UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS MATEMÁTICAS
FACULTAD DE CIENCIAS

Regular Variation, Extremes and Autocovariance Function Convergence for Multivariate GARCH processes.

T E S I S

QUE PARA OBTENER EL GRADO ACADÉMICO DE
DOCTOR EN CIENCIAS MATEMÁTICAS

P R E S E N T A :
NELSON OMAR MURIEL TORRERO

DIRECTOR DE TESIS:
DRA. BEGOÑA FERNÁNDEZ FERNANDEZ

MÉXICO, D.F. JULIO, 2009

Agradecimientos

Agradecemos a Ana Meda Guardiola, Federico O'Reilly, Graciela González Farías, Víctor Pérez Abreu por la revisión final del trabajo que ayudó a conformar una presentación más consistente y autocontenida. A Ana, Graciela y Víctor agradecemos también por sus valiosos comentarios que contribuyeron en el desarrollo de los resultados presentados en este trabajo. Nuestro agradecimiento a Paul Embrechts quien nos ayudó a comprender la relevancia de los problemas resueltos y nos impulsa a enfrentar nuevos retos en la Teoría de Valores Extremos para series de tiempo no–lineales.

A los árbitros del artículo "Regular Variation and related results for the multivariate GARCH(p,q) process with constant conditional correlations" les agradecemos muy sinceramente por sus comentarios adecuados y precisos. Los cambios sugeridos y las correcciones mejoraron sustancialmente la presentación del contenido del Capítulo 2.

Agradecemos también a Claudia Klüppelberg y Robert Stelzer por su amable invitación y su hospitalidad durante mi estancia en TUM, Insituto de Estudios Avanzados en München, donde los resultados del Capítulo 3 fueron concluídos. A Begoña Fernández y a Gerardo Rubio agradecemos todas las interesantes discusiones que ayudaron a conformar mejor la presentación del contenido de este Capítulo.

Por último, agradecemos a CONACYT por su apoyo durante todo el programa de Doctorado bajo el número de becario 180346 y al proyecto PAPIIT–DGAPA–UNAM 103606.

Contents

Introducción

La investigación sobre las propiedades probabilistas y estadísticas de procesos estocásticos con varianza y covarianzas dependientes del tiempo ha tenido un gran auge en los últimos años. Uno de los mayores impulsos para esta actividad ha sido la creciente evidencia de que gran parte de la teoría financiera actual se basa en el estudio de la volatilidad. De este reconocimiento surge la necesidad de desarrollar modelos empíricamente razonables para probar, aplicar y así profundizar este conocimiento teórico.

La manera más usual de definir un proceso $(Y_t)_{t\in T}$ con volatililidad aleatoria V_t es mediante

$$Y_t = \begin{cases} \int_0^t \sqrt{V_{s-}} dL_s, & \text{si } T = [0, \infty), \\ \sqrt{V_t} Z_t, & \text{si } T = \mathbb{N}. \end{cases} \tag{1}$$

En el caso de tiempo continuo, el proceso (L_t) es un proceso de Lévy y si el tiempo es discreto, la sucesión $\{Z_t\}$ es i.i.d. con media 0 y varianza unitaria. Pasar del modelo de volatilidad aleatoria (1) a un modelo de volatilidad estocástica, implica modelar el proceso (V_t).

En tiempo continuo, una posibilidad es considerar a (V_t) como un proceso de Ornstein–Uhlenbeck generalizado, es decir, como solución de la ecuación estocástica

$$dV_t = -\lambda V_t dt + dU(\lambda t)$$

donde el proceso U es un proceso de Lévy espectralmente positivo, con deriva nula. La elección inusual del parámetro temporal λt en la ecuación es una manera de garantizar que la distribución marginal de V_t no depende del parámetro λ. Puede encontrarse una explicación detallada del uso de este tipo de procesos en finanzas en Barndorff-Nielsen and Shephard (2001).

Como generalización de este planteamiento se propone que (V_t) tenga la forma

$$V_t = e^{-\xi_t}\left(V_0 + \int_0^t e^{\xi_{s-}} d\eta_s\right) \tag{2}$$

donde (ξ_t, η_t) es un proceso de Lévy bivariado independiente del valor inicial V_0 y η_t es un subordinador. Lindner and Maller (2005) contribuyeron de forma esencial en el conocimiento de las propiedades probabilistas de este tipo de procesos.

Los modelos en tiempo continuo son muy útiles para los desarrollos teóricos y han contribuído enormemente al entendimiento de la dinámica financiera y económica.

En tiempo discreto también hay muy diversas maneras de modelar el proceso $\{V_t\}$. Una de las formas más conocidas es el proceso ARCH (Autoregressive conditional heteroskedasticity), introducido en Engle (1982) que especifica

$$V_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i V_{t-i}.$$

Destaca el paralelismo de esta especificación con los modelos autoregresivos tan ampliamente utilizados para las variaciones de la media. Del mismo modo que los modelos autoregresivos pueden extenderse a modelos (autoregresivos y) de promedios móviles, el modelo ARCH puede generalizarse especificando

$$V_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i V_{t-i} + \sum_{j=1}^{q} Y_{t-j}^2$$

donde Y_t está dado en (1). Este proceso es conocido como GARCH (generalized autoregressive conditional heteroskedasticity) y fue por primera vez planteado por Bollerslev (1986). Como en el caso de modelación en tiempo continuo, existen muy diversos modelos para la volatilidad estocástica. Bollerslev et al. (1992, 1994); Engle and Ishida (2002) presentan un compendio de generalizaciones al modelo ARCH de Engle. Shephard (2005) ejemplifica la modelación de la volatilidad estocástica con otros métodos y establece una comparativa entre estos modelos y los modelos tipo ARCH.

Una de las razones del éxito de este tipo de modelos en las aplicaciones es su capacidad de reproducir las *propiedades empíricas estilizadas*. Las propiedades empíricas estilizadas son propiedades no triviales comunes a una amplia gama de instrumentos, un denominador común entre las propiedades observadas en datos históricos de diferentes mercados que abarcan más de medio siglo. Las siguientes son las más conocidas de ellas:

1. Autocorrelación idénticamente cero en el proceso $\{Y_t\}$.

2. Decaimiento lento de las autocorrelaciones del proceso $\{|\, Y_t\,|\}$ y también de $\{Y_t^2\}$.

3. Colas pesadas en las distribuciones marginales.

4. Agrupación de la volatilidad.

5. Colas condicionales pesadas.

Esta no es la única afinidad entre la modelación en tiempo discreto y continuo. De hecho, los modelos en tiempo discreto pueden entenderse como aproximaciones naturales de los modelos de tiempo continuo. Como ejemplo concreto, Nelson (1990) demuestra que los procesos ARCH son una aproximación discreta del modelo

$$dV_t = \lambda(a - V_t)dt + \sigma V_t dW_t$$

donde W_t es un movimiento Browniano. Puede demostrarse que en este caso V_t admite la representación

$$\begin{aligned} \xi_t =& -\sigma W_t + (\tfrac{\sigma^2}{2} + \lambda)t, \\ V_t =& e^{-\lambda t}\left(\lambda a \int_0^t e^{\xi_s} ds + V_0\right), \end{aligned}$$

de modo que es un caso particular de (2). Drost and Werker (1996) examinan los procesos de difusión que surgen como límites débiles de procesos GARCH(1,1) a escala temporal $h\mathbb{N}$ con $h \to 0$ y muestran que son soluciones a ecuaciones de la forma

$$\begin{aligned} dX_t =& \sqrt{V_t} dW_t^{(1)} \\ dV_t =& \theta(\gamma - V_t)dt + \varrho V_t dW_t^{(2)} \end{aligned} \tag{3}$$

donde $W_t^{(1)}$ y $W_t^{(2)}$ son movimientos Brownianos independientes. Este estudio tuvo por objetivo el construir un proceso estocástico en tiempo continuo que tuviese las propiedades estilizadas (1–5) mencionadas anteriormente y concluyó con el proceso (3), que fue llamado GARCH a tiempo continuo. Llama la atención que en el proceso GARCH a tiempo discreto se tiene una única fuente de aleatoriedad, la sucesión $\{Z_t\}$, en tanto que en la difusión límite aparecen dos fuentes de aleatoriedad independientes, $W_t^{(1)}$ y $W_t^{(2)}$.

Esto motivó a Klüppelberg et al. (2004) a construir un modelo en tiempo continuo análogo al GARCH en tiempo discreto, especificado mediante las ecuaciones

$$\begin{aligned} dY_t =& \sqrt{V_{t-}} dL_t, \\ dV_t =& -\beta(V_{t-} - c)dt + \alpha V_{t-} d[L, L]_t^{\delta}, \end{aligned}$$

donde (L_t) es un proceso de Lévy unidimensional, $[L, L]^{\delta}$ es la parte discontinua de su variación cuadrática y los parámetros α, β, c son positivos. Este modelo es conocido como COGARCH(1,1). Brockwell et al. (2006) han extendido el concepto al caso COGARCH(p,q). En ambos casos, la intención de los modelos es imitar las propiedades que tiene el proceso GARCH a tiempo discreto en un proceso de volatilidad estocástica a tiempo continuo. Klüppelberg et al. (2004) establecen una comparativa muy precisa entre las propiedades del GARCH y del COGARCH.

En este escenario resalta la importancia de conocer las propiedades de los procesos GARCH unidimensionales en tiempo discreto: Dan lugar al planteamiento de modelos en tiempo continuo con propiedades *deseables* para la modelación financiera.

A pesar de que los modelos unidimensionales han sido empleados con éxito para describir el comportamiento de activos financieros por separado, la interdependencia entre distinos precios, retornos e incluso mercados es evidente tanto de consideraciones teóricas como de observaciones empíricas. Por lo tanto, el uso de procesos de volatilidad estocástica multivariados es indispensable para mejorar la presición de los modelos en estos ámbitos.

La generalización del proceso (1) a dimensión superior es directa y da lugar a

$$Y_t = \begin{cases} \int_0^t H_{s-}^{1/2} dL_s, & \text{si } T = [0, \infty), \\ H_t^{1/2} Z_t, & \text{si } T = \mathbb{N}. \end{cases} \tag{4}$$

En este caso H_t es una matriz de dimensiones $d \times d$ y $H_t^{1/2}$ representa una matriz tal que

$$H_t^{1/2} (H_t^{1/2})^T = H_t. \tag{5}$$

Para $t \in [0, \infty)$ el proceso (L_t) es un proceso de Lévy d–dimensional y para $t \in \mathbb{N}$ la sucesión $\{Z_t\}$ es de vectores aleatorios independientes e idénticamente distribuídos, centrados y con matriz de varianzas y covarianzas igual a la identidad.

De nuevo, es necesario modelar estocásticamente la dinámica de H_t. En el caso unidimensional, es indispensable en toda modelación de volatilidad estocástica garantizar que $V_t \geq 0$ para todo valor de t, puesto que V_t representa la varianza condicional de Y_t dado el pasado del proceso. Este requerimiento adquiere mayor complejidad en el caso multidimensional puesto que H_t es la matriz de varianzas y covarianzas condicionales del vector Y_t dado el pasado del proceso, de modo que es necesario garantizar que es no–negativa definida para todo t.

Barndorff-Nielsen and Stelzer (2007) estudian las condiciones para representar por medio de ecuaciones diferenciales estocásticas procesos de Lévy y procesos tipo Ornstein–Uhlenbeck con valores en el cono cerrado de matrices no–negativas definidas. Similar al caso unidimensional, estos procesos pueden representarse como la solución de

$$dH_t = \mathbf{A}H_t dt + dL_t,$$

donde $\mathbf{A}$ es un operador matricial y (L_t) es un proceso de Lévy con valores en el espacio de matrices cuadradas de dimensión d denotado por $\mathcal{M}(d \times d)$. Si (L_t) es un subordinador matricial y H_0 es no–negativa definida, entonces H_t es no–negativa definida para toda t, lo cual permite utilizar este tipo de procesos como modelos de volatilidad estocástica.

En el caso de tiempo discreto, también se tienen modelos de volatilidad estocástica generalizados a mayores dimensiones. Para definirlos se han tomado dos alternativas, definir la dinámica de H_t directamente y buscar condiciones para que sea no–negativa definida o bien definir la dinámica de $H_t^{1/2}$ y garantizar que H_t es no–negativa definida mediante la ecuación (5).

Como ejemplo del primer método, y dado que H_t debe modelar la matriz de varianzas y covarianzas (condicional) de un proceso estocástico, podemos considerar la representación

$$H_t = D_t R_t D_t$$

donde D_t es una matriz diagonal no–negativa con $D_t(i,i) = \sqrt{H(i,i)}$. Obsérvese que $H_t(i,i)$ es la varianza (condicional) de $Y_t(i)$, de modo que es no–negativa y la raíz es tomada, también, no–negativa. La matriz R_t es la matriz de correlación condicional $R_t(i,j) = Corr(Y_t(i), Y_t(j)|\mathcal{F}_{t-1})$.

Podemos simplificar el problema de modelación suponiendo que R_t es independiente de t, es decir, que es constante en el tiempo. Esto da lugar a los modelos de correlaciones condicionales constantes para los cuáles

$$H_t = D_t R D_t,$$

de modo que para generar un modelo de volatilidad estocástica con correlaciones condicionales constantes, basta describir la dinámica de la matriz diagonal no–negativa D_t.

Entre los modelos con correlaciones condicionales constantes, destaca el GARCH, dado en Bollerslev (1990), que especifica

$$\delta(D_t) = c + \sum_{i=1}^{p} A_i \delta(D_{t-i}) + \sum_{j=1}^{q} \delta(Y_{t-j} Y_{t-j}^T). \tag{6}$$

El símbolo $\delta(M)$ representa el vector en $\mathbb{R}^d$ cuyas componentes son los elementos diagonales de la matriz M. Los parámetros $C \in \mathbb{R}^d$, $A_i \in \mathcal{M}(d \times d)$, $B_j \in \mathcal{M}(d \times d)$ tienen entradas no–negativas. Podemos ver que este planteamiento es análogo al del GARCH unidimensional al considerar algunos valores pasados del proceso de volatilidad $\{D_{t-i}\}$ y los cuadrados de algunos valores pasados del proceso $\{Y_t\}$.

Las matrices $\{H_t\}$ son no–negativas–definidas, ya que las entradas de D_t son no–negativas y R es no–negativa–definida al ser una matriz de correlaciones. Son definidas positivas si y sólo si todas las varianzas condicionales son positivas y la matriz R es positiva definida.

Como ejemplo del segundo método de definición, el describir la dinámica de $H_t^{1/2}$, tenemos los modelos basados en el factor de Cholesky. En este caso, para simplificar la modelación, se supone que $H_t^{1/2}$ es una matriz triangular inferior y se representa vectorialmente mediante la operación de sobreposición de columnas de una matriz (vech(M)), cuya definición precisa puede consultarse en Horn and Johnson (1990).

Entre las especificaciones de este tipo contamos con el GARCH, propuesto en Kawakatsu (2003) que especifica

$$\text{vech}(H_t^{1/2}) = c + \sum_{i=1}^{p} A_i \text{vech}(H_{t-i}^{1/2}) + \sum_{j=1}^{q} B_j Y_{t-j}.$$

Comparando esta ecuación con (6), vemos que la operación vech(M) toma el lugar de $\delta(\cdot)$, lo cuál es de esperarse puesto que ahora se busca dar un modelo para todas las entradas de la matriz H_t, no sólo su diagonal. En contraste con (6), se utiliza el proceso Y_{t-j} en lugar de $\mathrm{vech}(Y_{t-j}Y_{t-j}^T)$. Podemos justificar esto observando que Y_{t-j} es la *raíz cuadrada* de $Y_{t-j}Y_{t-j}^T$ y el modelo está dado para $H_t^{1/2}$, la raíz de H_t.

El parámetro $C \in \mathbb{R}^{\frac{d(d+1)}{2}}$ se supone distinto de cero. Los parámetros $A_i \in \mathcal{M}(\frac{d(d+1)}{2} \times \frac{d(d+1)}{2})$ tienen todos sus renglones distintos del vector 0 al igual que las matrices $B_j \in \mathcal{M}(\frac{d(d+1)}{2} \times d)$. Como se mencionó anteriormente H_t es automáticamente no–negativa definida en virtud de (5).

En este trabajo, demostramos que las propiedades asintóticas de estos modelos multivariados de volatilidad estocástica discretos son consistentes con las de sus análogos univariados. Las técnicas utilizadas en cada caso son distintas por las diferencias naturales entre ambos modelos. En el caso del GARCH con correlaciones condicionales constantes, mostramos que bajo las hipótesis

$\mathcal{H}0$: Las matrices A_i, B_j tienen todos sus renglones distintos del vector 0.

$\mathcal{H}1$: La distribución F de los ruidos admite una densidad f con soporte en todo el espacio $\mathbb{R}^d$.

$\mathcal{H}2$: Para todo $\theta \geq 1$ existe $h > 1$ tal que

$$\theta^h \leq \mathbb{E}\left[\eta_{t,j}^{2h}\right] \leq \infty \qquad \text{para todo } 1 \leq j \leq n,$$

las distribuciones finito dimensionales del proceso son de variación regular. Esta propiedad es la versión multivariada de la presencia de colas pesadas. Utilizando la teoría de valores extremos multivariados, mostramos también que el proceso $\{\|Y_t\|\}$ presenta aglomeraciones con tamaño promedio $1/\theta > 1$, donde θ representa el *índice extremo* de la sucesión $\{Y_t\}$, cuya definición puede consultarse en Leadbetter et al. (1983). Por último, demostramos que si los ruidos $\{Z_t\}$ tienen distribuciones marginales simétricas, entonces el comportamiento asintótico de la función muestral de autocovarianzas

$$\gamma_{Y,n}(h) = \frac{1}{n}\sum_{t=1}^{n-h} Y_t Y_{t+h}^T, \qquad h \geq 0, \tag{7}$$

depende del índice de variación regular del proceso $\{Y_t\}$. Denotando por $\alpha > 0$ este índice, distinguimos tres casos: Si $\alpha \in (0, 2)$ no tenemos consistencia, sino convergencia a una variable aleatoria $\alpha/2$ estable. Si $\alpha > 2$

tenemos consistencia, pero un Teorema Límite Central sólo es posible para valores $\alpha > 4$.

En este desarrollo se hace uso de algunas propiedades conocidas y demostradas en Boussama (1998) del proceso. En particular, la representación estocástica recursiva

$$\widetilde{Y}_t = A(\eta_t)\widetilde{Y}_{t-1} + G,$$

con

$$\begin{aligned}
\widetilde{Y}_t =& (\delta(H_{t+1})^T, \dots, \delta(H_{t-q+2})^T, \delta(Y_t Y_t^T)^T, \dots, \delta(Y_{t-p+2} Y_{t-p+2}^T)^T)^T \\
G =& (C^T, 0, \dots, 0)^T
\end{aligned}$$

$$A(\eta_t) = \begin{bmatrix}
A_1 \mathrm{diag}(\eta_t \eta_t^T) + B_1 & B_2 & \cdots & B_{q-1} & B_q & A_2 & A_3 & \cdots & A_p \\
I & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
0 & I & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & I & 0 & 0 & 0 & \cdots & 0 \\
\mathrm{diag}(\eta_t \eta_t^T) & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & I & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & I & 0
\end{bmatrix}$$

y sus implicaciones tanto en la existencia y unicidad de una distribución estacionaria para el proceso $\{Y_t\}$ como en las propiedades de "mixing" del proceso estacionario.

Para el modelo GARCH del factor de Cholesky, mostramos que bajo las hipótesis

$\overline{\mathcal{H}_0}$: $\varrho(\sum_{i=1}^p A_i) < 1$

$\overline{\mathcal{H}_1}$: Los vectores $\{Z_t\}$ tienen una densidad f positiva y continua en $\mathbb{R}^d$

$\overline{\mathcal{H}_3}$: Existe $h_0 \in (0, \infty]$ tal que

$$\mathbb{E}\left[| Z_t(i) |^h \right] < \infty \qquad \text{si } h < h_0,$$

$$\lim_{h \to h_0} \mathbb{E}\left[| Z_t(i) |^h \right] = \infty.$$

$\overline{\mathcal{H}_4}$: Existe $k \in \mathbb{N}$ tal que los vectores

$$\{\{V_i\}_i, \{\widetilde{V}_i\}_i, \{A(0)V_i\}_i, \{A(0)\widetilde{V}_i\}_i, \dots \{A(0)^r V_i\}_i, \{A(0)^r \widetilde{V}_i\}_i\}$$

generan $\mathbb{R}^D$.

el proceso $\{Y_t\}$ admite una única distribución estacionaria, que el proceso estacionario es "β–mixing" a taza geométrica y finalmente que sus distribuciones finito dimensionales son de variación regular. Esto está basado en la representación estocástica recursiva

$$\widetilde{Y}_t = A(Z_t)\widetilde{Y}_{t-1} + G$$

con

$$\begin{aligned}
\widetilde{Y}_t =& (\mathrm{vech}(L_{t+1})^T, \ldots, \mathrm{vech}(L_{t-p+2})^T, Y_t^T, \ldots, Y_{t-q+2}^T)^T \\
G =& (C^T, 0, \ldots, 0)^T
\end{aligned}$$

$$A(Z_t) = \begin{bmatrix}
A_1 + (Z_t^T \otimes B_1)K & A_2 & \cdots & A_{p-1} & A_p & B_2 & B_3 & \cdots & B_q \\
I & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
0 & I & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & I & 0 & 0 & 0 & \cdots & 0 \\
(Z_t^T \otimes I)K & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & I & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & I & 0
\end{bmatrix}$$

Obsérvese que en contraste con la representación recursiva para el GARCH con correlaciones condicionales constantes, las matrices involucradas en este modelo y sus productos pueden tener entradas negativas. Aunque esta diferencia puede parecer trivial a primera vista, en realidad no lo es y sortearla requiere del uso de metodologías muy distintas a las usadas en el caso de matrices con entradas no–negativas.

En particular, para mostrar la irreducibilidad de la cadena de Markov

$$W_0 = x_0, \qquad W_n = \frac{A(Z_n)W_{n-1}}{\|A(Z_n)W_{n-1}\|}$$

sobre los conjuntos abiertos de la esfera $\mathbb{S}_{d-1}$ fue necesario considerar una transformación del proceso W_n como un sistema dinámico estocástico. De la controlabilidad de este sistema se puede obtener una hipótesis suficiente para concluir que el proceso $\{W_n\}$ visita eventualmente cualquier abierto de $\mathbb{S}_{d-1}$ con probabilidad positiva.

Otra diferencia es que en este caso es necesario que el componente no–singular de la distribución de

$$vec(\Pi_n) = vec(A(Z_n)\dots A(Z_1))$$

con respecto a la medida de Lebesgue en $\mathbb{R}^d$ tenga una densidad uniformemente acotada por abajo sobre un cubo $C \subseteq \mathbb{R}^d$.

Para lograr esta cota escribimos

$$vec(\Pi_n) = (I \otimes A(Z_n))vec(\Pi_{n-1})$$

y mostramos que el Kernel de transición $P(\cdot,\cdot)$ del proceso Markoviano $\{vec(\Pi_n)\}$ admite un componente continuo, es decir, una función semicontinua inferiormente $T(\cdot,\cdot)$ tal que

$$P(x, A) \geq T(x, A).$$

De esta relación, y bajo la hipótesis de que la distribución de $\{Z_n\}$ admite una densidad continua, obtenemos la cota deseada.

También demostramos las propiedades de agrupación del proceso $\{\|Y_t\|\}$ a partir de la teoría de valores extremos. Encontramos que las propiedades asintóticas de la función de autocovarianzas dependen del índice de variación regular exactamente como en el caso de correlaciones condicionales constantes, con la salvedad de que en este caso no necesitamos suponer que las distribuciones marginales de la sucesión $\{Z_t\}$ son simétricas.

Nuestra mayor contribución con este trabajo es el estudio de las propiedades del modelo GARCH del factor de Cholesky. En este caso nuestra propuesta sobre los parámetros del modelo hace posible demostrar la existencia y unicidad de una distribución estacionaria y la presencia de las propiedades empíricas estilizadas.

El plan de la Tesis es el siguiente. En el Capítulo 1 se dan los resultados preliminares que sirven de base para el desarrollo posterior. El Capítulo 2 está dedicado al proceso GARCH con correlaciones condicionales constantes. El contenido de este capítulo ha sido aceptado para su publicación en Journal of Multivariate Analysis, vol. 100, no.7, agosto de 2009 y está disponible en línea en http://dx.doi.org/10.1016/j.jmva.2009.01.002. El modelo GARCH para el factor de Cholesky es estudiado en el capítulo 3. El contenido de este capítulo ha sido sometido para su publicación.

Se ha elegido redactar la Tesis en lengua inglesa por ser más apropiado para su revisión y uso internacionales.

Chapter 1

Preliminaries

This Chapter is devoted to the presentation of the theories that are the basis for the developement of our results. We briefly outline Multivariate Extreme Value Theory as applied to GARCH processes in Chapters 2 and 3. First, in Sections 1.1, through 1.3 we present the classical theory of independent, identically distributed vectors. The results developed in these Sections are then extended to stationary sequences of random variables in Section 1.4.

Next, we focus on an important class of stochastic processes, the ones that admit a recursive representation in Section 1.6. We state sufficient conditions for the stationary version of these processes to be multivariate regularly varying and finally, in Section 1.7 we explain further properties of stochastic recursions when embeded into Markov processes.

As was discussed in the Introduction, the stylized facts are most important properties when it comes to modelling stochastic volatility in discrete time. It is by using the concepts of *multivariate regular variation* and *multivariate extremal index*, which are explained in this Chapter, that a theoretically solid proof of these properties can be provided for multivariate GARCH processes in Chapters 2 and 3. Furthermore, stochastic recursions and point processes are a technical requirement of these developements.

1.1 Multivariate Extremes

In this section we overview some results on Multivariate Extreme Value Theory. In the multidimensional case the term "extreme" holds a degree of ambiguity. The notion of maxima is unclear, and how to define it should answer some of the needs the theory itself answers.

Consider the financial scenario of managing a portfolio and recording the values of the different assets in it in the vectors $X_i = (X_i^{(1)}, \ldots, X_i^{(d)})$ day after day. The maximum may be defined here componentwise suggesting that we are interested in the extremal behaviors of each of the assets $X_i^{(j)}$, taking into account the way they relate to each other.

As another example, we may think of $X_n^{(i)}$ as the n–th measurement of the temperature of a specific location. We then may be interested in the fluctuations of $\{X_n^{(i)}\}$. By considering the random vector $\{X_n^{(i)}, 1 \leq i \leq d\}$ and the componentwise maxima we will understand the temperature fluctuation in each region and the way they relate to each other too, which is more informative than marginal considerations.

Definition 1. *Given a sequence $\{X_n\}$ of random vectors in $\mathbb{R}^d$, with components $X_n^{(i)}$, set for $n \geq 1$*

$$M_n = \left(\max_{i \leq n} X_i^{(1)}, \ldots, \max_{i \leq n} X_i^{(d)} \right) = (M_n^{(1)}, \ldots, M_n^{(d)})$$

One of the main concerns of Multivariate Extreme Value theory is characterizing the possible asymptotic distribution functions of the normalized and centered maxima. In other words, given $X_n = (X_n^{(1)}, \ldots X_n^{(d)})$ a sequence of random vectors with common distribution function F, it is intended to find sequences of real vectors

$$a_n = (a_n^{(1)}, a_n^{(2)}, \ldots, a_n^{(d)}) \qquad b_n = (b_n^{(1)}, b_n^{(2)}, \ldots, b_n^{(d)})$$

such that $a_n^{(i)} > 0$ for all $i = 1, 2, \ldots, d$ and a multivariate distribution function G for which

$$\mathbb{P}\left[\frac{M_n^{(i)} - b_n^{(i)}}{a_n^{(i)}} \leq x^{(i)}, 1 \leq i \leq d \right] \xrightarrow[n \to \infty]{} G(x^{(1)}, \ldots, x^{(d)}). \tag{1.1}$$

We will first assume that the vectors $\{X_n\}$ are i.i.d.

Definition 2. *Let G be as in* (1.1) *and with non–degenerate marginal distribution functions G_i. We call G a multivariate extreme value distribution (MEVD) and the class of distribution functions F for which* (1.1) *holds, the Maximum Domain of Attraction of G, denoted by MDA(G).*

Remark 1. *The marginal distributions G_i are non–degenerate. Since joint convergence entails marginal convergence, the Fisher–Tippet Theorem (See Embrechts et al. (1997)) implies that for $i = 1, \ldots, d$, the distribution function G_i is a real extreme value distribution with norming constants $a_n^{(i)}, b_n^{(i)}$.*

We say a multivariate distribution G is max–stable if there exist functions $\alpha^{(i)}(t) > 0, \beta^{(i)}(t)$ such that

$$G^t(x) = G(\alpha^{(1)}(t)x^{(1)} + \beta^{(1)}(t), \ldots, \alpha^{(d)}(t)x^{(d)} + \beta^{(d)}) \tag{1.2}$$

for all $t > 0$.

Proposition 1. *The class of MEVDs is exactly the class of max–stable distribution functions with non–degenerate marginals.*

Proof. If G is max–stable with functions $\alpha^{(i)}(t) > 0$ and $\beta^{(i)}(t)$ and it has non–degenerate marginals, then $G \in MDA(G)$ with norming constants $a_n^{(i)} = (\alpha^{(i)}(n))^{-1}$ and $b_n^{(i)} = -\beta^{(i)}(n)/\alpha^{(i)}(n)$, so that G is MEVD.

To prove the converse, if G is MEVD then marginal convergence together with the Kintchine's convergence to types Theorem guarantee that

$$\alpha^{(i)}(t) = \lim_{n\to\infty} \frac{a_n^{(i)}}{a_{[nt]}^{(i)}}, \qquad \beta^{(i)}(t) = \lim_{n\to\infty} \frac{b_n^{(i)} - b_{[nt]}^{(i)}}{a_{[nt]}^{(i)}}$$

both exist. Since

$$\mathbb{P}\left[M_{[nt]} \leq a_n x + b_n\right] = F^{[nt]}(a_n x + b_n) = (F^n(a_n x + b_n))^{[nt]/n},$$

we have

$$(M_{[nt]}^{(i)} - b_n^{(i)})/a_n^{(i)} \xrightarrow{d} G^t. \tag{1.3}$$

On the other hand

$$(M_{[nt]}^{(i)} - b_{[nt]}^{i})/a_{[nt]}^{(i)} \xrightarrow{d} G. \tag{1.4}$$

Let $Y(t)$ be a vector with distribution function G^t (observe how equation (1.3) implies G^t is a distribution function) for each $t > 0$. Combining equations (1.3), (1.4) and the definition of the functions $\alpha^i(t), \beta^i(t)$ we obtain

$$\begin{aligned}
&((M_{[nt]}^{(i)} - b_{[nt]}^{(i)})/a_{[nt]}^{(i)}, 1 \leq i \leq d) \\
&\quad = \left(\left(\frac{M_{[nt]}^{(i)} - b_n^{(i)}}{a_n^{(i)}}\right)\frac{a_n^{(i)}}{a_{[nt]}^{(i)}} + \frac{b_n^{(i)} - b_{[nt]}^{(i)}}{a_{[nt]}^{(i)}}, 1 \leq i \leq d\right) \\
&\qquad \Rightarrow (\alpha^{(i)}(t)Y^{(i)}(t) + \beta^{(i)}(t), 1 \leq i \leq d) \stackrel{d}{=} Y(1),
\end{aligned}$$

so that G is max–stable. ∎

The most common approach to study multivariate domains of attraction is to transform the coordinates making them have the standard Fréchet distribution. For a detailed treatment see Resnick (1987), and for an alternative method, based on copulas, see Hsing (1989). By changing the marginal distribution functions, no difficulties arise as is shown in the next Proposition.

Proposition 2. *Let G be a multivariate distribution function with marginals G_i and let $X \sim G$. For $y \in \mathbb{R}, x \in \mathbb{R}^d$, define*

$$\varphi_i(y) = \left(\frac{1}{-\log G_i}\right)(y) \quad \text{and} \quad \varphi(x) = (\varphi_1(x^{(1)}), \ldots, \varphi_d(x^{(d)})).$$

Name G_* the distribution function of $\varphi(X)$. Then G_* has standard Fréchet marginals and G is a MEVD if and only if G_* is also a MEVD.

Furthermore, given a multivariate distribution function F with marginals F_i and $Y \sim F$, let

$$U_i(y) = \frac{1}{1-F_i}(y) \quad \text{and} \quad U(x) = (U_1(x^{(1)}), \ldots, U_d(x^{(d)})).$$

Name F_* the distribution function of $U(Y)$. Then if $F \in MDA(G)$ with norming norming vectors a_n, b_n, we have $F_* \in MDA(G_*)$ with norming vectors $a_{n*} = (n, n, \ldots, n), b_{n*} = 0$.

Conversely, if $F_* \in MDA(G_*)$ with the given norming vectors and $F_i \in MDA(G_i)$ with norming constants $a_n^{(i)}, b_n^{(i)}$ for each $i = 1, 2, \ldots, d$, then $F \in MDA(G)$.

In what follows the next notation is used for $a, b \in \mathbb{R}^d$

$$(a, b) = \{x \in \mathbb{R}^d : \ a^{(i)} < x^{(i)} < b^{(i)}, \ 1 \leq i \leq d\}$$

and analogously for $[a, b), [a, b]$ so that comparison between elements in $\mathbb{R}^d$ is defined componentwise. Also, given $l \in \mathbb{R}^d$ the set $E = [l, \infty] \setminus \{l\}$ is the compact set $[l, \infty]$ with the point l removed and topologized so that compact sets in E are those bounded away from l and E remains complete (See Linskog (2004); Resnick (1987)).

Proposition 3. *Let F be a multivariate distribution function. The following are equivalent:*

1. *F^t is a distribution function for all $t > 0$,*

2. *There exists* $l \in \mathbb{R}^d$ *and a Radon measure* μ *on the space* $E = [l, \infty]\backslash\{l\}$ *such that*
$$F(x) = \exp\{-\mu((-\infty, x]^c)\}.$$
The measure μ *satisfies*
$$\mu([-\infty, x]^c) \to 0, \qquad x \to \infty \tag{1.5}$$
$$l = -\infty, \quad or \quad x \geq l \text{ and } x^{(i)} = -\infty \Rightarrow \mu([-\infty, x]^c) = \infty. \tag{1.6}$$

3. *There exists* $l \in \mathbb{R}^d$ *and a Radon measure* μ *on the space* $E = [l, \infty]\backslash\{l\}$ *as in part 2 and* $PRM(dt \times d\mu)$ *on* $[0, \infty) \times E$*, given by* $\xi = \sum_k \delta(t_k, j_k)$ *such that*
$$F(y) = \mathbb{P}\left[\max_{t_k \leq t} j_k \vee l \leq y\right].$$

Proof. Call F max–id if F satisfies the hypothesis of part 1. We show 1 implies 2. Let F be max–id. It can be shown that the set $A = \{F > 0\}$ is a rectangle, $A = \otimes_{i=1}^d A_i$ where each A_i is of the form $[l_i, \infty)$ or $(-\infty, \infty)$. See Resnick (1987).

Set $l = \inf(A)$ and $E = [l, \infty] \setminus \{l\}$ and define the Radon measures $\mu_n = nF^{1/n}$. Observe that for $x \in E$
$$\mu_n([l, x]^c) = n(1 - F^{1/n}(x)) \sim -\log(F(x)) < \infty.$$
This means that for every $x \in E$ we have
$$\sup_{n \geq 1} \mu_n([l, x]^c) < \infty.$$
Since compact sets of E are bounded away from l it is evident that given a relatively compact set B, there exists $x \in E$ such that $B \subseteq [-\infty.x]^c$ so that for all relatively compact B
$$\sup_{n \geq 1} \mu_n(B) < \infty.$$
This condition is enough for μ_n to be relatively compact in the vague topology (see Kallenberg (1969)). Let $\mu_I, \mu_I I$ be two limit points, then
$$\mu_I([-\infty, x]^c) = \lim_{n \to \infty} \mu_n([-\infty, x]^c) = -\log(F(x)).$$

Since this is also true for μ_{II}, it follows that for all $x \in E$ we have

$$\mu_I([-\infty, x]^c) = \mu_{II}([-\infty, x]^c)$$

which is enough to have $\mu_I \equiv \mu_{II}$.

Thus, if μ is the vague limit of μ_n then μ is a Radon measure and

$$F(x) = \exp\{-\log F(x)\} = \exp\{-\mu([-\infty, x]^c\}.$$

That μ satisfies the conditions (1.5) and (1.6) follows from considering the limit properties of distribution functions. ■

As was said earlier, it is enough to focus on MEVD's with standard Fréchet marginals G_*. For G_* the max–stability condition becomes

$$G_*^t(tx) = G_*(x), \tag{1.7}$$

which follows from the marginal considerations, by which we may take $a_n^{(i)} = n, b_n^{(i)} = 0$. Let μ_* be the exponent measure of G_*. Since each marginal is Fréchet, and hence concentrated on $[0, \infty)$, it is appropriate to take $E = [0, \infty]^d \setminus \{0\}$ as the support of μ_*. In terms of μ_*, equation (1.7) is

$$\mu_*([0, x]^c) = t\mu_*([0, tx]^c) = t\mu_*(t[0, x]^c). \tag{1.8}$$

For a fixed $t > 0$ this relation holds for all rectangles in E, and so

$$\mu_*(B) = t\mu_*(tB) \tag{1.9}$$

for a generating π–system and thus for all Borel subsets of E. Thus (1.9) and (1.7) are equivalent.

From condition (1.5) it follows that

$$\mu_*\left(\bigcup_{i=1}^{d}\{y \in E :\ y^{(i)} = \infty\}\right) = 0. \tag{1.10}$$

Let $|\cdot|$ be any norm on $\mathbb{R}^d$. The fact that $|\cdot|$ is not defined in all of E is immaterial because of (1.10). Let $\mathbb{S}_{d-1}$ be the unit sphere in $\mathbb{R}_d$, which is compact on E because it is bounded away from 0, for any given norm. Define the measure σ on $\mathcal{B}(\mathbb{S}_{d-1})$ by

$$\sigma(A) = \mu_*\left(\{x :\ |x| > 1,\ x/|x| \in A)\right.$$

Since μ_* is a Radon measure, if $\mathbb{S}_{d-1} \cap E$ is compact, then σ defines a finite measure. Consider $T\colon E \to (0,\infty] \times \mathbb{S}_{d-1}$ the "polar" coordinate transformation

$$T(x) = (|x|, x/|x|)$$

With respect to this coordinates μ_* is a product measure: The first coordinate governs length while the second governs directions. Indeed, for $r > 0$ and $A \in \mathcal{B}(\mathbb{S}_{d-1})$ we have

$$\begin{aligned}
&\mu_*\left(\{y :\ |y| > r, y/|y| \in A\}\right) \\
&\quad = r^{-1}\mu_*\left(\{r^{-1}y :\ |y|, y/|y| \in A\}\right) \\
&\quad = r^{-1}\mu_*\left(\{x :\ |x| > 1, x/|x| \in A\}\right) = r^{-1}\sigma(A),
\end{aligned}$$

which implies that

$$\mu_* \circ T^{-1}(dr, da) = r^{-2} dr \sigma(da).$$

Therefore we can write

$$\begin{aligned}
\mu_*([0,x]^c) &= \iint_{T([0,x]^c)} r^{-2} dr \sigma(da) \\
&= \int_{\mathbb{S}_{d-1}} \sigma(da) \left(\int_{\left\{r > \min_i \frac{x^{(i)}}{a^{(i)}}\right\}} r^{-2} dr \right) \\
&= \int_{\mathbb{S}_{d-1}} \bigvee_{i=1}^{d} \left(\frac{a^{(i)}}{x^{(i)}} \right) \sigma(da).
\end{aligned}$$

Putting all this facts together we get the next Proposition.

Proposition 4 (Representation of MEVD's)**.** *The following are equivalent*

1. *G_* is a MEVD with standard Frechét marginals,*

2. *There exists a finite measure σ on $\mathbb{S}_{d-1}$ such that*

$$\int_{\mathbb{S}_{d-1}} a^{(i)} \sigma(da) = 1, \qquad 1 \le i \le d \tag{1.11}$$

and that for $x \in \mathbb{R}^d$

$$G_*(x) = \exp\left\{ - \int_{\mathbb{S}_{d-1}} \bigvee_{i=1}^{d} \left(\frac{a^{(i)}}{x^{(i)}} \right) \sigma(da) \right\}$$

3. *There exists* $\sum_k \delta_{(t_k,j_k)}$, *a PRM(*$dt \times d\mu_*$*) on* $[0,\infty) \times E$ *with*

$$\mu_* \left(\{y : \ |y| > r, y/|y| \in A\}\right) = r^{-1}\sigma(A)$$

and σ *a finite measure on* $\mathbb{S}_{d-1}$ *satisfying* (1.11) *such that for* $x \geq 0$

$$G_*(x) = \mathbb{P}\left[\bigvee_{t_k \leq 1} j_k \leq x\right]$$

Proof. It only remains to prove (1.11). Since G_* has Fréchet marginals we have

$$G_*^{(i)}(x^{(i)}) = \lim_{\substack{x^{(j)} \to \infty \\ j \neq i}} G_*(x) = \exp\{-(x^{(i)})^{-1}\} \tag{1.12}$$

On the other hand, as proved before

$$\mu_*([0,x]^c) = \int_{\mathbb{S}_{d-1}} \bigvee_{i=1}^{d} \left(\frac{a^{(i)}}{x^{(i)}}\right) \sigma(da)$$

so that

$$G_*^{(i)}(x^{(i)}) = \exp\left\{-(x^{(i)})^{-1} \int_{\mathbb{S}_{d-1}} a^{(i)} \sigma(da)\right\} \tag{1.13}$$

which completes the proof by comparing (1.12) with (1.13) ■

Remark 2. *The measure* σ *associated to the exponent measure* μ_* *is known as "spectral measure". It describes the distribution of the angular component of the points* $\{j_k\}$ *of the associated Point Process and is relevant in the study of stationary sequences with Point Processes as will be seen later.*

1.1.1 Regular variation of real valued functions

The notion of regular variation is very common in applications and particularly important in Extreme Value Theory. Basically, a regularly varying function of real variable behaves asymptotically as a power function.

Definition 3. *A function* f *is*

- *slowly varying if*

$$\lim_{x \to \infty} \frac{f(xt)}{f(x)} = 1$$

- *regularly varying of index α for some $\alpha \in \mathbb{R}$ if*

$$\lim_{x\to\infty} \frac{f(xt)}{f(x)} = t^{\alpha}$$

We call α the exponent of regular variation. We shall denote by $RV(\alpha)$ the set of all regularly varying functions with exponent α. It is immediate from the definition that a regularly varying function of index α can be expressed as $f(x) = x^{\alpha}L(x)$ where $L(x)$ is a slowly varying function. Of course, the canonical regularly varying function of index α is x^{α} itself.
The functions $\ln(1+x)$, $\ln(\ln(e+x))$, $\exp\{(\ln(x))^{\beta}\}$ with $0 < \beta < 1$ are slowly varying. Any function with $\lim_{x\to\infty} f(x) = f(\infty) < \infty$ is slowly varying. Examples of functions which are not regularly varying include e^x or $sin(x+2)$.
As we see in the definition, regular variation implies both, the existence of the limit

$$\lim_{x\to\infty} \frac{f(xt)}{f(x)} = h(t) \quad \text{and} \quad h(t) = t^{\alpha}$$

for some $\alpha \in \mathbb{R}$. Observe, however, that given that $h(t)$ exists as defined above we have for $s > 0, t > 0$

$$\frac{f(xst)}{f(x)} = \frac{f(xst)}{f(tx)} \frac{f(tx)}{f(x)}$$

which, letting $x \to \infty$, yields

$$h(st) = h(s)h(t)$$

so that h satisfies the Hamel equation and is therefore of the form x^{α} for some $\alpha \in \mathbb{R}$. This means we can weaken slightly the definition of regular variation.

Definition 4. *A function f is regularly varying if there exists a function h such that for all $t > 0$*

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = h(t)$$

In this case $h(x) = x^{\alpha}$ for some real number α which is called exponent of variation.

Some widely used properties of regularly varying functions include their integrability properties which lead to a usefull representation. We present the results without proofs. See Bingham et al. (1989), Resnick (1987) and the references therein.

Theorem 1 (Karamata's Theorem). *1. If $\alpha \geq -1$ then $f \in RV(\alpha)$ implies $\int_0^t f(x)dx \in RV(\alpha+1)$ and*

$$\lim_{t\to\infty} \frac{tf(t)}{\int_0^t f(x)dx} = \alpha + 1$$

If $\alpha < -1$ then $f \in RV(\alpha)$ implies $\int_t^\infty f(x)dx$ is finite and belongs to $RV(\alpha+1)$ and

$$\lim_{t\to\infty} \frac{tf(t)}{\int_t^\infty f(x)dx} = -\alpha - 1$$

2. If f satisfies

$$\lim_{t\to\infty} \frac{tf(t)}{\int_0^t f(x)dx} = \lambda \in (0,\infty)$$

then $f \in RV(\lambda-1)$. If $\int_t^\infty f(x)dx$ is finite and

$$\lim_{t\to\infty} \frac{tf(t)}{\int_t^\infty f(x)dx} = \lambda \in (0,\infty)$$

then $f \in RV(-\lambda-1)$.

Corollary 1 (Karamata's Representation). *The function L is slowly varying if and only if it can be represented as*

$$L(t) = c(t)\exp\left\{\int_1^t x^{-1}\varepsilon(x)dx\right\} \tag{1.14}$$

for $t > 0$ where $c\colon \mathbb{R}_+ \to \mathbb{R}_+$ and $\varepsilon\colon \mathbb{R}_+ \to \mathbb{R}_+$ and

$$\lim_{t\to\infty} c(t) = c > 0, \qquad \lim_{x\to\infty} \varepsilon(x) = 0$$

Observe that this Corollary can also be used to represent regularly varying functions of index α for real α: Given f a regularly varying function of index α, we know that $f(t) = t^\alpha L(t)$ where L is slowly varying. Using Karamata's Representation (1.14) on L we get

$$U(t) = c(t)\exp\left\{\int_1^t t^{-1}\alpha(x)dx\right\}$$

where c is as before and $\lim_{x\to\infty}\alpha(x) = \alpha$.

1.1.2 Multivariate regular variation

For multivariate valued functions the notion of regular variation has some different definitions. One of the most usefull ones follows.

Definition 5. *An $\mathbb{R}^d$ valued random vector X is regularly varying if there exist a sequence $\{a_n\}$ such that $a_n \nearrow \infty$ and a non–zero Radon measure μ on $\mathcal{B}(\overline{\mathbb{R}}^d \setminus \{0\})$ with $\mu(\overline{\mathbb{R}}^d \setminus \mathbb{R}^d) = 0$ such that, as $n \to \infty$*

$$n\mathbb{P}\left[a_n^{-1}X \in \cdot\,\right] \xrightarrow{v} \mu(\,\cdot\,), \qquad in\ \mathcal{B}(\overline{\mathbb{R}}^d \setminus \{0\}) \tag{1.15}$$

One feature of the limit measure μ appearing in definition 5 is scaling.

Theorem 2 (Linskog (2004))**.** *If the conditions of definition 5 hold, then there exists an $\alpha > 0$ such that $\mu(tB) = t^{-\alpha}\mu(B)$ for every $t > 0$ and measurable set $B \in \mathcal{B}(\overline{\mathbb{R}}^d \setminus \{0\})$.*

The best way to understand how this $\alpha > 0$ works is expressed in the next Theorem which gives an equivalent formulation of regular variation

Theorem 3 (Linskog (2004))**.** *Let X be an $\mathbb{R}^d$ valued vector. The following are equivalent*

1. *X is regularly varying in the sense of definition 5,*
2. *Let $|\cdot|$ denote any given, fixed norm on $\mathbb{R}^d$ and $\mathbb{S}_{d-1}$ be the ball of radius 1 for that norm. Then, there exists an $\alpha > 0$ and a probability measure σ on $\mathcal{B}(\mathbb{S}_{d-1})$, such that for every $x > 0$, as $t \to \infty$*

$$\frac{\mathbb{P}\left[|X| > tx, X/|X| \in \cdot\,\right]}{\mathbb{P}\left[|X| > t\right]} \xrightarrow{w} x^{-\alpha}\sigma(\,\cdot\,) \quad in\ \mathcal{B}(\mathbb{S}_{d-1}) \tag{1.16}$$

A very simple example of a regularly varying vector can be constructed as follows. Let $X = RU$ where R is a non–negative random variable and U is an $\mathbb{S}_{d-1}$–valued random element which are independent. If R is regularly varying of index $\alpha > 0$, then

$$\frac{\mathbb{P}\left[|X| > tx, X/|X| \in \cdot\,\right]}{\mathbb{P}\left[|X| > t\right]} = \frac{\mathbb{P}\left[R > tx\right]}{\mathbb{P}\left[R > t\right]}\mathbb{P}\left[U \in\ cdot\,\right] \xrightarrow{w} x^{-\alpha}\mathbb{P}\left[U \in \cdot\,\right].$$

Observe that the definition of regular variation given in Theorem 3 holds irrespective of the norm $|\cdot|$ in $\mathbb{R}^d$ chosen to specify $\mathbb{S}_{d-1}$. Therefore, given two different norms $|\cdot|_1$ and $|\cdot|_2$ on $\mathbb{R}^d$, the vector X is regularly varying with index α for the norm $|\cdot|_1$ if and only if it is also regularly varying with index α for the norm $|\cdot|_2$. However, the spectral measure differs for each choice of the norm.

Basrak et al. (2002a) explored the possibility of explaining regular variation of a given vector X as has been described with the regular variation of linear combinations of the entries of X. Formally, they explored the situation where there exists an $\alpha > 0$ and a slowly varying function L such that for all $x \in \mathbb{R}^d$ we have

$$\lim_{t\to\infty} \frac{\mathbb{P}\left[\langle x, X\rangle > t\right]}{t^{-\alpha}L(u)} = w(x) \quad \text{exists, and there exists a } x_0 \neq 0 \text{ with } w(x_0) > 0 \tag{1.17}$$

The choice of α plays an important role in the equivalence between (1.16) and (1.17) as shown in Basrak et al. (2002a).

Theorem 4. *[Basrak et al. (2002a)] Let X be an $\mathbb{R}^d$ valued vector*

1. *If X is regularly varying with index α, then condition* (1.17) *holds with the same value of α*

2. *If X satisfies condition* (1.17) *where α is positive and non–integer, then X is regularly varying with index α and the spectral measure σ is uniquely determined.*

3. *If X is non–negative in each entry and satisfies* (1.17) *where α is an odd integer, then X is regularly varying with index α and the spectral measure α is uniquely determined.*

Remark 3. *It is worth mentioning that for α even, the result doesn't hold in general. A counterexample can be found in Basrak et al. (2002a).*

1.2 Multivariate Domains of Attraction

Domains of attraction for multivariate extreme value distributions can be described by regular variation conditions similar to the real case.

As said before, we study only the distributions G_* with standard Frechét marginals.

Proposition 5. *[Domain of Attraction of G_*] Let G be a multivariate extreme value distribution and F a multivariate distribution function. Let G_* and F_* be defined as in Proposition 2. Then*

1. *$F_* \in D(G_*)$ if and only if $1 - F_*$ is regularly varying in $(0, \infty)^d$ with limit function*
$$w(x) = \frac{-\log(G_*(x))}{-\log(G_*(1))}$$
2. *$F \in D(G)$ if and only if marginal convergence to standard Frechét holds and $F_* \in D(G_*)$.*

Proof. We sketch the proof. For details see Resnick (1987).
For 1, assume first that $1 - F_*$ is regularly varying in $(0, \infty)^d$. Define a_n by

$$1 - F_*(a_n) \sim n^{-1}(-\log(G_*(1))),$$

so that replacing t with a_n one gets as $n \to \infty$

$$n(1 - F_*(a_n x)) \to -\log(G_*(x)), \qquad x > 0.$$

It follows immediately that

$$F_*^n(a_n x) \to G_*(x) \qquad x > 0,$$

from which $F_* \in D(G_*)$ with norming constants a_n. Conversely, if $F_* \in D(G_*)$ it can be shown by marginal considerations that we may take $a_n = n$ so that $F_* \in D(G_*)$ implies

$$F_*(nx) \to G_*(x), \qquad x > 0.$$

Taking logarithms and the usual Taylor approximation of $-\log(t)$ we get

$$n(1 - F_*(nx)) \to -\log(G_*(x)), \qquad x > 0.$$

Which in turn gives us

$$\frac{1 - F_*(nx)}{1 - F_*(n1)} \to \frac{-\log(G_*(x))}{-\log(G_*(1))}.$$

We now replace n by t by monotonicity and get the desired regular variaton of $1 - F_*$. ∎

Example 1
Consider the bivariate distribution given by

$$1 - F(x, y) = e^{-x} + e^{-y} - (e^x + e^y - 1)^{-1}$$

introduced by Marshall and Olkin in Marshal and Olkin page 176. Then, $F(x, y)$ has standard exponential marginals and, following the notation of Proposition 2, we have

$$U_i(x) = \frac{1}{1 - F_i(x)} = e^x \quad x > 0 \quad i = 1, 2$$
$$U_i^{-1}(y) = \log(y) \quad y > 1 \quad i = 1, 2$$

Therefore

$$\frac{1 - F_*(tx, ty)}{1 - F_*(t, t)} = \frac{x^{-1} + y^{-1} - (x + y - t^{-1})}{2 - (2 - t^{-1})} \to \frac{x^{-1} + y^{-1} - (x + y)^{-1}}{3/2}$$

as $t \to \infty$ verifying 1 of Proposition 5. We identify G_* by

$$- \log G_*(x, y) = x^{-1} + y^{-1} - (x + y)^{-1}$$

Now, to identify G we have, from marginal convergence that

$$n(1 - F_i(x + \log(n))) \to e^{-x}, \quad x \in \Re \quad i = 1, 2$$

from which it follows that, again in the notation of Proposition 2,

$$\varphi_i^{-1}(x) = 1/(-\log G_i(x)) = e^{-x}, \quad x \in \Re, \quad i = 1, 2$$

and

$$G(x, y) = G_*(\varphi_1^{-1}(x), \varphi_2^{-1}(y)) = \exp\{-(e^{-x} + e^{-y} - (e^x + e^y)^{-1}\}$$

In the next Proposition domains of attraction are explained in terms of the exponent of regular variation and the spectral mesure.

Proposition 6. *Let F, F_*, G, G_* be as in Proposition 5 and μ_*, σ, S_{d-1} as in Section 1.1. Let X_* be a random vector with distribution function F_*, then, the following are equivalent*

1. $F_* \in D(G_*)$

2. *For every* $x > 0$,

$$\lim_{t\to\infty} \frac{1 - F_*(tx)}{1 - F_*(t1)} = \frac{\mu_*([0,x]^c)}{\mu_*([0,1]^c)}$$

3. $n\mathbb{P}\left[n^{-1}X_* \in \cdot\,\right] \xrightarrow{v} \mu_*$ *on* $[0,\infty]^d \setminus \{0\}$

4. *As measures on* $\mathcal{B}(S_{d-1})$

$$\frac{\mathbb{P}\left[|\,X_*\,| > tx, X_*/\,|\,X_*\,| \in \cdot\,\right]}{\mathbb{P}\left[|\,X\,| > t\right]} \xrightarrow{v} x^{-1}\sigma(\,\cdot\,)$$

Proof. 1 is equivalent to 2 by propositon 5 and the representation of G_* by its exponent measure explained in section 1.1. To prove 3 we observe that $F_* \in D(G_*)$ means that $F_*^n(nx) \to G_*(x)$ for $x > 0$ which is equivalent to

$$n(1 - F_*(nx)) = n\mathbb{P}\left[n^{-1}X_* \in [0,x]^c\right] \to -\log G_*(x) = \mu_*([0,x]^c), \quad x > 0$$

Expanding this to convergence on rectangles we get the desired vague convergence on $[0,\infty]^d \setminus \{0\}$. For the equivalence between 3 and 4 see Resnick (1987). ■

We now restate this Proposition in a more appliable way, without the need to compute F_*.

Corollary 2. *Let* F *be a distribution funcion on* $\Re^d$*, then*

1. *If* F *satisties the regular variation condition*

$$\lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t1)} = W(x) > 0, \qquad x > 0$$

   and $W(cx) = c^{-\alpha}W(x)$*, for any* $c > 0, x > 0$ *and a certain* $\alpha > 0$*, then* $F \in D(G)$ *for*

$$G(x) = \exp\{-W(x)\}, \qquad x > 0\}$$

2. *If* F *has regular variation index* α *and spectral measure* σ *then* $F \in D(G)$ *for* $G(x) = \exp\{-\mu([-\infty,x]^c)\}$, $x > 0$ *where* μ *is the Radon measure described by*

$$\mu\left(\{x \in \Re^d \setminus \{0\}, |\,x\,| > r, x/\,|\,x\,| \in A\}\right) = r^{-\alpha}\sigma(A)$$

1.2.1 Point Processes

We follow the point processes theory developed in Kallenberg (1969). Let E be a locally compact second countable Hausorf (lcscH) space and $\mathcal{E}$ be its Borell sigma–field. Denote by $M(E)$ the set of all measures defined on $\mathcal{E}$ and endow it with the sigma–field $\mathcal{M}$ generated by the evaluation mappings $\{\mu \to \mu(B), B \in \mathcal{E}\}$.
Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we call any measurable mapping

$$\xi\colon (\Omega, \mathcal{F}) \to (M(E), \mathcal{M})$$

a random measure. Denote by $M_+(E)$ the subset of $M(E)$ of counting measures, that is, measures of the form

$$\mu(\,\cdot\,) = \sum_{i=1}^{\infty} \delta_{X_i}(\,\cdot\,)$$

where $X_i \in E$ and δ_{X_i} denotes the unit mass measure at X_i. A random measure μ with values in $M_+(E)$ and such that $\mu(K) < \infty$ for $K \in \mathcal{E}$ compact (μ is Radon) is called a point process.
For any random measure ξ and Borell function $f\colon E \to [0, \infty)$ we write

$$\xi(f) = \int_E f d\xi \qquad \textit{and} \qquad L_\xi(f) = \mathbb{E}\left[\xi(f)\right]$$

The function L_ξ is called the Laplace transform of ξ. In terms of the distribution of ξ, the measure $\mathbb{P}_\xi = \mathbb{P} \circ \xi^{-1}$, it can written as

$$L_\xi(f) = \int_{M(E)} \left(\exp\left\{-\int_E f(x)m(dx)\right\}\right) P_\xi(dm)$$

Remark 4. *Given any set $A \in \mathcal{E}$, we can compute the probability $\mathbb{P}\left[\xi(A) = 0\right]$ by knowing the Laplace transform of ξ alone.*

Proof. Let $A \in \mathcal{E}$ and $f_t(x) = t\mathbb{1}\,(x \in A)$. The Laplace transform of ξ applied to f_t yields

$$L_\xi(f) = \mathbb{E}\left[\exp\{-\xi(f)\}\right] = \mathbb{E}\left[\left\{-\int t\mathbb{1}\,(A)\, d\xi\right\}\right] = \mathbb{E}\left[\exp\{-t\xi(A)\}\right]$$

Observe that as t grows $\exp\{-t\xi(A)\}$ tends to 1 if $\xi(A) = 0$ and to 0 if $\xi(A) > 0$, that is

$$\lim_{t\to\infty} \exp\{-t\xi(A)\} = \mathbb{1}\,(\xi(A) = 0)$$

Therefore, by the bounded convergence theorem we have

$$\lim_{t\to\infty} \mathbb{E}\left[\exp\{-t\xi(A)\}\right] = \mathbb{P}\left[\xi(A) = 0\right]$$

■

The Laplace transform determines uniquely the distribution of the random measure ξ, see Kallenberg (1969) or Resnick (1987).
A particularly important class of point processes is the class of Poisson point processes. Given a Radon measure μ on $\mathcal{E}$, a Poisson point process with intensity measure μ is characterized by having the Laplace transform

$$L(f) = \exp\left\{\int_E 1 - e^{-f(x)}\mu(dx)\right\}$$

We will denote this process as $PRM(d\mu)$ which stands for Poisson Random Measure with intensity measure μ. The most important features of $PRM(d\mu)$ are explained in the next proposition.

Proposition 7. *Let N be a $PRM(d\mu)$ for a given Radon measure μ on $\mathcal{E}$, then*

1. *For any $F \in \mathcal{E}$ and $k \in \mathbb{N}$,*

$$\mathbb{P}\left[N(F) = k\right] = \begin{cases} \exp\{-\mu(F)\}(\mu(F))^k/k!, & \text{if } \mu(F) < \infty, \\ 0, & \text{if } \mu(F) = \infty. \end{cases}$$

2. *For any $k \geq 1$ and mutually disjoint sets $F_1, F_2, \ldots, F_k$ the random variables*

$$N(F_i), 1 \leq i \leq k$$

are independent.

■

In order to relate point processes to the stochastic behaviour of extremes of random variables or vectors, weak convergence is a basic tool. In order to study the weak convergence of point process, we topologize the space $M(E)$ with the vague convergence: Given μ_n random measures, we say μ_n converges vaguely to μ, written as $\mu_n \stackrel{v}{\to} \mu$ if for any non–negative, continuous function f defined on E for which $supp(f)$ is compact we have

$$\lim_{n\to\infty} \int_E f d\mu_n = \int f d\mu \tag{1.18}$$

Remark 5. *The definition above may be mistaken with weak convergence of measures. Nonetheless there's an important difference, namely that for weak convergence the condition 1.18 is imposed on every bounded continuous function f and not only on those which are compactly supported. Since all continuous, compactly supported functions are bounded it follows that weak convergence implies vague convergence.*
For the converse implication the additional hypothesis that the sequence (μ_n) is tight is needed. See Kallenberg (2002) Lemma 4.20.

The integral definition of vague convergence is suitable for many theoretical purposes, however, it is also very usefull to understand the convergence of measures in terms of sets, just as with Portmanteau's Theorem in the case of weak convergence. The next Theorem is the vague–convergence–version of this result.

Theorem 5 (Kallenberg (1969))**.** *Let $\{\mu_n\}_{n\geq 0}$ be a sequence of Radon measures on $(E, \mathcal{E})$. Then, the following statements are equivalent.*

1. $\mu_n \xrightarrow{v} \mu_0$,
2. $\mu_n(B) \to \mu_0(B)$ *for every relatively compact set* $B \in \mathcal{E}$ *with* $\mu_0(\partial B) = 0$,
3. $\limsup_{n\to\infty} \mu_n(F) \leq \mu_0(F)$ *and* $\liminf_{n\to\infty} \mu_n(G) \geq \mu_0(G)$ *for every compact set* $F \in \mathcal{E}$ *and every open, relatively compact set* $G \in \mathcal{E}$.

■

In the special case where the sequence $\{\mu_n\}$ is composed by point processes alone, it is clear that for all $n \geq 0$ there exists a sequence of points in E, say, $\{x_j^{(n)}\}$ such that $\mu_n = \sum_{j=1}^{\infty} \delta(x_j^{(n)})$. Then, as a consequence of the above Theorem, we have the next Corollary.

Corollary 3. *Suppose $\{\mu_n\}_{n\geq 0}$ are point processes such that $\mu_n \xrightarrow{v} \mu_0$. For K compact with $\mu_0(\partial K) = 0$ we have, for $n \geq n(K)$ a labeling of the points of $\mu_n, n \geq 0$ in K such that*

$$\mu_n(\,\cdot \cap K) = \sum_{j=1}^{P} \delta_{x_i^{(n)}}(\,\cdot\,), \qquad n \geq 0$$

and in E^P endowed with the product topology

$$(x_j^{(n)}, 1 \leq i \leq P) \to (x_j^{(0)}, 1 \leq i \leq P), \qquad \text{as } n \to \infty$$

1.3 Characterizing the Maximum Domains of attraction with Point Processes

The relation of weak convergence of sequences of Point Processes to extreme value theory is expressed in the next two results.

Theorem 6. *[Resnick (1987)] For each $n \in \mathbb{N}$, suppose $\{X_{n,j}, j \geq 1\}$ are i.i.d. random elements of $(E, \mathcal{E})$. Define*

$$N_n = \sum_{j=1}^{\infty} \delta(jn^{-1}, X_{n,j})$$

and suppose N is PRM on $[0, \infty) \times E$ with mean measure $dt \times d\mu$. Then $N_n \Rightarrow N$ in $M([0, \infty) \times E)$ if and only if

$$n\mathbb{P}\left[X_{n,1} \in \cdot\,\right] \xrightarrow{v} \mu \qquad \text{on } E \tag{1.19}$$

Proof. First, we prove a simpler result: If N is PRM(μ) on E, then

$$N_n = \sum_{j=1}^{\infty} \delta(X_{n,j}) \Rightarrow N \qquad \text{in } M(E)$$

if and only if (1.19) holds. We show for this propose convergence of Laplace functionals. Given a function $f \in C_K^+(E)$ we have

$$\begin{aligned} L_{N_n}(f) &= \mathbb{E}\left[\exp\left\{-\sum_{j=1}^{\infty} f(X_{n,j})\right\}\right] = \left(\mathbb{E}\left[-f(X_{n,1})\right]\right)^n \\ &= \left(1 - \frac{\int_E (1 - e^{-f(x)}) n\mathbb{P}\left[X_{n,1} \in dx\right]}{n}\right)^n \end{aligned}$$

which converges to the Laplace functional of PRM(μ)

$$\exp\{-\int_E (1 - e^{-f(x)})\mu(dx)$$

if and only if (1.19) holds.
With this same method, we now show the full result of the Proposition. For

$f \in C_K^+([0,\infty)\times E)$ we have

$$L_{N_n}(f) = \mathbb{E}\left[\exp\left\{-\sum_{j=1}^{\infty} f(jn^{-1}, X_{n,j})\right\}\right]$$
$$= \prod_j \left(1 - \int_E (1 - e^{-f(jn^{-1}, X_{n,j})}\mathbb{P}\left[X_{n,1} \in dx\right]\right)$$

Suppose (1.19) holds and define λ_n by

$$\lambda_n(ds, dx) = \sum_j \delta(jn^1) ds \mathbb{P}\left[X_{n,1} \in dx\right]$$

so that

$$\lambda_n(ds, dx) \xrightarrow{v} ds\mu(dx)$$

This implies that

$$\begin{aligned}\sum_j \int_E (1 - e^{-f(jn^{1-}, X_{n,j})})\mathbb{P}\left[X_{n,1} \in dx\right] &= \iint_{[0,\infty)\times E} (1 - e^{-f}) d\lambda_n \\ &\to \iint (1 - e^{-f(s,x)}) ds\mu(dx)\end{aligned} \tag{1.20}$$

Furthermore, since f is compactly supported, then there exists a compact set $A \subseteq E$ such that

$$\sup_j \int_E (1 - e^{-f(jn^{-1}, X_{n,j})})\mathbb{P}\left[X_{n,1} \in dx\right] \le \mathbb{P}\left[X_{n,1} \in A\right] \to 0 \tag{1.21}$$

where the last convergence is a byproduct of (1.19). Using the expansion

$$\ln(1+z) = z(1+\varepsilon(z)), \quad |\,\varepsilon(z)\,| \le |\,z\,|, \text{ if } |\,z\,| \le 1/2$$

we have

$$\left| -\ln L_{N_n}(f) - \sum_j \int_E (1 - e^{-f(jn^{-1}, X_{n,j})})\mathbb{P}\left[X_{n,1} \in dx\right] \right|$$
$$\le \sum_{j=1}^{n} \left(\int_E (1 - e^{-f(jn^{-1}, X_{n,j})})\mathbb{P}\left[X_{n,1} \in dx\right]\right)^2$$

for sufficiently large $n \in \mathbb{N}$, which may be bounded by

$$\leq \left(\sup_j \int_E (1 - e^{-f(jn^{-1}, X_{n,j})}) \mathbb{P}\left[X_{n,1} \in dx\right] \right) \times \sum_{j=1}^{\infty} \int_E (1 - e^{-f(jn^{-1}, X_{n,j})}) \mathbb{P}\left[X_{n,1} \in dx\right] \to 0$$

by (1.20) and (1.21). Therefore, if (1.19) holds, we have that $\ln L_{N_n} \to \ln L_N$ which shows the convergence $L_{N_n} \to L_N$ and thus $N_n \Rightarrow N$.
Conversely, if we know $L_{N_n} \to L_N$ we set $f(s, x) = \mathbb{1}\,(s \in [0, 1])\, g(x)$ where $g \in C_K^+(E)$ and we get

$$\mathbb{E}\left[\exp\left\{-\sum_{j=1}^{n} g(X_{n,j})\right\}\right] \to \exp\left\{-\int_{(} 1 - e^{-g}) d\mu\right\} \tag{1.22}$$

Of course, since f is not an element of $C_K^+([0,\infty) \times E)$ we need to use standard approximation arguments. Now, observe that (1.22) says that $\sum_j \delta(X_{n,j})$ converges weakly to PRM(μ) and so (1.19) holds by the discussion at the beginning of the proof. ■

Corollary 4 (Resnick (1987)). *Let $\{X_n\}$ be a sequence of i.i.d. random variables with distribution function $F \in D(G)$, where G is an EVD. Denote by M_n the sample maxima of the sequence $\{X_n\}$, so there exists norming constants $a_n > 0, b_n$ such that*

$$\mathbb{P}\left[a_n^{-1}(M_n - b_n) \leq x\right] = F^n(a_n x + b_n) \to G(x) \tag{1.23}$$

Suppose, for simplicity, that the norming constants are chosen canonically. Define the point process (in the different spaces mentioned for the different cases listed bellow)

$$N_n = \sum_{j=1}^{\infty} \delta(jn^{-1}, a_n^1(X_j - b_n))$$

1. *If $G = \Lambda$, set $E = (-\infty, \infty]$, $\mu(x, \infty] = e^{-x}$ for $x \in \Re$ then (1.23) is equivalent to*
$$N_n \Rightarrow N = PRM(dt \times d\mu)$$
in $M([0, \infty) \times (-\infty, \infty])$.

2. *If* $G = \phi_\alpha$ *and if* $F(0) = 0$*, then set* $E = (0, \infty]$*,* $\mu(x, \infty] = x^{-\alpha}$ *for* $x > 0$*. Then* (1.23) *is equivalent to*

$$N_n \Rightarrow N = PRM(dt \times d\mu)$$

in $M([0, \infty) \times (0, \infty])$*.*

3. *If* $G = \psi_\alpha$ *so that* $x_F = \sup\{\, x \in \Re \quad s.t. \quad F(x) < 1\,\} < \infty$*, set* $E = (-\infty, \times 0]$ *and* $\mu(x, 0] = (-x)^\alpha$ *for* $x < 0$*. Then* (1.23) *is equivalent to*

$$N_n \Rightarrow N = PRM(dt \times d\mu)$$

in $M([0, \infty) \times (-\infty, 0])$*.*

Proof. It follows directly from (1.23) that

$$n\mathbb{P}\left[a_n^{-1}(X_1 - b_n) > x\right] \to -\ln G(x)$$

for x such that $G(x) > 0$ and this last convergence is equivalent to (1.19) because of the choice of μ and the topology for E in each case. ■

As a consequence to this Corollary and the continuous mapping Theorem we know that given a real number x and the normalizing sequences $a_n > 0, b_n \in \mathbb{R}$, we have

$$\hat{N_n} = N_n([0, \infty) \times (x, \infty]) \Rightarrow N([0, \infty) \times (x, \infty]).$$

The Process $\hat{N_n}$ counts the number of exceedances over the thresholds $u_n = a_n x + b_n$ by the sequence X_n and is deeply studied in Leadbetter et al. (1983). It's relevance in EVT is also discussed in Embrechts et al. (1997) both theoretically and statistically.

The first advantage of the Point Process approach is that the dimension of the space in which the random variables are defined makes no difference, only its topological structure. Also, since it is based on weak convergence, when a result is made corollaries emerge based on arguments involving continuity. This, in turn, results theoretically powerfull but lacks the precision needed to take the results to practice: It doesn't seem to explain neither the necessity of regular variation nor the quality of convergence. Another important feature is that we can use the Point Processes theory in more general cases such as that of stationary sequences.

1.4 Stationary sequences

One way to extend the results of classical Extreme Value Theory is to change the independence assumption to something less restrictive. A natural hypothesis to consider is k–dependence which means X_j is independent of X_i whenever $|\,i-j\,| \geq k$. Another possibility is that the association degree decreases as the magnitude $|\,i-j\,|$ increases. For this, we have three distinct approaches. First, the correlational restriction $\mathrm{Corr}(X_i, X_{i+k}) \leq h(k)$ where $h(k) \to 0$ as $k \to \infty$. Obviously, this kind of restriction is most usefull for elliptical distributions for which correlation and independence characterize each other. Another possibility is the mixing condition of Rosenblatt

$$|\,\mathbb{P}\,[A \cap B] - \mathbb{P}\,[A]\,\mathbb{P}\,[B]\,| \leq h(k),$$

where $A \in \mathcal{F}(X_1, \ldots, X_p),\ B \in \mathcal{F}(X_{p+k+r}, r \geq 1)$ for any p and $h(k) \to 0$ as $k \to \infty$. This condition can be interpreted saying that the past up to time p is "almost independent" of the future time from $p+k+1$ onwards if k is large. The third approach, more general than the other ones, to be briefly accounted here, is the distributional mixing condition.

This condition is used in Leadbetter et al. (1983). It is intuitively based on the observation that for studying the distribution of the sample maxima M_n one studies sets of the form $\{M_n \leq x\} = \bigcap_{i=1}^{n}\{X_i \leq x\}$ and so, a mixing condition is necessary only for this family of events.

Definition 6. *The condition D will be said to hold if for any integers*

$$i_1 < i_2 < \ldots < i_p < i_p + k \leq j_1 < j_2 \cdots < j_q$$

and for any real u we have

$$\Big|\, F_{i_1,i_2,\ldots,i_p,j_1,j_2,\ldots,j_q}(u,u,\ldots,u) - F_{i_1,i_2,\ldots,i_p}(u,\ldots,u)F_{j_1,j_2,\ldots,j_q}(u,\ldots,u)\,\Big| \leq g(k)$$

with $g(k) \to 0$ as $k \to \infty$.

This condition can be weakened to be satisfied by only sequences $\{u_n\}$ of real numbers.

Definition 7. *The condition $D(u_n)$ is said to hold if for any integers*

$$i_1 < i_2 < \ldots < i_p < i_p + k \leq j_1 < j_2 \cdots < j_q \leq n$$

we have

$$\left| F_{i_1,i_2,\ldots,i_p,j_1,j_2,\ldots,j_q}(u_n,u_n,\ldots,u_n) - F_{i_1,i_2,\ldots,i_p}(u_n,\ldots,u_n)F_{j_1,j_2,\ldots,j_q}(u_n,\ldots,u_n) \right| \leq \alpha(n,k)$$

with $\alpha(n,k_n) \to 0$ as $n \to \infty$ for some sequence $k_n = o(n)$.

Observe that $\alpha(n,\ \cdot\ )$ may be taken as a decreasing function. Indeed, $\alpha(n,k)$ may be replaced by

$$\alpha'(n,k) = \max_{i_1<\cdots<i_p<i_p+k\leq j_1<\cdots<j_q\leq n} \big| F_{i_1,i_2,\ldots,i_p,j_1,j_2,\ldots,j_q}(u,u,\ldots,u) - F_{i_1,i_2,\ldots,i_p}(u,\ldots,u)F_{j_1,j_2,\ldots,j_q}(u,\ldots,u) \big|$$

so that $\alpha'(n,k) \leq \alpha(n,k)$ and $\alpha'(n,k_n) \to 0$ as $n \to \infty$ for $k_n = o(n)$. Furthermore $\alpha'(n,k)$ is decreasing in k because the possible choices of indexes with $j_1 - i_p \geq k$ include those with $j_1 - i_p \geq k+1$ for fixed n. This in turn implies there exists a sequence $k_n = o(n)$ such that $\alpha(n,k_n) \to 0$ as $n \to \infty$ if and only if for every $\lambda > 0$ we have $\alpha(n,[n\lambda]) \to 0$. This observations are sometimes convenient.

The role of the condition $D(u_n)$ is to confer a stationary sequence a degree of asymptotic independence so that the known results from the classical extreme value theory still hold. Particularly, $D(u_n)$ will guarantee that any limiting distribution for normalized maxima of a stationary sequence is an extreme value distribution. This is proven by a series of approximations.

In what follows, we denote by $E \subseteq \{1,2,\ldots,n\}$ the maximum value taken by the sequence X_n over E,

$$M(E) = \bigvee_{i\in E}\{X_i\}.$$

Also, any segment of consecutive integers $\{i_1, i_1+1,\ldots,i_2\}$ will be called an *interval* with length $i_2 - i_1 + 1$.

Lemma 1. *Let $\{X_n\}$ be a stationary sequence of random variables. Suppose $D(u_n)$ holds for some sequence $\{u_n\}$. Let n,r,k be fixed integers and $E_1,E_2,\ldots,E_r$ be subintervals of $\{1,\ldots,n\}$ such that for $i \neq j$ the intervals E_i, E_j are separated by at least k. Then*

$$\left| \mathbb{P}\left[\bigcap_{i=1}^{r}\{M(E_i)\leq u_n\}\right] - \Pi_{i=1}^{r}\mathbb{P}\left[M(E_j)\leq u_n\right] \right| \leq (r-1)\alpha(n,k) \tag{1.24}$$

Proof. Write $A_i = \{M(E_i) \le u_n\}$ and $E_i = \{k_i, \ldots, l_i\}$ with $k_1 \le l_1 < k_2 \le \cdots \le l_r$. Then

$$| \mathbb{P}[A_1 \cap A_2] - \mathbb{P}[A_1]\mathbb{P}[A_2] | =$$
$$| F_{k_1,\ldots,l_1,k_2,\ldots,l_2}(u_n,\ldots,u_n) - F_{k_1,\ldots,l_1}(u_n,\ldots,u_n)F_{k_2,\ldots,l_2}(u_n,\ldots,u_n) | \le \alpha(n,k)$$

because $k_2 - l_1 \ge k$ by hypothesis. Analogously,

$$\begin{aligned} | \mathbb{P}[A_1 \cap A_2 \cap A_3] - \mathbb{P}[A_1]\mathbb{P}[A_2]\mathbb{P}[A_3] | &\le | \mathbb{P}[A_1 \cap A_2 \cap A_3] - \mathbb{P}[A_1 \cap A_2]\mathbb{P}[A_3] | \\ &+ | \mathbb{P}[A_1 \cap A_2] - \mathbb{P}[A_1]\mathbb{P}[A_2] | \mathbb{P}[A_3] \\ &\le 2\alpha(n,k) \end{aligned}$$

because $E_1 \cup E_2 \subseteq \{k_1, \ldots, l_2\}$ and $k_3 - l_2 \ge k$. Proceeding in this way we obtain (1.24) as desired. ■

The next two Lemmas will show how condition $D(u_n)$ forces the limit distribution to be an extreme value distribution. Observe that in this development the existence of the limit distribution for the stationary sequence is assumed.

First, some notation. Let k be a fixed integer and $n' = [n/k]$. For large n, let $k < m < n'$, we divide the first $n'k$ integers in $2k$ consecutive intervals of lengths $n - m, m$ alternatively, as follows. Define

$$I_1 = \{1, 2, \ldots, n' - m\}, \quad J_1 = \{n' - m + 1, \ldots, n'\}$$

and similarly $I_2, J_2, \ldots, I_k, J_k$. Finally, let

$$I_{k+1} = \{(k-1)n' + m + 1, \ldots, kn'\} \quad J_{k+1} = \{kn' + 1, \ldots, kn' + m\}$$

Lemma 2. *Let $\{X_n\}$ be a stationary sequence, and assume $D(u_n)$ holds, then, with the notation above*

$$\begin{aligned} \left| \mathbb{P}[M_n \le u_n] - (\mathbb{P}[M_{n'} \le u_n])^k \right| &\le (2k+1)\mathbb{P}[M(I_1) \le u_n < M(J_1)] \\ &+ (k-1)\alpha(n,m) \end{aligned} \tag{1.25}$$

Proof. First, observe that $\{M_n \le u_n\} \subseteq \bigcap_{i=1}^k \{M(I_j) \le u_n\}$ and in their difference we have $\{M(I_j) \le u_n < M(J_j)$ for some $j \le k\}$. Therefore, by stationarity, we have

$$0 \le \mathbb{P}\left[\bigcap_{i=1}^k \{M(I_j) \le u_n\}\right] - \mathbb{P}[M_n \le u_n] \le (k+1)\mathbb{P}[M(I_1) \le u_n < M(J_1)] \tag{1.26}$$

Lemma 1 with I_j for E_j and the fact that $\mathbb{P}\left[M(I_j) \leq u_n\right]$ is independent of j give

$$\left| \mathbb{P}\left[\bigcap_{j=1}^{k}\{M(I_j) \leq u_n\}\right] - \left(\mathbb{P}\left[M(I_1) \leq u_n\right]\right)^k \right| \leq (k-1)\alpha(n,m) \tag{1.27}$$

Finally,

$$0 \leq \mathbb{P}\left[M(I_1) \leq u_n\right] - \mathbb{P}\left[M_{n'} \leq u_n\right] = \mathbb{P}\left[M(I_1) \leq u_n < M(J_1)\right]$$

From the fact that if $0 \leq x \leq y \leq 1$ then $0 \leq y^k - x^k \leq k(y-x)$ applied to $y = \mathbb{P}\left[\bigcap_{j=1}^{k}\{M(I_j) \leq u_n\}\right]$ and $x = \mathbb{P}\left[M(I_1) \leq u_n\right]$ we get

$$\left| \left(\mathbb{P}\left[M(I_1) \leq u_n\right]\right)^k - \left(\mathbb{P}\left[M_{n'} \leq u_n\right]\right)^k \right| \leq k\mathbb{P}\left[M(I_1) \leq u_n < M(J_1).\right] \tag{1.28}$$

Combining (1.26) with (1.27) and (1.28) we get the desired result. ■

Lemma 3. *If $D(u_n)$ holds, $r \geq 1$ is a fixed integer and if $n \geq (2r+1)mk$, then,*

$$\mathbb{P}\left[M(I_1) \leq u_n < M(J_1)\right] \leq \frac{1}{r} + 2r\alpha(n,m). \tag{1.29}$$

Proof. Being $n \geq (2r+1)mk$ we may pick $E_1, E_2, \ldots E_r$ subintervals of $I_1 = \{1, 2 \ldots, n' - m\}$ each with r elements and separated among them and of J_1 by at least m. Then

$$\begin{aligned} \mathbb{P}\left[M(I_1) \leq u_n < M(J_1)\right] &\leq \mathbb{P}\left[\bigcap_{j=1}^{r}\{M(E_j) \leq u_n\}, \{M(J_1) > u_n\}\right] \\ = \mathbb{P}\left[\bigcap_{j=1}^{r}\{M(E_j) \leq u_n\}\right] &- \mathbb{P}\left[\bigcap_{j=1}^{r}\{M(E_j) \leq u_n\}\{M(J_1) \leq u_n\}\right]. \end{aligned} \tag{1.30}$$

By stationarity, the quantity $\mathbb{P}\left[M(E_j) \leq u_n\right] = \mathbb{P}\left[M(J_1) \leq u_n\right]$ is independent of j. Name it p. By Lemma 1, we have

$$\left| \mathbb{P}\left[\bigcap_{j=1}^{r}\{M(E_j) \leq u_n\}\right] - p^r \right| \leq (r-1)\alpha(n,m),$$

$$\left| \mathbb{P}\left[\bigcap_{j=1}^{r}\{M(E_j) \leq u_n \{M(J_1) \leq u_n\}\right] - p^{r+1} \right| \leq r\alpha(n,m).$$

Therefore

$$\mathbb{P}\left[M(I_1) \leq u_n < M(J_1)\right] \leq p^r - p^{r+1} + 2r\alpha(n,m).$$

The bound $p^r - p^{r+1} \leq \frac{1}{r}$ for $0 \leq p \leq 1$ gives

$$\mathbb{P}\left[M(I_1) \leq u_n < M(J_1)\right] \leq \frac{1}{r} + 2r\alpha(n,m),$$

which completes the proof. ■

Let us now assume that there exist a non–degenerate distribution function G and norming constants $a_n > 0,\ b_n$ such that

$$\mathbb{P}\left[a_n^{-1}(M_n - b_n) \leq x\right] \to G(x), \tag{1.31}$$

for the stationary sequence $\{X_n\}$. From the development in section 1.1 we know that G is an extreme value distribution if and only if it is max–stable and from Kintchine's convergence to types theorem this will follow if

$$\mathbb{P}\left[a_{nk}^{-1}(M_n - b_{nk}) \leq x\right] \to G^{1/k}(x) \quad k \geq 1. \tag{1.32}$$

Since the case $k = 1$ is our assumption (1.31), it is enough to prove that if (1.32) is valid for $k = 1$ then it remains valid for $k \geq 2$. This is clearly the case if

$$\mathbb{P}\left[M_{nk} \leq a_{nk}x + b_{nk}\right] - \left(\mathbb{P}\left[M_n \leq a_{nk}x + b_{nk}\right]\right)^k \to 0. \tag{1.33}$$

Lemma 2 gives us a bound on this difference using $n = nk$ and appropriate u_n. What remains is to dominate the righthand side of (1.25) which is achieved via Lemma 3. We thus deduce that

$$\mathbb{P}\left[M_n \leq u_n\right] - \left(\mathbb{P}\left[M_{n'} \leq u_n\right]\right)^k \xrightarrow[n\to\infty]{} 0. \tag{1.34}$$

As a consequence of the foregoing development, we have the next Theorem.

Theorem 7. *Let $\{X_n\}$ be a stationary sequence and $a_n > 0,\ b_n \in \mathbb{R}$ constants such that $\mathbb{P}\left[a_n^{-1}(M_n - b_n)\right] \to G(x)$ for a non–degenerate distribution function G. Suppose $D(u_n)$ for each sequence $u_n = a_n x + b_n$ with $x \in \mathbb{R}$. Then G is an extreme value distribution.*

Proof. Simply use $n = nk$ in place of n in equation (1.34) to get equation (1.32) which implies that G is max–stable and thus an extreme value distribution. ■

As has been shown, the condition $D(u_n)$ plays an important role, namely, if there is a non–degenerate limit distribution G for the maxima of the stationary sequence $\{X_n\}$, then, under $D(u_n)$, for propper sequences u_n, G is an extreme value distribution.

Theorem 8. *Suppose $u_n(\tau)$ is defined for $\tau > 0$ in such a way that $n\bar{F}(u_n) \to \tau$, and $D(u_n(\tau))$ holds for every $\tau > 0$. Then, there exist constants $0 \le \theta \le \theta' \le 1$ such that*

$$\limsup_{n\to\infty} \mathbb{P}\left[M_n \le u_n(\tau)\right] = e^{-\theta\tau},$$
$$\liminf_{n\to\infty} \mathbb{P}\left[M_n \le u_n(\tau)\right] = e^{-\theta'\tau}.$$

Proof. Let $\limsup_{n\to\infty} \mathbb{P}\left[M_n \le u_n(\tau)\right] = h(\tau)$. Observe that $(1-n'\bar{F}(u_n(\tau))) \le \mathbb{P}\left[M_{n'} \le u_n(\tau)\right]$, and so

$$\left(1-\frac{\tau}{k}\right) \le \liminf_{n\to\infty} \mathbb{P}\left[M_{n'} \le u_n(\tau)\right] \le \limsup_{n\to\infty} \mathbb{P}\left[M_{n'} \le u_n(\tau)\right].$$

Taking kth power and using the approximation (1.34) we find that

$$\left(1-\frac{\tau}{k}\right)^k \le \limsup_{n\to\infty} \mathbb{P}\left[M_n \le u_n(\tau)\right],$$

and letting $k \to \infty$ we have

$$h(\tau) \ge e^{-\tau} \qquad \forall \tau > 0.$$

Now, if $\tau' < \tau$ then for n large enough $u_n(\tau') > u_n(\tau)$ and then $h(\tau)$ is a decreasing function of τ. As we have already proven in (1.34), for a fixed integer k and $n' = [n/k]$,

$$\mathbb{P}\left[M_n \le u_n(\tau)\right] - \left(\mathbb{P}\left[M_{n'} \le u_n(\tau)\right]\right)^k \xrightarrow[n\to\infty]{} 0,$$

therefore

$$\limsup_{n\to\infty} \mathbb{P}\left[M_{n'} \le u_n\right] = h^{1/k}(\tau).$$

Also, arguing by cases where $u_n(\tau) \ge u_{n'}(\tau/k)$ and $u_n(\tau) < u_{n'}(\tau/k)$, we see that

$$\begin{aligned}
\left|\,\mathbb{P}\left[M_{n'} \le u_n(\tau)\right] - \mathbb{P}\left[M_{n'} \le u_{n'}(\tau/k)\right]\right| &\le n' \left|\, F(u_{n'}(\tau/k)) - F(u_n(\tau))\,\right| \\
&= n' \left|\, \bar{F}(u_{n'}(\tau/k)) - \bar{F}(u_n(\tau))\,\right| \\
&= n' \left|\, \frac{\tau/k}{n'}(1+o(1)) - \frac{\tau}{n}(1+o(1)).\right|
\end{aligned}$$

Thus $|\,\mathbb{P}\left[M_{n'} \le u_n(\tau)\right] - \mathbb{P}\left[M_{n'} \le u_{n'}(\tau/k)\right]| \to 0$ being $n' \sim n/k$. This in turn implies that

$$\limsup_{n\to\infty} \mathbb{P}\left[M_{n'} \le u_n(\tau)\right] = h(\tau/k),$$

which yields

$$h^{1/k}(\tau) = h(\tau/k), \qquad \forall \tau > 0. \tag{1.35}$$

The only decreasing solution $h > 0$ to (1.4) and (1.35) is known to be $h(\tau) = e^{\theta\tau}$ for some $\theta \ge 0$. Since, as in (1.4), $h(\tau) \ge e^{-\tau}$ it follows that $0 \le \theta \le 1$.

Similarly, $\liminf_{n\to\infty} \mathbb{P}\left[M_n \le u_n(\tau)\right] = e^{-\theta'\tau}$ for some $0 \le \theta' \le 1$. It is evident that $\theta' \ge \theta$ completing the proof. ■

A simple but remarkable consequence of this Theorem is that if for a given $\tau > 0$ the sequence $\mathbb{P}\left[M_n \le u_n(\tau)\right]$ converges then $\theta = \theta'$ and therefore the sequence converges for every $\tau > 0$ to $e^{-\theta\tau}$.

Definition 8. *A stationary sequence of random variables $\{X_n\}$ is said to have extremal index $0 \le \theta \le 1$ if for every $\tau > 0$*

1. *There exists a sequence $u_n(\tau)$ such that $n\bar{F}(u_n(\tau)) \to \tau$, and*
2. *$\mathbb{P}\left[M_n \le u_n(\tau)\right] \to e^{-\theta\tau}$.*

Theorem 9. *Let $\{X_n\}$ be a stationary sequence with extremal index $\theta > 0$. Let $\{\hat{X}_n\}$ be a sequence of i.i.d. random variables with the same distribution F as $\{X_n\}$ and $\hat{M}_n$ its sample maxima. Then, M_n has a non–degenerate limit distribution if and only if $\hat{M}_n$ does and the limiting distributions are of the same type. The same norming constants may be used.*

Proof. First we will prove that for $0 \le \varrho \le 1$ the limit relations

$$\mathbb{P}\left[\hat{M}_n \le v_n\right] \to \varrho, \tag{1.36}$$

$$\mathbb{P}\left[M_n \le v_n\right] \to \varrho^{\theta}, \tag{1.37}$$

are equivalent. Suppose (1.36) true. First assume $0 < \varrho \le 1$ and write $\varrho = e^{-\tau}$ for some $\tau \in [0,\infty)$. Take $\tau' > \tau$ so that $e^{-\tau'} < e^{-\tau} = \varrho$. Condition 1 in the definition of extremal index tells us there exist a sequence $u_n(\tau')$ such that $n\bar{F}(u_n(\tau')) \to \tau'$ and also $\mathbb{P}\left[M_n \le u_n(\tau')\right] \to e^{-\tau'\theta}$. We then have

$$\mathbb{P}\left[\hat{M}_n \le u_n(\tau')\right] \to e^{-\tau}, \qquad \mathbb{P}\left[\hat{M}_n \le v_n\right] \to e^{-\tau} > e^{-\tau'}.$$

So, for sufficiently large n we have $v_n > u_n(\tau')$, giving

$$\liminf_{n\to\infty} \mathbb{P}\left[M_n \le v_n\right] \ge \lim_{n\to\infty} \mathbb{P}\left[M_n \le u_n(\tau')\right] = e^{-\tau'\theta}.$$

Since this is true for all $\tau' > \tau$, we have

$$\liminf_{n\to\infty} \mathbb{P}\left[M_n \le v_n\right] \ge \varrho^{\theta}.$$

For $\varrho = 1$ we get, in particular, $\mathbb{P}\left[M_n \le v_n\right] \to 1 = \varrho^{\theta}$. Now, for $0 < \varrho < 1$ it is similarly shown – by taking $e^{-\tau'} > e^{-\tau} = \varrho$ –, that

$$\limsup_{n\to\infty} \mathbb{P}\left[M_n \le v_n\right] \le \varrho^{\theta}.$$

This two relations put together are exactly (1.37). The proof that (1.37) implies (1.36) is analogous.

Now, assume that the sequence $\{\hat{M}_n\}$ has a non–degenerate limiting distribution G, that is

$$\mathbb{P}\left[a_n^{-1}(\hat{M}_n - b_n) \le x\right] \xrightarrow[n\to\infty]{} G(x), \qquad \forall x \in \mathbb{R}.$$

Then, the preceding with $v_n = a_n x + b_n$ shows that

$$\mathbb{P}\left[a_n^{-1}(M_n - b_n) \le x\right] \xrightarrow[n\to\infty]{} G^{\theta}(x), \qquad \forall x \in \mathbb{R}.$$

Since G is an extreme value distribution, it is max–stable and therefore G^{θ} and G are of the same type. To see the converse, assume $\{M_n\}$ has a non–degenerate limiting distribution H so that

$$\mathbb{P}\left[A_n^{-1}(M_n - B_n) \le x\right] \xrightarrow[n\to\infty]{w} H(x),$$

then, as before,

$$\mathbb{P}\left[A_n^{-1}(\hat{M}_n - B_n) \le x\right] \xrightarrow[n\to\infty]{w} H^{1/\theta}(x).$$

This means $H^{1/\theta}$ is an extreme value distribution and therefore max–stable. Again, this is enough to ensure both limiting distributions are of the same type. ■

There is one important aspect that condition $D(u_n)$ doesn't deal with making it unable to guarantee the existence of the limiting distribution for M_n. It's clustering.

In section 1.3 it was explained that in the independent case, the convergence $\mathbb{P}\left[a_n^{-1}(M_n - b_n) \leq x\right] \to G(x)$ is equivalent to weak convergence of the process

$$\sum_{j=1}^{\infty} \delta_{(j/n, a_n^{-1}(X_j - b_n))}(\,\cdot\,)$$

to a $PRM(dt \times d\mu)$ where μ is determined by G. It is clear from the very definition of Poisson Process that there is no clustering. We therefore need further conditions to extend the classical results concerning the Poisson Measures to the stationary case.

Definition 9. *The condition $D'(u_n)$ is said to hold for the stationary sequence $\{X_n\}$ and $\{u_n\}$ if*

$$\lim_{k\to\infty}\left(\limsup_{n\to\infty}\left(n\sum_{j=2}^{[n/k]}\mathbb{P}\left[X_1 > u_n, X_j > u_n\right]\right)\right) = 0 \tag{1.38}$$

Observe that under $D'(u_n)$

$$\mathbb{E}\left[\sum_{1\leq i<j\leq[n/k]} \mathbb{1}\left(X_i > u_n, X_j > u_n\right)\right] \leq n\sum_{j=1}^{[n/k]}\mathbb{E}\left[\mathbb{1}\left(X_i > u_n, X_j > u_n\right)\right] \to 0$$

so $D'(u_n)$ is a way to avoid joint exceedances over u_n among $X_1, \ldots, X_{[n/k]}$.

In Leadbetter et al. (1983) it is proven that under $D'(u_n)$ the extreme law for the normalized maxima exists so that under both $D(u_n)$ and $D'(u_n)$ the extremes of stationary sequences are just like those of i.i.d. sequences.

However, the condition $D'(u_n)$ is mainly thought to deal with the so–called point process of exceedances, formally defined for $x \in \mathbb{R}$ as

$$\sum_{j=1}^{\infty} \delta_{(j/n, a_n^{-1}(X_j - b_n))}([0,\infty) \times (u_n(x), \infty))$$

for $u_n(x) = a_n x + b_n$ with a_n, b_n the usual norming constants.

A more general approach was taken in Davis and Hsing where the weak convergence of

$$\sum_{j=1}^{\infty} \delta(j/n, a_n^{-1}(X_j - b_n))$$

is shown under a proper mixing condition for strictly stationary sequences of regularly varying random variables.

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables equal in distribution to a given random variable X. Assume that

$$\mathbb{P}\left[\mid X \mid > x\right] = x^{-\alpha} L(x) \tag{1.39}$$

for L a slowly varying function. Let us work only with the random part of the process, that is

$$\sum_{j=1}^{\infty} \delta(a_n^{-1}(X_j))$$

where $b_n = 0$ because of the regular variation condition.

Definition 10. *Given a strictly stationary sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ whose distribution satisfies* (1.39) *and the sequence of real numbers $\{a_n\}_{n\in\mathbb{N}}$ defined by*

$$n\mathbb{P}\left[\mid X_1 \mid > a_n\right] \to 1 \tag{1.40}$$

we say condition $\mathcal{A}(a_n)$ holds for $\{X_n\}_{n\in\mathbb{N}}$ if there exists a sequence of positive integers $\{r_n\}$ such that $r_n \to \infty$, $r_n/n \to 0$ as $n \to \infty$ and

$$\mathbb{E}\left[\exp\left\{-\sum_{j=1}^{n} f(a_n^{-1}X_j)\right\}\right] - \left(\mathbb{E}\left[\exp\left\{-\sum_{j=1}^{r_n} f(a_n^{-1}X_j)\right\}\right]\right)^{[n/r_n]} \to 0 \tag{1.41}$$

for all f step function with bounded support on $\mathbb{R} \setminus \{0\}$.

Observe that condition $\mathcal{A}(a_n)$ is closely related to the Laplace transforms of the processes

$$\hat{N}_n = \sum_{i=1}^{[n/r_n]} N_{r_n,i},$$

where each $N_{r_n,i}$ is independent of the rest and with the same distribution as

$$\sum_{i=1}^{r_n} \delta(a_n^{-1}X_j).$$

In fact, condition $\mathcal{A}(a_n)$ implies that N_n converges weakly if and only if $\hat{N}_n$ does and the limits coincide.

The first result given in Davis and Hsing is the description of all possible weak limits for N_n via the Laplace transform. Before stating the result, let us define for $y \geq 0$

$$M_y = \{\mu\colon \mu([-y,y]^c) > 0, \text{ and } \mu([-x,x]^c) = 0, \text{ for some} 0 < x = x_\mu < \infty\}.$$

For $\mu \in M_0$ with points $a_\mu^{(i)}$ let $\mu_+ = \max(0, \max(a_\mu^{(i)}))$ and $\mu_- = \max(0, \min(a_\mu^{(i)}))$ and $x_\mu = \max(\mu_+, \mu_-)$. Define the mapping

$$T\colon \mu \to (x_\mu, \mu(x_\mu \cdot\,))$$

Observe that T defines a continuous mapping with range $(0,\infty)\times\hat{M}$ where $\hat{M}$ is the set of measures with support contained in $[-1,1]$ and $\mu(\{-1\}\cup\{1\}) > 0$.

Theorem 10 (Davis and Hsing). *Assume condition $\mathcal{A}(a_n)$ for the stationary sequence $\{X_n\}$ of random variables and $N_n \xrightarrow{d} N$ for some not null random measure N. Then N is infinitely divisible with canonical measure λ such that $\lambda(M_0^c) = 0$ and $\lambda \circ T^{-1} = \vartheta \times Q$ where Q is a probability measure on $(\hat{M}, \mathcal{B}(\hat{M}))$ and*

$$\vartheta(dy) = \gamma\alpha y^{-\alpha-1} \mathbb{1}\,(y > 0)\, dy,$$

with $\gamma \in (0,1]$ defined by

$$\gamma = \lambda\left(\{\mu\colon \mu([-1,1]^c) > 0\}\right).$$

In this case the Laplace transform of N may be written for measurable f as

$$L_N(f) = \exp\left\{-\int_0^\infty \int_{\hat{M}} (1 - \exp(-\mu f(y\,\cdot\,)))Q(d\mu)\nu(dy)\right\}$$

Proof. For fixed $A \in \mathcal{B}(\hat{M})$, define the measure

$$Q_A(E) = \lambda \circ T^{-1}(E \times A) \qquad E \in \mathcal{B}((0,\infty)).$$

As shown in Lemma 2.2 of Davis and Hsing, for any given $u > 0$ and the mapping $\pi_u\colon \mu \to \mu(u^{-1}\,\cdot\,)$, the measure λ satisfies the relation $\lambda = u^\alpha \lambda \circ \pi_u$, so that for every $E \in \mathcal{B}((0,\infty))$ we have

$$Q_A(uE) = \lambda\circ T^{-1}(uE\times A) = \lambda\circ\pi_u\circ T^{-1}(E\times A) = u^{-\alpha}\lambda\circ T^{-1}(E\times A) = u^{-\alpha}Q_A(E).$$

It now follows that the measure Q_A admits a density with respect to Lebesgue measure for fixed $A \in \mathcal{B}(\hat{M})$ and

$$Q_A(dx) = \alpha x^{-\alpha-1} Q_A((1,\infty)) \mathbb{1}\,(x > 0)\,.$$

Since $\gamma \neq 0$ by Lemma 2.2 in Davis and Hsing, so we may as well write

$$Q_A(dx) = \gamma \alpha x^{-\alpha-1} \left(\frac{Q((1,\infty))}{\gamma} \right) \mathbb{1}\,(x > 0)\,,$$

so that for $Q(A) = Q((1,\infty))/\gamma$, we see that Q_A can be written as $Q_A(E) = Q(A) \times \nu$ and $Q(A)$ depends only on A. Regarding Q as a measure on $\hat{M}$ we obtain the desired decomposition

$$\lambda \circ T^{-1} = \nu \times Q.$$

Observe further that

$$\gamma = \lambda\left(\{\mu\colon \mu([-1,1]^c) > 0\}\right) = -\log(\mathbb{P}\left[N([-1,1])^c = 0\right])$$

is the extremal index of $|\,X_n\,|$. Indeed,

$$\begin{aligned} \mathbb{P}\left[a_n^{-1} \max_{1\leq j\leq n} |\,X_j\,| \leq x\right] =& \mathbb{P}\left[N_n([-x,x]^c) = 0\right] \\ &\to \exp\left\{-\lambda\left(\{\mu\colon \mu([-x,x]^c) > 0\}\right)\right\}. \end{aligned}$$

By the scaling property of λ we obtain

$$\mathbb{P}\left[a_n^{-1} \max_{1\leq j\leq n} |\,X_j\,| \leq x\right] \to \exp\left\{-\gamma x^{-\alpha}\right\}, \qquad x > 0$$

which implies that γ is the extremal index of $|\,X_j\,|$ as stated. We now conclude that $\gamma \in (0,1]$. Therefore

$$\gamma = \lambda\left(\{\mu\colon \mu([-1,1]^c) > 0\}\right) = \lambda \circ T^{-1}((1,\infty) \times \hat{M}) = \gamma Q(\hat{M})$$

which shows that Q is indeed a probability measure and ends the proof. ■

With this characterization we understand the limit point process N as a compound Poisson with intensity measure ϑ and spectral component distributed according to Q, more explicitly

Corollary 5 (Davis and Hsing)**.** *$N = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\delta(P_i Q_{i,j})$ where $\sum_{i=1}^{\infty}\delta(P_i)$ is Poisson with intensity measure ϑ and $\sum_{j=1}^{\infty} Q_{i,j}, i \geq 1$ are mutually independent point processes identically distributed according to Q.*

The next result gives a neater description of the components of the limit process N.

Theorem 11 (Davis and Hsing)**.** *Under the condition $\mathcal{A}(a_n)$ for $\{X_n\}$, the following are equivalent*

*(**1**) N_n converges weakly to some not–null N,*

*(**2**) For some finite, positive constant γ, and $k_n = [n/r_n]$,*

$$k_n \mathbb{P}\left[\bigvee_{j=1}^{r_n} \mid X_j \mid > a_n x\right] \rightarrow \gamma x^{-\alpha}, x > 0,$$

and for some probability measure Q on $\hat{M}$,

$$\mathbb{P}\left[\sum_{j=1}^{r_n}\delta\left(X_j / \bigvee_{j=1}^{r_n} \mid X_j \mid\right) \in \cdot \,\middle|\, \bigvee_{j=1}^{r_n} \mid X_j \mid > a_n x\right] \xrightarrow{w} Q,\ x > 0$$

In this case N is infinitely divisible with canonical measure λ confined to M_0 and such that

$$\lambda \circ T^{-1} = \nu \times Q$$

with $\nu(dy) = \gamma\alpha y^{-\alpha-1}dy$.

Proof. First, we assume (1) holds and that the canonical measure λ admits the representation $\lambda \circ T^{-1} = \nu \times Q$. Then, since N has no fixed atoms (Lemma 2.2 in Davis and Hsing), we have that for any $x > 0$

$$N_n([-x, x]^c) \xrightarrow{d} N([-x, x]^c)$$

Due to condition $\mathcal{A}(a_n)$ N_n may be replaced by $\hat{N}_n$. Thus, letting $k_n^{-1}\lambda_n$ be the common distribution of the processes $N_{r_n,i}$ we have

$$\begin{aligned}
k_n \mathbb{P}\left[\bigvee_{j=1}^{r_n} \mid X_j \mid > a_n x\right] &= \lambda_n(M_x) \\
&= -\log\left(\mathbb{P}\left[\hat{N}_n([-x, x]^c) = 0\right]\right) + o(1) \\
&\rightarrow -\log\left(\mathbb{P}\left[N([-x, x]^c) = 0\right]\right) = \lambda(M_x) = \lambda \circ T^{-1}((x, \infty) \times \hat{M}) \\
&= \vartheta(x, \infty) = \gamma x^{-\alpha}
\end{aligned}$$

The equality $\lambda_n(M_x) = -\log\left(\mathbb{P}\left[\hat{N}_n([-x,x]^c) = 0\right]\right) + o(1)$ is worth a few lines. Observe that, since $\hat{N}_n = \sum_{i=1}^{k_n} N_{r_n,i}$ with $N_{r_n,i}$ i.i.d.,

$$L_{\hat{N}_n}(f) = \left(L_{N_{r_n,1}}(f)\right)^{k_n}$$

From this we get

$$\begin{aligned}\mathbb{P}\left[\hat{N}_n(A) = 0\right] &= \mathbb{P}\left[N_{r_n,1}(A) = 0\right]^{k_n} = \left(1 - \mathbb{P}\left[N_{r_n,1}(A) > 0\right]\right)^{k_n} \\ &= \left(1 - \frac{\lambda_n(M_x)}{k_n}\right)^{k_n} = \exp\{-\lambda_n(M_x)\} + o(1)\end{aligned}$$

We have proven the first part in (2). For the second part, define the probability measures $P_{n,x}$ and P_x on M_0 by

$$P_{n,x} = \frac{\lambda_n(\,\cdot\, \cap M_x)}{\lambda_n(M_x)} \qquad P_x = \frac{\lambda(\,\cdot\, \cap M_x)}{\lambda(M_x)}$$

It can be shown that $P_{n,x} \xrightarrow{w} P_x$ as $n \to \infty$, see Davis and Hsing. Since the mapping T is continuous

$$P_{n,x} \circ T^{-1} \xrightarrow{w} P_x \circ T^{-1}$$

on $(0,\infty) \times \hat{M}$. Using the marginal convergence implied by the bivariate convergence and the facts that $\lambda(M_x) = \nu((x,\infty))$, and $T^{-1}((x,\infty) \times \hat{M}) = M_x$, we conclude that

$$\begin{aligned}P_{n,x} \circ T^{-1}((x,\infty) \times \,\cdot\,) &\xrightarrow{w} P_x \circ T^{-1}((x,\infty) \times \,\cdot\,) \\ &= \frac{\lambda \circ T^{-1}((x,\infty)\times)\,\cdot}{\gamma x^{-\alpha}} = Q(\,\cdot\,)\end{aligned}$$

on $\hat{M}$ which completes the proof (2). For the converse similar arguments may be used to prove that for all measurable f we have

$$\int (1 - e^{-\mu f})\lambda_n(d\mu) \to \int (1 - e^{-\mu f})\lambda(d\mu)$$

which is (1) expressed in terms of the corresponding Laplace transforms. See Davis and Hsing for further details.

■

One last result to add is a necessary condition for the existence of the limit N of the point processes N_n. See Davis and Hsing for the proof.

Theorem 12 (Davis and Hsing). *Suppose that $\{X_j\}$ is a stationary sequence of random variables for which all finite–dimensional distributions are jointly regularly varying with index $\alpha > 0$ and denote by $\eta_m = (\eta_m^{(i)}, |\,i\,| \leq m)$ the random vector corresponding to the spectral measure associated to the distribution of $\{X_i, |\,i\,| \leq m\}$. Assume condition $\mathcal{A}(a_n)$ and that*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\left[\bigvee_{m\leq |\,i\,|\leq r_n} |\,X_j\,| > ta_n \,\middle|\, |\,X_0\,| > ta_n \right] = 0, \quad t > 0 \tag{1.42}$$

Then the limit

$$\gamma = \lim_{m\to\infty} \frac{\mathbb{E}\left[\left|\,\eta_m^{(0)}\,\right|^\alpha - \bigvee_{j=1}^m \left|\,\eta_m^{(j)}\,\right|^\alpha \right]_+}{\mathbb{E}\left[\left|\,\eta_m^{(0)}\,\right|^\alpha \right]}$$

exists. If $\gamma = 0$ then $N_n \xrightarrow{d} 0$, but if $\gamma > 0$ then N_n converges to not–null N with canonical measure λ such that $\lambda \circ T^{-1} = \vartheta \times Q$ as described earlier and Q is the weak limit of

$$\frac{\mathbb{E}\left[\left(\left|\,\eta_m^{(0)}\,\right|^\alpha - \bigvee_{j=1}^m \left|\,\eta_m^{(j)}\,\right|^\alpha \right)_+ \mathbb{1}\left(\sum_{|\,i\,|\leq m} \delta(\eta_m^{(i)}) \in \;\cdot\; \right) \right]}{\mathbb{E}\left[\left|\,\eta_m^{(0)}\,\right|^\alpha - \bigvee_{j=1}^m \left|\,\eta_m^{(j)}\,\right|^\alpha \right]_+}$$

as $m \to \infty$ which exists.

The theory presented so far applies for stationary sequences of real random variables. However, the very same concepts can be taken to the multivariate setting without essential modifications in the proofs. We reffer the reader to Davis and Hsing, Hüsler (1990) and to Davis and Mikosch (1998) where the results are presented.

The results presented so far are applied to establish the presence of the stylized properties in two multivariate GARCH models in chapters 2 and 3. It is worth mentioning that in both cases the mixing properties of the process play a determining role. While sufficient conditions for the CCC-GARCH process to be β–mixing have been provided in Boussama (1998), this property, together with stationarity, is proven for the Cholesky factor GARCH in Chapter 3.

The one concept that changes radically with multidimensionality is the extremal index. In the multivariate setting the condition $D(u_n(x))$, for vector sequences[1] $u_n(x) = a_n \odot x + b_n$, doesn't imply the existence of a **constant** $\theta \in [0,1]$ such that

$$G(x) = \hat{G}(x)^{\theta} \tag{1.43}$$

for $x \in \mathbb{R}^d$, where G is the EVD associated to the independent sequence and $\hat{G}$ is the limit of $n(1 - F(u_n(x)))$ as $n \to \infty$. In fact, if we assume equation (1.43) holds, then we get that for any fixed $1 \leq i_0 \leq d$, by taking limit as $x^{(i)} \to \infty$ for $i \neq i_0$,

$$G_{i_0}(x^{(i_0)}) = \hat{G}_{i_0}^{\theta}(x^{(i_0)}) \tag{1.44}$$

so that every marginal distribution can be expressed with the same extremal index θ. Compare (1.44) with Definition 8. Nonetheless, examples can be constructed where this is not the case. See Hsing (1989) or Hüsler (1990).

In Nandagopalan (1994), the author proposes the multivariate extremal index as a function $\theta(x)$ with values on $(0,1]$ –by avoiding the case $\theta(x) = 0$ we clear out the degenerate limits– via the relation

$$G(x) = \hat{G}^{\theta(x)}(x) \tag{1.45}$$

where G and $\hat{G}$ are defined as before. In Nandagopalan (1994) the extremal index function $\theta(x)$ is explained with dependence functions or copulas, but we find it convenient to do it otherwise.

Following equation (1.45) and using the fact that we may normalize the marginal distributions to be Frechét, see Proposition 2, and that the corresponding distributions G_* and $\hat{G}_*$ are max–stable as implied by Propositions 1 and 2 [2], we may write equation (1.45) in terms of the extremal measures μ_* and $\hat{\mu_*}$ as

$$\mu_*([0,x]^c) = \theta(x)\hat{\mu_*}([0,x]^c), \qquad \forall x > 0 \tag{1.46}$$

With this definition, it is simple to calculate for fixed $1 \leq j \leq d$,

$$\theta_j = \lim_{x^{(i)} \to \infty, i \neq j} \theta(x) = \lim_{x^{(i)} \to \infty, i \neq j} \frac{-\log(G(x))}{-\log(\hat{G}(x))} = \frac{-\log(G_j(x^{(j)}))}{-\log(\hat{G}_j(x^{(j)}))} \tag{1.47}$$

[1]For vectors z, w in $\mathbb{R}^d$, z$\odot$w represents the direct product vector $(z^{(i)} \cdot w^{(i)}, 1 \leq i \leq d)$

[2]The proof of Proposition 1 tells us that if the marginal distributions converge to EVD's and the joint distribution converges to, say, G, then G is a MEVD. This remains valid for stationary sequences under condition $D(u_n)$.

The value θ_j exists under $D(u_n)$ and is the extremal index for the sequence $\{X_n^{(j)}\}$, so that the two concepts are indeed related.

Because of the homogeneous nature of μ_* and $\hat{\mu_*}$ shown in (1.9), we see that

$$\begin{aligned}\mu_*([0,x]^c) &= u^{-1}\mu_*(u[0,x]^c) = u^{-1}\mu_*([0,ux]^c)\\ &= u^{-1}\theta(ux)\hat{\mu_*}([0,ux]^c) = u^{-1}\theta(ux)u\hat{\mu_*}([0,x]^c)\end{aligned} \tag{1.48}$$

which shows that for any $u > 0$ and $x > 0$ we have

$$\theta(ux) = \theta(x)$$

This suggests that we may as well define the extremal index for $x \in S_{d-1}$ only and write (1.45) as

$$G(x) = \hat{G}^{\theta(x/|\,x\,|)}(x), \qquad \forall x > 0$$

Now, putting this together with (1.47) and letting $\{e_i, 1 \le i \le d\}$ be the canonical basis of $\mathbb{R}^d$ we see that for every fixed $1 \le j \le d$ we have

$$\theta(e_j) = \theta_j$$

so that on the axis the values of the function θ are the individual extremal indexes corresponding to the components of the vector X.

1.5 Autocovariance Function Convergence

Given a mean zero time series $\{X_n\}$ in $\mathbb{R}$, the autocovariance function is defined by

$$\gamma_X(h) = \mathbb{E}\left[X_0 X_h\right].$$

As a natural approximation to this function, we have the sample autocovariance function, which is given by

$$\gamma_{n,X}(h) = \frac{1}{n}\sum_{t=1}^{n-h} X_t X_{t+h}.$$

Assuming that the sequence $\{X_n\}$ is in $\mathcal{L}^2$, stationary and ergodic, the Ergodic Theorem of Von–Newman and Birkhoff (See Kallenberg (1969)) tells us that

$$\gamma_{n,X}(h) \xrightarrow[n\to\infty]{c.s.} \gamma_X(h), \qquad h \ge 0$$

Given this convergence, a Central Limit Theorem is expected. This result, however needs some extra hypotheses. In Doukhan (1994) it is shown that if we assume further that the sequence $\{X_n\}$ is strongly mixing and in $\mathcal{L}^4$, then

$$\lim_{n\to\infty} n^{-1/2}\left(\gamma_{n,X}(h) - \gamma_X(h)\right) \stackrel{d}{=} Y$$

where Y is a centered Normal random variable and $\stackrel{d}{=}$ means limit in distribution.

There are some interesting questions regarding the convergence of the Sample Autocovariance Function when the sequence $\{X_n\}$ is regularly varying. As is detailed in Bingham et al. (1989), Embrechts et al. (1997) and Resnick (1987), among others, a regularly varying random variable with index $\alpha > 0$ satisfies

$$\mathbb{E}\left[\mid X\mid^\beta\right] < \infty, \qquad \text{for } \beta < \alpha,$$

$$\mathbb{E}\left[\mid X\mid^\beta\right] = \infty, \qquad \text{for } \beta > \alpha.$$

Therefore, if $\alpha < 2$ the sequence is not in $\mathcal{L}^2$ so that the function $\gamma_X(h)$ is not defined. In this case the almost sure limit of $\gamma_{n,X}(h)$ provides no information. If $2 < \alpha < 4$ then the sequence is in $\mathcal{L}^2$ so that the Ergodic Theorem applies and $\gamma_{n,X}(h) \xrightarrow[n\to\infty]{a.s.} \gamma_X(h)$. However, the CLT result in Doukhan (1994) will not apply because the sequence $\{X_n\}$ is not in $\mathcal{L}^4$.

An explanation of the asymptotical behavior of the sample autocovariance function in these special cases was given in Davis and Mikosch (1998) for stationary sequences with regularly varying finite dimensional distributions. The notation in the following Theorem is the same as in section 1.4.

Theorem 13 (Theorem 3.5 in Davis and Mikosch (1998))**.** *Assume that $\{X_n\}$ is a strictly stationary sequence with regularly varying finite–dimensional distributions of index $\alpha > 0$ and such that*

$$N_n \xrightarrow[n\to\infty]{d} N$$

where N_n and N are the point processes described in section 1.4.

1. *If $\alpha \in (0,2)$, then*

$$(n^{1-2/\alpha}(\gamma_{n,X}(h)))_{h=0,1,\ldots,m} \xrightarrow{d} (V_h)_{h=0,1,\ldots,m} \qquad (1.49)$$

   where $(V_0, V_1, \ldots, V_m)$ is jointly $\alpha/2$ stable in $\mathbb{R}^m$.

2. *If* $\alpha \in (2,4)$, *then*

$$(n^{1-2/\alpha}(\gamma_{n,X}(h) - \gamma_X(h)))_{h=0,1,\dots,m} \xrightarrow{d} (V_h)_{h=0,1,\dots,m} \tag{1.50}$$

were $(V_0, V_1, \dots, V_m)$ *is jointly* $\alpha/2$ *stable in* $\mathbb{R}^m$.

The convergence of the autocovariance function is important for statistical analysis, particularly in time series. As is shown in Chapters 2 and 3 both, the CCC GARCH and the Cholesky factor GARCH, exhibit the behavior explained in the last Theorem as a consequence of regular variation.

1.6 Stochastic Recurrence Equations

Consider an $\mathbb{R}^d$ valued process $\{X_n\}_{n\in\mathbb{Z}}$ given as the solution to the stochastic recurrence equation

$$X_n = A_n X_{n-1} + B_n, \qquad n \in \mathbb{Z} \tag{1.51}$$

for some i.i.d. sequence $\Psi = (A_n, B_n)$ of random $d \times d$ matrices and $d \times 1$ vectors B_n. Let $|\cdot|$ denote a norm in $\mathbb{R}^d$ and $\|\cdot\|$ the corresponding operator norm in the matrix space given by

$$\|A\| = \sup_{|x|=1} |Ax|$$

Also, define the relation "$>$" componentwise for vectors and entrywise for matrices so that, for example $x > 0$ for a given vector $x \in \mathbb{R}^d$ means $x^{(i)} > 0$ for all $1 \le i \le d$ and $A > 0$ for a given $d \times d$ matrix means $A(i,j) > 0$ for all $1 \le i, j \le d$.
By recursively applying equation (1.51) we see that the process $X_n = X_n(\Psi)$ may be written as

$$X_n(X_0, \Psi) = \sum_{j=0}^{n-1} \left(\prod_{i=n-j}^{n-1} A_i \right) B_{n-j-1} + \left(\prod_{i=0}^{n-1} A_i \right) X_0$$

In Brandt (1986) conditions for the existence of a unique casual stationary solution to (1.51) are given.

Theorem 14 (Brandt (1986))**.** *If the sequence* $\Psi = \{(A_n, B_n)\}$ *is stationary and ergodic and*

$$\mathbb{E}\left[\ln \|A_0\|\right] < 0 \quad \textit{and} \quad \mathbb{E}\left[\ln^+ |\, B_0 \,|\right] < \infty \tag{1.52}$$

then

$$X_n(\Psi) = \sum_{j=0}^{\infty} \left(\prod_{i=n-j}^{n-1} A_i \right) B_{n-j-1}, \qquad n \in \mathbb{N} \tag{1.53}$$

is the only proper stationary solution of (1.51) *for the given* Ψ *(where we set* $\prod_{i=n}^{n-1} A_i = 1$*). The sum in equation* (1.53) *converges absolutely almost surely. Furthermore,*

$$\mathbb{P}\left[\lim_{n\to\infty} |\, X_n(Y, \Psi) - X_n(\Psi) \,| = 0\right] = 1 \tag{1.54}$$

for arbitrary random variables Y *on the same probability space as* Ψ*. In particular,*

$$X_n(X_0, \Psi) \xrightarrow[n\to\infty]{d} X_0(\Psi) \tag{1.55}$$

Proof. First, we prove the absolute convergence of (1.53). By assumptions (1.52) and the strong law of large numbers for stationary and ergodic sequences we have

$$\limsup_{k\to\infty} \frac{1}{k} \left(\sum_{i=-1}^{-k} \ln \|A_{n+i}\| + \ln |\, B_{n-k-1} \,| \right) < 0, \quad \text{a.s.}$$

which means that

$$\limsup_{k\to\infty} \log(\|A_{n-1}\| \, \|A_{n-2}\| \ldots \|A_{n-k}\| \, |\, B_{n-k-1} \,|)^{1/k} < 1, \quad \text{a.s.}$$

From this, almost sure absolute convergence of (1.53) follows by the Cauchy's root criterion.

We now prove (1.54). By definition

$$\begin{aligned} |\, X_n(Y, \Psi) - X_n(\Psi) \,| &= \left| -\sum_{j=n}^{\infty} \left(\prod_{i=n-j}^{n-1} A_i \right) B_{n-j-1} + \left(\prod_{i=0}^{n-1} A_i \right) Y \right| \\ &\le \left(\prod_{i=0}^{n-1} \|A_i\| \right) (|\, X_0(\Psi) \,| + |\, Y \,|) \end{aligned} \tag{1.56}$$

Again, by the strong law of large numbers, we get from (1.52)

$$\prod_{i=0}^{n-1} \|A_i\| = \left(\exp\left(\frac{1}{n}\sum_{i=0}^{n-1} \ln \|A_i\|\right)\right)^n \xrightarrow[n\to\infty]{} 0, \text{ a.s.}$$

Together with (1.56) this implies (1.54). Now, observe that the shift applied to Ψ on the right–hand side of (1.53) produces the shift of $\{X_n(\Psi)\}$ on the left–hand side so that $\{X_n(\Psi)\}$ is stationary.
Now, from (1.54) we get that $X_n(Y,\Psi)$ converges to $X_n(\Psi)$ in probability which together with the fact that $X_n(\Psi) \stackrel{d}{=} X_0(\Psi)$ implies (1.55). Moreover

$$\begin{aligned} A_n X_n(\Psi) + B_n &= A_n \left(\sum_{j=0}^{\infty}\left(\prod_{i=n-j}^{n-1} A_i\right) B_{n-j-1}\right) + B_n \\ &= \sum_{j=1}^{\infty}\left(\prod_{i=(n+1)-j}^{(n+1)-1} A_i\right) B_{(n+1)-j-1} + B_n = X_{n+1}(\Psi), \text{ a.s.} \end{aligned}$$

so that the sequence $\{X_n(\Psi)\}$ satisfies the relation (1.51) almost surely. Thus, $\{X_n(\Psi)\}$ is a stationary solution of the stochastic recurrence equation for Ψ. To complete the proof, let us assume $\{Y_n\}$ is another stationary solution of (1.51) for Ψ. Then

$$\begin{aligned} |\, Y_n - X_n(\Psi)\,| &= \|A_{n-1}\|\,|\,Y_{n-1} - X_{n-1}(\Psi)\,| = \dots \\ &= \left\|\prod_{i=1}^{k} A_{n-i}\right\| |\,Y_{n-k} - X_{n-k}(\Psi)\,| \\ &\le \left\|\prod_{i=1}^{k} A_{n-i}\right\| |\,Y_{n-k}\,| + \left\|\prod_{i=1}^{k} A_{n-i}\right\| |\,X_{n-k}(\Psi)\,| \end{aligned} \tag{1.57}$$

By the assumptions (1.52) and the strong law of large numbers, we obtain

$$\left\|\prod_{i=1}^{k} A_{n-i}\right\| = \left(\exp\left(\frac{1}{k}\sum_{i=1}^{k} \ln \|A_{n-1}\|\right)\right)^k \xrightarrow[k\to\infty]{} 0, \text{ a.s.}$$

Therefore, being $\{Y_n\}$ and $\{X_n(\Psi)\}$ stationary sequences, we see that

$$\left\|\prod_{i=1}^{k} A_{n-i}\right\| |\,Y_{n-k}\,| \xrightarrow{\mathbb{P}} 0, \qquad \left\|\prod_{i=1}^{k} A_{n-i}\right\| |\,X_{n-k}(\Psi)\,| \xrightarrow{\mathbb{P}} 0$$

as $k \to \infty$. Using this and (1.57) we find that $Y_n - X_n(\Psi) = 0$ almost surely for $n \in \mathbb{Z}$ which proves the uniqueness of the solution of (1.51) for Ψ. ∎

Remark 6. *In the proof of this Theorem it jumps out, while applying the strong law of large numbers, that the conditions* (1.52) *can be softened to*

$$\gamma = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\ln \|A_{n-k} \ldots A_{n-1}\|\right] < 0, \qquad \mathbb{E}\left[\ln^+ |\, B_0 \,|\right] < \infty \tag{1.58}$$

The coefficient γ is called top Lyapunov exponent of the sequence of random matrices $\{A_n\}$ and is generally hard to compute. In Furstenberg and Kesten (1960) it is shown that if $\mathbb{E}\left[\ln^+ \|A_1\|\right] < \infty$ then

$$\gamma = \lim_{n\to\infty} \frac{1}{n} \ln \|A_1 \ldots A_n\| \,, \qquad a.s..$$

The relation between the existence of the stationary solution and the strict negativity of the Top Lyapunov exponent has been further examined. In Bougerol and Picard (1992), the authors show that under an irreducibility condition of the model (1.51), both properties are equivalent.

Definition 11. *The model* (1.51) *is called irreducible if given that there exists $I \subseteq \{1, 2, \ldots, D\}$ such that the linear space V generated by the canonical vectors $\{e_i, i \in I\}$ satisfies $A_1 V + G \subseteq V$, then $V = \mathbb{R}^D$.*

Theorem 15 (Theorem 2.5 in Bougerol and Picard (1992))**.** *Suppose that the model* (1.51) *has i.i.d. coefficients and that both $\mathbb{E}\left[\log^+ \|A_1\|\right]$ and $\mathbb{E}\left[\log^+ \|B_1\|\right]$ are finite. Then there exists a unique casual stationary solution to* (1.51) *if and only if the Top Lyapunov exponent is strictly negative, that is*

$$\gamma = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\ln \|A_{n-k} \ldots A_{n-1}\|\right] < 0$$

Remark 7. *It is worth mentioning that without the assumption of irreducibility this equivalence does not hold.*

To study the extremes and point process convergence for the model (1.51), we focus on the stationary distribution of $\{X_n\}$. The heavy–tailedness of this distribution can be shown using some of the results in Kesten (1973). Two scenarios are possible:

1. Either the distribution of A_1 has its support on the cone of nonnegative matrices, or

2. the support of the distribution of A_1 contains matrices with both, positive and negative entries.

The first case is easier to handle, and less conditions on the matrices $\{A_n\}$ are required for X_1 to be, in some sense, regularly varying. Theorems 3 and 4 in Kesten (1973) put together deal with this case.

Theorem 16 (Theorems 3 and 4 in Kesten (1973))**.** *Let (A_n, B_n) be a sequence of i.i.d. random $d \times d$ matrices and $d \times 1$ vectors with non–negative entries such that $\mathbb{P}[| B_1 | = 0] = 0$. Assume the following*

1. *For some $\varepsilon > 0$, $\mathbb{E}[\|A_1\|^{\varepsilon}] < 1$,*
2. *A_1 has no zero rows with probability one,*
3. *The group generated by the set*
$$\{\ln \|A_1 \dots A_n\| : n \geq 1, A_i \in supp(\mathbb{P}_{A_1})\}$$
is dense in $\mathbb{R}$.
4. *There exists a γ_0 such that*
$$\mathbb{E}\left[\left(\min_{1 \leq i \leq d} \sum_{j=1}^{d} A(i,j)\right)^{\gamma_0/2}\right] \geq d^{\gamma_0/2}$$
5. *$\mathbb{E}[\ln \|A_1\|^{\gamma_0} \ln \|A_1\|] < \infty$,*

Then, the following hold

1. *There exists a unique solution $\gamma \in (0, \gamma_0]$ to the equation*
$$0 = \lim_{n \to \infty} \frac{1}{n} \ln \mathbb{E}[\|A_1 \dots A_n\|^{\gamma}]$$
2. *If $\mathbb{E}[| B_1 |^{\gamma}] < \infty$, then X satisfies the following regular variation condition*
$$\text{For all } x \in \mathbb{R}^d \setminus \{0\}, \qquad \lim_{u \to \infty} \mathbb{P}[\langle x, X_1 \rangle > u] = w(x) \tag{1.59}$$
exists and is positive for all non–negative vectors $x \in \mathbb{R}^d \setminus \{0\}$.

In the second case, that is, if the matrices A_1 have positive probability of having negative entries, more hypotheses are needed to guarantee regular variation. In this scenario, Kesten's theorem 6 is needed, and we present it next.

Theorem 17 (Theorem 6 in Kesten (1973))**.** *Let (A_n, B_n) be a sequence of i.i.d. random $d \times d$ matrices and $d \times 1$ vectors such that $\mathbb{E}\left[\log^+(\|A_1\|)\right]$ Assume the following conditions hold*

1. *The matrix A_1 is a.s. nonsingular.*

2. *For every given $x \in S_{d-1}$ define the Markov chain $\{W_n\}$ by*

$$W_0 = x, \qquad W_n = A_n W_{n-1}$$

   Then, the process

$$\widetilde{W}_0 = x, \quad \widetilde{W}_n = \frac{W_n}{\|W_n\|}$$

   is open set irreducible in S_{d-1}.

3. *There exists $n \in \mathbb{N}$ and a cube $C \subseteq \mathbb{R}^{d^2}$ such that the nonsingular component of the distribution of $A_n A_{n-1} \ldots A_1$ has a density bounded bellow by $\delta > 0$ on C.*

4. *The group generated by the set*

$$\{\ln \|A_1 \ldots A_n\| : n \geq 1, A_i \in supp(\mathbb{P}_{A_1})\}$$

   is dense in $\mathbb{R}$.

5. *For every fixed vector r we have*

$$\mathbb{P}\left[B_1 = (I - A_1)r\right] < 1$$

6. *There exists $\kappa_0 > 0$ such that*

$$0 < \mathbb{E}\left[\|B_1\|^{\kappa_0}\right] < \infty$$
$$\mathbb{E}\left[\|A_1\|^{\kappa_0} \log^+(\|A_1\|)\right] < \infty$$
$$\mathbb{E}\left[(\lambda(A_1))^{\kappa_0}\right] \geq 1$$

then there exists $\kappa_1 \in (0, \kappa_0]$ *such that* X *satisfies the following regular variation condition*

$$\lim_{u\to\infty} \mathbb{P}\left[\langle x, X_1\rangle > u\right] = w(x) \tag{1.60}$$

exists and is positive for all $x \in S_{d-1}$.

Observe that the properties (1.59) and (1.60) are not the regular variation of the vector X but the regular variation of any given linear combination of their entries. Remember from section 1.1.1 that this two regular variations are equivalente only in certain cases. See Theorem 4.

1.7 Dynamical systems and Markov Chains Analysis

In this section we outline some techniques used to analyze Markov Chains when they are stochastic realizations of a Dynamical System. In $\mathbb{R}^d$ a general dynamical system with state space A and control set O is defined by the relation

$$x_0 = x, \qquad x_n = F(x_{n-1}, z_n) \tag{1.61}$$

where

$$F\colon A \times O \to A$$

is a $\mathcal{C}^\infty$ function. The sets $A \subseteq \mathbb{R}^d$ and $O \subseteq \mathbb{R}^p$ are assumed open. Define the k–th iteration of F by

$$F^{(1)}(x,z) = F(x,z), \qquad F^{(k)}(x) = F(F^{(k-1)}(x, z_{k-1}), z_k).$$

Then, the orbit of the point x is defined as the set

$$A_+(x) = \bigcup_{k\in\mathbb{N}} \{y\colon y = F^{(k)}(x, z_1, \ldots, z_k),\ z_j \in O\}.$$

This set is also referred to as *the set of reachable states starting from* x. Now, consider an i.i.d. sequence $\{W_n\}$ of random variables in $\mathbb{R}^p$ and define

$$X_0 = x, \qquad X_n = F(X_{n-1}, W_n). \tag{1.62}$$

The process $\{X_n\}$ is then a Markov chain and it may be viewed as a stochastic realization of the dynamical system (1.61). The support of the (common) distribution of the sequence $\{W_n\}$ is taken to be the control set O and we call the sequence a *disturbance* sequence.

The stability properties of the deterministic system (1.61) are deeply linked with the (random) stability properties of the Markov chain (1.62). The model (1.61) is called *controllable* if for $x \in A$ the set $A_+(x)$ has non–empty interior. Observe that this property depends heavily on the choice of the control set O. This controllability for the deterministic system implies that the transition kernel of the random system (1.62) has a *continuous component*. To explain further, we first introduce some concepts on Markov Chains. Since a deeper treatment on Markov chains is out of the scope of this Thesis, we refer to Nummelin (1984), Meyn and Tweedie (1993) and the references therein.

Definition 12. *Let $\{X_n\}$ be a Markov chain defined on $\mathbb{R}^d$ for some $d \geq 1$. We call $\{X_n\}$ a T–chain if there exists a discrete random variable N, independent of $\{X_n\}$ such that for all $x \in \mathbb{R}^d$ and $A \subseteq \mathcal{B}(\mathbb{R}^d)$*

$$\mathbb{P}\left[X_N \in A \middle| X_0 = x\right] \geq T(x, A)$$

where the function $T(\,\cdot\,, A)$ satisfies

$$\liminf_{y \to x} T(y, A) \geq T(x, A).$$

The function $T(\,\cdot\,,\,\cdot\,)$ is then called a continous component of the sampled chain X_N.

One of the most important properties of T–chains is the link between the property of *petiteness* and topollogical compactness. In a sense, the existence of the continuous component $T(\,\cdot\,,\,\cdot\,)$ indicates that the chain $\{X_n\}$ has a transition *almost adapted* to the topollogy of its state space.

Theorem 18 (Theorem 6.0.1 in Meyn and Tweedie (1993))**.** *If $\{X_n\}$ is a T–chain, then every topologically compact set is petite. Conversely, if every compact set is petit and the chain $\{X_n\}$ is ψ–irreducible for some non–trivial measure ψ, then it is a T–chain.*

This irreducibility property can be linked to the topology of the space, as is to be expected. Call a point reachable if the chain $\{X_n\}$ can, with possitive probability, reach every open neighborhood of it, and call the chain $\{X_n\}$ open set irreducible if every point in its state space is reachable.

Proposition 8 (Proposition 6.2.1 in Meyn and Tweedie (1993))**.** *If $\{X_n\}$ is a T–chain with one reachable point x^*, then the chain is ψ irreducible for the measure $\psi(A) = T(x^*, A)$. Thus, if the T–chain $\{X_n\}$ is open set irreducible, then it is ψ–irreducible.*

Back to the link between Markov chains and dynamical systems, we have the following results.

Proposition 9 (Proposition 7.1.2 in Meyn and Tweedie (1993))**.** *If the deterministic model* (1.61) *is controllable and if the distribution of the disturbance sequence $\{W_n\}$ has a continuous density with respect to Lebesgue measure, then the process $\{X_n\}$ defined by* (1.62) *is a T–chain.*

The relationship between deterministic stability and random stability can now be further developed.

Definition 13. *A set E such that $A_+(E) \subseteq E$ is called invariant for the dynamical system* (1.61)*. A set is called minimal if it is (topologically) closed, invariant, and does not contain any other closed, invariant proper subset.*

It can be shown that an invariant M set has the property

$$M = \overline{A_+(x)}, \quad x \in M$$

As expected, this minimal set has a close relationship with irreducibility.

Proposition 10. *If M is a minimal set for* (1.61) *and the disturbance sequence has a continuous density with respect to Lebesgue measure, then the process $\{X_n\}$ defined by* (1.62) *restricted to M is an open set irreducible (hence ψ–irreducible) T–chain. In particular, every compact $K \subseteq M$ is petite.*

Chapter 2

The Constant Conditional Correlations GARCH model

Multivariate GARCH (M-GARCH) modelling has been one of the most successful and prominent tools to understand and predict the temporal dependence among the second order moments of financial returns in the last two decades. M-GARCH models arise naturally as an empirically more relevant explanation of this feature than working with separate univariate GARCH models for each asset. See Bauwens et al. (2006) for a survey of multivariate GARCH models and Engle and Kroner (1995) for a presentation of the theoretical formulation and estimation of such models within simultaneous equations systems.

M-GARCH models are far from being the only description for stochastic volatility studied to date. The reader is referred to Shephard (2005) to find several other possible models for stochastic volatility and a comparison of each of them with the M-GARCH model.

One of the shortfalls of M-GARCH models is that as the dimension grows larger, the number of parameters increases dramatically. To avoid this issue, Bollerslev introduced the M-GARCH with constant conditional correlations (CCC-GARCH) in Bollerslev (1990) which reduces the number of parameters involved in estimation. The reduction of parameters makes the CCC-GARCH an attractive model for empirical applications.

The research on asymptotic theory for multivariate GARCH models has been developed in two main areas, stationarity and estimation. By using Markov chain theory in Boussama (1998), the author proved the existence of a unique strictly stationary and ergodic solution for multivariate GARCH models. He also showed that if the noise sequence distribution has a strictly positive density around 0, then the stationary solution is also geometrically β–mixing.

Regarding estimation, the consistency and asymptotic normality of the quasi–maximum likelihood estimator were proved in Comte and Lieberman (2003) for square integrable BEKK–GARCH processes (see also Ling and McAleer (2003)). Later Hafner and Preminger obtained the same for a multivariate factor–GARCH model under the assumption of finiteness of the fourth moment of the noise distribution in Hafner and Preminger (2008).

For multivariate models with constant conditional correlations, the regular variation of the marginal distributions of the squared process is given in Stărică (1999). The author characterizes the spectral measure for this process and gives an empirical method to estimate it in the two–dimensional space.

The asymptotic theory of the sample autocovariance function of M-GARCH models has not been developed. For the one–dimensional case this asymptotics are realted to the regular variation of the finite–dimensional distributions of the GARCH$(1, 1)$ process in Mikosch and Stărică (2000). Later,the same relationship is established for the general GARCH(p, q) in Basrak et al. (2002b).

The regular variation properties are used to study the maximum domain of attraction of the stationary distribution of the one–dimensional GARCH(p, q) process in Davis and Mikosch (2006).

The results presented here generalize the ones given in Basrak et al. (2002b) and Davis and Mikosch (2006) for the one–dimensional case to the multidimensional CCC-GARCH model. Assuming that the generating noise sequence $\{\eta_t\}_{t\in\mathbb{Z}}$ satisfies:

1. For all $t \in \mathbb{Z}$ and $i = 1, 2, \ldots, d$ the random variable $\eta_t(i)$ has a symmetric distribution with a density which is strictly positive on $\mathbb{R}$.

2. The distribution of the noise sequence is such that for any given $\theta \geq 1$ and $1 \leq j \leq d$ there exists $h > 1$ for which

$$\theta^h \leq \mathbb{E}\left[\eta_{t,j}^{2h}\right] \leq \infty,$$

we establish the regular variation of the finite–dimensional distributions of the CCC-GARCH(p, q). Using this regular variation property we give two alternative expressions for the componentwise–maximum domain of attraction of the stationary distribution of the process.

We also show that the asymptotic behavior of the sample autocovariance function of the CCC-GARCH(p,q),

$$\gamma_{n,X}(h) = n^{-1} \sum_{t=1}^{n-h} X_t X_{t+h}^T, \qquad h \geq 0,$$

depends on the so–called index of regular variation of the finite–dimensional distributions of the process. Denoting this index by α, three cases are possible: If $\alpha \in (0, 2)$ we have no consistency but convergence to an $\alpha/2$–stable random variable. If $\alpha > 2$ we have consistency but a Central Limit Theorem is only achieved for $\alpha > 4$.

This Chapter is structured as follows:

In Section 2.1 we introduce the notation used throughout the paper. Section 2.2 is devoted to the regular variation properties of the CCC-GARCH(p,q) process. In this Section our main Theorem, which gives sufficient conditions on the noise distribution for the process to have regularly varying finite–dimensional distributions, is stated and proven, and some examples of possible noise distributions which satisfy these conditions are given. In Section 2.3 we provide two alternative expressions for the componentwise–maximum domain of attraction of the stationary distribution of the CCC-GARCH. The asymptotic properties of the sample autocovariance function are studied in Section 2.4. Finally, in Section 2.5 the reader will find some of the technical results used in the proof of our main theorem.

2.1 Notation

Let $\mathcal{M}(d \times d)$ be the space of square $d \times d$ matrices with real coefficients, and let $M \in \mathcal{M}(d \times d)$

1. $\delta(M)$ denotes the vector in $\mathbb{R}^d$ whose entries are $\delta(M)(i) = M(i,i)$ for $i = 1, 2, \ldots, d$, that is, the main diagonal of M.

2. diag(M) denotes the diagonal matrix with the same diagonal as M, namely,

$$\text{diag}(M)(i,j) = \begin{cases} 0, & \text{if } i \neq j \\ M(i,i), & \text{for } 1 \leq i \leq d. \end{cases}$$

3. The symbol vec (M) stands for the column–stack operation of the matrix M so that vec(M) is a vector in $\mathbb{R}^{d^2}$ whose entries are

$$\mathrm{vec}(M)(i) = \begin{cases} M(1,i), & \text{for } i = 1,2,\ldots,d, \\ M(2,i-d), & \text{for } i = d+1,\ldots,2d, \\ \vdots & \vdots \\ M(d,i+d-d^2), & \text{for } i = d^2-d+1,\ldots,d^2. \end{cases}$$

4. Given another matrix N of arbitrary dimensions, we define the Kronecker product of M and N by

$$M \otimes N = \begin{bmatrix} M(1,1)N & M(1,2)N & \cdots & M(1,d)N \\ M(2,1)N & M(2,2)N & \cdots & M(2,d)N \\ \vdots & \vdots & \ddots & \vdots \\ M(d,1)N & M(d,2)N & \cdots & M(d,d)N \end{bmatrix}.$$

   Given another matrix P, the vec operation and the $\otimes$ product are related by

$$\mathrm{vec}\,(MNP) = (P^T \otimes M)\mathrm{vec}\,(N)\,. \tag{2.1}$$

5. $\|M\|$ denotes the matrix norm of M,

$$\|M\| = \sup\left\{\|Mx\| \;:\; \|x\| = 1\right\},$$

   and take $\|x\|$ the max–norm in $\mathbb{R}^d$ from now on.

6. Given a sequence of square $d \times d$ random matrices $\{A_t\}_{t\in\mathbb{N}}$ defined on the same measurable space, we define its top Lyapunov exponent as

$$\gamma = \inf\left\{\frac{1}{t}\mathbb{E}\left[\ln(\|A_t A_{t-1}\ldots A_1\|)\right], t \in \mathbb{N}\right\}.$$

   If $\mathbb{E}\left[\ln^+\|A_1\|\right] < \infty$, then

$$\gamma = \lim_{t\to\infty}\frac{1}{t}\ln\|A_t A_{t-1}\ldots A_1\| \quad a.s.$$

   as proved in Furstenberg and Kesten (1960) which provides, potentially, a way to evaluate the index γ via simulation. On this subject, see Gol′dsheĭd (1991).

7. Given two vectors $x, y \in \mathbb{R}^d$ we denote by $\langle x, y\rangle$ their inner product $\langle x, y\rangle = \sum_{i=1}^d x_i y_i$.

2.2 Main Result

The multivariate GARCH(p,q) model with constant conditional correlations was introduced in Bollerslev (1990) and is defined as follows, see Bauwens et al. (2006).

Definition 14. *Given a sequence of i.i.d. random vectors $\{\eta_t\}_{t\in\mathbb{Z}}$ with mean vector 0 and covariance matrix R such that $R(i,i) = 1$ for all $i = 1, 2, \ldots, d$, we say the stochastic process $\{X_t\}_{t\in\mathbb{Z}}$ is a CCC-GARCH(p,q) if it satisfies the equations*

$$\begin{aligned} \delta(H_t) =& C + \sum_{i=1}^{p} A_i \delta(X_{t-i}X_{t-i}^T) + \sum_{j=1}^{q} B_j \delta(H_{t-j}), \\ D_t =& diag(H_t(1,1)^{1/2}, H_t(2,2)^{1/2}, \ldots, H_t(d,d)^{1/2}), \\ H_t =& D_t R D_t, \\ X_t =& D_t \eta_t. \end{aligned} \tag{2.2}$$

The vector C is assumed positive to avoid the trivial solution $X_t = 0$. The matrices A_i, B_j are assumed to be nonnegative for $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q$. Finally, the matrices A_p, B_q are supposed to be non–zero to avoid ambiguity about the order of the process.

Remark 8. *The matrices $\{H_t\}_{t\in\mathbb{Z}}$ are non–negative–definite because the entries of D_t are non–negative and R is non–negative–definite. They are positive–definite if and only if all the conditional variances are positive and the matrix R is positive–definite. Furthermore, H_t is the covariance matrix of X_t conditional on $\mathcal{F}_t = \sigma(X_s, s < t)$ as the following lines show.*

$$\begin{aligned} \mathbb{E}\left[X_t X_t^T \middle| \mathcal{F}_t\right] =& \mathbb{E}\left[H_t^{1/2} Z_t Z_t^t (H_t^{1/2})^T \middle| \mathcal{F}_t\right] \\ =& H_t^{1/2} \mathbb{E}\left[Z_t Z_t^T\right] (H_t^{1/2})^T \\ =& H_t^{1/2} I (H_t^{1/2})^T \\ =& H_t \end{aligned}$$

In Boussama (1998), it is shown that the CCC-GARCH process admits a Markovian state space representation which solves the stochastic recurrence equation (SRE)

$$Y_t = A(\eta_t) Y_{t-1} + G, \tag{2.3}$$

where

$$
\begin{aligned}
Y_t =&(\delta(H_{t+1})^T, \ldots, \delta(H_{t-q+2})^T, \delta(X_t X_t^T)^T, \ldots, \delta(X_{t-p+2}X_{t-p+2}^T)^T)^T \\
G =&(C^T, 0, \ldots, 0)^T
\end{aligned}
$$

$$
A(\eta_t) = \begin{bmatrix}
A_1 \mathrm{diag}(\eta_t \eta_t^T) + B_1 & B_2 & \cdots & B_{q-1} & B_q & A_2 & A_3 & \cdots & A_p \\
I & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
0 & I & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & I & 0 & 0 & 0 & \cdots & 0 \\
\mathrm{diag}(\eta_t \eta_t^T) & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & I & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & I & 0
\end{bmatrix} \tag{2.4}
$$

Remark 9. *It is shown in Boussama (1998) that a unique causal stationary solution* $\{Y_t\}$ *to* (2.3) *exists if and only if the top Lyapunov exponent of the sequence of matrices* $\{A(\eta_t)\}$ *satisfies*

$$\gamma(A(\eta_t)) < 0$$

Furthermore if the noise distribution has a density which is strictly positive on $\mathbb{R}^d$ *the stationary solution is geometrically* β*–mixing.*

Remark 10. *The matrices* $\{A(\eta_t)\}_{t\in\mathbb{Z}}$ *used in representation* (2.3) *are similar to those used in Basrak et al. (2002b) to embed the one–dimensional GARCH process into a SRE. In the multivariate case the matrices are defined blockwise and thus calculations involving them need to be done with appropriate matrix techniques.*

Using this representation the regular variation of the marginal distributions of the process $\{X_t\}$ is shown by Kesten's Theorem, see Theorem 2.4 in Basrak et al. (2002b), and the regular variation of the finite dimensional distributions is provided for noise distributions with symmetric marginals as a consequence. Three hypotheses are made for this purpose:

$\mathcal{H}0$: The parameters A_i, B_j have no zero rows.

$\mathcal{H}1$: The distribution F of the noise sequence admits a density f which is strictly positive on $\mathbb{R}^d$.

$\mathcal{H}2$: For any given $\theta \geq 1$ there exists $h > 1$ for which

$$\theta^h \leq \mathbb{E}\left[\eta_{t,j}^{2h}\right] \leq \infty \qquad \text{for } 1 \leq j \leq n.$$

Hypothesis $\mathcal{H}0$ is a natural one. It states that every vector $X_{t-i},\ i = 1, \ldots, p$ and every matrix $H_{t-j}\ j = 1, \ldots, q$ is involved in the definition of the matrix H_t. Hypotheses $\mathcal{H}1$ and $\mathcal{H}2$ imply that the noise sequence is sufficiently spread–out. It is worth mentioning that hypothesis $\mathcal{H}2$ is, as will be shown in section 2.2.1, rather mild.

Theorem 19. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a stationary CCC-GARCH(p,q) process with parameters $\{A_i, B_j, i \leq p, j \leq q\}$ and $\{Y_t\}_{t\in\mathbb{Z}}$ its SRE representation* (2.3). *Under $\mathcal{H}0$, $\mathcal{H}1$ and $\mathcal{H}2$, there exists $\kappa_1 > 0$ and $w(x) > 0$ such that*

$$\forall x \in \mathbb{R}^{d(p+q-1)} \setminus \{0\}, \quad \lim_{u\to\infty} u^{\kappa_1} \mathbb{P}\left[\langle x, Y_1\rangle > u\right] = w(x).$$

Proof. Let m be an arbitrary natural number and consider the sequence $\{Y_{tm}\}_{t\in\mathbb{Z}}$ which, as shown in Basrak et al. (2002b), satisfies the stochastic recurrence equation

$$Y_{tm} = \hat{A}_t^{(m)} Y_{(t-1)m} + \hat{B}_t^{(m)},$$

for

$$\begin{aligned}\hat{A}_t^{(m)} =& A_{tm}A_{tm-1}\ldots A_{tm-m+1},\\ \hat{B}_t^{(m)} =& \sum_{k=1}^{m-1} A_{tm}A_{tm-1}\ldots A_{tm-k+1}G.\end{aligned}$$

It is for this sequence that we will verify Kesten's hypothesis by choosing a suitable value for m. For ease of presentation, the required hypothesis are restated one by one with its respective proof.

1. $\mathbb{E}\left[\ln\|A_t\|\right] < \infty$ *and* $\mathbb{E}\left[\ln\|B_t\|\right] < \infty$ *and the top Lyapunov exponent of the sequence of random matrices $\{A_t\}$ is strictly negative.*

This was proved in Boussama (1998).

2. *Both* $\mathbb{E}\left[\ln(\left\|\hat{A}_t^{(m)}\right\|)\right]$ *and* $\mathbb{E}\left[\ln(\left\|\hat{B}_t^{(m)}\right\|)\right]$ *are finite, and there exists* $\varepsilon > 0$ *such that* $\mathbb{E}\left[\|A\|^{\varepsilon}\right] < 1$.

In Basrak et al. (2002b) these properties are shown to be satisfied for $m \geq M_0$ with M_0 a given natural number in the one–dimensional case. The argument rests on the negativity of the Lyapunov exponent associated to the SRE representation of the process. Remark 9 tells us that stationarity is equivalent to the fact that the top Lyapunov exponent of the sequence $A(\eta_t)$ in (2.3) is strictly negative. Therefore, we can follow exactly the same lines as in the one–dimensional case to prove Kesten's hypotheses 1 and 2.

3. *The matrix* $\hat{A}_t^{(m)}$ *has no zero rows a.s.* (*for any fixed* $m \in \mathbb{N}$)

The matrices A_{tm} and A_{tm-1} have no zero rows a.s. because of hypotheses $\mathcal{H}0$ and $\mathcal{H}1$. It follows that their product $A_{tm}A_{tm-1}$ shares this property as Lemma 4 shows. Using Lemma 4 recursively we conclude that the matrix $\hat{A}_t^{(m)}$ has no zero rows a.s. for any fixed $m \in \mathbb{N}$.

4. *The set* $\{\ln(\|a_n \dots a_1\|) : n \geq 1, a_n \dots a_1 > 0,\ a_i \in supp(\mathbb{P})\}$ *generates a dense group in* $\mathbb{R}$. $\mathbb{P}$ stands for the distribution of $\hat{A}_1^{(m)}$.

Observe that the support of the random variables $\{\eta_t^2(i)\}$ is the interval $(0, \infty)$. The entries of the matrix $\hat{A}_1^{(m)}$ are multilinear forms of the random variables

$$\{\eta_s^2(i), i = 1, \dots, d, s = tm - m + 1, \dots, tm\},$$

as will be shown below. Since multilinear forms are continuous functions, the support of $\hat{A}_1^{(m)}$ is connected. The same can be said about the support of $\left\|\hat{A}_1^{(m)}\right\|$. Therefore, the support of $\ln\left(\left\|\hat{A}_1^{(m)}\right\|\right)$ contains an interval, which yields the desired property.
Let us now argue that the entries of the matrix $\hat{A}_1^{(m)}$ have the aforementioned form.
We denote the blocks by square brackets, so that the block (i, j) of the matrix A_{tm} is denoted by $A_{tm}[i, j]$.
Begin observing that the blocks $A_{tm}[1, 1]$ and $A_{tm}[q + 1, 1]$ are already multilinear forms of the random variables

$$\{\eta_{tm}^2(i), i = 1, 2 \dots, d\}.$$

Block–matrix multiplication rules show that all the blocks

$$A_{tm}A_{tm-1}[1,j], \quad \text{and} \quad A_{tm}A_{tm-1}[q+1,j], j=1,2\ldots,d$$

are multilinear forms of the random variables

$$\{\eta^2_{tm-j}(i), i=1,2\ldots,d,\ j=1,2\}.$$

The particular arrangement of the identity blocks in the matrix A_{tm}, see (2.4), implies the following block equalities

$$A_{tm}A_{tm-1}[i,j] = A_{tm-1}[i-1,j], \quad \text{for } i \neq 1, q+1 \tag{2.5}$$

which show two things. First, that the blocks

$$A_{tm}A_{tm-1}[2,1] \quad \text{and} \quad A_{tm}A_{tm-1}[q+2,1]$$

are multilinear forms of the random variables

$$\{\eta^2_{tm-1}(i), i=1,2\ldots,d\}.$$

Second, and as a consequence, that the blocks

$$A_{tm}A_{tm-1}A_{tm-2}[i,j] \quad \text{and} \quad A_{tm}A_{tm-1}A_{tm-2}[q+i,j]$$

are in the desired form for $i=1,2$ and $j=1,2,\ldots,d$

We can now see that the entries of the product $A_{tm}A_{tm-1}\ldots A_{tm-k}$ are multilinear forms of the random variables

$$\{\eta^2_{tm-l}(i), l=1,2,\ldots,k; i=1,2\ldots,d\}$$

in all the blocks (i,j) and $(q+i,j)$ for $i=1,\ldots,k-1$ and $j=1,2\ldots,d$. This suffices to prove Assertion A.

5. *There exists a* $\kappa_0 > 0$ *such that*

$$\mathbb{E}\left[\left(\min_i \sum_j A_{ij}\right)^{\kappa_0}\right] \geq d^{\kappa_0/2} \quad \text{and} \quad \mathbb{E}\left[\|A_1\|^{\kappa_0}\ln(\|A_1\|)\right] < \infty.$$

Take $m = \max(p+1, q+1, M_0)$. For this particular choice of m all of the previously proved hypotheses hold. For ease of notation, let

$$\hat{Y}_t := \hat{Y}_t^{(m)},\ \hat{A}_t = \hat{A}_t^{(m)},\ \hat{B}_t = \hat{B}_t^{(m)}.$$

The block

$$\hat{A}_t[1,1] = A_{tm}A_{tm-1}\cdots A_{(t-1)m}[1,1]$$

contains the term

$$(A_1\mathrm{diag}(\eta_{tm}\eta_{tm}^T) + B_1)\cdots(A_1\mathrm{diag}(\eta_{(t-1)m}\eta_{(t-1)m}^T) + B_1)$$

added to others, which implies it also contains the term

$$B_1^{m-1}A_1\mathrm{diag}(\eta_{(t-1)m}\eta_{(t-1)m}^T).$$

The block equalities (2.5) show that the block $\hat{A}_t[k,1]$ contains the term

$$B_1^{m-k}A_1\mathrm{diag}(\eta_{tm-m}\eta_{tm-m}^T)$$

added to others for $k = 1, 2, \ldots, (q-1)$.
Observe that

$$B_1^{m-k}A_1\mathrm{diag}(\eta_{(t-1)m}\eta_{(t-1)m}^T)(i,1) = (B_1^{m-k}A_1)(i,1)\eta_{(t-1)m,1}^2$$

which implies

$$\sum_j \hat{A}_t(i,j) \geq (B_1^{m-k}A_1)(i,1)\eta_{(t-1)m,1}^2 \quad i = 1, 2, \ldots, dq-1.$$

Let us now analyze the blocks $(k,1)$ with $k \geq q+1$. The block $(q+1,1)$ contains the term

$$\mathrm{diag}(\eta_{tm}\eta_{tm}^T)(A_1\mathrm{diag}(\eta_{tm-1}\eta_{tm-1})^T + B_1)\cdots(A_1\mathrm{diag}(\eta_{(t-1)m}\eta_{(t-1)m}^T) + B_1)$$

added to others. Again, because of the equalities (2.5), we see that the block $\hat{A}_t[q+k,1]$ contains the term

$$\mathrm{diag}(\eta_{tm-k}\eta_{tm-k}^T)B_1^{m-k-1}.$$

Since

$$(\mathrm{diag}(\eta_{tm-k}\eta_{tm-k}^T)B_1^{m-k-1})(i,1) = \eta_{tm-k,i}^2 B_1^{m-k-1}(i,1),$$

then

$$\sum_j \hat{A}_t(i,j) \geq B_1^{m-k(i)-1}(i,1)\eta_{tm-k(i),i}^2 \quad i = dq, dq+1, \ldots, dq+dp-1$$

where $k(i) = \left[\frac{i}{qd}\right]$. Therefore upon using hypothesis $\mathcal{H}2$ for the random variables $\{\eta_{t,i}\}$ on each of the two parts of the matrix, we conclude that there exists $h > 1$ such that

$$\mathbb{E}\left[\left(\min_i \sum_j \hat{A}_t(i,j)\right)^h\right] \geq d^{h/2}.$$

For this value of $h > 1$, by the moment conditions on the variables $\{\eta_t\}$ and the independence of this sequence it can be seen that

$$\begin{aligned}\mathbb{E}\left[\left\|\hat{A}_1\right\|^h\right] &\leq \mathbb{E}\left[\left(\sum_{i,j} \hat{A}(i,j)\right)^h\right] \\ &< \mathbb{E}\left[\sum_{i,j} \hat{A}(i,j)^h\right] < \infty.\end{aligned}$$

Finally, as in Basrak et al. (2002b) for the one–dimensional case, it follows that

$$\mathbb{E}\left[\left\|\hat{A}_1\right\|^h \ln\left(\left\|\hat{A}_1\right\|\right)\right] < \infty.$$

Having proved all of Kesten's hypotheses we know there exists $\kappa_1 > 0$ and $w(x) > 0$ such that

$$\forall x \in \mathbb{R}^{(p+q-1)d} \setminus \{0\}, \quad \lim_{u\to\infty} u^{\kappa_1} \mathbb{P}\left[\langle x, Y_1\rangle > u\right] = w(x),$$

proving the Theorem. ■

Corollary 6. *Let $\{X_t\}$ be a stationary CCC–GARCH(p,q) process satisfying the hypotheses of Theorem 19 and let κ_1 be the regular variation index of $\langle x, Y_1\rangle$.*

1. *If κ_1 is not an even integer, then X_t is regularly varying with index $2\kappa_1$.*

2. *If, furthermore, for all t the vector η_t has symmetric marginal distributions, then the finite–dimensional distributions of the process $\{X_t\}$ are regularly varying with index $2\kappa_1$.*

Proof. Using Theorem 1.1 in Basrak et al. (2002a) and the sequence $\{Y_t\}$ being stationary we conclude that its marginal distributions are multivariate regularly varying if the number κ_1 is not an even integer. To show that X_1 is regularly varying write

$$X_1 = \mathrm{diag}(\eta_1)(\sigma_1(1), \ldots, \sigma_1(d))^T.$$

Observe that the vector $(\sigma_1(1), \ldots, \sigma_1(d))^T$ is regularly varying which follows because Y_t is regularly varying and $\sigma_1(i)$ is a.s. non–negative. Apply Proposition 5.1 in Basrak et al. (2002b) to finish the proof of 1.

To show that the finite–dimensional distributions of $\{X_t\}$ are multivariate regularly varying if the distribution F of the noise sequence has symmetric marginal distributions, begin by observing that by Corollary 2.7 in Basrak et al. (2002b) the finite dimensional distributions of the process $\{Y_t\}$ are regularly varying with index κ_1, which implies that for any given $k \in \mathbb{N}$ the vector

$$(\delta(H_1), \delta(X_1 X_1^T), \ldots, \delta(H_k), \delta(X_k X_k^T))^T$$

is regularly varying. Since for all $i = 1, 2, \ldots, d$ and $t = 1, 2, \ldots, k$ we have that $\sigma_t(i) \geq 0$ and $X_t(i)$ is symmetric, it follows that the vector

$$(\sigma_1(1), \ldots, \sigma_1(d), X_1(1), \ldots, X_1(d), \ldots, \sigma_k(1), \ldots, \sigma_k(d), X_k(1), \ldots X_k(d))^T$$

is also regularly varying, with index $2\kappa_1$ by Lemma 5 which completes the proof. ∎

Remark 11. *Hypothesis $\mathcal{H}2$ can be changed for the following one.*

$$\mathcal{H}2\text{': There exists } h_0 > 2 \text{ such that } \mathbb{E}\left[|\,\eta_t\,|^h\right] < \infty \text{ for } h < h_0 \text{ and } \mathbb{E}\left[|\,\eta_t\,|^{h_0}\right] = \infty.$$

This moment assumption implies that

$$\mathbb{E}\left[|\,\eta_t\,|^h\right] \xrightarrow[h \to h_0]{} \infty$$

from which hypothesis $\mathcal{H}2$ follows. It is some times more convenient to work with $\mathcal{H}2'$ than with $\mathcal{H}2$.

2.2.1 Examples

In this section we provide examples of distribution functions for the noise sequence $\{\eta_t\}_{t\in\mathbb{Z}}$ satisfying the moment conditions $\mathcal{H}2$ or $\mathcal{H}2'$. To do this, we consider different i.i.d. sequences $\{Z_t\}_{t\in\mathbb{Z}}$ with mean vector 0 and identity covariance matrix and a fixed correlation matrix R with $R(i,i) = 1$ for $i = 1, 2, \dots, d$. The noise sequence $\{\eta_t\}_{t\in\mathbb{Z}}$ is then defined by the relation

$$\eta_t = LZ_t. \tag{2.6}$$

The matrix L is the unique Cholesky factor of R which exists because R is non–negative definite.

Regarding the model's parameters, we assume that hypothesis $\mathcal{H}0$ is satisfied and also that

$\mathcal{H}3$: The spectral radius of $\sum_i A_i + \sum_j B_j$ is strictly less than one.

Assumption $\mathcal{H}3$ implies that the top Lyapunov exponent of the sequence $\{A(\eta_t)\}_{t\in\mathbb{Z}}$ is strictly negative so that a unique stationary CCC-GARCH process exists. See Remark 9. The reader is referred to Boussama (1998) for the proof of this implication.

The main purpose of this section is to show that the noise sequence $\{\eta_t\}_{t\in\mathbb{Z}}$ needs not be heavy–tailed for the generated CCC-GARCH process to be regularly varying as it could seem at first glance.

It is worth mentioning that the conditions of Theorem 19 also apply to the one–dimensional case, with obvious modifications, so that the examples given here provide regularly varying one–dimensional GARCH processes.

Example 2 Heavy–tailed noise sequence
Let $\{Z_t\}_{t\in\mathbb{Z}}$ be i.i.d. random vectors with independent entries and with Pareto–like tails, i.e.,

$$\mathbb{P}\left[Z_t(i) > x\right] \sim Kx^{-\alpha}, \qquad \alpha > 0, \quad x \to \infty \tag{2.7}$$

and with a positive density on $\mathbb{R}$. Define $\{\eta_t\}_{t\in\mathbb{Z}}$ as in (2.6). Then, for each $i = 1, 2, \dots, d$ the random variable $\eta_t(i)$ is regularly varying with index $\alpha > 0$ (because it is a linear combination of independent regularly varying random variables). Consequently

$$\mathbb{E}\left[\mid \eta_t(i) \mid^h\right] \begin{cases} < \infty, & h < \alpha, \\ = \infty, & h > \alpha. \end{cases}$$

By Remark 11 we conclude that the CCC-GARCH(p,q) process generated with this noise sequence has regularly varying marginal distributions.

If the noise distribution has symmetric marginals, then the process has regularly varying finite dimensional distributions as well. This is the case if we let $Z_t(i)$ have a Student's t distribution with ϑ degrees of freedom with density given by

$$f(x) = \frac{\Gamma((\vartheta+1)/2)}{\sqrt{\vartheta\pi}}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}$$

which, by Karamata's Theorem, satisfies (2.7) for $\alpha = \vartheta$.

This example is the easiest one to construct, but it may mislead the reader into thinking that the heavy–tailedness of the noise sequence is somehow necessary for the regular variation of the CCC-GARCH process. The following examples show that this is not the case.

Example 3 Medium–tailed noise sequence
Consider a family of random vectors $\{\eta_t\}_{t\in\mathbb{N}}$ such that for all $i = 1, 2 \ldots d$ we have $\eta_{t,i} \sim logN(0, 1)$ and such that its covariance matrix is R. Then, since

$$\mathbb{E}\left[\eta_{t,i}^{2n}\right] = \exp\{2n^2\}$$

it is easy to prove that for any given $c > 1$ there exist $n \in \mathbb{N}$ such that

$$\mathbb{E}\left[\eta_{t,i}^{2n}\right] \geq c^n.$$

Indeed, it suffices to take $n \geq \frac{1}{2}\ln(c)$. By Theorem 19 the CCC-GARCH(p,q) process generated by this sequence will be regularly varying.

Remark 12. *The support of the logNormal density doesn't contain a neighborhood of* 0*. Therefore, using the logNormal distribution for the random variables Z_t we generate a stationary CCC-GARCH process which may not be β–mixing. The process is, however, multivariate regularly varying. To understand why, observe that the aforementioned density assumption is only used in the proof of hypothesis 3 of Kesten's Theorem indirectly by showing it implies that the random variables $\{\eta_t^2(i)\}$ have the interval $(0,\infty)$ as support for its density. For the logNormal distribution this property is also true so that the marginal distributions of the CCC-GARCH process generated are regularly varying. For the finite–dimensional distributions observe that since the random variables η_t have strictly positive entries, so does the random vector X_t. Thus, by the same argument given in the proof of Theorem 19 the finite dimensional distributions of $\{X_t\}$ are regularly varying.*

Example 4 Gaussian noise
In time series applications Gaussian noise is the most common noise. It is therefore desirable to include this distribution in the list of the ones generating regularly varying CCC-GARCH processes. Consider a sequence $\{Z_t\}$ of i.i.d. standard normal random vectors and let $\{\eta_t\}$ be defined as in equation (2.6). Observe that each component of η_t is Gaussian with mean 0 and fixed variance $\sigma_i, i = 1, 2, \dots d$.

Therefore, for any given component we have

$$\mathbb{E}\left[\eta_{t,i}^{2n}\right] = \frac{\sigma_i^{2n}(2n)!}{2^n n!}.$$

From which, given that $c \geq 1$, it is enough to take

$$n \geq 2c - 1$$

to get

$$\mathbb{E}\left[\eta_{t,i}^{2n}\right] \geq c^n.$$

Therefore, the Gaussian generated CCC-GARCH has regularly varying finite–dimensional distributions.

2.3 Extreme Values

In this section we study the extreme values of the CCC-GARCH process. The main focus is on the asymptotic behavior of the norm–maximum and the componentwise–maximum of the process, which are defined as follows.

Definition 15. *The norm–maximum of the CCC-GARCH process* $\{X_t\}_{t\in\mathbb{Z}}$ *is defined by*

$$M_n^{\|X\|} = \max\{\|X_t\|\,,\ t \leq n\}$$

and its componentwise–maximum M_n *by the equations*

$$M_n(i) = \max\{X_t(i),\ t \leq n\}, \quad \textit{for } i = 1, 2 \dots, d$$
$$M_n = (M_n(1), M_n(2), \dots, M_n(d)).$$

The interest lies in the description of all distribution functions such that

$$\mathbb{P}\left[a_n^{-1} M_n^{\|X\|} \leq x\right] \to G(x), \tag{2.8}$$

$$\mathbb{P}\left[\left(a_n^{-1} M_n(1) \leq x(1), \dots, a_n^{-1} M_n(d) \leq x(d)\right)\right] \to H(x(1), \dots, x(d)), \tag{2.9}$$

where $0 < a_n \to \infty$ is the sequence of regular variation of the CCC-GARCH process. Of course, G in relation (2.8) is a real valued distribution function while H in relation (2.9) is $\mathbb{R}^d$–valued. To simplify notation, the limit (2.9) will be written

$$\mathbb{P}\left[a_n^{-1} M_n \leq x\right] \to H(x), \quad \text{for } x \in \mathbb{R}^d.$$

It is known from classical extreme value theory for stationary sequences that the convergence in distribution of the norm–maximum and the component–wise–maximum is related to that of the point process

$$N_n(\,\cdot\,) = \sum_{t=1}^{n} \varepsilon_{a_n^{-1} X_t}(\,\cdot\,) \tag{2.10}$$

where $\varepsilon_t(\,\cdot\,)$ is the Dirac measure at t. For a review of classical multivariate extreme value theory see Resnick (1987), Embrechts et al. (1997), Hsing (1989) and Hüsler (1990). It is also known that for a regularly varying sequence with index $\alpha > 0$ and a mixing property (see Leadbetter et al. (1983)) the coefficients $\{a_n\}$ satisfy

$$a_n \sim n^{1/\alpha} \tag{2.11}$$

and it is therefore customary to take $a_n = n^{1/\alpha}$ in this case. Since the CCC-GARCH process is stationary and β–mixing, relation (2.11) can be assumed to hold.

The next proposition gives the distributional limit for the point process (2.10) and is the basis for studying the extremes of the CCC-GARCH process.

Proposition 11. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a CCC-GARCH(p,q) process. Assume all the conditions in part 2 of Corollary 6 are satisfied. Then there exists a point process N such that $N_n \to N$ in distribution and N is identical in law to the point process*

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \varepsilon_{P_i Q_{i,j}}(\,\cdot\,)$$

where $\sum_{i=1}^{\infty} \varepsilon_{P_i}(\,\cdot\,)$ is a Poisson process with intensity measure

$$\nu(dy) = \theta\alpha y^{-\alpha-1} \mathbb{1}\,(y > 0)\,dy$$

for some $0 < \theta < 1$ and independent of the sequence of i.i.d. point processes $\sum_{j=1}^{\infty} \varepsilon_{Q_{i,j}}(\,\cdot\,), i \geq 1$.

Proof. By Theorem 19 we know that the finite–dimensional distributions of the process $\{X_t\}$ are regularly varying. By the β–mixing property we also know that there exists a sequence of positive integers r_n such that $r_n \to \infty, k_n = [n/r_n] \to \infty$ as $n \to \infty$ and

$$\mathbb{E}\left[\exp\left\{-\sum_{t=1}^{n} f(X_t a_n^{-1})\right\}\right] - \left(\mathbb{E}\left[\exp\left\{-\sum_{t=1}^{r_n} f(X_t a_n^{-1})\right\}\right]\right)^{k_n} \to 0,$$

as $n \to \infty$ for all simple, non–negative, measurable function f. See Basrak et al. (2002b).
Finally, it can be proved as is done in Basrak et al. (2002b) for the one dimensional case, that for all $y > 0$

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\left[\max_{k\le|\,t\,|\le r_n} \|X_t\| > a_n y \Big| \, \|X_0\| > a_n y\right] = 0.$$

These conditions enable us to apply Theorem 2.8 in Davis and Mikosch (1998) to finish the proof. ∎

We now deal with the extreme values of the CCC-GARCH process.

Theorem 20. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a stationary CCC-GARCH(p,q) process satisfying the hypotheses of part 2 of Corollary 6 and denote by α its index of regular variation.*

1. *The norm–maximum of $\{X_t\}_{t\in\mathbb{Z}}$ satisfies*

$$\mathbb{P}\left[n^{-1/\alpha} M_n^{\|X_t\|} \le x\right] \to \exp\{-\theta x^{-\alpha}\} \mathbb{1}\,(x > 0)\,,$$

where θ is the extremal index of $\|X_t\|$.

2. *The componentwise–maximum of the sequence $\{X_t\}_{t\in\mathbb{Z}}$ satisfies*

$$\mathbb{P}\left[n^{-1/\alpha} M_n \le x\right] \to \exp\{-\lambda\left(\{\mu\colon \mu(B_x) > 0\}\right)\},$$

where λ is the canonical measure of the point process N and

$$B_x = (-\infty, x]^c \quad \text{for any fixed } x \in \mathbb{R}_+^d.$$

Proof. The density $\nu(dy)$ in the description of the point process N given in Proposition 11 is the asymptotic density of the extremes of the sequence $\{\|X_t\|\}$, see Davis and Mikosch (2006) or Davis and Mikosch (1998). Part 1 is thus proved.
To prove part 2, first observe that

$$\mathbb{P}\left[n^{-1/\alpha}M_n \leq x\right] = \mathbb{P}\left[N_n(B_x) = 0\right]. \tag{2.12}$$

Since the set B_x is relatively compact, Proposition 11 gives us the convergence

$$\mathbb{P}\left[N_n(B_x) = 0\right] \xrightarrow[n\to\infty]{} \mathbb{P}\left[N(B_x) = 0\right]. \tag{2.13}$$

By dominated convergence, we have

$$\mathbb{P}\left[N(B_x) = 0\right] = \mathbb{E}\left[\lim_{t\to\infty} e^{-N(t\mathbb{1}(B_x))}\right] = \lim_{t\to\infty} \mathbb{E}\left[e^{-N(t\mathbb{1}(B_x))}\right].$$

Since λ is the canonical measure of N, see Kallenberg (1969), this expectation can be calculated as

$$\mathbb{E}\left[e^{-N(t\mathbb{1}(B_x))}\right] = \exp\left\{-\int(1 - e^{-\mu(t\mathbb{1}(B_x))})\lambda(d\mu)\right\}.$$

Because of this equality and, again, dominated convergence we now have

$$\begin{aligned}\mathbb{P}\left[N(B_x) = 0\right] &= \lim_{t\to\infty} \exp\left\{-\int(1 - e^{-\mu(t\mathbb{1}(B_x))})\lambda(d\mu)\right\}\\ &= \exp\left\{-\int(1 - \mathbb{1}\left(\mu(B_x) = 0\right))\lambda(d\mu)\right\},\end{aligned}$$

which shows that

$$\mathbb{P}\left[N(B_x) = 0\right] = \exp\left\{-\int \mathbb{1}\left(\mu(B_x) > 0\right)\lambda(d\mu)\right\}. \tag{2.14}$$

Putting equations (2.12), (2.13) and (2.14) together, it is seen that

$$\mathbb{P}\left[n^{-1/\alpha}M_n \leq x\right] \xrightarrow[n\to\infty]{} \exp\{-\lambda\left(\{\mu\colon \mu(B_x) > 0\}\right)\}.$$

■

Remark 13. *Three comments on the last result:*

1. *Despite the fact that the measure λ enables us to express the limit distribution of the normalized componentwise–maximum of the CCC-GARCH process $\{X_t\}_{t\in\mathbb{Z}}$, it is of no practical use for probability computations. However, by showing the existence of the limiting distribution it suggests that we can use non–parametrical statistical methods or resampling techniques to estimate it.*

2. *The number $\kappa_1 = \alpha/2$ is only known to solve equation*

$$0 = \lim_{n\to\infty} \frac{1}{n} \ln(\mathbb{E}\left[\|A_n \ldots A_1\|^{\kappa_1}\right]),$$

which cannot be solved explicitly.

3. *There's no closed known expression for calculating the extremal index θ of the sequence $\|X_t\|$. We only know that $\theta < 1$, which means the sequence $\|X_t\|$ has clusters with mean size $1/\theta > 1$.*

To obtain an alternative expression for the domain of attraction of the componentwise –maximum for the CCC-GARCH process, we follow section 6 in Perfekt (1997) to obtain:

Proposition 12. *Let $\{Y_t\}$ be the stationary solution to the CCC-GARCH(p,q) stochastic recurrence equation representation* (2.3)*. Let ν be such that*

$$n\mathbb{P}\left[a_n^{-1} Y_1 \in \cdot\,\right] \xrightarrow{v} \nu(\,\cdot\,).$$

Assume that ν satisfies

$$c_i = \nu\left(\{y \colon y_i > 1\}\right) > 0, \qquad \text{for all } i = 1, 2, \ldots, d^2(p+q-1).$$

Then, for $x \in (0,\infty)^d$ we have

$$\mathbb{P}\left[a_n^{-1} M_n \le x\right] \xrightarrow{n\to\infty} \exp\left\{ -\int_{[0,x]^c} \mathbb{P}\left[\prod_{i=1}^{j} A_i y \le x;\ j \ge 1\right] \nu(dy) \right\}.$$

Remark 14. *Observe that due to the definition of the multivariate extremal index, see Nandagopalan (1994), and the result in Proposition 12, the multivariate extremal index of the stationary process $\{Y_t\}$ may now be written as*

$$\theta(x) = \frac{\int_{[0,x]^c} \mathbb{P}\left[\prod_{i=1}^{j} A_i y \le x;\ j \ge 1\right] \nu(dy)}{\nu([0,x]^c)}.$$

This expression enables us to write the marginal extremal index for the ith component of Y_t as

$$\theta_i = \frac{\int_{\{y:\, y_i>1\}} \mathbb{P}\left[\prod_{i=1}^{j} A_i y \le x;\ j \ge 1\right] \nu(dy)}{c_i}.$$

This is the only closed form for this index known to the writers, and generalizes the one given in Mikosch and Stărică (2000) for the extremal index of the variances of the one–dimensional GARCH(1,1) process.

2.4 Autocovariance function

In this section we study the sample autocovariance function of the CCC-GARCH process defined as

$$\gamma_{n,X}(h) = n^{-1} \sum_{t=1}^{n-h} X_t X_{t+h}^T, \qquad h \ge 0.$$

One expects this sample function to converge a.s. to

$$\gamma_X(h) = \mathbb{E}\left[X_0 X_h^T\right],$$

which is the autocovariance function of the process, by the Law of Large Numbers. Furthermore, having this convergence the Central Limit Theorem is also expected to apply, that is, a convergence of $\gamma_{n,X}(h) - \gamma_X(h)$ to a Normal distribution at rate $n^{1/2}$ is expected.

For a regularly varying CCC-GARCH process, these limit relations need not be true. The reason is that the moments of the stationary distribution of the CCC-GARCH process depend on the index of regular variation. The relationship between the index of regular variation and the moments of a random variable is well known and can be found in Resnick (1987), Embrechts et al. (1997) or Bingham et al. (1989) among others.

The following result is the generalization of Theorem 3.6 in Basrak et al. (2002b) for the autocovariance function of the CCC-GARCH process.

Theorem 21. *Let $\{X_t\}$ be a stationary CCC-GARCH(p,q) process satisfying all the conditions in Theorem 19 and assume that the noise vectors η_t have symmetric marginal distributions. Let α denote the regular variation exponent of $\{X_t\}$, then*

1. *If* $\alpha \in (0, 2)$*, then*

$$(n^{1-2/\alpha} vec\,(\gamma_{n,X}(h))_{h=0,1,\dots,m} \xrightarrow{d} (V_h)_{h=0,1,\dots,m} \tag{2.15}$$

where $(V_0, V_1, \dots, V_m)$ *is jointly* $\alpha/2$ *stable in* $\mathbb{R}^{dm}$*.*

2. *If* $\alpha \in (2, 4)$*, then*

$$(n^{1-2/\alpha}(vec\,(\gamma_{n,X}(h) - \gamma_X(h))_{h=0,1,\dots,m} \xrightarrow{d} (V_h)_{h=0,1,\dots,m} \tag{2.16}$$

were $(V_0, V_1, \dots, V_m)$ *is jointly* $\alpha/2$ *stable in* $\mathbb{R}^{dm}$*.*

3. *If* $\alpha > 4$ *then equation* (2.16) *holds with normalization* $n^{1/2}$*, where* V *is multivariate normal with mean zero.*

Proof. 1. If $\alpha \in (0, 2)$ the result is immediate from Proposition 11 and Theorem 3.5 in Davis and Mikosch (1998).

2. For $\alpha \in (2, 4)$ it is necessary to show that the following condition holds.

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} var \left(n^{-2/\alpha} \sum_{t=1}^{n-h} X_{t,i}X_{t+h,j} \mathbb{1} \left(|\, X_{t,i}X_{t+h,j} \,| \le n^{2/\alpha}\varepsilon \right) \right) = 0$$

for all $i, j = 1, 2, \dots, d$.

To this end, choose $h \ge 1$. First, by symmetry of the noises $\{\eta_t\}$ the random variables $\{X_{t,i}X_{t+h,j} \mathbb{1} \left(\left|\, X_{t,i}X_{t+h,j} \le n^{2/\alpha}\varepsilon \,\right| \right)$ are uncorrelated. This, together with the fact that $X_{0,i}X_{h,j}$ is regularly varying with index $\alpha/2$ implies

$$\begin{aligned} var &\left(\sum_{t=1}^{n-h} X_{t,i}X_{t+h,j} \mathbb{1} \left(\left|\, X_{t,i}X_{t+h,j} \le n^{2/\alpha}\varepsilon \,\right| \right) \right) \\ &= (n-h)n^{-4/\alpha} \mathbb{E} \left[X_{0,i}^2 X_{h,j}^2 \mathbb{1} \left(\left|\, X_{t,i}X_{t+h,j} \le n^{2/\alpha}\varepsilon \,\right| \right) \right] \\ &\sim n^{1-4/\alpha} (n^{2/\alpha}\varepsilon)^2 (4-\alpha) \mathbb{P} \left[\left|\, X_{0,i}X_{h,j} > n^{2/\alpha}\varepsilon \,\right| \right] \\ &\to (4-\alpha)\varepsilon^{2-\alpha/2}, \qquad \text{as } n \to \infty \\ &\to 0, \qquad \text{as } \varepsilon \to 0. \end{aligned}$$

where the asymptotic equivalence follows from Karamata's Theorem, see Bingham et al. (1989).

3. For $\alpha \in (4, \infty)$ we see that the stationary time series

$$\{X_{t,i}^2, X_{t,i}X_{t+1,j}, \ldots X_{t,i}X_{t+m,j},\ i,j = 1,2,\ldots,d\}_t$$

has a finite $2+\theta$ moment (for some $\theta > 0$) and is geometrically β–mixing, and thus geometrically α–mixing. It follows from standard limit Theorems for mixing sequences (see Doukhan (1994)) that

$$(n^{1/2}(\mathrm{vec}\,(\gamma_{n,X}(h) - \gamma_X(h)))_{h=0,1,\ldots,m} \xrightarrow{d} (V_h)_{h=0,\ldots,m}$$

where V has a multivariate normal distribution with mean zero.

■

2.5 Technical Results

In this section we state some of the technical results used to prove Theorem 19 and present their proof when needed.

Lemma 4. *If two non–negative matrices have no zero rows then their product (when defined) has no zero rows either.*

Proof. Let p, q, r be fixed natural numbers. Consider the matrices A and B of respective dimensions $p \times q$ and $q \times r$ so that the matrix AB is well defined and of dimension $p \times r$. Assume that A and B have no zero rows.

To show that AB has no zero rows, first consider a fixed $i \in \{1, 2, \ldots, p\}$. Since A has no zero rows it follows that

$$\exists\ j \in \{1, 2, \ldots, q\} \quad \text{such that} \quad A(i,j) > 0$$

For this particular j, and since B has no zero rows,

$$\exists\ k \in \{1, 2, \ldots, r\} \quad \text{such that} \quad B(j,k) > 0$$

Therefore,

$$AB(i,k) = \sum_{l=1}^{q} A(i,l)B(l,k) \geq A(i,j)B(j,k) > 0$$

which shows that row i of AB is not a zero row. This being true for all $i = 1, 2, \ldots, p$, the matrix AB has no zero rows. ■

2.5.1 Multivariate regular variation

The notion of multivariate regular variation in $\mathbb{R}^d$ has been treated by many authors. See for example Resnick (1987), Embrechts et al. (1997), Bingham et al. (1989), Linskog (2004), Basrak et al. (2002a) and Jessen and Mikosch (2006). We only give some results used earlier and, if necessary, their proof.

Lemma 5. *Let $\{X_i\}_{i=1}^k$ and $\{Y_i\}_{i=1}^k$ be d–dimensional random vectors such that for all i, X_i is non–negative a.s. and Y_i has symmetric marginal distributions. If the vector*

$$Z^2 := \left(\delta(X_1X_1^T)^T, \delta(Y_1Y_1^T)^T, \ldots, \delta(X_kX_k^T)^T, \delta(Y_kY_k^T)^T\right)^T$$

is regularly varying with index 2α then the vector

$$Z := \left(X_1^T, Y_1^T, \ldots, X_k^T, Y_k^T\right)^T$$

is regularly varying with index α.

Proof. Denote by $\mathbb{S}_{dk-1}^+$ and by $\mathbb{S}$ the sets

$$\mathbb{S}_{dk-1}^+ = \left\{x \in \Re_+^{dk} \colon \|x\| = 1\right\}$$
$$\mathbb{S} = \left\{x \in \bigotimes_{i=1}^k \left(\mathbb{R}^d \cap supp(X_i) \cap supp(Y_i)\right) : \|x\| = 1\right\}$$

We need to show that there exists a Radon measure ν such that for any Borel–measurable subset S of $\mathbb{S}$ we have

$$\lim_{u\to\infty} \frac{\mathbb{P}\left[\|Z\| > ut, Z/\|Z\| \in S\right]}{\mathbb{P}\left[\|Z\| > u\right]} = \nu(S) \tag{2.17}$$

Consider first $S \subseteq \mathbb{S}$ with non–negative elements and the set of dk–dimensional vectors

$$I = \{(1, \pm 1, 1, \pm 1, \ldots, 1, \pm 1)\}.$$

For any given $\beta \in I$, let

$$S(\beta) = \{(x_1\beta_1, \ldots, x_{dk}\beta_{dk}) \colon x \in S\}.$$

By symmetry of the random vectors $\{Y_i\}_{i=1}^k$ we have that

$$\mathbb{P}\left[\|Z\| > ut, Z/\|Z\| \in S\right] = \mathbb{P}\left[\|Z\| > ut, Z/\|Z\| \in S(\beta)\right] \tag{2.18}$$

for all $\beta \in I$. This entails that

$$\frac{2^{dk}\mathbb{P}\left[\|Z\| > ut, Z/\|Z\| \in S\right]}{\mathbb{P}\left[\|Z\| > u\right]} = \frac{\mathbb{P}\left[\|Z^2\| > u^2t^2, Z^2/\|Z^2\| \in S^2\right]}{\mathbb{P}\left[\|Z^2\| > u^2\right]} \tag{2.19}$$

Letting $u \to \infty$ we find that equation (2.17) holds for S with non–negative elements. The measure ν is determined by the Radon measure of the squares, μ by the relation

$$\nu(S) = 2^{1-dk}\mu(S^2)$$

It follows from (2.18) that if $S \subseteq \mathbb{S}^+_{dk-1}(\beta)$ for some $\beta \in I$ then (2.17) also holds. To finish the proof, observe that for arbitrary $S \subseteq \mathbb{S}$ we may write

$$S = \bigcup_{\beta \in I}(S \cap \mathbb{S}^+_{dk-1}(\beta))$$

and use the previous argument. That the index of regular variation of Z is α is a consequence of equation (2.19). ■

Chapter 3

The Cholesky Factor GARCH

3.1 Introduction

Multivariate GARCH models are a way to describe the temporal dependence among the second order moments of financial returns capturing the nowadays widely accepted feature that financial volatilities move together over time across assets and markets. There are several ways to define multivariate GARCH models. We refer to Bauwens et al. (2006) for an excellent survey on the subject.

One approach to multivariate GARCH modelling is providing an autoregressive expression for the conditional autocovariance matrices $\{H_t\}$ of an $\mathbb{R}^d$–valued time series $\{X_t\}$. Observe that since H_t is intended to be a covariance matrix it must be nonnegative–definite. This is one of the difficulties of this approach.

Sufficient conditions have been given for specific processes guaranteeing the nonnegative–definiteness of H_t. For the VEC model these conditions can be found in Section 6.1 of Gouriéroux (1997). The BEKK model is studied in Engle and Kroner (1995). See also Stelzer for the relationships between these two models.

An alternative approach to guarantee that H_t is nonnegative–definite is to consider $H_t = L_t L_t^T$ where L_t is lower triangular and known as the Cholesky factor of H_t. In this case a GARCH type specification for the matrices L_t is given. Observe that the parameters involved in defining L_t need no restrictions for H_t to be nonnegative–definite. A GARCH model based on the Cholesky factor L_t can be found in Gallant and Tauchen, Kawakatsu

(2003) and Tsay (2005).

We will focus on the parametrization for the Cholesky factor L_t provided in Kawakatsu (2003) and defined by the equations

$$\begin{aligned}
\text{vech}(L_t) =& C + \sum_{i=1}^{p} A_i \text{vech}(L_{t-i}) + \sum_{j=1}^{q} B_j X_{t-j}, \\
H_t =& L_t L_t^T, \\
X_t =& L_t Z_t.
\end{aligned}$$

Where C, A_i, B_j are parameters and $\{Z_t\}$ is an i.i.d. sequence of random vectors. In specifying the model, the author restricts the parameters to identify L_t uniquely. He argues that the way to guarantee that L_t is uniquely defined is making its diagonal elements strictly positive and thus imposes the following identifiability condition:
Let r denote the rows in vech(L_t) corresponding to the diagonal elements in L_t. Then the Cholesky factor vech model is identified if

1. *rows r of A_i are positive in columns r and zero elsewhere,*

2. *rows r of B_j are all zeros.*

These conditions for the parameters imply that the diagonal elements of L_t do not depend on the values of the innovation vectors X_{t-j}. Since $H_t = L_t L_t^T$ the diagonal elements of H_t do depend on the values of X_{t-j} except for the entry $H_t(1,1)$, so that the identifiability requirements don't appear to be very restrictive. However, as is shown in Remarks 18 and 22 bellow, these conditions have important consequences.

We consider the Cholesky factor GARCH model restricting the parameters to have nonzero rows only and show that it can be uniquely defined in distribution by embedding it into a Stochastic Recurrence Equation of the type

$$Y_t = A(Z_t) Y_{t-1} + G, \tag{3.1}$$

Necessary and sufficient conditions for the existence and uniqueness of a stationary causal solution to the Cholesky Factor GARCH equations are derived from this representation using the well known results of Brandt (1986) and Bougerol and Picard (1992).

We then focus on this uniquely defined process and establish its regular variation. To study the tails of the finite dimensional distributions of

GARCH processes it is customary to use the representation (3.1) and Theorems 3 and 4 in Kesten (1973). See Mikosch and Stărică (2000) for the one–dimensional GARCH(1,1) case, Basrak et al. (2002b) for the one–dimensional GARCH(p,q) model and Fernández and Muriel (2008) for the multivariate CCC-GARCH(p,q) process. The reason for this method to work is that the matrices $A(Z_t)$ in the representation (3.1) for the mentioned models are almost surely non–negative. For the Cholesky factor GARCH model the matrices $A(Z_t)$ may be negative and thus the conventional approach can't be taken. Instead, we use Kesten's theorem 6 in Kesten (1973) which includes two extra hypotheses, namely,

1. the process defined by equation (3.1) with $G = 0$ is an open set irreducible Markov chain, and

2. the process defined by

$$W_0 = w, \qquad W_n = (I \otimes A(Z_n))W_{n-1}$$

   admits a Nummelin–small set.

To prove these conditions, we use a dynamical system approach to Markov chains analysis which is detailed in Chapter 7 of the book Meyn and Tweedie (1993).

Improving the results for other GARCH processes, we obtain the regular variation for the Cholesky factor GARCH model without assuming that the marginal distributions of the noise sequence are symmetric.

As is usual in stationary, mixing time series, regular variation is linked to extreme values and autocovariance function convergence. In accordance with the known results for the one–dimensional GARCH models (see Basrak et al. (2002b)) and the CCC-GARCH model (see Fernández and Muriel (2008)), we show that the asymptotic behavior of the sample autocovariance function of the Cholesky Factor GARCH(p,q),

$$\gamma_{n,X}(h) = n^{-1} \sum_{t=1}^{n-h} X_t X_{t+h}^T, \qquad h \geq 0,$$

depends on the so–called index of regular variation of the finite–dimensional distributions of the process. Denoting this index by κ, three cases are possible: If $\kappa \in (0, 2)$ we have no consistency but convergence to a $\kappa/2$–stable random variable. If $\kappa > 2$ we have consistency but a Central Limit Theorem is only achieved for $\kappa > 4$.

The paper is structured as follows: Notation and preliminary results are given in Section 3.2. Section 3.3 introduces the Cholesky Factor GARCH(p,q) model and characterizes the strictly stationary solutions to it. Simple conditions to guarantee geometrical β–mixing are also given in this Section. In Section 3.4 the regular variation for the finite dimensional distributions of the process is proven. Section 3.5 is devoted to the study of the extremes and the limit behaviour of the sample autocovariance function of the Cholesky Factor GARCH via the weak convergence of point processes generated by it.

3.2 Notation and Preliminary Results

We will denote the space of $d \times k$ matrices with real coefficients by $\mathcal{M}(d \times k)$. Given two matrices A and B, we denote by $A \otimes B$ their Kronecker (tensor) product. By vec and vech we denote the well–known column stack operations for general matrices and for symmetric matrices respectively. For the relationships between the vec operation and the Kronecker product, we refer to Chapter 4 of Horn and Johnson (1991). The operations vec and vech are related by

$\exists K$ such that for any given symmetric matrix A of dimensions $d \times d$, we have

$$\mathrm{vec}\left(()\, A\right) = K \mathrm{vech}(A). \tag{3.2}$$

A proof can be found in Appendix A of Boussama (1998).

Remark 15. *When vech is applied to lower triangular matrices,* (3.2) *remains true though for a different matrix $\hat{K}$. For ease of notation, we will still write K for this matrix.*

Given a square $d \times d$ matrix A we denote its spectral radius by $\varrho(A)$ and its minimum singular value by $\lambda(A)$. We refer to Horn and Johnson (1990). The column vectors of the matrix A will be denoted by $A[\,\cdot\,, i]$, $\;i = 1, 2, \ldots, d$. For a matrix $M \in \mathcal{M}(d \times k)$, the symbol $\|M\|$ denotes the matrix norm of M,

$$\|M\| = \sup\left\{\|Mx\| \;:\; \|x\| = 1\right\},$$

where $\|x\|$ is a vector norm in $\mathbb{R}^d$. When no comments are made on which norm is chosen for $\mathbb{R}^d$ it is because it is unimportant and we may use whichever norm.

Given a sequence of random $d \times d$ matrices $\{A_t\}$ defined on the same probability space, we denote their top Lyapunov exponent as

$$\gamma(A_t) = \inf \left\{ \frac{1}{t} \mathbb{E} \left[\ln(\|A_t A_{t-1} \dots A_1\|) \right], t \in \mathbb{N} \right\}.$$

We will consider d dimensional GARCH processes of orders (p, q) and use the notations

$$D := p\tfrac{d(d+1)}{2} + (q-1)d \qquad \text{and} \qquad d^* := \tfrac{d(d+1)}{2}$$

throughout the paper.

Given a vector $v \in \mathbb{R}^D$ we write $v(i), i = 1, 2, \dots, D$ for its components, and given a set of vectors $V = \{v_i, i \in I\}$ we write $\mathcal{S}(V)$ for the linear space generated by them. We will use the Zariski closure of this kind of sets and on the topic refer the reader to Boussama (1998), and Benedetti and Risler (1990).

Given a Markov chain $X = \{X_n\}$ with state space $\mathbb{R}^D$, we say that a Borel measurable set A is Nummelin–small if there exist a natural number $n > 0$ and a nontrivial measure ν_n defined on the Borel sets of $\mathbb{R}^D$ such that for any measurable set B

$$\mathbb{P}\left[X_n \in B \middle| X_0 = x\right] \geq \nu_n(B) \mathbb{1}\left(x \in A\right).$$

Finally, a Markov chain X is called T–chain if there exists a discrete random variable N independent of X and a substochastic kernel T such that for every measurable set A and for every $x \in \mathbb{R}^D$

$$\mathbb{P}\left[X_N \in A \middle| X_0 = x\right] \geq T(x, A) \qquad \text{and} \qquad T(x, \mathbb{R}^D) > 0,$$

and the function $T(\,\cdot\,, A)$ is lower semicontinuous. The reader is referred to Meyn and Tweedie (1993) for further details on Nummelin–small sets and T–chains..

3.3 Stationarity and Mixing

3.3.1 The Model

The dynamics for the Cholesky factor introduced in Kawakatsu (2003) define a GARCH type process which we will call the Cholesky factor GARCH(p,q)

(ChF–GARCH for short) and is defined as follows.

Definition 16. *Given a sequence of i.i.d. random vectors* $\{Z_t\}_{t\in\mathbb{Z}}$ *with mean vector* 0 *and identity covariance matrix, we say that the stochastic process* $\{X_t\}_{t\in\mathbb{Z}}$ *is a ChF–GARCH(p,q) if it satisfies the equations*

$$\begin{aligned} vech(L_t) =& C + \sum_{i=1}^{p} A_i vech(L_{t-i}) + \sum_{j=1}^{q} B_j X_{t-j}, \\ H_t =& L_t L_t^T, \\ X_t =& L_t Z_t. \end{aligned} \tag{3.3}$$

The vector C *is assumed nonzero to avoid the trivial solution* $X_t = 0$*. The matrices* A_i, B_j *are supposed to have nonzero rows and the matrix* L_t *is assumed lower triangular.*

Remark 16. *The matrix* H_t *is the covariance matrix of* X_t *conditional on* $\mathcal{F}_t = \sigma(X_s, s < t)$*, as can be easily shown. Furthermore,* H_t *is non–negative definite from the equality* $H_t = L_t L_t^T$*.*

We will embed this process into a Markovian process using Stochastic Recurrence Equations (SRE) as is done for the one–dimensional GARCH(p,q) in Basrak et al. (2002b) and for the multivariate CCC-GARCH(p,q) in Boussama (1998).

3.3.2 SRE representation and Stationarity

From Remark 15, there exists $K = K(d) \in \mathcal{M}(d^2 \times d^*)$ such that for any lower triangular matrix A we have $\text{vec}\,(()\,A) = K\text{vech}(A)$ which gives us

$$\begin{aligned} B_j X_{t-j} = B_j L_{t-j} Z_{t-j} =& (Z_{t-j}^T \otimes B_j)\text{vec}\,(()\,L_{t-j}) \\ =& (Z_{t-j}^T \otimes B_j) K \text{vech}(L_{t-j}) \end{aligned}$$

Substituting into equation (3.3) we get

$$\begin{aligned} \text{vech}(L_t) =& C + (A_1 + (Z_{t-1}^T \otimes B_1)K)\text{vech}(L_{t-1}) + \sum_{i=2}^{p} A_i \text{vech}(L_{t-i}) \\ &+ \sum_{j=2}^{q} B_j X_{t-j} \\ X_t =& (Z_t^T \otimes I) K \text{vech}(L_t) \end{aligned} \tag{3.4}$$

Letting

$$Y_t = (\text{vech}(L_{t+1})^T, \dots, \text{vech}(L_{t-p+2})^T, X_t^T, \dots, X_{t-q+2}^T)^T \tag{3.5}$$

we find that

$$Y_t = A(Z_t)Y_{t-1} + G \tag{3.6}$$

where

$$G = (C^T, 0, \dots, 0)^T$$

$$A(Z_t) = \begin{bmatrix} A_1 + (Z_t^T \otimes B_1)K & A_2 & \cdots & A_{p-1} & A_p & B_2 & B_3 & \cdots & B_q \\ I & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & I & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I & 0 & 0 & 0 & \cdots & 0 \\ (Z_t^T \otimes I)K & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & I & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & I & 0 \end{bmatrix} \tag{3.7}$$

Equation (3.6) is the SRE representation for the ChF–GARCH(p,q). The existence and uniqueness of a causal stationary solution to a given SRE are studied in Brandt (1986) and Bougerol and Picard (1992). We apply these results to the ChF–GARCH(p,q).

Proposition 13. *The SRE* (3.6) *is irreducible in the sense that if there exists* $I \subseteq \{1, 2, \dots, D\}$ *such that the linear space* V *generated by the canonical vectors* $\{e_i, i \in I\}$ *satisfies* $A(Z_1)V + G \subseteq V$*, then* $V = \mathbb{R}^D$*.*

Proof. The proof follows the lines of that of Theorem 5.4.3 in Boussama (1998) where the reader is referred for further details.
Let $\{e_i, i = 1, 2, \dots, D\}, \{f_i, i = 1, 2, \dots, d^*\}, \{g_i, i = 1, 2, \dots, d\}$ be the canonical bases in $\mathbb{R}^D, \mathbb{R}^{d^*}$ and $\mathbb{R}^d$ respectively. Let $I \subsetneq \{1, 2, \dots, D\}$ and $V = \mathcal{S}(\{e_i, i \in I\})$ and assume that $A(Z_t)V + G \subseteq V$. Since $0 \in V$ it follows that $G \in V$ and since $C > 0$ we get that

$$\{e_i, i = 1, 2, \dots, d^*\} \subseteq V$$

Furthermore, $G \in V$ also gives us $A(Z_t)V \subseteq V$ a.s. Computing $A(Z_t)e_i$ for $i = 1, 2, \ldots, d^*$ we see that

$$([(A_1 + (Z_t^T \otimes B_1)K)f_1]^T, f_1^T, 0, 0, \ldots, [(Z_t^T \otimes I)Kf_1]^T, 0, \ldots, 0)^T \in V$$

which is the same as

$$([(A_1 + (Z_t^T \otimes B_1)K)f_1]^T, 0, 0, \ldots, 0)^T + e_{d^*+i} + Z_{t,l(i)}e_{pd^*+j(i)+1} \in V$$

The particular values of $l(i)$ and $j(i)$ make no difference, so we don't provide them.
Since $(A_1 + (Z_t^T \otimes B_1)K)f_1]^T, 0, 0, \ldots, 0) \in V$ it follows that

$$e_{d^*+i} + Z_{t,l(i)}e_{pd^*+j(i)+1} \in V.$$

Observe that $Z_{t,l(i)}$ is not constant a.s. from which it follows that for at least two different values x, y we have

$$e_{d^*+i} + xe_{pd^*+j(i)+1} \in V \qquad \text{and} \qquad e_{d^*+i} + ye_{pd^*+j(i)+1} \in V.$$

Taking the difference between them, and rearranging the numbers $j(i)$ we get that

$$e_{pd^*+j} \in V \qquad \text{for } j = 1, 2, \ldots, d.$$

Finally, this implies that

$$e_{d^*+i} \in V \qquad \text{for } i = 1, 2, \ldots, d^*.$$

So far, we have shown that

$$\{e_i, i = 1, 2, \ldots, 2d*, pd^* + j, \;\; j = 1, 2, \ldots, d\} \subseteq V$$

Continuing to argue in this way, we next find that

$$\{e_i, i = 1, 2, \ldots, 3d*, pd^* + j, \;\; j = 1, 2, \ldots, 2d\} \subseteq V$$

Recursively, we finally get that

$$e_i \in V \qquad \text{for } i = 1, 2, \ldots, D$$

so that $V = \mathbb{R}^D$ as desired. ■

Corollary 7. *There exists a unique causal stationary solution to* (3.6) *if and only if the Top Lyapunov exponent associated to the sequence* $\{A(Z_t)\}$ *satisfies*

$$\gamma(A(Z_t)) < 0$$

Proof. By Theorem 1 in Brandt (1986) and Theorem 2.5 in Bougerol and Picard (1992) it follows that if

1. the matrices $\{A(Z_t)\}_{t\in\mathbb{Z}}$ are i.i.d.,
2. $\mathbb{E}\left[\log^+(\|G\|)\right] < \infty$,
3. $\mathbb{E}\left[\log^+\|A(Z_1)\|)\right] < \infty$, and
4. the SRE model is irreducible,

then the condition $\gamma(A(Z_t)) < 0$ is equivalent to the existence and uniqueness of a causal solution to the SRE (3.6).

Condition 1 is satisfied because $A(Z_t)$ depends only on Z_t and the sequence $\{Z_t\}_{t\in\mathbb{Z}}$ is i.i.d. Condition 2 is satisfied since G is a constant vector. Moreover, since $Z_t \in \mathcal{L}^2$, condition 3 is also satisfied. Finally, Proposition 13 shows that condition 4 is also met which gives us the result. ■

Remark 17. *The stationary causal solution is of the form*

$$Y_t = f(Z_s, s \le t)$$

so that it is ergodic in a measure theoretic way. See Chapter 9 of Kallenberg (2002)

Remark 18. *Observe that even if we do not impose the condition that the diagonal elements of* L_t *are strictly positive, we get the existence of a unique stationary causal solution to the ChF–GARCH equations. It is for this uniquely defined process that we will study the distributional properties in what follows.*
Furthermore, under the identifiability conditions imposed by Kawakatsu in Kawakatsu (2003), Proposition 13 doesn't hold. A simple example to this situation is given by the ChF–GARCH(1,1) in $\mathbb{R}^2$*. It is not hard to see that in this case the subspaces*

$$V = \{(0, \alpha x, 0)^T, \alpha \in \mathbb{R}\}, \qquad x \in \mathbb{R}$$

are invariant for Y_t*.*

Corollary 7 gives necessary and sufficient conditions for the ChF–GARCH model to have a unique causal stationary solution. However, the condition $\gamma(A(Z_t)) < 0$ is in general hard to prove for there is no closed known form for $\gamma(A(Z_t))$. For particular cases, we may use numerical analysis. See Gol′dsheĭd (1991) on this subject. The following Proposition provides a link between the coefficients A_i defining the ChF–GARCH and the negativity of the top Lyapunov exponent $\gamma(A(Z_t))$.

From now on we will write $A(0)$ for the matrix $A(Z_t)$ in (3.7) evaluated at $Z_t = 0$.

Proposition 14. *Consider a ChF–GARCH(p,q) with matrix parameters* (A_i, B_j). *If* $\varrho(\sum_{i=1}^p A_i) < 1$ *then* $\gamma(A(Z_t)) < 0$.

Proof. By Proposition 4.1 in page 118 of de Saporta it is enough to show that

$$\varrho\left(\mathbb{E}\left[A(Z_1) \otimes A(Z_1)^T\right]\right) < 1 \tag{3.8}$$

Since $\mathbb{E}[A(Z_t)] = A(0)$, it is not hard to see, by the definition of the tensor product $\otimes$, that

$$\mathbb{E}\left[A(Z_1) \otimes A(Z_1)^T\right] = A(0) \otimes A(0)^T.$$

Now, the proper values of the matrix $A(0) \otimes A(0)^T$ are

$$\{\lambda\mu : \lambda \text{ is a proper value of } A(0), \text{ and } \mu \text{ is a proper value of } A(0)^T\}.$$

Since the proper values of $A(0)^T$ are the same as those of $A(0)$, it follows that every proper value of $A(0) \otimes A(0)^T$ is the product of two proper values of $A(0)$. Consequently it is enough to have $\varrho(A(0)) < 1$.

This is indeed our case, because by Lemma A.1.1 in Boussama (1998) the assumption $\varrho(\sum_{i=1}^p A_i) < 1$ is equivalent to $\varrho(A(0)) < 1$. ■

Remark 19. *The condition* $\varrho(\sum A_i) < 1$ *is sufficient for* $\gamma(A(Z_t)) > 0$ *but not necessary. Therefore a stationary ChF–GARCH process may exist with* $\varrho(\sum A_i) \geq 1$.

3.3.3 Mixing

To study the mixing properties of the ChF–GARCH(p,q) process we will assume that

$\mathcal{H}_0$: $\varrho(\sum_{i=1}^{p} A_i) < 1$

$\mathcal{H}_1$: The random vectors $\{Z_t\}$ have a density f with respect to Lebesgue measure in $\mathbb{R}^d$ such that $E = \{x\colon f(x) > 0\}$ contains a neighborhood of $\{0\}$.

Theorem 22. *Under the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, any solution $\{Y_t\}$ to the SRE* (3.6) *is a positively Harris recurrent, geometrically ergodic Markov chain. Furthermore, the stationary solution is geometrically β–mixing.*

Proof. For $y \in \mathbb{R}_+^D$ define

$$Y_0^y = y, \qquad Y_n^y = A(0)Y_{n-1}^y + G.$$

From hypothesis $\mathcal{H}_0$ and Lemma A.1.1 in Boussama (1998) it follows that $\varrho(A(0)) < 1$ so that

$$Y_n^y \xrightarrow{n\to\infty} \sum_{j=0}^{\infty} A(0)^j G := T \qquad \text{for all } y \in \mathbb{R}_+^D.$$

Furthermore, T satisfies the fixed point equation

$$T = A(0)T + G.$$

Following Lemma 3.4.3 in page 54 of Boussama (1998) we find that the stationary process $\{Y_t\}$ satisfies the Foster–Lyapunov condition
"$\exists k \in \mathbb{N}$, a compact set K in $\mathbb{R}^D$, and $\alpha < 1, s < 1$ such that for $y \in \mathbb{R}_+^D$

$$\mathbb{E}\left[V(Y_k)\middle|Y_0 = y\right] \le \alpha^k \left(V(y) + b\mathbb{1}\left(y \in K\right)\right),$$

where V is the function defined by $V(y) = 1 + \|y\|^s$."
From the mixing criteria in Mokkadem (1990) (See also to Theorem 2.3.5 in page 41 of Boussama (1998) for a restatement of these criteria), it follows that $\{Y_t\}$ is a positively Harris recurrent, geometrically ergodic Markov chain and that the stationary solution is β–mixing. As is shown in Doukhan (1994), the β–mixing rate is the same as the ergodicity rate, so that the stationary process $\{Y_t\}$ is geometrically β–mixing which completes the proof. ∎

3.4 Regular Variation

In this section we give sufficient conditions for the stationary Cholesky Factor GARCH(p,q) to have regularly varying finite–dimensional distributions. We do so by using the SRE representation (3.6) and Kesten's theorem. We remark here that since the matrices $\{A(Z_t)\}$ in (3.7) may have negative entries, we can't use Theorems 3 and 4 in Kesten (1973) as is done for the one–dimensional case in Basrak et al. (2002b) and for the CCC-GARCH in Fernández and Muriel (2008). We will instead use Kesten's Theorem 6.

As mentioned in the introduction, we will take a dynamical system approach which is detailed in Chapter 7 of the book Meyn and Tweedie (1993). This will be done using the decomposition

$$A(Z_t) = A(0) + \sum_{i=1}^{d} Z_{t,j(i)} C_i + \sum_{i=1}^{d} Z_{t,k(i)} \widetilde{C}_i \tag{3.9}$$

where the columns of C_i and $\widetilde{C}_i$ are all, but one, equal to 0. Denote the nonzero column vector of C_i and $\widetilde{C}_i$ by V_i and $\widetilde{V}_i$ respectively. The particular values of $j(i)$ and $k(i)$ in (3.9) will be of no relevance for which we don't provide them.

Proposition 15. *Under $\mathcal{H}_0$, the deterministic model defined by*

$$x_0 = x, \qquad x_n = A(u_n)x_{n-1}, \qquad u_n \in \mathbb{R}^d$$

is controllable in the sense that the sets

$$A_+(x) = \bigcup_{k\in\mathbb{N}} \{y\colon y = A(Z_k)\dots A(Z_1)x, Z_t \in \mathbb{R}^d\}, \qquad x \neq 0.$$

have non–empty interior. Furthermore, there exists $r > 0$ such that the affine space generated by the vectors

$$\{\{V_i\}_i, \{\widetilde{V}_i\}_i, \{A(0)V_i\}_i, \{A(0)\widetilde{V}_i\}_i, \dots \{A(0)^r V_i\}_i, \{A(0)^r \widetilde{V}_i\}_i\}$$

is minimal for $\{x_n\}$.

Proof. To show that the model is controllable, let W_1 be the Zariski closure of the set

$$\{A(u_t)x, u_t \in \mathbb{R}^d\}.$$

We claim that $W_1 = A(0)x + E_1$ where E_1 is the linear subspace of $\mathbb{R}^D$ generated by the vectors

$$\{V_1, \ldots, V_d, \widetilde{V}_1, \ldots, \widetilde{V}_d\}$$

To see this, use (3.9) to obtain

$$\begin{aligned} A(Z_t)x =& A(0)x + \sum_{i=1}^{d} Z_t(j(i))C_i x + \sum_{i=1}^{d} Z_t(k(i))\widetilde{C}_i x \\ =& A(0)x + \sum_{i=1}^{d} Z_t(j(i))x_i V_i + \sum_{i=1}^{d} Z_t(k(i))x_i \widetilde{V}_i \end{aligned}$$

Therefore W_1 contains an open set of $A(0)x + E_1$. Consequently, since $W_1 \subseteq A(0)x + E_1$, it follows that $W_1 = A(0)x + E_1$.
Proceeding recursively, define W_k as the Zariski closure of the set

$$\{A(u_t)v, v \in W_{k-1}, u_t \in \mathbb{R}^d\}$$

and E_k as the linear subspace of $\mathbb{R}^d$ generated by the vectors

$$\{V_1, \ldots, V_d, A(0)V_1, \ldots, A(0)V_d, \ldots, A(0)^k V_1, \ldots, A(0)^k V_d, \ldots, A(0)^k \widetilde{V}_1, \ldots, A(0)^k \widetilde{V}_d\}.$$

Following the calculations in page 85 of Boussama (1998) it can be proven that for all $k \in \mathbb{N}$ the equality $W_k = A(0)^k x + E_k$ holds. This shows that for $x \neq 0$ the set $A_+(x)$ has nonempty interior, and thus that the model $\{x_n\}$ is controllable.

To complete the proof of the Proposition, observe that the sequence of linear subspaces $\{E_k\}$ is strictly increasing in $\mathbb{R}^D$ so that it is constant from a certain rank on, say $r \leq D$. Define

$$W := \lim_{k\to\infty} E_k = E_r.$$

By assumption $\mathcal{H}_0$ and Lemma A.1.1 in Boussama (1998) we have that $\varrho(A(0)) < 1$ which shows that $A(0)^k \to 0$, thus the topological closure of $A_+(x)$ satisfies

$$\overline{A_+(x)} = \lim_{k\to\infty} W_k = W \qquad \forall x \neq 0$$

. By Proposition 7.2.3 in page 162 of Meyn and Tweedie (1993) the set W is minimal for $\{x_n\}$ which yields the result. ■

Corollary 8. *If the distribution function of the i.i.d. sequence of random vectors* $\{Z_t\}$ *has a density* f *on* $\mathbb{R}^d$ *which is lower semicontinuous and whose support is* $\mathbb{R}^d$*, then the Markov process*

$$X_0 = x, \qquad X_n = A(Z_n)X_{n-1}$$

restricted to W *is an open set irreducible* T*–chain.*

Proof. This is an immediate consequence of Proposition 15 together with Theorem 7.2.4 in page 163 of Meyn and Tweedie (1993). ■

We now study the regular variation of the ChF–GARCH process using the following hypotheses:

$\mathcal{H}_2$: The distribution of the vectors $\{Z_t\}$ admits a continuous and everywhere positive density with respect to Lebesgue measure.

$\mathcal{H}_3$: There exists $h_0 \in (0, \infty]$ such that

$$\mathbb{E}\left[\mid Z_t(i)\mid^h\right] < \infty \qquad \text{if } h < h_0,$$
$$\lim_{h\to h_0} \mathbb{E}\left[\mid Z_t(i)\mid^h\right] = \infty.$$

$\mathcal{H}_4$: Hypothesis $\mathcal{H}_0$ is satisfied and there exists $k \in \mathbb{N}$ such that the vectors

$$\{\{V_i\}_i, \{\widetilde{V}_i\}_i, \{A(0)V_i\}_i, \{A(0)\widetilde{V}_i\}_i, \ldots \{A(0)^r V_i\}_i, \{A(0)^r \widetilde{V}_i\}_i\}$$

generate the space $\mathbb{R}^D$.

Theorem 23. *Let* $\{X_t\}$ *be a stationary ChF–GARCH(p,q) process with SRE representation* $\{Y_t\}$ *as given in* (3.6)*. Under* $\mathcal{H}_2, \mathcal{H}_3$ *and* $\mathcal{H}_4$ *there exists* $\kappa > 0$ *such that*

$$\forall x \in \mathbb{R}^D \setminus \{0\} \qquad \lim_{u\to\infty} u^\kappa \mathbb{P}\left[\langle x, Y_1\rangle > u\right] = w(x)$$

exists and $w(x) > 0$*. Furthermore, if* κ *is not an even integer, then the finite–dimensional distributions of* $\{X_t\}$ *are regularly varying.*

Proof. Let m be an arbitrary natural number and consider the sequence $\{Y_{tm}\}_{t\in\mathbb{Z}}$ which, as shown in Basrak et al. (2002b), satisfies the stochastic recurrence equation

$$Y_{tm} = \hat{A}_t^{(m)} Y_{(t-1)m} + \hat{B}_t^{(m)},$$

for

$$\begin{aligned}\hat{A}_t^{(m)} =& A_{tm} A_{tm-1} \ldots A_{tm-m+1},\\ \hat{B}_t^{(m)} =& \sum_{k=1}^{m-1} A_{tm} A_{tm-1} \ldots A_{tm-k+1} G.\end{aligned}$$

It is for this sequence that we will verify Kesten's hypotheses by choosing a suitable value for m. For ease of presentation, the required hypotheses are restated one by one with its respective proof.

1. *Both* $\mathbb{E}\left[\ln\left(\left\|\hat{A}_t^{(m)}\right\|\right)\right]$ *and* $\mathbb{E}\left[\ln\left(\left\|\hat{B}_t^{(m)}\right\|\right)\right]$ *are finite, and there exists* $\varepsilon > 0$ *such that* $\mathbb{E}\left[\left\|\hat{A}_t(m)\right\|^{\varepsilon}\right] < 1$.

In Basrak et al. (2002b) these properties are shown to be satisfied for $m \geq M_0$ with M_0 a given natural number in the one–dimensional case. The argument rests on the negativity of the Lyapunov exponent associated to the SRE representation of the process. Corollary 7 tells us that stationarity of the ChF–GARCH is equivalent to the fact that the top Lyapunov exponent of the sequence $\{A(Z_t)\}$ in (3.7) is strictly negative. Therefore, taking $m \geq M_0$, we can follow exactly the same lines as in the one–dimensional case to prove this Kesten's hypothesis.

2. *The matrix* $\hat{A}_t^{(m)}$ is nonsingular a.s.

Since the matrices A_i have no zero rows, it is not difficult to see that for all $t \in \mathbb{Z}$ we have $Rank(A_t) = D$ a.s. from which it follows that the matrix A_t is a.s. nonsingular. $\hat{A}_t^{(m)}$ is thus nonsingular, being a product of nonsingular matrices.

3. *For every open set* $U \subseteq \mathbb{S}_{d-1}$ and $x \in \mathbb{S}_{d-1}$ there exists $n \in \mathbb{N}$ such

that

$$\mathbb{P}\left[\frac{\hat{A}_n^{(m)}\dots\hat{A}_1^{(m)}x}{\left\|\hat{A}_n^{(m)}\dots\hat{A}_1^{(m)}x\right\|}\in U\right]>0. \tag{3.10}$$

By Corollary 8 and the continuity assumption on the density of the random vectors $\{Z_t\}$ in $\mathcal{H}_2$, it follows that the Markov Chain $V_n = \hat{A}_n^{(m)}V_{n-1}$ restricted to W is an open set irreducible T–chain.

Hypothesis $\mathcal{H}_4$ tells us that $W=\mathbb{R}^D$. This implies (3.10) by the equality

$$\mathbb{P}\left[\frac{\hat{A}_n^{(m)}\dots\hat{A}_1^{(m)}x}{\left\|\hat{A}_n^{(m)}\dots\hat{A}_1^{(m)}x\right\|}\in U\right]=\mathbb{P}\left[V_n\in\{y\colon\ \|y\|>0, y/\|y\|\in U\right]$$

4. *There exists* $n\in\mathbb{N}$, a cube $C\subseteq\mathbb{R}^{D^2}$ and $\gamma_0>0$ such that the density of the distribution of $\operatorname{vec}\left(()\,\hat{A}_n^{(m)}\cdots\hat{A}_1^{(m)}\right)$ with respect to Lebesgue measure is bounded from bellow by γ_0.

Let $U_n=\operatorname{vec}\left(()\,A(Z_n)\cdots A(Z_1)\right)$. Then we have the Markovian relationship

$$U_n=(I\otimes A(Z_n))U_{n-1}.$$

Observe that Kesten's hypothesis 4, which we wish to prove, is equivalent to the existence of a Nummelin–small cube for the chain $\{U_n\}$ (See Chapter 5 of Meyn and Tweedie (1993)).

Arguing as we did in the proof to Proposition 15 and using again our hypothesis $\mathcal{H}_4$ it is not hard to see that $\{U_n\}$ is an open set irreducible T–chain on $\mathbb{R}^{D^2}$. From Theorem 6.2.5 in page 138 of Meyn and Tweedie (1993) we conclude that every compact set is Nummelin–small. It follows that there exists a Nummelin-small cube C for the chain $\{U_n\}$ as desired.

5. *The set* $\{\ln(\|a_n\dots a_1\|): n\geq 1, a_n\dots a_1>0,\ a_i\in \mathit{supp}(\mathbb{P})\}$ *generates a dense group in* $\mathbb{R}$, where $\mathbb{P}$ stands for the distribution of $\hat{A}_1^{(m)}$.

If $m\geq\max(p,q)$, then the entries of the matrix $\hat{A}_1^m$ are multilinear forms of the random variables $\{Z_t(i)\}$, which can be proven as is done for the CCC-GARCH process in Fernández and Muriel (2008). Take $m\geq\max(M_0,p,q)$ and from now on, write $\hat{A}_t$ for $\hat{A}_t^{(m)}$.

The support of the random variables $\{Z_t(i)\}$ is $\mathbb{R}$, thus, by continuity of multilinear forms, the support of $\hat{A}_1$ is connected. By continuity of the norm, the same can be said about the support of $\left\|\hat{A}_1\right\|$. Therefore, the support of $\ln\left(\left\|\hat{A}_1\right\|\right)$ contains an interval, which yields the desired property.

6. *For every fixed vector r we have* $\mathbb{P}\left[G = (I - \hat{A}_1)r\right] < 1$

This follows from assumption $\mathcal{H}_2$. Since the random vectors Z_t have a density with $\mathbb{R}^d$ as support, it is straightforward that

$$\mathbb{P}\left[\hat{A}_1 x = w\right] < 1$$

for any fixed vectors x, w, in particular for $x = r$ and $w = r - G$ which gives the desired result.

7. *There exists* $\kappa_0 > 0$ such that

$$\mathbb{E}\left[\left\|\hat{A}_1\right\|^{\kappa_0} \ln\left(\left\|\hat{A}_1\right\|\right)\right] < \infty \qquad \text{and} \qquad \mathbb{E}\left[(\lambda(\hat{A}_1))^{\kappa_0}\right] \geq 1$$

To prove that $\mathbb{E}\left[(\lambda(\hat{A}_1))^{\kappa_0}\right] \geq 1$ we will focus on the first block, $\hat{A}_1[1,1]$ of the matrix $\hat{A}_1$. By Corollary 3.1.3 in Horn and Johnson (1991), page 149 it is enough to obtain the bound for $\lambda(\hat{A}_1[1,1])$.
From the Courant–Fischer Theorem for singular values, Theorem 3.1.2 in Horn and Johnson (1991) page 148, we need only show that $\mathbb{E}\left[\left\|\hat{A}_1[1,1]x\right\|_2^{\kappa_0}\right] \geq 1$ for $\|x\| = 1$.
As said earlier, all the entries in the block $\hat{A}_1[1,1]$ are multilinear forms of the random variables $\{Z_t(i)\}$, thus, by assumption $\mathcal{H}_3$ there exists $\kappa_0 > 0$, close enough to h_0, such that

$$\mathbb{E}\left[\left\|\hat{A}_1[1,1]x\right\|_2^{\kappa_0}\right] \geq 1.$$

The proof that $\mathbb{E}\left[\|A_1\|^{\kappa_0} \ln(\|A_1\|)\right] < \infty$ is done as in Basrak et al. (2002b) for the one–dimensional GARCH process.

Having proved all of Kesten's hypotheses we know that there exists $\kappa > 0$ and $w(x) > 0$ such that

$$\forall x \in \mathbb{R}^D \setminus \{0\}, \quad \lim_{u\to\infty} u^\kappa \mathbb{P}\left[\langle x, Y_1\rangle > u\right] = w(x),$$

proving the first part of the Theorem. To conclude the proof observe that if κ is not an even integer, then by Theorem 1.1 in page 910 of Basrak et al. (2002a) the random vector Y_1 is regularly varying. By Corollary 2.7 in page 100 of Basrak et al. (2002b) we find that the finite dimensional distributions of $\{Y_t\}$ are regularly varying. Since $\{X_t\}$ is embedded in $\{Y_t\}$ it follows that the finite dimensional distributions of $\{X_t\}$ are also regularly varying. The index of regular variation for these distributions is κ. ■

Remark 20. *The regular variation of the random vector Y_1 implies the existence of a sequence of real numbers*

$$0 < a_n \nearrow \infty,$$

and a Radon measure μ defined on $\mathcal{B}(\mathbb{R}^d)$, such that

$$\mathbb{P}\left[a_n^{-1} Y_1 \in \cdot\,\right] \xrightarrow{v} \mu(\,\cdot\,),$$

where "$\xrightarrow{v}$" stands for vague convergence (See Chapter 5 of Resnick (1987)).

Remark 21. *The vectors $\{X_t, vech(L_t)\}$ are also embedded into $\{Y_t\}$. The proof of Theorem 23 tells us that $\{X_t, vech(L_t)\}$ also has regularly varying finite dimensional distributions. In particular, the sequence $\{vech(L_t)\}$ shares this property.*

Remark 22. *Hypothesis $\mathcal{H}_4$ is not compatible with the identification rules proposed in Kawakatsu (2003). These identification rules restrict the parameters to ensure that the diagonal elements of L_t are strictly positive. As mentioned earlier in Remark 18, these identification rules are not important for our analysis since uniqueness of the stationary solution is guaranteed.*

3.5 Point Process Convergence

In this section we study the weak convergence of Point Processes

$$N_n(\,\cdot\,) = \sum_{t=1}^{n} \varepsilon_{a_n^{-1} X_t}(\,\cdot\,) \tag{3.11}$$

where $\varepsilon_x(\cdot)$ is the Dirac measure at x, $\{X_t\}$ is the stationary solution to the Cholesky Factor GARCH equations provided in Definition 16 and $\{a_n\}$ is the sequence given in Remark 20. Point Process convergence is the basis for the study of the extremes of the ChF–GARCH as well as of the asymptotics of the sample autocovariance function. The following result is based on the work Davis and Mikosch (1998).

Proposition 16. *Let $\{X_t\}$ be a stationary ChF–GARCH(p,q) process. Assume all hypotheses in Theorem 23 are satisfied. Then, there exists a point process N such that $N_n \to N$ in distribution and N is identical in law to the point process*

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \varepsilon_{P_i Q_{i,j}}(\cdot)$$

where $\sum_{i=1}^{\infty} \varepsilon_{P_i}(\cdot)$ is a Poisson process with intensity measure

$$\nu(dy) = \theta\alpha y^{-\alpha-1} \mathbb{1}\,(y>0)\,dy$$

for some $0<\theta<1$ and independent of the sequence of i.i.d. point processes $\sum_{j=1}^{\infty} \varepsilon_{Q_{i,j}}(\cdot), i \geq 1$.

Proof. By Theorem 23 we know that the finite–dimensional distributions of the process $\{X_t\}$ are regularly varying. By the geometrical β–mixing shown in Theorem 22, we also know that there exists a sequence of positive integers r_n such that $r_n \to \infty, k_n = [n/r_n] \to \infty$ as $n \to \infty$ and

$$\mathbb{E}\left[\exp\left\{-\sum_{t=1}^{n} f(X_t a_n^{-1})\right\}\right] - \left(\mathbb{E}\left[\exp\left\{-\sum_{t=1}^{r_n} f(X_t a_n^{-1})\right\}\right]\right)^{k_n} \to 0,$$

as $n \to \infty$ for all simple, non–negative, measurable function f. See Basrak et al. (2002b).
Finally, it can be proved as is done in Basrak et al. (2002b) for the one dimensional case, that for all $y > 0$

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\left[\max_{k\leq|\,t\,|\leq r_n} \|X_t\| > a_n y \,\Big|\, \|X_0\| > a_n y\right] = 0.$$

These conditions enable us to apply Theorem 2.8 in Davis and Mikosch (1998) to complete the proof. ■

3.5.1 Extreme Values

In this section we study the extreme values of the ChF–GARCH process. The main focus is on the asymptotic behavior of the norm–maximum and the componentwise–maximum of the process, which are defined as follows.

Definition 17. *The norm–maximum of the ChF–GARCH process* $\{X_t\}$ *is defined by*

$$M_n^{\|X\|} = \max\{\|X_t\|\,,\ t \le n\}$$

and its componentwise–maximum M_n *by the equations*

$$M_n(i) = \max\{X_t(i),\ t \le n\}, \quad \textit{for } i = 1, 2 \ldots, d$$
$$M_n = (M_n(1), M_n(2), \ldots, M_n(d)).$$

The interest lies in the description of all distribution functions such that

$$\mathbb{P}\left[a_n^{-1} M_n^{\|X\|} \le x\right] \to G(x), \tag{3.12}$$
$$\mathbb{P}\left[\left(a_n^{-1} M_n(1) \le x(1), \ldots, a_n^{-1} M_n(d) \le x(d)\right)\right] \to H(x(1), \ldots, x(d)), \tag{3.13}$$

where $0 < a_n \to \infty$ is the sequence of regular variation of the ChF–GARCH process. Of course, G in relation (3.12) is a real valued distribution function while H in relation (3.13) is $\mathbb{R}^d$–valued. To simplify notation, the limit (3.13) will be written

$$\mathbb{P}\left[a_n^{-1} M_n \le x\right] \to H(x), \quad \text{for } x \in \mathbb{R}^d.$$

As said earlier, Proposition 16 is the basis for the study of the extreme values for the ChF–GARCH process. Due to the regular variation proven in Theorem 23, the sequence $\{a_n\}$ in Proposition 16 satisfies $a_n \sim n^{1/\kappa}$, with $\kappa > 0$ the index of regular variation. We will, as is customary, take $a_n = n^{1/\kappa}$.

Theorem 24. *Let* $\{X_t\}_{t\in\mathbb{Z}}$ *be a ChF–GARCH(p,q) process satisfying the hypotheses of Theorem 23 and let* $\kappa > 0$ *be its index of regular variation.*

1. *The norm–maximum of* $\{X_t\}_{t\in\mathbb{Z}}$ *satisfies*

$$\mathbb{P}\left[n^{-1/\kappa} M_n^{\|X_t\|} \le x\right] \to \exp\{-\theta x^{-\kappa}\}\mathbb{1}\,(x > 0)\,,$$

where $0 < \theta < 1$ *is the extremal index of* $\|X_t\|$.

2. *The componentwise–maximum of the sequence* $\{X_t\}_{t\in\mathbb{Z}}$ *satisfies*

$$\mathbb{P}\left[n^{-1/\kappa}M_n \le x\right] \to \exp\{-\Lambda\left(\{\mu\colon \mu(B_x) > 0\}\right)\},$$

where Λ *is the canonical measure of the point process* N *and*

$$B_x = (-\infty, x]^c \quad \text{for any fixed } x \in \mathbb{R}_+^d.$$

Proof. The density $\nu(dy)$ in the description of the point process N given in Proposition 16 is the asymptotic density of the extremes of the sequence $\{\|X_t\|\}$, see Davis and Mikosch (1998, 2006). Part 1 is thus proved.
To prove part 2, first observe that

$$\mathbb{P}\left[n^{-1/\kappa}M_n \le x\right] = \mathbb{P}\left[N_n(B_x) = 0\right]. \tag{3.14}$$

Since the set B_x is relatively compact, Proposition 16 gives us the convergence

$$\mathbb{P}\left[N_n(B_x) = 0\right] \xrightarrow[n\to\infty]{} \mathbb{P}\left[N(B_x) = 0\right]. \tag{3.15}$$

By dominated convergence, we have

$$\mathbb{P}\left[N(B_x) = 0\right] = \mathbb{E}\left[\lim_{t\to\infty} e^{-N(t\mathbb{1}(B_x))}\right] = \lim_{t\to\infty} \mathbb{E}\left[e^{-N(t\mathbb{1}(B_x))}\right].$$

Since Λ is the canonical measure of N (see Kallenberg (1969)), this expectation can be calculated as

$$\mathbb{E}\left[e^{-N(t\mathbb{1}(B_x))}\right] = \exp\left\{-\int (1 - e^{-\mu(t\mathbb{1}(B_x))})\Lambda(d\mu)\right\}.$$

Because of this equality and, again, dominated convergence we now have

$$\begin{aligned}\mathbb{P}\left[N(B_x) = 0\right] &= \lim_{t\to\infty} \exp\left\{-\int (1 - e^{-\mu(t\mathbb{1}(B_x))})\Lambda(d\mu)\right\}\\ &= \exp\left\{-\int (1 - \mathbb{1}\left(\mu(B_x) = 0\right))\Lambda(d\mu)\right\},\end{aligned}$$

which shows that

$$\mathbb{P}\left[N(B_x) = 0\right] = \exp\left\{-\int \mathbb{1}\left(\mu(B_x) > 0\right)\Lambda(d\mu)\right\}. \tag{3.16}$$

Putting equations (3.14), (3.15) and (3.16) together, it is seen that

$$\mathbb{P}\left[n^{-1/\kappa}M_n \le x\right] \xrightarrow[n\to\infty]{} \exp\{-\Lambda\left(\{\mu\colon \mu(B_x) > 0\}\right)\}.$$

∎

Remark 23. *It is worth mentioning that*

1. *There's no closed known expression for calculating the extremal index θ of the sequence $\|X_t\|$. We only know that $0 < \theta < 1$, which means the sequence $\|X_t\|$ has clusters with mean size $1/\theta > 1$.*

2. *The measure λ provides a theoretical expression for the limiting distribution of the normalized componentwise–maximum of the ChF–GARCH process $\{X_t\}_{t\in\mathbb{Z}}$. However, this expression is of no use for probability computations.*

An alternative expression for the domain of attraction of the componentwise –maximum for the ChF-GARCH process can be found by following section 6 in Perfekt (1997) to obtain:

Proposition 17. *Let $\{Y_t\}$ be the stationary solution to the ChF–GARCH(p,q) SRE representation (3.6). Let ν be a measure such that*

$$n\mathbb{P}\left[a_n^{-1}Y_1 \in \cdot\,\right] \xrightarrow{v} \nu(\,\cdot\,).$$

Assume that ν satisfies

$$c_i := \nu\left(\{y\colon y_i > 1\}\right) > 0, \qquad \text{for all } i = 1, 2, \ldots, d^2(p+q-1).$$

Then, for $x \in (0,\infty)^d$ we have

$$\mathbb{P}\left[a_n^{-1}M_n \le x\right] \xrightarrow[n\to\infty]{} \exp\left\{-\int_{[0,x]^c} \mathbb{P}\left[\prod_{i=1}^{j} A_i y \le x;\ j \ge 1\right] \nu(dy)\right\}.$$

Remark 24. *In view of the definition of the multivariate extremal index, see Nandagopalan (1994), and the result in Proposition 17, the multivariate extremal index of the stationary process $\{Y_t\}$ may now be written as*

$$\theta(x) = \frac{\int_{[0,x]^c} \mathbb{P}\left[\prod_{i=1}^{j} A_i y \le x;\ j \ge 1\right] \nu(dy)}{\nu([0,x]^c)}.$$

This expression enables us to write the marginal extremal index for the ith component of Y_t as

$$\theta_i = \frac{\int_{\{y\colon y_i>1\}} \mathbb{P}\left[\prod_{i=1}^{j} A_i y \le x;\ j \ge 1\right] \nu(dy)}{c_i}.$$

These results are analogous to the ones found in Mikosch and Stărică (2000) for the one–dimensional GARCH(p,q) process and in Fernández and Muriel (2008) for the CCC-GARCH(p,q) process.

3.5.2 Sample Autocovariance Function

Using Proposition 16 we can describe the asymptotics of the sample autocovariance function for the stationary ChF–GARCH defined as

$$\gamma_{n,X}(h) = n^{-1} \sum_{t=1}^{n-h} X_t X_{t+h}^T, \qquad h \geq 0.$$

One expects this sample function to converge a.s. to

$$\gamma_X(h) = \mathbb{E}\left[X_0 X_h^T\right],$$

which is the autocovariance function of the process, by Remark 17 and the ergodic theorem. Furthermore, having this convergence the Central Limit Theorem is also expected to apply, that is, a convergence of $\gamma_{n,X}(h) - \gamma_X(h)$ to a Normal distribution at rate $n^{1/2}$ is expected.

For a regularly varying ChF–GARCH process, these limit relations need not be true. The reason is that the moments of the stationary distribution of the ChF–GARCH process depend on the index of regular variation. The relationship between the index of regular variation and the moments of a random variable is well known and can be found in Resnick (1987), Embrechts et al. (1997) or Bingham et al. (1989) among others.

In this section we will use the hypothesis

$\mathcal{H}_5$: The marginal densities of the distribution of the random vectors $\{Z_t\}$ have symmetric marginals.

We have the following result, which is the generalization of Theorem 3.6 in Basrak et al. (2002b) for the autocovariance function of the ChF–GARCH process.

Theorem 25. *Let $\{X_t\}$ be a stationary ChF-GARCH(p,q) process satisfying all the conditions in Theorem 23 and assume that hypothesis $\mathcal{H}_5$ is also satisfied. Let κ denote the index of regular variation of $\{X_t\}$, then*

1. *If $\kappa \in (0, 2)$, then*

$$(n^{1-2/\kappa} vec\,(()\,\gamma_{n,X}(h)))_{h=0,1,\ldots,m} \xrightarrow{d} (V_h)_{h=0,1,\ldots,m} \tag{3.17}$$

where $(V_0, V_1, \ldots, V_m)$ is jointly $\kappa/2$ stable in $\mathbb{R}^{dm}$.

2. *If* $\kappa \in (2,4)$, *then*

$$(n^{1-2/\kappa}(vec\left(()\,\gamma_{n,X}(h) - \gamma_X(h)\right)))_{h=0,1,\ldots,m} \xrightarrow{d} (V_h)_{h=0,1,\ldots,m} \qquad (3.18)$$

were $(V_0, V_1, \ldots, V_m)$ *is jointly* $\kappa/2$ *stable in* $\mathbb{R}^{dm}$.

3. *If* $\kappa > 4$ *then equation* (3.18) *holds with normalization* $n^{1/2}$, *where* V *is multivariate normal with mean zero.*

Proof. 1. If $\kappa \in (0,2)$ the result is immediate from Proposition 16 and Theorem 3.5 in Davis and Mikosch (1998).

2. For $\kappa \in (2,4)$ it is necessary to show that the following condition holds.

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty} var\left(n^{-2/\alpha}\sum_{t=1}^{n-h} X_t(i)X_{t+h}(j)\mathbb{1}\left(|\,X_t(i)X_{t+h}(j)\,| \le n^{2/\kappa}\varepsilon\right)\right) = 0$$

for all $i, j = 1, 2, \ldots, d$.

To this end, choose $h \ge 1$. First, by symmetry of the noises $\{\eta_t\}$ the random variables $\{X_t(i)X_{t+h}(j)\mathbb{1}\left(|\,X_t(i)X_{t+h}(j)\,| \le n^{2/\kappa}\varepsilon\right)\}$ are uncorrelated. This, together with the fact that $X_0(i)X_h(j)$ is regularly varying with index $\kappa/2$ implies

$$\begin{aligned}
var&\left(\sum_{t=1}^{n-h} X_t(i)X_{t+h}(j)\mathbb{1}\left(|\,X_t(i)X_{t+h}(j)\,| \le n^{2/\kappa}\varepsilon\right)\right)\\
&= (n-h)n^{-4/\kappa}\mathbb{E}\left[X_0(i)^2X_h(j)^2\mathbb{1}\left(|\,X_t(i)X_{t+h}(j)\,| \le n^{2/\kappa}\varepsilon\right)\right]\\
&\sim n^{1-4/\kappa}(n^{2/\kappa}\varepsilon)^2(4-\kappa)\mathbb{P}\left[|\,X_0(i)X_h(j)\,| > n^{2/\kappa}\varepsilon\right]\\
&\to (4-\kappa)\varepsilon^{2-\kappa/2}, \qquad \text{as } n \to \infty\\
&\to 0, \qquad \text{as } \varepsilon \to 0.
\end{aligned}$$

where the asymptotic equivalence follows from Karamata's Theorem, see Bingham et al. (1989).

3. For $\kappa \in (4,\infty)$ we see that the stationary time series

$$\{X_t(i)^2, X_t(i)X_{t+1}(j), \ldots X_t(i)X_{t+m}(j),\ i,j = 1,2,\ldots,d\}_t$$

has a finite $2+\theta$ moment (for some $\theta > 0$) and is geometrically β–mixing, and thus geometrically α–mixing. It follows from standard limit Theorems for mixing sequences (see Doukhan (1994)) that

$$(n^{1/2}(\text{vec}\,(()\,\gamma_{n,X}(h) - \gamma_X(h))))_{h=0,1,\dots,m} \xrightarrow{d} (V_h)_{h=0,\dots,m}$$

where V has a multivariate normal distribution with mean zero.

■

3.5.3 Examples

In this section we discuss some examples of Cholesky factor GARCH processes satisfying some or all of the hypotheses required for the regular variation and autocovariance function convergence results to hold. In each case, we analyze which hypotheses are satisfied and thus which results hold. We will focus on the two–dimensional ChF–GARCH(1,1) with parameters (A, B) given by the equations

$$\begin{aligned} \text{vech}(L_t) =& C + A\text{vech}(L_{t-1}) + BX_{t-1}, \\ H_t =& L_t L_t^T, \\ X_t =& L_t Z_t. \end{aligned} \tag{3.19}$$

where $\{Z_t\}$ is an i.i.d sequence of two–dimensional random vectors with distribution function F.

Example 5 Hypothesis $\mathcal{H}_4$
Since we are considering only two–dimensional ChF–GARCH(1,1) processes, we have that $A(Z_t) = A + (Z^T \otimes B)K$ so that the decomposition (3.9) is

$$A(0) = A, \qquad V_1 = \widetilde{V}_1 = B[\,\cdot\,,1], \qquad V_2 = \widetilde{V}_2 = B[\,\cdot\,,2].$$

Therefore, hypothesis $\mathcal{H}_4$ is satisfied if

$$\mathcal{H}_0 \text{ is satisfied, } A \text{ has no zero rows and } AV_1 \notin \mathcal{S}(V_1, V_2).$$

For a concrete example, consider the matrices

$$A = \begin{bmatrix} 1/2 & 0 & 0 \\ 1/3 & 1/3 & 0 \\ 1/6 & 1/6 & 1/6 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Since A is a triangular matrix, its diagonal elements are its eigenvalues, therefore

$$\varrho(A) = 1/2 < 1$$

so that $\mathcal{H}_0$ is satisfied. By proposition 14 this implies that the ChF–GARCH defined by the equations (3.19) admits a unique casual stationary solution.

Since $A(0)V_1 = (1/2, 1/3, 1/6)^T$ it follows that $AV_1 \notin \mathcal{S}(V_1, V_2)$, so that these vectors generate $\mathbb{R}^3$. Finally, the matrix A has no zero rows. This shows that $\mathcal{H}_4$ holds for this particular choice of A and B.

The next step is choosing a distribution function F satisfying hypotheses $\mathcal{H}_2, \mathcal{H}_3$ and $\mathcal{H}_5$. We assume that the parameter set (A, B) is the one given in Example 5.

Example 6 Heavy–tailed distribution function
Assume that the random variables $\{Z_t(i)\}$ have the Student's t distribution function with ν degrees of freedom whose density is given by

$$f(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}$$

This distribution function satisfies hypothesis $\mathcal{H}_2$ and has regularly varying tails with index $\alpha = \nu$, which follows from Karamata's theorem (See Bingham et al. (1989)). Therefore,

$$\mathbb{E}\left[\mid Z_t(i) \mid^h\right] \begin{cases} < \infty, & \text{if } h < \nu, \\ = \infty, & \text{if } h > \nu \end{cases}$$

Moreover, $\mathbb{E}\left[\mid Z_t(i) \mid^\nu\right] = \infty$ so that hypothesis $\mathcal{H}_3$ is fulfilled with $h_0 = \nu$. Finally, the Student's t distribution is symmetric so that hypothesis $\mathcal{H}_5$ is also satisfied.
Generally speaking, if we assume that the marginal distribution functions have Pareto–like tails, i.e.,

$$\mathbb{P}\left[Z_t(i) > x\right] \sim Kx^{-\alpha}, \qquad \alpha > 0, \quad x \to \infty,$$

and a continuous, everywhere positive density with respect to Lebesgue measure f, then the conclusions of Theorem 23 hold true. For the autocovariance function convergence detailed in Theorem 25 symmetry of the function f is necessary.

Example 7 Light–tailed distribution function
Take $\{Z_t\}$ Normally distributed with parameters $(0,0)$ and $\sigma^2 I$, where I is the 2×2 identity matrix and $\sigma^2 > 0$. This distribution function has an everywhere positive density with respect to Lebesgue measure, and thus satisfies $\mathcal{H}_2$. Since

$$\mathbb{E}\left[\mid Z_t(i) \mid^h \right] = \begin{cases} 0, & \text{if } h = 2k+1, \\ \frac{\sigma^{2n}(2n)!}{2^n n!}, & \text{if } h = 2k, \end{cases}$$

hypothesis $\mathcal{H}_3$ is satisfied with $h_0 = \infty$ (observe that h_0 may be ∞ in $\mathcal{H}_3$). Finally, the marginal distribution functions are standard Normal, so that they are symmetric and thus $\mathcal{H}_5$ holds.
More generally, if all the moments of the distribution function F exist and the support of its (continuous) density f is $\mathbb{R}$ then the hypotheses in Theorem 23 hold but symmetry is needed for the conclusions of Theorem 25 to be true.

Example 8 Mid–talied distribution function
Assume $\{Z_t(i)\}$ have a standard logNormal distribution, then

$$\mathbb{E}\left[Z_t(i)^n\right] = \exp\{n^2/2\}$$

so that hypothesis $\mathcal{H}_3$ is satisfied with $h_0 = \infty$.
Since the logNormal random variable has $\mathbb{R}_+$ as support, hypothesis $\mathcal{H}_2$ is not satisfied. However, the proof to Theorem 23 can still be achieved because hypothesis $\mathcal{H}_2$ was only used in this proof to show that the support of $A(Z_t)$ is connected. The argument is thus still true if the support of $\{Z_t(i)\}$ is $\mathbb{R}_+$. Therefore, the ChF–GARCH(1,1) (and indeed, the ChF–GARCH(p,q) in any other dimension $d > 2$) generated with logNormally distributed $\{Z_t\}$ has regularly varying finite dimensional distributions.
Observe that the point 0 is not in the interior of the support of the logNormal density, so that the β–mixing property stated in Theorem 22 can not be guaranteed in this case. Also, since the logNormal density is not symmetric the autocovariance function convergence rates given in Theorem 25 may not hold.
This example can be generalized to distribution functions with a continuous density supported in $\mathbb{R}_+$, finite moments of all orders, but no Laplace transform. In this generic case, the above argument shows that the regular variation of the finite–dimensional distribution functions of the ChF–GARCH(1,1) is guaranteed by Theorem 23. However, the β–mixing may not hold, so that

the Point Process convergence in Proposition 16 could fail and, thus, the rates of convergence of the sample autocovariance function need not agree with those given in Theorem 25.

Comentarios finales y trabajo futuro

Hemos visto que los procesos GARCH de correlaciones condicionales constantes y para el factor de Cholesky son procesos con colas pesadas bajo hipótesis generales sobre los ruidos con que los generamos y cómo esta propiedad tiene consecuencias en la conducta asintótica de los extremos y de la función muestral de autocovarianzas. Nuestro estudio tiene tres pilares:

1. La representación recursiva estocástica

$$Y_t = A(Z_t)Y_{t-1} + G_t \tag{3.20}$$

   y los Teoremas de Kesten (1973).

2. La convergencia de los procesos puntuales

$$N_n = \sum_{i=1}^{n} \delta_{a_n^{-1} X_i}$$

   para sucesiones estacionarias con propiedades de mixing y sus consecuencias para la función muestral de autocovarianzas dadas en Davis and Mikosch (1998).

3. La representación del índice extremo multivariado para procesos con SRE dada en Perfekt (1997).

No obstante, el uso de estos resultados para cada caso difiere sustancialmente. En el caso del CCC–GARCH la técnica utilizada es una generalización de la dada en Basrak et al. (2002b) para el caso unidimensional. Esto se debe a que tanto el planteamiento del modelo como las hipótesis a verificar para comprobar la variación regular son muy similares. Cabe destacar que para este proceso, nuestra mayor contribución fue dar una prueba rigurosa de las propiedades de variación regular.

En el caso del GARCH basado en el factor de Cholesky, nuestro estudio fue mucho más profundo en virtud de que se contaba únicamente con cierto conocimiento del estimador de log–máxima–verosimilitud. En este caso, la representación recursiva (3.20) fundamentó la relación entre los coeficientes del proceso y su estacionariedad fuerte. Aquí vemos una diferencia entre los distintos GARCH. Tanto para el GARCH unidimensional como para el CCC–GARCH, se conocen condiciones necesarias y suficientes para garantizar la estacionariedad, pero para el GARCH del factor de Cholesky sólo se tienen condiciones suficientes.

La condición necesaria involucra al índice de Lyapunov que es, en general, muy difícil de conocer. En primer lugar, no se tiene una expresión sencilla, cerrada para él y estimarlo numéricamente puede ser muy complicado. Véase Gol′dsheĭd (1991) a este respecto.

Hay que recordar que las reglas de identificación originales de Kawakatsu (2003) no permiten la existencia de esta distribución estacionaria en todo el espacio, es decir, restringen el proceso estacionario a un subespacio de dimensión menor. Esta limitante del modelo de Kawakatsu (2003) fue superada imponiendo condiciones distintas a los parámetros del proceso, mismas que no son restrictivas en modo alguno.

La variación regular también depende de la representación (3.20), pero a diferencia del proceso CCC–GARCH, las hipótesis a verificar son distintas a las del caso unidimensional. En particular, el hecho de que las matrices $A(Z_t)$ puedan tomar valores negativos nos llevó a utilizar ciertas técnicas de sistemas dinámicos perturbados por ruidos estocásticos y a encontrar componentes continuos para el Kernel de transición de ciertas cadenas de Markov asociadas al proceso (3.20).

Otra diferencia llamativa es que para el GARCH unidimensional y para el CCC–GARCH se requiere la hipótesis de simetría en las distribuciones marginales, tanto para la variación regular de las distribuciones finito dimensionales como para relacionarla con la convergencia de la función de autocovarianzas. Para el modelo de Cholesky esta hipótesis es innecesaria para todas las propiedades de variación regular, de modo que en este caso nuestros resultados son más generales.

En ambos casos, es importante observar que la sucesión de ruidos $\{Z_n\}$ **no** necesita ser de colas pesadas para que el proceso GARCH tenga esta propiedad. Esta característica es por demás interesante y llama a extender los resultados a más y distintos modelos GARCH.

Los métodos presentados en este trabajo no pueden utilizarse para culquier proceso GARCH multivariado ya que no todos ellos admiten la representación (3.20). Por ejemplo, el modelo *vech*–GARCH de Bollerslev,

$$vech(H_t) = C + \sum_{i=1}^{p} A_i vech(H_{t-i}) + \sum_{j=1}^{q} vech(X_{t-j}X_{t-j}^T), \tag{3.21}$$

no admite dicha representación. La razón de ello es que el producto tensorial $\otimes$, que es el modo natural de lidiar con la operación *vec* (es suficiente trabajar con *vec* en virtud de la relación entre ambas operaciones), nos lleva a las igualdades

$$\begin{aligned} vec(X_{t-j}X_{t-j}^T) =& H_t^{1/2} \otimes H_t^{1/2} vec(Z_{t-j}Z_{t-j}^T) \\ =& (vec(Z_t Z_t^T) \otimes I) vec(H_t^{1/2} \otimes H_t^{1/2}). \end{aligned} \tag{3.22}$$

De modo que no es posible obtener $vech(H_t^{1/2})$ en el término derecho de (3.22) con este procedimiento y no se puede plantear una ecuación recursiva de la forma (3.20).

Sin embargo, los métodos pueden extenderse a modelos de volatilidad estocástica con correlaciones condicionales constantes y modelos basados en el factor de Cholesky que admitan la representación (3.20) sin que sean GARCH multivariados. La diferencia será la forma de las matrices $A(Z_t)$ y de los vectores G_t y las hipótesis de momentos sobre las variables $\{Z_t\}$ necesarias en cada caso. Dar una explicación más concreta de cuáles procesos se pueden incluir en esta categoría y los resultados de variación regular y convergencia es un primer objetivo para mi trabajo futuro.

Otro problema interesante para el futuro es estudiar otros modelos GARCH multivariados puesto que se intuye que comparten las propiedades expuestas, pero requieren de metodologías distintas. Concretamente, es deseable contar con una teoría de variación regular para procesos Markovianos de la forma

$$X_0 = x, \qquad X_n = F(X_{n-1}, Z_n),$$

donde F es una función suave. La pregunta a abordar es: ¿Para qué clase $\mathcal{F}$ de funciones F se obtiene con esta recursión un proceso de estacionario, absolutamente regular y con colas pesadas? La respuesta no es trivial, incluso partiendo de que $\{Z_n\}$ tenga colas pesadas, cuanto menos si tiene colas ligeras.

Por último, se tiene contemplado abundar sobre las consecuencias prácticas de los resultados sobre la convergencia de la función de autocovarianzas en el campo de estimación de series de tiempo.

Bibliography

O. E. Barndorff-Nielsen and N. Shephard. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(2):167–241, 2001. ISSN 1369-7412.

O. E. Barndorff-Nielsen and R. Stelzer. Positive-definite matrix processes of finite variation. *Probab. Math. Statist.*, 27(1):3–43, 2007. ISSN 0208-4147.

B. Basrak, R. A. Davis, and T. Mikosch. A characterization of multivariate regular variation. *Ann. Appl. Probab.*, 12(3):908–920, 2002a.

B. Basrak, R. A. Davis, and T. Mikosch. Regular variation of GARCH processes. *Stochastic Process. Appl.*, 99(1):95–115, 2002b.

L. Bauwens, S. Laurent, and J. V. K. Rombouts. Multivariate GARCH models: a survey. *J. Appl. Econometrics*, 21(1):79–109, 2006.

R. Benedetti and J. Risler. *Real algebraic and semi–algebraic sets*. Herman Paris, 1990.

N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1989.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, 31(3):307–327, 1986. ISSN 0304-4076.

T. Bollerslev. Modelling the Coherence in Short–Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model. *Rev. Econom. Statist.*, 72(3):498–505, 1990.

T. Bollerslev, R. Chou, and K. Kroner. ARCH modelling in finance: A review of the theory and empirical evidence. 52:5–59, 1992.

T. Bollerslev, R. F. Engle, and D. B. Nelson. Arch models. In *Handbook of econometrics, Vol. IV*, volume 2 of *Handbooks in Econom.*, pages 2959–3038. North-Holland, Amsterdam, 1994.

P. Bougerol and N. Picard. Strict stationarity of generalized autoregressive processes. *Ann. Probab.*, 20(4):1714–1730, 1992.

F. Boussama. *Ergodicité, mélange et estimation dans les modelés GARCH.* PhD thesis, Université de Paris 7, 1998.

A. Brandt. The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. in Appl. Probab.*, 18(1):211–220, 1986.

P. Brockwell, C. Erdenebaatar, and A. Lindner. Continuous–time GARCH processes. *Ann. Appl. Probab.*, 16(2):790–826, 2006.

F. Comte and O. Lieberman. Asymptotic theory for multivariate GARCH processes. *J. Multivariate Anal.*, 84(1):61–84, 2003.

R. Davis and T. Mikosch. Extreme Value Theory for GARCH Processes, 2006. submitted.

R. A. Davis and T. Hsing. Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Probab.*

R. A. Davis and T. Mikosch. The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Ann. Statist.*, 26(5):2049–2080, 1998.

B. de Saporta. *Etude de la solution stationnaire de l'équation* $Y_{n+1} = a_n Y_n + b_n$ *á coefficients aléatoires.* PhD thesis.

P. Doukhan. *Mixing: properties and examples*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1994.

F. Drost and B. Werker. Closing the GARCH gap: Continuous time GARCH modelling. *J. Econometrics*, 74:31–57, 1996.

P. Embrechts, T. Mikosch, and C. Klüppelberg. *Modelling extremal events for insurance and finance.* Springer-Verlag, London, UK, 1997. ISBN 3-540-60931-8.

R. Engle and I. Ishida. Forecasting variance of variance: The square-root, the affine, and the CEV GARCH models. *Dept. Finance Working Papers, New York University*, 2002.

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007, 1982. ISSN 0012-9682.

R. F. Engle and K. F. Kroner. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(1):122–150, 1995. ISSN 0266-4666.

B. Fernández and N. Muriel. Regular variation and related results for the multivariate GARCH(p,q) process with constant conditional correlations. . 2008. To appear in Journal of Multivariate Analysis.

H. Furstenberg and H. Kesten. Products of random matrices. *Ann. Math. Statist*, 31:457–469, 1960.

R. Gallant and G. Tauchen. SNP. A program for nonparametric time series analysis. Available on line at http://www.econ.duke.edu/∼ get.

I. Y. Gol′dsheĭd. Lyapunov exponents and asymptotic behaviour of the product of random matrices. In *Lyapunov exponents (Oberwolfach, 1990)*, volume 1486 of *Lecture Notes in Math.*, pages 23–37. Springer, Berlin, 1991.

C. Gouriéroux. *ARCH models and financial applications*. Springer Series in Statistics. Springer-Verlag, New York, 1997. ISBN 0-387-94876-7.

C. M. Hafner and A. Preminger. Asymptotic theory for a factor GARCH model, 2008. to appear in Journal of Econometric Theory.

R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, 1990.

R. A. Horn and C. R. Johnson. *Topics in matrix analysis*. Cambridge University Press, 1991.

T. Hsing. Extreme value theory for multivariate stationary sequences. *J. Multivariate Anal.*, 29(2):274–291, 1989.

J. Hüsler. Multivariate extreme values in stationary random sequences. *Stochastic Process. Appl.*, 35(1):99–108, 1990.

A. H. Jessen and T. Mikosch. Regularly varying functions. *Publ. Inst. Math. (Beograd) (N.S.)*, 80(94):171–192, 2006. ISSN 0350-1302.

O. Kallenberg. *Random Measures.* Akademie-Verlag, Berlin, 1969. US publisher: Academic Press.

O. Kallenberg. *Foundations of Modern Probability.* Springer, January 2002. ISBN 0387953132.

H. Kawakatsu. Cholesky factor GARCH. Mimeo. *Quantitative Micro Software, Irvine, California*, 2003.

H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Math.*, 131:207–248, 1973.

C. Klüppelberg, A. Lindner, and R. Maller. A continuous time GARCH process driven by a lévy process: Stationarity and second order behaviour. *J. Appl. Probab.*, 41:601–622, 2004.

M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes.* Springer Series in Statistics. Springer-Verlag, New York, 1983. ISBN 0-387-90731-9.

A. Lindner and R. Maller. Lévy integrals and the stationarity of generalised ornstein–uhlebeck processes. *Stochastic Process. Appl.*, 115:1701–1722, 2005.

S. Ling and M. McAleer. Asymptotic theory for a vector ARMA–GARCH model. *Econometric Theory*, 19(2):280–310, 2003.

F. Linskog. *Multivariate Extremes and Regular Variation for Stochastic Processes.* PhD thesis, ETH, Zurich, 2004.

A. Marshal and I. Olkin. Domains of attraction of multivariate extreme value distributions. *Ann. Probab.*, 11:168–177.

S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability.* Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993. ISBN 3-540-19832-6.

T. Mikosch and C. Stărică. Limit theory for the sample autocorrelations and extremes of a GARCH $(1,1)$ process. *Ann. Statist.*, 28(5):1427–1451, 2000.

A. Mokkadem. Propriétés de mélange des processus autorégressifs polynomiaux. *Ann. Inst. H. Poincaré Probab. Statist.*, 26(2):219–260, 1990. ISSN 0246-0203.

S. Nandagopalan. On the Multivariate Extremal Index. *J. Res. Natl. Inst. Stand. Technol.*, 99(4):543–550, 1994.

D. Nelson. ARCH models as diffusion approximations. *J. Econometrics*, 45: 7–38, 1990.

E. Nummelin. *General irreducible Markov chains and nonnegative operators*, volume 83 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1984. ISBN 0-521-25005-6.

R. Perfekt. Extreme value theory for a class of Markov chains with values in $\mathbf{R^d}$. *Adv. in Appl. Probab.*, 29(1):138–164, 1997.

S. I. Resnick. Extreme Values, Regular Variation and Point Processes. Springer-Verlag, 1987.

N. Shephard. *Stochastic Volatility: Selected Readings (Advanced Texts in Econometrics)*. Oxford University Press, USA, May 2005.

C. Stărică. Multivariate extremes for models with constant conditional correlations. *Journal of Empirical Finance*, 6(5):515–553, December 1999. URL `http://ideas.repec.org/a/eee/empfin/v6y1999i5p515-553.html`.

R. Stelzer. On the relation between the vec and BEKK multivariate GARCH models. *Econometric Theory*, 24(4):1131–1136.

R. S. Tsay. *Analysis of financial time series*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2005. ISBN 978-0-471-69074-0; 0-471-69074-0.