José Antonio Duarte Mendieta

Temas Selectos de Estadística 21 de agosto de 2021

1. De acuerdo a la visto en clase desarrolle el álgebra para mostrar que la matriz de covarianza poblacional de R está dada por $\Sigma = \sigma_f^2 LL^t + \sigma_\epsilon^2 I_p$

R. Se tiene que

$$\Sigma = \frac{1}{n} RR^t \tag{1}$$

Usando el hecho de que $R = LF^t + \epsilon$ es posible obtener que

$$\begin{aligned}
\frac{1}{n}RR^t &= \frac{1}{n}(LF^t + \epsilon)(LF^t + \epsilon)^t \\
&= \frac{1}{n}(LF^t + \epsilon)((LF^t)^t + \epsilon^t) \\
&= \frac{1}{n}(LF^t + \epsilon)(FL^t + \epsilon^t) \\
&= \frac{1}{n}LF^tFL^t + \frac{1}{n}LF^t\epsilon^t + \frac{1}{n}\epsilon FL^t + \frac{1}{n}\epsilon\epsilon^t \\
&= L\frac{1}{n}F^tFL^t + \frac{1}{n}LF^t\epsilon^t + \frac{1}{n}\epsilon FL^t + \frac{1}{n}\epsilon\epsilon^t \\
&= L(\frac{1}{n}F^tF)L^t + L(\frac{1}{n}\epsilon F)^t + \frac{1}{n}\epsilon FL^t + \frac{1}{n}\epsilon\epsilon^t
\end{aligned} \tag{2}$$

Se tiene que

$$\frac{1}{n}F^tF = \sigma_f^2 I_k \tag{3}$$

$$\frac{1}{n}\epsilon F = 0 \tag{4}$$

Además, es posible observar que

$$\epsilon\epsilon^t = n\sigma_\epsilon^2 I_p \tag{5}$$

Utilizando (3), (4) y (5) en (2) se obtiene entonces que

$$\begin{aligned}
\frac{1}{2}RR^t &= L(\sigma_f^2 I_k)L^t + \sigma_\epsilon^2 I_p \\
&= \sigma_f^2 LL^t + \sigma_\epsilon^2 I_p
\end{aligned} \tag{6}$$

que es lo que se solicitó demostrar.

2. Obtenga los valores propios de Σ analíticamente como se hizo en las notas de clase.

R. Por el resultado del ejercicio 1 se tiene que

$$\Sigma = \sigma_f^2 LL^t + \sigma_\epsilon^2 I_p \tag{7}$$

y asumeiendo que $L \sim N(b11^t, \sigma_b^2 I)$, se tiene que la columna i (con $i \in \{1, 2, \ldots, k\}$) de L puede escribirse como:

$$L_i = b(1_p) + \sqrt{\sigma_b^2} e_i \tag{8}$$

considerando que $e_i \sim N_p(0, 1)$, lo que hace que la distribución de L sea la que se supuso.

Adicionalmente definimos la matriz X de la siguiente forma:

$$X = \begin{bmatrix} e_1 & e_2 & ... & e_k \end{bmatrix} \tag{9}$$

por lo que es posible definir la matriz L como

$$L = b1_p1_k^t + \sigma_b X \tag{10}$$

También es posible afirmar que LL^t es de tamaño pxp y L^tL es de tamaño kxk, y que ambas matrices tienen los mismos valores diferentes de cero (cuya cantidad se denotará por k), por lo tanto L^tL tambíen tiene k valores propios distintos de cero.

Desarrollando la multiplicación matricial L^tL, se obtiene lo siguiente:

$$\begin{aligned} L^tL &= (b1_k1_p^t + \sigma_b X^t)(b1_p1_k^t + \sigma_b X) \\ &= b^2 p(1_k1_k^t) + b\sigma_b(1_k1_p^t X) + b\sigma_b(X^t 1_p 1_k^t) + \sigma_b^2(X^t X) \end{aligned} \tag{11}$$

Para el segundo y tercer sumando, puede notarse que la entrada ij del producto matricial $1_k1_p^tX$ (y también la entrada de $X^t1_p1_k^t$) es la siguiente

$$\sum_{z=1}^{p} e_{zj} \tag{12}$$

Notando que $e \sim N_p(0, 1)$, cuando p tiende a infinito, la suma de valores muestrales en (12) va a tender a cero. por lo que las matrices $1_k1_p^tX$ y $X^t1_p1_k^t$ tenderán a la matriz cero (notar que este paso sólo aplica para valores grandes de p).

En el caso de la matriz X^tX se tiene que la entrada ij, está dada por $\sum_{z=1}^{p} e_{zi}^2$ si $i = j$ y $\sum_{z=1}^{p} e_{zi}e_{zj}$ si $i \neq j$. Recordando que $Var(e) = 1$ y que son independientes (i.e. $Cov(e_i, e_j) = 0$ cuando $i \neq j$), se tiene entonces que la matriz X^tX tendrá en la diagonal el valor p y 0 en las entradas fuera de la diagonal.

Con estas dos consideraciones, es posible continuar con el desarrollo en (11) de la siguiente forma

$$\begin{aligned} L^tL = b^2p(1_k1_k^t) + b\sigma_b(1_k1_p^tX) + b\sigma_b(X^t1_p1_k^t) + \sigma_b^2(X^tX) \\ = b^2p(1_k1_k^t) + \sigma_b^2pI_k \end{aligned} \tag{13}$$

Para poder obtener los valores propios de L^tL se puede utilizar la ecuación característica, es decir

$$|L^tL - \lambda I_k| = |b^2p(1_k1_k^t) + \sigma_b^2pI_k - \lambda I_k| = 0 \tag{14}$$

Si se define $\lambda^* = -\frac{\sigma_b^2+\lambda}{b^2p}$ es posible expresar (14) en términos de λ^* en lugar de λ, obteniéndose lo siguiente:

$$|L^tL - \lambda I_k| = |(1_k1_k^t) - \lambda^*I_k| = 0 \tag{15}$$

La identidad del determinante de Cauchy especifica que para toda matriz A que sea cuadrada y no singular (como lo es $-\lambda^*I_k$) y dos vectores x, y en C^n se cumple que

$$det(A + xy^t) = det(A) + y^t(adj(A))x \tag{16}$$

Si se identifica a A con $-\lambda^*I_k$ y a x con 1_k y y^t con 1_k^t se tiene entonces que

$$|(1_k1_k^t) - \lambda^*I_k| = |-\lambda^*I_k| + 1_k^tadj(-\lambda I_k)1_k \tag{17}$$

Como $-\lambda^*I_k$ es diagonal, es posible obtener que $adj(-\lambda^*I_k) = (-\lambda^*I_k)^{-1}|-\lambda^*I_k|$.
También puede observarse que $|-\lambda^*I_k| = (-\lambda^*)^k$ y que $(-\lambda^*I_k)^{-1} = -\frac{1}{\lambda^*}I_k$
Se tiene entonces que $adj(-\lambda I_k)1_k$ es lo siguiente:

$$1_k^tadj(-\lambda I_k)1_k = 1_k^t(-\frac{1}{\lambda^*}I_k)(-\lambda^*)^k1_k = \sum_{i=1}^{k} -\frac{1}{\lambda^*}(-\lambda^*)^k = k(\frac{(-\lambda^*)^k}{-\lambda^*}) \tag{18}$$

Sustituyendo (18) en (16) se obtiene entonces que

$$|(1_k1_k^t) - \lambda^*I_k| = (-\lambda^*)^k + k(\frac{(-\lambda^*)^k}{-\lambda^*}) = (-\lambda^*)^k(1 - \frac{k}{\lambda^*}) \tag{19}$$

Igualando a cero la última expresión en (19), se obtiene entonces que los valores propios de L^tL (que a su vez son los mismos valores propios que los de LL^t deben ser

$$\begin{aligned}\lambda_1 &= b^2pk + \sigma_b^2 p \\ \lambda_i &= \sigma_b^2 p \quad i = 2, \ldots, k\end{aligned} \tag{20}$$

Las veces que se presenta cada valor propio se obtienen de las potencias a las que están elevados los múltiplos en (19).

Con los valores propios de LL^t, es posible obtener los valores propios (desde el primero hasta el número k) de $\Sigma = \sigma_f^2 LL^t + \sigma_\epsilon^2 I_p$ de forma directa. Dichos valores propios serán:

$$\begin{aligned}\lambda_1 &= \sigma_f^2 p(b^2k + \sigma_b^2) + \sigma_e^2 \\ \lambda_i &= \sigma_f^2\sigma_b^2 p + \sigma_e^2 \quad i = 2, \ldots, k\end{aligned} \tag{21}$$

La matriz Σ puede escribirse como una matriz particionada con la siguiente estructura

$$\Sigma = \begin{bmatrix} \sigma_f^2 LL^t + \sigma_\epsilon^2 I_k & 0 \\ 0 & \sigma_\epsilon^2 I_{p-k} \end{bmatrix} \tag{22}$$

El determinante de una matriz particionada puede ser obtenido de la siguiente forma. Si M es una matriz particionada definida de la forma:

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

entonces el determinante puede ser obtenido:

$$|M| = |A - BD^{-1}C||D| \tag{23}$$

Utilizando entonces la estructura en (22), se tiene que el determinante de Σ puede obtenerse de la siguiente forma:

$$|\Sigma - \sigma I_k| = |\sigma_f^2 LL^t + \sigma_\epsilon^2 I_p||\sigma_e^2 I_{p-k} - \lambda I_{p-k}| = 0 \tag{24}$$

Para obtener entonces los valores propios de $k+1$ hasta p, puede notarse que el segundo múltiplo de (24) únicamente es cero si $\lambda = \sigma_e^2$ por lo que se obtiene que, utilizando este y los resultados obtenidos en la ecuación (20), es posible afirmar que los eigenvalores de Σ son:

$$\begin{aligned}\lambda_1 &= \sigma_f^2 p(b^2k + \sigma_b^2) + \sigma_e^2 \\ \lambda_i &= \sigma_f^2\sigma_b^2 p + \sigma_e^2 \quad i = 2, \ldots, k \\ \lambda_j &= \sigma_e^2 \quad j = k+1, \ldots, p\end{aligned} \tag{25}$$

3. Desarrolle los ejercicios propuestos en el taller visto en clase, dando respuesta a las siguientes preguntas:

- Describir metodologıa
- Anadir codigo y graficos con su respectiva descripcion y analisis
- Explicar comportamiento
- Sugerir posible solucion del fenómeno

R. La metodología para este ejercicio consiste en simular matrices de varianza covarianza y obtener (mediante simulación monte carlo) los valores propios de estas matrices para luego poder comparar los patrones que se presentan en la simulación contra los patrones que se presentan al utlizar las soluciones analíticas obtenidas en el ejercicio 2.

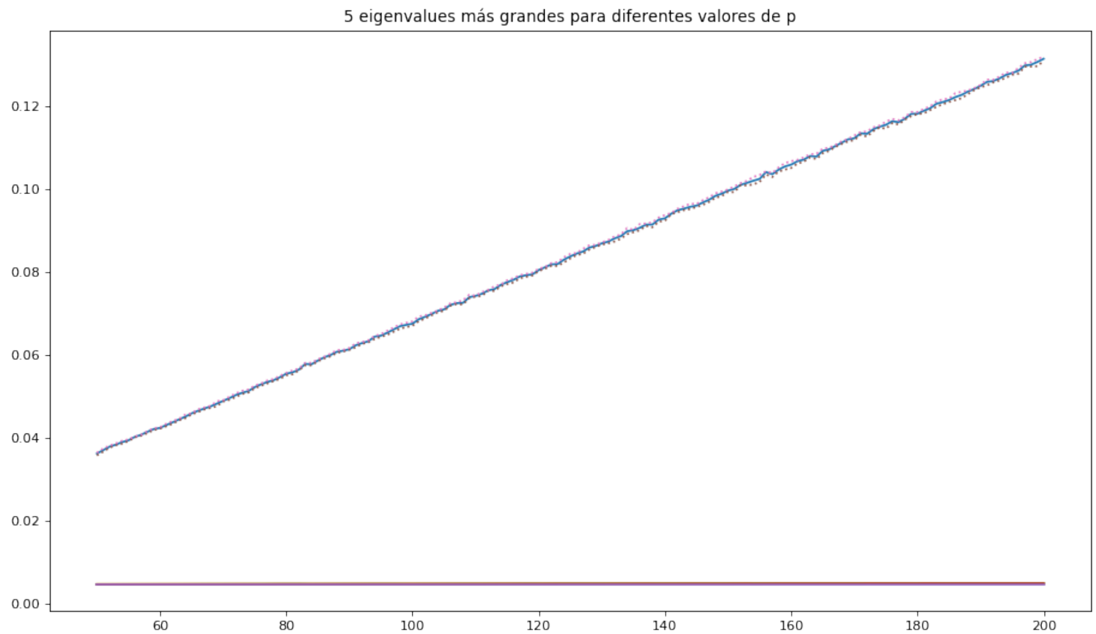
Las simulaciones se realizan considerando la siguiente igualdad :

$$\Sigma = \sigma_f^2 LL^t + \sigma_\epsilon^2 I_p$$

Se toman valores de p de 50 al 200, y para cada uno de estos valores de p se realizan 100 estimaciones de Σ, se obtienen los valores propios para cada estimación, se seleccionan los 5 valores propios más grandes y al final se promedian los valores propios de las 100 estmiaciones para estimar los valores propios de cada p.

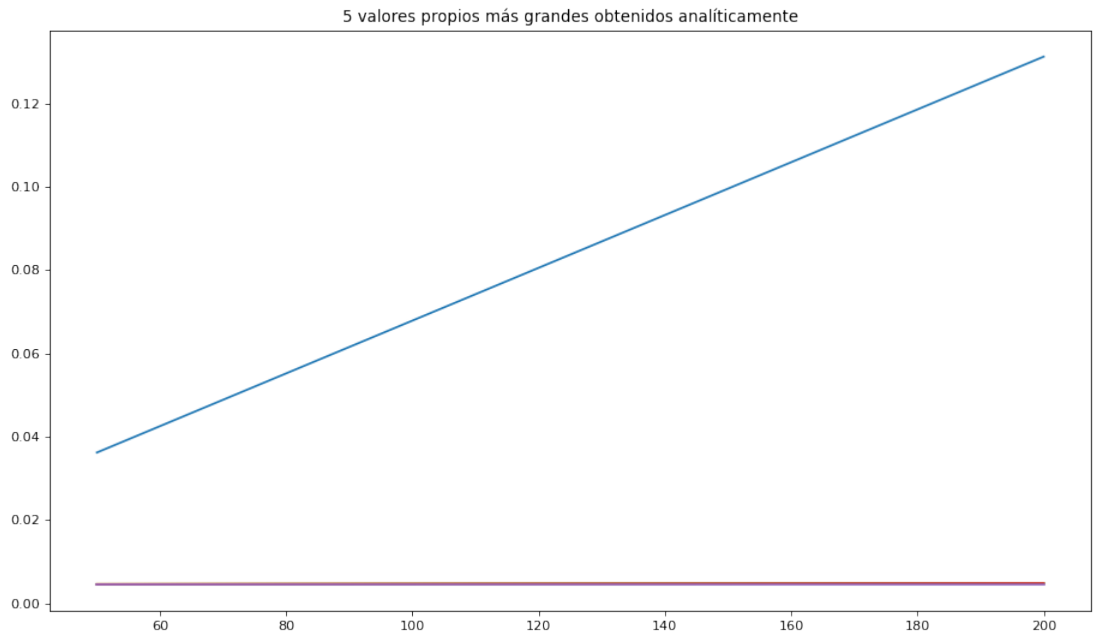
Se asume que $L \sim N(b11^t, \sigma_b^2 I)$ con $\sigma_b^2 = 0.01$, $b = 1$, $\sigma_f^2 = 0.000158$ y $\sigma_e^2 = 0.0045$.

Una vez definidos los parámetros y se elegida una semilla aleatoria, Se procede a realizar la simulación y obtener los valores propios más grandes para cada p. Se procede a graficar los 5 valores propios más grandes obtenidos mediante simulación monte carlo, y se ponen los interquantiles en líneas punteadas.



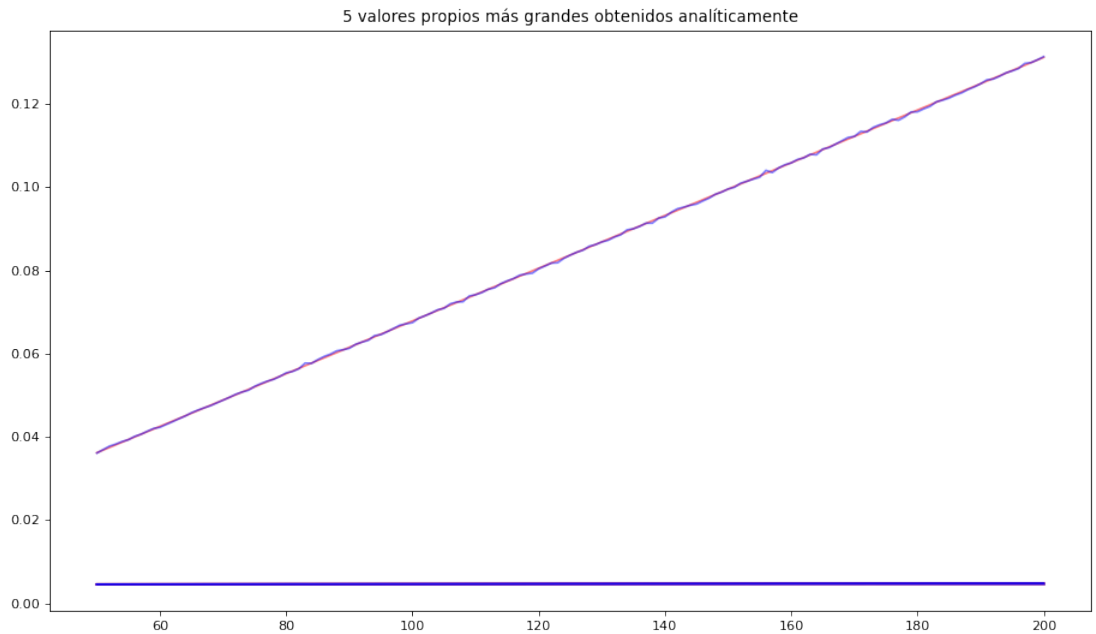
Se puede observar que a medida que incrementa p, el valor propio más grande de la matriz Σ aumenta de forma considerable, mientras que los valores propios del 2 al k también aumentan pero no de forma considerable y que los valores propios del $k+1$ al p prácticamente se mantienen constantes.

Ahora se procede a obtener los 5 valores propios utilizando los parámetros dados pero esta vez mediante la solución analítica y, una vez obtenidos los valores propios, se grafican los resultados para los distintos valores de p.



Se puede observar un comportamiento muy similar con respecto a la simulación hecha anteriormente, es decir, el valor propio más grande crece sin límite cuando p crece, del segundo al k valor propio hay un aumento, pero es pequeño comparado con el aumento del primer valor propio, y el $k+1$ al p valor propio no hay aumento, es decir, los valores propios se mantienen constantes.

También es posible sobre poner los valores propios obtenidos mediante simulación (color azul) y compararlos contra los valores propios obtenidos analíticamente(color rojo), como se hace a continuación.



Es posible observar que los valores obtenidos analíticamente se asemejan bastante a los valores simulados, lo que demuestra empíricamente que la solución analítica es correcta.

Dado que el modelo APT está basado en un análisis de factores, la cantidad que cada factor f_i contribuye a explicar la varianza total de los retornos R está dada por el valor propio asociado λ_i. Esto permite deducir que aún cuando se añaden más activos al portafolio (i.e. se aumenta el valor de p) siempre existirá un factor (asociado al valor propio más grande) que va a ser significativamente más determinante e importante para explicar la varianza de los rendimientos, sin importar que se añadan más y más activos al portafolio.

De forma más particular, Stephen Brown en **The Number of Factors in Security Returns** , publicado en 1989, afirma que en la literatura se ha encontrado que este factor sería el índice del mercado, lo cuál tendría sentido, pues dicho índice representa la tendencia general del desempeño de todos los activos y por lo tanto explicaría gran parte de los retornos.

Nota: El link al github es `https://github.com/jduarte00/est_tarea1`