son el NASDAQ100 y el NIKKEI225.

Por último, el análisis de dependencia con datos agrupados por quintiles reafirma las conclusiones previamente comentadas. En particular, muestra como todos los mercados financieros son dependientes entre sí y que dicha relación es positiva. Además, ordena a los pares de índices bursátiles según su grado de dependencia de igual modo que el análisis realizado a partir de los diferentes coeficientes de correlación y asociación: la relación más fuerte aparece entre los dos índices europeos (IBEX35 y EUROSTOXX50) y la relación más débil entre el mercado nipón y la americana.

5. BIBLIOGRAFÍA

Aparicio, F. y Estrada, J. (2001): "Empirical distributions of stock returns: European securities markets, 1990-95", *The European Journal of Finance*, 7:1, pp 1-21.

Cont, R. (2001): "Empirical properties of asset returns: stylized facts and statistical issues". *Institute of Physics Publishing*, 1, pp 223-236.

Drouet D. y Kotz S. (2001): *Correlation and dependence*, Imperial College Press, Londres.

Embrechts, P., McNeil, A., & Straumann, D. (2002). "Correlation and dependence in risk management: properties and pitfalls", en Dempster M. (Ed.), *Risk Management: Value at Risk and Beyond* Cambridge, University Press, Cambridge, (pp. 176-223).

Gnanadesikan, R. y Kettenring, J. R. (1972). "Robust estimates, residuals and outlier detection with multiresponse data". *Biometrics*, 28, 81-124

Hu, L. (2006) "Dependence patterns across financial markets: a mixed copula approach", *Applied Financial Economics*, 16:10, pp 717-729

Liebetreau A. (1983): *Measures of association*, Sage Publications, Beverly Hills.

Malevergne Y. y Sornette D. (2006): *Extreme financial risks. from dependence to risk management*, Springer, Berlín.

Peña, D. (2014): *Fundamentos de Estadística*. Editorial Alianza, Madrid.

Shevlyakov, G. y Smirnov, P. (2011): "Robust estimation of the correlation coefficient: an attempt of survey", *Austrian Journal of Statistics*, 40, pp 147-150.

Törnqvist, L., Vartia, P. y Vartia, Y. (1985): "How should relative changes be measured?", *The American Statistician*, 39:1, pp 43-46.

6. ANEXO

Para analizar cualquier cambio relativo existente podemos aproximarlos de muchas formas distintas. Una de ellas puede ser el cambio porcentual o el cambio logarítmico usado en este trabajo, pero también hay otros como medir el cambio respecto del valor final, por ejemplo. A la hora de decidir cuál de ellos utilizar podemos usar muchos aspectos de evaluación diferentes. No obstante, cuando queremos que puedan ser fácilmente operables matemáticamente, hay tres propiedades fundamentales que deben cumplirse: simetría, normalidad y aditividad.

Partimos de una función $H(x,y)$ que mide el cambio existente entre dos valores. Para nuestro caso serían el valor de cierre en t , x y el valor de cierre en $t+1$, y .

Para que esa función cumpla las propiedades se tiene que dar lo siguiente.

- Simétrico. $H(x,y) = -H(y,x)$.
- Aditivo. Si tenemos un cambio en dos etapas desde x hasta z pasando por y se dice que el medidor de la diferencia relativa es aditivo si

$$H(x, z) = H(x, y) + H(y, z)$$

- Normalizado. Se dice que $H(x, y)$ este normalizado si

$$\lim_{x \rightarrow y} \frac{H(x, y)}{H_1(x, y)} = \lim_{s \rightarrow 1} \left(\frac{H(x, y) - H(y, y)}{x - y} \right) = H'(1) = 1$$

donde $H_1(x, y) = (y - x) / x$.

Este parámetro prueba si la H generada se aproxima lo suficiente a la diferencia real entre x e y .

Si analizamos estos tres parámetros comprobamos que el único que los cumple es la diferenciación logarítmica puesto que la diferenciación porcentual es asimétrica y no aditiva. La última propiedad, la normalidad, la cumplen tanto la diferenciación porcentual usada generalmente como la diferenciación logarítmica escogida en este trabajo (Törnqvist, Vartia y Vartia 1985).