VI Congreso Galego de Estatística e Investigación de Operacións
Vigo 5-7 de Novembro de 2003

ILUSIONES, ESPERANZAS, MEDIAS, BETAS Y VARIANZAS

Juan Salazar Clavel[1], Susana Iglesias Antelo[2]
[1]Département des Sciences Administratives
Université du Québec en Outaouais, Canadá
[2]Departamento de Economía Financeira e Contabilidade
Universidade da Coruña, España

RESUMEN

En el ámbito de la teoría del mercado de capitales es habitual emplear el método MCO a la hora de estudiar la relación entre rentabilidad y riesgo de los activos financieros. En este estudio se exponen motivos teóricos y se ilustra mediante el método de simulación de Monte Carlo que los resultados favorables de estos tests pueden ser debidos a un fenómeno puramente numérico.

1. INTRODUCCIÓN

En numerosas investigaciones empíricas realizadas en el ámbito de la teoría del mercado de capitales – empezando por la pionera de Miller y Scholes (1972) – se analiza la relación entre la rentabilidad media y el riesgo – medido por la desviación típica de la rentabilidad o por la beta – de los activos financieros (títulos y carteras) mediante regresión lineal MCO. Y si bien diversos autores han hecho un esfuerzo por poner de relieve los problemas estadísticos inherentes a este tipo de tests (por ejemplo, Gómez-Bezares, Madariaga y Santibáñez, 1995), y se han propuesto otros métodos alternativos más robustos para realizar dichos contrastes (MCG, MCP, MV,…), el recurso al método MCO sigue siendo habitual.

Las razones de esta obstinación pueden ser de lo más diverso y no es propósito de este trabajo profundizar en ellas; sin embargo, a nuestro juicio, dos podrían ser fundamentales. Por un lado, el que los métodos más sofisticados y precisos no carezcan de otros defectos y que los resultados obtenidos tras su aplicación no sean claramente mejores que los alcanzados mediante técnicas más sencillas y atractivas, y, por otro lado, que los problemas econométricos asociados a los tests hayan sido presentados tradicionalmente de una forma un tanto críptica y, por consiguiente, de difícil comprensión para el investigador financiero.

Como consecuencia, en este estudio hemos puesto el énfasis en el aspecto pedagógico del tema, tratando de ofrecer una visión lo más clara posible de los mencionados problemas, así como de su gravedad, puesto que en un número de ocasiones muy superior al que sería deseable pueden conducir a resultados erróneos.

El punto que nos interesa resaltar en este trabajo es que el modelo de regresión lineal simple no se presta al estudio de las relaciones entre el riesgo y el rendimiento debido a que varias de las condiciones necesarias para su utilización no son razonablemente respetadas. En particular, porque en las regresiones estudiadas no todos los términos de error tienen la misma varianza, porque esos términos de error suelen estar fuertemente correlacionados entre ellos, porque existe en general una correlación positiva entre la varianza de los términos de error y los valores de la variable independiente, y porque los términos de error no siempre obedecen a leyes de probabilidad cercanas a la normalidad. Estos factores, que actúan simultáneamente, pueden *crear la ilusión* de que la relación entre rendimiento y riesgo es significativa.

Estas cuestiones se abordan desde una perspectiva teórica en los apartados 2 y 3, mientras que en el apartado 4 se presenta una simulación diseñada para evaluar la magnitud de los problemas señalados.

2. LA REGRESIÓN DE LAS MEDIAS SOBRE CUALQUIER VARIABLE NO ALEATORIA

Supongamos que estuviéramos interesados en saber si existe una relación entre el rendimiento esperado de los títulos y una variable no aleatoria cualquiera, X, y que, para comenzar nuestro estudio, hubiéramos decidido censar todos los títulos que existen en el mercado, a fin de evitar los problemas de validez externa asociados a la utilización de datos provenientes de una simple muestra de títulos, y explorar la posibilidad de que la relación entre el rendimiento esperado y la variable X para el conjunto de todos los títulos *i* sea estadísticamente lineal, tal como se expresa en la siguiente ecuación:

$$E_i = \gamma_0 + \gamma_1 X_i + e_i \qquad (1)$$

Es decir, nos planteamos la pregunta ¿$\gamma 1 \neq 0$? Además, para calcular los valores de los parámetros γ_0 y γ_1 utilizaremos el método de los mínimos cuadrados ordinarios.

Como se puede constatar, en este marco general no hay lugar para el azar. Dado que nos estamos refiriendo a la población de todos los títulos que existen en el mercado, en la ecuación (1) sólo caben parámetros y constantes. Por consiguiente, si pudiéramos conocer el rendimiento esperado de cada título que existe en el mercado, así como los valores de las constantes X_i, podríamos determinar sin dificultad si existe o no una relación entre esas dos series de constantes.

Dicho esto, supongamos que en el curso de nuestro censo hubiéramos logrado determinar los valores de las constantes X_i pero no los de los rendimientos esperados de los títulos. En este caso podríamos admitir que los rendimientos observados periódicamente de cada título *i* constituyen realizaciones independientes de una misma variable aleatoria $\tilde{R}_i$ cuyo valor esperado es igual al rendimiento esperado del título *i.* Aceptando esta hipótesis, podremos considerar que las medias aritméticas de los rendimientos observados constituyen un conjunto de estimaciones no sesgadas de los rendimientos esperados de los títulos.

Así, si definimos $\tilde{\bar{R}}_i$ como la media aritmética de n_o rendimientos observados del título *i*, y sumamos luego el término $\tilde{\bar{R}}_i - E_i$ a cada lado de la ecuación (1), obtenemos una ecuación que pone en relación la media de los rendimientos observados con la variable no aleatoria X:

$$\tilde{\bar{R}}_i \;=\; \gamma_0 + \gamma_1 X_i + e_i + \tilde{\bar{R}}_i - E_i \qquad (2)$$

Designando

$$\tilde{\xi}_i = e_i + \tilde{\bar{R}}_i - E_i \qquad (3)$$

se puede presentar la ecuación (2) de manera más concisa :

$$\tilde{\bar{R}}_i \;=\; \gamma_0 + \gamma_1 X_i + \tilde{\xi}_i \qquad (4)$$

Obsérvese que las ecuaciones (3) y (4) son indisociables. Por consiguiente, cualquier hipótesis con respecto a una de ellas tiene repercusiones sobre la otra. Por ejemplo, si se examina la ecuación (3), resulta claro que la varianza de cada término de error individual en la ecuación (4) es igual a la varianza del *error de estimación de la media* ($\tilde{\bar{R}}_i - E_i$). Ahora bien, como se sabe, esta varianza es igual a la varianza del rendimiento del título *i* correspondiente, dividida por el número de observaciones (n_o):

$$\sigma^2(\tilde{\xi}_i) \;=\; \sigma^2(\tilde{\bar{R}}_i - E_i) \;=\; \sigma_i^2 / n_o \qquad (5)$$

Por consiguiente, si se supusiera que la varianza de los términos de error individuales que aparecen en la ecuación (4) fuera constante, se estaría suponiendo, implícitamente, que los rendimientos de todos los títulos existentes en el mercado tienen la misma varianza.

De manera similar, si se supusiera que los términos de error que aparecen en la ecuación (4) no están correlacionados entre sí, se estaría suponiendo que los rendimientos de todos los títulos del mercado tampoco lo están. En efecto, si se examina la ecuación (3) se deduce que la covarianza de dos términos de error individuales, $\tilde{\tilde{\xi}}_i$ y $\tilde{\tilde{\xi}}_j$, es igual a la covarianza de los rendimientos de los títulos *i* y *j*, dividida por el número de observaciones:

$$\sigma(\tilde{\tilde{\xi}}_i, \tilde{\tilde{\xi}}_j) = \sigma((\bar{\tilde{\tilde{R}}}_i - E_i)(\bar{\tilde{\tilde{R}}}_j - E_j)) = \sigma_{ij} / n_o \qquad (6)$$

Ahora bien, teniendo en cuenta las varianzas que aparecen en la ecuación (5), todo eso implica que la correlación entre dos términos de error individuales, $\tilde{\tilde{\xi}}_i$ y $\tilde{\tilde{\xi}}_j$, es idéntica a la correlación de los rendimientos de los títulos *i* y *j*:

$$\rho(\tilde{\tilde{\xi}}_i, \tilde{\tilde{\xi}}_j) = \rho_{ij} \qquad (7)$$

Por otro lado, si examinamos la ecuación (3) a la luz del teorema central del límite, nos daremos cuenta de que la forma de la distribución de cada término de error individual en la ecuación (4) depende de la forma de la distribución del rendimiento del título *i* correspondiente y del número de observaciones utilizadas para calcular la media. Por consiguiente, si tomáramos una, dos o tres observaciones para calcular la media de rendimiento de cada título y las distribuciones de los rendimientos tuvieran formas diferentes, resultaría un tanto forzado suponer que todos los términos de error en la ecuación (4) obedecen a leyes de probabilidad normales.

En resumen, llegados a este punto, podemos afirmar que las características de la distribución conjunta de los rendimientos de los títulos tienen un efecto importante sobre las características de la distribución conjunta de los términos de error de la regresión de las medias de rendimiento de los títulos sobre un conjunto de constantes X_i. Y, como se puede constatar fácilmente, la validez estricta de esta afirmación no reposa en realidad sino sobre una sola hipótesis crítica; a saber, que los rendimientos periódicamente observados sobre cada título *i* son realizaciones independientes de una misma variable aleatoria $\tilde{R}_i$ cuyo valor esperado es igual al rendimiento esperado del título *i*.

3. LA REGRESIÓN DE LAS MEDIAS SOBRE LAS DESVIACIONES TÍPICAS Y LAS BETAS DE LAS CARTERAS

Dadas las características de la distribución conjunta de los rendimientos de los títulos, se puede entonces afirmar que en toda regresión de medias de rendimiento de títulos sobre constantes X_i, la varianza de los términos de error no es constante, los términos de error no son independientes unos de otros y esos términos no obedecen necesariamente a leyes de probabilidad normales. Esto amenaza la validez interna de las pruebas estadísticas que se efectúan siguiendo los procedimientos propios del modelo de regresión lineal simple.

Pero, evidentemente, la violación de esas condiciones no crea el mismo tipo de problema en todos los casos. Por ejemplo, si la varianza de los términos de error no estuviera correlacionada con los valores de la variable independiente, aquélla podría parecer constante. En ese caso, la desigualdad de las varianzas y aun la correlación de los términos de error no crearían probablemente problemas muy graves. Sin embargo, en el otro extremo, si la variable independiente fuera la desviación típica (σ_i) de las tasas de rendimiento de los títulos, el problema de heterocedasticidad sería muy agudo. Y, en un contexto en el que las correlaciones entre los rendimientos de los títulos son bastante fuertes y mayoritariamente positivas, la utilización del modelo de regresión lineal simple haría aparecer relaciones significativas cuando no las hay. Y esto, con una frecuencia mucho más elevada que el nivel de significación que se pudiera haber elegido para efectuar las pruebas estadísticas.

Para comprender mejor cómo se produce ese fenómeno, pensemos en un período alcista en el mercado financiero. Es evidente que los títulos con rendimientos más volátiles tenderían a tener rendimientos más elevados que los títulos con rendimientos menos volátiles. Y si el alza fuera suficientemente pronunciada, es casi seguro que se llegaría a encontrar en ese período una relación significativa entre las medias y las desviaciones típicas de las tasas de rendimiento de los títulos.

En resumen, en la regresión de las medias sobre las desviaciones típicas de los rendimientos de los títulos, la varianza de los términos de error no es constante, los términos de error están correlacionados entre sí, la desviación típica de los términos de error está *perfectamente* correlacionada con la variable independiente, y los efectos de esos tres fenómenos se combinan, produciendo así la *ilusión* de que las relaciones estadísticas entre la media y la desviación típica – estimadas según las reglas y procedimientos del modelo de regresión lineal simple – son significativas.

Y resulta evidente que si se utilizara una variable fuertemente correlacionada con la desviación típica como variable independiente, los problemas serían prácticamente los mismos; es decir, que toda variable fuertemente correlacionada con la desviación típica puede producir, en principio, el mismo tipo de ilusión estadística, prácticamente con la misma frecuencia y la misma intensidad. Y una variable tal lo es la beta calculada a partir de datos muestrales, ya que por definición la beta de un título i es:

$$\beta_{iM} = \sigma_i \rho_{iM} / \sigma_M \qquad (8)$$

Dicha correlación ha sido ampliamente contrastada, sobre todo en el caso de carteras (véase, por ejemplo, Black, Jensen y Scholes, 1972).

4. ILUSTRACIÓN NUMÉRICA DE LA INTENSIDAD DE LOS PROBLEMAS

La intensidad de los problemas relacionados con los fenómenos que acabamos de describir se puede ilustrar mediante la utilización del método de simulación de Monte Carlo.

Nuestro modelo de simulación responde a las especificaciones siguientes:

1. Los rendimientos de los títulos son mensuales y normalmente distribuidos.
2. Se estudiará el comportamiento del sistema durante 1.000 períodos de 60 meses cada uno.
3. Todos los títulos tienen el mismo rendimiento esperado (a fin de tener la seguridad de que no existe ninguna relación entre el rendimiento esperado y cualquier otra variable).
4. El coeficiente de correlación de los rendimientos de cada dos títulos es igual a una constante ($\rho_{ij}=\rho$).
5. El número de títulos que existen en el mercado (n_t) es divisible por 10 a fin de poder formar 10 carteras que contengan el mismo número de títulos.
6. Todos los títulos que conforman una cartera tienen la misma desviación típica de rendimiento.
7. El índice del mercado, *m*, es un índice equiponderado.
8. En todas las regresiones, la variable independiente es la *verdadera* beta (β_{pm}) de las 10 carteras con respecto al índice del mercado.
9. En todas las regresiones, la variable dependiente es la media ($\tilde{\tilde{R}}_p$) de 60 rendimientos obtenidos por cada una de las 10 carteras.
10. Se aplicarán las reglas y procedimientos propios del modelo de regresión lineal simple.
11. La hipótesis nula indica que no existe ninguna relación entre el rendimiento esperado y la beta de las 10 carteras.
12. El nivel de significación de las pruebas estadísticas es del 5%.

Nótese que en el sistema que acabamos de describir, la hipótesis nula es verdadera por construcción. Por consiguiente, consideraremos que hemos alcanzado el objetivo de nuestra simulación si, al hacer funcionar repetidamente ese sistema, llegamos a rechazar la hipótesis nula en una proporción muy superior al 5% de las veces.

Cuando concebimos nuestro sistema, supusimos que el rendimiento de cada título i podía expresarse de la manera siguiente:

$$\tilde{R}_i = E_i + \sigma_i \rho_{ig} \tilde{G} + \sigma_i \sqrt{1-\rho_{ig}^2}\ \tilde{C}_i \qquad (9)$$

donde $\tilde{G}$ y $\tilde{C}_i$ son variables aleatorias independientes normales (0,1). Más precisamente, supusimos también que esas variables eran independientes de las variables $\tilde{C}_j$ correspondientes a otros títulos j.

Teniendo en cuenta esta última hipótesis, el coeficiente de correlación de los rendimientos de dos títulos, es igual al producto de los coeficientes de correlación de los rendimientos de los títulos con la variable $\tilde{G}$:

$$\rho_{ij} = \rho_{ig}\,\rho_{jg} \qquad (10)$$

Ahora bien, en el punto 4 hemos especificado que todos los coeficientes ρ_{ij} eran iguales a una constante, ρ. Por ello, hemos supuesto también que los coeficientes ρ_{ig} eran constantes:

$$\rho_{ig} = \sqrt{\rho} \qquad (11)$$

Luego, teniendo en cuenta todas las hipótesis antes mencionadas, concluimos que el rendimiento de cada cartera *p* podría expresarse de la manera siguiente:

$$\tilde{R}_p = E_p + \sigma_p \rho_{pg} \tilde{G} + \sigma_p \sqrt{1-\rho_{pg}^2}\ \tilde{C}_p \qquad (12)$$

donde $\tilde{G}$ ya ha sido definida anteriormente y $\tilde{C}_p$ es una variable aleatoria normal (0,1) completamente independiente de $\tilde{G}$ y de las variables $\tilde{C}_q$ correspondientes a las otras carteras *q*.

Finalmente, teniendo en cuenta que las distribuciones de los rendimientos de los títulos y de las carteras son normales, decidimos utilizar la ecuación siguiente para generar las medias de los rendimientos de las carteras:

$$\tilde{\tilde{R}} = E_p + \sigma_p \rho_{pg} \tilde{\tilde{G}} + \sigma_p \sqrt{1-\rho_{pg}^2}\ \tilde{\tilde{C}}_p \qquad (13)$$

donde $\tilde{\tilde{G}}$ y $\tilde{\tilde{C}}_p$ son variables aleatorias normales (0,1/60), independientes entre sí y respecto a todas las variables $\tilde{\tilde{C}}_q$ correspondientes a las otras carteras *q*.

En cuanto a los datos, en primer lugar hay que mencionar que el rendimiento esperado de las carteras fue fijado a cero. Este valor fue escogido para simplificar los cálculos. Con todo, cualquier otro valor habría sido igualmente apropiado, pues el valor de ese parámetro no tiene ningún impacto sobre las pendientes de las regresiones y, por consiguiente, sobre los resultados de las simulaciones.

El coeficiente de correlación de los rendimientos de los títulos ($\rho_{ij}=\rho$) se fijó inicialmente en 0,16 y el número de títulos en el mercado (n_t), en 860. Esos valores son críticos y se verán modificados cuando llevemos a cabo el análisis de sensibilidad.

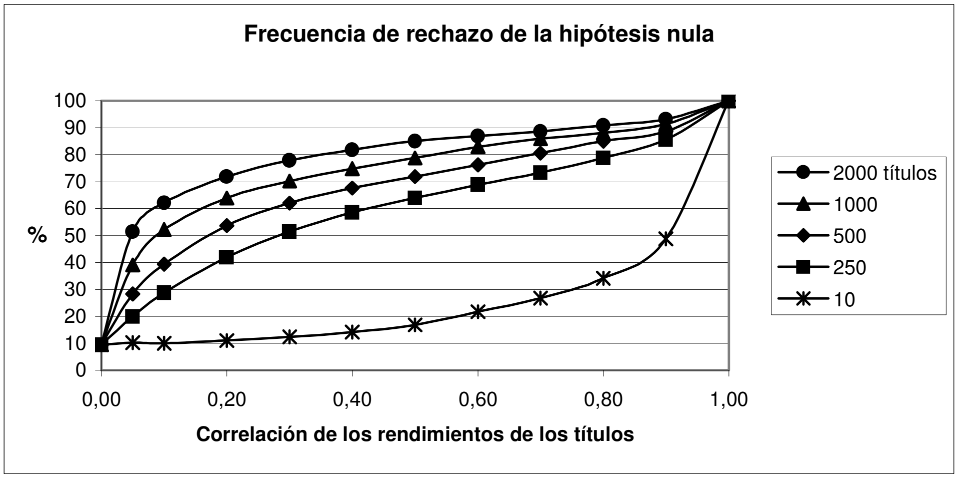
En cuanto a los valores de las desviaciones típicas de los rendimientos de los títulos, fueron escogidos a fin de que algunos de los parámetros de nuestras carteras se pareciesen, en la medida de lo posible, a los de las carteras del célebre estudio de Black, Jensen y Scholes (1972). Las desviaciones típicas de los rendimientos de los títulos (σ_i) en función de la cartera C_p a la que pertenecen y los parámetros de las carteras generados por nuestro modelo se recogen en el siguiente cuadro:

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
σ_i	0,1201	0,1422	0,1663	0,1874	0,2029	0,2306	0,2536	0,2733	0,3029	0,3507
ρ_{pm}	0,9708	0,9713	0,9720	0,9725	0,9729	0,9737	0,9743	0,9748	0,9756	0,9769
σ_p	0,0495	0,0586	0,0685	0,0772	0,0836	0,0950	0,1045	0,1126	0,1248	0,1445
β_{pm}	0,5369	0,6360	0,7439	0,8389	0,9089	1,0336	1,1376	1,2265	1,3604	1,5772

Finalmente, para calcular las medias simuladas generamos 11.000 números seudoaleatorios, que obedecen a una distribución normal (0,1/60), y consideramos que el 1º, el 12º,… y el 10.990º de la serie correspondían a realizaciones de la variable aleatoria $\tilde{\tilde{G}}$; que el 2º, el 13º,… y el 10.991º correspondían a realizaciones de la variable aleatoria $\tilde{\tilde{C}}_1$; y así sucesivamente.

Como resultado de la simulación se obtuvo un rechazo de la hipótesis nula el 58,2% de las veces. Recuérdese que la hipótesis nula era verdadera por construcción.

En el análisis de sensibilidad repetimos todo el proceso varias veces, modificando el valor del coeficiente de correlación de los rendimientos de los títulos y el número de títulos en el mercado. Los resultados obtenidos aparecen en el gráfico siguiente:



Como se puede constatar, la *probabilidad empírica* de cometer un error de tipo I es superior a 5% en todos los casos, inclusive cuando el coeficiente de correlación entre los rendimientos de los títulos es igual a cero. En este caso, la diferencia entre la frecuencia con que la hipótesis nula ha sido rechazada en el curso de las simulaciones (más o menos el 10%) y el nivel de significación elegido para las pruebas estadísticas (5%) puede ser atribuida a fluctuaciones aleatorias y a la desigualdad de varianzas. Pero, obviando las fluctuaciones aleatorias, se puede decir que los resultados que se observan en la parte central del gráfico ilustran los efectos combinados de la desigualdad de varianzas y de la falta de independencia de los términos de error de las regresiones de las medias sobre las betas de las carteras. Es decir, que, cuando se utilizan valores "realistas" para el coeficiente de correlación de los rendimientos de los títulos y para el número de títulos, la probabilidad de cometer un error estadístico de tipo I es aproximadamente igual a la probabilidad de ganar en un juego de cara o cruz.

5. CONCLUSIONES

Basándonos en los análisis que acabamos de presentar, aconsejamos tener mucha prudencia cuando se examinen los resultados de investigaciones empíricas sobre la relación entre el riesgo y el rendimiento. En particular, no hay que dejarse impresionar por el hecho de que algunos autores consideren que la relación entre las medias de rendimiento de las carteras y sus coeficientes beta es "sorprendentemente lineal". La linealidad de esta relación puede ser sin más el resultado de un fenómeno puramente numérico.

6. REFERENCIAS

Black, F.; Jensen, M. C. y Scholes, M. (1972): "The Capital Asset Pricing Model: some empirical tests". En Jensen, ed., *Studies in the Theory of Capital Markets*. Praeger, 79-124.

Gómez-Bezares, F.; Madariaga, J. A. y Santibáñez, J. (1995): "El CAPM: metodologías de contraste". *Boletín de Estudios Económicos*, 50, 557-582.

Miller, M. H. y Scholes, M. (1972): "Rates of return in relation to risk: a re-examination of some recent findings". En Jensen, ed., *Studies in the Theory of Capital Markets*, Praeger, 47-78.