

TIME-VARYING EXTREME VALUE DEPENDENCE WITH APPLICATION TO LEADING EUROPEAN STOCK MARKETS

BY DANIELA CASTRO CAMILO, MIGUEL DE CARVALHO AND JENNIFER WADSWORTH

*King Abdullah University of Science and Technology,
University of Edinburgh and
Lancaster University*

Extremal dependence between international stock markets is of particular interest in today's global financial landscape. However, previous studies have shown this dependence is not necessarily stationary over time. We concern ourselves with modeling extreme value dependence when that dependence is changing over time, or other suitable covariate. Working within a framework of asymptotic dependence, we introduce a regression model for the angular density of a bivariate extreme value distribution that allows us to assess how extremal dependence evolves over a covariate. We apply the proposed model to assess the dynamics governing extremal dependence of some leading European stock markets over the last three decades, and find evidence of an increase in extremal dependence over recent years.

1. Introduction. In recent years, international stock markets have been registering unprecedented levels of turbulence. Episodes such as the subprime crisis and the Greek debt crisis may have boosted this turbulence a little further, and led many to fear a financial doomsday. The situation has been extraordinarily delicate in Europe, where evidence of increasing extremal dependence was found by [Poon et al. \(2003, 2004\)](#) before the most recent financial crisis. We look to update suitable parts of their analysis and in particular analyze the time-varying extremal dependence in a more complete manner than has been done before. To achieve this goal, we propose an approach for modeling nonstationarity in the extreme value dependence structure.

Statistical modeling of univariate extreme values has been in development since the 1970s ([Natural Environment Research Council, 1975](#)). Fundamental to practical application to complex problems has been the development of methodology to account for nonstationarity in the distributions of interest, which was first strongly advocated by [Davison and Smith \(1990\)](#). Typical approaches to this problem are based around the generalized linear modeling idea of allowing the parameters of a marginal distribution to depend on covariates; more flexible approaches involving generalized additive modeling were introduced by [Chavez-Demoulin and Davison \(2005\)](#). [Eastoe and Tawn \(2009\)](#) present related ideas where data are preprocessed according to their dependence on covariates.

Statistical methods for modeling multivariate extreme values were introduced by [Tawn \(1988\)](#), and developed in [Tawn \(1990\)](#) and [Coles and Tawn \(1991\)](#). Since this time, much work has been done on developing dependence modeling frameworks for extremes, yet surprisingly little has focused on how to incorporate nonstationarity into the (extremal) dependence structure. Exceptions include [Eastoe \(2009\)](#), who introduces a conditionally independent hierarchical model, [Jonathan et al. \(2014\)](#), who develop methodology for including covariates in the model of [Heffernan and Tawn \(2004\)](#), and [de Carvalho and Davison \(2014\)](#), who develop a semiparametric model

Keywords and phrases: Angular measure, Bivariate extreme values, European stock market integration, Risk, Statistics of extremes.

for settings where several multivariate extremal distributions are linked through the action of a covariate on an unspecified baseline distribution. In addition, [Huser and Genton \(2016\)](#) developed nonstationary models for spatial extremes where covariates can be included. In this work we add to the literature on modeling nonstationarity in the dependence structure by proposing flexible methodology for a simple set-up. Working within a tail dependence framework known as *asymptotic dependence*, we suppose that the relevant bivariate extreme value distribution evolves over a certain covariate of interest. The approach that we take is fully nonparametric, which is advantageous since neither the form of the bivariate distribution at a given covariate, nor the form of dependence on the covariate can be parametrically specified.

Our methodology is particularly tailored for assessing temporal changes in extremal dependence, which is the situation that we would like to investigate in our motivating example. [Poon et al. \(2003, 2004\)](#) studied the dependence between stock market returns in the US, UK, France, Germany, and Japan. The main focus of their works was to highlight that not all markets exhibit a sufficient strength of tail dependence to be asymptotically dependent, and to propose alternative dependence summaries. However, considering only the European markets, they noted that there was evidence for relatively strong left tail dependence, and we also find evidence for asymptotic dependence in the left tails of these major European markets. As noted by [Poon et al. \(2003\)](#), the dependence is not stationary in time, and a main focus of this work is to explore this nonstationarity using a full model for the time-varying dependence structure, rather than simply summary statistics.

In the next section we provide a background on dependence modeling for extreme values, and introduce our proposed framework for incorporating nonstationarity. In [Section 3](#) we introduce our estimation and inference methods; numerical illustrations follow in [Section 4](#). The focus of [Section 5](#) is on applying the proposed methods to returns from three major European stock markets—using CAC, DAX, and FTSE—to assess the evolution of their extremal dependence structure over time. We conclude in [Section 6](#).

2. Conditional modeling for bivariate extremes.

2.1. *Bivariate statistics of extremes.* Let $\{(Y_{i,1}, Y_{i,2})\}_{i=1}^N$ be a collection of independent and identically distributed random vectors with continuous marginal distributions F_{Y_1} and F_{Y_2} . We are concerned with assessments of the extremal dependence between the components of the vectors, and thus without loss of generality we shall suppose that they have standard Fréchet margins, i.e., $P(Y_j > y) = \exp(-1/y)$, for $y > 0$ and $j = 1, 2$. Let

$$(M_{N,1}, M_{N,2}) = \frac{1}{N} \left(\max_{1 \leq i \leq N} \{Y_{i,1}\}, \max_{1 \leq i \leq N} \{Y_{i,2}\} \right)$$

be the standardized vector of componentwise maxima. Then if

$$(2.1) \quad P(M_{N,1} \leq y_1, M_{N,2} \leq y_2) \rightarrow G(y_1, y_2), \quad \text{as } N \rightarrow \infty,$$

where G is a non-degenerate distribution function, G has the form

$$(2.2) \quad G(y_1, y_2) = \exp \left\{ -2 \int_{[0,1]} \max \left(\frac{w}{y_1}, \frac{1-w}{y_2} \right) H(dw) \right\}, \quad y_1, y_2 > 0.$$

Here, $G(y_1, y_2)$ is the so-called *bivariate extreme value distribution* and H is a probability measure—known as the *angular measure*. A consequence of Pickands' ([1981](#)) representation

theorem is that the angular measure needs to obey the following marginal moment constraint

$$(2.3) \quad \int_{[0,1]} w H(dw) = 1/2;$$

see, for example, [Coles \(2001, Theorem 8.1\)](#). Let $R = Y_1 + Y_2$ and $W = Y_1/(Y_1 + Y_2)$. [de Haan and Resnick \(1977\)](#) have shown that the convergence in (2.1) is equivalent to

$$(2.4) \quad \mathbb{P}(W \in \cdot \mid R > u) \rightarrow H(\cdot), \quad u \rightarrow \infty.$$

In practice, convergence (2.4) is more often useful than (2.1) and tells us that when the ‘radius’ R is large, the ‘pseudo-angles’ W are approximately distributed according to H , and approximately independent of R . The distribution of mass of H on $[0, 1]$ describes the extremal dependence structure of the random vector (Y_1, Y_2) . The extreme cases of this distribution are given by *asymptotic independence*, whereby all mass is placed at the vertices of $[0, 1]$, giving $G(y_1, y_2) = \exp\{-(y_1^{-1} + y_2^{-1})\}$, and by complete dependence, whereby all mass is placed at the center of the interval, yielding $G(y_1, y_2) = \exp\{-\max(y_1^{-1}, y_2^{-1})\}$. We refer to situations where H has mass away from the vertices as *asymptotic dependence* and this will be the framework of our modeling. Nevertheless, asymptotic independence is a relatively common situation in practice, and can be detected when R and W are not found to be independent for any values of R , with the mass of W moving closer to 0 and 1 as events become more extreme. In this situation, no models for H will provide useful information on the extremal dependence structure. Finally, a standard assumption for statistical modeling is that H is absolutely continuous with *angular density* $h = dH/dw$, and this will be our framework.

Functionals of interest of the angular measure include the bivariate extreme value distribution (2.2), which also represents the extreme value *copula*, C_{EV} , (e.g. [Gudendorf and Segers, 2010](#)) in Fréchet margins, i.e., $G(y_1, y_2) = C_{EV}(e^{-1/y_1}, e^{-1/y_2})$. Other functionals include the [Pickands \(1981\)](#) dependence function $A(w) = 1 - w + 2 \int_0^w H(u) du$, and the extremal coefficient $C = 2A(1/2)$. Extreme value independence corresponds to $A(w) = 1$, whereas perfect dependence corresponds to $A(w) = \max(w, 1 - w)$.

2.2. Conditional modeling framework. We define the conditional bivariate extreme value (BEV) distribution as

$$(2.5) \quad G_x(y_1, y_2) \equiv G(y_1, y_2 \mid X = x) = \exp \left\{ -2 \int_{[0,1]} \max \left(\frac{w}{y_1}, \frac{1-w}{y_2} \right) H(dw \mid X = x) \right\},$$

for $x \in \mathcal{X} \subseteq \mathbb{R}$, and $y_1, y_2 > 0$. Here $H_x(\cdot) \equiv H(\cdot \mid X = x)$ are conditional probability measures satisfying

$$(2.6) \quad \int_{[0,1]} w H_x(dw) = 1/2, \quad x \in \mathcal{X}.$$

If $H_x(w) \equiv H_x[0, w]$ is absolutely continuous, its conditional angular density is $h_x = dH_x/dw$. Further aspects of conditional angular measures are discussed in [de Carvalho \(2016\)](#).

Our main modeling object of interest will be the set of conditional angular densities $\{h_x(w) : w \in [0, 1], x \in \mathcal{X}\}$, which we will refer to as the *angular surface*. A simple angular surface can be obtained with the conditional angular density $h_x(w) = \beta(w; \mu_x, \mu_x)$, where $\mu : \mathcal{X} \mapsto (0, \infty)$, and $\beta(\cdot; p, q)$ denotes the beta density with shape parameters $p, q > 0$. In Figure 1 (a), we represent

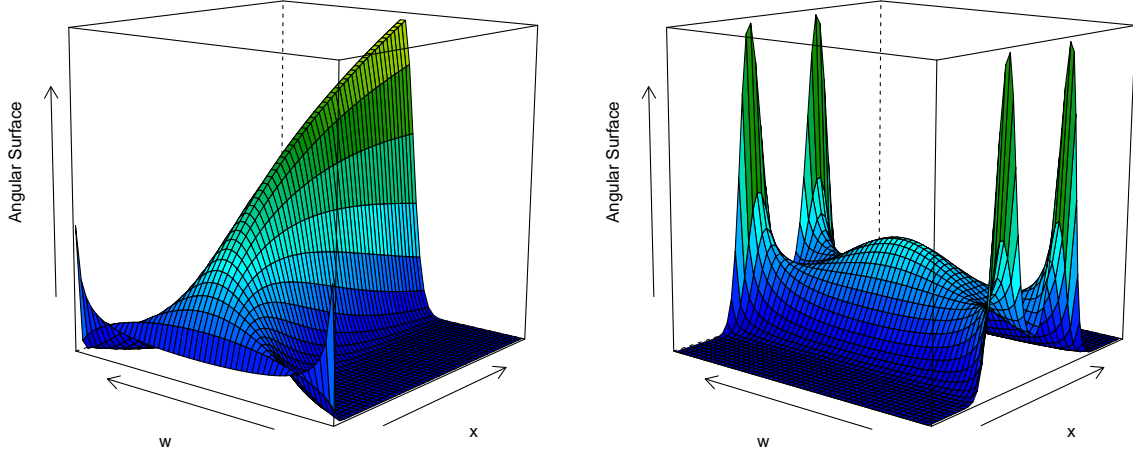


FIG 1. (a) Angular surface from a conditional beta family, with $\mu_x = x$, for $x \in \mathcal{X} = [0.5, 50]$. (b) Angular surface from a conditional logistic family, with $\alpha_x = \Phi(x^2)$, for $x \in \mathcal{X} = [-3, 3]$.

an angular surface based on this model, with $\mu_x = x$, for $x \in \mathcal{X} = [0.5, 50]$. As can be seen, larger values of the predictor x lead to stronger levels of extremal dependence. Other angular surfaces can be readily constructed from parametric models for the angular density.

EXAMPLE 1 (conditional logistic model). The logistic angular surface is a covariate-adjusted extension of the logistic model (Coles, 2001, p. 146), and it is based on the conditional angular density

$$(2.7) \quad h_x(w) = \frac{1}{2} \left(\frac{1}{\alpha_x} - 1 \right) \{w(1-w)\}^{-1-1/\alpha_x} \{w^{-1/\alpha_x} + (1-w)^{-1/\alpha_x}\}^{\alpha_x-2}, \quad w \in (0, 1),$$

where $\alpha : \mathcal{X} \mapsto (0, 1]$. The closer α_x is to 0, the higher the level of extremal dependence, while the closer α_x is to 1, the closer we get to independence. Angular surfaces with simple ‘shapes’ can be obtained by modeling α_x with either a distribution function, $F(x)$, or a survivor function, $1 - F(x)$. More sophisticated shapes can be obtained with $\alpha_x = (F \circ G)(x)$, for a certain continuous function $G : \mathcal{X} \mapsto \mathbb{R}$. In Figure 1 (b) we represent the logistic angular surface in (2.7) with $\alpha_x = \Phi(x^2)$, for $x \in \mathcal{X} = [-3, 3]$, where Φ denotes the standard normal distribution function.

EXAMPLE 2 (conditional Dirichlet model). The Dirichlet angular surface is a covariate-adjusted extension of the Dirichlet model (Coles and Tawn, 1991), and it is based on the conditional angular density

$$(2.8) \quad h_x(w) = \frac{a_x b_x \Gamma(a_x + b_x + 1) (a_x w)^{a_x - 1} \{b_x (1 - w)\}^{b_x - 1}}{2\Gamma(a_x)\Gamma(b_x) \{a_x w + b_x (1 - w)\}^{a_x + b_x + 1}}, \quad w \in (0, 1),$$

where $a : \mathcal{X} \mapsto (0, \infty)$ and $b : \mathcal{X} \mapsto (0, \infty)$. Angular surfaces with simple shapes can be obtained with $a_x = b_x = \exp(x)$, while if more complex dynamics are desirable, can be based on $a_x = \exp\{\mathcal{A}(x)\}$, $b_x = \exp\{\mathcal{B}(x)\}$, where $\mathcal{A} : \mathcal{X} \mapsto \mathbb{R}$ and $\mathcal{B} : \mathcal{X} \mapsto \mathbb{R}$ are continuous functions.

The basic idea of a conditional angular measure is not especially complicated, and inference for such would be simple if: (i) we knew our data conform to a particular parametric family, and (ii) we knew precisely how that family depended on x . However, since we do not have knowledge of either of these things, the natural approach to take is a nonparametric one. We assume that h_x varies smoothly with x , and thus kernel smoothing becomes a natural option. We describe our estimation strategy in Section 3.

2.3. Related conditional objects of interest. Our estimation target $\{h_x(w) : w \in [0, 1], x \in \mathcal{X}\}$ can be used for constructing other objects of interest when modeling bivariate extremes. For example, a conditional version of Pickands (1981) dependence function can be defined as

$$A_x(w) = 1 - w + 2 \int_0^w H_x(u) du, \quad x \in \mathcal{X}, \quad w \in [0, 1],$$

leading to the conditional extremal coefficient $C_x = 2A_x(1/2)$. A covariate-adjusted extreme value copula can be readily constructed from (2.5). Although much theoretical and applied work has been devoted to time-dependent copulas (Patton, 2006; Veraverbeke et al., 2007; Acar et al., 2011; Fermanian and Marten, 2012), the amount of work dedicated to time-varying extreme value copulas is by comparison fairly reduced, but of obvious relevance in a wealth of contexts of applied interest. The latter setup is the one of interest in the current manuscript.

EXAMPLE 3. Using the conditional angular density from Example 1, we obtain $A_x(w) = \{(1-w)^{1/\alpha_x} + w^{1/\alpha_x}\}^{\alpha_x}$ and $C_x = 2^{\alpha_x}$, while the logistic angular surface is based on the conditional BEV distribution,

$$G_x(y_1, y_2) = \exp\{-(y_1^{-1/\alpha_x} + y_2^{-1/\alpha_x})^{\alpha_x}\}, \quad x \in \mathcal{X}, \quad y_1, y_2 > 0.$$

3. Estimation and inference.

3.1. Derivation of pseudo-angles. Consider Eq. (2.4). We are now supposing nonstationarity in the dependence structure such that

$$(3.1) \quad \mathbf{P}(W \in \cdot \mid R > u, X = x) \rightarrow H_x(\cdot), \quad u \rightarrow \infty.$$

Note that we still assume that R and W are derived from Y_1, Y_2 with standard Fréchet margins. Typically, when stationarity in the extremal dependence structure is assumed, one searches for a high threshold in R , such that W and R are approximately independent above the threshold, and uses all W associated to threshold exceedances of R for inference. Supposing that x does not impact upon the rate of convergence in the limit (3.1), a similar approach is justified here. However, for prudence, we assess the dependence of R on x using quantile regression (Koenker, 2005). To be consistent with the nonparametric nature of our approach, we fit a nonparametric quantile regression using regression splines. This method flexibly fits a piecewise cubic polynomial to estimate the 95% quantile of R . If any relationship between R and x is detected, then we take the W associated to exceedances of the fitted threshold by R for inference. Below we use $n = o(N)$ to denote the number of pseudo-angles that resulted from thresholding $R_i = Y_{i,1} + Y_{i,2}$, for $i = 1, \dots, N$. Further details on the derivation of pseudo-angles for our data application can be found in Section 5.3.

We note that we are not allowing for the margins to change over the predictor. This is however a sensible modeling assumption for our data application, because (filtered) returns are known

to be approximately stationary. Indeed, as posed by [Resnick \(2007, p. 7\)](#) “Returns have more attractive statistical properties than prices such as stationarity.” See [Section 5.2](#) for details on the filtering methods used in our data application.

3.2. Conditional angular density estimation. Here we outline our estimator for the family of densities $\{h_x(w) : w \in [0, 1], x \in \mathcal{X}\}$. Assume observations $\{(X_i, W_i)\}_{i=1}^n$, where the covariates X_i are continuous and in $\mathcal{X} \subseteq \mathbb{R}$. Let $K_b(x) = (1/b)K(x/b)$ be a kernel with bandwidth $b > 0$. For any $x \in \mathcal{X}$, we define the estimator

$$(3.2) \quad \hat{h}_x(w) = \sum_{i=1}^n \pi_{b,i}(x) \beta(w; \nu W_i \theta_b(x) + \tau, \nu \{1 - W_i \theta_b(x)\} + \tau), \quad w \in (0, 1),$$

where

$$\theta_b(x) = \frac{1/2}{\sum_{i=1}^n \pi_{b,i}(x) W_i}, \quad \pi_{b,i}(x) = \frac{K_b(x - X_i)}{\sum_{j=1}^n K_b(x - X_j)}, \quad i = 1, \dots, n.$$

The moment constraint [\(2.6\)](#) is satisfied, since

$$\int_0^1 w \hat{h}_x(w) dw = \frac{\sum_{i=1}^n K_b(x - X_i) \{\nu W_i \theta_b(x) + \tau\}}{(\nu + 2\tau) \sum_{i=1}^n K_b(x - X_i)} = \frac{\nu/2 + \tau}{\nu + 2\tau} = 1/2,$$

for all valid $\tau \geq 0$, upon substitution of $\theta_b(x)$.

The two kernels (K_b and β) and the three parameters involved in our estimator can be interpreted as follows. The bandwidth $b > 0$ is the scale parameter of the kernel K_b and controls the amount of smoothing in the x -direction. The choice of the kernel K_b is subject to the typical considerations. In principle, K_b should be symmetric and unimodal, since there is a sense in which density estimators based on kernels that do not satisfy these requirements are inadmissible ([Cline, 1988](#)). While there are many kernel functions that do satisfy these basic requirements, it is well known that the choice of the kernel has little impact on the corresponding estimators; see [Wand and Jones \(1995, Ch. 2\)](#) and references therein. The parameter $\nu > 0$ is asymptotically inversely proportional to the variance of the kernel β and has the main role of controlling the amount of smoothing in the w -direction. The additional parameter $\tau \geq 0$ has the role of adjusting slightly the center of the kernel, allowing more flexible estimation, whilst not affecting the imposition of the moment constraint. Note that $\tau = 0$ yields a kernel with mean equal to W_i , whilst $\tau = 1$ yields a kernel with mode W_i . In addition, $\theta_b(x)$ assesses by how much we deviate from the moment constraint [\(2.6\)](#). To see this, note that $\theta_b(x) = (1/2)/\hat{E}(W | X = x)$, where $\hat{E}(W | X = x) = \sum_{i=1}^n \pi_{b,i}(x) W_i$ is the Nadaraya–Watson estimator ([Nadaraya, 1964](#); [Watson, 1964](#)) of $E(W | X = x) = \int_{[0,1]} w H_x(dw) = 1/2$, for all $x \in \mathcal{X}$.

Plug-in estimators for the related conditional objects of interest discussed in [Section 2.3](#) can be readily obtained; particularly

$$\hat{H}_x(w) = \sum_{i=1}^n \pi_{b,i}(x) B(w; \nu W_i \theta_b(x) + \tau, \nu \{1 - W_i \theta_b(x)\} + \tau), \quad w \in (0, 1),$$

where $B(w; p, q)$ is the regularized incomplete beta function, with $p, q > 0$; in addition, the plug-in estimators for the conditional Pickands dependence function, extremal coefficient, and bivariate extreme value distribution can be written as

(3.3)

$$\begin{aligned}\widehat{A}_x(w) &= 1 - w + 2 \sum_{i=1}^n \pi_{b,i}(x) \int_0^w B(u; \nu W_i \theta_b(x) + \tau, \nu \{1 - W_i \theta_b(x)\} + \tau) du, \\ \widehat{C}_x &= 2\widehat{A}_x(1/2) = 1 + 4 \sum_{i=1}^n \pi_{b,i}(x) \int_0^{1/2} B(u; \nu W_i \theta_b(x) + \tau, \nu \{1 - W_i \theta_b(x)\} + \tau) du, \\ \widehat{G}_x(y_1, y_2) &= \exp \left\{ -2 \int_0^1 \max \left(\frac{u}{y_1}, \frac{1-u}{y_2} \right) \right. \\ &\quad \left. \times \sum_{i=1}^n \pi_{b,i}(x) \beta(u; \nu W_i \theta_b(x) + \tau, \nu \{1 - W_i \theta_b(x)\} + \tau) du \right\},\end{aligned}$$

for $x \in \mathcal{X}$, and $y_1, y_2 > 0$.

3.3. Connections to smoothing on the unit interval. Kernel density estimation on the unit interval is a challenging problem; see [Chen \(1999\)](#), [Jones and Henderson \(2007\)](#), [de Carvalho et al. \(2013\)](#), [Geenens \(2014\)](#), and the references therein. In this section we contrast a stationary version of our estimator (3.2) with that of [Chen \(1999\)](#), and comment on the connections with the smooth Euclidean likelihood angular density of [de Carvalho et al. \(2013\)](#). The latter can be regarded as a moment constrained kernel density estimator on the unit interval, in the sense that it obeys (2.3).

If all covariates x take the same value, so that the estimation problem reduces to one of estimating the angular density for an identically-distributed set of pseudo-angles $\{W_i\}_{i=1}^n$, then (3.2) becomes

$$(3.4) \quad \widehat{h}(w) = \frac{1}{n} \sum_{i=1}^n \beta \left(w; \nu \frac{W_i}{2\overline{W}} + \tau, \nu \left\{ 1 - \frac{W_i}{2\overline{W}} \right\} + \tau \right), \quad w \in (0, 1).$$

The version of our estimator in Eq. (3.4) differs from Chen's beta kernel ([Chen, 1999](#)):

$$(3.5) \quad h^*(w) = \frac{1}{n} \sum_{i=1}^n \beta \left(W_i; \frac{w}{s} + 1, \frac{1-w}{s} + 1 \right),$$

where $s > 0$ is a bandwidth. Indeed, (3.5) puts the mode of the kernel at W_i and so does our estimator in (3.4), if we set $\tau = 1$. Yet, in (3.4) w is the argument of $\beta(\cdot)$, whereas in (3.5), W_i is the argument of $\beta(\cdot)$. Estimator (3.4) has closer connections with the smooth Euclidean angular density estimator in [de Carvalho et al. \(2013, p. 1190\)](#), and which is given by

(3.6)

$$\begin{aligned}\widetilde{h}(w) &= \frac{1}{n} \sum_{i=1}^n \{1 - (\overline{W} - 1/2)S^{-2}(W_i - \overline{W})\} \beta\{w; W_i\nu, (1 - W_i)\nu\} \\ &= \frac{1}{n} \sum_{i=1}^n \beta\{w; W_i\nu, (1 - W_i)\nu\} - \frac{1}{n} \sum_{i=1}^n (\overline{W} - 1/2)S^{-2}(W_i - \overline{W})\beta\{w; W_i\nu, (1 - W_i)\nu\},\end{aligned}$$

for $w \in (0, 1)$; here \overline{W} and S^2 are the sample mean and sample variance of W_1, \dots, W_n , that is,

$$\overline{W} = \frac{1}{n} \sum_{i=1}^n W_i, \quad S^2 = \frac{1}{n} \sum_{i=1}^n (W_i - \overline{W})^2.$$

A heuristic argument can be used to see this, by focusing on the case $\tau = 0$. The right-hand term in (3.6) enforces the moment constraint, and hence it is asymptotically negligible, so that for large n , we have $\tilde{h}(w) \approx (1/n) \sum_{i=1}^n \beta\{w; W_i\nu, (1 - W_i)\nu\}$; on the other hand, we also have that for large n , $\hat{h}(w) \approx (1/n) \sum_{i=1}^n \beta\{w; W_i\nu, (1 - W_i)\nu\}$, since $\bar{W} = 1/2 + o_p(1)$, as $n \rightarrow \infty$. While both (3.4) and (3.6) obey the moment constraint (2.6), they impose it through different approaches: our estimator enforces (2.3) by rescaling the pseudo-angles with a factor of $(2\bar{W})^{-1}$; the smooth Euclidean angular density enforces (2.3) additively, through the right-hand term in (3.6). To our knowledge, it is not straightforward to impose the moment constraint on Chen's kernel in (3.5).

3.4. Tuning parameter selection and bootstrap. We select the tuning parameters via maximum likelihood K -fold cross-validation (MLCV) (Hastie et al., 2001, Section 7.10.1). Specifically, let $\{\mathbf{W}_1, \dots, \mathbf{W}_K\}$ be the full sample of pseudo-angles split into K blocks. In the analyses in Sections 4 and 5, we split the blocks according to the values of the accompanying covariate x , so that each $\mathbf{W}_k = (W_{k,1}, \dots, W_{k,n_k})$ is in a similar part of the covariate space. Letting $\hat{h}_{x(-k)}$ denote the estimator leaving out the k th sample, \mathbf{W}_k , of length n_k , we select

$$(3.7) \quad (\hat{b}, \hat{\nu}, \hat{\tau}) = \arg \min_{(b, \nu, \tau) \in \mathcal{R}_{\mathcal{X}, n}} \sum_{k=1}^K \sum_{j=1}^{n_k} -\log \hat{h}_{X_{k,j}(-k)}(W_{k,j}),$$

with

$$(3.8) \quad \begin{aligned} \mathcal{R}_{\mathcal{X}, n} &= \{(b, \nu, \tau) \in (0, \infty)^3 : \nu W_i \theta_b(x) + \tau > 0, \nu \{1 - W_i \theta_b(x)\} + \tau > 0, \text{ for } i = 1, \dots, n; x \in \mathcal{X}\} \\ &= \{(b, \nu, \tau) \in (0, \infty)^3 : \nu \{1 - W_i \theta_b(x)\} + \tau > 0, \text{ for } i = 1, \dots, n; x \in \mathcal{X}\}. \end{aligned}$$

The constrained optimization yields well-defined estimates, since it guarantees the positivity of the beta parameters in our estimator. The latter equality in (3.8) follows from noticing that $\nu W_i \theta_b(x) + \tau > 0$, for all $x \in \mathcal{X}$; further details on practical implementation of tuning parameter selection are given in Section 4.2. It is known that for density estimation, MLCV can produce estimates with suboptimal performance leading to undersmoothed density estimates, especially when the true density has unbounded support (DasGupta, 2008, Section 32.10.1). Computational experiments in the supplementary materials show that the main findings in Section 5 are very similar regardless of whether we use MLCV or least-squares cross-validation (LSCV) (DasGupta, 2008, Section 32.10.2). Better results than the ones in Section 4 are to be expected if LSCV is used. However, LSCV would not be theoretically grounded for non-square integrable densities (e.g., $h_x(w) = \beta(w; x, x)$, for $x \in (0, 1/2)$).

An uncertainty assessment can be performed by simulating from kernel density estimates themselves—in the spirit of the so-called smoothed bootstrap (Silverman and Young, 1987). The procedure detailed below, allows us to generate B bootstrap angular surfaces. For $r \in \{1, \dots, B\}$:

1. Sample j^* from a discrete uniform distribution over $\{1, \dots, n\}$.
2. Sample $X_j^r \sim K_{\hat{b}}(\cdot - X_{j^*})$.
3. Sample $W_j^r \sim \hat{h}_{X_j^r}$ with

$$\hat{h}_{X_j^r}(w) = \sum_{i=1}^n \pi_{\hat{b}, i}(X_j^r) \beta(w; \hat{\nu} W_i \theta_{\hat{b}}(X_j^r) + \hat{\tau}, \hat{\nu} \{1 - W_i \theta_{\hat{b}}(X_j^r)\} + \hat{\tau}), \quad w \in (0, 1),$$

where

$$\theta_{\hat{b}}(X_j^r) = \frac{1/2}{\sum_{i=1}^n \pi_{\hat{b},i}(X_j^r) W_i}, \quad \pi_{\hat{b},i}(X_j^r) = \frac{K_{\hat{b}}(X_j^r - X_i)}{\sum_{k=1}^n K_{\hat{b}}(X_j^r - X_k)}, \quad i = 1, \dots, n.$$

4. Repeat Steps 1–3 n times to obtain the r th bootstrap sample $(\mathbf{X}^r, \mathbf{W}^r)$, with $\mathbf{X}^r = (X_1^r, \dots, X_n^r)^\top$ and $\mathbf{W}^r = (W_1^r, \dots, W_n^r)^\top$.
5. Use $(\mathbf{X}^r, \mathbf{W}^r)$ and (3.7) to obtain bootstrap estimates $(\hat{b}^r, \hat{\nu}^r, \hat{\tau}^r)$.

Using the bootstrap samples $\{(\mathbf{X}^r, \mathbf{W}^r)\}_{r=1}^B$, the bootstrap estimates $\{(\hat{b}^r, \hat{\nu}^r, \hat{\tau}^r)\}_{r=1}^B$, and (3.2), we can construct B bootstrap angular surfaces $\hat{h}_x^1, \dots, \hat{h}_x^B$. For computational convenience, Step 2 considers only a single bandwidth, \hat{b} , but it is known (see, e.g., Polanski, 2001) that smoothed bootstrap resamples need not be generated from the kernel density estimate with the same bandwidth. Indeed, the so-called calibration methods are known to perform well, but they require one to construct a resample over a sequence of bandwidths, and thus are computationally costlier. Visualizing uncertainty of angular surfaces can be awkward, but cross sections of the angular surface (i.e., conditional angular density estimates at fixed values of x) can be easily summarized using, for example, functional boxplots (Sun and Genton, 2011). Details on constructing functional boxplots for angular densities are given in Section 4.2.

3.5. *A local-linear version of the estimator.* A local linear version of our estimator can be readily constructed by replacing the Nadaraya–Watson weights in (3.2) with

$$(3.9) \quad \pi_{b,i}(x) = \frac{1}{n} \frac{\{\hat{s}_2(x; b) - \hat{s}_1(x; b)(X_i - x)\} K_b(X_i - x)}{\hat{s}_2(x; b) \hat{s}_0(x; b) - \hat{s}_1^2(x; b)},$$

where $\hat{s}_m(x; b) = n^{-1} \sum_{i=1}^n (X_i - x)^m K_b(X_i - x)$, for $m = 0, 1, 2$. Local linear regression is often presented as a solution to mitigate boundary bias issues of the Nadaraya–Watson estimator (Wand and Jones, 1995, Section 5.5). Throughout, we consider both Nadaraya–Watson and local linear weights to illustrate their relative performance.

4. Simulation study.

4.1. *Data-generating configurations and preliminary experiments.* We study the performance of our methods under the logistic and Dirichlet conditional models introduced in Examples 1 and 2. Regarding the logistic conditional model, we take $\alpha_x = \Phi(x)$ and consider $x \in \mathcal{X}_{\text{logistic}} = [\Phi^{-1}(0.2), \Phi^{-1}(0.4)]$. For the Dirichlet conditional model we consider two scenarios: a symmetric Dirichlet angular surface with $(a_x, b_x) = (x, x)$, for $x \in \mathcal{X}_{\text{sDir}} = [0.8, 4]$ and an asymmetric Dirichlet angular surface with $(a_x, b_x) = (x, 100)$, for $x \in \mathcal{X}_{\text{aDir}} = [0.5, 2]$. In Figure 2 we plot the true and estimated angular surfaces for the three cases described above on a single experiment with $n = 500$. The top panel of Figure 2 corresponds to the logistic angular surface, where extremal dependence decreases as a function of the predictor. The center panel shows the symmetric Dirichlet angular surface, where we observe weaker dependence for lower values of the covariate, whereas stronger dependence prevails for higher values. Finally, an increasing asymmetric dependence dynamic is displayed in the bottom panel, where we have plotted the asymmetric Dirichlet angular surface.

The single run experiment in Figure 2 allows us to illustrate strengths and limitations with the methods. Even though there is a good fit—which is discussed in further detail in Section 4.2—we can anticipate from this figure that our estimator suffers from limitations inherent to kernel-based

estimators. For example, pointwise estimation using the Nadaraya–Watson weights (middle column of Figure 2) underperforms when the angular surface peaks, but this is mostly due to the boundary bias of K_b which is a drawback of kernel-based estimators on bounded domains; see [Hardle \(1990, Section 4.4\)](#) and references therein. To mitigate this issue we also compute our estimator using local linear weights, as described in Section 3.5 (right column of Figure 2). We see that the performance in the upper boundaries of the covariate space is slightly improved for both Dirichlet angular surfaces, but it remains almost the same for the logistic angular surface. The estimator using local linear weights seems to produce smoother estimates for the asymmetric Dirichlet model. This relative improvement is corroborated in Table 1, where we assess the mean performance of both estimators. Estimates for the other two models tend to be better (in terms of mean performance) using the Nadaraya–Watson weights. In terms of computations, the runtime of the estimator using the Nadaraya–Watson weights outperforms its local linear counterpart by at least a factor of 10. In spite of these limitations, both estimators successfully recover the shape of the true angular surface, and thus are able to reproduce accurately the evolution of extremal dependence over the covariate.

4.2. Simulation results. To construct the simulation studies, we took 1000 samples of sizes 300 and 500 for the three conditional models presented in Section 4.1. For the samples of size 500, Figure 3 displays functional boxplots ([Sun and Genton, 2011](#)) of cross sections of the angular surface (conditional angular density estimates at fixed values of x) along with their Monte Carlo means. Functional boxplots are constructed introducing measures to define functional quantiles and the centrality or outlyingness of a curve. Specifically, [Sun and Genton \(2011\)](#) use band depths to order a sample of curves from the center outwards, defining $100\alpha\%$ central regions ($0 < \alpha < 1$). These central regions can be estimated using the α proportion of deepest curves; a formal definition of these regions can be found in [Sun and Genton \(2011, Section 3\)](#). The gray areas in Figure 3 show the sample 50%, 75%, and 95% central regions of the sampled curves. These plots allow us to illustrate the performance of our estimator in terms of variability, under different dependence dynamics. For example, the logistic model estimates presented in the top panel of Figure 3 turn out to be the most dispersed over all three scenarios. We argue that there are two related reasons for this: the limitations due to boundary bias that were discussed in Section 4.1, and the fact that the range of extremal dependence in the logistic conditional surface is greater compared to the two other scenarios. Both estimators seems to perform similarly, although the local–linear estimator is more variable when the extremal dependence is stronger. The center panel corresponds to the symmetric Dirichlet angular model, which displays a good mean performance in the last two cases, but some bias when the true angular density is U-shaped. Estimates using the Nadaraya–Watson weights seem to be slightly more variable than the ones using the local linear weights. Finally, the asymmetric Dirichlet angular model presented in the bottom panel, displays more dispersed estimates than its symmetric counterpart for both estimators (and between them it seems that the Nadaraya–Watson estimates are again more variable than the ones using local linear weights), although the Monte Carlo mean produces suitable approximations. The asymmetry does not seem to be a major issue. Overall, estimates for the three models display reasonable performance in recovering the different shapes of the densities, and Monte Carlo means produce reliable estimates. Monte Carlo mean surfaces for the three models and the two estimators can be found in the supplementary material.

We assess the performance of our estimator using the mean integrated absolute error (MIAE),

$$(4.1) \quad \text{MIAE} = \mathbb{E} \left(\int_{\mathcal{X}} \int_0^1 |\hat{h}_x(w) - h_x(w)| dw dx \right),$$

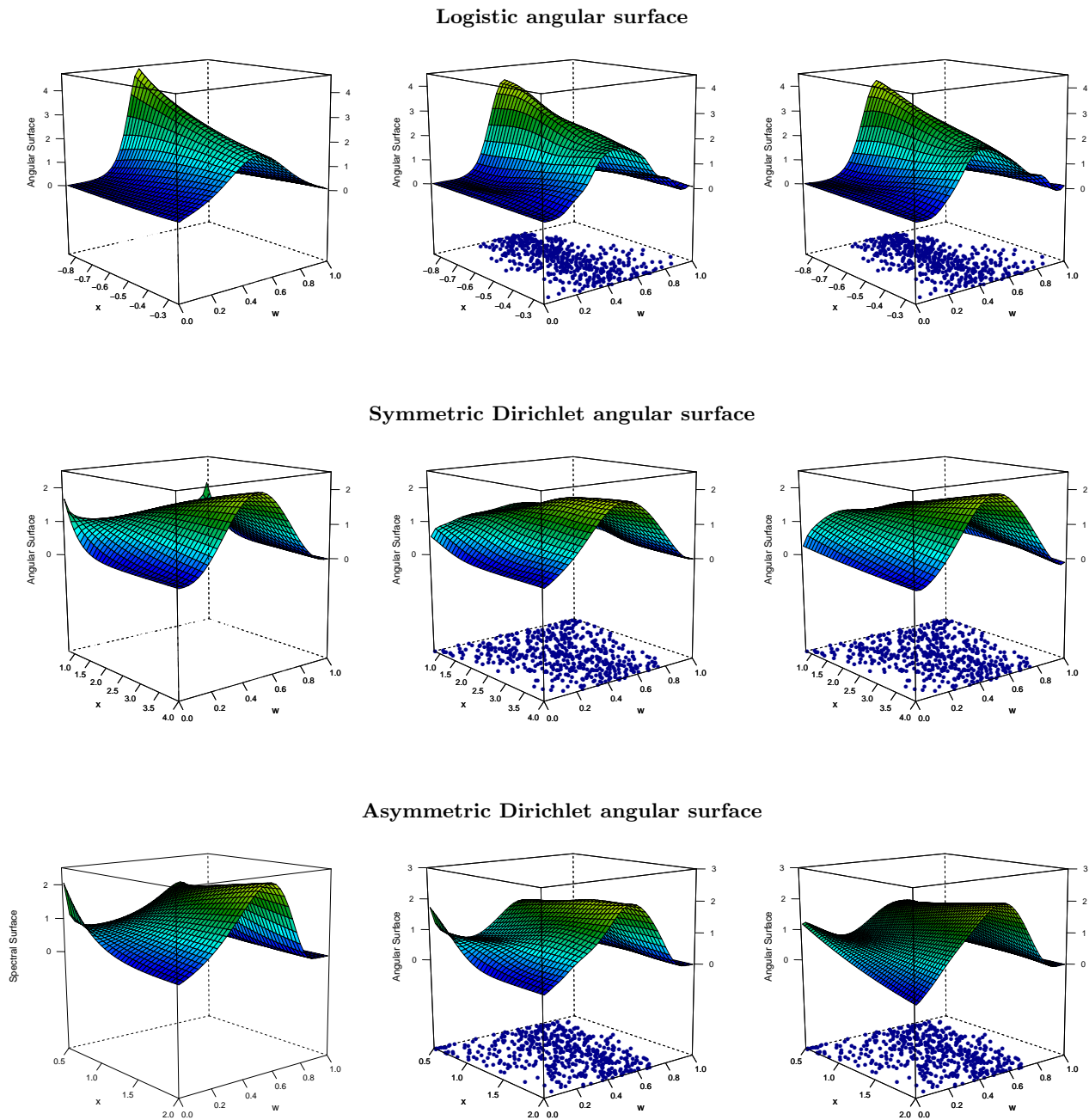


FIG 2. True angular surfaces (left) and corresponding estimates using Nadaraya-Watson weights (middle) and local linear weights (right). Top panel: conditional logistic model with $\alpha_x = \Phi(x)$, for $x \in \mathcal{X}_{logistic} = [\Phi^{-1}(0.2), \Phi^{-1}(0.4)]$. Center panel: conditional Symmetric Dirichlet model with $(a_x, b_x) = (x, x)$, for $x \in \mathcal{X}_{sDir} = [0.8, 4]$. Bottom panel: conditional Asymmetric Dirichlet model with $(a_x, b_x) = (x, 100)$, for $x \in \mathcal{X}_{aDir} \in [0.5, 2]$. The simulated pseudo-angles based on which the estimates are produced are overlaid on the bottom of the boxes.

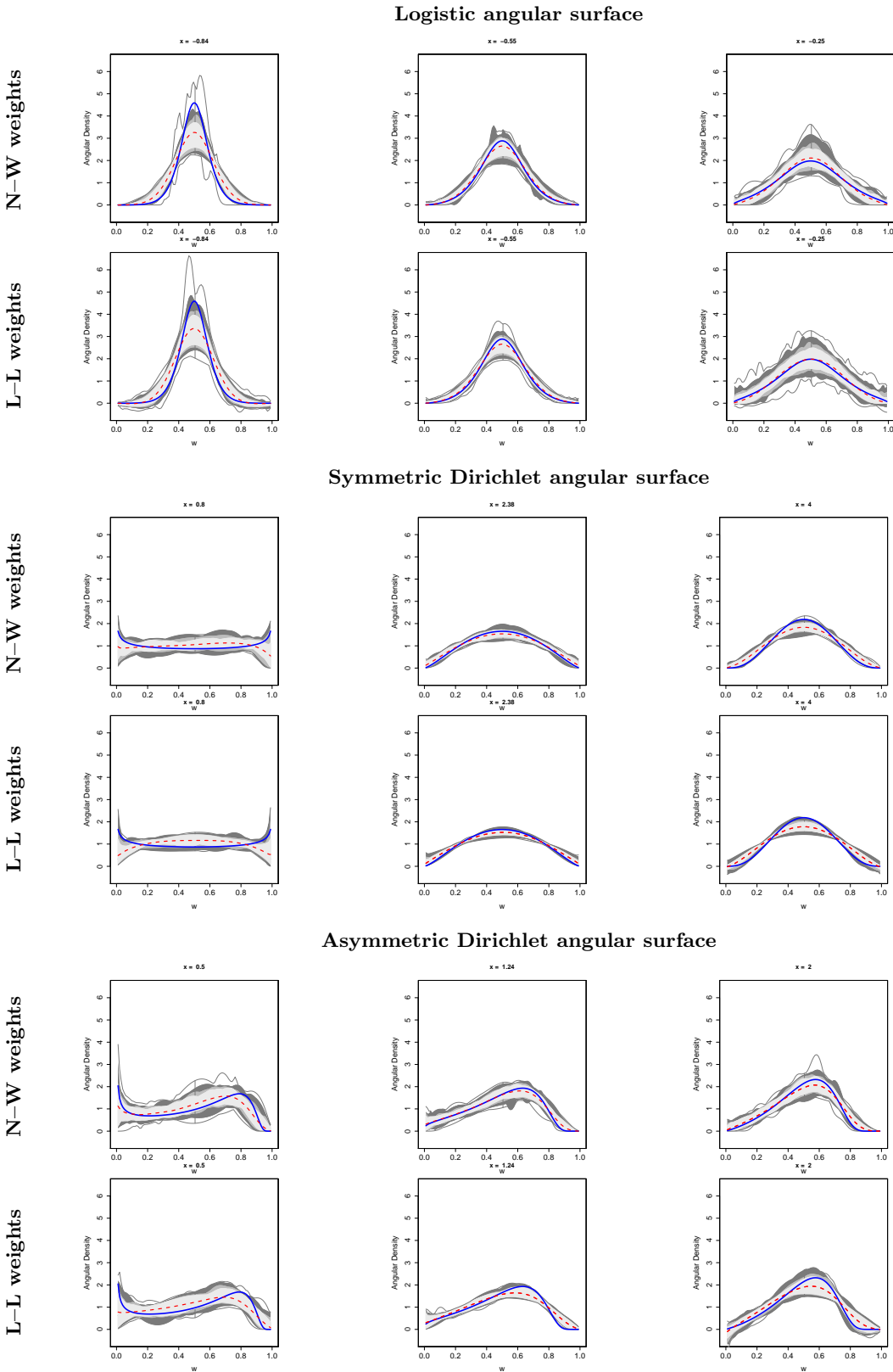


FIG 3. Functional boxplots (gray shadow) showing the 50%, 75%, and 95% central regions (as defined by *Sun and Genton (2011)*) of 1000 samples of size 500 for the conditional models presented in Section 4.1, as well as their corresponding true values (solid blue line) and Monte Carlo means (dashed red line).

TABLE 1

Mean integrated absolute error estimates computed over 1000 samples for the data-generating configurations discussed in Section 4.1 for the Nadaraya–Watson (N–W) and the local linear (L–L) weights.

n	conditional Model	Specification	MIAE	
			N–W weights	L–L weights
300	Logistic	$\alpha_x = \Phi(x)$	0.09	0.92
	Symmetric Dirichlet	$(a_x, b_x) = (x, x)$	0.42	0.60
	Asymmetric Dirichlet	$(a_x, b_x) = (x, 100)$	0.63	0.59
500	Logistic	$\alpha_x = \Phi(x)$	0.08	0.14
	Symmetric Dirichlet	$(a_x, b_x) = (x, x)$	0.39	0.55
	Asymmetric Dirichlet	$(a_x, b_x) = (x, 100)$	0.62	0.55

and report the results in Table 1. As mentioned before, we can see that the estimator using the Nadaraya–Watson weights outperforms the one using local linear weights in the logistic and symmetric Dirichlet models, but the local linear weights seem to be a better choice for the asymmetric Dirichlet model. In any case and except for the logistic model with sample size 300, the improvements of one estimator over the other are fairly modest. As we should expect, the results show that performance increases with sample size. Overall, simulations confirm that our methods produce acceptably accurate estimates of the angular surface.

We conclude this section providing some comments on implementation of the tuning parameter selection (Section (3.4)). Since in some cases optimization over $\mathcal{R}_{\mathcal{X},n}$ (defined in Eq. (3.8)) can be computationally expensive, our experiments suggest that optimization over \mathcal{R}_n defined as

$$\mathcal{R}_n = \{(b, \nu, \tau) \in (0, \infty)^3 : \nu\{1 - W_i\theta_b(X_j)\} + \tau > 0, \text{ for } i, j = 1, \dots, n\},$$

performs reasonably well. Note that \mathcal{R}_n is a version of $\mathcal{R}_{\mathcal{X},n}$ determined only by the observed covariate values, and not by the entire covariate space \mathcal{X} . Furthermore, for large n , unconstrained optimization over $(0, \infty)^3$ typically also performs well. We thus recommend the user to initially try unconstrained optimization for large n , or optimization over \mathcal{R}_n for moderate n . Only if the resulting parameter values do not yield a valid estimator over the study region of interest does one then need to implement the constrained optimization over $\mathcal{R}_{\mathcal{X},n}$.

5. Dynamics of joint extremal losses in leading European stock markets.

5.1. *Background and motivation for empirical analysis.* In 1999, 11 European Union (EU) countries formed the Economic and Monetary Union (EMU), which led them to adopt a common currency and monetary policy as well as the conduction of coordinated economic policies.

The process of creation of the EMU was the outcome of three stages of development, further details of which can be found on the European Central Bank website:

<https://www.ecb.europa.eu/ecb/history>

see also James (2012). To join the Eurozone (countries who adopted the Euro as their common currency) member states had to qualify by meeting the criteria of the Maastricht Treaty in terms of budget deficits, inflation, interest rates, and other monetary requirements. At the moment the Euro is the single currency shared by 19 of the 28 EU members. The remaining 9 countries, including the UK, are endowed with ‘opt-out’ clauses which exempts them from using the Euro as their currency. In recent years there have been several studies providing evidence for an increased integration of European stock markets, and the EMU has been frequently put forward

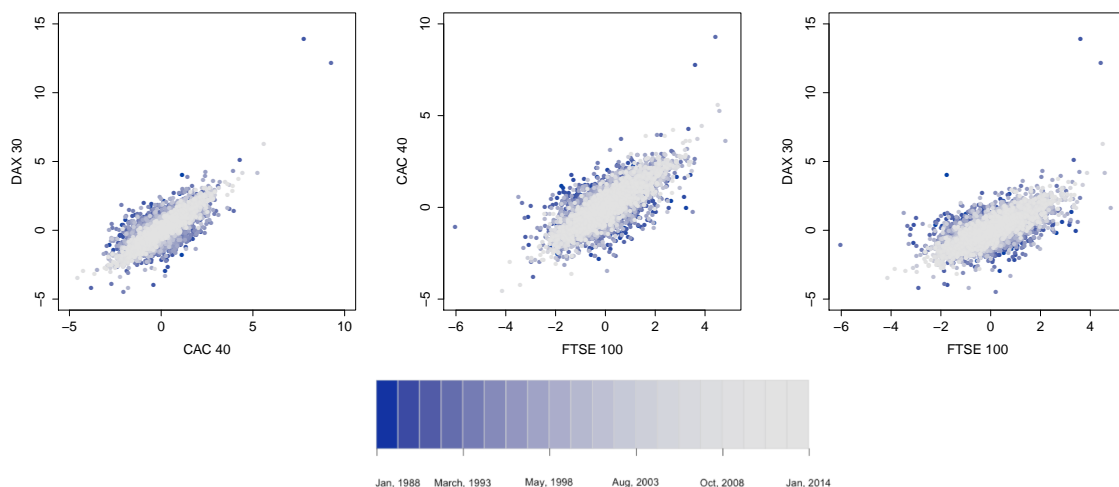


FIG 4. Scatterplots using a time-varying color palette for GARCH-filtered residuals for CAC 40 (FR), DAX 30 (DE) and FTSE 100 (UK) spanning the period from January 1, 1988 to January 1, 2014.

as the causal driver for this increase, along with some other determinants (Fratzscher, 2002; Kim et al., 2005; Hardouvelis et al., 2006; Büttner and Hayo, 2011, and the references therein). Hardouvelis et al. (2006) found however that the UK, who chose not to enter the eurozone, showed no increase in stock market integration by that time.

Although there is a wealth of studies analyzing stock market integration over time, few attempts have been made to ascertain the dynamics governing extreme value dependence of stock market returns over time. The huge literature looking into dependence of financial markets (see for example King et al., 1994; Longin and Solnik, 1995, 2001; Karolyi and Stulz, 1996; Forbes and Rigobon, 2002; Brooks and Del Negro, 2004, 2005; Rua and Nunes, 2009) has collected evidence compatible with the hypothesis that the comovement of returns has not remained constant over time. Yet, none of these papers has focused on tracking the dynamics of extremal dependence of returns, which is the object of the current inquiry. An exception in this respect is the seminal paper of Poon et al. (2003), which provides evidence of increasing levels of extremal dependence for three major stock markets within Europe [CAC (France), DAX (Germany), and FTSE (UK)]. The subperiod analysis of Poon et al. (2003, Section 3.3.2) is however exploratory, in the sense that they arbitrarily partitioned the sample period into three periods, and thus estimation of extremal dependence on each period only takes data from that period into account.

Below, we apply our methods to address a similar question to that of Poon et al. (2003, 2004). Specifically, one of our main interests is disentangling the dynamics governing the dependence of extreme losses on three leading European stock markets—using CAC, DAX, and FTSE—in recent years. The motivation for choosing these markets is twofold: these are the stock markets of the European members of G5; these are also the same European stock markets considered by Poon et al. (2003, 2004). Moreover they display a stronger type of extremal dependence than some of the other markets studied by Poon et al. (2003, 2004), i.e., asymptotic dependence as defined in Section 2.1.

5.2. *Data description, preprocessing, and exploratory considerations.* Our data were gathered from Datastream and consist of daily closing stock index levels of three leading European stock

markets: CAC 40, DAX 30, and FTSE 100 (henceforth CAC, DAX, and FTSE). The sample period spans from January 1, 1988 to January 1, 2014 ($N = 6784$ observations), and hence it includes the Great Moderation and Great Recession which are by all standards challenging modeling issues. Since we want to focus on extreme losses, we use daily negative returns as a unit of analysis. Daily negative returns are computed by taking the negative of the first differences of the logarithmic indices. Following the bivariate analysis in [Poon et al. \(2004\)](#) both observations of a particular day are removed if at least one of the two observations is a zero return (plots of the data and summary statistics can be found in the supplementary material). The *Engle* (1) statistic of [Engle \(1982\)](#) (not reported here) is large and significant for all three stock return series, indicating strong heteroskedasticity which can be removed by fitting volatility filters. In the spirit of [Poon et al. \(2004\)](#) we fit three different filters: GARCH(1,1) assuming t -distributed errors for CAC and normal for FTSE and DAX, NGARCH (also known as nonlinear asymmetric GARCH) with normal innovations, and the stochastic volatility model (SV) of [Kim et al. \(1998\)](#) with hyperparameters chosen according to the latter paper. Diagnostic plots (not shown here) suggest that the GARCH fits are superior than the NGARCH fits for the three stock markets, and heteroskedasticity is successfully removed with the GARCH and NGARCH filters, but not with the SV filter. The results shown below correspond to the GARCH-filtered residuals, but similar conclusions can be drawn using the NGARCH filter (angular surfaces based on the NGARCH-filtered residuals can be found in the supplementary material). Scatterplots of possible combinations of pairs of filtered residual series are displayed in [Figure 4](#), depicted using a time-varying color palette which allows us to uncover the nonstationary nature of joint extremes. This is in line with the findings of [Poon et al. \(2003, 2004\)](#).

To verify that our methods are a sensible approach for modeling these data, we need to assess whether the filtered residuals are asymptotically dependent. As mentioned in [Section 2.1](#), in the modeling of extreme events two different classes of extreme value dependence can arise: asymptotic dependence and asymptotic independence. Dependence between moderately large values can arise in both cases, but the very largest values from each variable can occur together only under asymptotic dependence. To make ideas concrete, let Y_1 and Y_2 be any two filtered residuals of interest, transformed to have unit Fréchet margins. Under an exploratory setting, two measures of tail dependence can be obtained to summarize the strength of extremal dependence:

$$\chi = \lim_{u \rightarrow \infty} \mathbb{P}(Y_1 > u \mid Y_2 > u), \quad \bar{\chi} = \lim_{u \rightarrow \infty} \frac{2 \log \mathbb{P}(Y_1 > u)}{\log \mathbb{P}(Y_1 > u, Y_2 > u)} - 1.$$

Here, $\chi \in [0, 1]$ measures the strength of dependence within the class of asymptotically dependent variables, whereas $\bar{\chi} \in [-1, 1]$ is often used to measure the strength of dependence within the class of asymptotically independent variables. Taken together, the pair $(\chi, \bar{\chi})$ provides a summary of extremal dependence for the vector (Y_1, Y_2) . For asymptotically dependent variables we have $\bar{\chi} = 1$ and the value of $\chi > 0$ increases with the strength of dependence at extreme levels. For asymptotically independent variables we have $\chi = 0$ and $\bar{\chi} \leq 1$ increases with the strength of dependence at extreme levels. Roughly speaking, if $\bar{\chi} > 0$ then we often speak about ‘positive extremal dependence,’ whereas if $\bar{\chi} < 0$ we use the expression ‘negative extremal dependence.’ Indeed, for the bivariate normal dependence structure $\bar{\chi}$ corresponds to Pearson correlation; see [Heffernan \(2000\)](#) for further examples.

In [Figure 5](#) we present rolling window estimates of χ and $\bar{\chi}$ with approximate 95% confidence intervals, which is tantamount to the subperiod analysis of [Poon et al. \(2003, Section 3.3.2\)](#). The rolling window estimates were computed using the empirical estimators of χ and $\bar{\chi}$ ([Beirlant et al., 2004, p. 348](#)) at the 95% quantile for moving windows of 600 observations. Given

the large uncertainty entailed in the estimation of χ , interpretation of these plots is far from straightforward. Nevertheless, pointwise estimation for χ seems reasonably different from 0 for the three pairs under study, and despite some drops around 1992 and 2000, there seems to be an increasing trend for the three cases. Moreover values for $\bar{\chi}$ are closer to 1 as time passes. This combined information indicates that the assumption of asymptotic dependence is certainly plausible for the later years, and might be adequate for earlier years. We discuss the asymptotic independence issue again in Section 6.

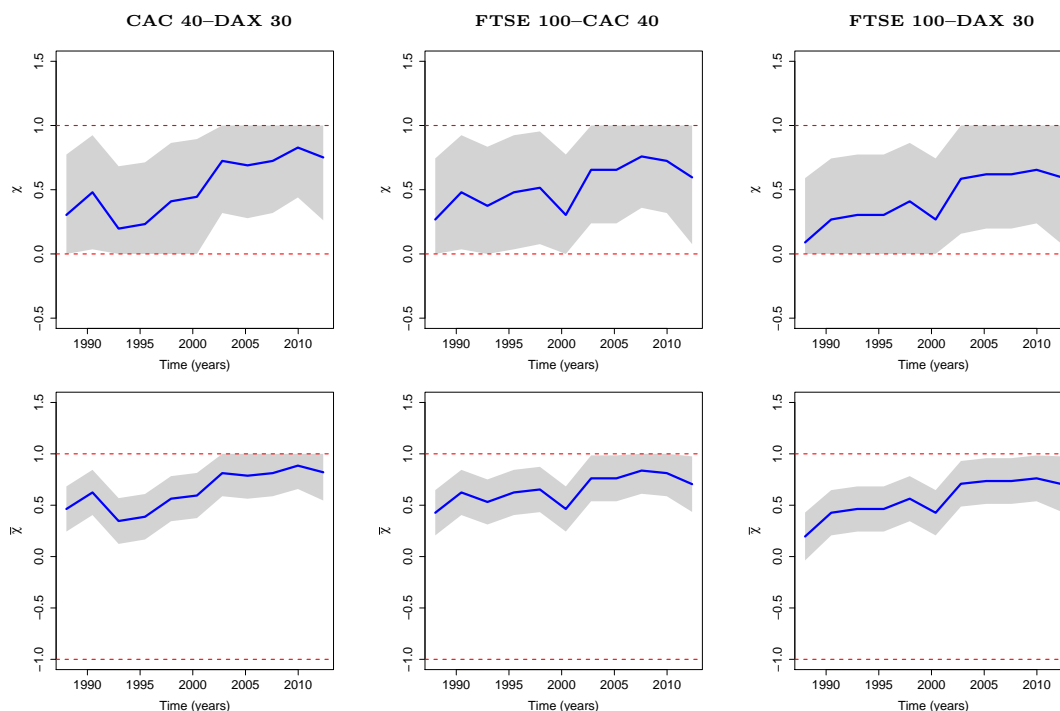


FIG 5. Rolling window estimates of χ (top) and $\bar{\chi}$ (bottom) at the 95% quantile using moving windows of size 600, applied to the three pairs under study.

5.3. *Modeling time-varying extremal dependence.* The time-varying color palette scatterplots in Figure 4 and the rolling window estimates in Figure 5 provide evidence of nonstationary extremal dependence, but they are only exploratory. In this section we complete the analysis from Section 5.2 by applying our conditional modeling approach to assess how the dependence structure of bivariate extreme losses in the three pairs has been evolving over recent years. Before we proceed any further, some comments regarding implementation are in order. As mentioned in Section 2, the data were transformed to have standard Fréchet margins. This was done as follows. Given a sample of pairs of filtered residuals $(r_{1,1}, r_{1,2}), \dots, (r_{N,1}, r_{N,2})$, we construct proxies for the unobservable pseudo-angles W_i by setting

$$W_i = \hat{Y}_{i,1}/(\hat{Y}_{i,1} + \hat{Y}_{i,2}), \quad R_i = \hat{Y}_{i,1} + \hat{Y}_{i,2},$$

where $\hat{Y}_{i,1} = -1/\log\{\hat{F}_{r_1}(r_{i,1})\}$ and $\hat{Y}_{i,2} = -1/\log\{\hat{F}_{r_2}(r_{i,2})\}$ and where \hat{F}_{r_1} and \hat{F}_{r_2} are estimates of the marginal distribution functions F_{r_1} and F_{r_2} . A robust choice for \hat{F}_{r_1} and \hat{F}_{r_2} is the pair of univariate empirical distribution functions, normalized by $N + 1$ rather than by

N to avoid division by zero. Following Section 3.1, after fitting a spline-based nonparametric quantile regression we found evidence of dependence of the pseudo-radii $\{R_1, \dots, R_N\}$ on time, and so we proceed under a nonstationary assumption. Specifically, we model the 95% quantile of the pseudo-radii through nonparametric quantile regression and threshold the pseudo-radii according to the fit. The tail region to study the extreme losses is therefore defined through the pseudo-angles associated with the threshold exceedances of the pseudo-radii. After thresholding, the number of pseudo-angles is 312 for CAC–DAX and FTSE–CAC and 314 for FTSE–DAX. The pseudo-angles corresponding to these observations are plotted in the two-dimensional bottom plane in Figure 7. The tuning parameters (b, ν, τ) were computed as discussed in Sections 3.4 and 4.2.

In Figure 6 we plot cross sections of the angular surface estimate, using both Nadaraya–Watson and local linear weights as described in Section 3, at three important periods on the EU agenda: I) Beginning of stage one of EMU (1 July, 1990); II) beginning of stage three of EMU (1 January, 1999); III) activation of the assistance package for Greece (2 May, 2010), the first country to be shut out of the bond market, which fostered the European sovereign debt crisis (Lane, 2012). The choice of landmarks I–III) is arbitrary, but recall that our main interest is in describing how extremal dependence may change, by comparing periods sufficiently apart in time. As can be observed from the first column in Figure 6, at around 1990 the dependence between extreme losses for the three pairs were similar, exhibiting some evidence of extremal independence, that is also reflected in Figure 5. The second column in Figure 6 reveals that about a decade later this dynamic changed, and that extreme losses started to show some mild signs of extremal dependence. These signs become stronger, and 11 years later (third column in Figure 6) we can clearly see evidence of extremal dependence of joint losses. Our findings may seem to contradict Hardouvelis et al. (2006)—who claimed that the UK showed no increase in stock market integration—however we note that Hardouvelis et al. (2006) did not assess extremal dependence. The functional boxplots in Figure 6 were obtained following the bootstrap procedure detailed in Section 3.4, with $B = 1000$ samples. We can clearly see some differences between the two types of estimators among the three pairs, but overall they report similar information in terms of the extremal dependence.

Figure 6 provides only a few snapshots corresponding to landmarks I–III). A more complete portrait of the temporal changes in extremal dependence is provided by the angular surface estimate in Figure 7, from which the cross-sections in Figure 6 are derived.

All in all, we can clearly see the change from weaker dependence around 1990 to strong dependence starting from 2005, thus suggesting that in recent decades there has been an increase in the extremal dependence in the losses for these leading European stock markets. The pair CAC–DAX is the one where extremal dependence peaks the most, thus suggesting a high level of synchronization and comovement of extreme losses in those markets over recent years.

Similar conclusions can be drawn from Figure 8, where we plot the conditional extremal coefficient, as defined in Section 2.3. The extremal coefficient is equal to $2 - \chi$, and as such is equal to 2 under asymptotic independence, and takes values in $[1, 2)$ under asymptotic dependence. Figure 8 permits comparison with the results of Poon et al. (2004), who calculated χ over subperiods. The red lines in Figure 8 represent the values from the analysis of Poon et al. for the subperiod November 1990–November 2001 (cf Poon et al., 2004, Table 3). Specifically, Poon et al. report the following values of χ for: CAC–DAX, 0.517 (0.037); FTSE–CAC, 0.532 (0.035) and FTSE–DAX, 0.459 (0.039), with standard errors in parentheses. As can be seen from Figure 8, the magnitudes of the extremal coefficients estimated by Poon et al. are in reasonable agreement

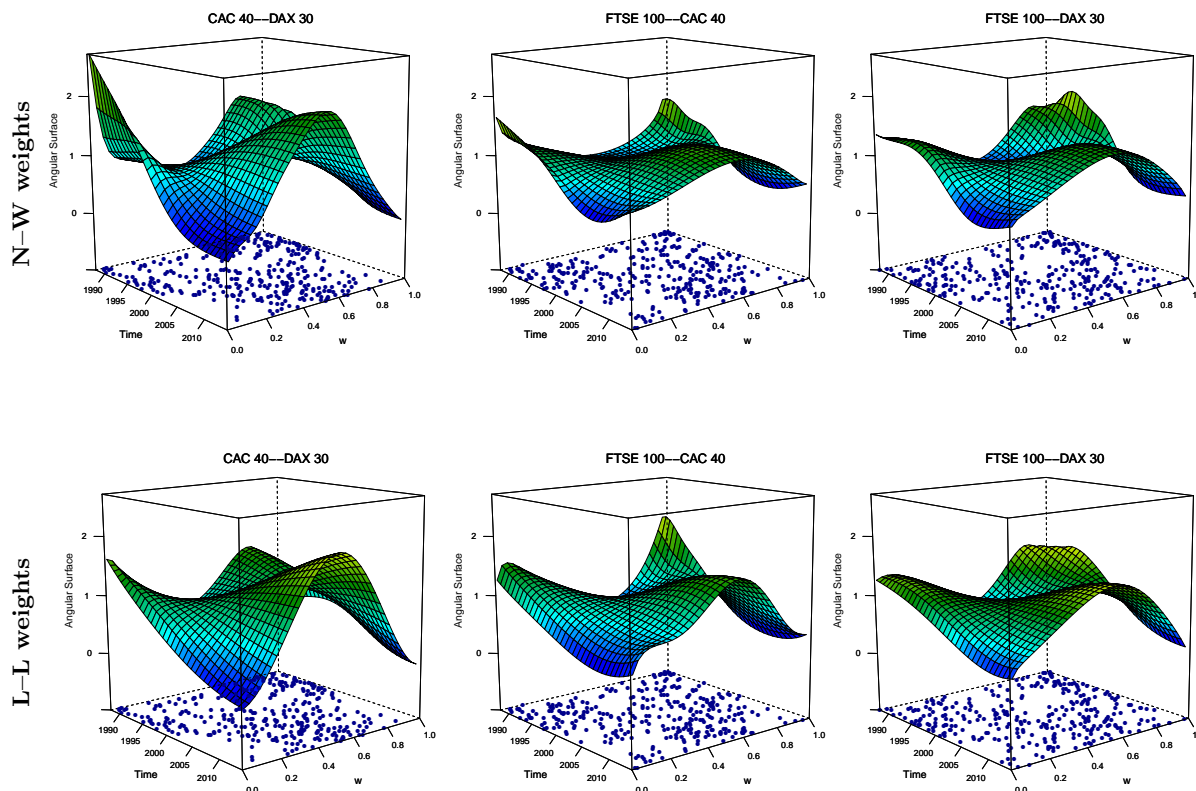


FIG 7. Angular surfaces estimates for CAC–DAX, FTSE–CAC and FTSE–DAX using Nadaraya–Watson (top) and local linear (bottom) weights, with pseudo-angles overlaid on the bottom of the box.

with the ones computed with our methods when uncertainty is taken into account.

6. Final comments. This paper develops methods for modeling nonstationary extremal dependence structures, motivated by the need to assess the comovement of extreme losses in some leading European stock markets over recent years. Although there are many studies analyzing stock market integration over time (see for example King et al., 1994; Longin and Solnik, 1995, 2001; Karolyi and Stulz, 1996; Forbes and Rigobon, 2002; Brooks and Del Negro, 2004, 2005; Rua and Nunes, 2009), few attempts have been made to assess the dynamics of extreme value dependence of stock market returns over time. An exception in this regard is the paper of Poon et al. (2003), which provides evidence suggesting increasing levels of extremal dependence for CAC, DAX and FTSE, although their analysis is essentially exploratory. The analysis performed in this paper reveals a more complete picture of this temporally-changing dependence.

Two related approaches to the current work are the so-called spectral density ratio model of de Carvalho and Davison (2014) and the spectral density regression model of Castro Camilo and de Carvalho (2016). While flexible, these approaches only apply to the setting where there are several pseudo-angles corresponding to the same value of the predictor—and thus they are inappropriate for our applied setting of interest. Our methods are more resilient in the sense that they do not require a sample of pseudo-angles for each value of the covariate, but apply more generally to a regression setting where each covariate value may only have a single corresponding

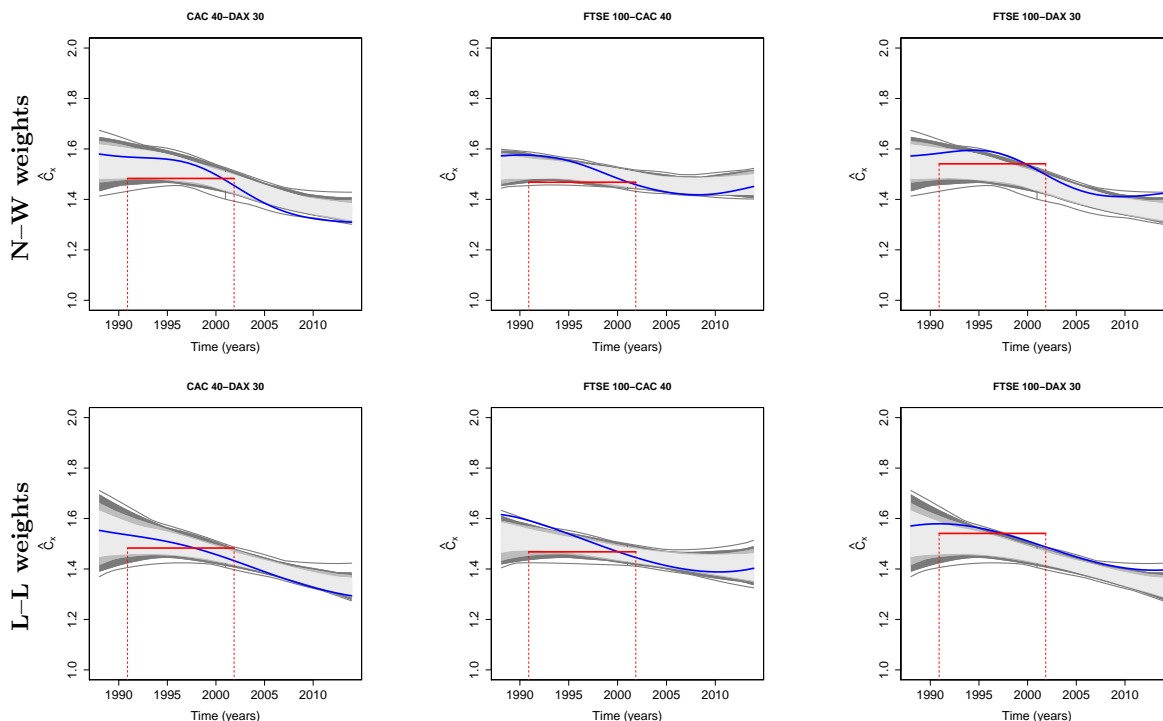


FIG 8. Conditional extremal coefficients (solid blue lines) and functional boxplots (gray shadows) showing the 50%, 75%, and 95% central regions (as defined by Sun and Genton (2011)) based on 1000 bootstrap samples. The red lines represent the values from the analysis of Poon et al. for the subperiod November 1990–November 2001 (cf Poon et al., 2004, Table 3).

pseudo-angle. Recent preprints of Escobar-Bach et al. (2016) and Mhalla et al. (2017) suggest methods for estimating Pickands dependence function under covariate dependence, offering alternative approaches to those presented herein.

Computational experiments suggest that U-shaped angular surfaces are much more difficult to fit. Whilst absolute errors may become large at the boundaries when we have an unbounded density, when this is translated to other quantities (H_x or A_x), the errors will be much less noticeable. Our methods have been developed with the setting of asymptotic dependence in mind, but certainly there is room for developing methodology for conditional modeling under asymptotic independence. Indeed, in common with any approach based on multivariate extreme value distributions, a limitation with our methods is that they will overestimate risk if data are asymptotically independent. Figure 5 gave some indication of possible asymptotic independence near the beginning of the analysis period. As such, the need for developing conditional models able to cope with both asymptotic dependence and asymptotic independence is of utmost importance.

Acknowledgements. We thank the Editor, AE, and two anonymous referees. We extend our thanks to António Rua, Vanda Inácio de Carvalho, and Claudia Wehrhahn for helpful discussions. Part of this work was written while D. C. was visiting the University of Cambridge—Statistical Laboratory, and while M. de C. was visiting *Banco de Portugal*. The research was partially funded by by Fundação para a Ciência e a Tecnologia, through UID/MAT/00006/2013

and by the Chilean National Science Foundation through Fondecyt 11121186, “Constrained Inference Problems in Extreme Value Modeling.”

SUPPLEMENTARY MATERIAL

Supplementary Monte Carlo evidence and empirical reports. The supplement includes additional simulation results, descriptive statistics for daily stock index negative returns, and further empirical analysis using the NGARCH-filtered residuals and LSCV bandwidths.

REFERENCES

- ACAR, E. F., CRAIU, R. V. and FANG, Y. (2011). Dependence calibration in conditional copulas: A nonparametric approach. *Biometrics* **67** 445–453.
- BEIRLANT, J., GOEGBEUR, Y., SEGERS, J. and TEUGELS, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley, New York.
- BROOKS, R. and DEL NEGRO, M. (2004). The rise in comovement across national stock markets: market integration or IT bubble? *J. Empir. Finance* **11** 659–680.
- BROOKS, R. and DEL NEGRO, M. (2005). Country versus region effects in international stock returns. *J. Portfolio Management* **31** 67–72.
- BÜTTNER, D. and HAYO, B. (2011). Determinants of European stock market integration. *Econ. Sys.* **35** 574–585.
- CASTRO CAMILO, D. and DE CARVALHO, M. (2016). Spectral density regression for bivariate extremes. *Stoch. Env. Res. Risk Ass.* DOI: 10.1007/s00477-016-1257-z.
- CHAVEZ-DEMOULIN, V. and DAVISON, A. C. (2005). Generalized additive modelling of sample extremes. *J. R. Statist. Soc. C* **54** 207–222.
- CHEN, S. X. (1999). Beta kernel estimators for density functions. *Comput. Statist. Data Anal.* **31** 131–145.
- CLINE, D. B. H. (1988). Admissible kernel estimators of a multivariate density. *Ann. Statist.* **16** 1421–1427.
- COLES, S. G. (2001). *An Introduction to the Statistical Modeling of Extreme Values*. Springer, London.
- COLES, S. G. and TAWN, J. A. (1991). Modelling extreme multivariate events. *J. R. Statist. Soc. B* **53** 377–392.
- DASGUPTA, A. (2008). *Asymptotic Theory of Statistics and Probability*. New York, Springer.
- DAVISON, A. C. and SMITH, R. L. (1990). Models for exceedances over high thresholds. *J. R. Statist. Soc. B* **52** 393–442.
- DE CARVALHO, M. (2016). Statistics of extremes: Challenges and opportunities. In: *Handbook of Extreme Value Theory and its Applications to Finance and Insurance*, Eds F. Longin. Wiley, Hoboken.
- DE CARVALHO, M. and DAVISON, A. C. (2014). Spectral density ratio models for multivariate extremes. *J. Amer. Statist. Assoc.* **109** 764–776.
- DE CARVALHO, M., OUMOW, B., SEGERS, J. and WARCHOL, M. (2013). A Euclidean likelihood estimator for bivariate tail dependence. *Comm. Statist. Theo. Meth.* **42** 1176–1192.
- DE HAAN, L. and RESNICK, S. I. (1977). Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **40** 317–377.
- EASTOE, E. and TAWN, J. (2009). Modelling non-stationary extremes with application to surface level ozone. *J. R. Statist. Soc. C* **58** 22–45.
- EASTOE, E. F. (2009). A hierarchical model for non-stationary multivariate extremes: A case study of surface-level ozone and NO_x data in the UK. *Environmetrics* **20** 428–444.
- ENGLE, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1007.
- ESCOBAR-BACH, M., GOEGBEUR, J. and GUILLOU, A. (2016). Local robust estimation of the Pickands dependence function. *Submitted*.
- FERMANIAN, J.-D. and MARTEN, H. (2012). Time-dependent copulas. *J. Mult. Anal.* **110** 19–29.
- FORBES, K. and RIGOBON, R. (2002). No contagion, only interdependence: Measuring stock market comovements. *J. Finance* **57** 2223–2261.
- FRATZSCHER, M. (2002). Financial market integration in Europe: On the effects of EMU on stock markets. *Int. J. Fin. Econ.* **7** 165–193.
- GUDENDORF, G. and SEGERS, J. (2010). Extreme-value copulas. In: *Copula Theory and Its Applications*, Eds P. Jaworski et al. Springer, New York.

- GEENENS, G. (2014). Probit transformation for kernel density estimation on the unit interval. *J. Amer. Statist. Assoc.* **109** 346–359.
- HARDLE, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Cambridge UK.
- HARDOUVELIS, G. A., MALLIAROPULOS, D. and PRIESTLEY, R. (2006). EMU and European stock market integration. *J. Bus.* **79** 365–392.
- HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, New York.
- HEFFERNAN, J. E. (2000). Directory of coefficients of tail dependence. *Extremes* **3** 279–290.
- HEFFERNAN, J. E. and TAWN, J. A. (2004). A conditional approach for multivariate extreme values (with Discussion). *J. R. Statist. Soc. B* **66** 497–546.
- HUSER, R. and GENTON, M. G. (2016). Non-stationary dependence structures for spatial extremes. *Journal of Agricultural, Biological, and Environmental Statistics* **21** 470–491.
- JAMES, H. (2012). *Making the European Monetary Union*. Harvard University Press, Cambridge MA.
- JONATHAN, P., EWANS, K. and RANDELL, D. (2014). Non-stationary conditional extremes of northern north sea storm characteristics. *Environmetrics* **25** 172–188.
- JONES, M. C. and HENDERSON, D. A. (2007). Kernel-type density estimation on the unit interval. *Biometrika* **94** 977–984.
- LONGIN, F. and SOLNIK, B. (1995). Is the correlation in international equity returns constant: 1960–1990? *J. Int. Money Finance* **14** 3–26.
- LONGIN, F. and SOLNIK, B. (2001). Extreme correlation of international equity market. *J. Finance* **56** 649–676.
- KAROLYI, G. A. and STULZ, R. M. (1996). Why do markets move together? An investigation of US–Japan stock return comovements. *J. Finance* **51** 951–986.
- KIM, S., SHEPHARD, N. and CHIB, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Rev. Econ. Stud.* **65** 361–393.
- KIM, S. J., MOSHIRIAN, F. and WU, E. (2005). Dynamic stock market integration driven by the European monetary union: An empirical analysis. *J. Bank. Fin.* **29** 2475–2502.
- KING, M., SENTANA, E. and SUSHIL, W. (1994). Volatility and links between national stock markets. *Econometrica* **62**, 901–933.
- KOENKER, R. (2005). *Quantile Regression*. Cambridge University Press, Cambridge UK.
- LANE, P. R. (2012). The European sovereign debt crisis. *J. Econ. Persp.* **26** 49–67.
- MHALLA, L., CHAVEZ-DEMOULIN, V. and NAVEAU, P. (2017). Non linear models for extremal dependence. Available at SSRN: <https://ssrn.com/abstract=2836587>.
- NADARAYA, E. A. (1964). On estimating regression. *Theo. Prob. Appl.* **9** 141–142.
- NATURAL ENVIRONMENT RESEARCH COUNCIL (1975). *Flood Studies Report*. Natural Environment Research Council, London.
- PATTON, A. J. (2006). Modelling asymmetric exchange rate dependence. *Int. Econ. Rev.* **47** 527–556.
- PICKANDS, J. (1981). Multivariate extreme value distributions. In: *Proceedings 43rd Session International Statistical Institute*, pp. 859–878.
- POLANSKI, A. M. (2001). Bandwidth selection for the smoothed bootstrap percentile method. *Comput. Statist. Data Anal.* **36** 333–349.
- POON, S.-H., ROCKINGER, M. and TAWN, J. A. (2003). Modelling extreme-value dependence in international stock markets. *Statist. Sinica* **13** 929–953.
- POON, S.-H., ROCKINGER, M. and TAWN, J. A. (2004). Extreme-value dependence in financial markets: Diagnostics, models and financial implications. *Rev. Fin. Stud.* **17** 581–610.
- RESNICK, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York.
- RUA, A. and NUNES, L. C. (2009). International comovement of stock market returns: A wavelet analysis. *J. Emp. Finance* **16** 632–639.
- SILVERMAN, B. W. and YOUNG, G. A. (1987). The bootstrap: To smooth or not to smooth? *Biometrika* **74** 469–479.
- SUN, Y. and GENTON, M. G. (2011). Functional boxplots. *J. Comp. Graph. Statist.* **20** 316–334.
- VERAVERBEKE, N., OMEKA, M. and GIJBELS, I. (2007). Estimation of a conditional copula and association measures. *Scand. J. Statist.* **38** 766–780.
- TAWN, J. A. (1988). Bivariate extreme value theory: Models and estimation. *Biometrika* **75** 397–415.
- TAWN, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika* **77** 245–253.
- WAND, M. P. and JONES, M. C. (1995). *Kernel Smoothing*. Chapman & Hall/CRC, Boca Raton FL.
- WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā A* **26** 359–372.

D. CASTRO CAMILO
CEMSE DIVISION
KING ABDULLAH UNIVERSITY OF SCIENCE AND TECHNOLOGY
THUWAL, KINGDOM OF SAUDI ARABIA
E-MAIL: daniela.castro@kaust.edu.sa

M. DE CARVALHO
SCHOOL OF MATHEMATICS
UNIVERSITY OF EDINBURGH
EDINBURGH, UK
E-MAIL: miguel.decarvalho@ed.ac.uk

J. WADSWORTH
DEPARTMENT OF MATHEMATICS AND STATISTICS
LANCASTER UNIVERSITY
LANCASTER, UK
E-MAIL: j.wadsworth@lancaster.ac.uk