

## CAPÍTULO II

**Modelos de selección de carteras con muchos activos**

Christian Brownlees  
Jordi Llorens  
Nuria Senar

Este capítulo trata el problema de selección de carteras de inversión con un gran número de activos financieros. En particular, se repasa la literatura en modelización de correlaciones condicionales dinámicas de elevadas dimensiones (*DCC*, por sus siglas en inglés). Consideramos diferentes tipos de especificaciones, en particular, la versión estándar del modelo *DCC*, el *DCC* con estructura de factores, y el *DCC* con regularización. Introducimos métodos de estimación específicamente diseñados para modelos de elevada dimensionalidad. Evaluamos su capacidad de predicción a través de una aplicación en selección de carteras de inversión con los constituyentes del índice S&P 500.

*Palabras clave:* volatilidad multivariante, correlaciones condicionales dinámicas, regularización de covarianzas, verosimilitud compuesta, selección de carteras.

## 1. INTRODUCCIÓN

El problema de selección de carteras es de suma importancia en el sector financiero. En la conocida teoría moderna de carteras –*MPT*, por sus siglas en inglés– desarrollada por Markowitz (1952), se demuestra que el problema de inversión únicamente depende de la media y la matriz de covarianzas de un vector de retornos de activos. Consecuentemente, el problema de selección de carteras en la *MPT* se traduce en un problema de estimación de la media y la matriz de covarianzas de los retornos. La evidencia empírica nos muestra que la matriz de covarianzas es un elemento clave en selección de carteras y su estimación ha despertado un gran interés tanto en la academia como en la industria. La estrategia más natural es construir la cartera mediante la matriz de covarianzas muestral, pero esto tiene algunas limitaciones. Por un lado –en la dimensión temporal– las volatilidades y correlaciones de los activos financieros son *dinámicas*, esto es, que varían a lo largo del tiempo. Por otro lado –en la sección cruzada– los métodos de estimación tradicionales de las covarianzas en contextos de *elevada dimensionalidad*, i.e. cuando el número de activos financieros es de un orden similar al número de observaciones en muestra– son imprecisos. Por ejemplo, véase Friedman, Hastie y Tibshirani (2008) o Pourahmadi (2013).

En este trabajo presentamos metodologías para lidiar con ambos tipos de retos, concretamente la familia de modelos GARCH-DCC. En las últimas dos décadas, la metodología GARCH-DCC introducida por Engle (2002) se ha establecido dentro de la literatura como una de las referencias clave tanto por su flexibilidad como su viabilidad de implementación. En síntesis, la estrategia consiste en modelizar por separado las varianzas y las correlaciones condicionales. Las varianzas condicionales de cada activo se modelizan individualmente vía GARCH (Bollerslev, 1986) mientras que la matriz de correlaciones condicionales se modeliza conjuntamente mediante el modelo de correlaciones condicionales dinámicas (*DCC*, por sus siglas en inglés). La modelización vía GARCH ha sido estudiada en profundidad y tiene ya una larga tradición en la literatura (Bollerslev, Engle y Nelson, 1994). En cambio, el modelo DCC fue desarrollado posteriormente y a día de hoy sigue estando en la frontera de la investigación en econometría financiera.

Este capítulo introduce el modelo DCC de Engle (2002) en su versión estándar y consideramos algunas de sus extensiones más importantes: el DCC con estructura de factores (*Factor DCC*) y el DCC con regularización lineal y no lineal (*Shrinkage DCC*). Una de las grandes ventajas de esta familia de modelos es que su estimación es escalable a dimensiones relevantes para aplicaciones prácticas.

En el modelo DCC estándar el proceso de correlaciones condicionales dinámicas viene determinado por los retornos estandarizados, es decir, los datos divididos por su volatilidad. El modelo se basa en una ecuación recursiva que garantiza que las matrices de correlaciones dinámicas son definidas positivas.

El modelo Factor DCC se basa en descomponer la matriz de covarianzas dinámica en un factor común y un componente idiosincrásico. En la literatura hay una larga tradición de trabajos que abordan la cuestión de la modelización de activos financieros mediante modelos de factores. Estos modelos se basan en el supuesto de que los precios y las volatilidades de diferentes activos vienen determinados por un pequeño número de factores, lo que determina su comovimiento. Los modelos de factores se pueden dividir en dos tipologías: factores latentes y factores observables. Algunos ejemplos que emplean factores latentes son Diebold y Nerlove (1986), Harvey, Ruiz y Sentana (1992), o el GARCH factorial ortogonal de Alexander y Chibumba (1996). Han (2006) y Aguilar (2009) combinan modelos multivariantes con factores latentes y volatilidad estocástica. Un ejemplo de modelo de factores latentes más reciente es Hallin *et al.* (2019), que emplea componentes principales dinámicos siguiendo la estrategia de Forni *et al.* (2015). Por el lado de los factores observables tenemos el reciente trabajo de De Nard, Ledoit y Wolf (2020), donde se utiliza el modelo DCC con estructura de factores. El modelo de factores que presentamos en detalle más adelante toma como referencia este último trabajo.

El modelo Shrinkage DCC combina la literatura en estimación regularizada de grandes matrices de covarianzas con la estimación dinámica de matrices de covarianzas. Es conocido, como mínimo desde Ledoit y Wolf (2004b), que cuando el número de variables es elevado con respecto al número de observaciones la matriz de covarianzas muestral suele tener un error de estimación preocupante. Esto ha motivado una amplia literatura que propone el uso de la regularización para aumentar la precisión en la estimación de la matriz de covarianzas (Bickel y Levina, 2008; Fan, Liao y Mincheva, 2013). Véase Pourahmadi (2013) para un repaso de la literatura relacionada. En una serie de trabajos, Olivier Ledoit y Michael Wolf han propuesto diferentes estimadores regularizados que han demostrado mejorar los resultados en aplicaciones financieras como la selección de carteras. Recientemente, Engle, Ledoit y Wolf (2019) incorporan esta metodología de regularización al modelo DCC. En este trabajo se ofrece un breve resumen de la versión del modelo con las técnicas de regularización lineal y no lineal.

Ilustramos los beneficios de las metodologías aquí introducidas a través de un ejercicio de selección de carteras de inversión con los constituyentes del índice S&P 500. El ejercicio en cuestión está estrechamente relacionado con los trabajos de Hautsch, Kyj y Malec (2015), Hautsch y Voigt (2019), Engle, Ledoit y Wolf (2019) y De Nard, Ledoit y Wolf (2020). En particular, evaluamos la precisión de los diversos estimadores de manera indirecta (Patton y Sheppard, 2009) mediante la volatilidad de la cartera (dinámica) de mínima varianza y mediante la ratio de información de la cartera de Markowitz con señal de *momentum*. Los resultados demuestran que las metodologías propuestas permiten construir carteras de inversión con rentabilidades ajustadas al riesgo superiores a los puntos de referencia. En particular, el valor diferencial que aportan los modelos Factor y Shrinkage DCC es significativo tanto desde el punto de vista económico como estadístico.

El modelo DCC pertenece a la familia de modelos generalizados de heteroscedasticidad condicional multivariante (MGARCH, por sus siglas en inglés), que ofrecen la posibilidad de estimar y predecir matrices de covarianzas dinámicas de un gran número de activos y que se han utilizado con éxito en aplicaciones de selección de carteras de inversión (Engle, Ledoit y Wolf, 2019), de gestión de riesgos (Ferreira, 2005) y de medición del riesgo sistémico (Brownlees y Engle, 2017). Fuera del sector financiero, estos modelos han sido empleados para modelizar sistemas de variables macroeconómicas como, entre otros, la interacción inflación-crecimiento (Conrad, Karanasos y Zeng, 2010) o la relación entre la evolución del petróleo y el oro (Bampinas y Panagiotidis, 2015). En Alessi, Barigozzi y Capasso (2009) se aplica, con éxito considerable, una combinación de modelo de factores dinámicos y MGARCH en series de inflación y en retornos de activos financieros. Para un repaso general de la literatura MGARCH, véanse, por ejemplo, Bauwens, Laurent, y Rombouts (2006), Engle (2009) Silvennoinen y Teräsvirta (2009) y de Almeida, Hotta, y Ruiz (2018).

Otra importante familia de modelos que ha suscitado un gran interés en la comunidad académica son los modelos de volatilidad estocástica, véanse Shephard (1996) y Broto y Ruiz (2004) para una revisión bibliográfica más detallada para modelos univariados, y Asai, McAleer y Yu (2006) para el caso multivariante. En resumen, la diferencia clave entre la familia de modelos MGARCH y la volatilidad estocástica multivariante reside en que los últimos modelizan la secuencia de matrices de covarianzas como un proceso estocástico, mientras que en los primeros el proceso es determinístico. En términos generales, se puede argumentar que los modelos de volatilidad estocástica ofrecen una mayor flexibilidad de modelización, pero pagan el precio de una mayor dificultad para llevar a cabo su estimación.

El resto del capítulo se estructura de la siguiente forma: la sección segunda sienta las bases metodológicas del capítulo; la subsección 2.1. introduce la notación de la versión canónica del modelo DCC y en la 2.2. tratamos su estimación; la subsección 2.3. explica cómo realizar predicciones con este modelo; la subsección 2.4. introduce el modelo DCC con estructura de factores y en la 2.5. consideramos el modelo DCC con regularización. La sección tercera presenta la metodología empleada para selección de carteras, en particular la cartera de mínima varianza (3.1.) y la cartera de Markowitz con señal de *momentum* (3.2.). La sección cuarta se dedica a la aplicación empírica; la sección 5 concluye el capítulo.

## 2. METODOLOGÍA

Esta sección fija la notación, así como las bases metodológicas sobre las que se asienta el capítulo. A lo largo de todo el capítulo, y salvo algunas excepciones donde no cabe ambigüedad, los escalares como  $x$  se denotan con minúsculas, los vectores como  $\mathbf{x}$  en negrita minúscula, y las matrices como  $\mathbf{X}$  con letra mayúscula. Un individuo de la sec-

ción cruzada (normalmente interpretado como el  $i$ -ésimo activo financiero) se denota con el subíndice  $i=1, \dots, N$ , y para la dimensión temporal se utiliza el subíndice  $t=1, \dots, T$ .

## 2.1. Modelo de correlaciones condicionales dinámicas (DCC)

Denotamos como  $\{\mathbf{r}_t\}_{t=1}^T$  a la serie de retornos de  $N$  activos con media cero<sup>1</sup>. Se asume que el proceso generativo de los retornos viene dado por:

$$\mathbf{r}_t = \Sigma_t^{1/2} \mathbf{z}_t, \quad \mathbf{z}_t \sim \mathcal{D}(\mathbf{0}, I) \quad \text{i.i.d.}$$

donde  $\Sigma_t \equiv \text{Var}(\mathbf{r}_t | \mathcal{F}_{t-1})$  es definida positiva,  $\mathcal{F}_{t-1}$  es el conjunto de información disponible en el período  $t-1$ , y  $\mathcal{D}$  es una distribución multivariante estandarizada de la familia localización-escala (e.g., Normal,  $t$ -Student, etc.). El objetivo principal es encontrar una buena predicción de  $\Sigma_t$ , es decir, de la matriz dinámica de covarianzas de los retornos. Dicha matriz puede ser reparametrizada como sigue:

$$\Sigma_t = D_t R_t D_t, \quad [1]$$

donde  $D_t$  es la matriz (diagonal) de volatilidades dinámicas, con su  $i$ -ésimo elemento dado por  $d_{i,t} \equiv \sqrt{\text{Var}(r_{i,t} | \mathcal{F}_{t-1})}$ , y siendo  $R_t \equiv \text{Var}(\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1}) = \mathbb{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' | \mathcal{F}_{t-1}]$  la matriz de correlaciones dinámicas, donde  $\boldsymbol{\varepsilon}_t = D_t^{-1} \mathbf{r}_t$ . Esta reparametrización permite la desagregación del problema en dos partes: por un lado, la estimación de las volatilidades y, por el otro, la de la matriz de correlaciones. Asumimos que la volatilidad de cada activo sigue un proceso estacionario GARCH (Bollerslev, 1986), tal y como se recoge en la ecuación [2], donde  $a_i$  y  $b_i$  son escalares positivos tales que  $a_i + b_i < 1$  y  $d_i \equiv \sqrt{\text{Var}(r_{i,t})}$  es la volatilidad incondicional del  $i$ -ésimo activo:

$$d_{i,t}^2 = (1 - a_i - b_i) d_i^2 + a_i r_{i,t-1}^2 + b_i d_{i,t-1}^2. \quad [2]$$

La dinámica de las correlaciones depende de los retornos estandarizados  $\boldsymbol{\varepsilon}_t$ , de acuerdo con la siguiente ecuación recursiva:

$$Q_t = (1 - \alpha - \beta) C + \alpha \boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}' + \beta Q_{t-1}, \quad [3]$$

donde  $C$  es una matriz  $N \times N$  positiva definida y  $\alpha$  y  $\beta$  son escalares positivos tales que  $\alpha + \beta < 1$ . Nótese que la dinámica de  $Q_t$  es muy similar a la del modelo VECH escalar (Engle, 2009).

Bajo estos supuestos,  $\{Q_t\}_{t=1}^T$  es una secuencia estacionaria de matrices definidas positivas. Sin embargo, su diagonal no es necesariamente unitaria y sus elementos no están

<sup>1</sup> Entendemos aquí  $\mathbf{r}_t$  como el vector de retornos. Al estar considerando retornos de acciones, el supuesto de media zero en frecuencias diarias es razonable. Sin embargo, queda entendido que también podríamos interpretar  $\mathbf{r}_t$  como los residuos de un modelo VARMA sin pérdida de generalidad en la exposición.

necesariamente acotados en el intervalo  $[-1,1]$ , lo que significa que  $Q_t$  no puede ser considerada como una matriz de correlaciones propiamente dicha. Por su similitud con la matriz  $R_t$ , a  $Q_t$  la denominamos matriz de pseudo-correlaciones. Es sencillo reescalar la misma para obtener la matriz de correlaciones condicional:

$$R_t = \text{diag}(Q_t)^{-1/2} Q_t \text{diag}(Q_t)^{-1/2}.$$

Aunque reescalar parece una opción bastante natural, no es obvio que  $R_t$  sea la matriz de correlaciones más próxima a  $Q_t$ . De otra manera,  $R_t$  no es una proyección de  $Q_t$  hacia el espacio de matrices de correlaciones, y existen otras estrategias propuestas en la literatura para llevar a cabo esta normalización (Tse y Tsui, 2002; Brownlees y Llorens, 2020). Es posible considerar especificaciones más generales con un mayor número de retardos tanto en  $\varepsilon_t$  como en  $Q_t$ , pero empíricamente se ha documentado que añadir más retardos no suele mejorar la capacidad de predicción de una manera significativa. Para abreviar, utilizamos la notación  $\mathbf{r}_t \sim \text{DCC}(1, 1)$  para expresar que los retornos siguen el modelo descrito por las ecuaciones [1] a [3].

## 2.2. Estimación en el modelo DCC

### 2.2.1. Focalización en covarianzas y estimación por (cuasi) máxima verosimilitud

Desde el punto de vista de la estimación, se han de obtener, primero, las volatilidades para, en un segundo paso, estimar las correlaciones, pues estas dependen de  $\varepsilon_t$ —que a su vez depende de  $D_t$ —. Es posible estimar  $\alpha$ ,  $\beta$  y  $C$  conjuntamente por máxima verosimilitud, pero esto supone resolver un problema de optimización en  $\binom{N}{2} + N + 2$  variables, lo cual resulta poco atractivo a medida que el número de activos crece. En su lugar, se recurre a la técnica de focalización en la covarianza (*covariance targeting*) que consiste en reemplazar  $C$  por un estimador  $f$  de la matriz de covarianzas de  $\varepsilon_t$  que no dependa de  $\alpha$  y  $\beta$ , es decir,

$$\hat{C} = f\left(\{\varepsilon_t\}_{t=1}^T\right). \quad [4]$$

Por ejemplo, en el modelo DCC canónico,  $f\left(\{\varepsilon_t\}_{t=1}^T\right) = \frac{1}{T} \sum_{t=1}^T \varepsilon_t \varepsilon_t'$ . La técnica de focalización tiene una justificación clara en los modelos estacionarios GARCH multivariantes (como, por ejemplo, el VECM escalar), puesto que, en ese caso, la matriz interceptor es equivalente al segundo momento de los datos. Desde el punto de vista teórico, la técnica de focalización en el modelo DCC es inconsistente puesto que en general  $C \neq \mathbb{E}[\varepsilon_t \varepsilon_t']$ , algo que fue señalado en primer lugar por Aielli (2013), quien a su vez propuso una versión corregida de ese modelo, abreviada como cDCC. Sin embargo, la inconsistencia

es menor y en la práctica se sigue empleando la versión canónica del modelo DCC con focalización en covarianzas, lo cual es, naturalmente, mucho más conveniente desde la perspectiva computacional.

Los parámetros  $\alpha$  y  $\beta$  se estiman maximizando el logaritmo de la cuasi-verosimilitud Gaussiana, que viene dada por:

$$\sum_{t=1}^T l_t(\theta) = \sum_{i=1}^N \mathcal{L}_{i,T}^G(\theta_i^G) + \mathcal{L}_T^{\text{DCC}}(\alpha, \beta),$$

donde  $\mathcal{L}_{i,T}^G(\theta_i^G)$  es la verosimilitud logarítmica del GARCH del  $i$ -ésimo activo,  $\theta_i^G := (d_i, a_i, b_i)'$ , y  $\theta = (\alpha, \beta, \theta_1^G, \dots, \theta_N^G)'$  es el vector que reúne todos los parámetros del modelo. Nótese que la dependencia en la matriz  $C$  se omite ya que  $C$  se estima como en la ecuación [4]. También es posible asumir que las innovaciones siguen otras distribuciones como, por ejemplo, la  $t$ -Student multivariante (Pesaran y Pesaran, 2007), cuyos extremos son notablemente más gruesos que los de la distribución Gaussiana.

### 2.2.2. Verosimilitud compuesta

La evaluación de la verosimilitud logarítmica de la familia de modelos DCC requiere el cálculo del determinante de la matriz de correlaciones dinámica  $R_t$ , así como de su inversa  $R_t^{-1}$ . La complejidad de las computaciones mencionadas es aproximadamente de  $\mathcal{O}(TN^3)$ , un cálculo progresivamente arduo a medida que la dimensión de la serie de retornos crece. Como el estimador de máxima verosimilitud se obtiene mediante métodos numéricos, la función de verosimilitud logarítmica ha de evaluarse múltiples veces, por lo tanto, la relevancia práctica de la cuestión es más que considerable.

El método de *verosimilitud compuesta* fue introducido por Pakel, Engle, Shephard y Sheppard (2017). Este se basa en la aproximación de la función de verosimilitud logarítmica conjunta empleando verosimilitudes marginales bivariantes. Definimos la verosimilitud logarítmica marginal del  $j$ -ésimo par de activos como:

$$l_{jT}(\alpha, \beta) = -\frac{1}{2} \sum_{t=1}^T \left( \log |R_{jt}| + \varepsilon'_{jt} R_{jt}^{-1} \varepsilon_{jt} \right), \quad [5]$$

donde se entiende que  $\varepsilon_{jt} = (r_{j1t}/d_{j1t}, r_{j2t}/d_{j2t})'$  y que  $R_{jt}$  es su correspondiente matriz de correlaciones condicionales. De tal manera, es posible aproximar la verosimilitud logarítmica tomando el promedio de todos los pares, como se representa en la ecuación [6] en la que  $L = \binom{N}{2}$  es el número de todas las posibles parejas de  $N$  activos. A esta aproximación se la denomina verosimilitud compuesta,

$$\mathcal{L}_{LT}^{\text{DCC}}(\alpha, \beta) = \frac{1}{L} \sum_{j=1}^L l_{jT}(\alpha, \beta). \quad [6]$$

La complejidad del problema queda, por tanto, reducida a  $\mathcal{O}(TN^2)$ . El proceso para una distribución  $t$ -Student es análogo al descrito en los párrafos previos. Cabe recalcar que naturalmente la verosimilitud logarítmica de cada pareja  $j$  depende de una matriz  $C_j$  constante  $2 \times 2$  análoga a la matriz  $C$  de la ecuación [3], que estimamos vía focalización en covarianzas como en [4].

La versión más común de la verosimilitud compuesta utiliza solamente parejas *contiguas*, esto es,  $X_{1t} = (r_{1t}, r_{2t})', \dots, X_{N-1t} = (r_{N-1t}, r_{Nt})'$ , simplificándose el problema en cuestión en  $\mathcal{O}(TN)$ . Pakel, Engle, Shephard y Sheppard (2017) demuestran que maximizar la verosimilitud logarítmica compuesta resulta en estimaciones consistentes (e incluso eficientes) de los parámetros  $\alpha$  y  $\beta$ .

### 2.3. Predicción en el modelo DCC

La predicción a  $k$  períodos vista desde la fecha  $h$  de construcción de una cartera basada en un modelo DCC se puede realizar en tres pasos, dados por las ecuaciones [7], [8] y [9] que explicamos a continuación.

La predicción a  $k$  períodos vista desde la fecha  $h$  de las varianzas condicionales del proceso GARCH(1,1) para el  $i$ -ésimo activo viene dada por:

$$\mathbb{E}[d_{i,h+k}^2 | \mathcal{F}_h] = \begin{cases} (1 - a_i - b_i)d_i^2 + a_i r_{i,h}^2 + b_i d_{i,h}^2 & k = 1 \\ d_i^2 + (a_i + b_i)^{k-1} (\mathbb{E}[d_{i,h+1}^2 | \mathcal{F}_h] - d_i^2) & k > 1 \end{cases} \quad [7]$$

Para simplificar la notación recogemos las predicciones de cada activo a  $k$  períodos vista en la matriz diagonal  $\hat{D}_{h+k}$  de dimensiones  $N \times N$ .

Para la predicción de las correlaciones condicionales a  $k$  períodos vista desde  $h$  adoptamos la estrategia de Engle y Shephard (2001) donde se usa la aproximación  $\mathbb{E}[R_{h+k} | \mathcal{F}_h] = \mathbb{E}[\boldsymbol{\varepsilon}_{h+k} \boldsymbol{\varepsilon}_{h+k}' | \mathcal{F}_h] \approx \mathbb{E}[Q_{h+k} | \mathcal{F}_h]$ . De esta manera, la predicción a  $k$  períodos vista puede escribirse como:

$$\hat{R}_{h+k} \equiv \mathbb{E}[R_{h+k} | \mathcal{F}_h] = \begin{cases} (1 - \alpha - \beta)C + \alpha \boldsymbol{\varepsilon}_h \boldsymbol{\varepsilon}_h' + \beta Q_h & k = 1 \\ C + (\alpha + \beta)^{k-1} (\mathbb{E}[R_{h+1} | \mathcal{F}_h] - C) & K > 1 \end{cases} \quad [8]$$

Combinando ecuaciones [7] y [8], la predicción a  $k$  períodos vista de la matriz de covarianzas condicionales queda determinada por la siguiente ecuación:

$$\mathbb{E}[R_{h+k} | \mathcal{F}_h] = \mathbb{E}[\boldsymbol{\varepsilon}_{h+k} \boldsymbol{\varepsilon}_{h+k}' | \mathcal{F}_h] \approx \mathbb{E}[Q_{h+k} | \mathcal{F}_h]. \quad [9]$$

En la selección de carteras de inversión, es común construir la matriz de covarianzas dinámicas en la fecha  $h$  en base a una única predicción de la matriz de covarianzas

para no realizar ninguna transacción hasta la próxima fecha de inversión  $h + \mathcal{K}$ , donde  $\mathcal{K}$  es un número entero positivo. De esta manera, uno evita incurrir en los costes de transacción provenientes de una rotación excesiva de la cartera. Una estrategia común –véanse, por ejemplo, Engle, Ledoit y Wolf (2019) y De Nard, Ledoit y Wolf (2020) – es calcular el promedio de las predicciones a  $k$  períodos vista, esto es,

$$\hat{\Sigma}_h \equiv \frac{1}{\mathcal{K}} \sum_{k=1}^{\mathcal{K}} \hat{\Sigma}_{h+k}.$$

## 2.4. DCC con estructura de factores

Hay una larga tradición de modelos factoriales dinámicos en macroeconomía y finanzas. En macroeconomía, los modelos factoriales dinámicos se usan frecuentemente para ejercicios de predicción con *big data* (Forni *et al.*, 2000; Stock y Watson, 2002b; Peña y Poncela, 2004). En finanzas, se suele emplear modelos de factores para explicar el comportamiento de los retornos de las acciones. En particular, la teoría del arbitraje de Ross (1976) implica que los retornos de los activos se explican por un pequeño número de factores.

En la literatura financiera se han considerado tanto modelos de factores latentes como observables. En este trabajo nos centramos en estos últimos. Un ejemplo tradicional es el modelo CAPM donde el único factor es la prima de riesgo de mercado  $r_{m,t} - r_{f,t}$ , donde  $r_{m,t}$  es el retorno del mercado y  $r_{f,t}$  el retorno libre de riesgo. Para la serie  $\{r_{m,t}\}$  se suele emplear la media ponderada de los retornos de todos los activos del universo en consideración (por ejemplo, todas las compañías disponibles en la base de datos CRSP cuyas acciones se comercian en NYSE, AMEX o NASDAQ) y para  $\{r_{f,t}\}$  se suele usar el retorno de algún activo que se considere seguro como, por ejemplo, los *Treasury bills* a un mes en EE. UU. o el *Bund* alemán en Europa. Naturalmente, se pueden considerar más factores como el “Small Minus Big” o el “High Minus Low”, pero en la aplicación empírica que consideramos en este trabajo éstos no aportan un gran valor añadido.

Es posible incorporar la estructura de factores al modelo DCC visto anteriormente. En este caso, el vector de retornos  $\mathbf{r}_t$  viene dado por la siguiente ecuación:

$$\mathbf{r}_t = B\mathbf{f}_t + \mathbf{u}_t, \mathbf{u}_t \sim \text{DCC}(1,1)$$

donde  $\mathbf{f}_t$  es un vector que recoge los  $K$  factores y  $B$  es una matriz  $N \times K$ .

Es fácil ver que la matriz de covarianza de los retornos  $\Sigma_t$  se basa en la covarianza de los factores  $\Sigma_f$  y la covarianza de los residuos  $\Sigma_{u,t}$  como sigue a continuación:

$$\Sigma_t = B\Sigma_f B' + \Sigma_{u,t}. \quad [10]$$

Por lo tanto, para implementar el modelo DCC con estructura de factores es suficiente estimar primero  $B$  mediante  $N$  regresiones lineales por mínimos cuadrados ordinarios de las cuales se extraen los residuos  $\mathbf{u}_t$ . Luego, uno puede estimar  $\Sigma_{\mathbf{u},t}$  mediante la metodología DCC estándar. La matriz  $\Sigma_f$  se puede estimar naturalmente vía la covarianza muestral de la serie de factores.

Nótese que también se podría haber considerado un modelo dinámico para los factores, y en ese caso  $\Sigma_{f,t}$  se puede estimar fácilmente también mediante la metodología DCC. También es fácil lidiar con el caso en el que  $B$  es dinámica. Por ejemplo, en el caso  $K = 1$ ,  $B = Cov(\mathbf{r}_t, f_t) / Var(f_t)$ , donde  $Cov(\mathbf{r}_t, f_t)$  es un vector de  $N$  dimensiones con las correspondientes covarianzas. Si se cree que  $B$  es dinámica, entonces  $B_t$  se puede construir estimando  $Cov(\mathbf{r}_t, f_t | \mathcal{F}_{t-1})$  y  $Var(f_t | \mathcal{F}_{t-1})$  mediante modelos GARCH. Sin embargo, ninguna de estas generalizaciones parece aportar un gran valor diferencial en nuestra aplicación práctica (De Nard, Ledoit y Wolf, 2020), de manera que para ahorrar espacio únicamente se reportan los resultados del modelo descrito en [10].

## 2.5. DCC con regularización

La cuestión que aborda esta sección está estrechamente relacionada con el problema de predicción de alta dimensionalidad, también denominado como predicción con big data. Cuando la estructura de los datos es *ancha* –es decir, más variables que observaciones– solemos encontrarlos con problemas de imprecisión y sobreajuste (*overfitting*). En estos contextos, las metodologías de estimación con regularización han demostrado ser extremadamente efectivas en aprendizaje automático supervisado – clasificación y regresión. La razón detrás de esto es que hay un compromiso entre *sesgo* y *varianza* para cada metodología estadística que utilizamos. Dicho de otra forma, un método demasiado simple es generalmente incapaz de capturar patrones interesantes en nuestro conjunto de datos, pero por otro lado es más robusto a ligeras perturbaciones en los datos: su sesgo es elevado y su *varianza* es reducida; por otro lado, un método más complejo puede capturar prácticamente cualquier patrón y si es demasiado complejo también encontrará patrones donde en realidad únicamente hay ruido: en este caso, su sesgo es reducido pero su *varianza* es elevada, ya que el resultado de la estimación puede variar considerablemente dada una pequeña perturbación en los datos. Para un tratamiento más riguroso de la predicción con big data y los problemas de elevada dimensionalidad, véanse Hastie, Tibshirani y Friedman (2008) y Wainwright (2019).

Volviendo al contexto de la metodología DCC, la técnica de focalización en covarianzas requiere un estimador adecuado de la matriz  $C$  basado únicamente en los residuos estandarizados  $\varepsilon_t$  (ecuación [4]). La covarianza muestral es un estimador bastante defectuoso cuando la dimensionalidad del problema es elevada. Esto se debe a que no es posible estimar con exactitud  $\mathcal{O}(N^2)$  parámetros con  $\mathcal{O}(N^2)$  observaciones. En términos técnicos, cuando la *ratio de concentración*  $N / T$  aumenta, la teoría asintótica estándar

no describe adecuadamente el comportamiento real de la matriz de covarianzas. Esto ha motivado una amplia literatura que ha propuesto el uso de estimadores regularizados que funcionan bien cuando la ratio de concentración es elevada. En esta sección introducimos una clase de estimadores regularizados para aumentar la precisión en la estimación con respecto a la covarianza muestral.

El modelo DCC con regularización se define igual que su versión estándar con focalización en covarianzas salvo por el hecho de que la estimación de la matriz  $C$  se realiza mediante un estimador regularizado.

### 2.5.1. Regularización lineal

El método de regularización lineal de Ledoit y Wolf (2004b) cuenta con una fórmula clara, fácil de estimar e interpretar. De hecho, ésta es la combinación lineal convexa asintóticamente óptima de la matriz de covarianza muestral y de la matriz identidad.

Supongamos que  $C = \mathbb{E}[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t']$  —aunque esto no sea del todo correcto en el modelo DCC. El método de regularización lineal consiste en encontrar la combinación lineal:

$$C^* = \rho_1 \mu I + (1 - \rho_1) S, \quad [11]$$

entre la matriz escalar  $\mu I$  y la matriz de covarianzas muestral  $S = \frac{1}{T} \sum_{t=1}^T \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'$  que minimiza la pérdida cuadrática esperada. El coeficiente  $\rho_1$  es la denominada intensidad de regularización (*shrinkage intensity*) y  $\mu I$  el objetivo de regularización (*shrinkage target*). Nótese que, a su vez, la intensidad de regularización en este caso coincide con la ratio de mejora con respecto a la covarianza muestral  $S$  en términos de pérdida cuadrática esperada.

Es posible demostrar que este estimador es consistente bajo teoría asintótica general, es decir, cuando dejamos que  $N$  y  $T$  tiendan a infinito simultáneamente de manera que  $N/T \rightarrow c < \infty$ .

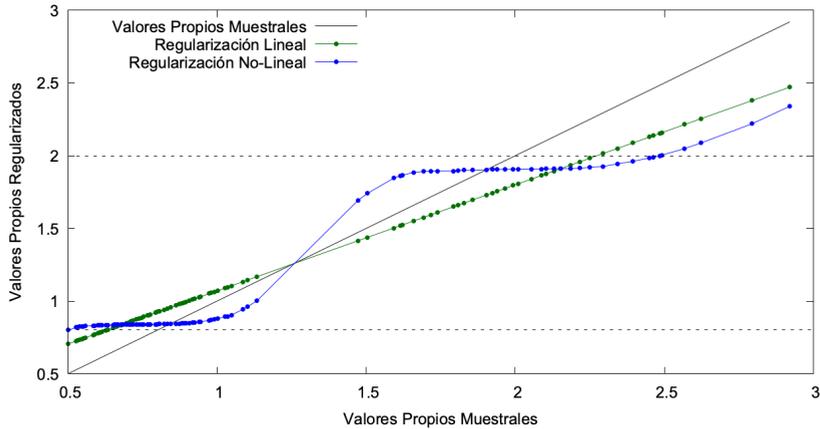
### 2.5.2. Regularización no lineal

La regularización lineal tiene la limitación de que modifica todos los valores propios hacia un mismo objetivo y con la misma intensidad, por lo tanto no permite capturar adecuadamente el hecho de que la regularización es un fenómeno esencialmente local y no global. La figura 1 muestra como el uso de la atracción local proporciona una mejor estimación de los valores propios de la matriz de covarianzas poblacional. En este ejemplo, la matriz covarianzas poblacional con  $N = 100$  tiene 62 valores propios iguales a 0,8, y los 38 restantes son iguales a 2 (líneas punteadas horizontales). El tamaño muestral es  $T = 1260$ . La regularización lineal se calcula como en la sección 2.5.1 y

la regularización no-lineal mediante el procedimiento que explicamos a continuación.

FIGURA 1

REGULARIZACIÓN LINEAL Y NO LINEAL



*Notas:* Efecto de Atracción Local. La matriz de covarianzas poblacional con  $N = 100$  tiene 62 valores propios iguales a 0,8, y los 38 restantes son iguales a 2 (líneas punteadas horizontales). El tamaño muestral es  $T = 1260$ . La regularización lineal se calcula como en la sección 2.5.1, y la regularización no-lineal como se describe en la sección 2.5.2, esto es, mediante el uso de la transformada de Hilbert con kernel de Epanechnikov.

*Fuente:* Elaboración propia.

Ledito y Wolf (2020) introducen una fórmula analítica para la regularización no lineal óptima de matrices de covarianzas de elevadas dimensiones. Dicha fórmula se obtiene observando que el problema está estrechamente relacionado con la estimación (no paramétrica) de la función de densidad de los valores propios de la matriz de covarianzas muestral y de su transformada de Hilbert. En primer lugar, se calcula la descomposición espectral de la matriz de covarianzas muestral  $S$  como se define en la subsección 2.5.1., esto es,

$$S = \frac{1}{T} \sum_{i=1}^T \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i' = \sum_{i=1}^N \hat{\lambda}_i \mathbf{u}_i \mathbf{u}_i', \quad [12]$$

donde  $\lambda_1, \dots, \lambda_N$  son los valores propios muestrales en orden descendente sin pérdida de generalidad y  $\mathbf{u}_1, \dots, \mathbf{u}_N$  son los respectivos vectores propios. Se considera la clase de estimadores *invariantes bajo rotación*, un marco conceptual desarrollado por Stein (1986). Bajo la pérdida de mínima varianza de Engle y Colacito (2006) y la pérdida cuadrática, la Proposición 2.1. en Ledito y Wolf (2020) establece que el estimador óptimo dentro de esa clase se obtiene reemplazando  $\lambda_i$  por  $\hat{d}_i = \mathbf{u}_i' C \mathbf{u}_i$  en [12]. Naturalmente esta cuantía es irrealizable ya que depende de la verdadera matriz  $C$ .

Bajo los supuestos 3.1.-3.4. de Leoit y Wolf (2020), el estimador “oráculo”

$$\forall i=1,\dots,N \quad d^o(\lambda_i) = \frac{\lambda_i}{\left[ \pi \frac{N}{T} \lambda_i f(\lambda_i) \right]^2 + \left[ 1 - \frac{N}{T} - \pi \frac{N}{T} \lambda_i \mathcal{H}_f(\lambda_i) \right]^2} \quad [13]$$

minimiza el límite (con probabilidad 1) de la pérdida de mínima varianza bajo teoría asintótica general. Este estimador sigue siendo irrealizable pero representa un gran progreso con respecto a  $\hat{d}_i$ , ya que éste depende de  $N(N+1)/2$  parámetros (libres) de  $C$ , mientras que  $d^o(\lambda_i)$  depende solamente de  $N$  parámetros (los valores propios de la matriz  $C$ ). Dados los supuestos mencionados, la función  $f(\lambda_i) \equiv \frac{dF}{d\lambda}$  y su transformada de Hilbert ( $\mathcal{H}_f$ ) existen, donde  $F$  es el límite (con probabilidad 1) de la función de distribución empírica de los valores propios de  $S$ .

En la transformada de Hilbert reside gran parte de la clave, ya que es esta transformada la que actúa como una *fuerza de atracción local*, devolviendo valores muy positivos cuando hay una gran proporción de otros valores propios muestrales en puntos ligeramente mayores que  $\lambda_i$ . Similarmente, la transformada devuelve valores muy negativos cuando hay una gran proporción de otros valores propios muestrales en puntos ligeramente menores que  $\lambda_i$ . Si inspeccionamos la ecuación [13], vemos que cuando hay una gran densidad de otros valores propios muestrales ligeramente por encima de  $\lambda_i$ , el valor de la transformada de Hilbert aumenta y el oráculo empuja a  $\lambda_i$  hacia arriba, y viceversa.

Un estimador realizable de la matriz de covarianzas se puede obtener como se representa en la ecuación [14], donde  $\tilde{d}_i$  está definida por la función de regularización no lineal asintóticamente óptima recogida en la ecuación [15].

$$\hat{C} \equiv \sum_{i=1}^N \tilde{d}_i \mathbf{u}_i \mathbf{u}_i' \quad [14]$$

$$\forall i=1,\dots,N \quad \tilde{d}_i(\lambda_i) \equiv \frac{\lambda_i}{\left[ \pi \frac{N}{T} \lambda_i \tilde{f}(\lambda_i) \right]^2 + \left[ 1 - \frac{N}{T} - \pi \frac{N}{T} \lambda_i \mathcal{H}_{\tilde{f}}(\lambda_i) \right]^2}, \quad [15]$$

donde  $\tilde{f}$  y  $\mathcal{H}_{\tilde{f}}$  son estimadores no paramétricos con una amplitud de banda local que es proporcional a cada valor propio  $\lambda_i$ .

Cabe observar que este estimador depende exclusivamente de fórmulas analíticas y por lo tanto no requiere de ningún cálculo numérico para su implementación práctica, lo que lo hace especialmente atractivo en aplicaciones financieras con centenares o incluso miles de activos financieros.

### 3. SELECCIÓN DE CARTERAS

Tal y como mencionamos en la introducción, este trabajo utiliza la conocida teoría moderna de carteras –o MPT, por sus siglas en inglés– como marco matemático para construir una cartera de activos. También se lo conoce como análisis media-varianza, ya que el problema de inversión se formula como un problema de optimización donde el vector de  $N$  pesos de la cartera  $\mathbf{w}$  se escoge con el objetivo de maximizar el retorno esperado (media) para un nivel de riesgo dado (varianza). Usando notación matemática, el objetivo del inversor es minimizar la expresión:

$$\mathbf{w}'\Sigma\mathbf{w} = \gamma\boldsymbol{\mu}'\mathbf{w}$$

donde  $\boldsymbol{\mu}$  es el vector de retornos esperados,  $\Sigma$  es la matriz de varianzas y covarianzas de los retornos y  $\gamma \in [0, +\infty)$  representa la tolerancia al riesgo del inversor. Si  $\gamma = 0$  la tolerancia es cero y como resultado se obtiene la cartera de mínima varianza, mientras que a medida que  $\gamma$  tiende a infinito, se obtienen carteras con rentabilidad esperada y riesgo que tienden a infinito simultáneamente a lo largo de la frontera eficiente. Entre los dos extremos, se obtienen carteras que encuentran un compromiso entre retorno esperado y riesgo.

Para poder implementar la estrategia en la práctica, es necesario usar datos históricos de retornos financieros para estimar los parámetros  $\boldsymbol{\mu}$  y  $\Sigma$  que son generalmente desconocidos. También es fácil acomodar el caso en que la media o varianza son dinámicas, i.e.  $\boldsymbol{\mu}_h$  y  $\Sigma_h$ , donde  $h$  representa una cierta fecha de inversión. Repitiendo el algoritmo para las fechas  $h = 1, \dots, H$ , se obtiene la secuencia de pesos  $\{\mathbf{w}_h\}_{h=1}^H$ . Puesto que en frecuencias diarias no hay una gran predecibilidad en la media condicional  $\boldsymbol{\mu}_h$ , en este trabajo nos centramos en metodologías de predicción de la matriz de covarianzas condicional  $\Sigma_h$ , que además es un ejercicio más interesante desde el punto de vista estadístico. Una opción natural en este contexto es analizar la cartera de mínima varianza (caso  $\gamma = 0$ ). Otra opción es utilizar un proxy para el vector de retornos esperados como por ejemplo un índice de *momentum*. Los índices de *momentum* se emplean con frecuencia en estrategias de *trading* que consisten en comprar activos que han experimentado retornos muy positivos durante los últimos meses (generalmente se excluye el mes inmediatamente anterior), y vender aquellos que por el contrario han tenido un mal rendimiento en el mismo periodo.

En las dos subsecciones que siguen a continuación explicamos ambas estrategias en mayor detalle.

#### 3.1. Cartera de mínima varianza

Consideramos el problema de minimización de la varianza de una cartera con  $N$  activos sin restricciones a la venta en corto. El problema de inversión en la fecha  $h$  queda formulado como,

$$\min_{\mathbf{w}_h} \quad \mathbf{w}_h' \Sigma_h \mathbf{w}_h \quad [16]$$

$$\text{sujeto a} \quad \mathbf{w}_h' \mathbf{1} = 1, \quad [17]$$

donde  $\mathbf{1}$  denota el vector unitario de dimensión  $N$ . Este problema tiene la solución analítica:

$$\mathbf{w}_h = \frac{\Sigma_h^{-1} \mathbf{1}}{\mathbf{1}' \Sigma_h^{-1} \mathbf{1}}.$$

Naturalmente, la estrategia consiste en obtener el vector  $\mathbf{w}_h$  de pesos para cada fecha de actualización  $h$  reemplazando  $\Sigma_h$  por su predicción,  $\hat{\Sigma}_h$ . De esta manera, el vector de pesos factible viene dado por,

$$\hat{\mathbf{w}}_h = \frac{\hat{\Sigma}_h^{-1} \mathbf{1}}{\mathbf{1}' \hat{\Sigma}_h^{-1} \mathbf{1}}.$$

Estimar la cartera de mínima varianza presenta la ventaja evidente de ser un problema "limpio" en el sentido de que no es necesario realizar una estimación del vector de retornos esperados. Además, diversos autores han señalado que las carteras de mínima varianza no sólo tienen propiedades deseables a la hora de reducir el riesgo sino también en términos del binomio rentabilidad-riesgo, cuantificado mediante la ratio de Información.

### 3.2. Cartera de Markowitz con señal de momentum

En esta subsección adoptamos un enfoque más completo en tanto que consideramos un problema media-varianza (caso  $\gamma \neq 0$ ), en particular, la cartera con señal predictiva basada en *momentum*. Al igual que en la sub-sección anterior, la solución a este problema tiene una expresión analítica sencilla que depende de (la inversa de) la matriz de covarianzas dinámicas. Sin embargo, a diferencia de la cartera de mínima varianza, la solución también depende del vector de retornos esperados, que en este contexto se estima mediante el factor de momentum como se describe en Jegadeesh y Titman (1993). Concretamente, para cada fecha  $h$  y cada activo, el factor *momentum* se calcula como la media geométrica de los retornos del último año excluyendo el último mes (*i.e.*, los últimos 21 días). Los *momentums* de los  $N$  activos de nuestro universo se recogen en el vector  $\hat{\mathbf{m}}_h$ .

En ausencia de restricciones a la venta en corto, el problema de inversión en la fecha  $h$  se formula como:

$$\min_{\mathbf{w}_h} \quad \mathbf{w}_h' \Sigma_h \mathbf{w}_h \quad [18]$$

$$\text{sujeto a} \quad \mathbf{w}_h' \mathbf{m}_h = b_h, \quad [19]$$

$$\mathbf{w}_h' \mathbf{1} = 1, \quad [20]$$

donde  $b_h$  es el objetivo de retorno esperado, que calculamos mediante la media aritmética del *momentum* de los activos en el quintil superior – donde el orden se determina también por *momentum*. Como se ve claramente, la restricción [19] obliga a que la solución pase por dar un peso suficientemente elevado a los activos con mayor *momentum* para alcanzar una rentabilidad objetivo  $b_h$ . Escribiendo el Lagrangiano y resolviendo un sistema de ecuaciones lineal 2 x 2, vemos que la solución a este problema viene dada por:

$$\mathbf{w}_h = c_1 \cdot \Sigma_h^{-1} \mathbf{m}_h + c_2 \cdot \Sigma_h^{-1} \mathbf{1},$$

donde,

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \frac{1}{AC - B^2} \begin{bmatrix} A \cdot b_h - B \\ C - B \cdot b_h \end{bmatrix}$$

$$A = \mathbf{1}' \Sigma_h^{-1} \mathbf{1}$$

$$B = \mathbf{1}' \Sigma_h^{-1} \mathbf{m}_h, \text{ y}$$

$$C = \mathbf{m}_h' \Sigma_h^{-1} \mathbf{m}_h.$$

Nuevamente, la estrategia consiste en reemplazar  $\Sigma_h$  por  $\hat{\Sigma}_h$  y  $\mathbf{m}_h$  por  $\hat{\mathbf{m}}_h$ , con lo que se obtiene el vector de pesos (factible):

$$\hat{\mathbf{w}}_h = \hat{c}_1 \cdot \hat{\Sigma}_h^{-1} \hat{\mathbf{m}}_h + \hat{c}_2 \cdot \hat{\Sigma}_h^{-1} \mathbf{1}.$$

#### 4. APLICACIÓN EMPÍRICA

En esta aplicación realizamos un ejercicio de selección de carteras de inversión. Se compara el rendimiento de algunas de las metodologías mencionadas anteriormente de manera similar a Engle, Ledoit y Wolf (2019) y De Nard, Ledoit y Wolf (2020). Los datos utilizados son los precios de cierre diarios ajustados por dividendos y *splits*,  $\{P_{i,t}\}$ , de las acciones que constituyen el índice S&P 500 en el período que va de 04-01-2000 a 30-06-2019 (fuente: AlphaVantage). En cada fecha de inversión  $h$ , seleccionamos los

100 activos con mayor capitalización bursátil (por lo tanto,  $N=100$ ). Construimos los retornos para cada activo  $i$  como  $r_{i,t} = 100 \cdot \log(P_{i,t}/P_{i,t-1})$ .

Para la familia de modelos DCC que incorporan estructura de factores utilizamos los datos de retornos diarios del factor “Prima de Riesgo de Mercado” de Fama y French (2015) de la página web de Ken French durante el mismo período<sup>2</sup>. Por lo tanto, el número de factores  $K=1$ . Este factor se calcula como el retorno ponderado por valor de todas las empresas de la base de datos CRSP registradas en los Estados Unidos que cotizan en los mercados NYSE, AMEX y NASDAQ, neto del tipo de interés a un mes del *US Treasury bill*.

Para esta aplicación, adoptamos la convención de que un “mes” equivale a 21 días consecutivos de comercialización. El período de validación de las diferentes estrategias va de 07-01-2005 a 30-06-2019, lo que representa un total de 173 meses (es decir,  $h = 1, \dots, 173$ ). La cartera se actualiza a principios de cada mes y se mantiene constante durante todo el mes. En cada fecha de actualización  $h$ , realizamos una predicción de la matriz de covarianzas  $\hat{\Sigma}_h$  como se explica en la subsección 2.3. basada en los datos de los últimos cinco años hasta la fecha correspondiente –lo que equivale a un total de 1260 observaciones–. Los pesos de la cartera  $\mathbf{w}_h$  son una función de (la inversa de)  $\hat{\Sigma}_h$ . En particular, consideramos dos estrategias conocidas: la cartera de mínima varianza (3.1.) y la cartera de Markowitz con señal de *momentum* (3.2.). Consecuentemente, el retorno de la cartera  $c$  en el día  $t$  del mes  $h$  se calcula como,

$$r_{c,t} = \mathbf{w}_h' \mathbf{r}_t.$$

Para predecir dicha matriz de covarianzas, los modelos de la familia DCC “estándar” que consideramos son los siguientes: DCC-SC, que utiliza la covarianza muestral de los residuos estandarizados para estimar la matriz  $C$  en la ecuación [3], DCC-LS, que en lugar de la covarianza muestral emplea el estimador de regularización lineal (2.5.1.), y DCC-NLS, que emplea regularización no lineal (2.5.2.). De manera similar, denotamos los modelos DCC que incorporan estructura de factores con las siglas AFM-DCC-SC, AFM-DCC-LS y AFM-DCC-NLS.

Para ampliar el rango de la comparativa, consideramos la versión escalar del modelo VECH (sVECH) con focalización en covarianzas. La matriz de covarianzas dinámica para este modelo se define como:

$$\Sigma_t = (1 - \gamma_1 - \gamma_2)\Sigma + \gamma_1 \mathbf{r}_{t-1} \mathbf{r}' + \gamma_2 \Sigma_{-1}, \quad [21]$$

donde aquí la igualdad  $\Sigma = \mathbb{E}[\mathbf{r}_t \mathbf{r}'_t]$  es exacta siempre y cuando  $\gamma_1 + \gamma_2 < 1$ . Similarmente, utilizamos los acrónimos sVECH-SC, sVECH-LS y sVECH-NLS para identificar la versión

<sup>2</sup> [http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html)

del modelo que utiliza la covarianza muestral, el estimador con regularización lineal y el estimador con regularización no lineal, respectivamente. Todos los modelos introducidos en esta sección se estiman mediante verosimilitud compuesta (2.2.2).

En las tablas 1 y 3 reportamos las siguientes medidas de desempeño para cada cartera en consideración. El primer grupo de métricas nos proporcionan el resumen numérico de los retornos de la cartera, y son 1) AV: Retorno medio de la cartera en el período de validación. Añalizamos multiplicando el resultado por 252. 2) SD: Desviación estándar (o volatilidad) de la cartera en el período de validación. Añalizamos multiplicando el resultado por  $\sqrt{252}$ . Para evitar ambigüedad, la calculamos como sigue:

$$SD = \sqrt{252} \times \sqrt{\frac{1}{173 \cdot 21} \sum_{t=1}^{173 \cdot 21} r_{c,t}^2},$$

y 3) IR: Ratio de información, *i.e.* AV / SD.

El segundo grupo de métricas nos informa sobre la evolución y distribución de los pesos a lo largo del ejercicio empírico. En particular, estas son 4) TO: Rotación (turnover) mensual de la cartera que calculamos como  $\frac{1}{172} \sum_{h=1}^{172} \|\hat{\mathbf{w}}_{h+1} - \hat{\mathbf{w}}_h^*\|_1$ , donde  $\|\cdot\|_1$  es la norma  $L^1$  y utilizamos  $\mathbf{w}_h^*$  para denotar el vector de pesos a final del mes  $h$  - que puede diferir de  $\hat{\mathbf{w}}_h$  pues el primero tiene en cuenta la evolución de precios de cada activo durante el mes. Este índice cuantifica si los pesos de la cartera experimentan cambios grandes a lo largo del ejercicio, lo que en la práctica se traduce en mayores costes de transacción; 5) PL: Proporción de apalancamiento (*leverage*), calculada como  $\frac{1}{173 \times N} \sum_{h=1}^{173} \sum_{i=1}^N \mathbf{1}_{\{\hat{w}_{i,h} < 0\}}$ . Esta medida cuantifica la proporción activos que hemos vendido en corto a lo largo del ejercicio, 6) Max: Máximo peso de la cartera  $\max_{i,t} \hat{w}_{i,t}$ , 7) Min: Mínimo peso de la cartera  $\min_{i,t} \hat{w}_{i,t}$ , y 8) HI: Índice de concentración Herfindahl corrigiendo por posibles valores negativos. Un índice elevado se traduce en una cartera donde la riqueza está invertida en un número relativamente pequeño de activos.

Es legítimo preguntarse entonces por qué no considerar datos mensuales. En primer lugar, con frecuencia mensual no disponemos de suficientes observaciones para estimar modelos GARCH multivariantes con suficientes garantías. Por otro lado, usar datos diarios parece ofrecer mejores resultados incluso cuando la frecuencia de inversión es mensual (véase, por ejemplo, Ledoit y Wolf, 2017).

Para evaluar los resultados del ejercicio de mínima varianza (subsección 3.1.), tomamos como referencia la cartera equiponderada también conocida en la literatura como 1/N (De Miguel, Garlappi y Uppal, 2009). De acuerdo con este trabajo, en carteras con un número moderado de activos (entre 3 y 50) y una ventana de estimación de 120 meses, es difícil batir la cartera equiponderada con carteras Markowitz en términos de ratio de

información. De las medidas de desempeño presentadas al inicio de esta sección, la más importante en este contexto es sin duda SD, puesto que esta cartera está especialmente diseñada con el objetivo de minimizar esta métrica. Naturalmente, también son de interés el retorno medio de la cartera y su ratio de información, pero en este contexto son de importancia secundaria.

TABLA 1.

**CARTERA DE MÍNIMA VARIANZA:  $N=100$   
PERÍODO DE VALIDACIÓN DE 07-01-2005 HASTA 30-06-2019**

Modelo	AV	SD	IR	TO	PL	Max	Min	HI
DCC-LS	12.060	11.848	1.018	0.154	0.460	0.372	-0.159	0.013
DCC-NLS	12.185	11.732	1.039	0.109	0.463	0.373	-0.130	0.013
DCC-SC	12.057	11.874	1.015	0.171	0.461	0.373	-0.164	0.013
AFM-DCC-LS	14.145	11.649	1.214	0.139	0.425	0.427	-0.187	0.012
AFM-DCC-NLS	14.620	11.648	1.255	0.686	0.425	0.448	-0.192	0.012
AFM-DCC-SC	14.094	11.803	1.194	0.566	0.427	0.462	-0.209	0.012
sVECH-LS	12.966	11.881	1.091	0.198	0.443	0.353	-0.242	0.012
sVECH-NLS	12.955	11.834	1.095	0.151	0.441	0.353	-0.237	0.012
sVECH-SC	12.961	11.934	1.086	0.871	0.446	0.353	-0.245	0.012
1/N	10.757	19.645	0.548	0.056	0.000	0.010	0.010	0.010

*Nota:* Medidas de desempeño para varios estimadores de la cartera de mínima varianza. Las columnas AV y SD se reportan en términos anualizados (en porcentaje), es decir, multiplicando por  $\sqrt{252}$ . AV denota el retorno medio; SD es su desviación estándar; IR es la ratio de información; TO mide la rotación de la cartera; PL es la proporción de apalancamiento; Max y Min son el peso máximo y mínimo, respectivamente; y HI es el índice Herfindahl corrigiendo por valores negativos.

*Fuente:* Elaboración propia.

La significación estadística de los resultados se basa en el test de la diferencia de varianzas propuesto por Ledoit y Wolf (2011), que es consistente en presencia de heterocedasticidad y autocorrelación. En la tabla 2 se reportan los resultados del citado test. Las diferencias se dan en términos anualizados para hacerlos equiparables a los de la tabla 1, pero para el cálculo de significación se han utilizado los retornos diarios de la cartera  $\{r_{c,t}\}$  en el período de validación sin anualizar.

En este ejercicio, la cartera que obtiene una menor volatilidad es AFM-DCC-NLS, lo cual pone en valor el uso de modelos de factores y la estimación por regularización no lineal. Por un lado, la regularización aporta un valor diferencial que es significativo tanto desde el punto de vista económico como estadístico para modelos DCC con y sin estructura de factores. La regularización no lineal ofrece mejores resultados que la regularización lineal, pero esta diferencia desaparece cuando se incorpora la estructura de factores en la ecuación. Por otro lado, las carteras basadas en modelos de factores parecen aportar un valor diferencial que es significativo desde el punto de vista económico pero no desde el punto de vista estadístico.

TABLA 2.

TEST DE DIFERENCIA DE VARIANZAS (LEDOIT Y WOLF, 2011) PARA VARIOS ESTIMADORES DE LA CARTERA DE MÍNIMA VARIANZA (N=100)

Modelo	DCC-LS	DCC-NLS	DCC-SC	AFM-DCC-LS	AFM-DCC-NLS	AFM-DCC-SC	sVECH-LS	sVECH-NLS	sVECH-SC	1/N
DCC-LS	-	-0.12***	0.03***	-0.20	-0.20	-0.05	0.03	-0.01	0.09	7.80***
DCC-NLS	0.12***	-	0.14***	-0.08	-0.08	0.07	0.15	0.10	0.20	7.91***
DCC-SC	-0.03***	-0.14***	-	-0.23*	-0.23*	-0.07	0.01	-0.04	0.06	7.77***
AFM-DCC-LS	0.20	0.08	0.23*	-	-0.00	0.15***	0.23	0.18	0.28	8.00***
AFM-DCC-NLS	0.20	0.08	0.23*	0.00	-	0.15***	0.23	0.19	0.29	8.00***
AFM-DCC-SC	0.05	-0.07	0.07	-0.15***	-0.15***	-	0.08	0.03	0.13	7.84***
sVECH-LS	-0.03	-0.15	-0.01	-0.23	-0.23	-0.08	-	-0.05***	0.05***	7.76***
sVECH-NLS	0.01	-0.10	0.04	-0.18	-0.19	-0.03	0.05***	-	0.10***	7.81***
sVECH-SC	-0.09	-0.20	-0.06	-0.28	-0.29	-0.13	-0.05***	-0.10***	-	7.71***
1/N	-7.80***	-7.91***	-7.77***	-8.00***	-8.00***	-7.84***	-7.76***	-7.81***	-7.71***	-

Nota: En cada celda se calcula la diferencia de la desviación estándar de la cartera de la columna menos la desviación estándar de la cartera de la fila. Las cifras se reportan en términos anualizados (es decir, multiplicando por  $\sqrt{252}$ ). Los símbolos \*, \*\*, y \*\*\* indican que los resultados son significativos al nivel 10, 5 y 1%, respectivamente.

Fuente: Elaboración propia.

Si comparamos la familia de modelos DCC con la alternativa del VECH escalar, vemos que los modelos DCC ofrecen resultados significativamente mejores desde el punto de vista económico pero no estadístico. Por un lado, que los modelos DCC tengan un mejor desempeño tiene sentido porque permiten capturar la potencial heterogeneidad en la dinámica de las volatilidades de cada activo – algo que no permite el modelo VECH escalar. Sin embargo, el error que resulta de estimar un número de parámetros que crece linealmente en  $N$  podría eclipsar los beneficios de modelizar esa heterogeneidad. Por lo tanto, no sería adecuado concluir que hay que descartar los modelos VECH.

Finalmente, los resultados confirman que, independientemente de qué especificación utilicemos, la familia de modelos GARCH multivariante ofrece unos resultados que son significativamente mucho mejores que utilizar una cartera equiponderada.

TABLA 3.

**CARTERA DE MARKOWITZ CON SEÑAL DE MOMENTUM:  $N=100$   
PERÍODO DE VALIDACIÓN DE 07-01-2005 HASTA 30-06-2019**

Modelo	AV	SD	IR	TO	PL	Max	Min	HI
DCC-LS	14.339	13.556	1.058	0.373	0.466	0.445	-0.240	0.012
DCC-NLS	14.801	13.458	1.100	0.161	0.467	0.412	-0.204	0.013
DCC-SC	14.311	13.579	1.054	0.315	0.467	0.457	-0.245	0.012
AFM-DCC-LS	15.579	13.386	1.164	0.195	0.435	0.380	-0.301	0.012
AFM-DCC-NLS	16.207	13.370	1.212	2.119	0.436	0.379	-0.305	0.012
AFM-DCC-SC	15.168	13.512	1.123	0.523	0.440	0.397	-0.317	0.012
sVECH-LS	14.205	13.646	1.041	0.194	0.451	0.340	-0.277	0.012
sVECH-NLS	14.321	13.600	1.053	0.182	0.449	0.340	-0.237	0.012
sVECH-SC	14.133	13.698	1.032	0.213	0.452	0.340	-0.245	0.012
EW-TQ	10.697	21.075	0.508	0.072	0.000	0.050	0.010	0.010

*Nota:* Medidas de desempeño para varios estimadores de la cartera de Markowitz con señal de momentum. Las columnas AV y SD se reportan en términos anualizados (en porcentaje), es decir, multiplicando por  $\sqrt{252}$ . AV denota el retorno medio; SD es su desviación estándar; IR es la ratio de información; TO mide la rotación de la cartera; PL es la proporción de apalancamiento; Max y Min son el peso máximo y mínimo, respectivamente; y HI es el índice Herfindahl corrigiendo por valores negativos.

*Fuente:* Elaboración propia.

Para el ejercicio de cartera Markowitz con señal de *momentum* (subsección 3.2), consideramos los mismos estimadores que para el ejercicio de mínima varianza, y como referencia esta vez utilizamos la cartera equiponderada de las acciones en el quintil superior por *momentum* (EW-TQ). En este caso, tiene mucho más sentido evaluar el desempeño de las diferentes estrategias primordialmente en base a la ratio de información IR, que cuantifica la relación entre rentabilidad y riesgo. En un contexto “ideal”, minimizar la varianza para un objetivo fijo de rentabilidad  $b_n$ , debería ser equivalente a maximizar la

TABLA 4.

TEST DE DIFERENCIA DE RATIO DE INFORMACIÓN (LEDOIT Y WOLF, 2011) PARA VARIOS ESTIMADORES DE LA CARTERA DE MARKOWITZ CON SENAL DE MOMENTUM (N=100)

Modelo	DCC-LS	DCC-NLS	DCC-SC	AFM-DCC-LS	AFM-DCC-NLS	AFM-DCC-SC	sVECH-LS	sVECH-NLS	sVECH-SC	EW-TQ
DCC-LS	-	-0.04**	0.00	-0.11	-0.15	-0.06	0.02	0.00	0.03	0.55**
DCC-NLS	0.04**	-	0.05**	-0.06	-0.11	-0.02	0.06	0.05	0.07	0.59**
DCC-SC	-0.00	-0.05**	-	-0.11	-0.16	-0.07	0.01	0.00	0.02	0.55**
AFM-DCC-LS	0.11	0.06	0.11	-	-0.05***	0.04*	0.12	0.11	0.13	0.66***
AFM-DCC-NLS	0.15	0.11	0.16	0.05***	-	0.09***	0.17	0.16	0.18	0.70***
AFM-DCC-SC	0.06	0.02	0.07	-0.04*	-0.09***	-	0.08	0.07	0.09	0.61***
sVECH-LS	-0.02	-0.06	-0.01	-0.12	-0.17	-0.08	-	-0.01*	0.01*	0.53**
sVECH-NLS	-0.00	-0.05	-0.00	-0.11	-0.16	-0.07	0.01*	-	0.02*	0.55**
sVECH-SC	-0.03	-0.07	-0.02	-0.13	-0.18	-0.09	-0.01*	-0.02*	-	0.52**
EW-TQ	-0.55**	-0.59**	-0.55**	-0.66***	-0.70***	-0.61***	-0.53**	-0.55**	-0.52**	-

Nota: En cada celda se calcula la diferencia de la ratio de información de la cartera de la fila menos la ratio de información de la cartera de la columna. Las cifras se multiplican por  $\sqrt{252}$  para facilitar la visualización. Los símbolos \*, \*\*, y \*\*\* indican que los resultados son significativos al nivel 10, 5 y 1%, respectivamente.

Fuente: Elaboración propia.

ratio de información IR, pero en la práctica esto no es así debido al error de estimación en el vector  $m_n$ . Por lo tanto, enfocarse en la métrica SD no es tan apropiado en este contexto.

De manera análoga al ejercicio anterior, la significación estadística de los resultados se basa en el test de la diferencia de la ratio de información (Ledoit y Wolf, 2008), que es consistente en presencia de heterocedasticidad y autocorrelación. En la tabla 4 se recogen los resultados del citado test. Con el objetivo de facilitar la lectura de los resultados a través de las tablas 3 y 4, se reporta la diferencia en la ratio de información anualizado (es decir, multiplicando por  $\sqrt{252}$  porque en  $IR = AV / SD$  el numerador se anualiza multiplicando por 252 y el denominador se anualiza multiplicando por  $\sqrt{252}$ ). Sin embargo, para el test se han utilizado los retornos diarios de la cartera  $\{r_{c,t}\}$  en el período de validación sin anualizar.

En términos generales, la cartera de Markowitz con señal de *momentum* ofrece una rentabilidad superior a la cartera de mínima varianza a expensas de una mayor volatilidad –como cabía esperar–. En términos de la ratio de información, ambas carteras ofrecen resultados comparables.

El mejor modelo por ratio de información es AFM-DCC-NLS. Como en el caso de la cartera de mínima varianza, estos resultados ponen en valor el uso de modelos de factores aunque los resultados no son estadísticamente significativos. Por otro lado, la regularización proporciona resultados significativamente mejores (económica y estadísticamente), y en este caso la regularización no lineal es significativamente superior a la lineal. El modelo VECH escalar ofrece resultados comparables desde el punto de vista estadístico, y ligeramente inferiores desde el punto de vista económico. El mismo comentario sobre la heterogeneidad que se hizo en la anterior subsección es de aplicación aquí. Finalmente, concluimos que, con independencia de la estrategia que utilizemos, todos los modelos GARCH multivariantes aquí considerados ofrecen resultados significativamente mucho mejores que los de una cartera que da el mismo peso a los activos que se encuentran en el quintil superior por *momentum*.

## 5. CONCLUSIONES

En este capítulo se han cubierto algunos de los últimos desarrollos en la literatura de modelización de covarianzas dinámicas de elevada dimensionalidad. Se ha hecho especial énfasis en el uso de técnicas computacionalmente escalables (verosimilitud compuesta, focalización en covarianzas) y en cómo introducir la estructura de factores de una manera sencilla. Hemos presentado los métodos de regularización de los valores propios de la matriz de covarianzas muestral y cómo pueden aplicarse de manera natural en el modelo DCC a través del método de focalización en covarianzas. Esto es de especial relevancia en contextos de elevada dimensionalidad, es decir, cuando la sección cruzada es de un tamaño comparable (o incluso superior) al número de observaciones.

Finalmente, a través de un ejercicio empírico con datos de acciones del índice S&P 500 se ha demostrado cómo la implementación de estos métodos permite construir carteras de inversión con mejores rentabilidades ajustadas por riesgo.

Ni que decir tiene que éste no es el primer trabajo en aplicar estimadores de covarianzas dinámicas de elevada dimensionalidad al problema de selección de carteras— y muy probablemente no será el último. Los resultados de este trabajo siguen la línea de los hallados por Engle, Ledoit y Wolf (2019) y De Nard, Ledoit y Wolf (2020), concluyendo que los modelos DCC factoriales conducen a carteras con menor volatilidad.

Cabe destacar que en la práctica los inversores deben tener en cuenta los costes de transacción para decidir qué estrategia seguir, y por lo tanto hay que encontrar el balance óptimo entre estrategias que aporten buenas rentabilidades ajustadas al riesgo y carteras que aporten un poco menos pero conlleven una menor rotación y con ello menores costes de transacción (véase, entre otros, Hautsch y Voigt, 2019)). En el reciente trabajo de Moura, Santos y Ruiz (2020) se hace una comparativa más amplia en la que también se incluyen modelos de volatilidad estocástica multivariante que resultan en carteras con mayor ratio de información después de costes de transacción. Los resultados de este trabajo también aportan evidencia de que los modelos DCC son los que mejor minimizan la volatilidad de la cartera. Sin embargo, según este trabajo, modelizar la matriz de covarianzas como un proceso Wishart con regularización hacia una matriz diagonal conduce a carteras más estables con menor rotación y mayor ratio de información una vez tenidos en cuenta los costes de transacción.

## Referencias

- AGUILAR, M. (2009). A latent factor model of multivariate conditional heteroscedasticity. *Journal of Financial Econometrics*, 7(4), pp. 481–503.
- AIELLI, G. P. (2013). Dynamic conditional correlation: On properties and estimation. *Journal of Business and Economic Statistics*, 31.
- ALESSI, L., BARIGOZZI, M. y CAPASSO, M. (2009). Estimation and forecasting in large datasets with conditionally heteroskedastic dynamic common factors. *ECB Working Paper*, 1115.
- ALEXANDER, C. y CHIBUMBA, A. (1996). Multivariate orthogonal factor garch. University of Sussex *Discussion Papers in Mathematics*.
- ASAI, M., McALEER, M. y YU, J. (2006). Multivariate stochastic volatility: A review. *Econometric Reviews*, 25(2-3), pp. 145–175.
- BAILLIE, R. y BOLLERSLEV, T. (1992). Prediction in dynamic models with time-dependent conditional variances. *Journal of Econometrics*, 52(1-2), pp. 91–113.

- BAMPINAS, G. y PANAGIOTIDIS, T. (2015). On the relationship between oil and gold before and after financial crisis: Linear, nonlinear and time-varying causality testing. *Working Paper series from Rimini Centre for Economic Analysis*.
- BAUWENS, L., LAURENT, S. y ROMBOUTS, J. (2006). Multivariate garch models: a survey. *Journal of Applied Econometrics*, 21, pp. 79–109.
- BICKEL, P. J. y LEVINA, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics*, 36, pp. 2577–2604.
- BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, pp. 307–327.
- BOLLERSLEV, T., ENGLE, R. F. y NELSON, D. B. (1994). ARCH models. En: R. F. ENGLE y D. McFADDEN, editors, *Handbook of Econometrics*, pages 2959–3038. Elsevier.
- BOLLERSLEV, T., ENGLE, R. y WOOLDRIDGE, J. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1), pp. 116–31.
- BOLLERSLEV, T. y WOOLDRIDGE, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 11(2), pp. 143–172.
- BROTO, C. y RUIZ, E. (2004). Estimation methods for stochastic volatility models: a survey. *Journal of Economic Surveys*, 18(5), pp. 613–649.
- BROWNLEES, C. y ENGLE, R. (2017). SRISK: A Conditional Capital Shortfall Measure of Systemic Risk. *Review of Financial Studies*, 30(1), pp. 48–79.
- BROWNLEES, C. y LLORENS, J. (2020). Projected dynamic conditional correlations. *Disponible en SSRN*.
- CONRAD, C., KARANASOS, M. y ZENG, N. (2010). The link between macroeconomic performance and variability in the uk. *Economics Letters*, 106(3), pp. 154–157.
- DEMIGUEL, V., GARLAPPI, L. y UPPAL, R. (2009). Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *The Review of Financial Studies*, 22(5), pp. 1915–1953.
- DE ALMEIDA, D., HOTTA, L. y RUIZ, E. (2018). Mgarch models: Trade-off between feasibility and flexibility. *International Journal of Forecasting*, 34(1), pp. 45–63.
- DE NARD, G., LEDOIT, O. y WOLF, M. (2020). Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *Journal of Financial Econometrics (forthcoming)*.
- DIEBOLD, F. y NERLOVE, M. (1986). The dynamics of exchange rate volatility: a multivariate latent factor arch model. *Board of Governors of the Federal Reserve System (U.S.)*, 205.
- DING, Z. y ENGLE, R. (2001). Large scale conditional covariance matrix modeling, estimation and testing. *Academia Economic Papers*.

ENGLE, R. (2002). Dynamic conditional correlation. *Journal of Business & Economic Statistics*, 20(3), pp. 339–350.

— (2009). *Anticipating Correlations: A New Paradigm for Risk Management*. Princeton University Press.

ENGLE, R. y COLACITO, R. (2006). Testing and valuing dynamic correlations for asset allocation. *Journal of Business & Economic Statistics*, 24(2), pp. 238–253.

ENGLE, R. y SHEPHARD, K. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate garch. *Working Paper National Bureau of Economic Research*, 8554.

ENGLE, R. F., LEDOIT, O. y WOLF, M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37(2).

FAN, J., LIAO, Y. y MINCHEVA, M. (2013). Large Covariance Estimation by Thresholding Principal Orthogonal Complements. *Journal of the Royal Statistical Society, Series B*, 75, pp. 603–680.

FERREIRA, M. A. (2005). Evaluating interest rate covariance models within a value-at-risk framework. *Journal of Financial Econometrics*, 3(1), pp. 126–168.

FORNI, M., HALLIN, M., LIPPI, M. y REICHLIN, L. (2000). The Generalized Dynamic-Factor Model: Identification and Estimation. *The Review of Economics and Statistics*, 82, pp. 540–554.

FORNI, M., HALLIN, M., LIPPI, M. y ZAFFARONI, P. (2015). Dynamic factor models with infinite-dimensional factor spaces: One-sided representations. *Journal of Econometrics*, 185(2), pp. 359–371.

FRIEDMAN, J., HASTIE, T. y TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3), pp. 432–441.

HAFNER, C. M. y HERWARTZ, H. (2008). Testing for causality in variance using multivariate garch models. *Annales d'Economie et de Statistique*, 89, pp. 215–241.

HAFNER, C. M. y PREMINGER, A. (2009). Asymptotic theory for a factor garch model. *Econometric Theory*, 25(2), pp. 336–363.

HALLIN, M., HOTTA, L. K., MAZZEU, J. H. G., TRUCIOS-MAZA, C. C., PEREIRA, P. L. V. y ZEVALLOS, M. (2019). Forecasting conditional covariance matrices in high-dimensional time series: a general dynamic factor approach. *Working Papers ECARES*, 2019-14.

HAN, Y. (2006). Asset allocation with a high dimensional latent factor stochastic volatility model. *The Review of Financial Studies*, 19(1), pp. 237–271.

HARVEY, A., RUIZ, E. y SENTANA, E. (1992). Unobserved component time series models with Arch disturbances. *Journal of Econometrics*, 52(1-2), pp. 129–157.

HASTIE, T., TIBSHIRANI, R. y FRIEDMAN, J. (2008). *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY: Springer New York.

HAUGEN, R. A. y BAKER, N. L. (1991). The efficient market inefficiency of capitalization-weighted stock portfolios. *The Journal of Portfolio Management*, 17(3), pp. 35–40.

HAUTSCH, N., KYJ, L. M. y MALEC, P. (2015). Do high-frequency data improve high-dimensional portfolio allocations? *Journal of Applied Econometrics*, 30(2), pp. 263–290.

HAUTSCH, N. y VOIGT, S. (2019). Large-scale portfolio allocation under transaction costs and model uncertainty. *Journal of Econometrics*, 212(1), pp. 221–240. Big Data in Dynamic Predictive Econometric Modeling.

JEGADEESH, N. y TITMAN, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1), pp. 65–91.

LEDOIT, O. y WOLF, M. (2004a). Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4), pp. 110–119.

— (2004b). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2), pp. 365 – 411.

— (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, 15(5), pp. 850–859.

— (2011). Robust performances hypothesis testing with the variance. *Wilmott*, 2011(55), pp. 86–89.

— (2017). Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks. *The Review of Financial Studies*, 30(12), pp. 4349–4388.

— (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *Annals of Statistics (forthcoming)*.

MARKOWITZ, H. (1952). Portfolio selection. *Journal of Finance*, 7(1), pp. 77–91.

MOURA, G. V., SANTOS, A. A. y RUIZ, E. (2020). Comparing high-dimensional conditional covariance matrices: Implications for portfolio selection. *Journal of Banking & Finance*, 118, 105882.

PAKEL, C., ENGLE, R. F., SHEPHARD, N. y SHEPPARD, K. (2017). Fitting vast dimensional time-varying covariance models. *NYU Working Paper*, No. FIN-08-009.

PATTON, A. J. y SHEPPARD, K. (2009). *Evaluating Volatility and Correlation Forecasts*, pp. 801–838. Berlin: Springer Berlin Heidelberg, Berlin.

PEÑA, D. y PONCELA, P. (2004). Forecasting with nonstationary dynamic factor models. *Journal of Econometrics*, 119(2), pp. 291–321.

PESARAN, B. y PESARAN, H. (2007). Modelling volatilities and conditional correlations in futures markets with a multivariate t distribution. *Technical report, IEPR Working Paper*, No. 07.19.

- POURAHMADI, M. (2013). *High-Dimensional Covariance Estimation*. John Wiley & Sons, Inc.
- ROSS, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3), pp. 341 – 360.
- SHEPHARD, N. (1996). *Statistical aspects of ARCH and stochastic volatility*, pages 1–67. Chapman & Hall, London, (edited by d.r. cox, david v. hinkley and ole e. barndorff-neilsen) edition. Reprinted in the Survey of Applied and Industrial Mathematics, issue on Financial and insurance mathematics, 3, 764-826, Scientific Publisher TVP, Moscow, 1996 (in Russian).
- SILVENNOINEN, A. y TERÄSVIRTA, T. (2009). *Multivariate GARCH Models*, pp. 201–229. Springer Berlin Heidelberg, Berlin, Heidelberg.
- STEIN, C. (1986). Lectures on the theory of estimation of many parameters. *Journal of Soviet Mathematics*, 34(1), p. 1373.
- STOCK, J. H. y WATSON, M. W. (2002a). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, 97, pp. 1167–1179.
- (2002b). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics*, 220, pp. 147–162.
- TSE, L. D. Y. K. y TSUI, A. K. C. (2002). Evaluating the hedging performance of the constant-correlation garch model. *Applied Financial Economics*, 12, pp. 791–798.
- WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.