Nonstationary extremes and the US business cycle

M. B. de Carvalho

Ecole Polytechnique Fédérale de Lausanne, Switzerland

K. F. Turkman

Universidade de Lisboa, Portugal

A. Rua

Banco de Portugal, Portugal

Abstract. In this paper we focus on modelling extremes of nonstationary processes. Given that our approach is linked to the celebrated Box–Jenkins method, we refer to the procedure proposed and applied herein as Box–Jenkins–Pareto. Our procedure is particularly appropriate for settings where the application of a parameter covariate model is non-trivial or when covariates are unavailable. We apply our approach to the weekly number of unemployment insurance claims in the US and explore the connection between threshold exceedances and the US business cycle.

Keywords: Box–Jenkins method; Generalized Pareto distribution; Nonstationary process; Peaks over threshold; Statistics of extremes; Unemployment data; US business cycle

1. Introduction

The statistical analysis of extreme events is of central importance in a wide variety of scenarios. A broad share of this statistical paradigm is founded on Karamata's regularly variation which places the methods at an elegant mathematical support (de Haan and Ferreira, 2006; Resnick, 2007). The domain of application of extreme-value statistics is quite extensive, with modern methods being applied to problems arising in finance (Poon *et al.*, 2003), sports (Einmahl and Magnus, 2008), environmetrics (Eastoe and Tawn, 2009), forestry (Turkman *et al.*, 2010), hydrology (Padoan *et al.*, 2010), among many others. If only block maxima—say, annual maxima—are available, a natural approach to model extremes is to fit a Generalized Extreme Value Distribution (GEVD), but in cases where an entire series is available restricting the analysis to block maxima is a wasteful of the data. Peaks over threshold methods provide an alternative approach by considering as extreme all observations above a large threshold, and this yields a Generalized Pareto Distribution (GPD) as limiting distribution of the threshold exceedances (Balkema and de Haan, 1974; Pickands, 1975). By construction both approaches fail to cover nonstationarity—a feature routinely found in the data.

In this paper we concern ourselves with the peaks over threshold paradigm for nonstationary series. A seminal paper in nonstationary extremes is Davison and Smith (1990), but the quest for alternative modelling approaches is far from complete (Davison and Ramesh, 2000; Chavez-Demoulin and Davison, 2005; Yee and Stephenson, 2007; Padoan and Wand, 2008; Laurini and Pauli, 2009; Eastoe and Tawn, 2009; Northrop and Jonathan, 2011). Most available models introduce covariates in the parameters of the threshold model to overcome

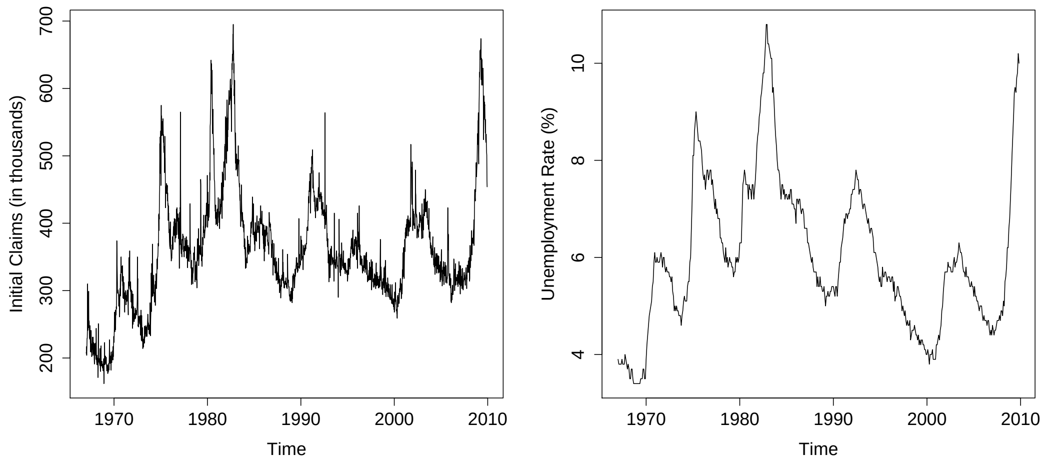
Address for correspondence: Miguel de Carvalho, Institute of Mathematics, Analysis, and Applications; Ecole Polytechnique Fédérale de Lausanne, Station 8, 1015 Lausanne, Switzerland.
E-mail: Miguel.Carvalho@epfl.ch

the lack of stationarity, but even though such approaches are appealing, some hindrances exist. *First*, from a practical standpoint, in some cases the parameter covariate model may be non-trivial or appropriate covariates may be unavailable. *Second*, there is a serious risk of establishing spurious associations linking the parameters and the corresponding covariate models. As it is well known, the similitude of trending mechanisms in the data can easily lead to spurious regressions, a problem which dates back to Yule (Phillips, 1998). *Third*, as recently pointed out by Eastoe and Tawn (2009), parameter covariate approaches based on Davison and Smith (1990) are unable to preserve one of the most important features of the GPD distribution, viz.: threshold stability; this implies that we are unable to guarantee that the distribution of exceedances remains unchanged if a larger threshold is selected.

Our motivation is driven by a well-known economic time series—the weekly number of unemployment insurance claims in the US (henceforth initial claims). This series is often considered as a reference leading indicator for several key macroeconomic variables, being even accredited to be able to forestall recessions (Montgomery *et al.*, 1998; Choi and Varian, 2009, and references therein). As it can be observed in Figure 1, there is a natural proclivity for the number of initial claims to second-guess the unemployment rate. This is clear for example at the end of the observation period where the initial claims peaked before the unemployment rate. Hence, from a practical stance, a peaks over threshold analysis could be of interest for assessing the risk of entering into an unemployment surge, given the most recent information available on initial claims. A stylized fact is that unemployment is known to behave asymmetrically, in the sense that the probability of a decrease in unemployment, given two previous decreases, is greater than the probability of an increase conditional on two preceding increases (Milas and Rothman, 2008). It is also common knowledge that unemployment is supposed to move countercyclically—upward in slowdowns and contractions and downward in speedups and expansions (Rothman, 1998; Caporale and Gil-Alana, 2008). Thus, the definition of a suitable dynamic threshold could be extremely helpful for recognizing the eruption of those surges and ultimately to help counteract them. As the harshness of latest unemployment episode testifies, the understanding of the law of motion of such thresholds is of real value for policy-making decision support. As it will be shown later, threshold exceedances of initial claims can be a valuable tool for the assessment of the US business cycle in real time.

Classical peaks over threshold methods are by no means a good modelling choice for the initial claims, since the series is clearly nonstationary. This can be conjectured from inspection of Figure 1 and is confirmed with the aid of results (not reported here) obtained from stationarity and unit root tests. The initial claims series is representative of the modelling flaws discussed above. *First*, the above-mentioned leading attributes of this series make parameter covariate based strategies, to be discussed in §2, non-trivial, and it is unquestionably difficult to obtain appropriate covariates which are also released on a weekly basis, given that most economic data are available on a monthly or quarterly frequency and moreover released with a large lag. *Second*, but related, some prudence with spurious associations should be taken into account. *Third*, the application of parameter covariate models to the initial claims raises problems in model selection, given the lack of threshold stability.

In this paper we present a modelling strategy for dealing with the above-mentioned difficulties. We propose an approach which can be applied to integrated processes of order α, with α denoting *any* real number. These encompass fractionally integrated processes which have their roots in the works of Granger and Joyeux (1980) and Hosking (1981). If the process is integrated of order α, then although the series of interest may not be stationary, it can be converted into an I(0) series by differencing α-times. Since after differencing α-times stationarity is acquired, classical peaks over threshold models can then be applied to the series which results from such preprocessing step. Binomial series expansions then allow us to naturally build a dynamic threshold for the time series of interest. Given that our approach



(a) (b)

Figure 1. (a) Weekly number of unemployment insurance claims in the US (initial claims). The 2239 weekly observations are seasonally adjusted and range from 7 January 1967, to 28 November 2009; (b) US monthly unemployment rate. The 515 monthly observations are seasonally adjusted and range from January 1967 to November 2009.

is linked to the celebrated Box–Jenkins method (Box *et al.*, 2008) and the GPD model, we designate the modelling strategy proposed herein as the Box–Jenkins–Pareto approach. This is assuredly not the first occasion wherein concepts borrowed from classical time series analysis have proven to be useful in modelling statistics of extremes. For example, in a context distinct from ours, Davis and Mikosch (2009) recently proposed the extremogram—a correlogram for extreme events.

The structure of this paper is as follows. The next section overviews the most frequently applied peaks over threshold approaches for stationary and nonstationary series. In §3 we introduce our modelling strategy and provide guidelines for implementation. In §4 we examine the weekly number of unemployment insurance claims in the US and exploit the connection between threshold exceedances and the US economy contraction and expansion periods. We conclude in §5.

2. A rundown on threshold models

In this section we revisit three threshold models of interest. The first model is for stationary series, while the remainder cover the nonstationary case. In the latter, we restrict our attention to linear models; other approaches can be found elsewhere (Davison and Ramesh, 2000; Chavez-Demoulin and Davison, 2005).

2.1. Models for stationary series

Suppose that the true series of interest $\{Y_t\}$ is stationary with univariate marginal survivor function S_Y. Threshold models consider as extreme the observations $y_1, y_2, \ldots$ which exceed, by a certain amount $y > 0$, a fixed threshold u. These observations are frequently known as exceedances. To draw a distinction between exceedances and non-exceedances, we use $\delta_{u,t} = \mathbb{I}(y_t < u)$ to denote a non-exceedance indicator. The centerpiece of threshold models is based on earlier asymptotic developments (Balkema and de Haan, 1974; Pickands, 1975). Essentially these establish that, for a fixed large threshold u, the conditional survivor of an exceedance by the amount $y > 0$, follows a GPD, with scale parameter $\varphi_u > 0$ and shape parameter γ, i.e.:

$$\Pr\{Y > u + y \mid Y > u\} = \left[1 + \frac{\gamma y}{\varphi_u}\right]_+^{-1/\gamma}, \tag{1}$$

where $a_+ = a \vee 0$. For $\gamma = 0$, equation (1) should be interpreted by taking the limit $\gamma \to 0$, giving an exponential distribution with parameter $1/\varphi_u$, viz.:

$$\Pr\{Y > u + y \mid Y > u\} = \exp\left(-y/\varphi_u\right).$$

After threshold selection has been executed, parameter estimation should be conducted. Some comments are in order. We focus our attention on parameter estimation via likelihood methods. Let $y_1, \ldots, y_n$ denote a random sample from S_Y. Then the likelihood of the model can be written as

$$\mathrm{L}(S_Y, \varphi_u, \gamma) = \prod_{t=1}^{n} (1 - S_Y(u))^{\delta_{u,t}} \left(\frac{S_Y(u)}{\varphi_u} \left[1 + \frac{\gamma y_t}{\varphi_u}\right]_+^{-1/\gamma - 1} \right)^{1-\delta_{u,t}}. \tag{2}$$

One of the most important measures in risk evaluation is the m-observation return level. Roughly speaking, the m-observation return level, here denoted by τ_m, is given by the value which is expected to be exceeded once in every m observations. This can be obtained from the quantiles of the GPD distribution, that is to say

$$\tau_m = u + \frac{\varphi_u}{\gamma}\left[(m S_Y(u))^{\gamma} - 1\right]. \tag{3}$$

Again, we remark that for $\gamma = 0$, this should be interpreted by taking the limit $\gamma \to 0$, so that in such case the m-observation return level is given by

$$\tau_m = u + \varphi_u \log(m S_Y(u)).$$

2.2. The parameter covariate model approach

Suppose now that $\{Y_t\}$ is nonstationary, but that some covariates $\{\mathbf{X}_t\}$ are available. Under this scenario, a natural approach entails considering a linear covariate model for the parameters of the GPD distribution, i.e., to assume that

$$\{Y \mid \mathbf{X} = \mathbf{x}\} \sim \mathrm{GPD}(\varphi_u(\mathbf{x}), \gamma(\mathbf{x})). \tag{4}$$

This modelling approach is due to Davison and Smith (1990) and it is routinely used in applications wherein nonstationarity is verified in the data. In this case the conditional survivor of the exceedances of u is given by

$$\Pr\{Y > u + y \mid Y > u, \mathbf{X} = \mathbf{x}\} = \left[1 + \frac{\gamma(\mathbf{x}) y}{\varphi_u(\mathbf{x})}\right]_+^{-1/\gamma(\mathbf{x})}. \tag{5}$$

Typically, in applications, the parameter covariate models for $\varphi_u(\mathbf{x})$, $\gamma(\mathbf{x})$, and $S_{Y|\mathbf{X}=\mathbf{x}}(u)$, are structured through generalized linear models with suitable link functions. The most natural choices rely upon a log-link, an identity link and a logistic link for the parameter covariate model of the scale, the shape and the rate of exceedances above the threshold u, respectively.

The likelihood of this model follows the same lines as (2). The due adjustments are however necessary

$$\mathrm{L}\left(S_{Y|\mathbf{X}=\mathbf{x}}, \varphi_u(\mathbf{x}), \gamma(\mathbf{x})\right) = \prod_{t=1}^{n}(1-S_{Y|\mathbf{X}=\mathbf{x}}(u))^{\delta_{u,t}} \left(\frac{S_{Y|\mathbf{X}=\mathbf{x}}(u)}{\varphi_u(\mathbf{x})}\left[1+\frac{\gamma(\mathbf{x})y_t}{\varphi_u(\mathbf{x})}\right]_+^{-1/\gamma(\mathbf{x})-1}\right)^{1-\delta_{u,t}}. \tag{6}$$

2.3. The Box–Cox–Pareto method for nonstationary series

A more recent approach for modelling nonstationary extremes is due to Eastoe and Tawn (2009). Given that this modelling strategy is based on the Box–Cox transformation we refer to their approach as the Box–Cox–Pareto method. This transformation is on the basis of the preprocessing approach of the Box–Cox–Pareto wherein the nonstationary process $\{Y_t\}$ is written through the following Box–Cox location-scale model

$$\frac{Y_t^{\lambda(\mathbf{x}_t)}-1}{\lambda(\mathbf{x}_t)} = \beta_1(\mathbf{x}_t) + \beta_2(\mathbf{x}_t)Z_t, \tag{7}$$

where $\{Z_t\}$ denotes an approximately stationary series and $\beta_1(\mathbf{x}_t)$, $\log\beta_2(\mathbf{x}_t)$ and $\lambda(\mathbf{x}_t)$ are linear functions of the covariates. The inference is then conducted through a two step procedure. Firstly, estimate the preprocessing parameters $(\beta_1(\mathbf{x}_t), \beta_2(\mathbf{x}_t), \lambda(\mathbf{x}_t))$. For estimating the preprocessing parameters, one can rely in an approach similar to quasi-likelihood assuming that the data is normally distributed. Secondly, apply the parameter covariate model approach, proposed in §2.2, to the approximately stationary series $\{Z_t\}$.

To give some intuition on how to interpret $\{Z_t\}$, suppose that $\lambda(\mathbf{x}_t) < \epsilon$, for $\epsilon > 0$ sufficiently small. Then, transformation (7) can be recast through the following rough approximation, via a Taylor expansion

$$\log(Y_t) \approx \frac{Y_t^{\lambda(\mathbf{x}_t)}-1}{\lambda(\mathbf{x}_t)} = \beta_1(\mathbf{x}_t) + \beta_2(\mathbf{x}_t)Z_t.$$

Hence $\{Z_t\}$ can be conceptually thought as a sort of residual of a linear model. Given that earlier applications of the Box–Cox transformation were meant to ensure that the classical assumptions of the linear model hold, one can arguably hope that the 'residuals' $\{Z_t\}$ are approximately stationary. The method proposed by Eastoe and Tawn (2009) introduced modelling advantages into the state-of-the-art. Such advantanges are put forward by the authors through an example in environment with emphasis on the maxima of hourly ozone concentrations. This method potentially supports the applications on a more proper theoretical support and greater efficiency can in effect be achieved. However, given that the second step of this method makes direct use of the parameter covariate model, introduced in §2.2, a broad part of the hindrances mentioned above are also potentially shared by the Box–Cox–Pareto approach.

3. The Box–Jenkins–Pareto approach

This section introduces our modelling strategy. Given that our approach is based on the celebrated Box–Jenkins time series method (Box *et al.*, 2008), we refer to the procedure proposed herein as the Box–Jenkins–Pareto approach, in opposition to the Box–Cox–Pareto method of Eastoe and Tawn (2009).

3.1. The Box–Jenkins–Pareto method for nonstationary series

Data preparation techniques can be very convenient for subsequent data analysis. One of the most common data preparation methods is given by differencing, i.e., to consider the differences between consecutive observations. The classical Box–Jenkins method is representative of the advantages that differencing can bring into the analysis.

Suppose that the nonstationary series $\{Y_t\}$ can be converted into a stationary series by differencing once, i.e.,

$$(1 - \mathbb{L})Y_t = Z_t, \tag{8}$$

for some stationary series $\{Z_t\}$ with survivor S_Z. Here and below $\mathbb{L}$ is the lag operator and $(1 - \mathbb{L}) \equiv \Delta$ is the difference operator. A series which satisfies (8) is said to be integrated of order 1 and will be denoted by I(1). In some cases there is no need to difference, given that we already have a stationary process. A particular case where such occurrence takes place is when we have an I(0) process; formally, a covariance stationary processes with autocovariance function γ_u is an I(0) process if

$$\sum_{u=-\infty}^{\infty} |\gamma_u| < \infty.$$

More generally, the series $\{Y_t\}$ is said to be integrated of order α (to be denoted by I(α)), for $\alpha \in \mathbb{R}$, if

$$\Delta^{\alpha} Y_t = Z_t,$$

for some I(0) series $\{Z_t\}$ with survivor S_Z. A comprehensive discussion on these series can be found in Robinson and Marinucci (2001). This general class encompasses fractionally integrated processes which have their roots in the seminal works of Granger and Joyeux (1980) and Hosking (1981). We emphasize that the memory parameter α is allowed to be *any* real number. This parameter condenses useful information regarding the stationarity of the sequence: if $\alpha \in [0; 0.5[$ then the series is stationary and mean-reverting; for $\alpha \in [0.5; 1[$ the series is no longer stationary although it is still mean-reverting; finally, if $\alpha \geq 1$ the series is neither stationary nor mean-reverting.

The following functional central limit theorem establishes the link between integrated series and fractional Brownian motion—a continuous stochastic process with known applications in extreme value modelling (Mikosch *et al.*, 2002; Buchamann and Klüppelberg, 2005). Here we use $\Rightarrow$, $W(x)$, and $\lceil . \rceil$ for denoting weak convergence, Brownian motion, and the ceiling function, respectively.

THEOREM 1. *(Sowell, 1990) Let $\{Y_t\}$ be an integrated series of order $-1/2 < \alpha < 1/2$. Suppose that $Z_t \equiv \Delta^{\alpha} Y_t$ are independent and identically distributed with $E\{Z_t\} = 0$ and $E\{|Z_t|^r\} < \infty$, for $r \geq \{4 \vee (-8\alpha/(1 + 2\alpha))\}$. Then*

$$\sigma_n^{-1} \sum_{\tau=1}^{\lceil nt \rceil} Y_\tau \Rightarrow W_\alpha(t),$$

where $\sigma_n \equiv \mathrm{var}\{\sum_{\tau=1}^{n} Y_\tau\}$ and $W_\alpha(t)$ is fractional Brownian motion, i.e. the stochastic integral

$$W_\alpha(t) = \frac{1}{\Gamma(\alpha + 1)} \int_0^t (t - x)\mathrm{d}W(x).$$

From the extreme value modelling standpoint, the question of interest is the following: suppose that the series of interest $\{Y_t\}$ is nonstationary, but it is I(α) for some real number α; is it still possible to build directly a threshold model for $\{Y_t\}$?

To give an answer to this question, assume by now that the differencing parameter α is a positive integer; later we let α be any real number. Since $\{Y_t\}$ is $\mathrm{I}(\alpha)$, the exceedances of Z in the amount $y > 0$, above a fixed high threshold u, can be modelled through a GPD, i.e.

$$\Pr\{Z > u + y \,|Z > u\} = \left[1 + \frac{\gamma y}{\varphi_u}\right]^{-1/\gamma}, \tag{9}$$

for every $y > 0$. Hence, the likelihood of the model is essentially the same as given in (2). Observe further that for every period t we have

$$\begin{aligned} \Pr\{Z_t > u + y \,|Z_t > u\} &= \Pr\{\Delta^\alpha Y_t > u + y \,|\Delta^\alpha Y_t > u\} \\ &= \Pr\left\{\left(\sum_{i=0}^{\alpha} \tbinom{\alpha}{i}(-\mathbb{L})^i\right) Y_t > u + y \,\middle|\, \left(\sum_{i=0}^{\alpha} \tbinom{\alpha}{i}(-\mathbb{L})^i\right) Y_t > u\right\} \\ &= \Pr\{Y_t > \widetilde{u}_t + y \,|Y_t > \widetilde{u}_t\}, \end{aligned} \tag{10}$$

where $\widetilde{u}_t$ defines the dynamic threshold given by

$$\widetilde{u}_t = u + \sum_{i=1}^{\alpha} \tbinom{\alpha}{i} \left(Y_{t-i}\mathbb{I}(i \text{ odd}) - Y_{t-i}\mathbb{I}(i \text{ even})\right), \quad t \geq \alpha + 1, \tag{11}$$

with $\binom{\alpha}{i} = \Gamma(\alpha+1)/(\Gamma(i+1)\Gamma(\alpha+i-1))$ denoting the binomial coefficient and $\mathbb{I}(.)$ the indicator function. Here and below, $\Gamma(.)$ is the gamma function with the customary conventions $\Gamma(0) = \infty$ and $\Gamma(0)/\Gamma(0) = 1$. Observe that the dynamic threshold, given in (11), is composed by a building block (u) and a remainder time-varying part which makes use of the previous α values of the series. From a practical stance this implies that it will only be possible to make the dynamic threshold start at the $(\alpha+1)$-observation. Nevertheless, this is not as critical as it might appear since in general α assumes values with order of magnitude below 1. This is strengthened by the fact that (9) follows from large sample results, so that $(\alpha+1)/T$ is negligible overall.

In the simplest case where the series is difference-stationary with $\alpha = 1$, it holds that $\widetilde{u}_t = u + Y_{t-1}$, for $t \geq 2$. The simple relationship established in (10) suggests a natural way for constructing a dynamic threshold for $\mathrm{I}(1)$ series, namely: first, obtain u from the first differences of the series of interest; second, add u to the lagged version of the series.

If α is any real number the more general series expansion should be taken into account

$$\Delta^\alpha = \sum_{i=0}^{\infty} \frac{\langle\alpha\rangle_i}{\Gamma(i+1)}(-\mathbb{L})^i, \tag{12}$$

where

$$\langle\alpha\rangle_i = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha-i+1)}\mathbb{I}(\alpha \neq 0) = \alpha(\alpha-1)\cdots(\alpha-i+1),$$

is the Pochhammer's symbol for the falling factorial. We recall that for any positive integer α, (12) is tantamount to the classical binomial expansion. Thus, if α is *any* real number, similarly to (10) we still have

$$\Pr\{Z_t > u + y \,|Z_t > u\} = \Pr\{Y_t > \widetilde{u}_t + y \,|Y_t > \widetilde{u}_t\},$$

but now the dynamic threshold $\widetilde{u}_t$ is more broadly defined as

$$\widetilde{u}_t = u + \sum_{i=1}^{t-1} \frac{\langle\alpha\rangle_i}{\Gamma(i+1)} \left(Y_{t-i}\mathbb{I}(i \text{ odd}) - Y_{t-i}\mathbb{I}(i \text{ even})\right), \quad t \geq 2 + \mathbb{I}(\alpha \in \mathbb{N}_0)(\alpha - 1). \tag{13}$$

Some comments are in order. If α is positive and integer we recover the threshold given in (11). The more general version of the dynamic threshold now obtained is similar to the one obtained in (11) being also composed by a building block and a remainder time-varying part. However, the threshold $\widetilde{u}_t$ now obtained makes *potential* use of all previous $(t-1)$ observations. The pathological case wherein $\alpha = 0$ is quite interesting. If $\alpha = 0$ then $\{Y_t\}$ is stationary so that we would expect the classical threshold model for stationary series to hold. From the inspection of (13) we can confirm that this is the case, since $\widetilde{u}_t = u$, for $t \geq 1$.

For completeness we discuss below how the dynamic threshold proposed above can be used for return level modelling; we also discuss the case for integrated series with a polynomial trend.

3.2. Path return level with dynamic threshold

Consider again the case wherein the series $\{Y_t\}$ is $\mathrm{I}(\alpha)$. In this case, the exceedances of Z, in the amount $y > 0$, above a fixed high threshold u, can be modelled through a $\mathrm{GPD}(\varphi_u, \gamma)$. Using the dynamic threshold given above, the following return level can be obtained

$$\widetilde{\tau}_m(t) = \widetilde{u}_t + \frac{\varphi_u}{\gamma}\left[(mS_{Y_t}(\widetilde{u}_t))^{\gamma} - 1\right].$$

We refer to this time-varying level as m-observation path return level. Just as the return level, presented in (3), yields the fixed level τ_m which is expected to be exceeded once in every m observations, the path return level defines a route level expected to be exceeded once in every m observations. Again, the case for $\gamma = 0$ should be interpreted by taking the limit $\gamma \to 0$, so that the m-observation return level is given as

$$\tau_m(t) = \widetilde{u}_t + \varphi_u \log(mS_{Y_t}(\widetilde{u}_t)).$$

3.3. The Box–Jenkins–Pareto approach for series with a polynomial trend

The foregoing subsections were devoted to the threshold modelling for nonstationary series. In applications we are also frequently confronted with the need to model a nonstationary series with a deterministic time trend. Formally, a process $\{Y_t\}$ is said to be integrated of order α, for $\alpha \in \mathbb{R}$, with a polynomial time trend of degree $\beta \in \mathbb{R}$, if

$$\Delta^{\alpha} Y_t - \lambda t^{\beta} = Z_t,$$

for some stationary $\mathrm{I}(0)$ series $\{Z_t\}$ with survivor S_Z. These processes will be denoted by $\mathrm{IT}(\alpha, \beta)$, where the "T" is used to denote trend. Of course the chief interest relies in the unexplored case of $\beta \neq 0$, since the remainder case was examined in §3.1. Observe that for any real number α it holds that

$$\Pr\{Z_t > u + y \,|Z_t > u\} = \Pr\{Y_t > \widetilde{u}_t + y \,|Y_t > \widetilde{u}_t\},$$

with the dynamic threshold $\widetilde{u}_t$ now defined as

$$\widetilde{u}_t = u + \lambda t^{\beta} + \sum_{i=1}^{t-1} \frac{\langle \alpha \rangle_i}{\Gamma(i+1)} \left(Y_{t-i}\mathbb{I}(i \text{ odd}) - Y_{t-i}\mathbb{I}(i \text{ even})\right), \tag{14}$$

for $t \geq 2 + \mathbb{I}(\alpha \in \mathbb{N}_0)(\alpha - 1)$. This roughly means the following: the polynomial time trend enters additively into the dynamic threshold. As expected the dynamic threshold is now composed by two time-varying components: one due to the trend; the remainder due to the memory of the series. The case wherein $\alpha = 0$ is again illustrative. In such case, the process is trend stationary and we obtain the threshold $\widetilde{u}_t = u + \lambda t^{\beta}$, for $t \geq 1$. This simple observation provides guidance for a simple alternative, to the parameter covariate model introduced above, for trend stationary processes. Thus, (14) suggests estimating the trend and computing the dynamic threshold in lieu of considering $\varphi_u(t) = \exp(\lambda t^{\beta})$ and/or $\gamma = \lambda t^{\beta}$ and estimating the model via (6). For testing if the process is trend-stationary one can consider the stationarity test of Kwiatkowski *et al.* (1992). This procedure is very popular and accessible in several statistical packages. A more complete portrait of this literature, including more avant-garde and powerful tests, can be found, for example, in Cavaliere and Taylor (2008).

4. The initial claims and the US business cycle

In this section we model initial claims using the proposed Box–Jenkins–Pareto approach. We intend to examine what connection the resulting threshold exceedances may have with the US economy contraction and expansion periods dated by the Business Cycle Dating Committee of the National Bureau of Economic Research (NBER). Given the hardship which results from economic contractions we are mainly interested in recessive periods and thus focus the analysis on right tail corresponding threshold exceedances of the initial claims. The period under analysis for the weekly number of unemployment insurance claims in the US (initial claims) ranges from 7 January 1967, to 28 November 2009. The 2239 observations from this seasonally adjusted series were gathered from the United States Department of Labor—Employment & Training Administration and can be downloaded from

`http://www.ows.doleta.gov`

One could think of using the exceedances resulting from the dynamic threshold scheme introduced above, as an indicator of whether an economy is entering or crossing a recession period. Several reasons anticipate however the difficulties with such inquiry, and one of such complications lies in the data itself. As pointed by the Business Cycle Dating Committee of the NBER (see Frequently Asked Questions NBER, 2008), there is a marked week-to-week noise in the initial claims. Moreover, it should be stressed that it is not our goal to design a dating procedure, or an ideal 'alarm' mechanism (Antunes *et al.*, 2003), but merely to provide economic interpretation to the dynamic threshold proposed above, in terms of a business cycle analysis.

To apply the Box–Jenkins–Pareto approach we first need to be apprised of the order of differentiation α to be used in the analysis. There has long been an interest in fractionally integrated models and in the estimation of the fractional differencing parameter; for a recent review see Gil-Alana and Hualde (2009). Among the estimation procedures available, parametric and semi-parametric approaches are the most employed in practice. In the former, a full parametric model is specified, and so there is the risk of misspecification which can yield biased estimates of the long memory parameter (Fox and Taqqu, 1986; Sowell, 1992). A semi-parametric approach to estimate the fractional differencing parameter is here pursued. We use the well-known GPH estimator (Geweke and Porter-Hudak, 1983); this method is based on the linear regression of the log-periodogram on a deterministic regressor and seeks to estimate α without specifying a finite parameter model for the α-th difference of the time series. A practical problem in the implementation of the GPH estimator is the selection of the number of frequencies to be used in the regression—a choice that entails a bias–variance trade-off. Geweke and Porter-Hudak suggested choosing the number of frequencies by $\mathsf{m} = T^{1/2}$, and this is the rule most commonly used in practice. Applying this rule to the initial claims series yields the estimate 0.96, with a standard error of 0.11. Since the choice regarding the number of frequencies to be used in the regression is not clear cut (Hurvich and Deo, 1999), we also present the GPH estimate for a wide range of frequencies; following Perron and Qu (2010), we consider m ranging from 10 up to $T^{3/4}$. From Figure 2,

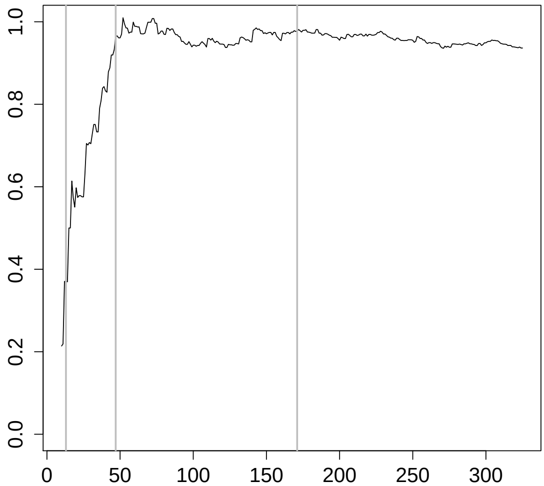


Figure 2. The GPH estimates of α with m ranging from 10 up to $T^{3/4}$ for the weekly number of unemployment insurance claims in the US; the gray vertical lines correspond to $T^{1/3}$, $T^{1/2}$, and $T^{2/3}$.

except for the case where only extremely low frequencies are used—where the GPH estimate reveals to be unstable—, the GPH estimate is always close to one.

For the sake of exposition, in the sequel we consider the memory parameter α to be equal to 1 and hence the first differences of the initial claims are examined below. The construction of the dynamic threshold, defined according to either (11) or (13), was as made as follows. Firstly, the time-varying part is simply given by the one week lagged initial claims. Secondly, the fixed part of the dynamic threshold (u) was obtained from the first differences of the initial claims. As usual, the selection of threshold is a debatable step. If a too low threshold is selected then the asymptotic rationale of the model is not justified and bias is generated. On the other hand, if a too high threshold is chosen few exceedances are available so that higher variance is obtained. Detailed recommendations on threshold selection can be found, for instance, in Bermudez *et al.* (2001). As guidance, we use the mean residual life plot and plotted parameter estimates of the peaks over threshold model, of the first differences of initial claims, at a variety of thresholds. Estimation results, not reported here, confirmed the graphical suggestion that the tail index estimate is quite stable if small perturbations are induced in the fixed threshold of $u^{+} = 48$. This implies that a weekly increase above 48.000 initial claims is considered as a right tail exceedance. The analysis was supplemented by probability plots, quantile plots, and density plots. In Figure 3 we depict the dynamic threshold obtained.

To give some interpretation to the sequence of exceedances generated by the dynamic threshold presented in Figure 3, we introduce in the subsequent figures shaded areas representing the US economic activity contractions dated by the Business Cycle Dating Committee of the NBER. Seven peak (P) to trough (T) movements occured from January 1967 to November 2009. Thus during the period under analysis seven contractions were acknowledged by the NBER Business Cycle Dating Committee, viz.: i) December 1969 – November

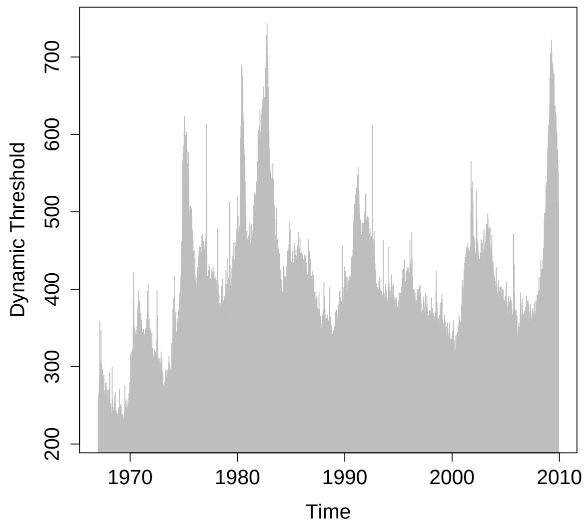
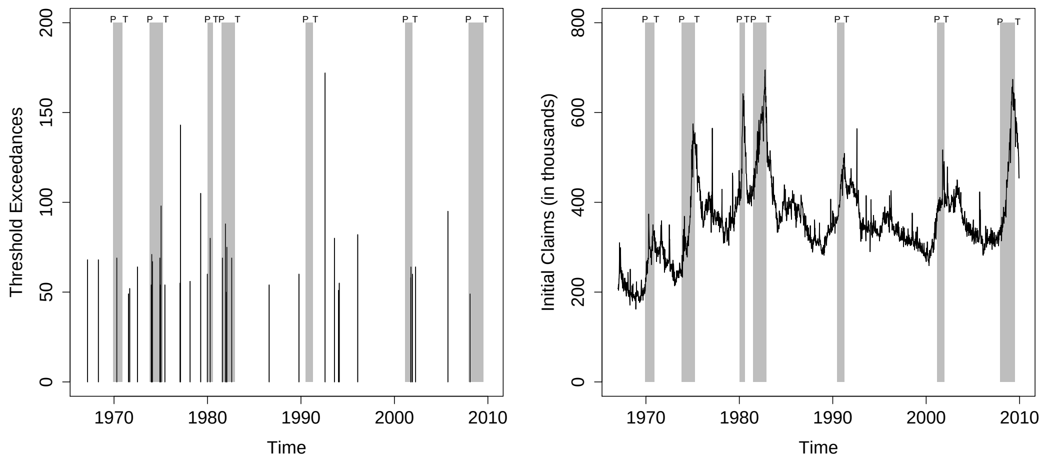


Figure 3. The dynamic threshold for the weekly number of unemployment insurance claims in the US.

1970; *ii*) November 1973 – March 1975; *iii*) January 1980 – July 1980; *iv*) July 1981 – November 1982; *v*) July 1990 – March 1991; *vi*) March 2001 – November 2001; *vii*) December 2007 – June 2009. Observe further that there is some lag in the identification of peaks by NBER. For example, the economic activity peak of December 2007 was only determined in December 2008 (NBER, 2008).

Figure 4 represents the threshold exceedances and the original series. In the sequel we assess the information content that the threshold exceedances of the initial claims possess for tracking contraction periods. From the inspection of Figure 4 we can ascertain that among the 2239 weekly observations such mechanism would have been activated only 37 times. It is somehow promising that such naive mechanism is consistent with several contraction episodes and particularly with the eruption of the latest economic activity peak determined by NBER. This is reinforced by the fact that in only 17.6% of the period under analysis contractions ocurred, so that it is substantially more difficult to spot recessive periods simply by chance. However, the analysis of Figure 4 also reveals that several exceedances occured during expansions. As argued above, it is recognized by the NBER (See Frequently Asked Questions NBER, 2008) that there is a noticeable week-to-week noise in the initial claims series which complicates its analysis. As it can be observed in Figure 4 the larger exceedances in (a) correspond to isolated spikes in (b) so that they are most probably due to week-to-week noise. In general, these spikes are immediately reverted in the following week. Therefore, one possible way to sieve plausible exceedances from noisy ones is to inspect which exceedances were followed in the next week by a left tail exceedance. This involves performing an analogous threshold analysis as performed above for the right tail of the first differences of the initial claims. The same approach now yields, for the left tail, a fixed threshold $u^- = -38$; we refer to the exceedances which result from the the latter analysis as left tail exceedances, and to the exceedances depicted in Figure 4 as right tail exceedances, or simply as exceedances whenever there is no possibility of confusion.



(a) (b)

Figure 4. (a) Threshold exceedances; (b) Weekly number of unemployment insurance claims in the US (inital claims). Shaded areas represent the US economic activity contractions dated by the Business Cycle Dating Committee of the NBER.

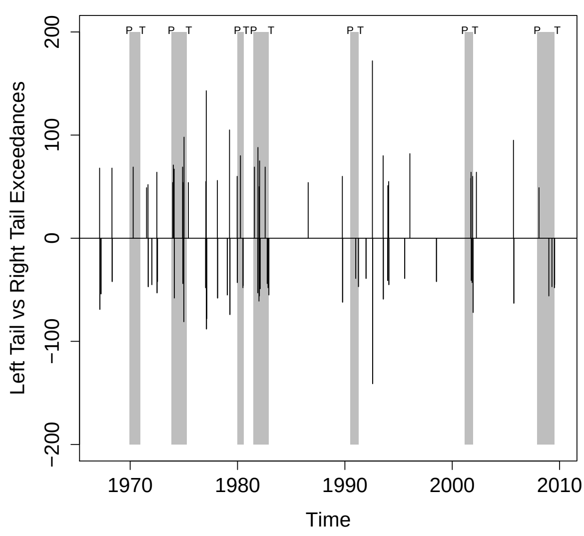


Figure 5. Mirror plot.

Figure 5 depicts the right and left tail exceedances—a representation which we denominate as the mirror plot. The analogy here is that the lines correspoding to noisy exceedances should be immediately followed by left tail exceedances creating the visual effect of a mirror image. The mirror plot can then be thought as an exploratory tool for examining which right tail exceedances are followed by left tail exceedances in the next week. Observe that the filtering procedure suggested by the mirror plot is congruous with the earlier discussed dynamic asymmetry, according to which unemployment exhibits abrupt increases in opposition to longer and gradual declines (Milas and Rothman, 2008). In particular this implies that right tail exceedances are not expected to be immediately followed by left tail exceedances. The right tail exceedances which are not followed by a left tail exceedance in the upcoming week are represented in Figure 6 and are here denominated as mirror filtered exceedances.

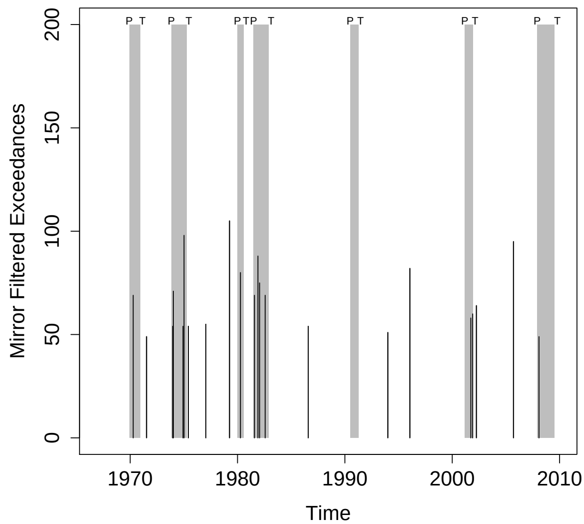


Figure 6. Mirror filtered exceedances.

The number of mirror filtered exceedances is 22, from which 13 occured during contraction periods and 9 during expansion periods. It is important to note that in only circa 4/22 of the period under analysis contractions ocurred. This implies that it is much more difficult to randomly spot contraction periods, so that a proportion of 13/22 is considerably satisfying. It should also be pointed out that two of the mirror filtered exceedances which occurred out of contraction periods are only a few weeks apart from the trough, and among the remainder only five are clearly distant from any contraction period.

5. Summary and conclusions

Classical approaches for modelling extremes are unable to cope with nonstationarity. Most popular modelling approaches introduce covariates in the parameters of the threshold model to overcome the lack of stationarity in the series of interest. In diverse contexts of interest the covariate model may be however nontrivial, so that impelling the introduction of covariates may lead to spurious associations which can seriously prejudice the analysis. Further, in some cases appropriated covariates may be simply unavailable or only at one's disposal at undesirable frequencies or horizons. The lack of threshold stability of some of these methods is also an important modelling issue with obvious practical implications.

This paper suggests an alternative approach for modelling nonstationary extremes which circumvents these difficulties. The modelling strategy proposed herein can be applied to integrated processes of order α, with α denoting *any* real number. Given that our procedure is linked to both the celebrated Box–Jenkins time series method and the GPD model, we designate the modelling strategy proposed in this paper as the Box–Jenkins–Pareto approach. The application enclosed herein examines the weekly number of unemployment insurance claims in the US and exploits the connection between the threshold exceedances and the US business cycle. During the course of the analysis we resorted to what we call the mirror plot as a means to deal with the week-to-week noise which is well-known to be present in the initial claims. This exploratory tool suggests a natural filtering approach which has shown to be effective in this empirical application. Our results suggest that the mirror filtered exceedances resulting from the Box–Jenkins–Pareto analysis are strongly related with the US business cycle.

Acknowledgments

We are grateful to Anthony Davison, Valérie Chavez-Demoulin, and Vanda Inácio for helpful discussions and recommendations that led to a significant improvement of this paper. We also thank comments from participants at a 2010 SAMSI workshop chaired by Richard Smith. Part of this paper was written while M. B. de C. was visiting the Economics Research Department of *Banco de Portugal*. Financial support from the *Centro de Matemática e Aplicações*, *Universidade Nova de Lisboa* is acknowledged.

References

Antunes, M., Turkman, M. A. and Turkman, K. F. (2003) A Bayesian approach to event prediction. *J. Time Ser. Anal.*, **24**, 631–646.

Balkema, A. A. and de Haan, L. (1974) Residual life time at great age. *Ann. Probab.*, **2**, 792–804.

Bermudez, P., Turkman, M. A. and Turkman, K. F. (2001) A predictive approach to tail probability estimation. *Extremes*, **4**, 295–314.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. (2008) *Time Series Analysis: Forecasting and Control.* 4th edn. New Jersey: Wiley.

Buchamann, B. and Klüppelberg, C. (2005) Maxima of stochastic processes driven by fractional Brownian motion. *Adv. Appl. Probab.*, **37**, 743–764.

Caporale, G. M. and Gil-Alana, L. (2008) Modelling the US, UK and Japanese unemployment rates: fractional integration and structural breaks. *Computnl Statist. Data Anal.*, **52**, 4998–5013.

Cavaliere, G. and Taylor, A. M. R. (2008) Testing for a change in persistence in the presence of non-stationary volatility. *J. Econometrics*, **147**, 84–98.

Chavez-Demoulin, V. and Davison, A. C. (2005) Generalized additive modelling of sample extremes. *J. R. Statist. Soc. C*, **54**, 207–222.

Choi, H. and Varian, H. (2009) Predicting initial claims for unemployment benefits. *Technical Report—Google.*

Davis, R. A. and Mikosch, T. (2009) The extremogram: a correlogram for extreme events. *Bernoulli*, **15**, 977–1009.

Davison, A. C. and Ramesh, N. I. (2000) Local likelihood smoothing of sample extremes. *J. R. Statist. Soc. B*, **62**,191–208.

Davison, A. C. and Smith, R. (1990) Models for exceedances over high thresholds (with discussion). *J. R. Statist. Soc. B*, **52**, 393–442.

de Haan, L. and Ferreira, A. (2006) *Extreme Value Theory: an Introduction.* New York: Springer.

Eastoe, E. F. and Tawn, J. A. (2009) Modelling non-stationary extremes with application to surface level ozone. *J. R. Statist. Soc. C*, **58**, 25–45.

Einmahl, J. H. J. and Magnus, J. R. (2008) Records in athletics through extreme-value theory. *J. Am. Statist. Ass.*, **103**, 1382–1391.

Fox, R. and Taqqu, M. S. (1986) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.*, **14**, 517–532.

Geweke, J. and Porter-Hudak, S. (1983) The estimation and application of long-memory time series models. *J. Time. Ser. Anal.*, **4**, 221–238.

Gil-Alana, L. and Hualde, J. (2009) Fractional integration and cointegration. An overview and an empirical application. *Palgrave Handbook in Econometrics*, **2**, (Eds. C. Mills and K. Patterson) pp. 434–472, MacMillan Publishers.

Granger, C. W. J. and Joyeux, R. (1980) An introduction to long memory time series and fractional differencing. *J. Time Ser. Anal.*, **1**, 5–39.

Hosking, J. R. M. (1981) Fractional differencing. *Biometrika*, **68**, 165–176.

Hurvich, C. M. and Deo, R. S. (1999) Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long memory time series. *J. Time Ser. Anal.*, **20**, 331–341.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992) How sure are we that economic time series have a unit root? *J. Econometrics*, **54**, 159–178.

Laurini, F. and Pauli, F. (2009) Smoothing sample extremes: the mixed model approach. *Computnl Statist. Data Anal.*, **53**, 3842–3854.

Mikosch, T., Resnick, S., Rootzén, H. and Stegeman, A. W. (2002) Is network traffic approximated by stable Lévy motion or fractional Brownian motion? *Ann. Appl. Probab.*, **12**, 23–68.

Milas, C. and Rothman, P. (2008) Out-of-sample forecasting of unemployment rates with pooled STVECM forecasts. *Int. J. Forecasting*, **24**, 101–121.

Montgomery, A. L., Zarnowitz, V., Tsay, R. S. and Tiao, G. C. (1998) Forecasting the US unemployment rate. *J. Am. Statist. Ass.*, **93**, 478–493.

NBER (2008) Determination of the December peak in economic activity. Announcement of the NBER's Business Cycle Dating Committee (`http://www.nber.org/cycles/`).

Northrop, P. J. and Jonathan, P. (2011) Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*, in press.

Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010) Likelihood-based inference for max-stable processes. *J. Am. Statist. Ass.*, **105**, 263–277.

Padoan, S. A. and M. P. Wand (2008) Mixed model-based additive models for sample extremes. *Statist. Probab. Lett.*, **78**, 2850–2858.

Perron, P. and Qu, Z. (2010) Long-memory and level shifts in the volatility of stock market return indices. *J. Bus. Econ. Statist.*, **28**, 275–290.

Phillips, P. C. B. (1998) New tools for understanding spurious regressions. *Econometrica*, **66**, 1299–1325.

Pickands, J. (1975) Statistical inference using extreme order statistics. *Ann. Statist.*, **3**, 119–131.

Poon, S.-H., Rockinger, M. and Tawn, J. A. (2003) Modelling extreme-value dependence in international stock markets. *Statist. Sin.*, **13**, 929–953.

Resnick, S. (2007) *Heavy-Tail Phenomena—Probabilistic and Statistical Modeling.* New York: Springer.

Robinson, P. M. and Marinucci, D. (2001) Narrow-band analysis of nonstationary processes. *Ann. Statist.*, **29**, 947–986.

Rothman, P. (1998) Forecasting asymmetric unemployment rates. *Rev. Econ. Statist.*, **80**, 164–168.

Sowell, F. (1990) The fractional unit root distribution. *Econometrica*, **58**, 495–505.

Sowell, F. (1992) Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *J. Econometrics*, **53**, 165–188.

Turkman, K. F., Turkman, M. A. and Pereira, J. M. (2010) Asymptotic models and inference for extremes of spatio-temporal data extremes. *Extremes*, **13**, 375–397.

Yee, T. W. and A. G. Stephenson (2007) Vector generalized linear and additive extreme value models. *Extremes*, **10**, 1–19.