



EL COLEGIO DE MÉXICO CENTRO DE ESTUDIOS ECONÓMICOS

MAESTRÍA EN ECONOMÍA

TRABAJO DE INVESTIGACIÓN PARA OBTENER EL GRADO DE
MAESTRO EN ECONOMÍA

**EL MÉTODO KERNEL PARA ESTIMAR
FUNCIONES DE DENSIDAD: UNA
APLICACIÓN MULTIVARIADA**

EDGARDO ESAÚ LÓPEZ HERRERA

PROMOCIÓN 1996-1998

Asesor:

ENEAS A. CALDIÑO GARCÍA

JULIO DE 2000

Dedico este gran esfuerzo a Vanessa donde quiera que esté.

CONTENIDO

	PÁGINA
1. INTRODUCCIÓN	1
2. MÉTODOS DE ESTIMACIÓN NO PARAMÉTRICA	2
2.1 El histograma	2
2.2 El estimador kernel	3
2.2.1 La elección del kernel	4
2.2.2 Kernels de orden superior	5
2.2.3 La elección del parámetro de suavización	6
2.2.3.1 La escala normal	7
2.2.3.1.2 La regla de sobresuavización	8
2.2.3.2 Métodos high tech	8
2.3 Otros métodos de estimación	9
2.3.1 El estimador del vecino más cercano	9
2.3.2 El estimador kernel variable	9
2.3.3 Comentarios finales	10
2.4 Aplicación	10
3. ESTIMADOR KERNEL MULTIVARIADO	12
3.1 El estimador kernel multivariado	12
3.2 La elección del kernel multivariado	12
3.3 El parámetro de suavización H	13
4. APLICACIÓN PARA EL CASO BIVARIADO	15
4.1 Antecedentes teóricos	15
4.2 El modelo de mercado	15
4.2.1 Propiedades del modelo de mercado	16
4.3 Estimaciones	17
4.3.1 Datos	17
4.3.2 Consideraciones técnicas	18
4.3.3 Estimación de los rendimientos diarios	18
4.3.4 Rendimientos semanales	21
4.4 Implicaciones teóricas y prácticas	23
5. COMENTARIOS FINALES	25
APÉNDICE A: PROPIEDADES ESTADÍSTICAS DE LOS ESTIMADORES DE FUNCIONES DE DENSIDAD	
APÉNDICE B1: ESTIMACIONES DIARIAS	
APÉNDICE B2: ESTIMACIONES SEMANALES	

RESUMEN

La necesidad de modelos que expliquen y describan mejor la realidad, ha dado lugar a la búsqueda de técnicas que se adapten mejor a cada caso. Los modelos paramétricos comúnmente utilizados son muchas veces ineficaces para este fin, sin embargo resultan convenientes debido a su relativa simplicidad práctica. Desde hace algunos años se han desarrollado alternativas no paramétricas para modelar la realidad. Dentro de la gama de modelos descritos se encuentran aquéllos encaminados a estimar funciones de densidad, los cuales con los recientes avances en la tecnología computacional, se han convertido en una verdadera alternativa a los modelos paramétricos.

CAPÍTULO 1: INTRODUCCIÓN

El concepto de función de densidad (o de probabilidad) es uno de los más importantes en el ámbito de la estadística. Muchos de los análisis y desarrollos dentro de esta área del conocimiento, y en especial en las aplicaciones a otras ciencias como la economía y las finanzas, están basados en las ideas derivadas de dicho concepto. Es por esto que el estudio de este tema ha permanecido presente en la literatura, siempre buscando nuevos y mejores enfoques encaminados a proporcionar herramientas más precisas sobre todo para el proceso de inferencia y estimación.

Los recientes avances en la tecnología computacional, han tenido un impacto significativo en el desarrollo de investigación dentro de la ciencia estadística. En este sentido, el campo de la "no paramétrica" se ha beneficiado y ha incrementado su participación dentro de las herramientas modernas de análisis estadístico. Estas nuevas herramientas ofrecen alternativas sofisticadas a las comúnmente utilizadas dentro del ámbito de los modelos paramétricos, ya que permite explorar grandes cantidades de datos multivariados y univariados, sin hacer ningún supuesto en cuanto a la distribución de éstos.

Así pues, este trabajo presenta un análisis de las técnicas no paramétricas que se han utilizado para estimar funciones de densidad. También se presenta una aplicación multivariada, con el objetivo de motivar y entender la forma de implementar la técnica en la práctica, y las implicaciones que puede tener el utilizar modelos inadecuados, ante la falta de una alternativa viable.

Finalmente, la estructura del trabajo comprende un segundo capítulo en el que se presenta el caso univariado y se introduce la terminología propia de la técnica. En el capítulo tercero se generalizan las ideas al caso multivariado, para pasar así al cuarto capítulo que presenta la aplicación. Finalmente un quinto capítulo de comentarios finales.

CAPÍTULO 2: MÉTODOS DE ESTIMACIÓN NO PARAMÉTRICA

A lo largo de este capítulo se presentará una visión general de los métodos no paramétricos de estimación, poniendo especial atención en el método Kernel de estimación univariado. Se introducirá la terminología propia de la técnica, presentando además algunas de las propiedades estadísticas de los estimadores.

2.1 El Histograma

Para poder entender las ideas detrás de los métodos de estimación no paramétrica, resulta conveniente estudiar el más simple de estos métodos: el histograma.

Para esto considere alguna variable aleatoria X , a la cual se le desea estimar su función de densidad $f(x)$ a partir de una muestra aleatoria $x_i, i=1,2,\dots,n$. Si dicha variable es discreta, la estimación de $f(x)$ será simplemente $f(x) = n^{-1} \sum_{i=1}^n I(x_i = x)$, donde $I(x_i = x)$ es una función "indicador" que toma el valor de 1 si $x_i = x$ y cero en otro caso. En otras palabras, se obtiene la proporción de observaciones de la muestra que toman algún valor particular, y ese es el estimado de la probabilidad para ese valor.

Sin embargo para el caso en que X es una variable aleatoria continua, hacer una estimación igual a la anterior no tiene mucho sentido, ya que la probabilidad en un punto es igual a cero, así que se estima la densidad por intervalos. De esta manera el estimador propuesto sería $f(x) = (nh)^{-1} \sum_{i=1}^n I\left(x - \frac{h}{2} \leq x_i \leq x + \frac{h}{2}\right)$ donde se introduce el parámetro h , el cual indica la longitud del intervalo a estimar. Alternativamente, se puede replantear el estimado como

$$\begin{aligned} f(x) &= (nh)^{-1} \sum_{i=1}^n I\left(-\frac{1}{2} \leq \frac{x_i - x}{h} \leq \frac{1}{2}\right) \\ &= (nh)^{-1} \sum_{i=1}^n I\left(-\frac{1}{2} \leq \psi_i \leq \frac{1}{2}\right) \end{aligned}$$

donde, $\psi_i = \frac{x_i - x}{h}$.

Note como en realidad se está calculando la frecuencia relativa en cada intervalo centrado en x , en otras palabras, $f(x)$ es el valor de la ordenada del histograma en x . Además la función indicador en este caso opera como un ponderador de las observaciones alrededor de x , ya que toma el valor de 1 si la distancia de x_i con x es menor en términos absolutos a $\frac{1}{2}h$, y el valor cero en otro caso.

Así, este primer estimador se puede interpretar como un histograma basado en las observaciones cercanas o "locales" a x , donde el parámetro h juega un papel fundamental, ya que es la medida que controla que tan "suave" será el estimador. Si h es relativamente

grande, significa que algunas observaciones “lejanas” a x serán tomadas en cuenta al calcular la frecuencia relativa del intervalo centrado en este valor.

Del análisis anterior, es evidente que el valor de h es fundamental en la estimación mediante histogramas. El hecho de que el histograma sea altamente sensible al valor de h es también una de las principales debilidades de éste como estimador de la densidad de alguna variable aleatoria. El problema es que una misma serie de datos se puede ver completamente distinta al tomar diferentes valores de h ¹. Sin embargo la idea del parámetro de suavización será muy útil en los desarrollos que serán tratados a lo largo del trabajo, por lo que esta primera aproximación aunque simple, permitirá entender más fácilmente las implicaciones que puede tener el parámetro h .

También vale la pena resaltar el papel que juega la función indicador en este caso como ponderador. Dicha función le da un peso específico a cada una de las observaciones alrededor de x . En el caso del histograma, la función indicador opera como una función uniforme, ya que a todas las observaciones x_i que caen dentro del intervalo señalado anteriormente se les pondera con el valor de 1 por igual. Esta es otra de las debilidades del histograma como estimador de la función de densidad, ya que es demasiado “rudo”, en otras palabras, no es continuo, y esto dificulta su manejo estadístico en especial al intentar realizar algunos cálculos, como pueden ser las derivadas.

Así pues, de este breve análisis del histograma como estimador de una función de densidad se pueden rescatar dos ideas fundamentales: el ponderador, y el parámetro de suavización h .

2.2 El estimador Kernel

Como se comentó anteriormente, el histograma es un estimador que no resulta ser muy adecuado en sus características, ya que presenta saltos discretos en los puntos $x_i \pm \frac{h}{2}$, por lo cual es discontinuo. Este problema nos lleva a la necesidad de buscar un estimador más “suave”² el cual sea continuo y a la vez más preciso de la verdadera densidad de X . Así Rosenblatt propuso el siguiente estimador, llamado el estimador kernel.

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_i - x}{h}\right) = \frac{1}{nh} \sum_{i=1}^n K(\psi_i)$$

donde K es una función que satisface $\int K(x)dx = 1$, a la que llamaremos función kernel, o simplemente el kernel, y h es un número positivo que funciona como parámetro de suavización, el cual es función del tamaño de muestra n , y tiende a cero conforme $n \rightarrow \infty$.

¹ Ejemplos de este caso pueden verse en Wand and Jones 1995

² Se utiliza esta palabra como traducción de smooth

2.2.1 La elección del Kernel

La clase más simple de funciones kernel está compuesta por aquellas que satisfacen además de la condición de la integral, la condición $K(x) \geq 0$. Esta condición garantiza que el estimador $f(x)$ sea en sí mismo una función de densidad. Dentro de esta clase generalmente se selecciona K de tal manera que sea simétrica y unimodal, por ejemplo la distribución normal.

Sin embargo existen muchas funciones kernel que satisfacen estos requerimientos básicos, por lo que vale la pena preguntarse si algunas opciones son mejores que otras. En Wand y Jones (1995) se desarrolla a partir de criterios de error cuadrado medio y convergencia asintótica las condiciones que debe satisfacer el kernel canónico³. Dentro de esta clase de kernels canónicos, la versión más simple es

$$K^*(x) = \frac{3}{4} (1 - x^2) I_{|x| \leq 1}$$

que recibe el nombre de kernel de Epanechnikov⁴, al haber sido él quien caracterizó todas sus propiedades. La gráfica de este kernel se muestra en la siguiente figura.

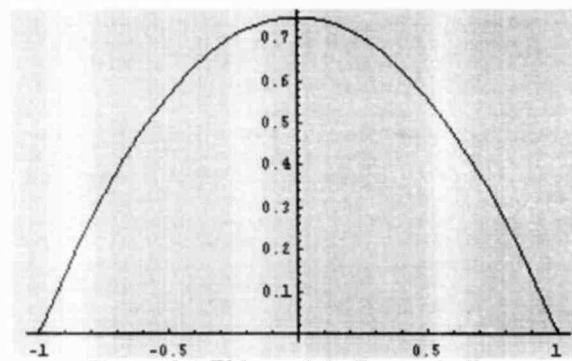


Figura 2.1 El kernel de Epanechnikov

Teniendo este kernel canónico como referencia se puede calcular una medida de eficiencia relativa entre éste y otros que satisfacen las mismas condiciones. En Wand y Jones (1995) se presenta la siguiente tabla que reporta la eficiencia relativa de kernels alternativos al Epanechnikov

KERNEL	EFICIENCIA RELATIVA
Epanechnikov	1.000
Normal	0.951
Triangular	0.986
Uniforme	0.930

Tabla 2.1 Eficiencia relativa de diferentes funciones que pueden utilizarse como kernel

³ El término canónico se refiere a que K^* minimiza el error cuadrático medio integrado sujeto a

$$\int K(x)dx = 1, \int xK(x)dx = 0, \int x^2 K(x)dx = a^2, K(x) \geq 0$$

⁴ Citado en Wand & Jones (1995)

El valor de la eficiencia relativa me indica el porcentaje de observaciones con las que el kernel canónico puede obtener el mismo valor mínimo para el error cuadrado medio integrado (ECMI⁵). Por ejemplo, para la normal cuya eficiencia relativa es de 0.951, se concluye que el kernel óptimo K^* puede obtener el mismo valor mínimo para ECMI con el 95% de los datos utilizados con el kernel Gaussiano.

Sin embargo, la principal conclusión que se puede extraer de la tabla 2.1 es que los kernels subóptimos en realidad no son tan ineficientes. Este resultado implica que no es tan importante qué función sea elegida como ponderador, así que se puede llevar la decisión a otros campos, como el de la eficiencia computacional.

2.2.2 Kernels de orden superior

Existe otra clase de funciones que pueden ser consideradas como kernels, y que no tienen que ser en sí mismas una función de densidad. Este tipo de kernels se conocen como kernels de orden superior, y su principal utilidad radica en que pueden reducir el sesgo muestral en la estimación de la función de densidad⁶.

Cuando la función K se restringe a ser una función de densidad, es necesario que $\mu_0(K) > 0$; en otras palabras que el segundo momento sea positivo. Sin embargo sin esta restricción es posible construir K de tal manera que $\mu_2(K) = 0$, lo cual tendrá el efecto de reducir el sesgo, y mejorar la tasa de convergencia del estimador kernel⁷.

Así se define que el kernel K es de orden j si dado

$$\mu_k = \int x^k K(x) dx,$$

$$\mu_0 = 1, \mu_k = 0 \forall k = 1, \dots, j-1, \text{ y } \mu_j \neq 0$$

Existen varias reglas para construir kernels de este tipo. Jones y Foster (1993) proporcionan la siguiente,

$$K_{j+2}(x) = \frac{3}{2} K_j(x) + \frac{1}{2} x^2 K_j(x)$$

donde K_j se refiere al kernel de orden L .

Con esta regla podemos definir un kernel superior de orden 4, utilizando uno de orden 2, por ejemplo la normal estándar (cuya notación es $\phi(x)$), así el kernel de orden 4, basado en $\phi(x)$ será,

$$K_4(x) = \left(\frac{3}{2} - x^2 \right) \phi(x)$$

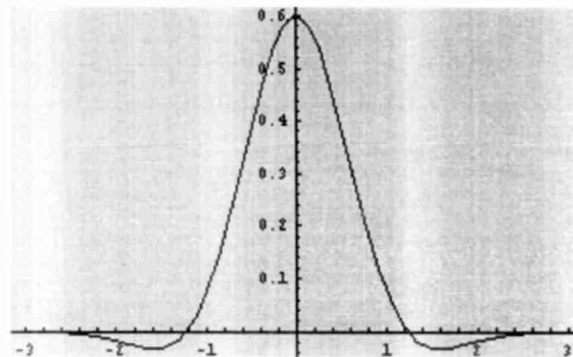
⁵ Ver apéndice A para abundar sobre esta medida de eficiencia.

⁶ Pagan & Ullah 1999

⁷ Wand & Jones 1995

La gráfica de este kernel superior se presenta en la siguiente figura, en la que es claro que no es una función de densidad, ya que toma valores negativos sobre algunos intervalos, sin embargo esto es necesario para que el segundo momento sea igual a cero.

Figura 2.2 Kernel superior de orden 4



Así pues se han comentado las cuestiones más relevantes respecto a la elección de la función kernel.

2.2.3 La elección del parámetro de suavización h

Como ya se comentó anteriormente, el parámetro de suavización es una medida de qué tan local será la estimación de la función de densidad en algún punto. Si este valor es relativamente pequeño, la estimación considera una vecindad limitada y puede generar una estimación poco suavizada, de tal manera que en el límite cuando $h \rightarrow 0$ el estimado será la unión de todas las observaciones. Por otro lado, si el valor de este parámetro es relativamente grande, genera estimaciones demasiado suavizadas, y pueden ser poco sensibles a cambios en las observaciones locales.

Así pues, la implementación práctica del estimador kernel requiere la especificación de un valor para h . En muchas circunstancias es posible escoger el valor de h simplemente por inspección visual, graficando varios estimados de f y seleccionando el valor que a nuestro criterio, sea el mejor. Sin embargo es conveniente contar con métodos automáticos, basados en los datos para determinar el valor del parámetro de suavización.

Para determinar el valor h de manera automática, existen dos clases de métodos: la primera clase está compuesta por métodos que son “fáciles y rápidos” de implementar en la práctica, ya que no contienen demasiados argumentos matemáticos, por lo que son desde el punto de vista computacional, fáciles de calcular. La segunda clase de métodos son los llamados “high tech”, ya que requieren un tratamiento matemático más complejo, y también requieren un esfuerzo computacional mayor que los métodos llamados fáciles.

2.2.3.1 Métodos Simples

Dentro de la clase de métodos automáticos simples para seleccionar el valor de h se considerarán dos en especial: la escala normal, y la regla de sobre-suavización.

2.2.3.1.1 La escala normal

Este método parte del valor de h que minimiza el ECMI asintóticamente. En Wand & Jones (1995), e Izenman (1991), se prueba que considerando el ECMI como criterio de optimización, derivando igualando a cero y despejando h , se obtiene que el valor de h que minimiza el ECMI asintóticamente (ECMIA) es:

$$h_{ECMIA} = \left[\frac{R(K)}{\mu_2(K)^2 R(f'')n} \right]^{1/5}$$

donde $R(K) = \int K(x)^2 dx$, $\mu_2(K) = \int z^2 K(z) dz$, f es la densidad que se desea estimar, K es el kernel, y n es el número de observaciones.

Si se considera f como la función de densidad normal con varianza σ^2 , aplicando su segunda derivada como argumento de R se obtiene que

$$h_{ECMIA} = \left[\frac{8\pi^{-1/2} R(K)}{3\mu_2(K)^2 n} \right]^{1/5} \sigma$$

A partir de esta última ecuación se puede estimar el valor h de escala normal, únicamente proponiendo un estimador para σ . A este respecto, un estimador lógico sería la desviación estándar de la muestra, sin embargo éste es muy sensible a las observaciones extremas, ya que al alejarse demasiado de la media su distancia al cuadrado se hace mayor. A este respecto, Silverman (1986) propone alternativamente que se utilice el rango intercuartil estandarizado,

$$\sigma_{RI} = \frac{\text{Rango_intercuartil_de_la_muestra}}{\Phi^{-1}(.75) - \Phi^{-1}(.25)}$$

donde $\Phi^{-1}(x)$ es la función que calcula los percentiles para la normal estándar, por lo que el valor del denominador de esta medida será el rango intercuartil de dicha densidad, cuyo valor aproximado a partir de tablas es 1.349.

Así pues, la regla general de selección entre el rango intercuartil estandarizado y la desviación estándar de la muestra, se reduce a escoger el menor de ellos de tal manera que el estimado h de escala normal se define como,

$$h_{EN} = \left[\frac{8\pi^{-1/2} R(K)}{3\mu_2(K)^2 n} \right]^{1/5} \min \left\{ S, \sigma_{RI} \right\}$$

donde se escoge el mínimo entre la desviación estándar y el rango intercuartil estandarizado para evitar con esto una sobresuavización y pasar por alto características importantes, como lo puede ser la existencia de varias modas.

2.2.3.1.2 La regla de sobre-suavización

La idea intuitiva de esta regla se basa en el llamado principio de máxima suavización introducido por Terrel (1990). Él prueba que el valor de h que minimiza el ECMIA está acotado por un valor máximo, de tal manera que

$$h_{ECMIA} \leq \left[\frac{243R(K)}{35\mu_2(K)^2 n} \right]^{1/5} \sigma$$

Así podemos estimar esta cota máxima simplemente sustituyendo el valor poblacional de σ por la desviación estándar de la muestra S , por lo que

$$h_{VS} = \left[\frac{243R(K)}{35\mu_2(K)^2 n} \right]^{1/5} S$$

A partir de este valor máximo se hace la estimación de la densidad, la cual de acuerdo a la manera en que se estima h estará seguramente sobreesuavizada. Sin embargo es un buen punto de partida para comenzar a utilizar el método de inspección visual, ya que a partir de este valor máximo se pueden ir probando valores más pequeños, hasta que encontremos de acuerdo con al criterio de inspección visual el valor adecuado para h .

Finalmente, vale la pena comentar sobre los casos en que estos métodos de fácil cálculo son útiles. El principio de máxima suavización es útil como ya se comentó para determinar un punto de partida para llevar a cabo la inspección visual. El método de la escala normal suele ser adecuado cuando la verdadera densidad se aproxima al menos en características a la normal: esto es, cuando la densidad a estimar es unimodal y más o menos simétrica, ya que la pieza clave de esta técnica es suponer que f es normal. También cabe mencionar que ninguna de estas técnicas garantiza por su simplicidad que el valor de h que se estime sea el que minimice el ECMIA, sin embargo resultan adecuadas por su simplicidad práctica.

2.2.3.2 Métodos high-tech

Para la implementación de este tipo de métodos, se requiere la utilización de algún criterio de optimización. Se puede utilizar el error cuadrado medio, el error cuadrado integrado, el error cuadrado medio integrado, el error integrado absoluto, en fin, debido a que se requiere alguna medida para calcular el valor óptimo para h , la complicación comienza desde la elección del criterio de optimización que se va a utilizar. En general las medidas de error cuadrado son un poco más simples que las medidas de error absoluto, al menos en su tratamiento matemático, por lo que suelen ser más usadas.

Dentro de los métodos más comunes se encuentran, la validación cruzada, los métodos "plug-in", la validación cruzada de máxima verosimilitud, por mencionar algunos. La idea de todos ellos en general, consiste en estimar la función f que invariablemente aparece en cualquier medida de optimización, y que es desconocida. Sin embargo no se abundará en el desarrollo de estos métodos en este trabajo, ya que en las aplicaciones que se presentarán posteriormente se recurre a los métodos de fácil cálculo.

Finalmente para concluir esta sección, vale la pena comentar acerca del “trade-off” que existe entre sesgo y varianza al elegir algún valor de h . Si h es pequeño la estimación tendrá muchos “picos”, los cuales al obtener varias muestras de f pueden aparecer en diferentes lugares, provocando que la varianza sea alta; sin embargo el sesgo será pequeño, ya que el estimado se adapta mucho más a los cambios en la muestra. Por otro lado, si h se incrementa, la variabilidad se reduce al ser el estimado más suave, sin embargo esto implica que sea menos sensible a cambios en la muestra, generando un mayor sesgo.

2.3 Otros métodos de estimación no paramétrica

En esta sección se comentarán de manera breve algunos métodos de estimación no paramétrica distintos al método kernel. Para esto seguiremos en especial a Izenman (1991), Pagan y Ullah (1999), Lofstgarden y Quesenberry (1965), Gessaman (1970).

2.3.1 El estimador del vecino más cercano

Este estimador está relacionado con el histograma, pero parte de la idea de que el parámetro de suavización debería ser mayor en las regiones alejadas de la media, es decir en las colas. La idea intuitiva de este estimador es que debido a que en las colas hay menos observaciones, las estimaciones serán demasiado pequeñas en esas regiones, por lo que el grado de suavizamiento se debe adaptar a la función de densidad que se está estimando.

Para poder plantear este estimador, se define $d(x_i, x)$ como la distancia del punto x_i al punto x , y para cada x el valor $d_k(x)$ representa la distancia entre x y su k -ésimo vecino más cercano de entre las observaciones de la muestra. Así definiendo $h = 2d_k(x)$ y aplicando el mismo razonamiento que en el histograma, se define

$$f(x) = \frac{k}{2nd_k(x)} = \frac{1}{2nd_k(x)} \sum_{i=1}^n I\left(\frac{x_i - x}{2d_k(x)}\right)$$

de tal manera que el grado de suavizamiento está controlado por k , ya que en las colas que es donde hay pocas observaciones $d_k(x)$ será mayor, y por lo tanto el valor de h también.

Acerca del valor adecuado para k generalmente se considera $k \approx n^{1/2}$ con lo que se tienen los elementos necesarios para construir este estimador.

Finalmente, al ser este en realidad un estimador similar al histograma, hereda los defectos de este último. En especial el problema de la discontinuidad permanece, aunque el problema de la variabilidad respecto a h disminuye, ya que como se comentó anteriormente el valor de h se adapta dependiendo la región de estimación en que se encuentre.

2.3.2 El estimador kernel variable

Este estimador surge a partir de las mismas ideas del estimador del vecino más cercano, sólo que en lugar de seleccionar el parámetro de suavización respecto a qué tan cerca está x de otras observaciones, se hace variar respecto a cada dato considerando alguna regla preestablecida. De esta manera el estimador kernel variable se define

$$f(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_{ii}} K\left(\frac{x_i - x}{h_{ii}}\right)$$

donde $h_m = d_k(x_i)$ es la distancia de la observación x_i respecto a su k -ésimo vecino más cercano de entre las $n-1$ observaciones restantes. El valor del parámetro k e puede aproximar como en el caso del estimador del vecino más cercano.

El hecho de que el valor de h_m varía con la muestra asegura que las regiones con pocas observaciones tendrán kernels mejor suavizados, la estimación será continua, y como se comentó en la sección referente al método kernel, si K es función de densidad, se garantiza que el estimador también lo será.

2.3.3 Comentarios finales

Existe otro tipo de métodos que utilizan ideas un tanto distintas que las que se han manejado hasta este momento para la estimación no paramétrica de funciones de densidad. Por mencionar algunos está el propuesto por Gessaman (1970) basado en bloques estadísticamente equivalentes. También se encuentran los estimadores basados en series ortogonales, trabajados por Watson (1970), los estimadores máximo verosímiles penalizados, etcétera.

Sin embargo el método kernel sobresale por su relativa simplicidad técnica y teórica, y las características de los estimadores encontrados mediante esta técnica, ya que además de ser continuos y diferenciables sobre todo el dominio, son en sí mismos funciones de densidad.

2.4 Aplicación

A continuación se presentará una pequeña aplicación de las ideas presentadas en este capítulo, poniendo especial atención en el efecto que tiene el parámetro de suavización en la estimación obtenida.

El caso que se presenta consiste en la estimación de la función de densidad para la variable aleatoria que describe los rendimientos semanales del índice DJ. Se cuenta con una muestra de 299 observaciones, y las estimaciones se hacen en E-Views.

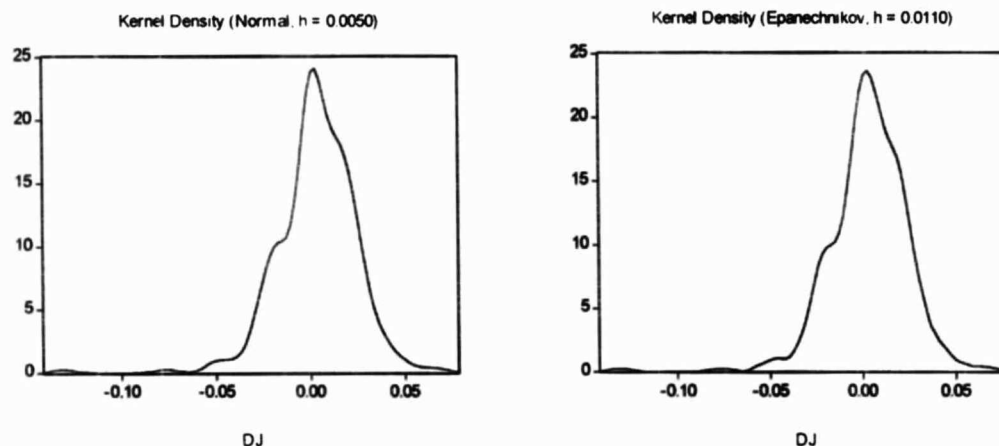


Figura 2.3 Estimaciones de la función de densidad del índice DJ, con el kernel normal, y Epanechnikov.

En estas gráficas se presentan las estimaciones de la función de densidad basadas en la muestra, pero utilizando diferente función kernel⁸. Como se puede notar, la elección del kernel no hace gran diferencia, ya que presentan básicamente las mismas características de curtosis, sesgo, media, etc. En otras palabras, tal y como se comentó anteriormente, la elección del kernel (dentro de la clase de los que son función de densidad) no implica cambios sustanciales en la estimación, ya que las estimaciones son casi idénticas.

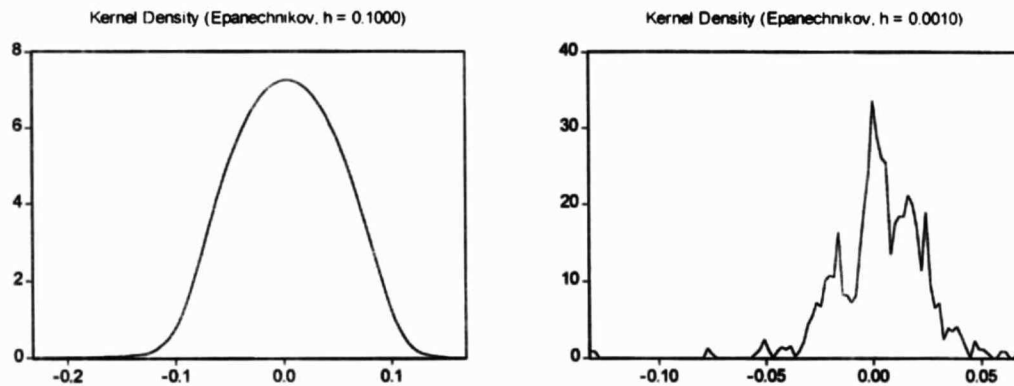


Figura 2.4 Estimaciones de la función de densidad, utilizando diferentes valores de h

En la figura 2.4 se presentan dos estimaciones basadas en la misma muestra, el mismo kernel (Epanechnikov), pero con diferente valor para h . Aquí es evidente el efecto que puede tener en la estimación utilizar tanto un valor demasiado grande (gráfico izquierdo), como un valor demasiado pequeño (gráfico derecho) para el parámetro h . Una está demasiado suavizada, la otra es demasiado “picuda”, mientras que en la figura 2.3 en la que el valor de h se calcula con el criterio de Silverman, se estima una función de densidad intermedia que no está en los casos extremos, y que por lo tanto es más eficiente.

⁸ E-views estima el valor de h utilizando el criterio de Silverman $h = 0.9kN^{-1/5} \min\{s, R_{1.34}\}$, donde s es la desv. Std, R es el rango intercuartil, y k es un factor de ajuste canónico dependiendo del kernel utilizado.

CAPÍTULO 3: ESTIMADOR KERNEL MULTIVARIADO

En este capítulo se extenderán las ideas principales de la estimación univariada mediante el método kernel, para el caso en el que se busca estimar una función de densidad multivariada, con el objetivo de desarrollar la aplicación bivariada del capítulo siguiente. Como veremos, no resulta ser muy problemático generalizar las ideas vistas anteriormente para el caso de la función kernel, sin embargo el parámetro de suavización si presenta complicaciones, sobre todo en cuanto al número opciones disponibles como medida de suavización que se pueden utilizar.

3.1 El estimador Kernel Multivariado

Nuevamente comenzamos planteando el contexto en el que se definirán las ideas a lo largo de este capítulo. Sea $\mathbf{X}_1, \dots, \mathbf{X}_n$ una muestra aleatoria d-variada con función de densidad a estimar f . Sea $\mathbf{X}_i = (X_{i1}, \dots, X_{id})'$ la notación utilizada para definir los componentes de \mathbf{X}_i , y el vector $\mathbf{x} \in \mathcal{R}^d$ la representación de $\mathbf{x} = (x_1, \dots, x_d)'$. También consideraremos por simplicidad $\int \mathbf{f}(\mathbf{x}) d\mathbf{x} = \int \dots \int_{\mathcal{R}^d} f(x_1, \dots, x_d) dx_1 dx_2 \dots dx_d$.

Así, en su forma general el estimador kernel d-dimensional, es el siguiente

$$\hat{f}(\mathbf{x}; \mathbf{H}) = n^{-1} \sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i)$$

$$\text{donde } K_{\mathbf{H}} = \mathbf{H}^{-1/2} K(\mathbf{H}^{-1/2} \mathbf{x})$$

es una función kernel d-variada que satisface

$$\int \mathbf{K}(\mathbf{x}) d\mathbf{x} = 1$$

y donde \mathbf{H} es una matriz simétrica, positiva definida de dimensión $d \times d$, llamada matriz de suavización.

Podemos ver que en la fórmula están las principales generalizaciones del caso univariado. Por un lado el kernel tiene que ser una función d-variada, y por otro el parámetro de suavización es ahora una matriz.

3.2 La elección del kernel multivariado

Las funciones kernel multivariadas pueden construirse a partir de las univariadas de varias maneras. Las dos formas más populares para este fin son: el producto de kernel

$$K^p(\mathbf{x}) = \prod_{i=1}^d k(x_i),$$

y el kernel esféricamente simétrico

$$K^s(\mathbf{x}) = c_{k,d} k\left[\left(\mathbf{x}^T \mathbf{x}\right)^{1/2}\right],$$

donde k es una función kernel univariada generalmente simétrica, y $c_{k,d}^{-1} = \int k\left[\left(\mathbf{x}^T \mathbf{x}\right)^{1/2}\right] d\mathbf{x}$.

Nuevamente una de las opciones más utilizadas como función kernel para este caso, es la función de densidad normal estándar d-variada

$$K(\mathbf{x}) = (2\pi)^{-d/2} e^{-\frac{1}{2}(\mathbf{x}^T \mathbf{x})}$$

de tal manera que $K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i)$ es la densidad normal multivariada, valuada en \mathbf{X}_i y con matriz de varianzas-covarianzas \mathbf{H} . Así, el kernel normal puede construirse a partir de funciones normales estándar univariadas usando tanto la técnica del producto como la esférica, ya que aunque en general $K^P \neq K^S$, para el caso de la normal ambos kernels son iguales.

Finalmente es posible llevar a cabo una discusión en la que se quiera determinar la eficiencia relativa de ambas técnicas. En general los kernels esféricos son más eficientes, sin embargo, como se prueba en Wand y Jones (1995) las diferencias no son sustantivas, sobre todo en el caso bivariado que es el que nos interesa en función de la aplicación que se presentará en el próximo capítulo.

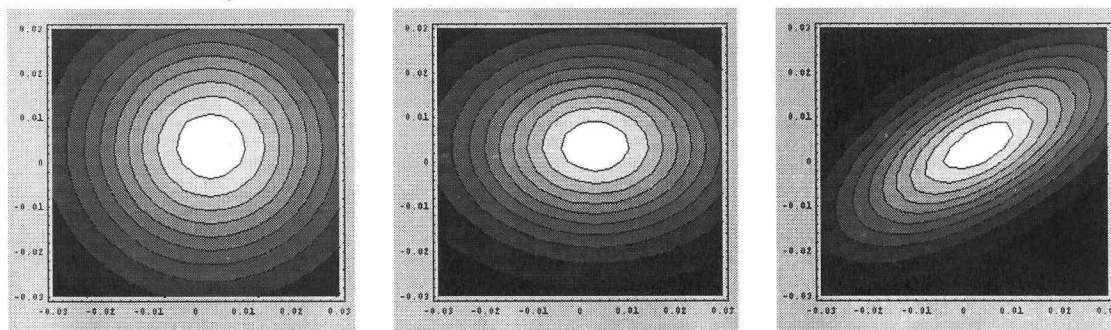
3.3 El parámetro de suavización \mathbf{H}

Como se comentó en la introducción del capítulo, la selección del parámetro de suavización no es tan simple como la del kernel multivariado, ya que al ser ahora una matriz cuadrada de dimensión d , el número de parámetros a elegir para construir la matriz es mucho mayor.

Existen varias posibilidades para definir \mathbf{H} . El caso más simple es aquél en que esta matriz es diagonal y con todas sus entradas iguales, por lo que en este caso únicamente se tiene que elegir un parámetro h , de tal manera que $\mathbf{H} = h^2 \mathbf{I}$ ⁹. Un caso menos simple es en el que la matriz es diagonal pero donde las entradas pueden ser diferentes, así para cada dimensión del vector \mathbf{x} se tendrá un parámetro de suavización diferente, de tal manera que $\mathbf{H} = \text{diagonal}(h_1^2, \dots, h_d^2)$. Finalmente, el caso más general es en el que \mathbf{H} puede ser cualquier matriz no diagonal, simétrica y positiva definida.

Para el caso en que $d = 2$ es fácil ver la diferencia entre los tres tipos de matrices de suavización. En la siguiente figura se presentan las curvas de nivel de tres kernels bivariados, en los que la matriz \mathbf{H} pertenece a cada uno de los tipos presentados anteriormente.

Figura 3.1 Diferentes valores para la matriz de suavización \mathbf{H}



⁹ Donde \mathbf{I} es la matriz identidad de dimensión d .

En el caso de la primera figura, la matriz es diagonal con los valores de h iguales, por lo que los diagramas de contorno son círculos concéntricos. En la segunda figura la matriz es diagonal, pero con diferentes entradas, por lo que se generan elipses con dirección hacia los ejes coordenados. Finalmente el tercer caso es el más general, en el que dados los parámetros especificados las curvas de nivel son elípticas, pero con una orientación diferente a la de los ejes coordenados.

Así pues, la elección de \mathbf{H} puede resultar algo muy complicado, ya que conforme aumenta la dimensión se incrementa el número de parámetros a elegir. En este sentido se ha desarrollado una serie de métodos encaminados a determinar \mathbf{H} basados en alguna de las ya mencionadas medidas de eficiencia. Sin embargo, se pueden utilizar las técnicas denominadas rápidas para estimar valores de h univariados, y tomar éstos como una buena aproximación de los valores que se deben considerar en la matriz \mathbf{H} . En especial los valores estimados por el criterio de Silverman son una buena aproximación, tanto en términos prácticos, como en función de la aplicación que se llevará a cabo en el capítulo 4.

CAPÍTULO 4: APLICACIÓN PARA EL CASO BIVARIADO

Durante el presente capítulo se aplicarán las ideas expuestas anteriormente para estimar la función de densidad de un vector de rendimientos bivariado. Aunque en el capítulo 2 ya se llevó a cabo una pequeña aplicación, en este capítulo se desarrollarán los aspectos teóricos y las implicaciones estadísticas que se pueden derivar de los hallazgos empíricos derivados de la estimación

4.1 *Antecedentes teóricos*

Uno de los principales objetivos de los modelos teóricos, es que puedan utilizarse para poder explicar y/o describir la realidad. Cuando un modelo es propuesto para describir una serie de datos, en general no se puede esperar que éste sea la representación exacta de la realidad que quiere explicar, sin embargo se debe buscar aquel modelo que sea más conveniente y útil para poder describirla.

En el caso de los rendimientos de activos financieros, en especial acciones e índices, se ha recurrido a la función de densidad normal para describir su comportamiento. Este modelo sin embargo, ha probado no ser una aproximación tan precisa como se puede pensar. Fama (1976) calculó las frecuencias relativas de rendimientos de las 30 acciones que componen el índice Dow Jones para algunos intervalos relevantes, y luego comparó sus resultados con las proporciones esperadas para la distribución normal. Sus conclusiones dejan de manifiesto dos fallas fundamentales de la distribución normal para describir la realidad: primero, se acumula una mucho mayor proporción de observaciones muy cerca de la media, y segundo, en las colas también se acumula más probabilidad de la esperada si los datos siguieran una distribución normal.

Pagan y Ullah (1999) estimaron con técnicas no paramétricas la densidad para los rendimientos mensuales de un índice en particular. Los resultados de su aplicación confirman lo que ya anteriormente Fama había establecido con su análisis de proporciones, esto es, la densidad estimada difiere de la normal en dos cuestiones básicas: las colas son muy anchas, y es demasiado picuda alrededor de cero. En otras palabras, hay muchas observaciones demasiado pequeñas y demasiado grandes para ser consistente con la normal. Pagan y Ullah (1999) también detectaron cierto grado de asimetría hacia los rendimientos positivos, resultado que tampoco concuerda con la característica de simetría de la normal.

4.2 *El modelo de mercado*

En esta sección se analizará el llamado modelo de mercado. Este análisis será apropiado para motivar la utilidad de la aplicación, y poder entender las implicaciones de los resultados a partir de la estimación. Para abundar sobre la construcción teórica de este modelo, se puede referir a Fama(1976) y Ross (1995).

Partimos considerando un vector de rendimientos $\mathbf{r} = (R_{1,t}, \dots, R_{n,t})$ compuesto por n activos, el cual tiene una distribución normal multivariada. Sabemos por propiedad de la normal, que cualquier combinación lineal de un vector con distribución multinormal,

también se distribuye como normal. Este resultado implica que la distribución de cualquier par de portafolios construidos a partir del vector \mathbf{r} se distribuye como normal bivariada. En especial podemos formar una par de portafolios, uno compuesto por un activo en particular con rendimiento R_i , y el otro portafolio compuesto por el resto de los activos, con rendimiento R_m ¹⁰, al que llamaremos portafolio de mercado, y que en conjunto forman el llamado modelo de mercado.

4.2.1 Propiedades del modelo de mercado

Ahora se analizarán algunas de las principales propiedades estadísticas del modelo de mercado, con el objeto de clarificar las implicaciones que tienen los supuestos básicos de dicho modelo en los resultados.

La idea fundamental de dicho modelo, es establecer una relación entre el rendimiento del portafolio formado por el activo i respecto de las realizaciones de un portafolio representativo del mercado m . Para esto es necesario obtener la expresión que defina la esperanza condicional del rendimiento del activo i dada alguna realización de la variable aleatoria R_m (a la que llamaremos r), por lo que se plantea la función de esperanza condicional.

$$E(R_i | r) = \int_{R_i} R_i f(R_i | r) dR_i$$

donde se utiliza la densidad condicional en lugar de la marginal, debido a que la esperanza que tratamos de calcular es condicionada a alguna realización de la variable aleatoria R_m . Ahora, retomando el supuesto de normalidad del vector que contiene los portafolios con rendimientos R_i y R_m , en Goldberg (1997) se prueba que la función de densidad normal bivariada, tiene función de esperanza condicional lineal, por lo que la función que relaciona a R_i dado $R_m = r$ es la función lineal,

$$E(R_i | r) = \alpha_i + \beta_i r$$

donde,

$$\beta_i = \frac{COV(R_i, R_m)}{V(R_m)},$$

$$\alpha_i = E(R_i) - \beta_i E(R_m).$$

Estos resultados coinciden con los de la regresión de R_i con R_m , pero sólo para el caso en el que la distribución de los rendimientos es normal.

Otro resultado importante que se deriva del supuesto de normalidad, es que la función de densidad condicional $f(R_i | r)$ se distribuye como normal, con media condicional

$$E(R_i | r) = \alpha_i + \beta_i r$$

y varianza condicional

$$V(R_i | r) = V(R_i) (1 - \rho_{im}^2),$$

¹⁰ El rendimiento de este portafolio se calcula con la combinación lineal utilizada para formarlo a partir del vector \mathbf{r} .

donde ρ_{mm}^2 es el coeficiente de correlación entre R_{ii} y R_{mm} , calculado de manera habitual.

De toda esta manipulación, se puede concluir que la varianza condicional no depende del valor que tome r , ya que tanto la varianza como el coeficiente de correlación son parámetros y tienen un valor dado y es constante.

Ahora podemos a partir de los cálculos hechos hasta ahora, encontrar una distribución teórica para el error.

$$e_{ii} = R_{ii} - \alpha_i - \beta_i r.$$

De aquí resulta casi evidente que el valor esperado del error condicionado a r es igual a cero. También, al ser la parte estimada por la regresión lineal determinística, la varianza condicional del error coincide con la varianza condicional de R_{ii} dado r . Nuevamente se concluye que dicha varianza condicional no depende del valor que tome r por lo que por propiedad de la distribución normal, se puede concluir que e_{ii} y R_{mm} son independientes.

En resumen, si la distribución conjunta del vector de rendimientos (R_{ii}, R_{mm}) es normal bivariada, la relación entre los componentes de dicho vector se puede expresar como

$$R_{ii} = \alpha_i + \beta_i R_{mm} + e_{ii}$$

donde e_{ii} y R_{mm} son independientes.

Finalmente, el modelo de mercado es uno de los modelos teóricos más utilizados en las finanzas para describir relaciones empíricas. De acuerdo a la manera en que se construyó, todos los resultados descansan sobre el supuesto fundamental de normalidad en el vector de rendimientos. También vale la pena comentar, que cada resultado implica a los demás. En otras palabras, si se utiliza el supuesto de que el error e_{ii} es normal, automáticamente se están implicando todos los demás supuestos. Si se establece que la relación entre los rendimientos de los portafolios es lineal, automáticamente se está asumiendo que el vector de rendimientos se distribuye como normal. Así pues, el supuesto de normalidad se encuentra esparcido en todas y cada una de las características del modelo de mercado, y obviamente en todos aquéllos derivados de éste.

4.3 Estimaciones

A partir de ahora se establecerán las condiciones bajo las cuales se hicieron las estimaciones, además de los resultados obtenidos no paramétricamente acerca de la función de densidad conjunta del vector compuesto por 2 portafolios: el Dow Jones y el Nasdaq.

4.3.1 Datos

Como se comentó anteriormente, para poner a prueba el modelo de mercado se utilizarán los índices del Nasdaq y Dow Jones. La descripción de la muestra utilizada se encuentra en la siguiente tabla:

NOMBRE DE LAS SERIES UTILIZADAS: Nasdaq y Dow Jones
PERIODICIDAD: Diaria
RANGO DE LA MUESTRA: 3-enero-1994 al 10-septiembre-1999
FUENTE: México Analítica

Tabla 4.1 Descripción de datos utilizados

La utilización de estos índices tiene una utilidad adicional. Ambas series se construyen como promedio ponderado de los valores de los activos que los componen. A la luz del teorema del límite central, que establece que en límite sin importar la distribución de la variable aleatoria en cuestión, la distribución de la media converge a la normal, es de esperarse que la distribución conjunta del vector formado por los rendimientos de los portafolios representados por estos índices, se distribuya como normal.

Se harán, además de las estimaciones a partir de los rendimientos diarios, las estimaciones considerando rendimientos semanales. Dichos rendimientos se calculan a partir de éstos utilizando promedios de cinco días. Nuevamente el teorema del límite central puede ser considerado como una expectativa de normalidad para este tipo de rendimientos, ya que ahora se estimará la densidad de la media de los promedios.

4.3.2 Consideraciones técnicas

Las estimaciones no paramétricas para ambos casos se hicieron en el software *Excel*^{VR}. Para esto fue necesario hacer una partición del dominio sobre el cual se hizo la estimación, para determinar los puntos en los cuales se evaluaría la función kernel. A este respecto, tanto en la estimación diaria como en la semanal, se parte el rango seleccionado en un total de 21^3 puntos equidistantes.

Una vez calculados los valores numéricos, o en otras palabras, los estimados para cada punto de la partición, se procedió a graficarlos en el software *Mathematica*^{VR}. En este sentido es necesario aclarar un punto técnico acerca de la manera en que se tiene que ingresar la información al software para que éste la grafique. Mathematica no admite coordenadas tridimensionales, por lo que no se puede introducir la coordenada (x, y) para hacer referencia a cada valor de z . En cambio se introducen arreglos rectangulares ordenados z_{ij} que representan las alturas para $z_{11}, z_{12}, \dots, z_{1n}$. Vale la pena aclarar esto, porque en algunas de las gráficas que se presentarán en las secciones siguientes el dominio de las estimaciones será de $1-21$ sobre cada eje coordenado. Sin embargo en realidad los valores de z que se grafican hacen referencia a los puntos de la partición sobre los cuales fueron calculados los valores estimados de la densidad.

4.3.3 Estimación de los rendimientos diarios

Una vez establecidas las condiciones técnicas necesarias para llevar a cabo las estimaciones, se implementarán las correspondientes a los rendimientos diarios. Lo primero

que se debe hacer es determinar el rango sobre el cual se llevarán a cabo. Para esto resulta útil analizar el diagrama de dispersión de los rendimientos diarios, para que a partir de la información proporcionada por éste se determine dicho rango.

Como se puede observa en la figura 4.1, la mayoría de las observaciones se encuentran en el rango comprendido de $-.02$ a $.02$, ya que a pesar de que la gráfica toma valores cercanos a $.10$ y $-.10$, éstos pueden ser considerados atípicos y pueden por lo tanto quedar fuera del rango de estimación. Finalmente, como se comentó en la nota técnica, la partición del intervalo relevante debe constar de puntos equidistantes, por lo que se fija una distancia de $.002$ entre cada punto a estimar.

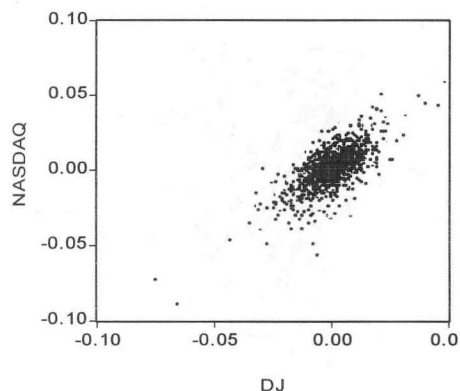


Figura 4.1 Diagrama de dispersión de los rendimientos diarios del Dow Jones y Nasdaq

Así pues, una vez seleccionado el rango de estimación, y los puntos sobre los cuales se harán las estimaciones, se presenta la gráfica resultante en la siguiente figura.

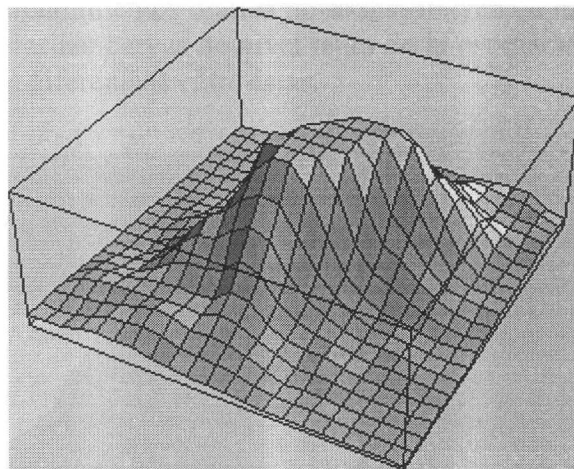


Figura 4.2 Estimación de la función de densidad para los rendimientos diarios

En la figura 4.2 se puede ver la gráfica de la función de densidad estimada para el caso en que los rendimientos se consideran diarios. Como función kernel se utilizó la normal bivariada, y como matriz de suavización \mathbf{H} se utilizó una matriz diagonal, en la que se encuentran los valores de h calculados con el criterio de Silverman.

La gráfica se distorsiona en las regiones cercanas al centro. Esto es debido a que los estimados en esta región son relativamente grandes respecto a los estimados en las regiones extremas, por lo que este resultado parece evidenciar que el estimado es demasiado puntiagudo en esta región. Para poder comprobar esta apreciación, resulta adecuado analizar la gráfica de la función normal bivariada, estimada a partir de los parámetros calculados a partir de los datos. La gráfica 4.3 que se presenta a continuación, corresponde a dicha densidad.

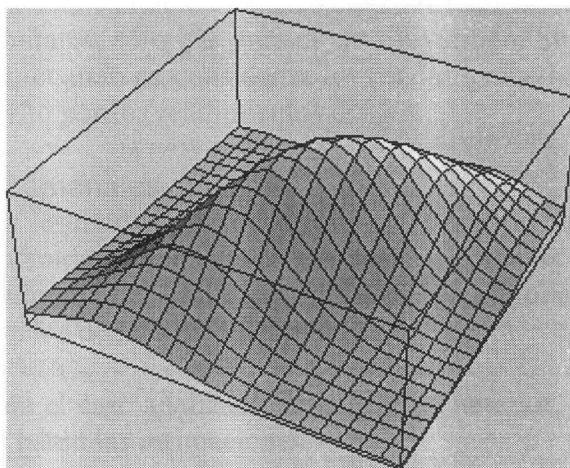
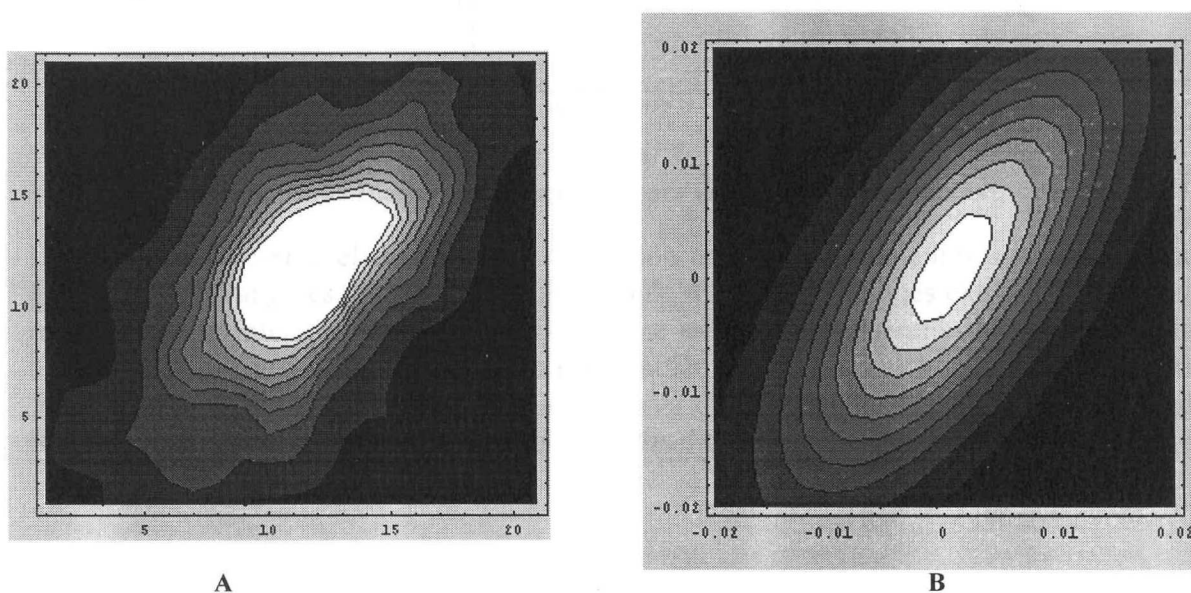


Figura 4.3 Función de densidad normal bivariada

Como se puede ver a partir de las gráficas 4.2 y 4.3 hay diferencias substanciales entre ellas. La densidad normal no se distorsiona en los valores centrales, ya que éstos no son tan grandes respecto a los valores extremos. Este breve análisis gráfico parece evidenciar sin duda alguna, que los rendimientos diarios no se distribuyen de manera normal. Vale la pena sin embargo, considerar las curvas de nivel tanto de la estimación como de la normal para tener muy en claro las diferencias entre éstas.



A
B
Figura 4.4 Diagramas de contorno de la estimación (A) y de la normal (B)

La figura 4.4 presenta los diagramas de contorno (curvas de nivel) tanto de la densidad estimada, como de la normal bivariada. En estas gráficas se puede ver que la probabilidad se distribuye de distinta manera, ya que en el estimado la región de mucha densidad es mayor que en la normal. También es evidente que fuera de la región de alta densidad, el estimado se desvanece rápidamente, mientras que la normal lo hace de una manera mucho más paulatina.

Así pues, parece que con este ejercicio empírico se confirman los hallazgos ya presentados en otras aplicaciones similares, sólo que para el caso bivariado. Sin embargo se analizarán estas mismas ideas, para cuando se consideran los rendimientos semanales, para ver si se puede confirmar lo descrito por los rendimientos diarios.

4.3.4 Rendimientos Semanales

Ahora se presentarán los resultados obtenidos para la estimación de la función de densidad para cuando los rendimientos se consideran semanales. Como se comentó en la sección relativa a los datos, el teorema del límite central incentiva a probar si los resultados de la estimación coinciden con los de la distribución normal.

De igual manera que para el caso analizado en la sección anterior, es necesario determinar el rango sobre el cual se harán las estimaciones.

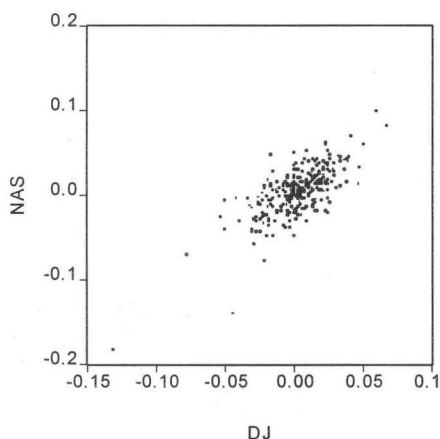


Figura 4.5 Diagrama de dispersión para los rendimientos semanales

La figura 4.5 presenta el diagrama de dispersión para los rendimientos semanales. Aquí nuevamente el rango es demasiado amplio debido a las observaciones extremas, así que lo reduciremos al intervalo de -0.03 a 0.03 , ya que en éste se encuentran la mayoría de las observaciones. La función kernel utilizada también fue la normal bivariada, y la matriz \mathbf{H} es la matriz diagonal que contiene los valores de h obtenidos mediante el criterio de Silverman, al igual que en el caso de rendimientos diarios.

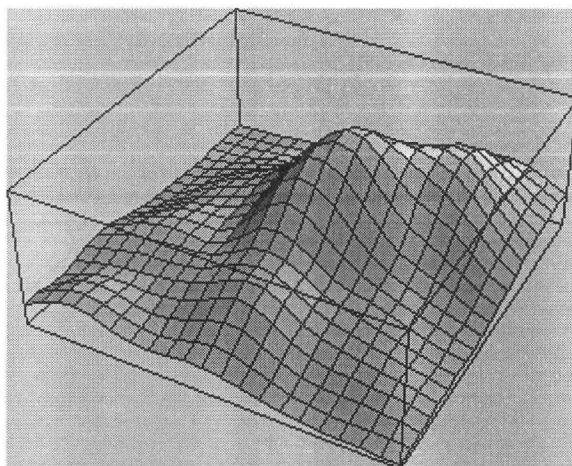


Figura 4.6 Estimación de la función de densidad con rendimientos semanales

La figura 4.6 presenta la estimación de la densidad en el caso que nos ocupa. Como se puede ver, esta estimación no presenta distorsiones como en el caso anterior, ya que es una estimación más suavizada al menos en apariencia. En otras palabras, el haber calculado los promedios fue una medida útil para estabilizar la forma de la densidad.

La pregunta que queda aún sin responder, es si esta estimación se aproxima a la normal. Para esto se presenta en la figura 4.7 la función normal bivariada estimada a partir de los parámetros calculados con la muestra.

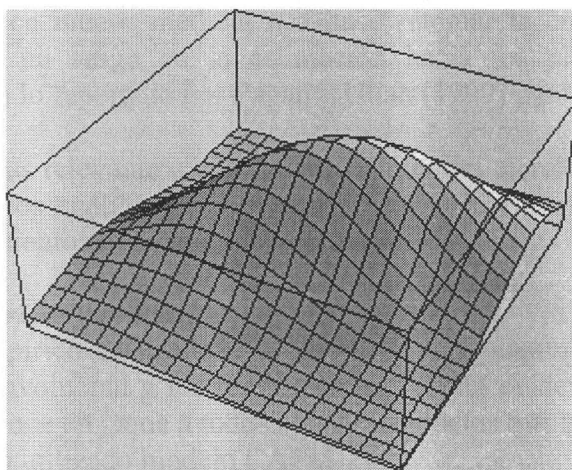


Figura 4.7 Función de densidad normal bivariada

Al analizar con detenimiento las figuras 4.6 y 4.7, parece lógico pensar que los rendimientos semanales se aproximan a la normal. Sin embargo debido a la forma en que se presentan las gráficas, puede haber algunas cuestiones que se escapen a la simple inspección visual, por lo que resulta conveniente analizar los diagramas de contorno de ambas figuras, y determinar a partir de ellos si efectivamente son coincidentes en su comportamiento.

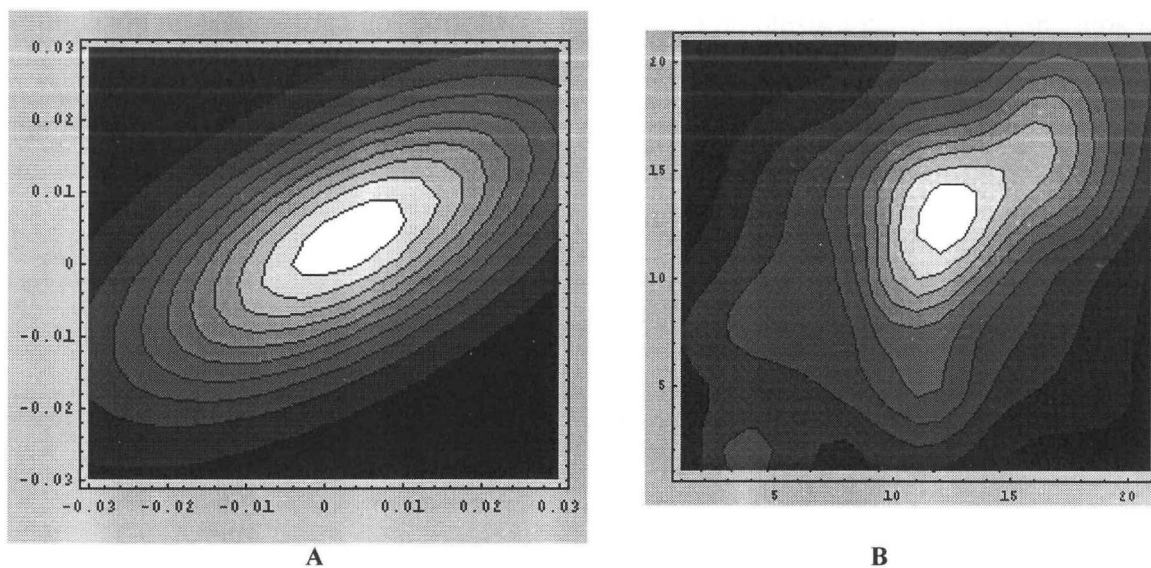


Figura 4.8 Diagramas de contorno de la normal (A), y de la estimación (B)

La figura 4.8 presenta los diagramas de contorno tanto de la figura 4.6, como de la 4.7. Nuevamente se pueden detectar diferencias importantes entre ambas densidades. Por ejemplo, en la estimación se acumula mayor proporción de la probabilidad en las regiones extremas (rendimientos muy positivos y muy negativos) que en la distribución normal. También en la región central de bajos rendimientos (cerca de cero), la estimación acumula o toma los valores más altos en una región más pequeña que en la normal (elipse central), lo cual aunque en menos medida, nos hace retomar la cuestión de la curtosis. Finalmente, es evidente un sesgo en la estimación hacia los rendimientos positivos, hallazgo que coincide con lo reportado por Pagan y Ullah (1999) para el caso univariado.

Así pues la conclusión más relevante que se puede extraer del ejercicio empírico llevado a cabo, es que a partir de los análisis gráficos realizados, la hipótesis de normalidad en la distribución de los rendimientos tiene que ser rechazada.

4.4 Implicaciones teóricas y prácticas

A partir del análisis empírico realizado a lo largo de este capítulo, y a la luz de las cuestiones teóricas que involucran a los rendimientos, resulta evidente que el rechazar la hipótesis de normalidad pone en jaque a todos aquellos modelos que se han basado en dicha premisa, como lo es el tan utilizado modelo CAPM.

A este respecto vale la pena recordar que en el tiempo en que se desarrolló el modelo de mercado, no se contaba con alternativas viables para llevar a cabo un análisis más preciso; esto es, un análisis basado en los datos y que permitiera que el proceso de inferencia estadística fuera menos manipulado a partir del supuesto de normalidad.

Hoy en día la posibilidad de llevar a cabo estimaciones no paramétricas, y con ellas cálculos precisos de derivadas, probabilidades, integrales y demás cuestiones relevantes, ya no es más un simple buen deseo, ya que a diferencia de todos aquellos análisis que

simplemente rechazaban la hipótesis de normalidad basados en cálculos de proporciones. la estimación no paramétrica nos permite rechazar dicha hipótesis a partir de argumentos más claros. pero sobre todo da una alternativa viable al proporcionar la función de densidad no paramétrica, y con ello el reto de redefinir los modelos y teorías basadas en supuestos limitantes de la realidad.

CAPÍTULO 5: COMENTARIOS FINALES

Durante el desarrollo de este trabajo, se ha intentado cumplir con dos objetivos primordiales: el primero presentar una visión no tan técnica, pero si formal de la metodología kernel para estimar funciones de densidad, y segundo, exponer mediante una aplicación con fundamento teórico y práctico, la utilidad que puede tener dentro de la ciencia económica la utilización de dichas técnicas con el objetivo de mejorar la manera en que se aplican las ideas estadísticas en la economía cuantitativa.

Queda de manifiesto que los recursos computacionales disponibles hoy en día, imponen un reto para redefinir toda una serie de teorías, ideas, conclusiones y enfoques empleados hasta ahora, y que ante la falta de alternativas viables habían quedado al margen de la literatura práctica en Economía. A este respecto vale la pena resaltar algo de trabajo reciente en este sentido como el llevado a cabo por Viviana Fernández (1999), en el que modela con un enfoque no paramétrico la estructura de tasas de interés en Chile. También podría citar el trabajo de Ridder y Van den Berg (1997) en el que proponen modelos de duración y búsqueda de empleo, estimando las funciones de densidad no paramétricamente.

Finalmente cabe mencionar que dentro del campo de la no paramétrica, existen técnicas encaminadas a la estimación de regresiones no paramétricas, estimación de derivadas, momentos condicionales, análisis de series de tiempo, modelos censurados, en fin toda una gama de alternativas disponibles dentro de este campo para desarrollar el análisis empírico a la luz de estas técnicas y sin necesidad de imponer una estructura paramétrica a priori.

REFERENCIAS

1. Fama, E. (1976), *Foundations of Finance*, Basic Books, Inc.
2. Fernández, V. (1999), "Un Enfoque No Paramétrico para Modelar la Estructura de Tasas de Interés en Chile", presentado en la reunión anual de la Sociedad Econométrica.
3. Gessaman, M (1970), "A Consistent Nonparametric Multivariate Density Estimator. Based on Statistically Equivalent Blocks", *Annals of Mathematical Statistics*, 41, 1344-1346
4. Goldberg, A. (1997), *A Course in Econometrics*, Harvard University Press.
5. Huang, C. y P. Crooke (1999), *Mathematics and Mathematica for Economists*, Blackwell Publishers.
6. Izenman, A. (1991), "Recent Developments in Nonparametric Density Estimation", *Journal of the American Statistical Association*, 86, 205-224.
7. Jones, M y Foster, P. (1993), "Generalized Jackknifing and Higher Order Kernels", *Journal of Nonparametric Statistics*, 3, 81-94.
8. Loftsgaarden, D y C. Quesenberry (1965), "A Nonparametric Estimate of a Multivariate Density Function", *Annals of Mathematical Statistics*, 36, 1049-1051.
9. Mood, A y F. Graybill (1963), *Introduction to the Theory of Statistics*, tercera edición, Mc Graw Hill
10. Pagan, A y A. Ullah (1999), *Nonparametric Econometrics*, Cambridge University Press
11. Perkulis, D. (1993), *Aplicaciones del Método Kernel Univariado en la Estimación de Densidades*, ITAM.
12. Ridder, G. y Van de Berg, G (1997), "Empirical Equilibrium Search Models", en *Advances in Econometrics: Theory and Applications*, Vol III, Cambridge.
13. Robinson, P (1988), "Semiparametric Econometrics: A Survey", *Journal of Applied Econometrics*, 3, 35-51.
14. Rosenblatt, M (1956), "Remarks on Some Nonparametric Estimates of a Density Function", *Annals of Mathematical Statistics*, 27, 832-837.
15. Ross, A., Westerfield, R. y Jaffe, J.(1995), *Finanzas Corporativas*, Irwin.
16. Silverman, B. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall.

17. Terrel, G (1990). "The Maximal Smoothing Principle in Density Estimation". *Journal of the American Statistical Association*, 85, 470-477.
18. Wand, M y M. Jones (1995), *Kernel Smoothing*, Chapman & Hall.
19. Watson, G. (1969), "Density Estimation by Orthogonal Series", *Annals of Mathematical Statistics*, 40, 1496-1498.

APÉNDICE A: PROPIEDADES ESTADÍSTICAS DE LOS ESTIMADORES DE FUNCIONES DE DENSIDAD

Como cualquier procedimiento estadístico, los estimadores no paramétricos son aceptables, sí y sólo sí poseen algunas propiedades deseables.

1. Sesgo

La primera propiedad deseable tiene que ver con el sesgo. Un estimador f de una función de densidad f es insesgado, si $\forall \mathbf{x} \in \mathbf{R}^d$, $E_f[f(\mathbf{x})] = f(\mathbf{x})$. Aunque existen estimadores insesgados para algunas distribuciones paramétricas como la normal o poisson, para las estimaciones no paramétricas, se tiene que establecer un criterio menos estricto, como lo es el sesgo asintótico. Así una estimación f de f es asintóticamente insesgada sí y sólo sí $\forall \mathbf{x} \in \mathbf{R}^d$, $E_f[f(\mathbf{x})] \rightarrow f(\mathbf{x})$, conforme $n \rightarrow \infty$.

2. Consistencia

Una propiedad más importante es la consistencia. La idea más simple de consistencia de un estimador de una función de densidad, es la consistencia débil. Un estimador es consistente débilmente si $f \xrightarrow{p} f$. En otras palabras, si f converge en probabilidad a f . Así un estimador es consistente de manera fuerte si la convergencia se mantiene casi con certeza.

Existen varias maneras de probar consistencia tanto débil como fuerte. Una de ellas es el enfoque L_2 . Si f es integrable, entonces el desempeño de f en algún punto del dominio se mide por el error cuadrado medio,

$$ECM(x) = E_f \left[f(x) - f(x) \right]^2$$
$$ECM(x) = Var[f(x)] + sesgo^2(x)$$

Si el $ECM(x) \rightarrow 0 \quad \forall x \in \mathbf{R}$, se dice que f es un estimador consistente débilmente en media cuadrada de f .

Un criterio más importante de desempeño del estimador, tiene que ver con cómo la curva generada por f se ajusta a f . Dicha medida es el Error Cuadrado Medio Integrado (ECMI), que se calcula a partir del ECM, pero integrando sobre todo el dominio de f , así

$$ECMI = \int_{-\infty}^{\infty} E_f \left[f(x) - f(x) \right]^2 dx$$

3. Propiedades estadísticas del estimador kernel

Utilizando criterios L_2 , y bajo condiciones de regularidad tanto en K como en f , en Izenman (1991) se prueba que si $h_n \rightarrow 0$ conforme $n \rightarrow \infty$, entonces el estimador kernel univariado es insesgado asintóticamente, y asintóticamente normal.

También se prueba que la expresión

$$h_n^{ECMI} = \alpha(K)\beta(f)n^{-1(d+4)}$$

donde $\alpha(K)$ depende únicamente del kernel, y $\beta(f)$ únicamente de f , minimiza de manera asintótica el ECMI. Más aún se prueba que $ECMI \rightarrow 0$ a la tasa $O(n^{-1(d+4)})$ mostrando claramente el problema de la dimensionalidad, ya que la tasa de convergencia es menor conforme aumenta el valor de d .

