Matrices de Covarianzas Grandes

Colegio de Economistas de Loreto

Hugo Vega de la Cruz

hugo.vega.dlc@gmail.com

Marzo 2021

Contenido

- Estimando matrices de covarianza
- Descomposición de la matriz de covarianzas
- Propiedades del estimador muestral de covarianzas y su número de condición en muestras pequeñas
- Estimación de covarianzas mediante reducción lineal y no lineal
- Aplicaciones
- Motivación
- Precursores: ventana móvil y RM 2006
- GARCH Multivariado: El Modelo BEKK
- El modelo de correlación condicional dinámica (DCC) de Engle (2002)
- Verosimilitud compuesta y estimación por reducción: Engle (2017)
- Aplicaciones

El problema de la estimación de Σ

- Estimar matrices de varianzas y covarianzas enfrenta dos problemas principales.
- El primero es conocido en la literatura: se trata de un problema de grados de libertad. Supongamos que tenemos N activos. La matriz de varianzas y covarianzas respectiva tendría la forma:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{21} & \cdots & \sigma_{N1} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_N^2 \end{bmatrix}$$

- La simetría implica que se requiere estimar $N(N+1)/2$ parámetros usando una muestra de T observaciones. Dado que el número de parámetros crece cuadráticamente en el número de activos, los grados de libertad se reducen drásticamente cuando aumenta N.

Descomponiendo la matriz de covarianzas

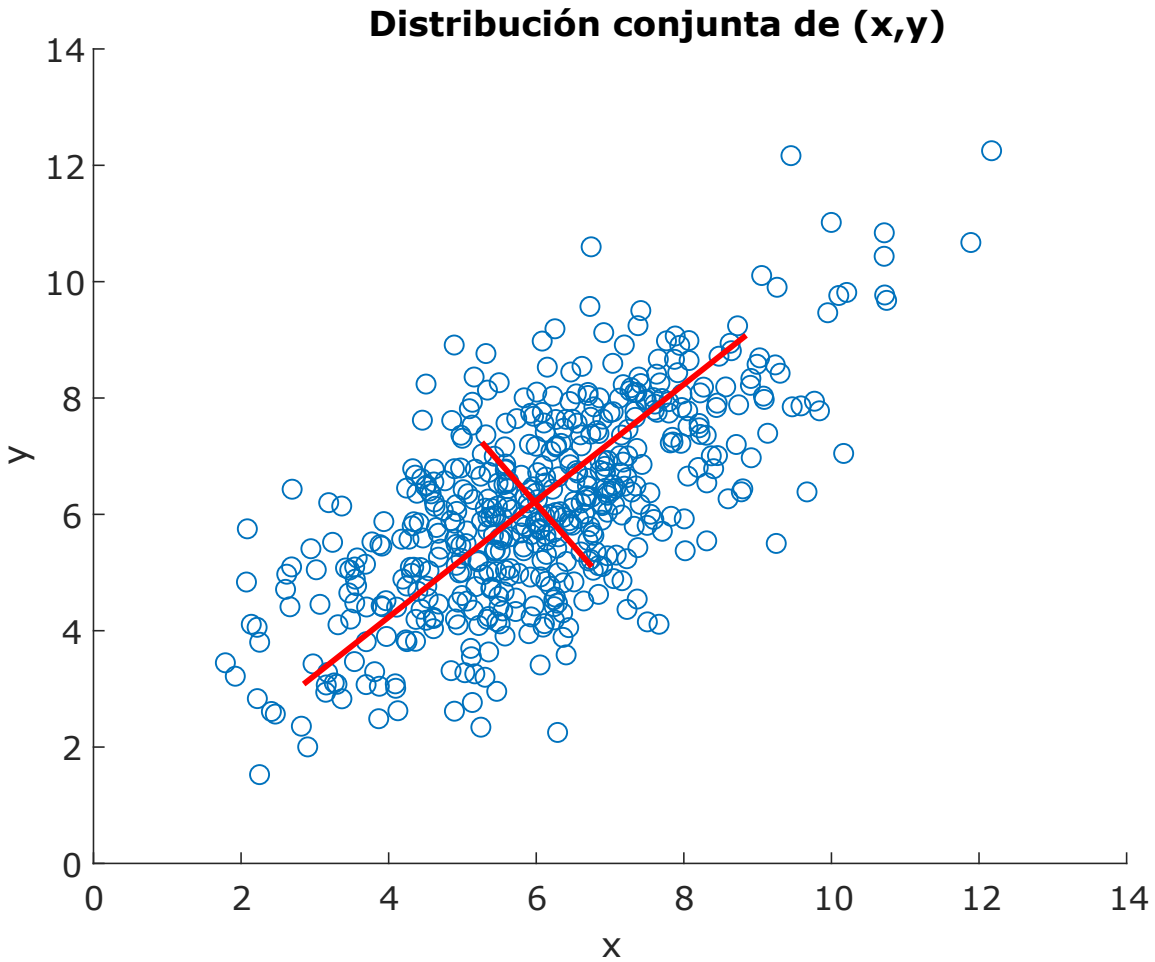
- Para comprender mejor este problema, hay que dar un paso atrás y estudiar en detalle las matrices de covarianzas. Para ello, recordemos que toda matriz puede descomponerse en valores y vectores propios:

$$\Sigma = V \Lambda V'$$

  donde V es la matriz que contiene los vectores propios de Σ y Λ es una matriz diagonal que contiene los valores propios de la misma.
- Toda matriz de covarianzas es simétrica y positivo (semi) definida. Por lo tanto, sus valores propios son siempre no-negativos.
- Intuitivamente, los vectores propios señalan la dirección de dispersión de los datos y los valores propios denotan la magnitud de la dispersión (asociada a cada vector). Por ello no pueden ser negativos.

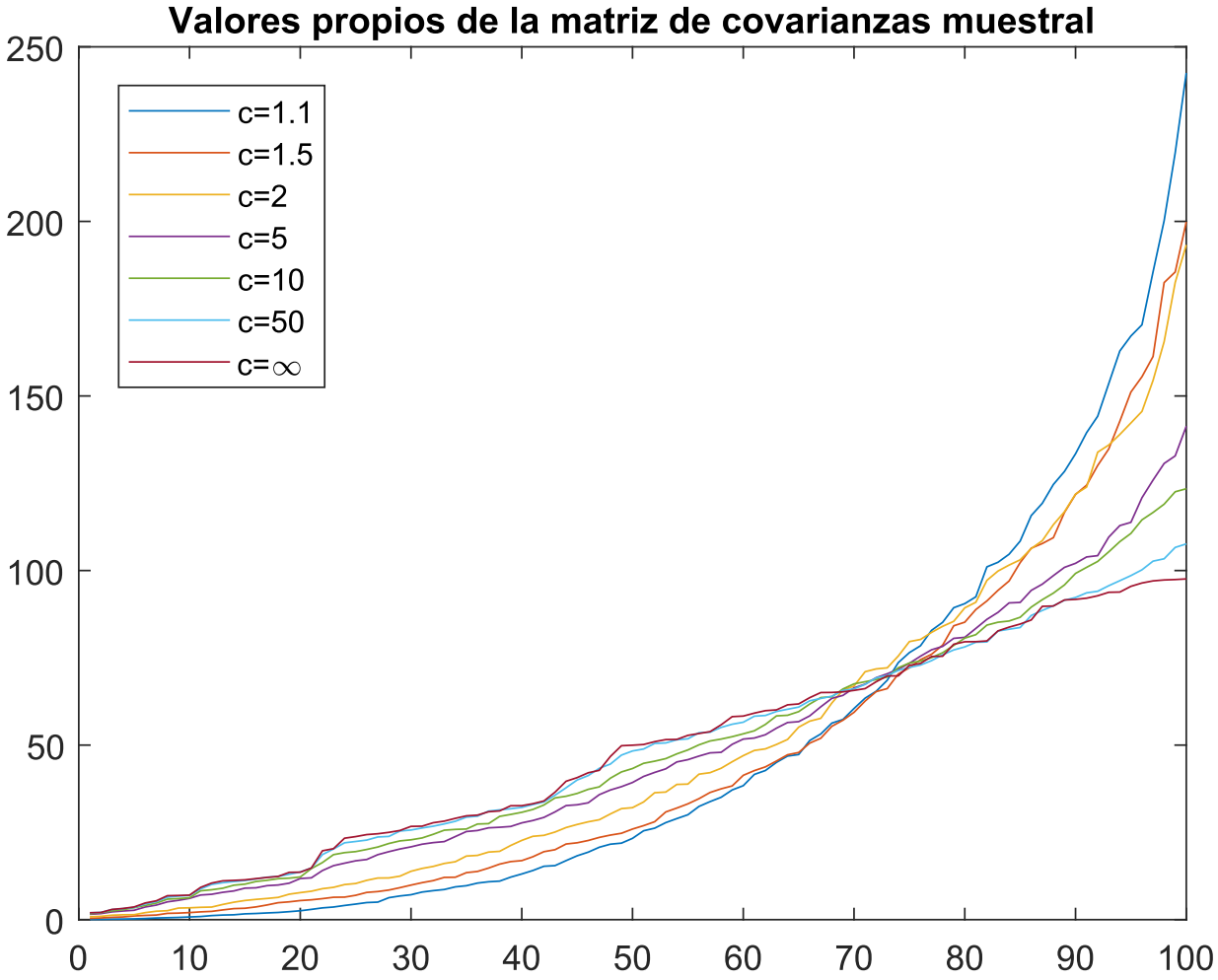
Ejemplo

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 6 \\ 6 \end{bmatrix}, \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}\right) \quad \Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix} \quad V = \begin{bmatrix} -0{,}7 & 0{,}7 \\ 0{,}7 & 0{,}7 \end{bmatrix}$$



Estimación de Σ y tamaño de muestra

¿Cómo afecta el tamaño de muestra la estimación de Σ? El siguiente gráfico muestra los valores propios que se obtienen de la matriz de covarianzas estimada para distintos tamaños de muestra con $N = 100$:



Estimación de Σ y tamaño de muestra (II)

- En el gráfico, c denota el cociente entre el tamaño de muestra y el número de variables (T/N).
- Los valores propios verdaderos son aquellos que se muestran en la línea $c = \infty$.
- Cuando c es pequeño, aparecen valores propios estimados extremos: muy pequeños (tendiendo a cero) y muy grandes (dos o tres veces la magnitud de los reales).
- Esto implica que la magnitud de la dispersión en varias dimensiones se estima incorrectamente.
- Además, aparece un problema adicional: ¡el determinante de cualquier matriz es el producto de sus valores propios!
- Este último punto es un problema serio para la econometría en general: prácticamente todos los estimadores econométricos usan la inversa de una matriz de covarianzas.

Número de condición de una matriz

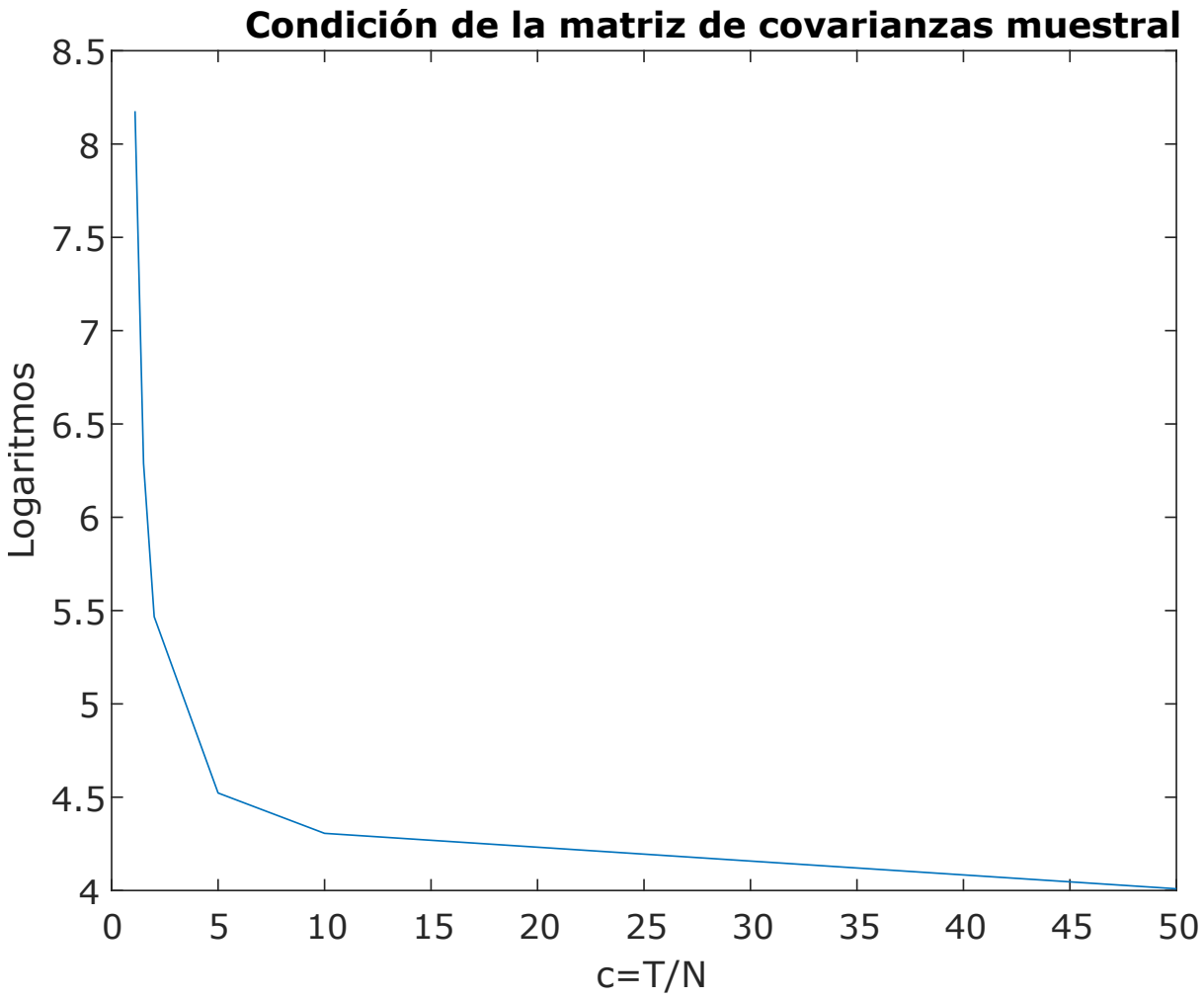
- Supongamos que tenemos el siguiente sistema de ecuaciones:

$$Ax = b$$

  donde A es una matriz, b es un vector y x es el vector de incógnitas del sistema.
- Sabemos que la solución del sistema es $x^* = A^{-1}b$.
- El número de condición de la matriz A, $\kappa(A)$, es la respuesta a la pregunta: ¿cuál es la tasa de cambio (porcentual) de x^* ante una variación (porcentual) en b?
- Intuitivamente, un número de condición cercano a 1 indica que la matriz A es "bien comportada". Cuando $\kappa(A)$ se aleja significativamente de 1, la matriz A genera problemas al invertirla.

Número de condición de Σ y tamaño de muestra

Siguiendo con nuestro ejemplo, el siguiente gráfico muestra como cambia el número de condición de la matriz de covarianzas estimada para distintos tamaños de muestra:



Número de condición de Σ y tamaño de muestra (II)

- Claramente, valores pequeños de c resultan en números de condición varios órdenes de magnitud mayores a lo ideal (note que el eje vertical en la figura anterior muestra $\log \kappa(\hat{\Sigma})$ y no $\kappa(\hat{\Sigma})$).
- Esto implica que cuando el tamaño de muestra es relativamente pequeño, la matriz $\hat{\Sigma}$ arroja resultados incorrectos al invertirla.
- Por lo tanto, portafolios construidos usando esta matriz sufren de una serie de problemas (inestabilidad, concentración).
- Para construir portafolios, se pueden llegar a usar cientos de activos (en el caso de selección de acciones individuales a partir de un índice, por ejemplo). Típicamente los datos que se usan tienen frecuencia mensual (mayor frecuencia puede enfrentar problemas de liquidez o disponibilidad de información).

Número de condición de Σ y tamaño de muestra (III)

- Si consideramos 10 años de datos (120 meses) estaríamos restringidos a 24 activos como máximo si queremos mantener $c \geq 5$. Ampliar el número de años puede ser factible pero podría volver a la muestra poco relevante.
- ¿Por qué ampliar la muestra puede volverla poco relevante? Esto hace referencia al segundo problema que se enfrenta al estimar matrices de covarianzas: estas cambian en el tiempo. Si se usa una muestra con datos muy lejanos en el pasado, podría no reflejar adecuadamente las condiciones actuales. Volveremos a este punto más adelante.

Estimación por reducción lineal

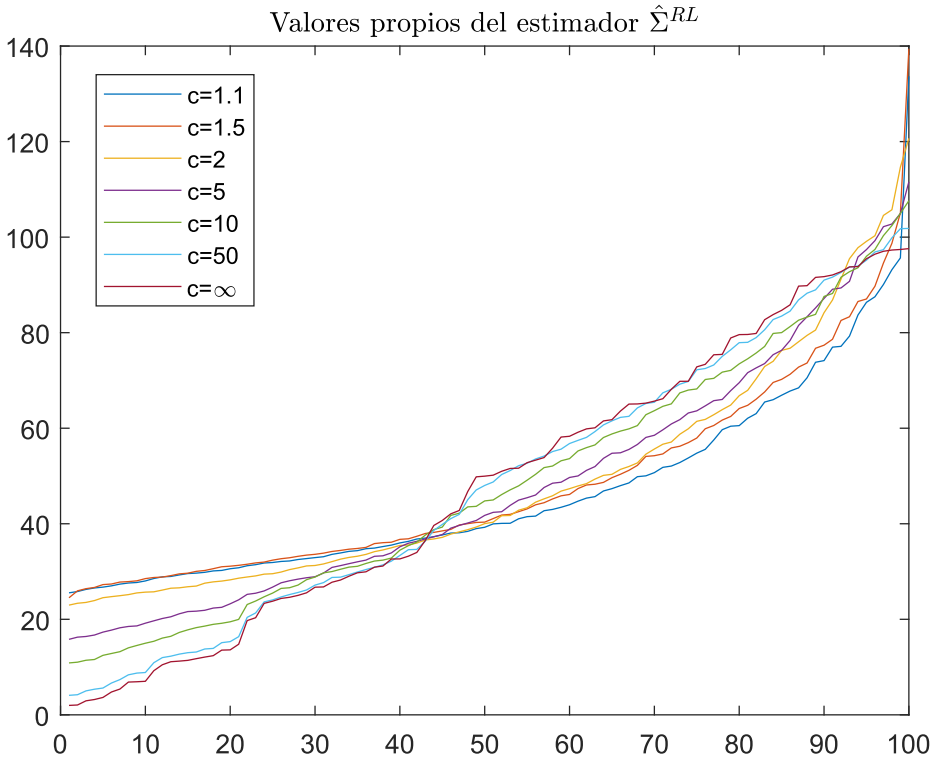
- Demos un paso atrás para pensar en el primer problema: ¿cómo mejoramos las propiedades de la matriz de covarianzas estimada? ¿Cómo evitamos extremos en los valores propios estimados?
- Una primera aproximación al problema la constituye el estimador por reducción (*shrinkage*) lineal:

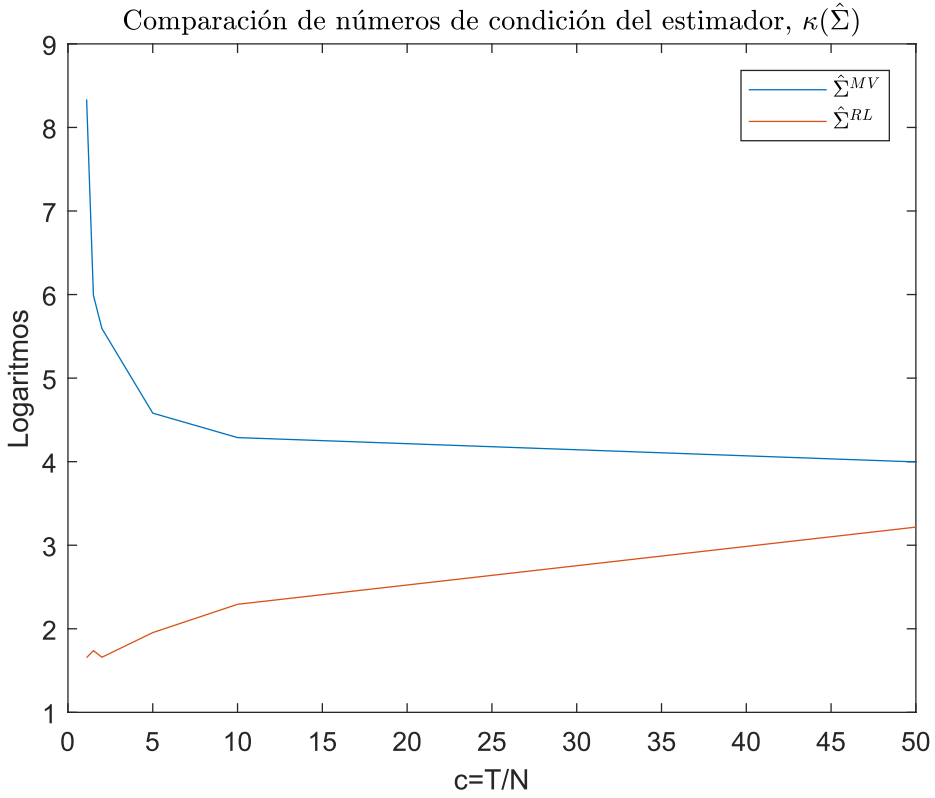
$$\hat{\Sigma}^{RL} = s\bar{\sigma}^2 I + (1-s)\hat{\Sigma}^{MV}$$

  donde s es la magnitud de la reducción, $\bar{\sigma}^2$ es la varianza promedio de $\hat{\Sigma}^{MV}$ y este último es el estimador muestral (que coincide con el de máxima verosimilitud) de Σ.
- Cabe señalar que, dado que la traza de $\hat{\Sigma}^{MV}$ es igual a la suma de sus valores propios, $\bar{\sigma}^2$ es igual al promedio de los valores propios de $\hat{\Sigma}^{MV}$.
- De esta forma, el estimador por reducción lineal disminuye la dispersión de los valores propios, "¡alándolos" hacia la media.

Estimador por reducción lineal, $\hat{\Sigma}^{RL}$

Las figuras muestran los nuevos valores propios y números de condición del estimador para distintos tamaños de muestra.





Discusión

- El estimador por reducción lineal hace un mejor trabajo que el estimador estándar para la matriz de covarianzas.
- La reducción lineal (hacia una matriz identidad multiplicada por la varianza promedio) arroja un estimador mejor comportado: sus valores propios dejan de ser extremos, garantizando que la matriz de covarianzas estimada pueda ser invertida.
- Sin embargo, la magnitud de la reducción óptima (s) no es única y depende del ratio entre el tamaño de muestra y el número de variables ($c = T/N$). Dicha relación no es trivial. Además, los valores propios estimados por reducción lineal tienen sesgos claros.
- Por esta razón, Ledoit y Wolf (2012, 2014) proponen un estimador de matriz de covarianzas por reducción no lineal como alternativa.

Estimación por reducción no lineal

- El estimador por reducción no lineal, $\hat{\Sigma}^{RN}$, pertenece a la clase de estimadores construidos usando la fórmula:

$$\hat{\Sigma}^{RN} = V \Lambda V'$$

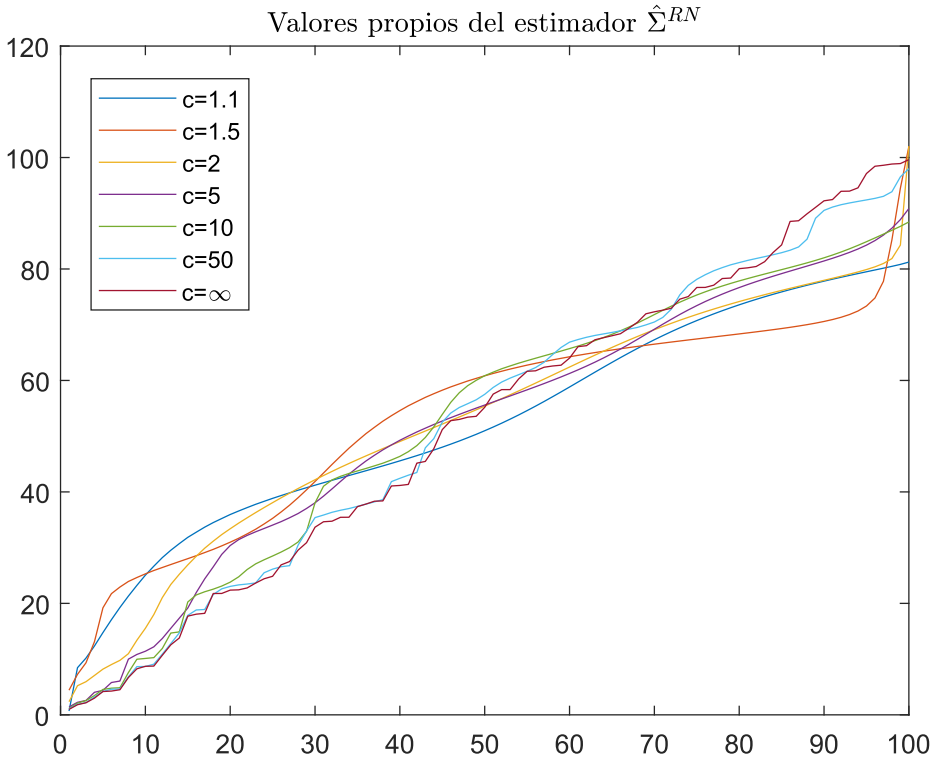
- La matriz V mantiene los vectores propios obtenidos de la descomposición del estimador muestral estándar, $\hat{\Sigma}^{MV}$. La matriz Λ se construye reduciendo los valores propios de $\hat{\Sigma}^{MV}$:

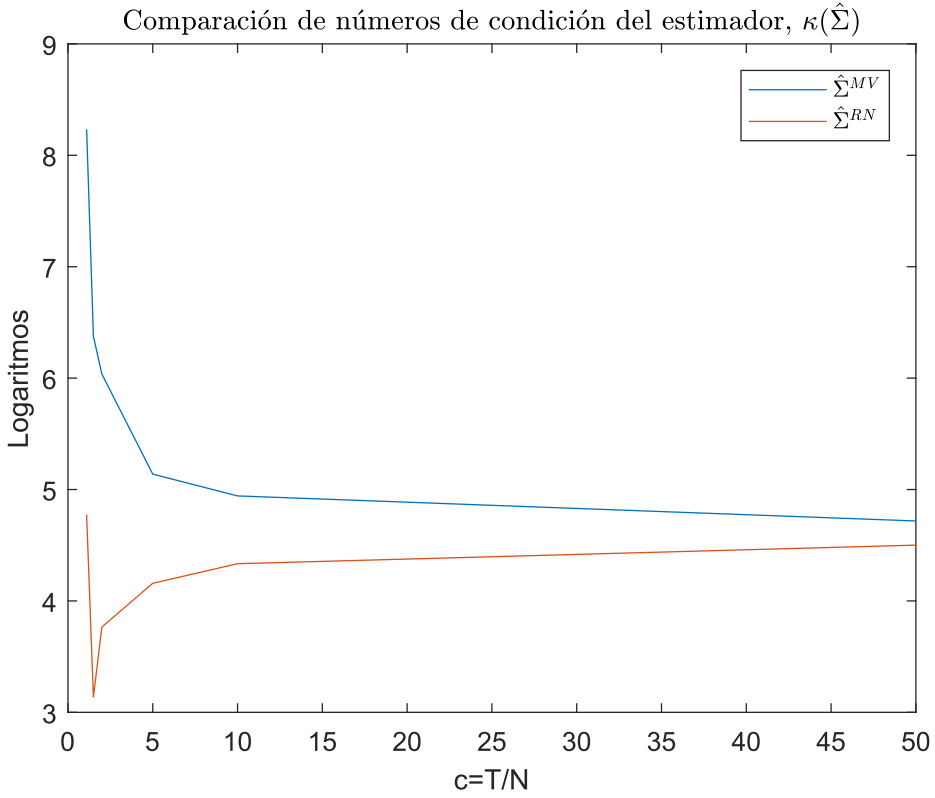
$$\Lambda := \text{Diag}\left(d(\hat{\lambda}_1), \ldots, d(\hat{\lambda}_N)\right)$$

  donde $d : \mathbb{R} \to \mathbb{R}$ es la función de reducción no lineal desarrollada por Ledoit y Wolf (2012, 2014) y $\hat{\lambda}_i$ son los valores propios estimados de $\hat{\Sigma}^{MV}$.
- Ledoit y Wolf han implementado la función d en el paquete QuEST, disponible en la web de Michael Wolf en UZH.

Estimador por reducción no lineal, $\hat{\Sigma}^{RN}$

Las figuras muestran los nuevos valores propios y números de condición del estimador para distintos tamaños de muestra.





Discusión

- El estimador de covarianzas por reducción no lineal representa el estado del arte actualmente. Puede manejar fácilmente problemas con cientos de variables.
- Muestra un desempeño superior a todos los estimadores previos, siendo su única desventaja el mayor costo computacional que hay que afrontar para poder obtenerlo.
- Sin embargo, Ledoit y Wolf (2017) ya han propuesto una alternativa con un costo computacional similar al de $\hat{\Sigma}^{MV}$ y $\hat{\Sigma}^{RL}$, con una pérdida de precisión mínima respecto a $\hat{\Sigma}^{RN}$. Este último estimador se vuelve necesario cuando el problema involucra miles de variables, pero esta situación es poco común en la práctica (aún).
- A continuación presentaremos algunas aplicaciones en las que se utiliza el estimador $\hat{\Sigma}^{RN}$.

Aplicaciones: Regresión lineal

- Una de las primeras (y más generales) aplicaciones que vamos a revisar es la regresión lineal.
- Considere el siguiente modelo:

$$Y = \mu + X\beta + \epsilon; \quad X \sim \mathcal{N}\left(\mu_X, \Sigma_{XX}\right); \quad \epsilon \sim \mathcal{N}\left(0, I\right).$$

Por definición,

$$\Sigma_{XY} = E\left[\left(X - \mu_X\right)\left(Y - E\left[Y\right]\right)\right]$$

entonces,

$$\begin{aligned}\Sigma_{XY} &= E\left[\left(X - \mu_X\right)\left(\mu + X\beta + \epsilon - \mu - \mu_X\beta\right)\right] \\ &= E\left[\left(X - \mu_X\right)\left(X - \mu_X\right)\right]\beta + E\left[\left(X - \mu_X\right)\epsilon\right] \\ &= \Sigma_{XX}\beta\end{aligned}$$

- Por lo tanto, $\beta = \Sigma_{XX}^{-1}\Sigma_{XY}$.

Aplicaciones: Regresión lineal (II)

- El problema de regresión simple consiste en estimar β, descubrir la relación entre X e Y.
- El estimador por excelencia para lidiar con este problema es el de mínimos cuadrados: $\hat{\beta} = (X'X)^{-1} X'Y$.
- Pero note que lo que estamos tratando de hacer en realidad es estimar dos cantidades: una varianza (Σ_{XX}) y una covarianza (Σ_{XY}).
- Podemos usar el estimador de covarianzas por reducción no lineal, $\hat{\Sigma}^{RN}$, para este propósito. Para ello, creamos una matriz con los datos de X e Y y le aplicamos el estimador:

$$Z = [X \mid Y] \quad \rightarrow \quad \hat{\Sigma}^{RN}_{ZZ} = \begin{bmatrix} \hat{\Sigma}^{RN}_{XX} & \hat{\Sigma}^{RN}_{XY} \\ \hat{\Sigma}^{RN}_{YX} & \hat{\Sigma}^{RN}_{YY} \end{bmatrix}$$

- Nuestro nuevo estimador de β sería: $\hat{\beta}^{RN} = \left(\hat{\Sigma}^{RN}_{XX}\right)^{-1} \hat{\Sigma}^{RN}_{XY}$.

Aplicaciones: Hamilton (2017)

Why You Should Never Use the Hodrick-Prescott Filter*

James D. Hamilton

jhamilton@ucsd.edu

Department of Economics, UC San Diego

July 30, 2016

Revised: June 22, 2017

ABSTRACT

Here's why. (1) HP introduces spurious dynamic relations that have no basis in the underlying data-generating process. (2) Filtered values at the end of the sample are very different from those in the middle, and are also characterized by spurious dynamics. (3) A statistical formalization of the problem typically produces values for the smoothing parameter vastly at odds with common practice. (4) There's a better alternative. A regression of the variable at date t on the four most recent values as of date $t-h$ achieves all the objectives sought by users of the HP filter with none of its drawbacks.

Aplicaciones: Hamilton (2017) (II)

- Hamilton propone dejar de lado el filtro HP para estimar la brecha de producto. Su propuesta consiste en estimar una regresión de la forma:

$$y_{t+h} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \beta_4 y_{t-3} + v_{t+h}$$

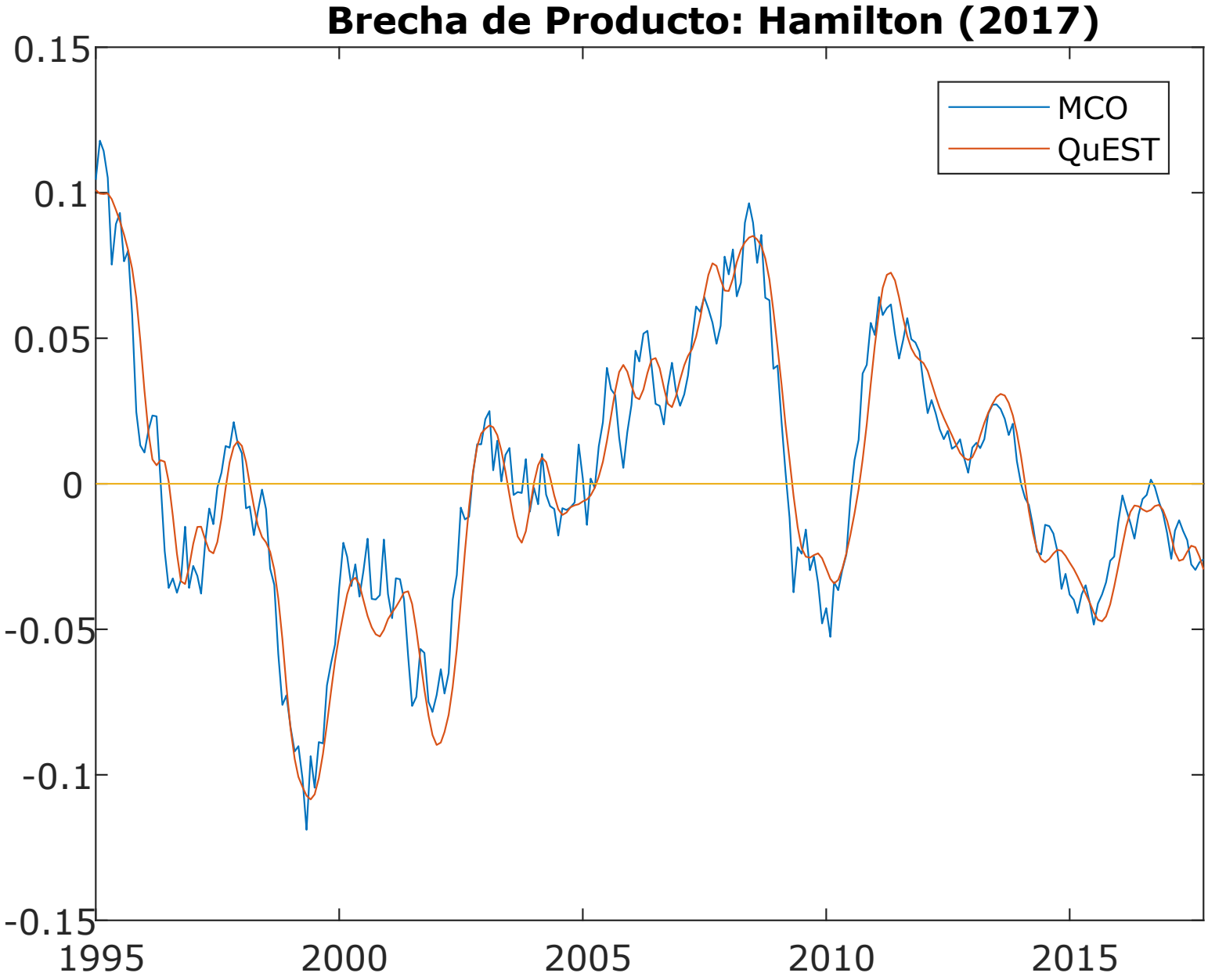
  y usar el residuo estimado mediante MCO, $\hat{v}_{t+h}$, como una proxy del componente transitorio (cíclico) del producto.
- En su artículo, Hamilton demuestra que esta estrategia tiene mejores propiedades que el filtro HP para estimar el componente cíclico del producto, usando datos trimestrales de Estados Unidos.
- Para aplicar esta metodología al PBI peruano, estimaremos la siguiente regresión con datos mensuales:

$$y_{t+h} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \ldots + \beta_{12} y_{t-11} + v_{t+h}$$

Aplicaciones: Hamilton (2017) (III)

- Lamentablemente, los resultados de aplicar MCO en el caso de los datos mensuales de Perú son pobres.
- El problema es que el método de Hamilton requiere un año completo de rezagos en los regresores (que vienen a ser 12 rezagos del PBI para el caso mensual). Estimar los betas de esa regresión enfrenta un serio problema de multicolinealidad.
- Afortunadamente, podemos estimar los betas usando reducción no lineal (como se detalló anteriormente). Este procedimiento corrige el problema elegantemente y arroja una brecha de producto bien comportada.

Aplicaciones: Hamilton (2017) (IV)



Aplicaciones: Portafolios mínima varianza

- Como última aplicación, construiremos un portafolio global defensivo de mínima varianza usando índices de acciones.
- Compararemos los resultados que se obtienen usando los estimadores $\hat{\Sigma}^{MV}$ y $\hat{\Sigma}^{RN}$.
- El portafolio mínima varianza se obtiene resolviendo el siguiente problema:

$$\min_{w} U = \frac{1}{2} w' \Sigma w$$

  sujeto a $\sum_{i=1}^{i=N} w_i = 1$ donde w es un vector (columna) de pesos para los N activos y Σ es la matriz de covarianzas de los activos (de orden N).
- Dado que desconocemos la "verdadera" matriz Σ, la reemplazamos por su contraparte estimada. Veremos que usar $\hat{\Sigma}^{RN}$ arroja resultados muy distintos a $\hat{\Sigma}^{MV}$

Aplicaciones: Portafolios mínima varianza (II)

- La condición de primer orden para hallar el portafolio de mínima varianza es:

$$\frac{\partial U}{\partial w} = \Sigma w + \lambda \mathbf{1} = 0$$

Despejando w, obtenemos:

$$w = -\lambda \Sigma^{-1} \mathbf{1}$$

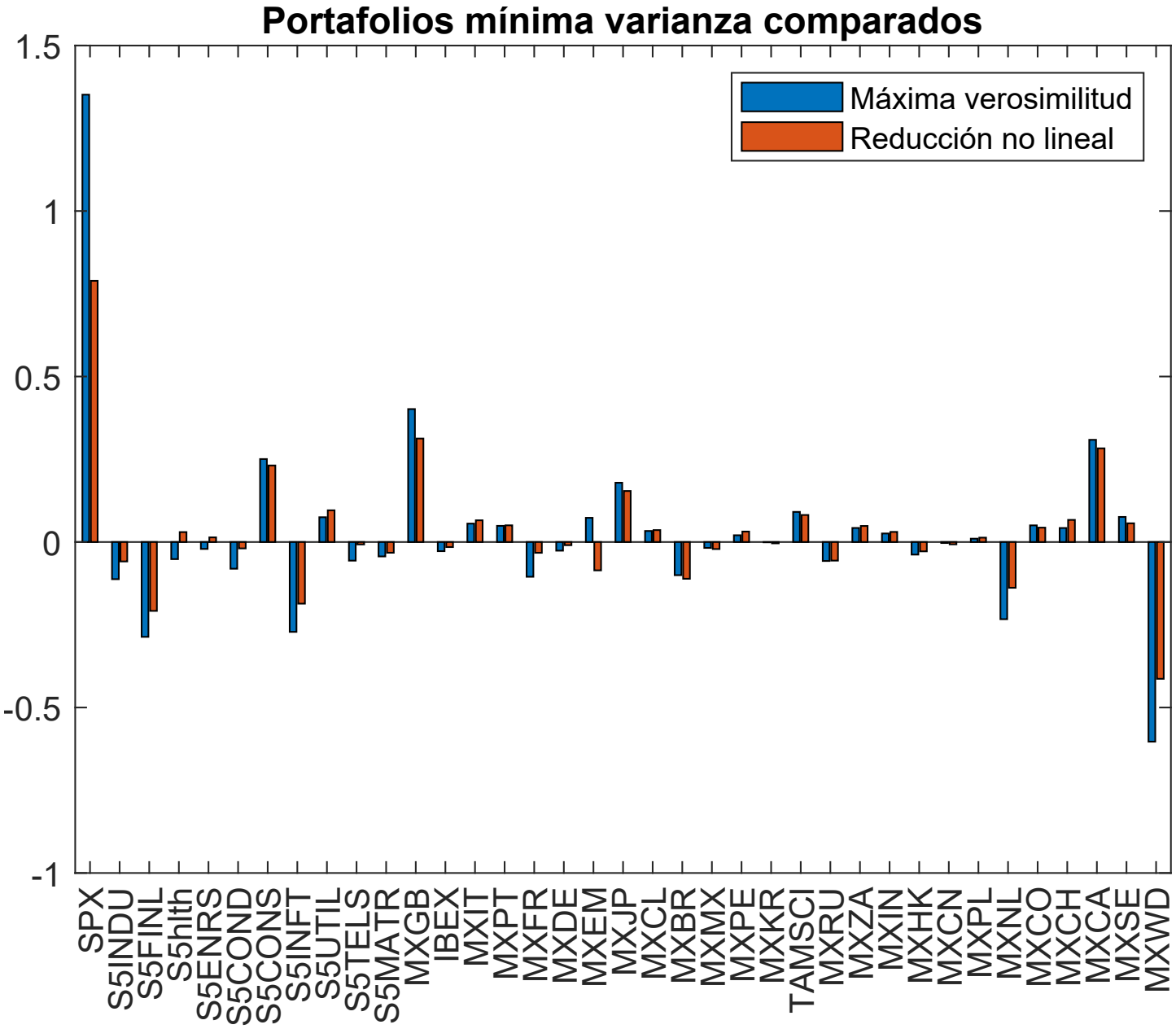
Reemplazando en la restricción ($w'\mathbf{1} = 1$):

$$-\lambda \mathbf{1}' \Sigma^{-1} \mathbf{1} = 1$$

por lo tanto,

$$\lambda^* = -\frac{1}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \quad \rightarrow \quad w^* = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}$$

Aplicaciones: Portafolios mínima varianza (III)



¿Qué es una (co)varianza?

- Si tenemos dos variables aleatorias, X e Y, la covarianza entre ellas se define como:

$$\text{cov}\left(X, Y\right) = E\left[\left(X - E\left[X\right]\right)\left(Y - E\left[Y\right]\right)\right]$$

  Esta es una medida del grado de "comovimiento" entre X e Y.
- La varianza de una variable no es más que su "covarianza" consigo misma:

$$\text{var}\left(X\right) = E\left[\left(X - E\left[X\right]\right)\left(X - E\left[X\right]\right)\right]$$

  Mide la dispersión de una variable aleatoria.

¿Qué es una (co)varianza? (II)

- Las contrapartes muestrales de la covarianza y la varianza son, respectivamente:

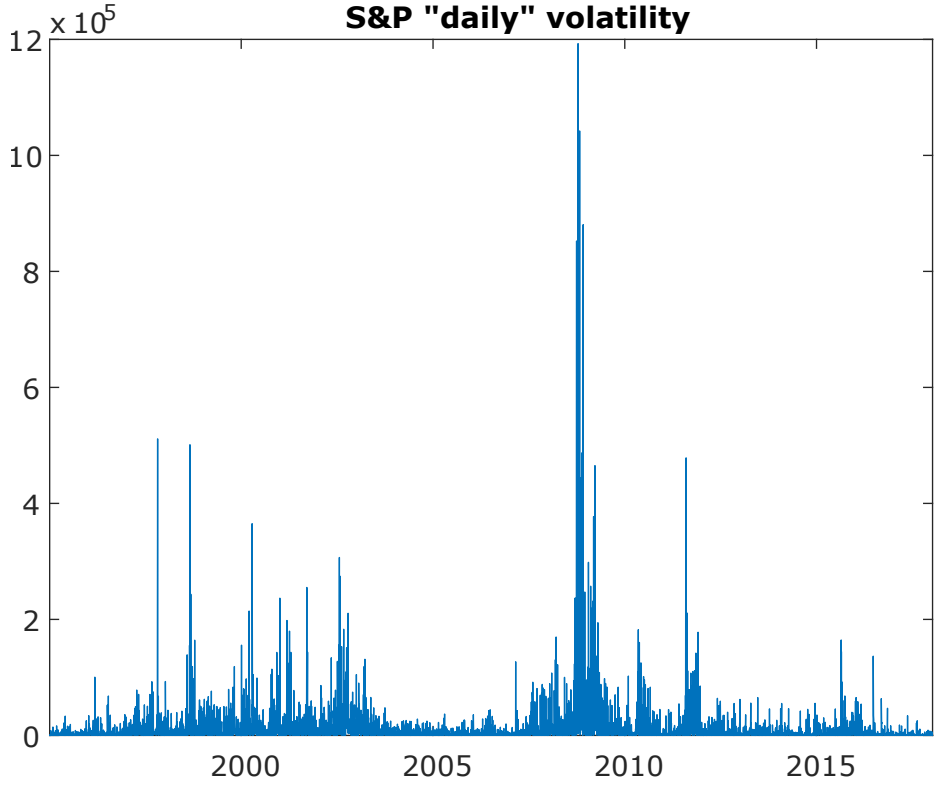
$$\hat{\text{cov}}\left(X,Y\right)=\frac{1}{T}\sum_{i=1}^{T}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)$$

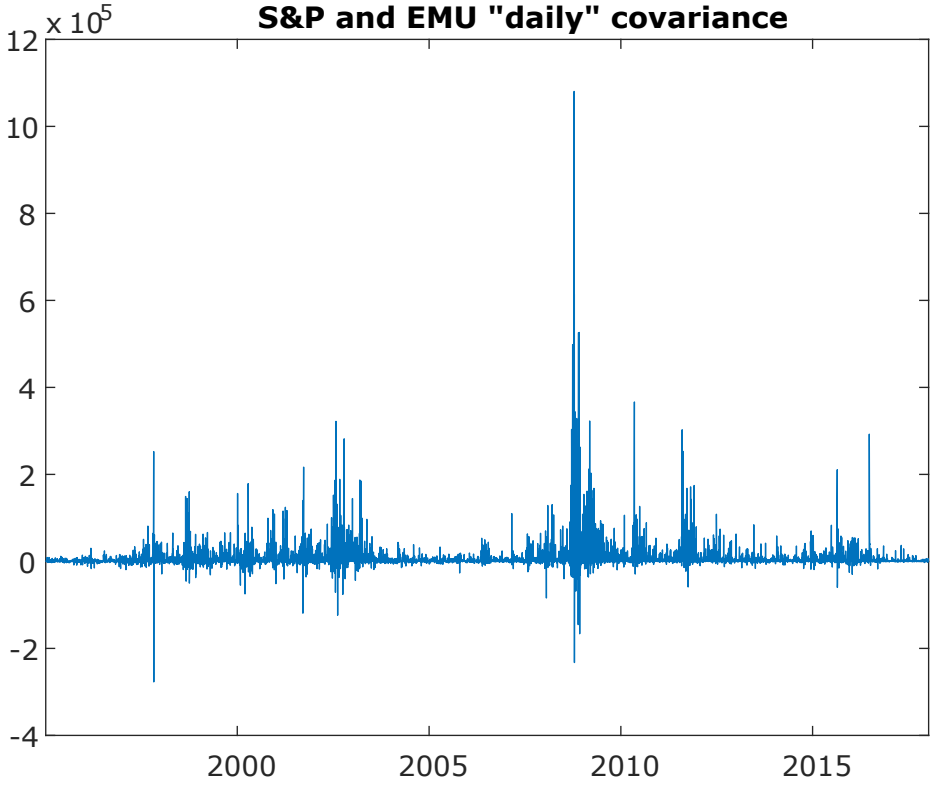
$$\hat{\text{var}}\left(X\right)=\frac{1}{T}\sum_{i=1}^{T}\left(x_i-\bar{x}\right)^2$$

  donde las barras encima de una variable representan la media muestral y T es el tamaño de la muestra.
- Note que la covarianza y varianza muestral son básicamente promedios de $\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)$ y $\left(x_i-\bar{x}\right)^2$, respectivamente. ¿Cómo se comportan estas variables en el tiempo?

Varianzas y covarianzas dinámicas

Las figuras muestran la evolución de $(x_i - \bar{x})^2$ y de $(x_i - \bar{x})(y_i - \bar{y})$ donde los x_i son los retornos diarios del S&P y y_i los retornos diarios de EMU.





Discusión

- Las varianzas y covarianzas NO son constantes en el tiempo.
- Además, presentan "clustering": en determinados periodos se observan aglomeraciones de valores más altos (o más bajos en el caso de la covarianza).
- La observación que la varianza no es constante en el tiempo y tiene "clusters" motivó el desarrollo de los modelos GARCH.
- Pero el hecho que las covarianzas también cambien en el tiempo implica que, en general, las matrices de covarianzas son dinámicas. Esto representa un reto para la estimación.
- Para construir un portafolio para un horizonte de 1 año o más puede ser razonable estimar una matriz de covarianzas con 10 o más años de datos. ¿Pero qué pasa cuando quiero revisar los portafolios cada mes? ¿Cada semana?

Ventanas móviles

- Una primera solución potencial al problema de estimación dinámica de covarianzas consiste en usar ventanas móviles.
- Al usar ventanas móviles enfrentamos dos posibilidades: con o sin traslape.
- Evidentemente, sin traslape será imposible estimar. Con muestras semanales tendríamos un problema de grados de libertad inmediatamente y si bien la estimación mediante reducción no lineal arrojaría un resultado con propiedades aceptables, los resultados no serían buenos (ningún estimador puede "crear" información que no está en la muestra y la muestra tendría insuficiente información).
- Con traslape, el tamaño de muestra tendría que ser lo suficientemente grande para tener un ratio T/N aceptable. Sin embargo, ello implicaría que el estimador no variaría significativamente con cada revisión.

Estimación con pesos exponenciales: Risk Metrics 2006

- Una alternativa a las ventanas móviles es la estimación mediante medias móviles con pesos exponenciales (EWMA por sus siglas en inglés). Esta metodología fue popularizada por el grupo Risk Metrics (1994, modificada en 2006).
- La idea consiste en usar todos los datos disponibles para estimar la matriz de covarianzas corriente, pero los datos más recientes pesan más en la estimación que aquellos que se encuentran más lejos en el pasado.
- Los pesos asignados a datos pasados decaen exponencialmente a una tasa dada por el parámetro λ, calibrado para una determinada "vida media" de la muestra.

Estimación con pesos exponenciales (II)

- Los pesos se calculan de la siguiente forma:

$$w_i = \frac{(1-\lambda)\,\lambda^{T-i}}{\sum_{i=1}^{T}(1-\lambda)\,\lambda^{T-i}} \quad 0 < \lambda < 1 \quad i = \{1, 2, \ldots, T\}$$

- Una vez calculados los pesos, la matriz de covarianzas se estima multiplicando cada observación por su respectivo peso.
- El estimador EWMA está basado en el concepto GARCH generalizado a un contexto de N variables. Sin embargo, constituye una solución que deja mucho que desear.

Discusión

- Resulta evidente que para estimar matrices de covarianzas dinámicas será necesario imponer algún tipo de estructura.
- El ejercicio está destinado al fracaso sin estructura alguna, simplemente no contamos con suficiente información.
- El problema de EWMA es que impone demasiada estructura: asume que todos los elementos de la matriz de covarianzas siguen un mismo tipo de proceso, con el mismo (único) parámetro.
- La clave para lograr una buena estimación de covarianzas dinámicas está en el balance: muy poca estructura nos lleva a un problema de grados de libertad. Mucha estructura implica supuestos demasiado irreales para los datos.

Baba, Engle, Kraft y Kroner: El Modelo BEKK

- En 1995, Engle y Kroner propusieron una extensión natural al GARCH para el caso multivariado: el llamado modelo BEKK.

$$r_t \sim \mathcal{N}(0, \Sigma_t)$$
$$\Sigma_t = C'C + \sum_{j=1}^{q} B_j' r_{t-j} r_{t-j}' B_j + \sum_{i=1}^{p} A_i' \Sigma_{t-i} A_i$$

  donde A_i y B_i son matrices de coeficientes de dimensión $N \times N$ y C es una matriz triangular con $N(N+1)/2$ coeficientes.
- Por lo tanto, el BEKK(p,q) es un modelo altamente parametrizado con un total de $(p+q)N^2 + N(N+1)/2$ parámetros.

El Modelo BEKK (II)

- En la práctica, el BEKK original sufría de una serie de problemas.
- Ya que se estimaba mediante máxima verosimilitud, el grán número de parámetros en BEKK y los máximos locales en la función de verosimilitud resultan frecuentemente en sobreajuste ("overfitting").
- Además, estimar los parámetros por MV implica un proceso de optimización no lineal computacionalmente costoso y hasta prohibitivo cuando se trabaja en grandes dimensiones (de N).
- Para mitigar estos problemas, en la práctica se usan versiones restringidas del BEKK. El orden de los parámetros p y q se asume 1 y las matrices A_1 y B_1 se asumen diagonales. La expresión para Σ_t pasa a ser:

$$\Sigma_t = C'C + B'r_{t-1}r'_{t-1}B + A'\Sigma_{t-1}A$$

- Complementariamente, se han venido desarrollando métodos bayesianos para estimar el BEKK.

Discusión

- La literatura relacionada a los modelos BEKK continúa desarrollándose.
- Conforme la capacidad computacional de nuestros ordenadores aumenta y (más importante) las variantes del modelo y estrategias de estimación se van refinando, estos modelos han ido ganando terreno. Los métodos bayesianos aplicados a estos modelos son particularmente prometedores.
- Sin embargo, el modelo BEKK continúa siendo, hasta cierto punto, un intento por estimar covarianzas dinámicas ignorando su estructura interna. Es poco eficiente ya que, incluso en sus versiones más recientes, sigue contando con muchos parámetros.
- En el año 2002, Engle lanzó una propuesta alternativa que ha rendido muchos frutos en los últimos años: correlaciones condicionales dinámicas (DCC).

Correlación condicional dinámica: Engle (2002)

- Toda matriz de covarianzas puede escribirse de la siguiente forma:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 & \cdots & \rho_{1N}\sigma_1\sigma_N \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 & \cdots & \rho_{2N}\sigma_2\sigma_N \\ \rho_{31}\sigma_3\sigma_1 & \rho_{32}\sigma_3\sigma_2 & \sigma_3^2 & \cdots & \rho_{3N}\sigma_3\sigma_N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{N1}\sigma_N\sigma_1 & \rho_{N2}\sigma_N\sigma_2 & \rho_{N3}\sigma_N\sigma_3 & \cdots & \sigma_N^2 \end{bmatrix}$$

  donde ρ_{ij} denota la correlación entre la variable i y la j (note que $\rho_{ij} = \rho_{ji}$ siempre).
- Expresar la matriz de covarianzas de esta manera permite intuir que podemos entender Σ como un producto de correlaciones y desviaciones estándar:

$$\Sigma = \text{Diag}\left\{\sigma_1, \sigma_2, \ldots, \sigma_N\right\} \times C \times \text{Diag}\left\{\sigma_1, \sigma_2, \ldots, \sigma_N\right\}$$

Correlación condicional dinámica (II)

- En la expresión anterior, Diag $\{\sigma_1, \sigma_2, \ldots, \sigma_N\}$ representa una matriz diagonal con las desviaciones estándar de las N variables en la diagonal.
- C es una matriz (simétrica) de correlaciones y tiene la forma siguiente:

$$C = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1N} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2N} \\ \rho_{31} & \rho_{32} & 1 & \cdots & \rho_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{N1} & \rho_{N2} & \rho_{N3} & \cdots & 1 \end{bmatrix}$$

- En el modelo DCC, la dinámica de cada una de las N desviaciones estándar se modela independientemente, con un GARCH(1,1). La evolución dinámica de la matriz de correlaciones se modela separadamente en una segunda etapa, usando los resultados de los GARCH.

Modelo DCC

- Sea r_t el vector de retornos de N activos financieros:

$$
\begin{aligned}
& r_t \mid I_{t-1} \sim \mathcal{N}\left(0, D_t R_t D_t\right) \\
& \sigma_{i,t}^2 = \omega_i + \alpha_i r_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2 \\
& D_t = \text{Diag}\left\{\sigma_{1,t}, \sigma_{2,t}, \ldots, \sigma_{N,t}\right\} \\
& \varepsilon_t = D_t^{-1} r_t \\
& Q_t = \left(1 - \alpha_q - \beta_q\right) S + \alpha_q \varepsilon_{t-1} \varepsilon_{t-1}' + \beta_q Q_{t-1} \\
& R_t = \text{diag}\left\{Q_t\right\}^{-\frac{1}{2}} Q_t \, \text{diag}\left\{Q_t\right\}^{-\frac{1}{2}}
\end{aligned}
$$

  donde S denota la matriz de correlaciones de equilibrio y el operador "diag" extrae la diagonal de una matriz.
- La segunda línea define procesos GARCH(1,1) para las varianzas, la cuarta línea es una estandarización, la quinta línea especifica la dinámica de la matriz de correlaciones y la última línea asegura que la matriz de correlaciones tenga unos en la diagonal.

Problemas prácticos

- En su artículo de 2002, Engle demostró que la versión bivariada del modelo DCC provee una muy buena aproximación a una variedad de procesos dinámicos de correlación.
- Si bien el modelo está diseñado para estudiar múltiples variables, es evidente que en ese entonces la capacidad computacional disponible y la especificación propuesta no permitían estudiar conjuntos de datos con N grande.
- Además, cabe señalar que en la propuesta original de Engle, α_q y β_q eran matrices, no escalares, esto incrementaba aún más el número de parámetros a estimar.
- Por estas razones, el modelo DCC no se popularizó mucho. La literatura de covarianzas dinámicas avanzó básicamente por la ruta del BEKK.

Verosimilitud compuesta y estimación por reducción: el nuevo DCC de Engle, Ledoit & Wolf (2017)

- El modelo DCC se estima por máxima verosimilitud, optimizando la función de verosimilitud que se obtiene de la especificación.
- La función de verosimilitud del DCC crece en complejidad y costo computacional en proporción a N^2. Cuando N es grande, la función de verosimilitud se vuelve rápidamente inmanejable.
- En un artículo publicado por Pakel et al. (2014), se propuso una solución a este problema: la función de verosimilitud compuesta.
- La idea detrás de la verosimilitud compuesta es la siguiente: en lugar de calcular la función de verosimilitud para los $N(N-1)/2$ pares de activos presentes en la muestra, usaremos sólo los $N-1$ pares contiguos. Pakel et al. (2014) demuestran que la función de verosimilitud "compuesta" resultante arroja estimadores consistentes.

Engle, Ledoit & Wolf (2017) (II)

- El estimador DCC tiene la propiedad que su función de máxima verosimilitud puede dividirse en dos partes: la primera comprende los procesos GARCH(1,1) y la segunda las correlaciones.
- Los parámetros de la primera parte son independientes de los de la segunda así que se pueden estimar por separado.
- Luego, la verosimilitud compuesta se usa para la estimación del modelo para las correlaciones.
- Para mejorar el proceso todavía más, se aplica la estimación mediante reducción no lineal de Ledoit & Wolf para obtener un estimado de la matriz de correlaciones de equilibrio, $\hat{S}$.
- El resultado es un estimador de covarianzas dinámicas consistente y robusto a dimensiones grandes (de N).

Engle et al. (2017): Resumen

1. Estimar un modelo GARCH(1,1) para cada activo.
2. Usar las varianzas dinámicas estimadas para estandarizar los retornos y obtener $\hat{\varepsilon}_t$.
3. Usar el estimador mediante reducción no lineal de Ledoit & Wolf en los retornos estandarizados para obtener $\hat{S}$ (hay que corregir el estimador para asegurarse que sea una verdadera matriz de correlaciones).
4. Maximizar la verosimilitud compuesta para estimar (α_q, β_q).

Evaluando estimadores

- La potencia de un estimador de matriz de covarianzas puede estudiarse usándolo para construir portafolios de mínima varianza cuyo desempeño refleja la capacidad del estimador para identificar la mejor forma de diversificar riesgo, independiente del *view* de mercado.
- Usaremos distintos estimadores para crear estos portafolios, usando datos semanales de 36 índices de RV.
- La muestra de datos va desde Enero de 1995 a Junio de 2017. El *back - test* comprende el periodo Enero 2006 – Junio 2017. Los portafolios se rebalancean semanalmente.
- Los retornos *ex-post* de los portafolios se usan para calcular retorno promedio, volatilidad y ratio de Sharpe.
- Este ejercicio es similar al que llevan a cabo Engle et al. (2017) en su artículo.

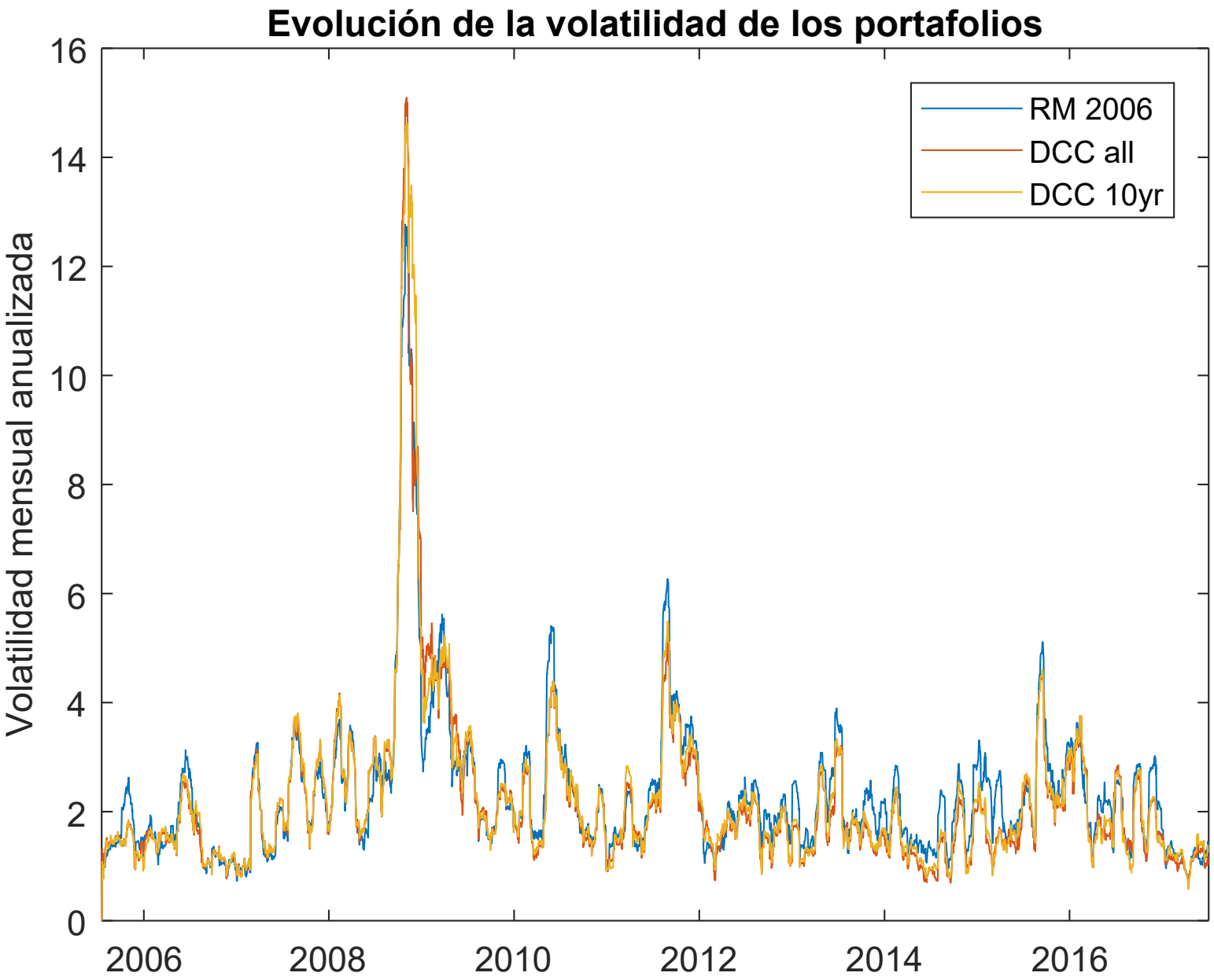
Resultados

Desempeño de Portafolios Mínima Varianza: Back - Test 2006 - 2017*

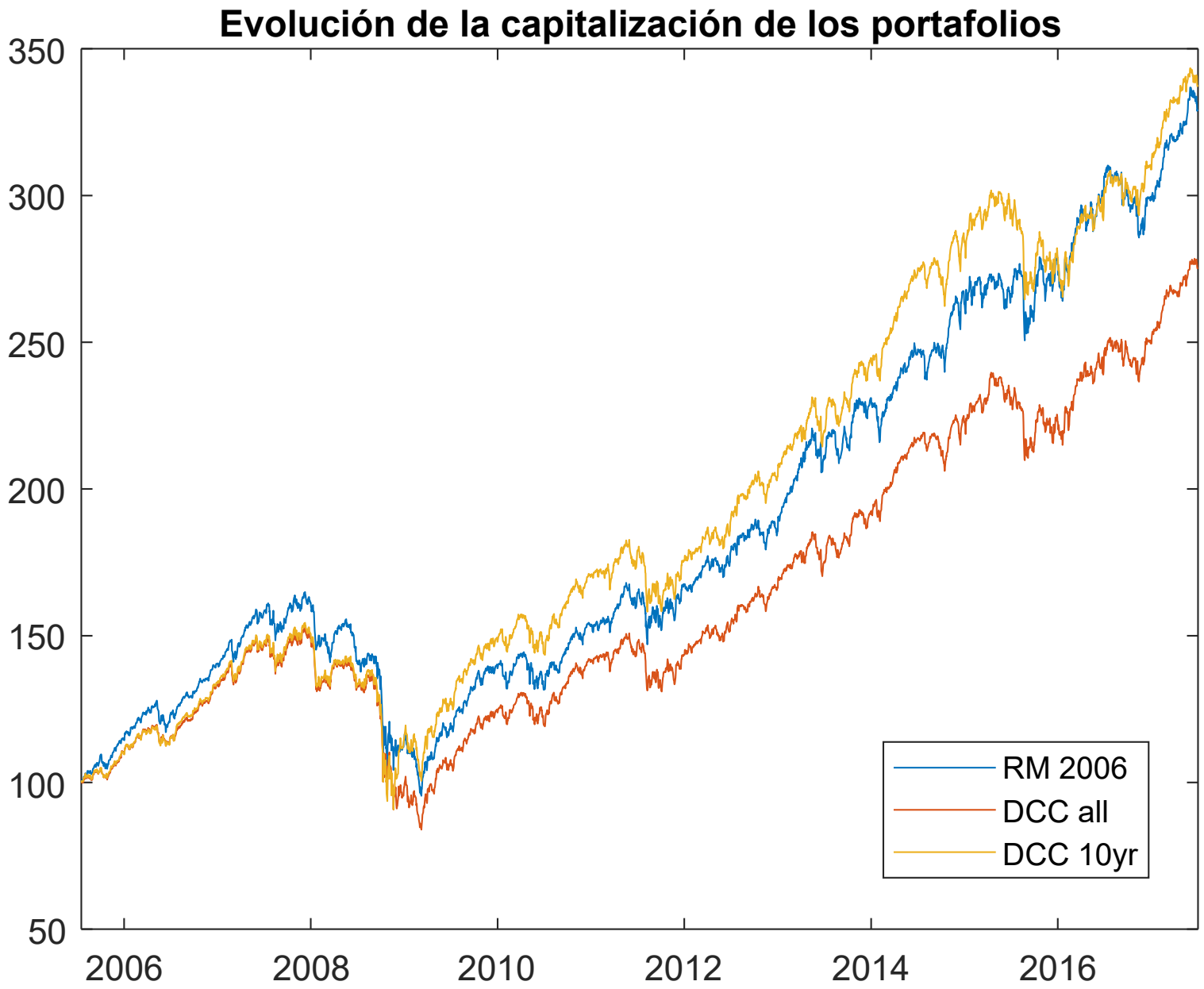
	Portafolio	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017*	'06-'10	'11-'15	'10-'17*	'06-'17*
Retorno	ACWI	19.5	11.6	-54.2	30.4	12.5	-7.1	14.8	21.1	4.6	-1.8	8.2	**11.2**	4.0	6.5	8.6	6.2
	1/N	**23.7**	**15.8**	-49.9	**36.8**	**14.3**	-8.4	**15.5**	14.4	4.3	-3.0	**12.9**	9.0	8.1	4.7	8.0	7.5
	Hist.	20.7	15.7	-39.9	22.1	8.7	6.3	8.4	21.4	11.8	**6.0**	6.0	7.8	5.5	11.0	10.3	8.3
	RM 2006	20.3	13.8	-34.3	19.4	9.8	**8.6**	9.8	**21.8**	13.0	5.0	7.6	9.8	5.8	**11.8**	**11.5**	9.2
	DCC (all)	19.2	11.3	-41.6	23.0	13.0	1.7	14.2	14.5	14.6	0.1	11.3	9.8	5.0	9.2	10.6	8.0
	DCC (10yr)	20.1	11.7	**-27.0**	25.2	14.0	3.2	14.4	18.0	**14.9**	-1.9	10.2	8.8	**8.8**	9.9	11.0	**9.8**
Volatilidad	ACWI	10.1	13.0	32.4	22.7	16.2	20.6	12.3	9.7	8.7	12.8	12.7	6.1	20.6	13.5	13.4	16.7
	1/N	11.3	13.6	31.5	21.5	14.9	18.8	11.6	9.6	9.6	13.6	13.6	6.3	20.1	13.1	13.1	16.4
	Hist.	8.5	10.5	24.6	14.3	12.1	13.6	8.5	10.1	9.0	12.7	10.9	6.2	15.2	11.0	10.9	12.8
	RM 2006	7.5	**10.0**	**24.2**	**14.3**	12.0	13.8	8.4	10.1	8.7	12.0	10.2	5.6	**14.9**	10.8	10.6	12.5
	DCC (all)	**7.0**	10.2	26.9	14.9	**10.7**	**12.2**	**7.6**	**8.4**	**6.7**	**10.3**	9.5	**5.1**	15.7	**9.3**	**9.3**	**12.4**
	DCC (10yr)	7.3	10.3	28.4	14.7	10.9	12.8	7.8	8.7	7.2	10.6	**9.5**	5.6	16.2	9.7	9.6	12.8
Sharpe	ACWI	1.9	0.9	-1.7	1.3	0.8	-0.3	1.2	**2.2**	0.5	-0.1	0.6	1.8	0.2	0.5	0.6	0.4
	1/N	2.1	1.2	-1.6	1.7	1.0	-0.4	1.3	1.5	0.5	-0.2	0.9	1.4	0.4	0.4	0.6	0.5
	Hist.	2.4	**1.5**	-1.6	1.5	0.7	0.5	1.0	2.1	1.3	**0.5**	0.6	1.3	0.4	1.0	0.9	0.6
	RM 2006	2.7	1.4	-1.4	1.4	0.8	**0.6**	1.2	2.1	1.5	0.4	0.7	1.8	0.4	**1.1**	1.1	0.7
	DCC (all)	2.7	1.1	-1.5	1.5	1.2	0.1	**1.9**	1.7	**2.2**	0.0	**1.2**	**1.9**	0.3	1.0	**1.1**	0.6
	DCC (10yr)	**2.8**	1.1	**-0.9**	**1.7**	**1.3**	0.2	1.8	2.1	2.1	-0.2	1.1	1.6	**0.5**	1.0	1.1	**0.8**

* Al cierre de junio

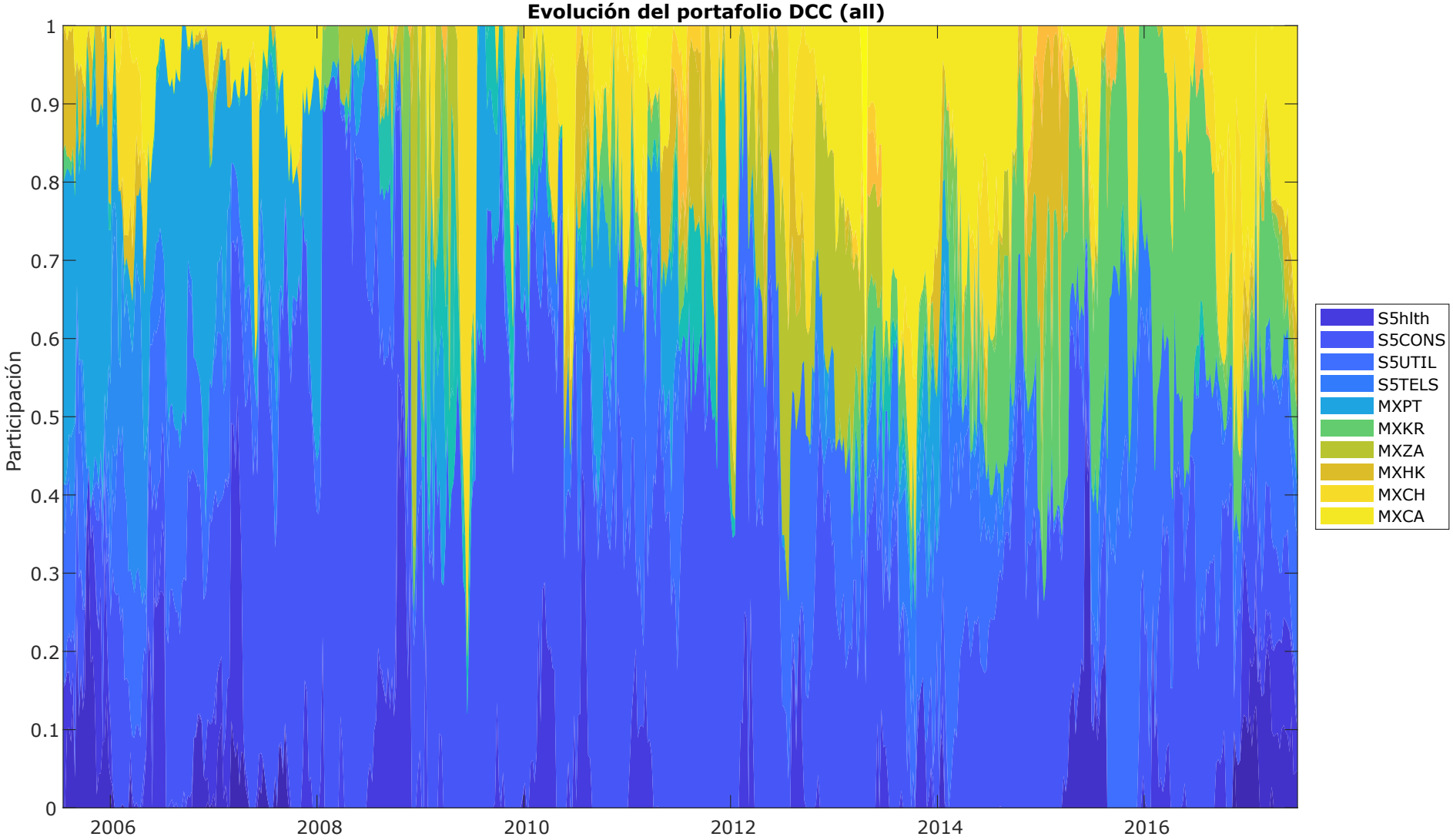
Volatilidad de los portafolios



Capitalización de los portafolios



Composición dinámica del portafolio DCC

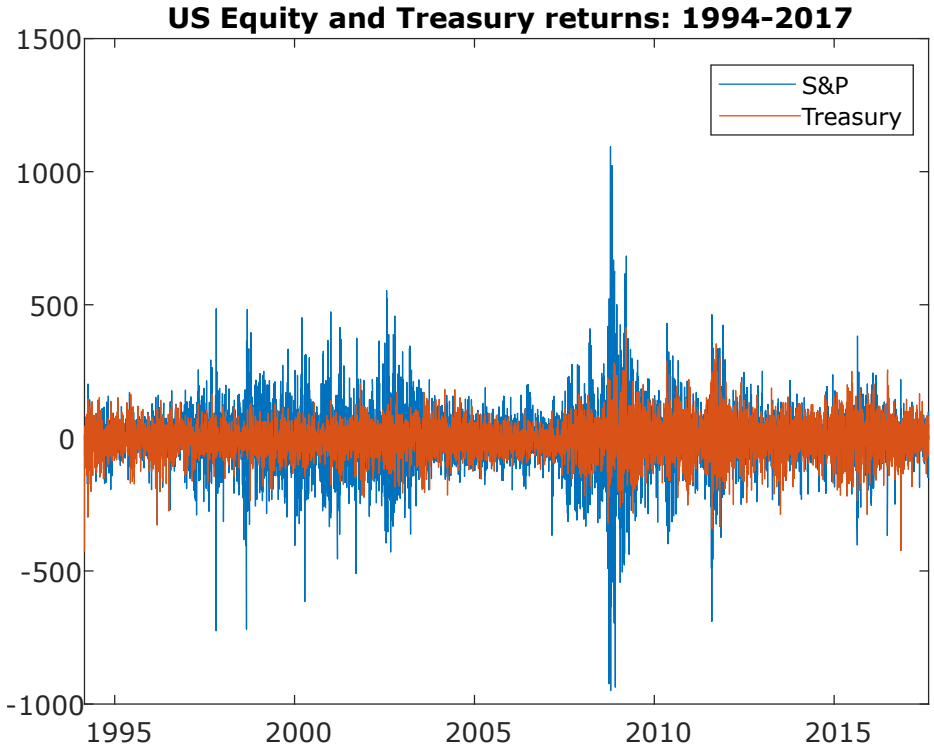


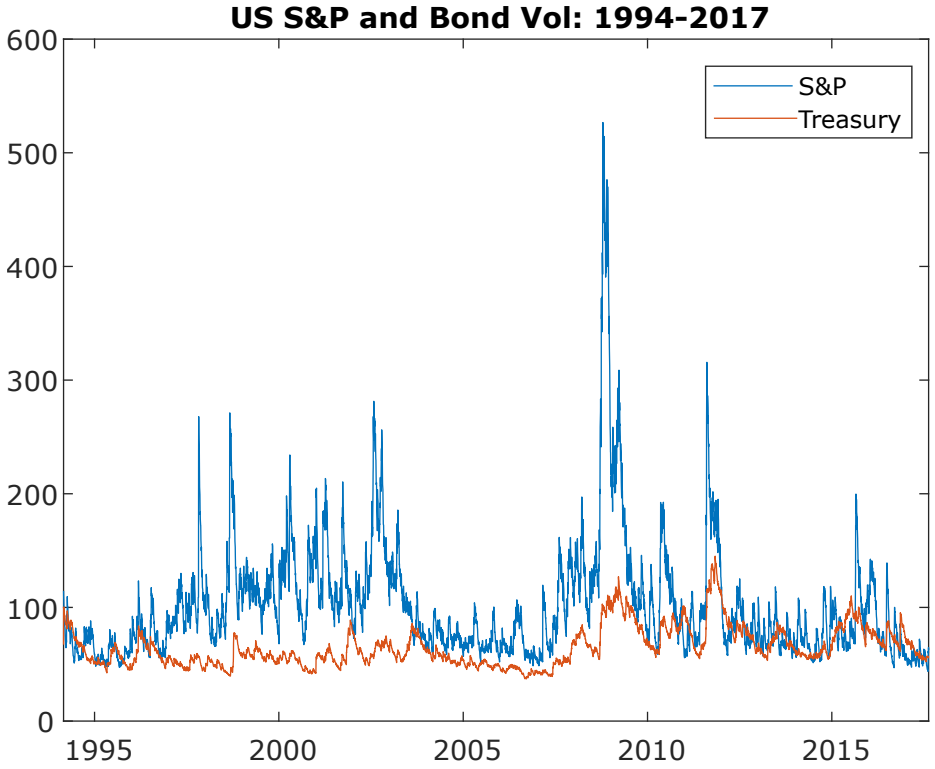
Discusión

- El estimador DCC muestra un desempeño superior (menor volatilidad) a otros estimadores en el *back - test*.
- Además, tiene las ventajas de ser aplicable a conjuntos más grandes de activos ($N > 1000$) y manejar alta frecuencia (diaria) sin problemas.
- En la práctica, el tamaño de la ventana sobre la cual conviene aplicar el estimador DCC es relevante para obtener un desempeño óptimo (en el ejercicio se considera una ventana de 10 años). Dicha ventana determina algunos componentes clave, como la matriz de correlaciones de equilibrio, S.

Aplicaciones: Correlaciones dinámicas

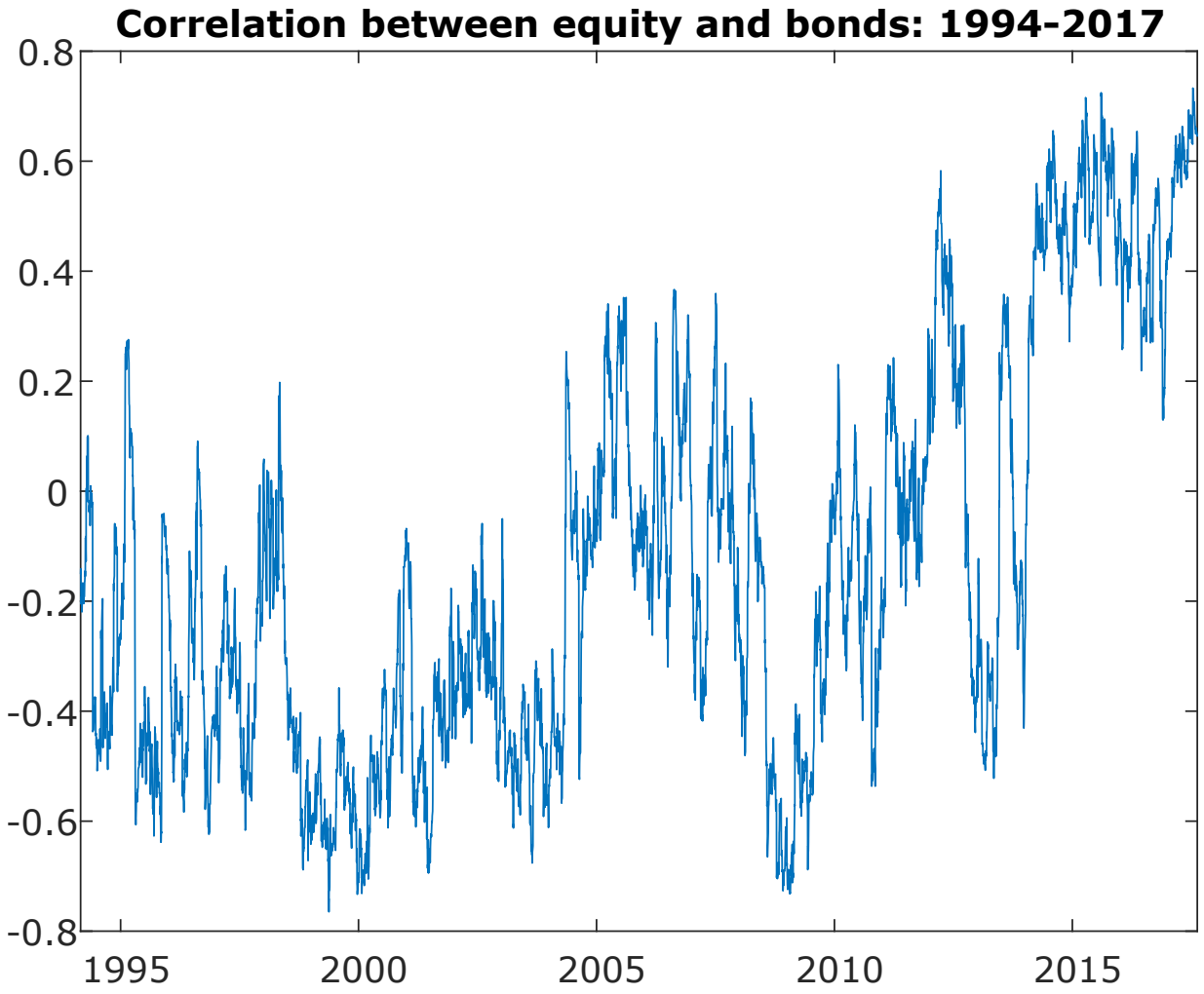
Una de las primeras aplicaciones de esta metodología ha sido el análisis de la correlación entre el índice de la bolsa de Estados Unidos y el retorno que arrojan los bonos del Tesoro.





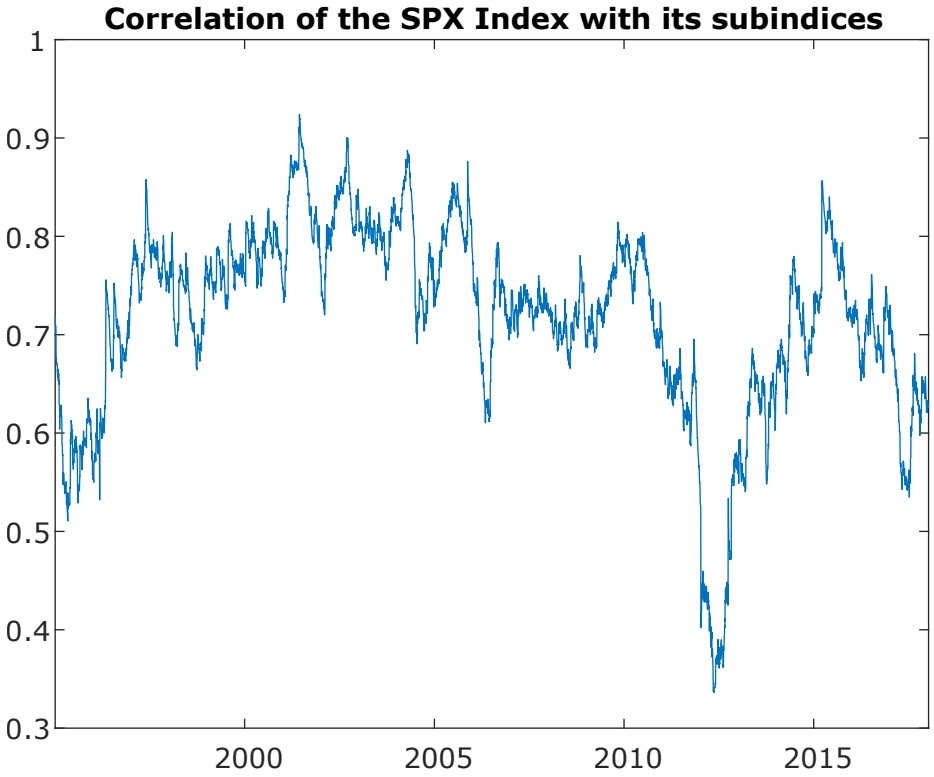
Aplicaciones: Correlaciones dinámicas (II)

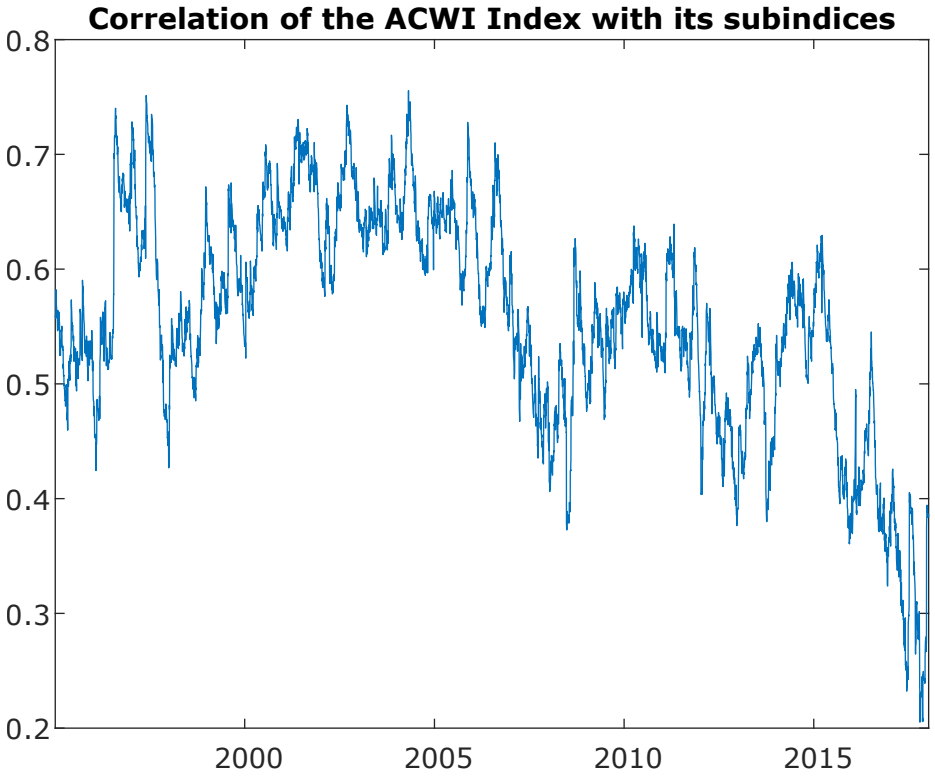
Aplicando el estimador DCC a los retornos del S&P y los bonos del Tesoro, obtenemos una serie de correlaciones que muestran una evolución interesante en los últimos años.



Aplicaciones: Correlaciones dinámicas (III)

Podemos hacer un ejercicio similar para observar la dinámica de la correlación promedio del S&P con sus subíndices y de un índice de renta variable global con sus componentes.





Aplicaciones: Betas dinámicos

- Como última aplicación, consideraremos el ejercicio clásico de identificar el beta de un índice de renta variable global contra sus componentes.
- La estimación de covarianzas dinámicas nos permite calcular betas dinámicos, dándonos una idea de cómo evolucionan estas relaciones en el tiempo.
- Recuerde que si

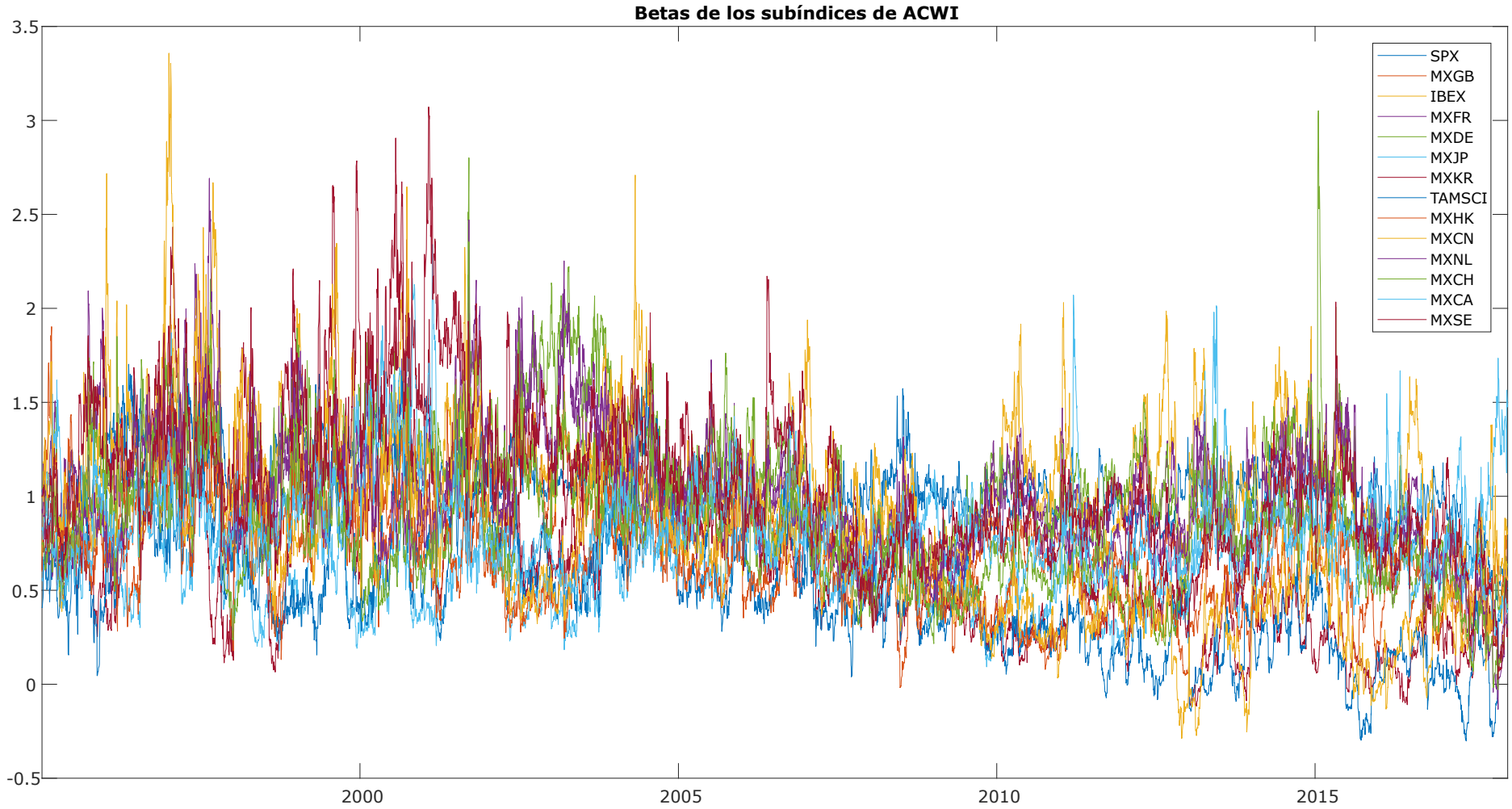
$$y = \beta x + \varepsilon$$

entonces

$$\hat{\beta} = \frac{\text{cov}\,(x, y)}{\sigma_x^2}$$

- El estimador DCC nos permite obtener series de tiempo para cov(x, y) y σ_x^2.

Aplicaciones: Betas dinámicos (II)



Aplicaciones: Betas dinámicos (III)

- Como se puede observar, la estimación arroja información interesante respecto a la forma cómo vienen evolucionando los betas de los diferentes países respecto al mercado global.
- Sin embargo, la estimación se ve bastante ruidosa (esto también puede observarse en las correlaciones dinámicas) lo cual indica que podría mejorarse la estimación dándole más estabilidad al proceso dinámico de correlaciones (quizás también al de volatilidades).
- ¿Qué técnicas podrían usarse para este fin? Quizás sería necesario disciplinar los estimadores usando *priors* bayesianos. Pero este es el estado actual en el que se encuentra la literatura.