Article

When the U.S. Stock Market Becomes Extreme?

Sofiane Aboura

Department of Finance, DRM-Finance, Université de Paris-Dauphine, Place du Maréchal de Lattre de Tassigny, 75775 Paris CEDEX 16, France; E-Mail: sofiane.aboura@dauphine.fr;
Tel.: +33-1-4405-4565.

Abstract: Over the last three decades, the world economy has been facing stock market crashes, currency crisis, the dot-com and real estate bubble burst, credit crunch and banking panics. As a response, extreme value theory (EVT) provides a set of ready-made approaches to risk management analysis. However, EVT is usually applied to standardized returns to offer more reliable results, but remains difficult to interpret in the real world. This paper proposes a quantile regression to transform standardized returns into theoretical raw returns making them economically interpretable. An empirical test is carried out on the S&P500 stock index from 1950 to 2013. The main results indicate that the U.S stock market becomes extreme from a price variation of $\pm 1.5\%$ and the largest one-day decline of the 2007–2008 period is likely, on average, to be exceeded one every 27 years.

Keywords: extreme value theory; volatility; risk management

JEL classifications: C4, G13, G32

1. Introduction

The tail behavior of the financial series has been largely examined by many studies[1] with various applications. The key attraction of extreme value theory (EVT) is that it offers a set of ready-made approaches to risk management analysis. A general discussion of the application of EVT to risk management is proposed by [12,15,16]. However, if the oldest studies made use of raw returns

---

[1] Some of the main studies of univariate EVT in finance include [1–14].

(e.g., [9,10] *etc.*) in their EVT analysis, since then, the literature has applied EVT on independent and identically distributed data such as standardized returns (e.g., [13]). Therefore, many results stemming from EVT applications remain quiet difficult to interpret economically because of the use of these standardized returns and conclusion can be vague. Indeed, the practice is to extract standardized returns by filtering the raw returns with a GARCH family model; then run some standard tests to check for the approximate independent and identically distributed nature of the standardized returns and finally apply EVT on these data. However, these standardized returns have no economic meaning in the real world. In clear, they cannot be understood in terms of monetary units. One naive possibility would have been to apply an OLS procedure to study the relationship between a response variable (raw returns) and an explanatory variable (standardized returns), in order to numerically recompute all the raw returns corresponding to the standardized returns derived from the EVT analysis and forecasting. Recall that the least-squares regression describes how the mean of the response variable changes with the vector of covariates. But standard regression analysis has several limits such as not being robust to extreme observations whereas EVT is based on observations located in the tails. For that reason, this article proposes an empirical methodology to convert the results into monetary units such as stock index returns. This issue seems not to have been addressed previously, perhaps because the typical EVT paper is generally devoted to the sole statistical perspective used to describe the extreme behavior of financial returns, but not to explain the economic meaning. The method is a simple linear transformation of the standardized returns into raw returns using quantile regressions [17]. Quantile regression generalizes the standard regression model to conditional quantiles of the response variable. Hence, it can be be viewed as a natural extension of standard least squares estimation of conditional mean models to the estimation of a series of models for conditional quantile functions [18]. To our best knowledge, no such statistical test has been implemented so far to address this problem of economic interpretation of EVT results. This research is relevant because it gives a new approach for interpreting the EVT results. The main contribution consists in proposing an empirical methodology based on quantile regression for transforming EVT results expressed in the form of standardized returns into raw returns. An application is provided on the U.S. market from 1950 to 2013. The article is organized as follows. Section 2 opens with a brief review of the extreme value theory applications in finance. Section 3 presents an analysis of the data while Section 4 analyses the empirical results. Section 5 summarizes the main findings and concludes.

2. Methodology

2.1. Tail Distribution

A theorem from [19,20] shows that when the threshold u is sufficiently high, the distribution function F_u of the excess beyond this threshold can be approximated by the Generalized Pareto Distribution (GPD):

$$F_u(x) \approx G_{\xi,\beta}(x). \tag{1}$$

This limit distribution has a general form given by:

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi}, & \text{for } \xi \neq 0 \\ 1 - \exp(-x/\beta), & \text{for } \xi = 0 \end{cases} \tag{2}$$

where $\beta \geq 0$ and where $x \geq 0$ when $\xi \geq 0$ and where $0 \leq x \leq -\beta/\xi$ when $\xi < 0$. β is a scaling parameter and ξ is the tail index. The tail index is an indication of the tail heaviness, the larger ξ, the heavier the tail. This distribution encompasses other type of distributions. If $\xi > 0$ ($\xi < 0$) then it is a reformulated version of the ordinary Pareto distribution and if $\xi = 0$, it corresponds to the exponential distribution.

2.2. Threshold Selection

Threshold selection is subject to a trade off between finding a high threshold where the tail estimate has a low bias with a high variance or finding a low threshold where the tail estimate has a high bias with a low variance. Low (high) bias refers to a (non-reliable) reliable estimate for the tail index.

There are two approaches for threshold selection. The first approach is visual inspection and the second one is automatic detection. The first approach of visual selection [16,21,22] denotes a plausible threshold choice based on the results of the two given plot methods, namely, the mean residual life plot and the threshold plot. The mean residual life plot is the first visual method:

$$\left[u, \frac{1}{n_u} \sum_{i=1}^{u} (x_i - u) \right] \tag{3}$$

with n_u representing the number of exceedances over the threshold u. The threshold detection begins with a slope is positive. The threshold plot is the second visual inspection method that consists in fitting the GPD over a range of thresholds allowing observation of both the parameter estimate stability and variance:

$$\left[\widehat{\xi}, G_{\xi,\beta}(x) = 1 - (1 + \xi x/\beta)^{-1/\xi} \right]. \tag{4}$$

The second approach, of optimal threshold, corresponds to the application of an automated method which aims at minimizing the Asymptotic Mean Squared Error (AMSE)[2]. For the optimal threshold selection method, let's consider an ordered sample of size n_u, $X_{n_u} \leq ... \leq X_1$ with X_{n_u} being the n_u^{th} upper order statistic. The [24] estimator is defined by:

$$\left[n_u, \widehat{\xi}_{n_u}^H = \frac{1}{n_u} \sum_{i=1}^{n_u} (log X_{n-i+1} - log X_{n-n_u}) \right]. \tag{5}$$

It has been popular for the optimal threshold to be estimated such that the bias and variance of the estimated Hill tail index vanish at the same rate where the mean squared error is asymptotically minimized. Usually, AMSE is obtained throughout a sub-sample bootstrap procedure. This paper

---

[2] See [15], Section 4.7 ii. More general discussion on AMSE includes [23].

follows [15] who develop a criterion for which the AMSE of the Hill estimator is minimal for the optimal number of observations in the tail. In clear, this approach computes the optimal sample fraction needed to apply the tail index estimator.

2.3. Three Extreme Risk Management Measures

Every day risk management practices require evaluating the potential risk of loss. Three risk management measures can be used to adress this evaluation problem. Most of the recent literature [12,25–29] confirms the superiority of the in and out-the-sample performance of the risk management models when combining a heavy-tailed GARCH filter with an extreme value theory-based approach. For that reason, these three risk management measures are based on returns standardized by GARCH models.

First, the value-at-risk (VaR) is a common risk measure on any portfolio of financial assets with a given probability and time horizon. For example, a one-day 1% VaR of 1 million euros means that there is a 0.01 probability that the portfolio will fall in value by more than 1 million euros over the next day period. However, there is no need to fully identify the probability distribution because only the extreme quantiles are of interest. Therefore, in our setting, the value-at-risk is computed by inverting the tail estimation formula based on the loss distribution:

$$VaR_q = u + \frac{\beta}{\xi} \left[\left(\frac{n_u}{n}(1-q) \right)^{-\xi} - 1 \right]. \tag{6}$$

Nevertheless, VaR models have been criticized for their partial inadequacy. First, VaR can be misleading during volatile periods [30,31]. Second, the VaR of a portfolio can be higher than the sum of the respective VaRs of each individual asset in the portfolio [32]. Third, VaR disregards any loss beyond the VaR level.

Second, a complementary measure known as the expected shortfall (ES) is usually used, for example, in margin requirements. ES particularly focuses on the loss beyond the VaR level. It accounts for the size of tail losses since it evaluates the expected loss size given that VaR is exceeded:

$$ES_q = VaR_q + E\left(X - VaR_q | X > VaR_q\right) \tag{7}$$

Third, the return level (RL) [33] highlights the effect of extrapolation, which is useful for forecasting, even if scarcity produces large variance estimates. Let's consider x_m as the return level that is exceeded on average once every m observations. Let ζ_u be the probability of exceeding the threshold u. Return level is expressed in annual scale so that the N-year return level is the level expected to be exceeded once every N years. From a risk management perspective, RL is a waiting time period before observing a minimum daily loss, while VaR is the minimum daily loss that can occur with a given probability over a certain horizon. Both measures are threfore closed and complementary. n_{250} is the average number of trading days per year with $m = N \times n_{250}$. N-year return level is an average waiting time period. For example, the N-year return level is the level expected to be exceeded once every N years. It comes for the N-year return level:

$$x_m = u + \frac{\beta}{\xi} \left[(N \times n_{250} \zeta_u)^{\xi} - 1 \right] \tag{8}$$

3. Data Analysis

3.1. Data Description

Let's define the market log-returns as $\{R_t\}_{t=1,...,T}$ with $T = 15,950$ raw daily log-returns computed from closing price of the U.S. S&P 500 stock index. The time period begins on 3 January 1950 and ends on 28 May 2013. When applying EVT analysis, parameters estimates along with standard errors can be quite unreliable. This underlies the importance of having long time-series with an adaptive econometric filter analysis to obtain identically distributed innovations. Indeed, the presence of autocorrelation and hetereskedasticity in log-returns series requires implementing filters for avoiding any type of bias in the analysis since the data need to be independent and identically distributed before applying extreme value theory. Therefore, this study makes exclusively use of filtered data.

3.2. Data Filtering Process

We examine all the possible specifications within five lags. We test 25 specifications of ARMA(p,q) + GARCH(1,1) models with $p = 1, ..., 5$ and $q = 1, ..., 5$. We select the more parsimonious model. Four criteria are used for comparison: the Schwarz criterion, the autocorrelogram of residuals and squared residuals and the ARCH effect test. We take care of the trade off between parsimony and maximizing criteria. We find that the ARMA(1,1) + GARCH(1,1) model produces the best fit. We then test an alternative model, ARMA(1,1) + TGARCH(1,1), that allows for leverage effects by considering the contribution of the negative residuals in the ARCH effect. Finally, the ARMA(1,1) + TGARCH(1,1) specification has the best fit considering the four criteria. The ARMA (1,1) component takes into account the mean reversion nature of the returns, while the TGARCH specification allows for leverage effects by considering the contribution of the negative residuals in the ARCH effect. Table 1 displays the descriptive statistics showing that except the AR(1) term, all the parameters are statistically significant. The AR(1) term was not removed because withdrawing ϕ_1 did not clearly improved all the criteria. The ARMA(1,1) + TGARCH (1,1) specification is given by:

$$R_t = \mu + \phi_1 R_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t \tag{9}$$

where the innovations ϵ_t being functions of Z_t and σ_t:

$$\epsilon_t = Z_t \sigma_t. \tag{10}$$

The standardized returns Z_t are independent and identically distributed with $Z_t \sim F_Z(0,1)$. The purpose of the time-varying σ_t is to capture as much of the conditional variance in the residual ϵ_t in order to leave Z_t approximately independent and identically distributed:

$$\sigma_t^2 = \omega + \alpha \left(Z_{t-1} \sigma_{t-1} \right)^2 + \gamma \left(Z_{t-1} \sigma_{t-1} \right)^2 I_{Z_{t-1}\sigma_{t-1}<0} + \beta \sigma_{t-1}^2. \tag{11}$$

Table 1. Descriptive statistics. The Table presents the descriptive statistics of the S&P 500 stock index standardized daily log-returns (Z) from 3 January 1950 to 31 May 2013. Sample: 15,950 observations.

	Z		Z
Mean	−0.0032	Q5 (residual) (p-value)	2.9403 (0.401)
Median	0.0202	Q5 (squared residual) (p-value)	7.2983 * (0.063)
Maximum	7.0191	Q10 (residual) (p-value)	11.504 (0.175)
Minimum	−13.1417	Q10 (squared residual) (p-value)	11.546 (0.173)
Std. Dev.	1.0001	Q20 (residual) (p-value)	20.090 (0.328)
Skewness (Z-statistic, p-value)	−0.5062 ($-16.2613, 2.2 \times 10^{-16}$)	Q20 (squared residual) (p-value)	16.385 (0.566)
Kurtosis (Z-statistic, p-value)	7.8954 ($36.6956, 2.2 \times 10^{-16}$)	μ (Z-statistic)	0.0002 *** (4.7722)
Jarque–Bera (p-value)	16607.86 *** (0.0000)	ϕ_1 (Z-statistic)	−0.0812 (−1.0679)
Engle LM (1) (p-value)	6.0183 ** (0.0142)	θ_1 (Z-statistic)	0.1853 ** (2.4722)
Engle LM (2) (p-value)	6.3745 ** (0.0413)	ω (Z-statistic)	9.37e-07 *** (15.5038)
Engle LM (5) (p-value)	7.2949 * (0.1996)	α (Z-statistic)	0.0306 *** (13.0564)
Engle LM (10) (p-value)	11.4924 (0.3205)	γ (Z-statistic)	0.0884 *** (28.0765)
q1%	−2.5373	β (Z-statistic)	0.9154 *** (415.2990)
q5%	−1.6240	Log-likelihood	54439.85
q95%	1.5658	Akaike criterion	−6.8220
q99%	2.3426	Number	15950

*, ** and *** denotes parameter statistically significant at the 90%, 95% and 99% confidence level. q1%, q5%, q95% and q99% represent the empirical quantile measures at respectively 1%, 5%, 95% and 99%.

The results for the maximum likelihood estimation of this model are displayed in Table 1. This model provides very good fit according to the selected criteria; all the parameters of the TGARCH(1,1) specification are statistically significant. We therefore extract the maxima and minima from the return shocks $\{Z_t\}_{t=1,...,T}$ corresponding to standardized returns using a time-varying volatility model. Our discussion is hereafter restricted to the standardized (raw) return maxima $+Z$ ($+R$) and minima

$-Z(-R)$. Figure 1 shows the evolution of the S&P 500 stock index (1) prices; (2) volatility; (3) raw returns and (4) standardized returns.

Figure 1. S&P 500 stock index. The Figure displays graphics from S&P 500 stock index from 3 January 1950 to 28 May 2013. From upper left to lower right corner: The stock index prices. The ARMA(1,1)-TGARCH historical volatility. The raw returns. The next four plots for visual threshold selections from the S&P 500 stock price standardized log-returns. $(+Z)$ stands for upper tail and $(-Z)$ stands for lower tail. The mean residual life plots are approximately linear with positively sloped, which indicates Pareto behavior overall for the $-Z$ lower tail. The threshold plots display the maximum likelihood estimates for the tail index ξ against a range of thresholds with 95% confidence limits. The stability in the parameter estimates is checked when the parameter estimates are approximately constant above the threshold range. The last two plots display for the lower tail $-Z$, the 100-year return level for the fitted GPD with 95% confidence intervals and tail plot based on a generalized Pareto model fitted to losses over the $-Z$-threshold where the three vertical lines (from left to right) locate the 99%, 99.90% and 99.99% expected shortfall level.
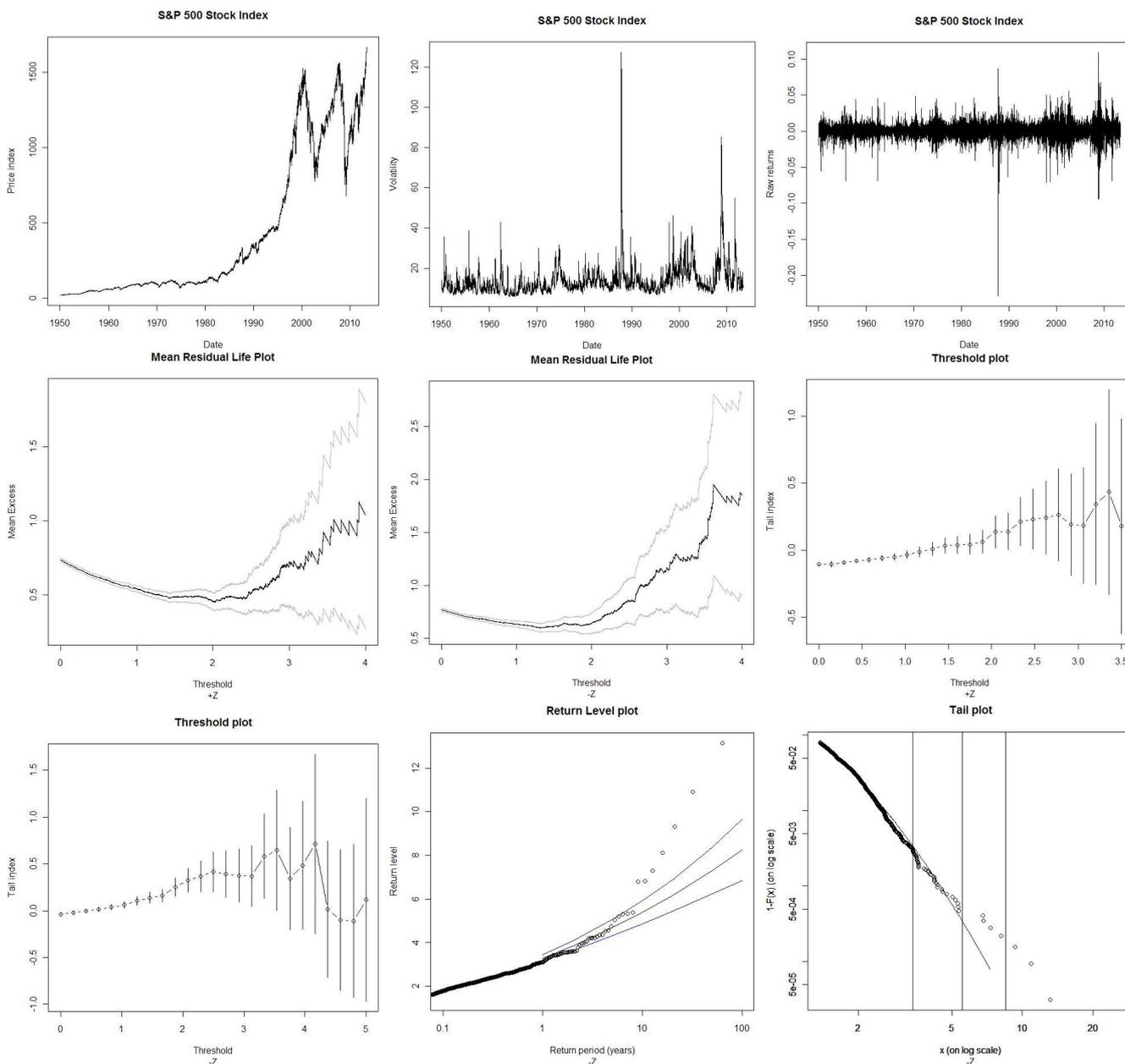
3.3. Descriptive Statistics

S&P 500 standardized returns have excess skewness and kurtosis. In all cases, the Jarque–Bera statistics induces a strong rejection of the normality hypothesis. When considering various percentiles from 1% to 99% as a comparison with those implied by the Normal distribution, it shows a clear departure. The Q-statistics, distributed as a χ_5^2, χ_{10}^2 and χ_{20}^2 with 95% critical value, is 11.07, 18.31 and 31.41. The correlogram for the residuals exhibits no more dependence since the Q-statistics for the series are lower than the critical values. The Engle's Lagrange multiplier test statistic measures the ARCH effect in the residuals; it shows no more evidence of remaining ARCH effects at lag five at the 95% confidence level.

3.4. Quantile Regression

The transformation for recomputing "theoretical raw returns" from standardized returns is very useful for understanding the economic meaning of the statistical inference drawn from EVT results. Indeed, many articles applying EVT to standardized returns leave the reader with few economic interpretations of the results extracted from the filtered series. Therefore, it is important to transform the standardized returns computed from EVT into theoretical raw returns. As recent literature, to our knowledge, does not propose any solution, this article applies a linear transformation based on a semi-parametric technique that extends the ordinary least squares regression model to conditional quantiles. Indeed, standard ordinary least squares modelize the relationship between one or more covariates X and the conditional mean of the response variable Y given X. The quantile regression [17,18] permits a more complete description of the conditional distribution since it generalises the regression model to conditional quantiles of the response variable, such as the 95th or 99th percentile. It appears to be the appropriate approach for the estimation of conditional quantiles of a response variable Y, given a vector of covariates X. Hence, it can be used to measure the effect of covariates, not only in the center of a distribution, but also in the upper and lower tails. Quantile regression is useful when the rate of change in the conditional quantile depends on the quantile. In fact, it can be used with heterogeneous data for which the tails and the center of the conditional distribution vary differently with the covariates. Moreover, it is viewed as a natural extension of OLS estimation with a better consideration brought to extreme observations. It is more robust then OLS because there is no need to do any distributional assumption about the error term, contrary to the normality assumption in the standard regression model. In clear, quantile regression is, by construction, particularly robust to extreme values ([34], p. 7). The linear conditional quantile function can be estimated by solving the following minimization problem:

$$\hat{\gamma}_{n_u}(q) = argmin_{\gamma(q)} E\left\{ \varsigma_q\left(Y - X^{'}\gamma(q)\right)\right\} \qquad (12)$$

The check function which weights positive and negative values asymmetrically for any quantile $0 < q < 1$ is $\varsigma_q(v) = v\left(q - I(v < 0)\right)$, where $I(.)$ denotes the indicator function. The vector of coefficients $\gamma(q) = (\gamma_{1q}, \gamma_{2q}, ..., \gamma_{pq})^{'}$ corresponds to p explanatory variables $X = (c, X_1, X_2...X_p)^{'}$ that take the value of the intercept c and the standardized returns Z. Y is the variable of interest corresponding the the raw returns R.

A quantile regression is implemented for the left tail and another one for the right tail. The choice of the percentile level corresponds exactly to the respective selected threshold levels for each tail, which are computed in Section 4.1.

For the left tail (respectively the right tail), the intercept is -0.0032 (respectively 0.0022), the slope coefficient is 0.0084 (respectively 0.0078). The parameters are all statistically significant at the level of 99%. The adjusted R^2 value is 0.5978 (respectively 0.6185). Hereafter, any transformation of standardized returns Z into theoretical raw returns $\widehat{R}$ will be computed such as $\widehat{R} = -0.0032 + Z \times 0.0084, \forall Z < 0$, or $\widehat{R} = 0.0022 + Z \times 0.0078, \forall Z > 0$.

4. Empirical Findings

4.1. Threshold Selection Results

This article proposes a complete approach for threshold detection mixing visual inspection and automatic detection. The mean residual life plot, in Figure 1, shows that the critical threshold above which the slope is positive is around +1.20 for $+Z$ and -1.40 for $-Z$. According to the threshold plot, a possible range of stability belongs between +0.09 and +1.50 for $+Z$ and between -0.80 and -1.80 for $-Z$. For the optimal threshold detection, we follow [15] who propose a criterion for which the AMSE of the Hill estimator is minimal for the optimal number of observations in the tail. Optimal threshold is around +0.93% for $+Z$ and -1.37% for $-Z$. It corresponds respectively to given number of upper order statistics of 2443 and 1278 observations. The critical threshold values computed from the optimal algorithm respond to the criteria of stability and sufficient exceedances with minimum variance. Table 2 summarizes the results for the threshold selection. Due to the convergence between the three approaches, threshold optimal values are considered in this study.

Table 2. Threshold choice. This Table presents the results for the three approaches for the upper $(+Z)$ and lower $(-Z)$ tail of the S&P500 stock index standardized log-returns. Visual guidance denotes a plausible threshold choice. Mean residual life plot is the first visual inspection where selection is made around linear region. Threshold plot is the second visual inspection method where selection is made around stability region. The given intervals denote the range of acceptable threshold. The optimal selection method is an automated method consisting in minimizing asymptotic mean squared error.

	$+Z$	$-Z$
Mean residual life plot	+1.20	-1.40
Threshold plot	$[+0.09; +1.50]$	$[-0.80; -1.80]$
Optimal selection	+0.9388	-1.3735

4.2. Tail Characteristics

The critical thresholds of +0.93% for $+Z$ and -1.37% for $-Z$ correspond to the entry point of the right and left tail standardized distribution. These locations in the tail area are relatively closed to those inferred from the raw return series. The transformation from Section 3.4 allows to determine the virtual location entry points for both left and right tail. Indeed, the corresponding critical threshold for the

right tail is +0.96% (upper order statistics of 1769) and for the left tail is -1.49% (upper order statistics of 755). This means that approximately beyond $\pm 1.5\%$ price variation, the U.S. stock market enters into a tail area. These thresholds of $\pm 1.5\%$ are the more important results of this study, because they delimit the calm period from the extreme period; they should be analyzed in the light of the associated volatility regime. Indeed, the return–volatility equilibrium requires to examine the stock index returns in the light of the volatility given the asymmetric nature of volatility [35–38]. In clear, when the market returns go below/beyond these thresholds, we should expect to switch into a different volatility regime. Therefore, let's consider three volatility regimes corresponding to the raw returns $<-1.5\%$, belonging to the interval -1.5% and $+1.5\%$ and $>+1.5\%$. We observe that the average volatility is 22.79% for $\widehat{R} < -1.5\%$, which includes 744 observations in the left tail; it is 12.91% for $-1.5\% < \widehat{R} < +1.5\%$; it is 21.86% for $\widehat{R} > +1.5\%$, which includes 732 observations in the right tail. The volatility level is therefore twice the average level of 12.91% when $\widehat{R}$ is superior in absolute value to 1.5%.

Table 3 displays the results for the GPD when considering the critical threshold values. The maximum likelihood estimators of the GPD are the values of the two parameters (ξ, β) that maximize the log-likelihood. The tail index value of $+0.13$ for $-Z$ is positive and statistically significant. The positive sign confirms the presence of fat-tailness for the lower tail $-Z$. The upper tail $+Z$ has a tail index slightly negative, indicating moderate tail behavior corresponding to the Gumbel-type domain of attraction. The scale parameters of $+Z$ and $-Z$ are statistically significant and have the same magnitude.

Table 3. Parameters estimates for the GPD model. Table 3 gives parameter estimates of the general Pareto distribution fitted to the upper $(+Z)$ and lower tails $(-Z)$ of the S&P500 stock index standardized log-returns. The generalized Pareto distribution is fitted to excesses over the optimal threshold. The vector parameters are estimated by the maximum likelihood method. Nb. Exceedances corresponds to the number of observations in the tail. Percentile is the percentage of observations below the threshold. Neg. Lik is the negative logarithm of the likelihood evaluated at the maximum likelihood estimates.

	$+Z$	$-Z$
ξ	-0.0411 ***	0.1359 ***
(s.e)	(0.0143)	(0.0274)
β	0.5671 ***	0.5168 ***
(s.e)	(0.0140)	(0.0201)
Threshold	+0.9388	-1.3735
Nb. Exceedances	2443	1278
Percentile	0.8468	0.9198
Neg. Lik.	957.047	608.1859

*** denotes parameter significantly different from zero at the 99% confidence level.

4.3. The Extreme Downside Risk

Table 4 displays some tailed-related risk measures such as the value-at-risk and expected shortfall measures for both Gaussian and General Pareto distributions. Probability level of 99%, 99.5%, 99.9%,

99.95% and 99.99% are considered. The conservative choice of the 99.99% probability level corresponds to a worst possible movement in 10,000 days or approximately 40 years. The reported 0.01%-GPD-VaR is -7.0% and the 0.01%-GPD-ES is -8.49%. The transformation from Section 3.4 indicates that the 0.01%-GPD-VaR and the 0.01%-GPD-ES would respectively represent -6.27% and -7.53% theoretical raw returns. Figure 1 also displays the associate tail plot in log-log scale where the expected shortfall at the level of 99%, 99.90% and 99.99% are visually located; it shows that the GPD largely underestimates these lowest quantiles even the 99.90% level.

Table 4. Value-at-Risk and expected shortfall. Table 4 gives for each probability level of 99%, 99.50%, 99.90%, 99.95%, 99.99% the VaR and the associated expected shortfall estimates based on a (i) GPD model fitted to the S&P 500 stock index negative standardized log-returns ($-Z$) and a (ii) normal distribution.

Probability	VaR-GPD	ES-GPD	VaR-normal	ES-normal
0.9900	-2.6166	-3.4104	-2.3263	-2.6652
0.9950	-3.1151	-3.9873	-2.5758	-2.8919
0.9990	-4.4709	-5.5564	-3.0902	-3.3670
0.9995	-5.1526	-6.3454	-3.2905	-3.5543
0.9999	-7.0068	-8.4912	-3.7190	-3.9584

Table 5. Return levels. Table 5 displays the results of the return levels for 1, 2, 5, 10, 20, 50 and 100 years. This table displays the result of the fit of the upper tail ($+Z$) and lower tail ($-Z$) of the S&P500 stock index standardized log-returns. 2×2 columns represent upper and lower bound with 95% confidence interval.

Period	$+Z$	Lower bound	Upper bound	$-Z$	Lower bound	Upper bound
1	2.8586	2.7746	2.9425	-3.2857	-3.1384	-3.4330
2	3.1921	3.0833	3.3008	-3.8502	-3.6283	-4.0720
5	3.6186	3.4683	3.7689	-4.6829	-4.3138	-5.0519
10	3.9307	3.7426	4.1188	-5.3854	-4.8609	-5.9098
20	4.2340	4.0031	4.4649	-6.1572	-5.4325	-6.8819
50	4.6219	4.3271	4.9166	-7.2958	-6.2255	-8.3661
100	4.9057	4.5577	5.2537	-8.2564	-6.8526	-9.6601

Table 5 displays the return levels for both positive and negative standardized returns with confidence intervals. The return level plot consists of plotting the theoretical quantiles as a function of the return period with a logarithmic scale for the x-axis so that the effect of extrapolation is highlighted. We note that the return levels for negative shocks ($-Z$) are higher in comparison with the positive shocks ($+Z$). Figure 1 additionally shows the associate plot for negative shocks. The 100-year return level is the level expected to be exceeded once every 100 years. It corresponds to a -8.25% standardized return (or -7.33% theoretical raw return) with an asymmetric confidence interval. This asymmetry reflects the greater uncertainty about large values. The 95% confidence interval is obtained from the profile log-likelihood as $[-6.85\%, -9.66\%]$. This upper limit of -9.66% is greater in absolute value than the magnitude of the Black Monday standardized return of -9.31%, which is, on average, likely to be exceeded in one every 84 years. As a comparison, the worst standardized return from 2007–2008

financial crisis (-7.31%, 27 February 2007) is, on average, likely to be exceeded in one every 26.5 years (or approximately 27 years) [3].

4.4. EVT with Raw Returns

We apply the EVT methodology directly on raw returns. There is a convergence of the results for extreme identifications but less on extreme predictions. Indeed, the critical thresholds are $+1.29\%$ for $+R$ and -1.30% for $-R$, which means that approximately, the S&P 500 stock index becomes extreme from $\pm 1.5\%$ like in the previous section. In addition, concerning the left tail, the 0.01%-GPD-VaR is -7.05% and the 0.01%-GPD-ES is -8.58%. However, we note some differences with the predictions. Indeed, the 100-year return level is -13.62%, which represent a level never reached by the index. The occurence of the Black October crash (19 October 1987, -9.31% with order statistics of 3) is likely, on average, to be exceeded in one every 24.7 years. The occurence of the worst crash from the 2007–2008 financial crisis (22 February 2007, -7.30% with order statistics of 5) is likely, on average, to be exceeded in one every 10.4 years. These predictions are more severe than what is found in Section 4.3.

4.5. Robustness Check

We test the robustness of the methodology that employed EVT combined with quantile regression. We select a sub period (1950–2008) including the financial crisis in order to check for the stability of the parameters. The results are very stable. Indeed, the critical thresholds are $+1.16\%$ for $+Z$ and -1.32% for $-Z$, which corresponds to respective theoretical raw returns of $+1.21\%$ and -1.39% (or approximately $\pm 1.50\%$). In addition for $-Z$, the 0.01%-GPD-VaR is -7.05% and the 0.01%-GPD-ES is -8.58%, which corresponds to respective theoretical raw returns of -6.11% and -7.37%. Finally, the 100-year return level is -8.33%, which corresponds to a -7.17% theoretical raw return and the occurrence of the crashes (19 October 1987, 22 February 2007) remains the same.

5. Conclusions

Extreme value theory is usually applied on standardized returns to offer more reliable results, but this makes the economic interpretation more complicated in the real world.

This paper contributes to the literature by proposing an empirical methodology based on quantile regression for transforming EVT results, expressed in the form of standardized returns, into raw returns. This linear transformation converts EVT results in terms of monetary units easy to interpret.

An empirical test on the S&P500 stock index is made from 1950 to 2013. The main empirical results are the following:

First, entering the tail area begins from $+0.96\%$ (-1.49%) for the right tail (left tail); in other words, U.S. stock market becomes extreme from a price variation of $\pm 1.5\%$.

---

[3] Note that for the sake of prudence, the return periods of 84 and 27 years correspond to their respective return level absolute upper bound

Second, the highest one-day decline from 2007–2008 is likely, on average, to be exceeded one every 27 years versus 84 years for the Black-Monday.

Third, the tailed related risk measures show that the 0.01% GPD VaR/ES correspond to a one-day decline of $-6.27\%/-7.53\%$. Future research will extend this methodology to commodity markets given their high level of volatility.

Conflicts of Interest

The author declares no conflict of interest.

References

1. Danielsson, J.; de Vries, C.G. Tail index quantile estimation with very high frequency data. *J. Empir. Financ.* **1997**, *4*, 241–257.
2. Danielsson, J.; de Haan, L.; Peng, L.; de Vries, C.G. Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivar. Anal.* **2001**, *76*, 226–248.
3. Danielsson, J.; Morimoto, Y. Forecasting extreme financial risk: A critical analysis of practical methods for the Japanese market. *Monet. Econ. Stud.* **2000**, *12*, 25–48.
4. Gencay, R.; Selcuk, F.; Ulugulyagci, A. High volatility, thick tails and extreme value theory in value-at-risk estimation. *Math. Econ.* **2003**, *33*, 337–356.
5. Gettinby, G.; Sinclair, C.D.; Power, D.M.; Brown, R.A. An analysis of the distribution of extremes in indices of share returns in the US, UK and Japan from 1963 to 2000. *Int. J. Financ. Econ.* **2006**, *11*, 97–113.
6. Gilli, M.; Kellezi, E. Extreme Value Theory for Tail-Related Risk Measure; Working Paper; University of Geneva: Geneva, Switzerland, 2000.
7. Jondeau, E.; Rockinger, M. Testing for differences in the tails of stock market returns. *J. Empir. Financ.* **2003**, *10*, 559–581.
8. LeBaron, B.; Samanta, R. Extreme Value Theory and Fat Tails in Equity Markets; Working Paper; Brandeis University: Waltham, MA, USA, 2004.
9. Longin, F. The asymptotic distribution of extreme stock market returns. *J. Bus.* **1996**, *69*, 383–408.
10. Longin, F. Stock market crashes: Some quantitative results based on extreme value theory. *Deriv. Use Trading Regul.* **2001**, *7*, 197–205.
11. McNeil, A.J.; Saladin, T. Developing scenarios for future extreme losses using the POT model. In *Extremes and Integrated Risk Management*; RISK Books: London, UK, 1998.
12. McNeil, A.J. Extreme Value Theory for Risk Managers; Working Paper; University of Zurich: Zürich, Switzerland, 1999.
13. McNeil, A.J.; Frey, R. Estimation of tail-related risk measures for heteroskedastic financial time series: An extreme value approach. *J. Empir. Financ.* **2000**, *7*, 271–300.
14. Tolikas, K.; Brown, R.A. The Distribution of the Extreme Daily Returns in the Athens Stock Exchange; Working Paper; University of Dundee: Dundee, UK, 2005.
15. Beirlant, J.; Goegebeur, Y.; Segers, J.; Teugels, J. *Statistics of Extremes: Theory and Applications*; Wiley: Chichester, UK, 2004.

16. Embrechts, P.; Klüppelberg, C.; Mikosch, T. *Modelling Extremal Events for Insurance and Finance*. Springer: New York, NY, USA, 1997.

17. Koenker R.; Bassett, G. Regression quantiles. *Econometrica* **1978**, *46*, 33–50.

18. Koenker, R.; Hallock, K. Quantile Regression. *J. Econ. Perspect.* **2001**, *15*, 143–156.

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

20. Pickands, J.I. Statistical inference using extreme value order statistics. *Ann. Stat.* **1975**, *3*, 119–131.

21. Coles, S.G. *An Introduction to Statistical Modelling of Extreme Values*; Springer-Verlag: London, UK, 2001.

22. Gumbel, E.J. *Statistics of Extremes*; Columbia University Press: New-York, NY, USA, 1958.

23. Baillie, R. The Asymptotic Mean Square Error of multistep prediction from the regression model with autoregressive errors. *J. Am. Stat. Assoc.* **1979**, *74*, 175–184.

24. Hill, B.M. A simple general approach to inference about the tail of a distribution. *Ann. Stat.* **1975**, *3*, 1163–1174.

25. Acerbi, C. Spectral measures of risk: A coherent representation of subjective risk aversion. *J. Bank. Financ.* **2002**, *26*, 1505–1518.

26. Inui, K.; Kijima, M. On the significance of expected shortfall as a coherent risk measure. *J. Bank. Financ.* **2005**, *29*, 853–864.

27. Kuester, K.; Mittik, S.; Paolella, M.S. Value-at-risk prediction: A comparison of alternative strategies. *J. Financ. Econ.* **2005**, *4*, 53–89.

28. Martins, F.C.; Yao, F. Estimation of value-at-risk and expected shortfall based on nonlinear models of return dynamics and extreme value theory. *Stud. Nonlinear Dyn. Econ.* **2006**, *10*, 107–149.

29. Ozun, A.; Cifter, A.; Yilmazer, S. Filtered Extreme Value Theory for Value-at-Risk Estimation; Working Paper; Marmara University: Kadikoy/Istanbul, Turkey, 2007.

30. Bao, Y.; Lee, T.-H.; Saltoglu, B. Evaluating predictive performance of value-at-risk models in emerging markets: A reality check. *J. Forecast.* **2006**, *25*, 101–128.

31. Yamai, Y.; Yoshiba, T. Value-at-risk *versus* expected shortfall: A practical perspective. *J. Bank. Financ.* **2005**, *29*, 997–1015.

32. Artzner, P.; Delbaen, F.; Eber, J.; Heath, D. Coherent measure of risk. *Math. Financ.* **1999**, *9*, 203–228.

33. Gumbel, E.J. The return period of flood flows. *Ann. Math. Stat.* **1941**, *12*, 163–190.

34. D'Haultfœuille, X.; Givord, P. La Regression de Quantile en Pratique; Working Paper; Insee: Paris, France, 2010.

35. Bekaert, G.; Wu, G. Asymmetric volatility and risk in equity markets. *Rev. Financ. Stud.* **2000**, *13*, 1–42.

36. Black, F. Studies in stock price volatility changes, American Statistical Association. In Proceedings of the Business and Economic Statistics Section, Boston, MA, USA, 23–26 August, 1976; pp. 177–181.

37. Campbell, J.Y.; Hentchel, L. No news is good news: An asymmetric model of changing volatility in stock returns. *J. Financ. Econ.* **1992**, *31*, 281–318.

38. Christie, A. The stochastic behavior of common stock variances—Value, leverage, and interest rate effects. *J. Financ. Econ. Theory* **1982**, *10*, 407–432.