Estimating and Forecasting Conditional Risk Measures with Extreme Value Theory: A Review

Marco Bee [1]* [ID] **and Luca Trapin** [2]

[1] Department of Economics and Management, University of Trento, via Inama 5, 38122 Trento, Italy
[2] Department of Mathematical Sciences, Mathematical Finance and Econometrics, Università Cattolica del Sacro Cuore, Largo Gemelli 1, 20123 Milano, Italy; luca.trapin@unicatt.it
* Correspondence: marco.bee@unitn.it; Tel.: +39-0461-28-2296

Abstract: One of the key components of financial risk management is risk measurement. This typically requires modeling, estimating and forecasting tail-related quantities of the asset returns' conditional distribution. Recent advances in the financial econometrics literature have developed several models based on Extreme Value Theory (EVT) to carry out these tasks. The purpose of this paper is to review these methods.

Keywords: Extreme Value Theory; volatility; risk; quantile

1. Introduction

Quantitative financial risk management has become a fundamental tool for investment decisions, capital allocation and regulation. These tasks usually require estimating and forecasting statistical risk measures based on a time series of asset returns. Several measures of market risk have been proposed in the literature, and the best known are probably the Value-at-Risk (VaR) and the Expected Shortfall (ES).

Introduced by JP Morgan's RiskMetrics in the mid-nineties for the bank internal risk control, the VaR has soon been adopted as the official risk measure for regulatory purposes. This measure is an estimate of how much a financial institution can lose on a portfolio of assets, with a given confidence level, over a given time horizon. Statistically speaking, the VaR coincides with the quantile of the portfolio return distribution over a prescribed holding period. Artzner et al. (1999) show that the VaR has various theoretical flaws as a measure of market risk and suggest using the ES instead. This measures the expected loss of a portfolio given that the loss exceeds the VaR. The Basel Committee on Banking Supervision has recently proposed switching from VaR to ES for the regulatory capital requirements (Basel Committee on Banking Supervision 2013).

In this paper, we are concerned with the estimation of conditional risk measures, where the conditioning is with respect to the available history of the financial time series. As asset returns typically present dynamic dependencies (Cont 2001), the conditional return distribution can strongly differ from the unconditional distribution, thus requiring appropriate time-varying models to compute the conditional risk measures. Let $\{r_t\}_{t\in\mathbb{N}}$ be a time series of portfolio returns. The conditional VaR and ES at level α are defined as

$$VaR_t^\alpha := \inf\left\{x \in \mathbb{R} : \Pr(r_t < x | \mathcal{F}_{t-1}) \geq \alpha\right\},$$

and

$$ES_t^\alpha := \mathbb{E}\left(r_t | r_t \leq VaR_t^\alpha, \mathcal{F}_{t-1}\right),$$

with $\mathcal{F}_{t-1}$ the information set available at time $t-1$. Note that these definitions are related to the lower tail of the conditional return distribution. Without loss of generality, throughout the paper, we will

consider the more convenient definitions based on the upper tail of the conditional loss distribution, i.e., $l_t = -r_t$.

There is a wide variety of methods for estimating and forecasting the conditional VaR and ES in the literature (see Manganelli and Engle (2004); Kuester et al. (2006) and Nieto and Ruiz (2016) among others). The purpose of this paper is to provide a critical discussion of the models based on Extreme Value Theory (EVT) that have been proposed over the last several years to model the tails of the conditional return distribution. EVT provides limit laws for the sample extremes of an *iid* sequence, and these results can be used to derive statistical models for the tails of the sample distribution. These models are unsuitable for estimating unconditional risk measures, as the time dependence that characterizes financial returns fails to match the assumptions underlying the use of EVT (Diebold et al. 2000). Nonetheless, they can be considered valid tools in a conditional setting.

There exist several reviews on EVT in finance (Brodin and Klüppelberg 2008; Diebold et al. 2000; Jondeau et al. 2007; Rocco 2014), but they mostly limit the discussion to unconditional applications or very early applications of conditional EVT. Following the extreme events that characterized the beginning of the 2000s era, scholars in statistics and financial econometrics have proposed novel dynamic extreme value models that quickly react to the changes in the tails of the conditional return distribution. We believe these models deserve a separate treatment.

To provide a clear and self-organized discussion, we classify the available models in three classes according to the mechanism driving the dynamics:

1. **Volatility-EVT**. This class of models proposes a two step procedure that pre-whitens the returns with a model for the volatility and then applies a model based on EVT to the tails of the estimated residuals (Bee et al. 2016; McNeil and Frey 2000).
2. **Quantile-EVT**. This class proposes using time-varying quantile models to obtain a dynamic threshold for the extremes. An extreme value model can then be applied to the exceedances over this threshold (Bee et al. 2018; Engle and Manganelli 2004).
3. **Time-varying EVT**. This class models the returns exceeding a high constant threshold, letting the parameters of the extreme value model to be time-varying to account for the dependence in the exceedancees (Bee et al. 2015, Chavez-Demoulin et al. 2005, 2014).

We put these models in perspective, providing a critical discussion aimed at detailing their strengths and limits. We expect this review to be a valid starting point to set forth a future research agenda on this topic.

2. Extreme Value Theory

This section overviews the main EVT results and presents the Peaks over Threshold method of Davison and Smith (1990), which is the baseline statistical tool underlying the construction of the dynamic extreme value models.

2.1. Main Results

Let $Y_1, \ldots, Y_n$ be a sequence of *iid* random variables with distribution F and upper-end point $\omega_F = \sup\{y : F(y) < 1\}$, and let $M_n = \max\{Y_1, \ldots, Y_n\}$ be the maximum of the sequence. The first important theorem of EVT states that if there exist sequences of real constants d_n and c_n and a distribution H such that

$$\Pr\left(\frac{M_n - c_n}{d_n} \leq m\right) \xrightarrow{d} H(m), \tag{1}$$

then F is in the maximum domain of attraction (MDA) of H, i.e., $F \in \text{MDA}(H)$, and H must be an extreme value distribution. Fisher and Tippett (1928) show that the limiting distribution H can only assume three specifications, depending on the characterization of the tail distribution $\overline{F}(x)$:

- Let $\Phi(m) = e^{-m^{-1/\xi}}$ with $\xi > 0$ be the cumulative distribution function (cdf) of the *Frechét distribution*. As $x \to \infty$,

$$F \in MDA(\Phi) \iff \overline{F}(x) = x^{-\frac{1}{\xi}}L(x),$$

  where $L(x)$ is a slowly varying function
- Let $\Lambda(m) = e^{-e^{-m}}$ being the cdf of the *Gumbel distribution*. As $x \to \infty$,

$$F \in MDA(\Lambda) \iff \overline{F}(x) = e^{-x}.$$

- Let $\Psi(m) = e^{-(-m)^{-1/\xi}}$ with $\xi < 0$ be the cdf of the *Weibull distribution*. As $x \to \infty$,

$$F \in MDA(\Psi) \iff \overline{F}(\omega_F - x^{-1}) = x^{\frac{1}{\xi}}L(x),$$

  where $L(x)$ is a slowly varying function.

Pickands III (1975) obtains the second theorem of EVT, establishing a limiting result for the observations exceeding an increasing threshold u. Consider the following conditional excess distribution above the threshold u,

$$F_u(y) = \Pr\left(Y - u \le y | Y > u\right) = \frac{F(y + u) - F(u)}{1 - F(u)},$$

for $0 \le y \le \omega_F - u$. If the result in Equation (1) holds, a positive measurable function can be found $\nu(u)$ such that

$$\lim_{u \to \omega_F} \sup_{0 \le y \le \omega_F - u} |F_u(y) - GP_{\xi,\nu(u)}(y)| = 0, \tag{2}$$

where $GP_{\xi,\nu(u)}(y)$ is the Generalized Pareto distribution, whose cdf is defined as

$$GP_{\xi,\nu(u)}(y) = \begin{cases} 1 - \left(1 + \xi\frac{y}{\nu}\right)^{-\frac{1}{\xi}}, & \xi \neq 0, \\ 1 - \exp\left(-\frac{y}{\nu}\right), & \xi = 0, \end{cases} \tag{3}$$

for $y \ge 0$ if $\xi \ge 0$ and $0 \le y \le -\frac{\nu}{\xi}$ if $\xi < 0$, where $\nu > 0$ is the scale parameter, and $\xi \in \mathbb{R}$ is the tail parameter.

Consider a sample of N *iid* observations and fix a threshold u such that N_u observations exceed the threshold. Under the assumption that the exceedances exactly follow a GP distribution, Smith (1987) established the asymptotic normality of the maximum likelihood estimator $\left(\widehat{\nu}_{ML}, \widehat{\xi}_{ML}\right)$ as $N_u \to \infty$, provided $\xi > -0.5$. An analogous result can be obtained under the weaker assumption that the distribution of the excesses $F_u(y)$ is only approximately *GP* provided that $N_u \to \infty$ as $N \to \infty$, and $u \to \omega_F$ sufficiently fast. The choice of u thus becomes important as one needs to trade off bias against variance. See the discussion of Dupuis (1999) regarding the threshold selection.

2.2. The Peaks over Threshold Method

Davison and Smith (1990) exploit the result of Pickands III (1975) to obtain a tail estimator for the distribution F. Their approach is called Peaks over Threshold (POT) and relies on the following decomposition for $y > u$:

$$1 - F(y) = \underbrace{(1 - F(u))}_{\phi}\underbrace{(1 - F_u(y - u))}_{\approx GP(\nu,\xi)}. \tag{4}$$

In an *iid* setting, a non-parametric estimator of ϕ can be obtained as $\widehat{\phi} = \frac{N_u}{N}$, and approximating $(1 - F_u(y - u))$ with the GP distribution, an estimate of the conditional excess probability is obtained

by Maximum Likelihood. A tail estimator of F for the observations above $y > u$ can then be obtained by the following closed form formula:

$$\widehat{F}(y) = 1 - \widehat{\phi}\left(1 + \widehat{\xi}\frac{y - u}{\widehat{v}}\right)^{-\frac{1}{\widehat{\xi}}}.\tag{5}$$

In a conditional setting, one is required to develop an econometric model that can be used in conjunction with the tail estimator (5) to obtain an estimate of the tail of the conditional distribution. The following section describes the three approaches available in the financial econometrics literature to obtain an estimate of the conditional return distribution.

3. Estimating Conditional Risk Measures with EVT

The extreme value results outlined in Section 2 rely on the assumption that the underling sequence is *iid*. In this setting, conditioning on the available history is not informative of the next observations, thus the conditional and unconditional distribution coincide, i.e., $\Pr(Y_i \leq y) = \Pr(Y_i \leq y | Y_{i-1}, Y_{i-2}, \dots)$.

The *iid* assumption is certainly not appropriate for financial time series, and to apply the POT method, one needs to take into account the time-variation affecting the tail of the return distribution. In what follows, we describe the three approaches available to do so.

3.1. Volatility-EVT

McNeil and Frey (2000) and Lauridsen (2000) address the conditional quantile problem suggesting a two-step procedure that pre-whitens the asset returns with a dynamic model for the conditional volatility, and then applies the Peaks over Threshold method to the tails of the residuals. If the volatility model is correctly specified, the residuals should be an approximately *iid* sequence, and thus EVT results can be applied.

Letting p_t be the logarithmic price of an asset at time t, and define the returns as $r_t = p_t - p_{t-1}$. We assume that the return process evolves according to the following general equations:

$$\begin{aligned} r_t &= \mu_t + \sigma_t \epsilon_t, \\ \mu_t &= m(\mathcal{F}_{t-1}), \\ \sigma_t^2 &= v(\mathcal{F}_{t-1}), \end{aligned}\tag{6}$$

where the conditional mean μ_t and the conditional variance σ_t^2 are both measurable functions of the available information set $\mathcal{F}_{t-1}$, and $\epsilon_t \sim D(0,1)$ is an *iid* process with mean zero and unit variance. One can assume very general specifications for the dynamics of μ_t and σ_t. If the model is correctly specified, estimates $\widehat{\mu}_t$ and $\widehat{\sigma}_t$ can be computed with Quasi-Maximum Likelihood (Bollerslev and Wooldridge 1992) and used to obtain residuals $\widehat{\epsilon}_t = \frac{r_t - \widehat{\mu}_t}{\widehat{\sigma}_t}$. These should represent an approximately *iid* sequence, and the POT method can be applied to the tails. Fixing a high threshold u and using the tail estimator in Equation (5) on the upper tail of the (negated) residuals, one obtains the following VaR and ES estimators by inversion:

$$\widehat{VaR}_t^\alpha = \widehat{\mu}_t + \widehat{\sigma}_t\left(u + \frac{\widehat{v}}{\widehat{\xi}}\left(\left(\frac{1-\alpha}{N_u/N}\right)^{-\widehat{\xi}} - 1\right)\right),\tag{7}$$

$$\widehat{ES}_t^\alpha = \widehat{\mu}_t + \widehat{\sigma}_t\widehat{\epsilon}_\alpha\left(\frac{1}{1+\widehat{\xi}} + \frac{\widehat{v} - \widehat{\xi}u}{\left(1 - \widehat{\xi}\right)\widehat{\epsilon}_\alpha}\right),\tag{8}$$

where $\widehat{\xi}$ and $\widehat{v}$ are the ML estimates of the GP distribution fitted to $-\widehat{\epsilon}_t$, N_u is the number of observations exceeding the threshold u, and $\widehat{\epsilon}_\alpha$ is the $(1-\alpha)$-quantile of the residuals.

Refinements and improvements of this approach have been proposed over the years. After pre-whitening the returns with a GARCH model (Bollerslev 1986); Laurini and Tawn (2008) noted that the scaled residuals still present some dependence in the extremes, and proposed a declustering procedure to account for this effect. Following recent trends in financial econometrics, Bee et al. (2016) propose extending the information set $\mathcal{F}_{t-1}$ to include intra-day observations and use volatility models based on high-frequency (HF) data to pre-whiten the returns.

3.2. Quantile-EVT

The Volatility-EVT approach assumes that a model for the conditional mean and variance can produce approximately *iid* residuals. However, empirical evidence suggests that time-variation can also characterize higher moments of the conditional return distribution (Hansen 1994; Harvey and Siddique 1999), implying that some form of heterogeneity may be left in the residuals.

To solve this issue, Manganelli and Engle (2004) propose using quantile regression to capture the dynamic dependence at a high quantile of the loss distribution, and applying the POT method to the quantile residuals to extrapolate extreme quantiles. This approach only requires the exceedances to be homogeneous, remaining agnostic about the behavior of the observations below the threshold. It has received increasing attention over the last years, and several quantile regression models used in the financial context have been extended to the Quantile-EVT framework (Bee et al. 2018; Li and Wang 2017; Yi et al. 2014).

Assume the loss process l_t evolves according to the following system of equations:

$$
\begin{aligned}
l_t &= u_t^\theta + \epsilon_t^\theta, \\
u_t^\theta &= q(\mathcal{F}_{t-1}),
\end{aligned}
$$

where $q(\cdot)$ is a measurable function of the information set $\mathcal{F}_{t-1}$, and ϵ_t^θ is such that $\mathcal{Q}_\theta(\epsilon^\theta|\mathcal{F}_{t-1}) = 0$, with $\mathcal{Q}_\theta(\cdot)$ the quantile function evaluated at the probability level θ. Although unusual in a quantile regression context, we make explicit the dynamic of l_t to make the exposition more intuitive. The function $q(\cdot)$ allows for a very general specification of the dynamics of u_t^θ. Depending on the specification, model estimation can be performed by either asymmetric least squares (Koenker and Bassett 1978) or Quasi-Maximum Likelihood (Komunjer 2005). Once an estimate $\hat{u}_t^\theta$ is obtained, the Peaks over Threshold method can be applied to the quantile residuals exceeding zero, i.e., $\hat{\epsilon}_t^\theta = \left\{ \frac{l_t}{\hat{u}_t^\theta} - 1 : l_t > \hat{u}_t^\theta \right\}$. Expressions for the VaR and ES at level α can then be written as

$$
\widehat{VaR}_t^\alpha = \hat{u}_t^\theta \left(1 + \frac{\hat{\nu}}{\hat{\xi}} \left(\left(\frac{1-\alpha}{1-\theta} \right)^{\hat{\xi}} - 1 \right) \right), \tag{9}
$$

$$
\widehat{ES}_t^\alpha = \hat{u}_t^\theta \frac{\frac{\hat{\nu}}{\hat{\xi}} \left(\left(\frac{1-\alpha}{1-\theta} \right)^{-\hat{\xi}} - 1 \right) + \hat{\nu} + 1 - \hat{\xi}}{1 - \hat{\xi}}, \tag{10}
$$

where $\hat{\xi}$ and $\hat{\nu}$ are the ML estimates of the GP distribution fitted to $\hat{\epsilon}_t^\theta > 0$.

3.3. Time-Varying EVT

While the Volatility-EVT and Quantile-EVT approaches work well in practice, they have the drawback of being two-stage procedures, so the results of the EVT analysis are sensitive to the model fitting in the first step. Considering the EVT adage "*Let the tails speak for themselves*", Chavez-Demoulin et al. (2005) propose applying the POT method directly to the upper tail of the loss

series, without pre-whitening with an econometric model, and letting the model parameters vary over time. They consider a conditional version of the tail decomposition in (4)

$$1 - F_t(l_t|\mathcal{F}_{t-1}) = \underbrace{(1 - F_t(u|\mathcal{F}_{t-1}))}_{\phi_t}\underbrace{(1 - F_{u,t}(l_t - u|\mathcal{F}_{t-1}))}_{\approx GP(v_t,\xi_t)} \qquad (11)$$

and specify the time-varying parameters as parametric functions of the available information set

$$\begin{aligned}\phi_t &= \phi\left(\mathcal{F}_{t-1}\right),\\ v_t &= v\left(\mathcal{F}_{t-1}\right),\\ \xi_t &= \xi\left(\mathcal{F}_{t-1}\right).\end{aligned}$$

This modeling approach has attracted a lot of attention over the years. Standard practice keeps the ξ parameter constant to gain stability, and focuses on the specification of ϕ_t and v_t: Herrera and Schipp (2009); Chavez-Demoulin and McGill (2012) and Herrera and Clements (2018) consider non-homogeneous point-processes; Santos and Alves (2013); Santos et al. (2013) and Herrera and Schipp (2013) propose duration-based POT models including the inter-exceedance times in the dynamic parameters; Taylor and Yu (2016) focus on the exceedance probability ϕ_t and advance an autoregressive logit model; Bee et al. (2015) combine realized measures of volatility with standard extreme value regression to induce the dynamic in ϕ_t and σ_t. Model estimation is usually performed by Maximum Likelihood.

The conditional decomposition in Equation (11) relies on the assumption that the exceedance probability and the size of the exceedances are independent conditionally on the available information set. Moreover, it makes difficult the inclusion of autoregressive terms, which are usually found to be important in financial econometrics. A possible solution is specifying a censored model that allows for modeling simultaneously the information below and above the threshold u (Bee et al. 2015; Massacci 2016). Letting $x_t = \max(l_t - u, 0)$, such a model can be written in the current notation as

$$1 - F_t(x_t|\mathcal{F}_{t-1}) = \begin{cases} 1 - \phi_t, & x_t = 0, \\ 1 - \phi_t\left(1 + \xi_t\frac{x_t}{v_t}\right)^{-\frac{1}{\xi_t}}, & x_t > 0, \end{cases} \qquad (12)$$

with ϕ_t, v_t, and ξ_t always functions of the available information set $\mathcal{F}_{t-1}$. Model parameters can be estimated by Maximum Likelihood.

The VaR and ES estimates at level α for these models can then be written as,

$$\widehat{VaR}_t^\alpha = u + \frac{\widehat{v}_t}{\widehat{\xi}_t}\left(\left(\frac{\widehat{\phi}_t}{1-\alpha}\right)^{\widehat{\xi}_t} - 1\right),$$

$$\widehat{ES}_t^\alpha = \frac{\widehat{VaR}_t^\alpha + \widehat{v}_t - \widehat{\xi}_t u}{1 - \widehat{\xi}_t}.$$

4. Discussion

We discuss the Volatility-EVT, Quantile-EVT and Time-Varying EVT approaches performing a risk management analysis on the S&P500 from 2000 to 2014. The data for this series are obtained from the Oxford-Man Institute "Realized Library" version 0.2 (Heber et al. 2009). Figure 1 shows that the returns series of the S&P500 presents the usual stylized facts, such as different volatility regimes depending on the market conditions, a heavy-tailed and asymmetric unconditional return distribution, and strongly autocorrelated squared returns. Trapin (2018) also shows that asset returns typically present persistent *extremal dependence*, i.e., serial dependence in extreme returns, and that this dependence tends to be

stronger in the lower than in the upper tail. This remark holds true also for the S&P500, as can be seen from the plot of the estimated upper- and lower-tail extremograms. Recall that the extremogram can be interpreted as the extreme value analog of the autocorrelation function of a stationary process, and it is defined as the conditional probability of observing two exceedances at h lags apart given that an exceedance occurs at time t (Davis and Mikosch 2009).
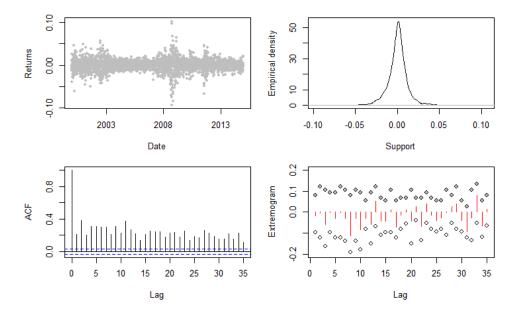

Figure 1. Time series of the S&P500 returns (*Upper-left*), empirical kernel density of the returns (*Upper-right*), autocorrelation of the squared returns (*Lower-left*), and extremogram computed on both the upper (grey bullets) and lower tail (white bullets) of the returns, and their differences in red (*Lower-right*).

The risk management comparison is performed considering several models in the three classes, and computing one-day ahead VaR and ES using a rolling window scheme with daily updates. Our sample consists of $n = 3743$ observations, and fixing the window size to $s = 2000$ allows us to obtain $n^{oos} = 1743$ observations for the validation sample. To evaluate the accuracy of the VaR predictions, we use the tests of Unconditional Coverage (UC), Independent violations (IND) and Conditional Coverage (CC) of Christoffersen (1998). To compare the ES forecasts, we consider the bootstrap test proposed by McNeil and Frey (2000). It tests the hypothesis that, conditional upon exceeding the VaR, the difference between the realized return and the predicted ES has mean zero against the alternative that the mean is greater than zero.

4.1. Volatility-EVT

The Volatility-EVT approach fully relies on the specification of the volatility model. To the extent that this is able to capture the dynamic of the return process and produce nearly independent residuals, good forecasts of the risk measures can be obtained. While volatility models typically capture the serial dependence in the squared returns, Trapin (2018) provides an extensive empirical analysis showing that several of them fail to capture extremal dependence. This important aspect can jeopardize the application of EVT, though the dependence in the body of the distribution is removed. He suggests that models generating extremal dependence and endowed with a leverage component (Glosten et al. 1993) should be used in order to capture the persistent and asymmetric dependence in the two tails. We consider both models that match and fail to match these characteristics, and see the implications of these differences from a risk management perspective. The extreme value analysis on the model residuals is performed fixing the threshold u at the 95th quantile of the negated residual distribution.

We start considering the GARCH(1,1) model of Bollerslev (1986) for the volatility, as in McNeil and Frey (2000) and Lauridsen (2000). In the general setting in (6), we fix μ constant and equal to zero and let σ_t^2 vary according to the following autoregressive structure,

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2.$$

As shown in Basrak et al. (2002), this process is able to generate extremal dependence in the two tails, but it fails to account for the asymmetric extremal dependence as it lacks of a leverage component. To this end, we consider the leveraged GARCH model of Glosten et al. (1993), defined as

$$\sigma_t^2 = \omega + \alpha \left(|r_{t-1}| - \gamma r_{t-1} \right)^2 + \beta \sigma_{t-1}^2,$$

with γ capturing the asymmetric effect of the returns on the future volatility.

Over the last 20 years, the financial econometrics literature has emphasized how better proxies of the volatility can be obtained building realized measures from HF data (see McAleer and Medeiros (2008) for a review). Building on this literature, several HF-based volatility models for the returns and the realized volatility have been proposed to improve model fitting (Engle and Gallo 2006; Hansen et al. 2012; Shephard and Sheppard 2010). Bee et al. (2016) suggest to extend the framework of McNeil and Frey (2000) using this new class of models to capture the volatility dynamic and applying the POT method to the tails of the residuals.

Let RV_t be a realized measure of volatility computed on day t, the first specification we consider is based on the Heterogeneous AutoRegressive (HAR) model of Corsi (2009). This simple three-factor model well captures the seemingly long-range dependence pattern observed in the volatility decay, and in our setting it takes the following form:

$$
\begin{aligned}
\sigma_t^2 &= \exp(RV_t + 0.5\sigma_\eta^2), \\
\log RV_t &= \beta_0 + \beta_d \log RV_{t-1} + \beta_w \tfrac{1}{5} \sum_{j=1}^{5} \log RV_{t-j} + \beta_m \tfrac{1}{22} \sum_{j=1}^{22} \log RV_{t-j} + \eta_t,
\end{aligned}
$$

where η_t is a Gaussian random variable with mean zero and variance σ_η^2. As suggested in Patton and Sheppard (2015), a leveraged extension of this model can be obtained using the realized semi-variance built from the negative HF returns

$$
\begin{aligned}
\sigma_t^2 &= \exp(RV_t + 0.5\sigma_\eta^2), \\
\log RV_t &= \beta_0 + \beta_d \log RV_{t-1} + \beta_w \tfrac{1}{5} \sum_{j=1}^{5} \log RV_{t-j} + \beta_m \tfrac{1}{22} \sum_{j=1}^{22} \log RV_{t-j} \\
&\quad + \gamma_d \log RS_{t-1} + \gamma_w \tfrac{1}{5} \sum_{j=1}^{5} \log RS_{t-j} + \gamma_m \tfrac{1}{22} \sum_{j=1}^{22} \log RS_{t-j} + \eta_t,
\end{aligned}
$$

with RS_t the realized semi-variance computed on day t, and γ_d, γ_w and γ_m capturing the leverage effect at several lags.

Both these processes are Gaussian stationary sequences, and the extreme value results of Breidt and Davis (1998) can be used to show that they cannot produce extremal dependence. In contrast, the High frEquency bAsed VolatilitY models (HEAVY) model of Shephard and Sheppard (2010) can be shown to generate extremal dependence (Trapin 2018). We consider both the non-leveraged and leveraged versions of this model. To keep the discussion contained, we write here the leveraged HEAVY model based on the realized semi-variance,

$$
\begin{aligned}
\sigma_t^2 &= \omega + \kappa RV_{t-1} + \varrho RS_{t-1} + \delta \sigma_{t-1}^2, \\
RV_t &= \mu_t^{RV} \varepsilon_{RV,t}^2, \\
RS_t &= \mu_T^{RS} \varepsilon_{RS,t}^2, \\
\mu_t^{RV} &= \psi_0 + \psi_1 RV_{t-1} + \psi_2 \mu_{t-1}, \\
\mu_t^{RS} &= \varsigma_0 + \varsigma_1 RS_{t-1} + \varsigma_2 \mu_{t-1}^{RS},
\end{aligned}
$$

where $\varepsilon_{RV,t}$ and $\varepsilon_{RS,t}$ are standard Gaussian random variables. The non-leveraged version of this model is obtained removing the equations for the realized semi-variance.

To evaluate the ability of the different volatility models of appropriately capturing the dependence in the extremes, we estimate the extremal index on the upper tail of the negated residuals. The extremal index, denoted with θ, is a measure of the degree of the dependence in the extremes of a stationary series (Embrechts et al. 1997). When $\theta < 1$, the extremes are dependent, conversely with $\theta = 1$, they are independent. Figure 2 displays the estimates of θ obtained with the Ferro and Segers (2003) intervals estimator computed on the models residuals from each moving window. The results are similar to those of Bee et al. (2016), and are consistent with the conclusions of Trapin (2018). Extremal dependence and leverage are essential to capture the serial dependence observed in the extreme returns. This might also explain why Laurini and Tawn (2008) find that the extremes of the residuals of a simple GARCH(1,1) model are still somewhat correlated, requiring further declustering.
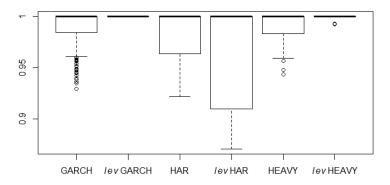

Figure 2. Estimates of the extremal index on the volatility residuals.

Figure 3 depicts the one-day ahead forecasts of the VaR and ES at the $\alpha = 0.01$ level using the volatility predictions and the formulas in (7) and (8). Table 1 reports the p-values of the Christoffersen (1998) tests on the VaR and the bootstrap test on the ES. The tests do not reject the null hypothesis over any model at the 5% confidence level. These results confirm the validity of the Volatility-EVT approach. Moreover, they confirm that the small amount of extremal dependence left in the residuals is not an issue from a risk management perspective. Finally, the HF-based volatility models do not seem to outperform the GARCH(1,1) at the one-day horizon. However, Bee et al. (2016) find that they do for longer horizons, such as 10-day.
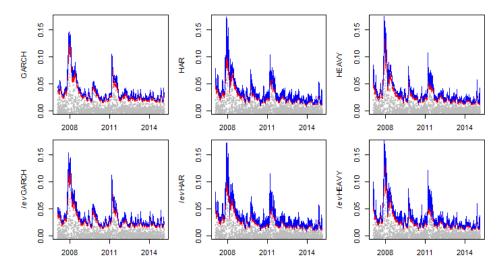

Figure 3. Predicted VaR (*red*) and ES (*blue*) at the $\alpha = 0.01$ level for the different volatility models, and losses of the validation sample (*grey*).

Table 1. *p*-values of the Unconditional Coverage (UC), Independent violations (IND) and Conditional Coverage (CC) tests of Christoffersen (1998) on the VaR predictions, and of the bootstrap test (BOOT) on the ES predictions. Rejections at the 5% level are in bold.

	GARCH	*lev*GARCH	HAR	*lev*HAR	HEAVY	*lev*HEAVY
UC	0.71	0.09	0.20	0.09	0.20	0.09
IND	0.52	0.39	0.44	0.39	0.43	0.39
CC	0.75	0.16	0.32	0.16	0.32	0.16
BOOT	0.62	0.96	0.99	1.00	0.95	1.00

4.2. Quantile-EVT

The quantile-EVT approach uses quantile regression to drive the dynamic of the extreme value model and capture the temporal dependence in the tail. The first quantile model we consider is the Conditional Autoregressive Value at Risk by Regression Quantiles (CAViaR) of Engle and Manganelli (2004). In particular, we adopt the Symmetric Absolute Value (SAV) specification, taking the form

$$u_t^\theta = \beta_1 + \beta_2 u_{t-1}^\theta + \beta_3 |r_{t-1}|,$$

with the past absolute return updating the quantile equation.

Next, we consider a model based on HF data as in the Volatility-EVT approach. As discussed in Žikeš and Baruník (2014) and Bee et al. (2018), including realized measures of volatility in the quantile regression framework can provide more responsive reactions to the quick changes in the market regimes. To this end, we consider a simple modification of the SAV specification where the update is driven by the realized semi-variance, obtaining a *Realized* CAViaR model,

$$u_t^\theta = \beta_1 + \beta_2 u_{t-1}^\theta + \beta_3 RS_{t-1}.$$

The Quantile-EVT approach only requires the quantile residuals $\widehat{\epsilon}_t^\theta$ greater than zero to be approximately independent. To check the validity of this assumption, we estimate the quantile models at the level θ corresponding to the 90th quantile, and compute the extremal index over each moving window with the Ferro and Segers (2003) intervals estimator. Figure 4 reports these estimates for both quantile models considered. One can note that, though the amount of extreme dependence left in the residuals is very small, the Realized-CAViaR model does a better job in cleaning this dependence.
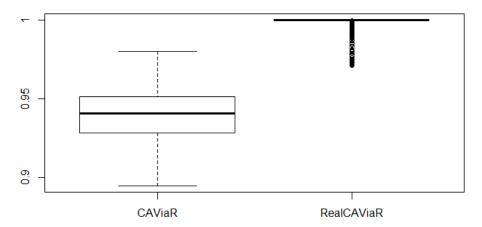

Figure 4. Estimates of the extremal index on the volatility residuals.

Figure 5 depicts the one-day ahead forecasts of the VaR and ES at the $\alpha = 0.01$ level using the forecasts of the next 90th quantile and the formulas in (9) and (10). Table 2 reports the *p*-values of the

Christoffersen (1998) tests on the VaR and the bootstrap test on the ES. Both models provide accurate forecasts of the risk measures and the null hypothesis of the tests are not rejected in any case.
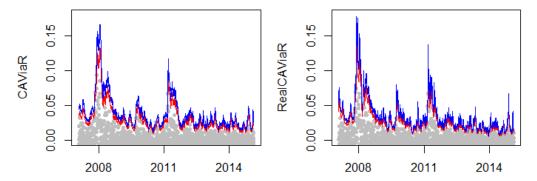

Figure 5. Predicted VaR (*red*) and ES (*blue*) at the $\alpha = 0.01$ level for the CAViaR and Realized CAViaR models, and losses of the validation sample (*grey*).

Table 2. *p*-values of the UC, IND and CC tests of Christoffersen (1998) on the VaR predictions, and of the bootstrap test (BOOT) on the ES predictions. Rejections at the 5% level are in bold.

	CAViaR	RealCAVIAR
UC	0.41	0.09
IND	0.47	0.40
CC	0.54	0.16
BOOT	0.81	1.00

4.3. Time-Varying EVT

Section 3.3 outlines two possible methods to make the standard POT method time-varying. We consider models with the censored structure, as they are more general and allow for easily introducing autoregressive components. The first model we consider, termed Dynamic POT (DPOT), has been proposed in Bee et al. (2015), and uses the Generalized Autoregressive Score method of Creal et al. (2013) to make the parameters ϕ_t and ν_t time-varying, leaving $\xi_t = \xi$ constant to gain stability. Starting from the model in (12), we can write the logarithm of the density of H_t as

$$\log h_t(x_t|\nu_t, \phi_t, \xi) = \begin{cases} \log(1 - \phi_t), & x_t = 0, \\ \log \phi_t - \log \nu_t - \left(1 + \frac{1}{\xi}\right)\log\left(1 + \frac{\xi}{\nu_t}x_t\right), & x_t > 0. \end{cases}$$

Defining the following relationships to constrain the parameter space,

$$\nu_t = \exp(\sigma_t) \quad \Rightarrow \quad \sigma_t = \log \nu_t,$$
$$\phi_t = \frac{1}{1 + \exp(p_t)} \quad \Rightarrow \quad p_t = \log\left(\frac{\phi_t}{1 - \phi_t}\right),$$

we let the dynamics of σ_t and p_t follow

$$\begin{pmatrix} \sigma_t \\ p_t \end{pmatrix} = \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} + \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix} s_{t-1} + \begin{pmatrix} B_{11} & 0 \\ 0 & B_{22} \end{pmatrix} \begin{pmatrix} \sigma_{t-1} \\ p_{t-1} \end{pmatrix},$$

where $\omega_1, \omega_2, A_{11}, A_{22}, B_{11}, B_{22}$ are parameters, and $s_t = S_t \nabla_t$ with

$$\nabla_t = \begin{pmatrix} \frac{\partial \log h_t(x|\nu_t,\phi_t,\xi)}{\partial \sigma_t} \\ \frac{\partial \log h_t(x|\nu_t,\phi_t,\xi)}{\partial p_t} \end{pmatrix}, \qquad S_t = \mathbb{E}_{t-1} \begin{pmatrix} \frac{\partial^2 \log h_t(x|\nu_t,\phi_t,\xi)}{\partial \sigma_t^2} & \frac{\partial^2 \log h_t(x|\nu_t,\phi_t,\xi)}{\partial \sigma_t \partial p_t} \\ \frac{\partial^2 \log h_t(x|\nu_t,\phi_t,\xi)}{\partial p_t \partial \sigma_t} & \frac{\partial^2 \log h_t(x|\nu_t,\phi_t,\xi)}{\partial p_t^2} \end{pmatrix}^{-1}.$$

This model uses the score of the likelihood to efficiently update the dynamic parameters. The second model we consider, proposed in Bee et al. (2015), uses the realized semi-variance instead of the score to drive the dynamics of the time-varying parameters. This *Realized* POT (RPOT) model can be specified as

$$\begin{aligned} \sigma_t &= \omega_1 + \alpha_1 RS_{t-1} + \beta_1 \sigma_{t-1}, \\ p_t &= \omega_2 + \alpha_2 RS_{t-1} + \beta_2 p_{t-1}. \end{aligned}$$

Figure 6 plots the VaR and ES prediction of the DPOT and RPOT models over the validation sample. One can note that the RPOT model tends to react more quickly than the DPOT model to the changes in the tail. This is consistent with the findings in Bee et al. (2015) and can be explained by the fact that the DPOT has a very limited amount of information to update the time-varying parameters. This is confirmed by the results of the risk management example in Table 3, where the DPOT rejects the null of independent violations.
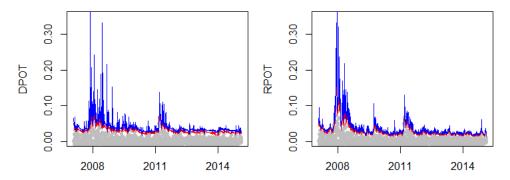

Figure 6. Predicted VaR (*red*) and ES (*blue*) at the $\alpha = 0.01$ level for the DPOT and RPOT models, and losses of the validation sample (*grey*).

Table 3. *p*-values of the UC, IND and CC tests of Christoffersen (1998) on the VaR predictions, and of the bootstrap test (BOOT) on the ES predictions. Rejections at the 5% level are in bold.

	DPOT	**RPOT**
UC	0.41	0.71
IND	**0.02**	0.52
CC	0.06	0.75
BOOT	0.70	0.94

5. Conclusions

This paper reviews three classes of dynamic models that use Extreme Value Theory to estimate and forecast conditional tail-risk measures. For each class, we describe the econometric setting, explain the estimation procedure, and provide closed-form formulas for the VaR and ES. A risk management analysis based on the S&P500 suggests that models in these classes perform well and that an overall winner is difficult to identify. Nonetheless, we can comment on the strengths and the weaknesses of each approach.

The volatility-EVT approach makes the stringent assumption that the dynamic of the conditional distribution can be explained just modeling the second moment, though this does not seem to be an issue in the risk management exercise. On the other hand, it is probably the easier-to-implement

approach. The financial econometrics literature provides an extremely long list of volatility models that can be used within this approach, and, for many of them, routines for estimation and prediction are available in R (R Core Team 2018) and Matlab$^{©}$ (MATLAB and Statistics Toolbox Release 2018a, The MathWorks, Inc., Natick, Massachusetts, United States). Once a model is chosen, EVT is easily applied to the tails of the residuals.

The quantile-EVT approach makes a less stringent assumption than the volatility-EVT approach, and has analogous complexity. However, it has only been recently investigated, and routines for the quantile regression models are not as widespread as for the volatility models. We implemented the R codes for the CAViaR and Realized CAViaR models used in this paper following the instructions of Engle and Manganelli (2004). Note, however, that the estimation of these models is more involved than that of a volatility model, such as a GARCH, as they use the *tick loss* as objective function.

The Time-varying EVT approach fully embraces the spirit of EVT, but it is also the least investigated of the three approaches. If one is willing to exclude autoregressive components and consider only information from exogenous covariates, then this approach is very easy to implement. Standard probit or logit routines can be used for the exceedance probability, and extreme value regression routines can be used for the size of the exceedances. Bee et al. (2015) show that the realized measures are highly informative of the behavior of the extremes, and very simple dynamic extreme value models can be obtained using them as covariates. Including autoregressive components comes at the cost of specifying a censored model. This can be very effective as seen in Figure 6, especially when used in combination with a realized measure. However, routines for these models are not available to the best of our knowledge, and their estimation is more challenging, particularly when a score-driven update is considered to induce the dynamic in the parameters.

All three of the classes contemplate both models based on past daily observations and models based on past intra-daily observations. HF data have made a breakthrough in the volatility literature, allowing to build more accurate measures of volatility and improve volatility forecasts. From the risk management analysis on the S&P500, the benefit of using HF data does not show off when it comes to forecast tail-risk measures. However, one can note from Figures 3, 5 and 6 that models based on the realized measures tend to be much more reactive. As long as a rich stream of intra-daily observations is available, one should consider introducing these data in its model specification.

Finally, we only consider econometric models that are formulated in terms of measurable functions of the available information set $\mathcal{F}_{t-1}$. This is a convenient choice as these models are easy to estimate and are frequently used in the risk management literature. Note, however, that, for each class, models with latent dynamic parameters can be considered. For instance, one can use a stochastic volatility model (Taylor 1994) in the Volatility-EVT approach. Analogously, Chavez-Demoulin et al. (2014) develop a time-Varying EVT latent framework in a Bayesian setting. Alternatively, one can adopt a non-parametric approach (see Schaumburg (2012) for an example in the quantile-EVT setting).

The pitfalls and opportunities of EVT in finance are now well known, and several scholars over the last twenty years have tried to overcome the pitfalls to seize the opportunities of EVT. Guidance on how to use EVT in finance is now available both in the unconditional and conditional settings. However, the current solutions are more directed toward the univariate problem. Future research should be more focused on the multivariate issues, i.e., how to use EVT for the risk assessment of a portfolio of assets.

Author Contributions: M. Bee and L. Trapin conceived and designed the review of the theory; L. Trapin analyzed the data and wrote the paper.

References

Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber, and David Heath. 1999. Coherent measures of risk. *Mathematical Finance* 9: 203–28. [CrossRef]

Basel Committee on Banking Supervision. 2013. Fundamental Review of the Trading Book: A Revised Market Risk Framework. Bank for International Settlement. Available online: https://www.bis.org/publ/bcbs265.pdf (accessed on 13 April 2018).

Basrak, Bojan, Richard A. Davis, and Thomas Mikosch. 2002. Regular variation of GARCH processes. *Stochastic Processes and Their Applications* 99: 95–115. [CrossRef]

Bee, Marco, Dupuis Debbie Janice, and Luca Trapin. 2015. *The Realized Peaks over Threshold: A High-Frequency Based Extreme Value Approach for Financial Time Series*. Cahiers du Gerad G-2015-104. Montreal: GERAD.

Bee, Marco, Debbie J. Dupuis, and Luca Trapin. 2016. Realizing the extremes: Estimation of tail-risk measures from a high-frequency perspective. *Journal of Empirical Finance* 36: 86–99. [CrossRef]

Bee, Marco, Debbie J. Dupuis, and Luca Trapin. 2018. Realized extreme quantile: A joint model for conditional quantiles and measures of volatility with evt refinements. *Journal of Applied Econometrics* 33: 398–415. [CrossRef]

Bollerslev, Tim. 1986 Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307–27. [CrossRef]

Bollerslev, Tim, and Jeffrey M. Wooldridge. 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* 11: 143–72. [CrossRef]

Breidt, F. Jay, and Richard A. Davis. 1998. Extremes of stochastic volatility models. *Annals of Applied Probability* 8: 664–75. [CrossRef]

Brodin, Erik, and Claudia Klüppelberg. 2008. *Extreme Value Theory in Finance*. Hoboken: John Wiley & Sons, Ltd.

Chavez-Demoulin, Valerie, and J. A. McGill. 2012. High-frequency financial data modeling using Hawkes processes. *Journal of Banking & Finance* 36: 3415–26.

Chavez-Demoulin, Valerie, Anthony C. Davison, and Alexander J. McNeil. 2005. Estimating value-at-risk: A point process approach. *Quantitative Finance* 5: 227–34, [CrossRef]

Chavez-Demoulin, Valerie, Paul Embrechts, and Sylvain Sardy. 2014. Extreme-quantile tracking for financial time series. *Journal of Econometrics* 181: 44–52 [CrossRef]

Christoffersen, Peter F. 1998. Evaluating interval forecasts. *International Economic Review* 32: 841–62. [CrossRef]

Cont, Rama. 2001. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance* 1: 223–36. [CrossRef]

Corsi, Fulvio. 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7: 174–96. [CrossRef]

Creal, Drew, Siem Jan Koopman, and Andre Lucas. 2013. Generalized autoregressive score models with applications. *Journal of Applied Econometrics* 28: 777–95. [CrossRef]

Davis, Richard A., and Thomas Mikosch. 2009. The extremogram: A correlogram for extreme events. *Bernoulli* 15: 977–1009. [CrossRef]

Davison, Anthony C., and Richard L. Smith. 1990. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B* 52: 393–442.

Diebold, Francis X., Til Schuermann, and John D. Stroughair. 2000. Pitfalls and opportunities in the use of extreme value theory in risk management. *The Journal of Risk Finance* 1: 30–35. [CrossRef]

Dupuis, Debbie J. 1999. Exceedances over high thresholds: A guide to threshold selection. *Extremes* 1: 251–61. [CrossRef]

Embrechts, Paul, Claudia Klüppelberg, and Thomas Mikosch. 1997. *Modelling Extremal Events: For Insurance and Finance*. New York: Springer Science & Business Media, vol. 33.

Engle, Robert F., and Giampiero M. Gallo. 2006. A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics* 131: 3–27. [CrossRef]

Engle, Robert F., and Simone Manganelli. 2004. Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22: 367–81.

Ferro, Christopher AT, and Johan Segers. 2003. Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B* 65: 545–56. [CrossRef]

Fisher, Ronald Aylmer, and Leonard Henry Caleb Tippett. 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*. Cambridge: Cambridge University Press, vol. 24, pp. 180–90.

Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance* 48: 1779–801. [CrossRef]

Hansen, Bruce E. 1994. Autoregressive conditional density estimation. *International Economic Review* 35: 705–30. [CrossRef]

Hansen, Peter Reinhard, Zhuo Huang, and Howard Howan Shek. 2012. Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics* 27: 877–906. [CrossRef]

Harvey, Campbell R., and Akhtar Siddique. 1999. Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis* 34: 465–87. [CrossRef]

Heber, Gerd, Asger Lunde, Neil Shephard, and Kevin Sheppard. 2009. *Oxford-Man Institute Realized Library*, Version 0.2. Avaliable online: https://realized.oxford-man.ox.ac.uk/ (accessed on 10 March 2018).

Herrera, Rodrigo, and Adam Clements. 2018. Point process models for extreme returns: Harnessing implied volatility. *Journal of Banking & Finance* 88: 161–75.

Herrera, Rodrigo, and Bernhard Schipp. 2009. Self-exciting extreme value models for stock market crashes. In *Statistical Inference, Econometric Analysis and Matrix Algebra*. New York: Springer, pp. 209–31.

Herrera, Rodrigo, and Bernhard Schipp. 2013. Value at risk forecasts by extreme value models in a conditional duration framework. *Journal of Empirical Finance* 23: 33–47. [CrossRef]

Jondeau, Eric, Ser-Huang Poon, and Michael Rockinger. 2007. *Financial Modeling under Non-Gaussian Distributions*. New York: Springer Science & Business Media.

Koenker, Roger, and Gilbert Bassett. 1978. Regression quantiles. *Econometrica* 46: 33–50. [CrossRef]

Komunjer, Ivana. 2005. Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics* 128: 137–64. [CrossRef]

Kuester, Keith, Stefan Mittnik, and Marc S. Paolella. 2006. Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics* 4: 53–89. [CrossRef]

Lauridsen, Sarah. 2000. Estimation of value at risk by extreme value methods. *Extremes* 3: 107–44. [CrossRef]

Laurini, Fabrizio, and Jonathan A. Tawn. 2008. Regular variation and extremal dependence of garch residuals with application to market risk measures. *Econometric Reviews* 28: 146–69. [CrossRef]

Li, Deyuan, and Huixia Judy Wang. 2017. Extreme quantile estimation for autoregressive models. *Journal of Business & Economic Statistics*. In Press.

Simone Manganelli, and Robert F. Engle. 2004. *A Comparison of Value at Risk Models in Finance*. Chichester: Wiley.

Massacci, Daniele. 2016. Tail risk dynamics in stock returns: Links to the macroeconomy and global markets connectedness. *Management Science* 63: 3072–89. [CrossRef]

McAleer, Michael, and Marcelo C. Medeiros. 2008. Realized volatility: A review. *Econometric Reviews* 27: 10–45. [CrossRef]

McNeil, Alexander J., and Rüdiger Frey. 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance* 7: 271–300. [CrossRef]

Nieto, Maria Rosa, and Esther Ruiz. 2016. Frontiers in var forecasting and backtesting. *International Journal of Forecasting* 32: 475–501. [CrossRef]

Patton, Andrew J., and Kevin Sheppard. 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics* 97: 683–97. [CrossRef]

Pickands III, James. 1975. Statistical inference using extreme order statistics. *Annals of Statistics* 3: 119–31.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna: R Core Team. Available online: https://www.R-project.org/ (accessed on 10 April 2018).

Rocco, Marco. 2014. Extreme Value Theory in finance: A survey. *Journal of Economic Surveys* 28: 82–108. [CrossRef]

Santos, P. Araújo, and M. I. Fraga Alves. 2013. Forecasting value-at-risk with a duration-based pot method. *Mathematics and Computers in Simulation* 94: 295–309. [CrossRef]

Santos, Paulo Araújo, Isabel Fraga Alves, and Shawkat Hammoudeh. 2013. High quantiles estimation with quasi-port and dpot: An application to value-at-risk for financial variables. *The North American Journal of Economics and Finance* 26: 487–96. [CrossRef]

Schaumburg, Julia. 2012. Predicting extreme value at risk: Nonparametric quantile regression with refinements from extreme value theory. *Computational Statistics & Data Analysis*, 56: 4081–96.

Shephard, Neil, and Kevin Sheppard. 2010. Realising the future: Forecasting with high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics* 25: 197–231. [CrossRef]

Smith, Richard L. 1987. Estimating tails of probability distributions. *Annals of Statistics* 15: 1174–207. [CrossRef]

Taylor, James W., and Keming Yu. 2016. Using auto-regressive logit models to forecast the exceedance probability for financial risk management. *Journal of the Royal Statistical Society: Series A* 179: 1069–92. [CrossRef]

Taylor, Stephen J. 1994. Modeling stochastic volatility: A review and comparative study. *Mathematical Finance* 4: 183–204. [CrossRef]

Trapin, Luca. 2018. Can volatility models explain extreme events? *Journal of Financial Econometrics* 16: 297–315. [CrossRef]

Yi, Yanping, Xingdong Feng, and Zhuo Huang. 2014. Estimation of extreme value-at-risk: An evt approach for quantile garch model. *Economics Letters* 124: 378–81. [CrossRef]

Žikeš, Filip, and Jozef Baruník. 2014. Semi-parametric conditional quantile models for financial returns and realized volatility. *Journal of Financial Econometrics* 14: 185–226. [CrossRef]