# Nonstationary Extremes and the US Business Cycle

Miguel de Carvalho† , K. Feridun Turkman†, António Rua†

**Abstract**

Considerable attention has been devoted to the statistical analysis of extreme events. Classical peaks over threshold methods are a popular modelling strategy for extreme value statistics of stationary data. For nonstationary series a variant of the peaks over threshold analysis is routinely applied using covariates as a means to overcome the lack of stationarity in the series of interest. In this paper we concern ourselves with extremes of possibly nonstationary processes. Given that our approach is, in some way, 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 parameter covariate model is non-trivial or when well qualified covariates are simply unavailable. We apply the Box-Jenkins-Pareto approach to the weekly number of unemployment insurance claims in the US and exploit the connection between threshold exceedances and the US business cycle.

KEY WORDS: Box-Jenkins method; Generalized Pareto distribution; Nonstationary process; Peaks over threshold; Statistics of extremes; Unemployment data; US business cycle.

† Miguel de Carvalho (Email: *mb.carvalho@fct.unl.pt*) is full researcher at CMA – *Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa*, Portugal. K. Feridun Turkman (Email: *kfturkman@fc.ul.pt*) is full professor at the *Faculdade de Ciências, Universidade de Lisboa*, Portugal and a research member of CEAUL. António Rua is head of the Conjuntural Unit of the Forecasting and Conjuntural Division of the Economics Research Department of the Portuguese Central Bank (*Banco de Portugal*). This paper was written during a visit of Miguel de Carvalho to the Economics Research Department of *Banco de Portugal*. The usual disclaimer applies.

## 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 regular variation, which places the methods at an elegant theoretical support (de Haan and Ferreira, 2006; Resnick, 2007). From the applied stance, the domain of application of extreme value statistics is quite extensive. In effect, modern methods are illustrated by means of problems of interest which arise in the contexts of hidrology (Padoan et al., 2010), portfolio management (Poon et al., 2003, 2004), clean steel production (Bortot et al., 2007), wildfire analysis (Turkman et al., 2009), records in athletics (Einmahl and Magnus, 2008), among others. At the crux of extreme value modelling relies the extremal types theorem, a classical result first established by Fisher and Tippet (1928) and Gnedenko (1943). Roughly speaking, this cornerstone result proclaims that the properly standardized maxima of a sequence of independent and identically distributed random variables, converges to one element of the trinity composed by the Fréchet, Gumbel and Weibull distributions. The trinity in unity is then provided by the Generalized Extreme Value Distribution (GEVD), which brings together all the above mentioned distributions. If only block maxima (*say*, annual maxima) are available, a natural modelling approach is to fit a GEVD to the data. However, in cases wherein an entire series is available restricting the analysis to block maxima is a wasteful of the data. Under such circumstances, the twin approach of peaks over threshold methods is potentially more effective given that it collects further information from the tail of the distribution of interest, considering as extreme all the exceedances above a fixed large threshold. This dual procedure yields the Generalized Pareto Distribution (GPD) as the limiting distribution of the threshold exceedances (Balkema and de Haan, 1974; Pickands, 1975). Unfortunately, by construction both of these approaches fail to cover nonstationarity – a feature routinely claimed by the data.

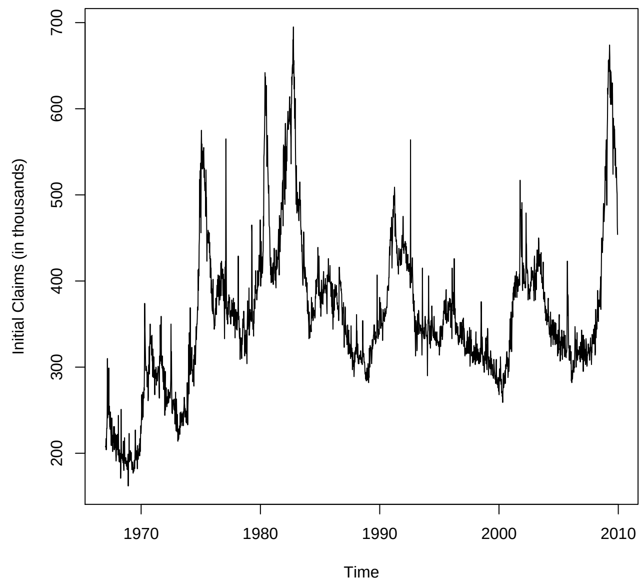
This paper is mainly concerned with the peaks over threshold paradigm for possibly nonstationary processes. In what regards nonstationary extremes a seminal reference is Davison and Smith (1990), but as some recent papers put forward, the quest for alternative modelling approaches is far from being completed (see for example Davison and Ramesh, 2000; Chavez-Demoulin and Davison, 2005; Eastoe and Tawn, 2009). Most of the available modelling approaches for nonstationary extremes make use of the introduction of covariates in the parameters of the threshold model as a means to overcome the lack of stationarity in the series of interest. Even though such approaches are quite appealing some hindrances should be pointed out. *First*, from the practical stance, in some circumstances the parameter covariate model may be non-trivial or well qualified covariates may be simply unavailable. *Second*, there is a serious risk of establishing spurious associations linking the parameters and the corresponding covariate models. In effect, 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 (see Phillips, 1998, and references therein). One of the most comical of such spurious connections led to the identification of an association between the number of ordained ministers and the rate of alcoholism in Britain (Phillips, 1998). *Third*, as recently pinpointed out by Eastoe and Tawn (2009) parameter covariate approaches based on Davison and Smith (1990) are unable to preserve one of the most notorious features of the GPD distribution, viz.: threshold stability. In particular this means that, in such models, we are unable to guarantee that the form of the distribution of the threshold exceedances remains unchanged if a larger threshold is selected.

Our driving example for illustrating the modelling flaws mentioned above makes use of a well known economic time series – the weekly number of unemployment insurance claims in the US (henceforth initial claims). This series is oftentimes considered as a reference leading indicator for several key macroeconomic variables of interest 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 from the examination of Figure 1, there is a natural proclivity for the number of initial claims to second-guess the unemployment rate. For example, this is clear at the end of the observation period wherein the initial claims peaked before the unemployment rate. Hence, from the pratical stance, a peaks over threshold analysis could be of considerable 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, 2001; Caporale and Gil-Alana, 2008). Hence, 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. In effect, as it will be shown later, threshold exceedances of initial claims can be a valuable tool for the assessment of the US business cycle.

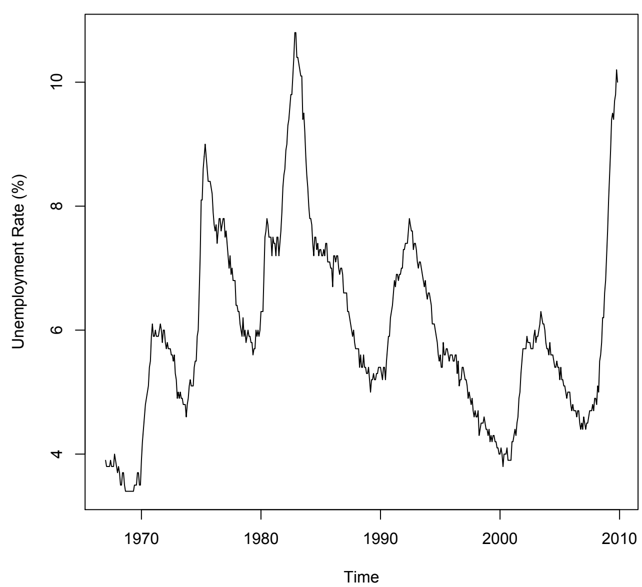
Classical peaks over threshold methods are by no means a good modelling choice for the initial claims. In fact, the lack of stationary in the series is clear. This can be definitely conjectured from the inspection of Figure 1 and easily confirmed with the aid of results (not reported here) obtained from stationarity and unit root tests. The long-range dependence of this type of data is self-evident, as corroborated by Figure 2.

The initial claims series is also 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 Section 2, highly non-trivial. *Second*, but related, some prudency with spurious associations should also be taken into account. In addition, 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 noteworthy lag. *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 propose a modelling stratagem for dealing with the set of difficulties mentioned above. More specifically, this paper proposes an approach which can be applied to integrated processes of order $\alpha$, with $\alpha$ denoting *any* real number. Hence, these also encompass fractionally integrated processes which have their roots in the works of Granger and Joyeux (1980) and Hosking (1981). We note that if the process is integrated of order $\alpha$, then although the series of interest may not be stationary, it can be converted into a stationary series by differencing $\alpha$-times. Since after differencing $\alpha$-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 original series of interest. Given that our approach is linked to both the celebrated Box-Jenkins time series method (Box et al., 2008) and the GPD model, we designate the modelling stratagem 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,




(a)

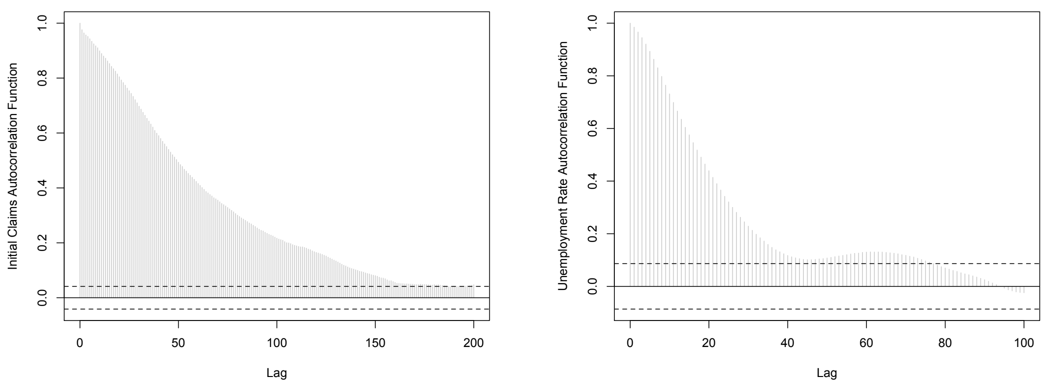



(b)

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.

in a context distinct from ours, it was recently proposed by Davis and Mikosch (2009) the extremogram – a correlogram for extreme events.

The layout of this paper is as follows. The next section overviews the most frequently applied peaks over threshold approaches for stationary and nonstationary series. Section 3 introduces our modelling stratagem and provides concise guidelines for implementation. In Section 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. The paper closes in Section 5.




(a) (b)

Figure 2: (a) Sample autocorrelation function for the initial claims; (b) Sample autocorrelation function for the unemployment data.

## 2. A RUNDOWN ON THRESHOLD MODELS

This section revisits three threshold models of interest. The first model presented here is for stationary series, while the remainder cover the nonstationary case. In what concerns 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

In the sequel 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. In order 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 $\gamma$, i.e.:

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

where $a_+ = a \vee 0$. It should be pointed out that for $\gamma = 0$, the relation given in (1) should be interpreted by taking the limit $\gamma \to 0$, so that under such circumstances it is obtained an exponential distribution with parameter $1/\varphi_u$, viz.:

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

After threshold selection has been executed, parameter estimation should be conducted. Some comments are in order. We will 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

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

For the sake of curiosity, observe that this likelihood resembles the case wherein random censoring is present (Einmahl et al., 2008). Yet, in such case, a non-censoring indicator is used in lieu of a non-exceedance indicator .

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 $\tau_m$, is given by the value which is expected to be exceeded once in every $m$ observations. This can be easily obtained from the quantiles of the GPD distribution, that is to say

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

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

$$\tau_m = u + \varphi_u \log(mS_Y(u)).$$

## 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

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

This modelling approach is due to Davison and Smith (1990) and it is routinely used in applications wherein stationarity is not claimed by the data. In this case the conditional survivor of the exceedances of $u$ is given by

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

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

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

As discussed above, although this approach is quite appealing some drawbacks should pinpointed. In fact, from the practical stance, in a variety of circumstances the parameter covariate model may be non-trivial or well qualified covariates may be simply unavailable. There is also a serious risk of establishment of spurious associations linking the parameter and the corresponding covariate model. In effect the similitude of trending mechanisms in the data can lead to spurious regressions – a problem which dates back to Yule (see Phillips, 1998, and references therein). Moreover, as recently pointed out by Eastoe and Tawn (2009) the parameter covariate approach is unable to preserve one of the most notorious features of the GPD distribution, viz.: threshold stability. In particular this implies that, in this model, we are not able to assure that the form of the distribution of the threshold exceedances remains unaltered if a larger threshold is chosen.

## 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 (Box and

Cox, 1964) 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, *mirabile dictu*, 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 the preceding subsection, to the approximately stationary series $\{Z_t\}$.

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

$$\log(Y_t) \approx \frac{Y_t^{\lambda(\mathbf{x}_t)} - 1}{\lambda(\mathbf{x}_t)} = \beta_1(\mathbf{x}_t) + \beta_2(\mathbf{x}_t) Z_t. \tag{8}$$

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 some nice modelling advantages into the state-of-the-art. It potentially supports the applications on a more proper theoretical support and greater efficiency can in effect be achieved. Nevertheless, given that the second step of this method makes direct use of the parameter covariate model, introduced in the previous subsection, 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 stratagem. Given that our approach is, in some way, 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 Possibly 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.,

$$(1 - \mathbb{L})Y_t = Z_t, \tag{9}$$

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 verifies (9) is said to be integrated of order 1 and will be denoted by $\mathrm{I}(1)$. In the same spirit, the notation $\mathrm{I}(0)$ is used for stationary series. Given that some series require additional differencing before reaching stationarity a broader definition is necessary. More generally, the series $\{Y_t\}$ is said to be integrated of order $\alpha$ (to be denoted by $\mathrm{I}(\alpha)$), for $\alpha \in \mathbb{R}$, if

$$\Delta^{\alpha} Y_t = Z_t, \tag{10}$$

for some stationary 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 underscore that the memory parameter $\alpha$ 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 make use of the notations $\Rightarrow$ and $W(x)$ for denoting weak convergence and Brownian motion, respectively. In addition, $\lceil . \rceil$ denotes the ceiling function.

**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 $\sigma_n \equiv Var\{\sum_{\tau=1}^{n} Y_\tau\}$ 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 theory stance, one question of interest is the following: Suppose that the series of interest $\{Y_t\}$ is possibly nonstationary, but that it is $\mathrm{I}(\alpha)$ for some real number $\alpha$. Is it still possible to build directly a threshold model for $\{Y_t\}$?

In order to give an answer to such question, assume by now that the differencing parameter $\alpha$ is a positive integer; later we let $\alpha$ 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.

$$\Pr\{Z > u + y \,|Z > u\} = \left[1 + \frac{\gamma y}{\varphi_u}\right]^{-1/\gamma}, \tag{11}$$

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

$$\begin{aligned}
\Pr\{Z_t > u + y \,|Z_t > u\} &= \Pr\{\Delta^\alpha Y_t > u + y \,|\Delta^\alpha Y_t > u\} \\
&= \Pr\left\{\left(\sum_{i=0}^{\alpha} \tbinom{\alpha}{i} (-\mathbb{L})^i\right) Y_t > u + y \,\middle|\, \left(\sum_{i=0}^{\alpha} \tbinom{\alpha}{i} (-\mathbb{L})^i\right) Y_t > u\right\} \\
&= \Pr\{Y_t > \widetilde{u}_t + y \,|Y_t > \widetilde{u}_t\}, \qquad (12)
\end{aligned}$$

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

$$\widetilde{u}_t = u + \sum_{i=1}^{\alpha} \tbinom{\alpha}{i} \left(Y_{t-i}\mathbb{I}(i \text{ odd}) - Y_{t-i}\mathbb{I}(i \text{ even})\right), \quad \text{for } t \geq \alpha + 1, \tag{13}$$

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 (13), is composed by a building block ($u$) and a remainder time-varying part which makes use of the previous $\alpha$ values of the series. From the 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 *prima facie* appear since in general $\alpha$ assumes values with order of magnitude below 1. This is strengthened by the fact that, as mentioned above, (11) follows from large sample results, so that $(\alpha + 1)/n$ is negligible in the 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 (12) suggests a natural way for constructing a dynamic threshold of I(1) series, namely: first, obtain $u$ from the first differences of the series of interest; secondly, sum $u$ to the lagged version of the series.

For the case wherein $\alpha$ is any real number the more general series expansion should be taken into account

$$\Delta^{\alpha} = \sum_{i=0}^{\infty} \frac{\langle \alpha \rangle_i}{\Gamma(i+1)} (-\mathbb{L})^i, \tag{14}$$

where

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

is the Pochhammer's symbol for the falling factorial. We bring to mind that for any positive integer $\alpha$, (14) is tantamount to the classical binomial expansion. Thus, if $\alpha$ is *any* real number, similarly to (12) 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

$$\widetilde{u}_t = u + \sum_{i=1}^{t-1} \frac{\langle \alpha \rangle_i}{\Gamma(i+1)} \left(Y_{t-i}\mathbb{I}(i \text{ odd}) - Y_{t-i}\mathbb{I}(i \text{ even})\right), \quad \text{for } t \geq 2 + \mathbb{I}(\alpha \in \mathbb{N}_0)(\alpha - 1). \tag{15}$$

Some comments are in order. As a first preliminary observation, note that if $\alpha$ is positive and integer we recover the threshold given in (13). The more general version of the dynamic threshold now obtained is similar to the one obtained in (13) being also composed by a building block and a remainder time-varying part. Nevertheless, the threshold $\widetilde{u}_t$ now obtained makes *potential* use of all the 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 (15) we can in fact confirm that this is the case, since we obtain $\widetilde{u}_t = u$, for $t \geq 1$.

For the sake of completeness we discuss below how the dynamic threshold proposed above can be used for return level modelling. In addition, for the sake of generality 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 $\mathrm{I}(\alpha)$. Given that we are considering that $\{Y_t\}$ 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$(\varphi_u, \gamma)$. Using the dynamic threshold given above, the following return level can be easily obtained

$$\widetilde{\tau}_m(t) = \widetilde{u}_t + \frac{\varphi_u}{\gamma}\left[(mS_{Y_t}(\widetilde{u}_t))^{\gamma} - 1\right]. \tag{16}$$

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 $\tau_m$ which is expected to be beat 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 under those circumstances the $m$-observation return level is given as

$$\tau_m(t) = \widetilde{u}_t + \varphi_u \log(mS_{Y_t}(\widetilde{u}_t)). \tag{17}$$

## 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 one is 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 $\alpha$, for $\alpha \in \mathbb{R}$, with a polynomial time trend of degree $\beta \in \mathbb{R}$, if

$$\Delta^{\alpha} Y_t - \lambda t^{\beta} = Z_t, \tag{18}$$

for some stationary 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 a preceding subsection. Observe that for any real number $\alpha$ 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

$$\widetilde{u}_t = u + \lambda t^{\beta} + \sum_{i=1}^{t-1} \frac{\langle \alpha \rangle_i}{\Gamma(i+1)} \left(Y_{t-i}\mathbb{I}(i \text{ odd}) - Y_{t-i}\mathbb{I}(i \text{ even})\right), \tag{19}$$

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. Note that 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 quite interesting. In fact, under such conditions the process is trend stationary and we now 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, (19) 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 easily 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 may the resulting threshold exceedances have with the US economy contraction and expansion periods dated by the Business Cycle Dating Committee of the National Bureau of Economic Research (hereinafter NBER). Given the hardship which results from economic contractions we are chiefly 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 easily downloaded from the web site: `http://www.ows.doleta.gov`. The R code developed during the implementation of this application is available from the authors upon request.

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. Notwithstanding, several reasons anticipate the difficulties with such an inquiry. For example one of such complications lies in the data itself. In effect, 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 point here to design an ideal alarm mechanism, but merely to provide some insight on how to ponder over the dynamic threshold proposed above, in the current context. For optimal alarm systems see, for example, Antunes et al. (2003) and references therein.

In order to apply the Box-Jenkins-Pareto approach we first need to be apprised of the order of differentiation $\alpha$ to be used in the analysis. Here we make use of the well known GPH estimator proposed by Geweke et al. (1983). This yields $\widehat{\alpha}_{GPH} = 0.9643$ with corresponding standard error 0.1069. For the sake of exposition, in the sequel we consider the memory parameter $\alpha$ 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 (13) or (15), 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. Quoting Davis and Mikosch (2009) "the choice of threshold is always a thorny issue in extreme value theory," entailing a balance between bias and variance. 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 made use of 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 $u^{+} = 48$. The corresponding tail index estimate is $\widehat{\gamma} = 0.0717$, with standard error 0.1649. The analysis was supplemented by probability plots, quantile plots, and density plots. In Figure 3 we depict the dynamic threshold obtained.

In order to give some flavor to the sequence of exceedances generated by the dynamic threshold presented in Figure 3, we introduce in the subsequent figures shaded areas rep-

[†] The case wherein the exact value of the estimate $\widehat{\alpha}_{GPH}$ was used is available from the authors upon request. Not surprisingly, the results are similar in the overall.

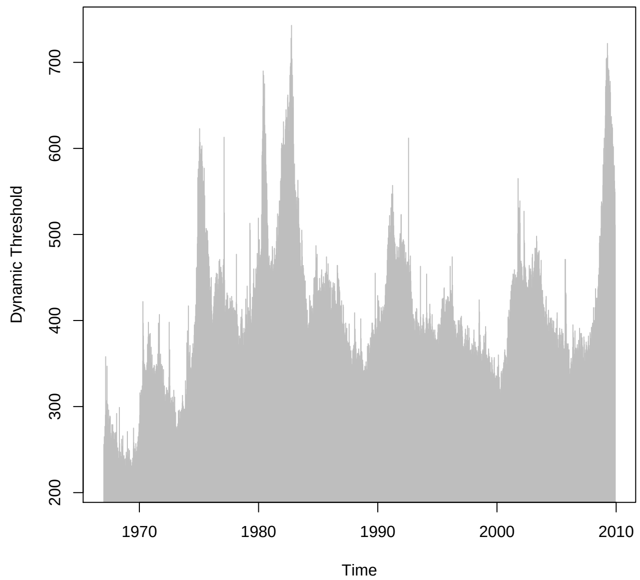
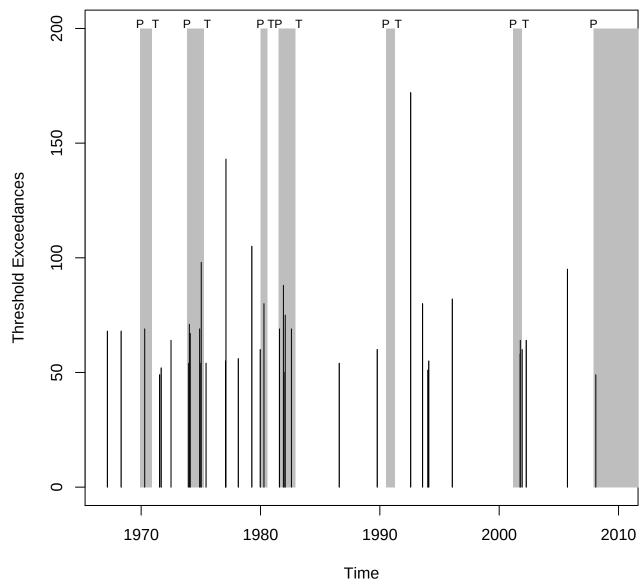


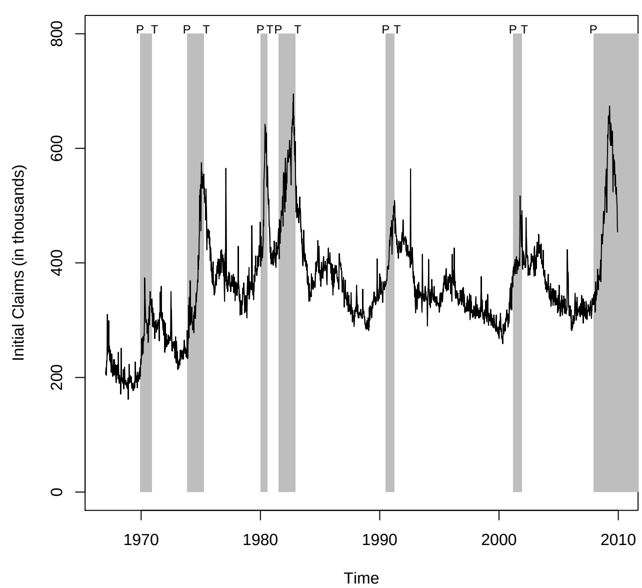

Figure 3: The dynamic threshold for the weekly number of unemployment insurance claims in the US.

resenting 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 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. Note that in the latest contraction episode the trough was not yet determined by NBER. 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




(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.

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 enthusiastic 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 22.6% of the period under analysis contractions ocurred, so that it is substantially more difficult to spot recessive periods simply by chance. Nevertheless, 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 difficults its analysis. In effect, 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 the overall, 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 $u^{-} = -38$. For simplicity, 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.

Figure 5 depicts the right and left tail exceedances – a representation which we denominate below as 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

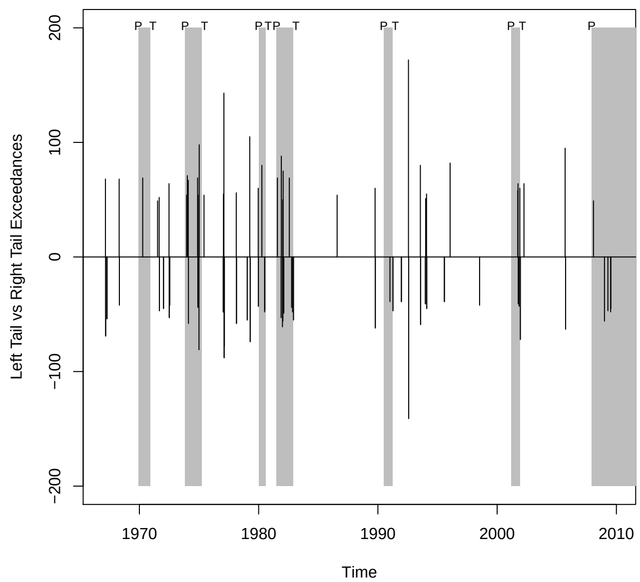



Figure 5: Mirror plot.

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 certainly 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 straightaway 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.

The number of mirror filtered exceedances is 22, from which 13 occured during contraction periods and 9 during expansion periods. Some comments are in order. We bring to mind that it is important to note in only circa 5/22 of the period under analysis contractions ocurred. Roughly speaking, this implies that it is much more difficult to randomly spot

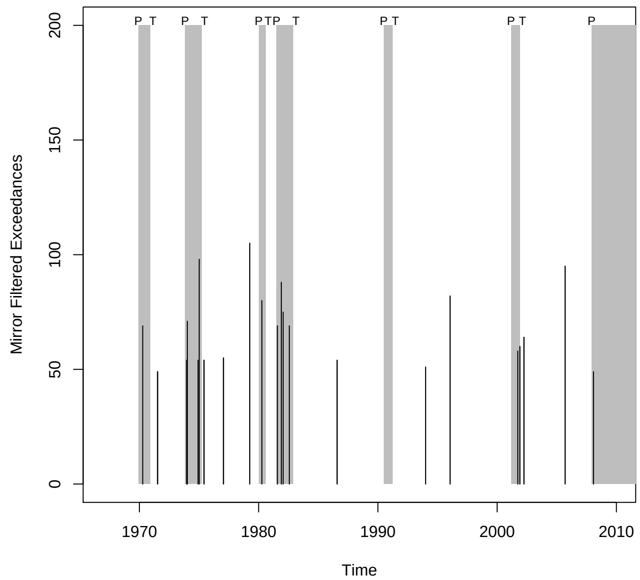



Figure 6: Mirror filtered exceedances.

contraction periods, so that a proportion of 13/22 is considerably satisfying. It should also be pointed out that two of mirror filtered exceedances which occurred out of contraction periods are only a few weeks apart from the trough, and among the remainder only 5 are clearly distant from any contraction period.

The obtained results are encouraging, evidencing the information content that the intial claims exceedances possess for tracking contraction periods in the US business cycle.

## 5. SUMMARY AND CONCLUSIONS

The statistical modelling of extreme events is a subject of noteworthy significance both at the theoretical and practical levels. Unfortunately, the classical approaches are unable cope with nonstationarity – a feature routinely claimed by the data. For dealing with this problem most popular modelling approaches make use of the introduction of covariates in the parameters of the threshold model in order to overcome the lack of stationarity in the series of interest.

As argued above, in diverse contexts of interest, several modelling hindrances may arise with the application of these methods. In effect, there is a wide variety of scenarios wherein the covariate model may not be trivial so that impelling the introduction of covariates may lead to spurious associations which can seriously prejudice the analysis. Further, in some cases well suited covariates may be simply unavailable or only at one's disposal at undesirable frequencies or horizons. The lack of threshold stability of these methods is also an important modelling issue with obvious practical implications.

This paper suggests an alternative approach for modelling possibly nonstationary extremes which circumvents the difficulties discussed above. The modelling stratagem proposed herein can be applied to integrated processes of order $\alpha$, with $\alpha$ 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 stratagem 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 very effective in this empirical application. Our results put forward 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 Vanda Inácio, Paulo M.M. Rodrigues and Pedro Portugal for helpful comments and suggestions.

## REFERENCES


Antunes, M., Turkman, M. A., and Turkman, K. F. (2003), "A Bayesian Approach to Event Prediction," *Journal of Time Series Analysis*, 24, 631–646.

Balkema, A. A., and de Haan, L. (1974), "Residual Life Time at Great Age," *The Annals of Probability*, 2, 792–804.

Bermudez, P., Turkman, M. A., and Turkman, K. F. (2001), "A Predictive Approach to Tail Probability Estimation," *Extremes*, 4, 295–314.

Bortot, P., Coles, S. G., and Sisson, S. A. (2007), "Inference for Steorological Extremes," *Journal of the American Statistical Association*, 102, 84–92.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. (2008), *Time Series Analysis: Forecasting and Control*, 4th Edition, Wiley Series in Statistics.

Box, G. E. P., and Cox, D. R. (1964), "An Analysis of Transformations," *Journal of the Royal Statistical Society, Ser. B*, 26, 211–252.

Buchamann, B., and Klüppelberg C. (2005), "Maxima of Stochastic Processes Driven by Fractional Brownian Motion," *Advances in Applied Probability*, 37, 743–764.

Caporale, G. M., and Gil-Alana, L. (2008), "Modelling the US, UK and Japanese Unemployment Rates: Fractional Integration and Structural Breaks," *Computational Statistics and Data Analysis*, 52, 4998–5013.

Cavaliere, G., and Taylor, A. M. R. (2008), "Testing for a Change in Persistance in the Presence of Non-Stationary Volatility," *Journal of Econometrics*, 147, 84–98.

Chavez-Demoulin, V., and Davison, A. C. (2005), "Generalized Additive Modelling of Sample Extremes," *Applied Statistics*, 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 Smith, R. (1990), "Models for Exceedances Over High Thresholds," *Journal of the Royal Statistical Society, Ser. B*, 52, 393–442.

Davison, A. C., and Ramesh, N. I. (2000), "Local Likelihood Smoothing of Sample Extremes" *Journal of the Royal Statistical Society, Ser. B*, 62, 191–208.

de Haan, L., and Ferreira, A. (2006), *Extreme Value Theory an Introduction.* Springer Series in Operations Research and Financial Engineering.

Einmahl, J. H. J., Fils-Villetard, A., and Guillou, A. (2008), "Statistics of Extremes Under Random Censoring," *Bernoulli*, 14, 207–227.

Einmahl, J. H. J., and Magnus, J. R. (2008), "Records in Athletics Through Extreme-Value Theory," *Journal of the American Statistical Association*, 103, 1382–1391.

Eastoe, E. F., and Tawn, J. A. (2009), "Modelling Non-Stationary Extremes with Application to Surface Level Ozone," *Journal of the Royal Statistical Society, Ser. C*, 58, 25–45.

Fisher, R. A., and Tippet, L. H. (1928), "On the Estimation of the Frequency of Distributions of the Largest or Smallest Member of a Sample," *Proceedings of the Cambridge Philosophical Society*, 24, 180–190.

Geweke, J., and Porter-Hudak, S. (1983), "The Estimation and Application of Long-Memory Time Series Models," *Journal of Time Series Analysis*, 4, 221–238.

Gnedenko, B. V. (1943), "Sur la Distribution Limite du Terme Maximum d'une Série Aléatoire," *Annals of Mathematics*, 44, 423–453.

Granger, C. W. J., and Joyeux, R. (1980), "An Introduction to Long Memory Time Series and Fractional Differencing," *Journal of Time Series Analysis*, 1, 5–39.

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

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992), "How Sure are we that Economic Time Series Have a Unit Root?," *Journal of Econometrics*, 54, 159–178.

Milas, C., and Rothman, P. (2008), "Out-of-Sample Forecasting of Unemployment Rates with Pooled STVECM Forecasts," *International Journal of Forecasting*, 24, 101–121.

Montgomery, A. L., Zarnowitz, V., Tsay, R. S., and Tiao, G. C. (1998), "Forecasting the U.S. unemployment rate," *Journal of the American Statistical Association*, 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/`).

Padaon, S. A., Ribatet, M., and Sisson, S. A. (2010), "Likelihood-Based Inference for Max-Stable Processes," *Journal of the American Statistical Association*, (to appear).

Phillips, P. C. B. (1998), "New Tools for Understanding Spurious Regressions," *Econometrica*, 66, 1299–1325.

Phillips, P. C. B., and Shimotsu, K. (2004), "Local Whittle Estimation in Nonstationary and Unit Root Cases," *The Annals of Statistics*, 32, 656–692.

Poon, S.-H., Rockinger, M., and Tawn, J. A. (2003), "Modelling Extreme-Value Dependence in International Stock Markets," *Statistica 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," *The Review of Financial Studies*, 17, 581–610.

Pickands, J. (1975), "Statistical Inference Using Extreme Order Statistics," *The Annals of Statistics*, 3, 119–131.

Resnick,S., Mikosch, T., Stegeman, A. W., and Rootzén, H. (2002), "Is Network Traffic Approximated by Stable Lévy Motion or Fractional Brownian Motion?," *The Annals of Applied Probability*, 12, 23–68.

Resnick, S. (2007), *Heavy-Tail Phenomena – Probabilistic and Statistical Modeling.* Springer Series in Operations Research and Financial Engineering.

Robinson, P. M., and Marinucci, D. (2001), "Narrow-Band Analysis of Nonstationary Processes," *The Annals of Statistics*, 29, 947–986.

Rothman, P. (2001), "Forecasting Asymmetric Unemployment Rates," *Review of Economics and Statistics*, 80, 164–168.

Sowell, F. (1990), "The Fractional Unit Root Distribution," *Econometrica*, 58, 495–505.

Turkman, K. F., Turkman, M. A., and Pereira, J. M. (2009), "Asymptotic Models and Inference for Extremes of Spatio-Temporal Data Extremes," *Extremes*, (to appear) DOI 10.1007/s10687-009-0092-8.