Journal of Financial Econometrics, 2019, Vol. 17, No. 2, 254–283
doi: 10.1093/jjfinec/nbz003
Advance Access Publication Date: 5 March 2019
Article

Realized Peaks over Threshold: A Time-Varying Extreme Value Approach with High-Frequency-Based Measures*

Marco Bee[1], Debbie J. Dupuis[2] and Luca Trapin[3]

[1]University of Trento, [2]HEC Montréal and [3]Università Cattolica del Sacro Cuore

Address correspondence to Debbie J. Dupuis, Department of Decision Sciences, HEC Montréal, 3000 Chemin de la Côte-Sainte-Catherine, Montréal, Québec, Canada H3T 2A7, or email: debbie.dupuis@hec.ca

Received September 1, 2016; revised January 23, 2019; editorial decision January 28, 2019; accepted February 4, 2019

Abstract

Recent contributions to the financial econometrics literature exploit high-frequency (HF) data to improve models for daily asset returns. This paper proposes a new class of dynamic extreme value models that profit from HF data when estimating the tails of daily asset returns. Our *realized peaks-over-threshold* approach provides estimates for the tails of the time-varying conditional return distribution. An in-sample fit to the S&P 500 index returns suggests that HF data convey information on daily extreme returns beyond that included in low frequency (LF) data. Finally, out-of-sample forecasts of conditional risk measures obtained with HF measures outperform those obtained with LF measures.

Key words: conditional risk measures, forecasting, peaks-over-threshold, realized volatility, tail risk

JEL classification: C1, C51, C43

Tail risk has been at the heart of discussions among economists, bankers, and world leaders in the aftermath of the 2008 stock market crash. Despite being an elusive notion, tail risk tends to be associated to large negative events that have a positive but rather small probability of occurrence. Appropriate management of this kind of risk is of the utmost importance from both policy and regulatory perspectives and for the internal risk control of financial institutions. For this purpose, several risk measures have been defined which require forecasting quantiles deep in the lower tail of the asset return distribution.

* The authors wish to thank two anonymous referees and the Associate Editor for helpful comments that improved the paper. The second author acknowledges the support of the Natural Sciences and Engineering Research Council of Canada RGPIN-2016-04114 and the Fondation HEC.

Traditional parametric methods based on estimation of entire densities are mostly ill-suited for the assessment of extreme quantiles. These parametric methods strive to produce a good fit in regions where most of the data fall, potentially at the expense of a good fit in the tails. Similarly, it is well-known that non-parametric methods of density estimation such as kernel smoothing perform poorly in the tails.

Extreme value theory (EVT) is a branch of probability theory that focuses on extreme outcomes and provides models for them. In particular, instead of forcing a single distribution on the entire sample, this theory allows for the investigation of only the tails of the sample distribution using limit laws. Estimates of probabilities associated with quantiles even higher than the most extreme observations are then obtained by extrapolation.

The use of EVT in financial applications has become more and more common over the last 20 years. Danielsson and de Vries (1997) and Longin (2000) use EVT to model the *unconditional* return distribution and emphasize its accuracy in predicting tail-risk. In a critical discussion of the use of EVT in risk management, Diebold, Schuermann, and Stroughair (2000) outline both the opportunities and the pitfalls of such applications. Their main criticism regards the time dependence that characterizes financial returns. Specifically, while the probabilistic results underlying the theory hold for *iid* observations, time series in economics and finance usually do not satisfy this requirement. Despite that, these authors support the approach and foster its application to the tails of the *conditional* return distribution.

To model the tails of the time-varying conditional return distribution, two different paths have been taken. One consists in specifying a model for the conditional mean and variance, and then applying an EVT-based model to the tails of the standardized residuals (McNeil and Frey, 2000). If the model for the first two conditional moments completely characterizes the dependence structure, then the standardized residuals should be approximately *iid*. A second strategy is to fit a dynamic extreme value model to account for the dependence in the original data (Chavez-Demoulin, Davison, and McNeil, 2005; Chavez-Demoulin, Embrechts, and Sardy, 2014). The benefit of the first strategy is that all the available observations are exploited in the estimation of the dynamic model. However, if the extremes of the estimated residuals present any form of heterogeneity, it will be necessary to model them directly.

The claim that a model for the conditional mean and variance can produce *iid* residuals implies that higher conditional moments are constant over time, but the evidence seems to argue otherwise (Hansen, 1994; Harvey and Siddique, 1999). Furthermore, the possible presence of switching-regimes, as suggested by Mikosch and Stărică (2004), makes the task of pre-whitening the return time series even more difficult. These considerations support direct modeling of the extremes of the original sample, and the current paper proposes a novel dynamic extreme value approach to do so.

Dynamic EVT modeling in finance requires finding both an economically sound source of information that can be used to explain the time-varying behavior of the extremes, and an appropriate way of employing this information within a suitable model. Chavez-Demoulin, Davison, and McNeil (2005) suggests using a self-exciting process to model the probability of exceeding a high threshold of the negated return distribution (*loss distribution*), and use a time-varying generalized Pareto (GP) distribution with the past exceedances as covariates to model the size of the excesses. Chavez-Demoulin, Embrechts, and Sardy (2014) models the intensity parameter of the non-homogeneous Poisson process describing

the exceedance rate and the time-varying scale parameter of the GP with non-parametric Bayesian smoothers. Massacci (2017) and Zhang and Bernd (2016) propose two different joint dynamic specifications for the exceedance probability and the distribution of the size of the exceedances based on a generalized autoregressive score model (Creal, Koopman, and Lucas, 2013).

The approaches listed in the previous paragraph exploit information in past daily returns. We consider a completely different perspective grounded on the results of Andersen et al. (2003). They show that, if the continuous-time return process of an asset traded on a frictionless market evolves according to an Itô semi-martingale with a drift and a diffusion component then, under certain conditions, the daily return distribution conditional on the available information is Gaussian with mean equal to the integrated drift and variance equal to the integrated variance. Empirical evidence, however, points toward the existence of market frictions causing phenomena, such as jumps and leverage, inducing heaviness in the tails. Moreover, jumps and leverage present dynamics that are often found to be related to the level of volatility, see for instance Bandi and Renò (2012) and Bollerslev and Todorov (2011). Based on this argument, we find sensible to link the dynamic of daily extremes to that of the daily variance. Our intuition is well illustrated in Figure 1 where losses for the S&P 500 are shown for the 2000–2014 period. Points indicate the days on which the 98th quantile of the loss distribution is exceeded. They tend to be concentrated in periods of high realized variance (RV).

We build models for the extremes based on realized measures of the daily asset price variation obtained from high-frequency (HF) data, that is, intra-daily returns. From a methodological perspective, we follow the Peaks-over-threshold (POT) method of Davison and Smith (1990) and propose a *realized* POT (RPOT) approach. We model the probability of exceeding a high threshold with a Logit model with realized measures as covariates. The size of the excesses is modeled using a GP distribution with time-varying parameters that are functions of the realized measure. We show that measures of the asset price variation obtained from HF data hold information about the tail beyond that contained in the daily exceedances. This result is robust across several proxies of the daily variance and it is not affected by the inclusion of jump and illiquidity measures. Finally, we find that out-of-sample risk measures forecasts from our model outperform those obtained from competing models based on past exceedances.

There is a growing literature that attempts to exploit HF data to enrich models for lower frequency (LF) data. Inclusion of HF data to model the conditional second moment of the return distribution has been proposed by Shephard and Sheppard (2010); Noureldin, Shephard, and Sheppard (2012); Hansen, Huang, and Shek (2012); Hansen, Lunde, and Voev (2014); and Hansen and Huang (2016). De Lira Salvatierra and Patton (2015) propose incorporating HF data into models for the time-varying dependence in a copula function, while Oh and Patton (2016) use a HF-based measure of correlation to disentangle the linear from the non-linear dependence in a portfolio of stocks and then model the non-linear dependence with joint-symmetric copulas. We propose to model the tails of the conditional return distribution with a class of EVT models that incorporate HF information.

The remainder of the paper is organized as follows. Section 1 presents an overview of the standard and the conditional POT approaches. Section 2 outlines our RPOT approach and shows how realized measures are employed. Section 3 contains simulations assessing

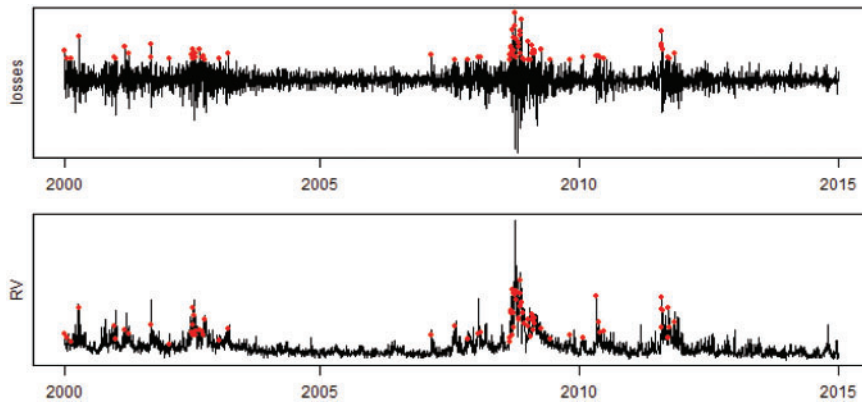

Figure 1 S&P 500. Daily negative returns (*losses*) and RV (RV) from January 1, 2000 to December 31, 2014. Dots indicate days for which a threshold set at the 98th quantile of the loss distribution is exceeded.

the small sample properties of our maximum-likelihood (ML) estimator, the effect of microstructure noise, and the validity of our approach for a general return process. The method is applied to daily S&P 500 index returns in Section 4 where in-sample results show that HF data are more informative than LF data with respect to the tail behavior. Section 5 shows that out-of-sample forecasts of standard risk measures obtained with the RPOT approach are more accurate than those obtained with dynamic extreme value models that only include LF realized measures. Section 6 concludes. A proof is relegated to the Appendix and the Supplementary Material contains additional details and results.

1 Extreme Value Theory

1.1 The Peaks over Threshold Approach

Let $\{Y_t\}_{t=1}^T$ be a sequence of *iid* random variables from a distribution function F with upper end point $v_F := \sup\{Y_t : F(Y_t) < 1\}$. Define the extremes of $\{Y_t\}_{t=1}^T$ to be the exceedances of a high threshold u, $u < v_F$. As $u \to v_F$, Pickands (1975) shows that the distribution of the excesses $(Y_t - u)_+$ converges to a GP distribution G with shape parameter ξ and scale parameter $\nu > 0$. That is, $\Pr(Y - u \le y | Y > u)$ goes to

$$G(y;\xi,\nu) = \begin{cases} 1 - \{1 + \xi y/\nu\}^{-\frac{1}{\xi}} & \text{for} \quad \xi \neq 0 \\ 1 - \exp\{-y/\nu\} & \text{for} \quad \xi = 0\,. \end{cases} \tag{1}$$

When $\xi > 0$, F has a Pareto-type upper tail with tail index $1/\xi$. For a given threshold u, the POT approach is based on the decomposition of the tail of F as

$$1 - F(u+y) = (1 - F(u))\big(1 - F_u(y)\big), \tag{2}$$

where $\phi \equiv (1 - F(u)) = \Pr(Y > u)$ and $F_u(y) = \Pr(Y - u \le y | Y > u)$. Letting $\mathcal{E} = \{t \in \{1,\ldots,T\} | Y_t > u\}$ be the set of times at which an exceedance occurs, the number of

exceedances $N_u(T) = \text{card}\{\mathcal{E}\}$ can be modeled as a binomial random variable[1] with exceedance probability ϕ, and the size of excesses $\{W_j = Y_j - u | j \in \mathcal{E}\}$ with the limiting GP distribution (Coles, 2001). An estimate of the tail probability in Equation (2) can thus be obtained as

$$\widehat{\overline{F}}(y) = \widehat{\phi}\left(1 + \widehat{\xi}\frac{y - \widehat{u}}{\widehat{\nu}}\right)^{-1/\widehat{\xi}}, \tag{3}$$

where $\overline{F} = 1 - F$, $\widehat{u}$ is an appropriately chosen threshold, $\widehat{\xi}$ and $\widehat{\nu}$ are estimates of the GP parameters, and $\widehat{\phi}$ is an estimate of the exceedance probability. The joint likelihood function for the POT model can be written as

$$\begin{aligned}\mathcal{L}(\nu, \xi, \phi) &= \prod_{t=1}^{T}(1-\phi)^{1-I_t}\left(\frac{\phi}{\nu}\left[1 + \frac{\xi(y_t - u)}{\nu}\right]_+^{-1/\xi-1}\right)^{I_t}, \\ &= \left[\prod_{t=1}^{T}(1-\phi)^{1-I_t}\phi^{I_t}\right]\left[\prod_{t=1}^{T}\left(\frac{1}{\nu}\left[1 + \frac{\xi(y_t - u)}{\nu}\right]_+^{-1/\xi-1}\right)^{I_t}\right], \\ &= \mathcal{L}(\phi)\mathcal{L}(\xi, \nu).\end{aligned} \tag{4}$$

where $\mathcal{L}(\phi)$ and $\mathcal{L}(\xi, \nu)$ are the Binomial and GP likelihoods, respectively. Separate maximization leads to the ML estimators $\widehat{\phi}$, $\widehat{\xi}$, and $\widehat{\nu}$. We have the closed-form estimator $\widehat{\phi} = N_u/T$. Numerical optimization (Hosking and Wallis, 1987) is required to find $\widehat{\xi}$ and $\widehat{\nu}$ and their asymptotic normality is established for $\xi > -0.5$ (Smith, 1985).

1.2 The *Conditional* Peaks over Threshold Approach

Following Davison and Smith (1990), the tail of the conditional distribution of Y_t can be decomposed as

$$\Pr(Y_t > u + y | \mathcal{F}_{t-1}) = \Pr(Y_t > u | \mathcal{F}_{t-1}) \Pr(Y_t - u > y | Y_t > u, \mathcal{F}_{t-1}), \tag{5}$$

where $\mathcal{F}_{t-1}$ is the information set of the process up to time $t-1$. An estimate of the conditional tail probability at time t can be obtained combining a dynamic model for $\phi_t = \Pr(Y_t > u | \mathcal{F}_{t-1})$ such as a generalized linear model for the counts $N_u(t)$ and a GP distribution with parameters depending on covariates for $\Pr(Y_t - u > y | Y_t > u, \mathcal{F}_{t-1})$.

With a slight misuse of notation, the joint likelihood of the conditional tail can be written as

$$\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\xi}, \boldsymbol{\nu}) = f_{\phi_1, \xi_1, \nu_1}(\phi_1, \xi_1, \nu_1)\prod_{t=2}^{T} f_{\phi_t, \xi_t, \nu_t | \mathcal{F}_{t-1}}(\phi_t, \xi_t, \nu_t | \mathcal{F}_{t-1}), \tag{6}$$

where $f_{\phi_t, \xi_t, \nu_t | \mathcal{F}_{t-1}}(\phi_t, \xi_t, \nu_t | \mathcal{F}_{t-1})$ is the joint density of the model at time t conditional on information at time $t-1$. Assuming conditional independence between the rate and the magnitude of the exceedances, one obtains

$$f(\phi_t, \xi_t, \nu_t | \mathcal{F}_{t-1}) = f(\phi_t | \mathcal{F}_{t-1}) f(\xi_t, \nu_t | \mathcal{F}_{t-1}), \tag{7}$$

1 An alternative modeling strategy for the exceedance probability is based on the Poisson approximation.

where $f(\phi_t|\mathcal{F}_{t-1})$ and $f(\xi_t, \nu_t|\mathcal{F}_{t-1})$ are, respectively, the density of the model for the exceedance rate and the density of the GP distribution. The joint likelihood can thus be separated into two components that can be maximized separately,

$$\mathcal{L}(\phi, \boldsymbol{\xi}, \boldsymbol{\nu}) = \left\{ f(\phi_1) \prod_{t=2}^{T} f(\phi_t|\mathcal{F}_{t-1}) \right\} \left\{ f(\xi_1, \nu_1) \prod_{t=2}^{T} f(\xi_t, \nu_t|\mathcal{F}_{t-1}) \right\}, \tag{8}$$

leading to the sequence of ML estimators $\widehat{\phi}$, $\widehat{\boldsymbol{\xi}}$, and $\widehat{\boldsymbol{\nu}}$. The validity of the assumptions of conditional independence of rates of exceedances, sizes of exceedances, and joint conditional independence of rate and size of exceedances is assessed empirically in Section 4.4 and we find that the assumptions are justified.

2 The *Realized* Peaks over Threshold Approach

In order to model the tail of the conditional return distribution, Chavez-Demoulin, Davison, and McNeil (2005) and Chavez-Demoulin, Embrechts, and Sardy (2014) develop, respectively, fully parametric and non-parametric extensions of the POT approach, considering $\mathcal{F}_t$ to be the information set generated by the daily price path. We propose to augment the available information set with HF data and model daily extremes with measures built upon these data. We denote this new information set by $\mathcal{H}_t$, where $\mathcal{F}_t \subset \mathcal{H}_t$.

Realized measures are non-parametric estimators of the variation of the price path of an asset. They ignore the variation of prices overnight and sometimes the variation in the first few minutes of the trading day when recorded prices may contain large errors. A good background for realized measures can be found in the survey articles by Barndorff-Nielsen and Shephard (2007) and Andersen et al. (2009).

As our approach includes realized measures as covariates in the models for the exceedances, throughout we refer to it as RPOT. In what follows, we discuss the two components of the conditional likelihood in Equation (8) individually and show how realized measures can be incorporated in each. We present the model for the exceedance rate, then the one for excess size. In both cases, we rely on parametric specifications of the dynamic parameters and base the inference on standard asymptotic arguments from ML theory.

2.1 Modeling the Exceedance Rate

We start following Davison and Smith (1990) who propose combining the approach for stationary data with regression modeling. We model the time-varying exceedance probability with a Logit function,[2]

$$\phi_t = \frac{1}{1 + \exp\left(\varphi_0 + \varphi_1 \mathrm{RM}_{t-1}\right)}, \tag{9}$$

where RM_{t-1} is a generic realized measure. More generally, p realized measures can be included in a $p+1$ vector $\mathbf{RM}_t$ of regressors at time t. The parameters $\boldsymbol{\varphi} = [\varphi_0, \varphi_1, \ldots, \varphi_p]$

2 An alternative approach to obtain an estimate of the time-varying exceedance rate is to model the number of exceedances $N_u(t)$ with a non-homogeneous Poisson process (Chavez-Demoulin, Davison, and McNeil, 2005).

are easily estimated through maximization of the likelihood function

$$\mathcal{L}(\boldsymbol{\varphi}; I_t, \mathbf{RM}_t) = \prod_{t=l+1}^{T} \left(\exp\left(\mathbf{RM}_t'\boldsymbol{\varphi}\right)\right)^{I_t} \frac{1}{1+\exp\left(\mathbf{RM}_t'\boldsymbol{\varphi}\right)},$$

where l is the lag at which realized measures in $\mathbf{RM}_t'$ become available,[3] and I_t takes value 1 if $t \in \mathcal{E} = \{t \in \{1,\ldots,T\} | Y_t > u\}$ and zero otherwise. ML estimates of the parameters are obtained through numerical techniques and asymptotic results follow from the standard arguments of Newey and McFadden (1994).

2.2 Modeling the Excess Size

The natural model for the excesses of a high threshold is the GP distribution. In the case of a non-stationary process, Davison and Smith (1990) suggest adding linear covariates in the scale ν_t and shape ξ_t parameters. We model ν_t as a log-linear function of the realized measure, that is, with one covariate we have

$$\nu_t(\kappa_0, \kappa_1) = \exp\left(\kappa_0 + \kappa_1 \mathrm{RM}_{t-1}\right). \tag{10}$$

We keep ξ constant to gain stability as in Chavez-Demoulin, Davison, and McNeil (2005; Chavez-Demoulin, Embrechts, and Sardy, 2014), and check the empirical validity of this assumption in Section 4.2. Define the exceedance $Z_t = Y_t - u$. For p realized measures the dynamic GP distribution has the functional form,

$$G_t(z_t; \mathbf{RM}_t, \boldsymbol{\kappa}, \xi) = \Pr(Y_t - u < z_t | Y_t > u, \mathcal{H}_{t-1}) = 1 - \left(1 + \frac{\xi}{\exp\left(\mathbf{RM}_t'\boldsymbol{\kappa}\right)} z_t\right)^{-1/\xi}, \tag{11}$$

with $z_t \geq 0$ when $\xi > 0$ and $0 \leq z \leq -\exp\left(\mathbf{RM}_t'\boldsymbol{\kappa}\right)/\xi$ when $\xi < 0$. Moreover, the functional form is different if $\xi = 0$ zero, see Equation (1). The corresponding likelihood function can be written as

$$\mathcal{L}(\boldsymbol{\kappa}, \xi; z_t, \mathbf{RM}_t) = \prod_{t=l+1}^{T} \left(\frac{1}{\exp\left(\mathbf{RM}_t'\boldsymbol{\kappa}\right)} \left[1 + \frac{\xi}{\exp\left(\mathbf{RM}_t'\boldsymbol{\kappa}\right)} z_t\right]_+^{-1/\xi-1}\right)^{I_t}, \tag{12}$$

where $\boldsymbol{\kappa} = [\kappa_0, \kappa_1, \ldots, \kappa_p]$ and ξ are the parameters that must be estimated. ML estimates of the parameters are obtained through numerical techniques.

Hall and Tajvidi (2000) and Beirlant and Goegebeur (2004) establish the asymptotic normality of several semi-parametric classes of estimators $\widehat{\xi}$ and $\widehat{\nu}$, both when the true conditional distribution is GP and when the conditional distribution of the excesses converges to the GP. In the following proposition, we establish the consistency and asymptotic normality of our ML estimators under the assumption that the exceedances $(z_1, \ldots, z_k)$ are independent with distribution (11) and $\xi > 0$. We relegate the asymptotic properties for the exponential case ($\xi = 0$) to the Supplementary Material, as this requires changing the functional form of the distribution in Equation (11). Note that these are the relevant cases with financial data, as the case $\xi < 0$ implies that the log-losses have finite support. Despite

3 In the simplest case, $\mathbf{RM}_t = [1, RM_{t-1}]$ as in Equation (9). When the regressors are functions of past observations, it may require several lags before all are available. For example, for a sequence of observations $y_1, \ldots, y_T$, if we use a 30-day moving average of the y_t's as a regressor, then the first observation will be available at $l = 30$.

this, throughout the analysis we fit the GP model without constraining $\xi \geq 0$, as this allows a better fit in finite samples (Cohen, 1982). Note that the randomness of the binary variable I_t may have some effects on the asymptotic variance of the ML estimator, but they are not considered in Proposition 1.

Proposition 1. *Let* $\boldsymbol{\theta} = (\boldsymbol{\kappa}, \xi) \in \Theta \subset \mathbb{R}^{p+1} \times$ *(0, 1) and* $(z_1, \ldots, z_k)$ *be an independent sequence. Define the ML estimator* $\widehat{\boldsymbol{\theta}}_k$ *as*

$$\widehat{\boldsymbol{\theta}}_k = \arg \max_{\boldsymbol{\theta} \in \Theta} \log \mathcal{L}(\boldsymbol{\theta}; z_t, \mathbf{RM}_t),$$

where $\mathcal{L}(\boldsymbol{\theta}; z_t, \mathbf{RM}_t)$ *is given in* Equation (12). *Assuming that (i) the true vector of parameters* $\boldsymbol{\theta}_0$ *is identifiable and interior to the parameter space* Θ *which is compact, (ii)* $\mathbb{E}[\mathbf{RM}_t \mathbf{RM}_t']$ *is positive definite, and (iii)* $\mathbb{E}\big[C \| \exp(\mathbf{RM}_t) \|\big] < \infty$ *for any* $C > 0$, *then*

$$\sqrt{k}\left(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}_0\right) \xrightarrow{d} N(0, J^{-1}),$$

where $J = \mathbb{E}\big[\nabla_{\boldsymbol{\theta}} \log g_t(z_t, \mathbf{RM}_t; \boldsymbol{\theta}_0) \nabla_{\boldsymbol{\theta}} \log g_t(z_t, \mathbf{RM}_t; \boldsymbol{\theta}_0)'\big]$ *with* $g_t(z_t, \mathbf{RM}_t; \boldsymbol{\theta})$ *the density of the dynamic GP in* Equation (11).

Proof. See Appendix. □

We derive the asymptotic properties of ML estimators under the assumption of independent extremes. To make sure that the estimated standard errors in the empirical analysis are not biased by any dependence left in the observations after conditioning, we use a robust "sandwich" estimator as in Chavez-Demoulin and Davison (2005). In Section 4.4, we verify empirically that the independence assumption holds, and find that robust estimates are almost indistinguishable from standard ones.

2.3 Estimation of the Conditional Risk Measures

Quantile-based risk measures such as the Value-at-Risk (VaR) and the Expected Shortfall (ES) are central tools for quantitative risk management in the financial industry. Denote by $L_t(x) = -F_t(x)$ the *loss distribution* at time t. The one-day-ahead VaR and ES at level α are, respectively, defined[4] as $\mathrm{VaR}^{\alpha}_{t|t-1} = \inf\{x \in \mathbb{R} : L_{t|t-1}(x) \geq 1 - \alpha\}$ and $\mathrm{ES}^{\alpha}_{t|t-1} = \frac{1}{\alpha}\int_0^{\alpha} \mathrm{VaR}^{\gamma}_{t|t-1} \mathrm{d}\gamma$, where $L_{t|t-1}(x)$ denotes the loss cumulative distribution function conditional on the information at time $t-1$. To obtain a forecast of the VaR and ES with the RPOT approach, it is necessary to invert Equation (3) and plug-in the forecast of the parameters obtained for the loss distribution, that is,

$$\widehat{\mathrm{VaR}}^{\alpha}_{t|t-1} = \widehat{u} + \frac{\widehat{\nu}_{t|t-1}}{\widehat{\xi}}\left[\left(\frac{\widehat{\phi}_{t|t-1}}{\alpha}\right)^{\widehat{\xi}} - 1\right], \tag{13}$$

$$\widehat{\mathrm{ES}}^{\alpha}_{t|t-1} = \frac{\widehat{\mathrm{VaR}}^{\alpha}_{t|t-1}}{1 - \widehat{\xi}} + \frac{\widehat{\nu}_{t|t-1} - \widehat{\xi}\widehat{u}}{1 - \widehat{\xi}}, \tag{14}$$

4 In Section 4, we compute the VaR and ES at level $\alpha = 0.01$. They are often called the 99%-VaR and the 99%-ES in the literature.

where $\widehat{\nu}_{t|t-1}$ and $\widehat{\xi}$ are the estimates from the dynamic GP distribution and $\widehat{\phi}_{t|t-1}$ is the estimate from the threshold exceedance model: $\widehat{\phi}_{t|t-1} = 1/\left(1+\exp\left(\widehat{\varphi_0}+\widehat{\varphi_1}\mathrm{RM}_{t-1}\right)\right)$ for the model in Equation (9).

$\widehat{\mathrm{VaR}}^{\alpha}_{t|t-1}$ and $\widehat{\mathrm{ES}}^{\alpha}_{t|t-1}$ are point forecasts. To obtain the confidence interval estimates, we use a post-blackened bootstrap (Davison and Hinkley, 1997). First, we fit a RPOT model to the observations and obtain the residuals $R_j = \widehat{\xi}_j^{-1}\log\left(1+\widehat{\xi}_j\left(Z_j-\widehat{u}\right)/\widehat{\nu}_j\right)$ for each exceedance $Z_j, j\in\mathcal{E}$. Then, we obtain B bootstrap samples of the residuals R_j and apply the RPOT model to each sample to obtain a percentile-based confidence interval of the VaR and ES.

3 Simulations

We perform Monte Carlo simulations with the purpose of: (i) assessing the finite sample properties of the ML estimator of both the Logit model for the exceedance rate and the dynamic GP model for the size of the exceedances; (ii) evaluating the impact of microstructure noise on the ML estimators and (iii) assessing the ability of the RPOT model to estimate tail quantiles when observations are generated from a general return process.

3.1 Finite Sample Properties and Microstructure Noise

Let σ_t^2 be the return variance on day t and E_t a random variable taking a value in $\mathbb{R}_+$ if an exceedance occurs and 0 otherwise. A random sample of exceedances is generated according to the following system of equations:

$$\begin{aligned}
\log\sigma_t^2 &= \beta_0+\beta_d\log\sigma_{t-1}^2+\beta_w\log\tilde{\sigma}_{t-1:t-5}^2+\beta_m\log\tilde{\sigma}_{t-1:t-22}^2+\epsilon_t,\\
\mathrm{logit}(\phi_t) &= \varphi_0+\varphi_1\log\sigma_t^2,\\
\log\nu_t &= \kappa_0+\kappa_1\log\sigma_t^2,\\
E_t &= W_t\mathbb{1}_{\{U_t<\phi_t\}},
\end{aligned} \tag{15}$$

where $\epsilon_t\sim N\left(0,\sigma_\epsilon^2\right)$, $W_t\sim\mathrm{GP}(\nu_t,\xi)$, $U_t\sim\mathrm{Unif}(0,1)$ and $\log\tilde{\sigma}_{t-1:t-s}^2=\frac{1}{s-1}\sum_{i=1}^{s}\log\sigma_{t-i}^2$.

The variance process follows a stochastic volatility model akin to the HAR model of Corsi (2009) which can replicate the typical persistence pattern of the volatility. The parameters of this model are set in order to obtain realistic dynamics of the variance. The parameters of the extreme value models are similar to empirical estimates obtained in Section 4 with a threshold at the 90th quantile. We generate $B=500$ samples of $T=2000$ observations and estimate the Logit and GP model parameters with the ML estimators set forth in Section 2.

Table 1 reports the results using as covariate the true volatility and the realized measure, defined here as σ_t^2 and $\mathrm{RM}_t=\sigma_t^2+\eta_t$, where $\eta_t\sim N\left(\mu_{\eta,t},\sigma_{\eta,t}^2\right)$, respectively. The random variable η_t reflects the uncertainty of the realized measure with respect to the true variance. In Table 1, we assume that RM_t is unbiased, $\mu_{\eta,t}=0$, but subject to uncertainty, $\sigma_{\eta,t}^2=0.15\sigma_t^2$. The results show that the estimates obtained using the realized measure instead of the true variance are only slightly affected by the uncertainty characterizing the observed proxy.

Table 2 reports the results adding microstructure noise to the realized measure. The presence of microstructure noise induces an upward bias in the realized measure. We inflate

Table 1 Simulation with realized measure versus true volatility

	True variance					Realized measure				
	ϕ_0	ϕ_1	κ_0	κ_1	ξ	ϕ_0	ϕ_1	κ_0	κ_1	ξ
True	9.460	0.840	−2.270	0.310	0.100	9.460	0.840	−2.270	0.310	0.100
Mean	9.455	0.839	−2.286	0.308	0.099	9.294	0.818	−2.352	0.300	0.101
Std. Dev.	0.791	0.090	0.222	0.027	0.026	0.871	0.100	0.215	0.026	0.026
Mean ASE	0.802	0.091	0.223	0.027	0.026	0.796	0.091	0.222	0.027	0.025

Notes: True parameter values along with mean of estimates (Mean), mean of standard errors based on plug-in estimate of the asymptotic covariance matrix (Mean ASE), and standard deviation of the estimates (Std.dev.) over $B = 500$ replications.

Table 2 Simulation with microstructure noise

	Constant noise					Time-varying noise				
	ϕ_0	ϕ_1	κ_0	κ_1	ξ	ϕ_0	ϕ_1	κ_0	κ_1	ξ
True	9.460	0.840	−2.270	0.310	0.100	9.460	0.840	−2.270	0.310	0.100
Mean	13.660	1.435	−1.431	0.438	0.102	9.171	0.831	−2.374	0.307	0.099
Std. dev.	1.822	0.231	0.334	0.044	0.025	0.793	0.093	0.223	0.028	0.027
Mean ASE	1.537	0.191	0.308	0.040	0.025	0.775	0.091	0.215	0.027	0.025

Notes: True parameter values along with mean of estimates (Mean), mean of standard errors based on plug-in estimate of the asymptotic covariance matrix (Mean ASE), and standard deviation of the estimates (Std. dev.) over $B = 500$ replications.

RV_t by fixing $\mu_{\eta,t}$ to a positive value. In particular, in Table 2 we consider two types of microstructure noise: constant noise, $\mu_{\eta,t} = 0.3\mathbb{E}(\sigma_t^2)$, and a time-varying noise, $\mu_{\eta,t} = 0.3\sigma_t^2$. The results suggest that the two types impact the estimated parameters differently: when the noise is time-varying, only the constants (ϕ_0, κ_0) are affected by the microstructure noise, while the parameters capturing the dynamics (ϕ_1, κ_1) continue to be well estimated; in contrast, a constant noise has a deleterious effect on all the parameters.

In conclusion, as long as the dynamic of the extremes depends on the latent variance process, a good proxy of the latter is able to unveil it. The effects of microstructure noise can be mitigated by careful selection of the realized measure. In the empirical analysis, we assess whether the microstructure noise is an issue considering several robust realized measures.

3.2 RPOT under Misspecification

We assume that the returns r_t are generated according to the following t-HAR process,

$$r_t = \sigma_t \epsilon_t,$$
$$\log \sigma_t^2 = \beta_0 + \beta_D \log \sigma_{t-1}^2 + \beta_W \log \sigma_{t-1:t-5}^2 + \beta_M \log \sigma_{t-1:t-22}^2 + \eta_t,$$

where $\eta_t \sim N\left(0, \sigma_\eta^2\right)$ and $\epsilon_t \sim t(10)$, with $t(\nu)$ denoting the Student's t with ν degrees of freedom. We found informative to consider two scenarios: high volatility-of-volatility

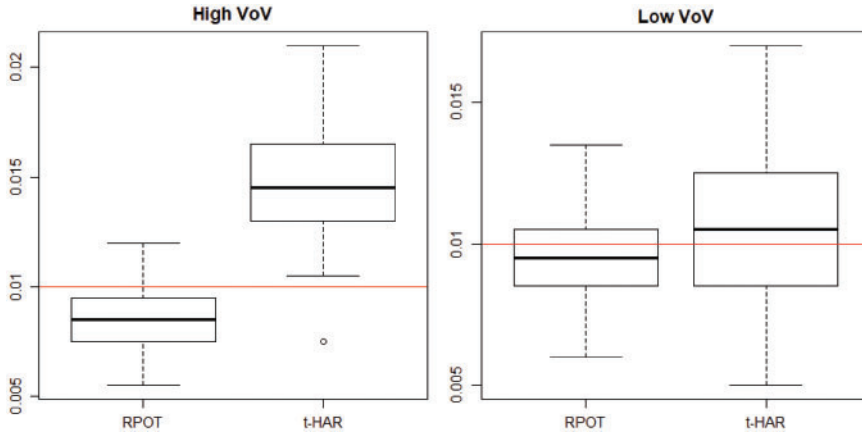

Figure 2 Simulation under misspecification. Proportion of exceedances of the estimated quantiles under high volatility-of-volatility and low volatility-of-volatility scenarios. Results from $B = 100$ replications are shown. The horizontal line shows the correct $\alpha = 0.01$ level.

($\sigma_\eta^2 = 0.56$) and low volatility-of-volatility ($\sigma_\eta^2 = 0.18$). For each scenario, we generate $B = 100$ samples of $T = 2000$ observations. We perform model estimation of the *t*-HAR model and RPOT model using σ_t^2 to drive the dynamic of the exceedance probability and the size of the exceedances. We estimate the conditional tail quantiles at level α for both models, and compute the proportion of times that the estimated quantiles are exceeded. If the model is correctly specified, the latter proportion should be α. Figure 2 reports the proportion of exceedances of the estimated quantiles at level $\alpha = 0.01$ over the $B = 100$ replications. These results confirm that the RPOT model is able to recover the tail quantiles under a general *t*-HAR process. When the volatility-of-volatility is high, the simple log-linear volatility model is unable to accurately capture the high variation in the tail, leading to many over-rejections, while the RPOT is somewhat conservative. When the volatility-of-volatility is low, both RPOT and *t*-HAR yield the correct average number of exceedances, with RPOT results being less variable.

4 Empirical Analysis

4.1 Data Description

The empirical analysis is based on the S&P 500 index from January 1, 2000 to December 31, 2014. The data come from the Oxford-Man Institute "Realised Library" version 0.2[5] (Heber et al., 2009). We consider open-to-close returns and overnight returns are not included in the analysis.

We consider three 5-year sub-samples: 2000–2004, 2005–2009, and 2010–2014. This allows us to discuss the fitted models in three different regimes of the stock market. The first sub-sample contains the downward trend that followed the dot-com bubble explosion,

5 This dataset provides a large number of realized measures, but does not include certain interesting measures that could be useful predictors of the extreme dynamics. We leave exploration of the informational content of such measures to future research.

and part of the subsequent rebound. The second sub-sample contains the 2008 crash after several banks and insurance companies went bankrupt. The last sub-sample contains the recovery and the bullish trend of the most recent years.

We start the analysis with the most common realized measure, that is, the 5-min RV (RV_t), see Andersen et al. (2001). We assess how various lags of this proxy affect the dynamic of the extremes. We then use the 5-min bipower variation (BV_t) of Barndorff-Nielsen and Shephard (2004) to disentangle the contribution of the jumps to the realized variation from that of the Brownian path. We obtain the realized jumps measure as $J_t = \max(RV_t - BV_t, 0)$, where the censoring is needed to avoid negative values due to finite sample effects. We also assess the effect of microstructure noise by considering realized measures sampled at a frequency lower than 5-min and noise-robust realized measures. Finally, we assess whether other market frictions can affect the analysis by considering two proxies for illiquidity.

The sequence of extremes is obtained by fixing a threshold level u. The choice of the threshold is important, implying a balance between bias and variance. On the one hand, a smaller u means more observations used for inference. On the other hand, probability theory suggests choosing a larger u for limiting results to apply. We use graphical analysis and tests to accurately choose the threshold when fitting a model for the extremes. Throughout the analysis, unless otherwise stated, we consider a threshold level corresponding to the 90th quantile of the loss distribution.

4.2 Model Specification

This section is devoted to the estimation of different specifications of the models presented in Section 2. For each model, we consider four different specifications with sets of covariates of increasing size. The sets of covariates at time t are based on the following realized measures: RV_t, BV_t, J_t, $\overline{RV}_t^W = \frac{1}{4}\sum_{i=1}^{4} RV_{t-i}$, and $\overline{RV}_t^M = \frac{1}{17}\sum_{i=5}^{21} RV_{t-i}$.

The first set contains the RV, and it constitutes the baseline model to assess the relationship between extremes and the asset price variation. The second set contains the bipower variation and the realized jumps. This choice follows the argument of Andersen, Bollerslev, and Diebold (2007), who suggest that distinguishing between the information coming from the continuous and discontinuous sample paths could be valuable. The third and fourth sets can be considered extensions where we add the information coming from the average weekly and monthly RV. This gives a HAR structure (Corsi, 2009) that allows us to see whether realized measures at different time-horizons are useful to predict the behavior of the extremes. We consider the logarithmic transformation of these realized measures as it is preferable from a modeling perspective.[6]

The threshold level defining the sequence of extremes is set at the 90th quantile of the loss distribution in the first two sub-samples (2000–2004 and 2005–2009). The third sub-sample (2010–2014) presents tail decay close to exponential that requires increasing the threshold level to obtain a good fit. Over this period, we set the threshold at the 97th, 98th, 96th, and 96th quantile for the first, second, third, and fourth specification, respectively. For each specification in the third sub-sample, we use a different threshold, the lowest one to yield a good fit.

6 In Section 4, we scale the jump component in order to have covariates of similar magnitude. Indeed, while $\log(RV_t)$ is much greater, in absolute value, than RV_t, $\log(1 + J_t) \approx J_t$, with the consequence that the parameter associated to the latter would be huge.

Table 3 Fitted Logit models

	φ_0	φ_1	φ_2	φ_3	φ_4	φ_5
			2000–2004			
I	5.46***	0.84***				
	(0.96)	(0.10)				
II	6.78***		0.97***	−0.07		
	(1.12)		(0.12)	(0.08)		
III	8.15***	0.38*			0.52*	0.25
	(1.34)	(0.18)			(0.24)	(0.23)
IV	9.07***		0.62***	−0.10	0.37	0.24
	(1.44)		(0.18)	(0.09)	(0.25)	(0.23)
			2005–2009			
I	5.02***	0.79***				
	(0.68)	(0.08)				
II	5.13***		0.78***	0.00		
	(0.78)		(0.08)	(0.07)		
III	5.85***	0.30			0.30	0.29
	(0.74)	(0.17)			(0.22)	(0.17)
IV	5.86***		0.27	0.00	0.31	0.29
	(0.81)		(0.18)	(0.07)	(0.22)	(0.17)
			2010–2014			
I	5.63***	0.96***				
	(1.13)	(0.12)				
II	10.28***		1.46***	−0.30		
	(1.75)		(0.19)	(0.19)		
III	7.29***	0.52**			0.21	0.36
	(1.56)	(0.20)			(0.22)	(0.23)
IV	8.88***		0.64**	−0.11	0.21	0.39
	(1.74)		(0.22)	(0.12)	(0.23)	(0.22)

Notes: Specifications are:
I $\text{logit}(\phi_t) = \varphi_0 + \varphi_1 \log \text{RV}_{t-1}$
II $\text{logit}(\phi_t) = \varphi_0 + \varphi_2 \log \text{BV}_{t-1} + \varphi_3 \log (1 + J_{t-1})$
III $\text{logit}(\phi_t) = \varphi_0 + \varphi_1 \log \text{RV}_{t-1} + \varphi_4 \log \overline{\text{RV}}^{W}_{t-1} + \varphi_5 \log \overline{\text{RV}}^{M}_{t-1}$
IV $\text{logit}(\phi_t) = \varphi_0 + \varphi_2 \log \text{BV}_{t-1} + \varphi_3 \log (1 + J_{t-1}) + \varphi_4 \log \overline{\text{RV}}^{W}_{t-1} + \varphi_5 \log \overline{\text{RV}}^{M}_{t-1}$
on S&P 500 returns. *,**,***Significance at the 5%, 1%, and 0.1% levels, respectively.

Table 3 reports the estimated parameters for the Logit model in Equation (9) and more elaborate specifications. Results for specification I show that the coefficient for RV is strongly significant across the three windows. This substantial statistical evidence confirms the usefulness of HF data for modeling the exceedance rate. Decomposing the RV in the contribution from the continuous and the discontinuous sample paths in specification II allows us to see that the jump component plays a negligible role on the exceedance rate. In particular, while the coefficient for BV is strongly significant across the windows, the jump coefficient is never significant. Finally, information on the variance at further lags in a

Table 4 Fitted dynamic GP models

	κ_0	κ_1	κ_2	κ_3	κ_4	κ_5	ξ
				2000–2004			
I	−2.27	0.31**					0.02
	(1.19)	(0.14)					(0.09)
II	−1.84		0.35**	−0.04			0.01
	(1.21)		(0.14)	(0.08)			(0.09)
III	−0.37	0.05			0.33	0.15	0.00
	(1.46)	(0.15)			(0.20)	(0.23)	(0.09)
IV	0.82		0.09	−0.06	0.38	0.19	0.00
	(1.54)		(0.15)	(0.08)	(0.20)	(0.23)	(0.09)
				2005–2009			
I	−0.95	0.42***					0.00
	(0.54)	(0.07)					(0.08)
II	−0.83		0.42***	−0.01			−0.02
	(0.62)		(0.08)	(0.03)			(0.07)
III	−0.51	0.06			0.48**	−0.07	−0.10
	(0.56)	(0.12)			(0.18)	(0.14)	(0.07)
IV	0.03		0.11	0.22	0.19	0.07	−0.86
	(0.46)		(0.20)	(1.58)	(0.36)	(0.51)	(0.73)
				2010–2014			
I	1.24	0.68***					−0.17
	(1.33)	(0.15)					(0.16)
II	4.89		1.09***	−0.42			−0.14
	(2.99)		(0.35)	(0.22)			(0.25)
III	0.92**	0.23**			0.75***	0.39***	−0.70
	(0.17)	(0.09)			(0.09)	(0.06)	(0.43)
IV	2.18		0.43*	−0.13	0.63*	−0.32	−0.60
	(1.30)		(0.20)	(0.10)	(0.27)	(0.24)	(0.91)

Notes: The parameter ξ is constant, while ν_t is allowed to vary according to
I $\log \nu_t = \kappa_0 + \kappa_1 \log \mathrm{RV}_{t-1}$
II $\log \nu_t = \kappa_0 + \kappa_2 \log \mathrm{BV}_{t-1} + \kappa_3 \log (1 + J_{t-1})$
III $\log \nu_t = \kappa_0 + \kappa_1 \log \mathrm{RV}_{t-1} + \kappa_4 \log \overline{\mathrm{RV}}^{W}_{t-1} + \kappa_5 \log \overline{\mathrm{RV}}^{M}_{t-1}$
IV $\log \nu_t = \kappa_0 + \kappa_2 \log \mathrm{BV}_{t-1} + \kappa_3 \log (1 + J_{t-1}) + \kappa_4 \log \overline{\mathrm{RV}}^{W}_{t-1} + \kappa_5 \log \overline{\mathrm{RV}}^{M}_{t-1}$
*,**,***Significance at the 5%, 1%, and 0.1% levels, respectively.

HAR fashion does not seem to add significantly to the prediction of the exceedance rate. Indeed, in specifications III and IV, the coefficients for $\overline{\mathrm{RV}}^{M}$ are never significant and those for $\overline{\mathrm{RV}}^{W}$ are significant on only one occasion.

Table 4 shows the estimated parameters of the dynamic GP distribution with ν_t as in Equation (10) and more elaborate specifications. The inclusion of HF data contributes significantly to explain the size of the excesses. The coefficients for RV and BV are positive and strongly significant, though adding further lags of the RV reduces their significance. $\overline{\mathrm{RV}}^{W}$ is a significant predictor in the third period and once in the second period, while $\overline{\mathrm{RV}}^{M}$ is

significant only once in the third period. This suggests that the inclusion of additional lags of realized variation can improve the fit of the model in some cases. The coefficient for the jump component J is never significant. One might find this counterintuitive, as jumps are typically found to contribute substantially to explain the tails of the return distribution (Bollerslev and Todorov, 2011). However, it is important to distinguish the explanatory power from the forecasting power. We consider the jump variation as a predictive component of future extremes. Repeating the analysis using J_t instead of J_{t-1}, we find that the jump variation significantly contributes in explaining the size of daily extremes (results in Supplementary Material).

As for the parameter ξ in Table 4, it is very close to zero or not significantly different from zero. As we are trying to estimate the tail of the conditional return distribution, we do not expect the parameter ξ to be strongly positive as it is usually observed for the unconditional return distribution.

In an option pricing setting, Bollerslev and Todorov (2014) find that the tail index of the risk-neutral distribution presents time variation. Similarly, using a panel structure, Kelly and Jiang (2014) find that the tail of a market-wide factor presents time variation. We inspect whether assuming a constant tail index ξ is a sensible decision, and fit a GP distribution where we allow ξ_t to change linearly with RV_t. Results in the Supplementary Material show that a constant ξ is justified.

Overall, it appears that HF data apport a meaningful contribution to understanding the behavior of the excesses. While the 1-day lagged realized variation is useful for modeling both the exceedance rate and the size of excesses, the contributions of jumps and past volatilities are limited.

4.3 A Comparison of Realized Measures

In this section, we question whether the lack of efficiency of the realized measure and the microstructure noise can have an impact on the explanatory power of the realized measure on the dynamics of the extremes. We estimate the Logit and the dynamic GP models with one lagged realized measure as covariate. We consider the 5-min (already shown in Tables 3–4, but displayed again here for ease of comparison) and 10-min RV, respectively, RV_t and RV10_t, and the corresponding sub-sampled version of these two measures, RVss_t and RV10ss_t. Sub-sampling provides a simple way to obtain more efficient estimators of the integrated variance, though they are still biased in the presence of microstructure noise (Zhang, Mykland, and Aït-Sahalia, 2005). Finally, we consider the realized kernel RK_t of Barndorff-Nielsen et al. (2008) which is robust to microstructure noise.

Tables 5–6 report the estimates for the Logit and the dynamic GP models, respectively. In both cases, the estimated parameters vary only slightly across the different realized measures, suggesting that the microstructure noise is not a concern in explaining the extremes dynamic. This conclusion is supported by the log-likelihood plots in Figure 3, where an overall winner does not appear.

4.4 Model Diagnostics

In this section, we make use of graphical methods and formal testing procedures to assess the validity of the models fitted in Section 4.2.

First, we perform a deviance χ^2-test on the residuals of the Logit model, a standard goodness-of-fit test in Logit regression literature (Hosmer and Lemeshow, 2004). Given a

Table 5 Logit models with other HF measures

	2000–2004		2005–2009		2010–2014	
RM_{t-1}	φ_0	φ_1	φ_0	φ_1	φ_0	φ_1
$\log RV_{t-1}$	5.46***	0.84***	5.02***	0.79***	5.62***	0.96***
	(0.96)	(0.11)	(0.68)	(0.08)	(1.14)	(0.12)
$\log RVss_{t-1}$	6.39***	0.93***	5.28***	0.80***	7.33***	1.11***
	(1.03)	(0.11)	(0.71)	(0.08)	(1.35)	(0.14)
$\log RV10_{t-1}$	4.63***	0.75***	4.84***	0.77***	5.77***	0.97***
	(0.90)	(0.10)	(0.66)	(0.07)	(1.10)	(0.11)
$\log RV10ss_{t-1}$	6.09***	0.90***	5.24***	0.80***	7.66***	1.13***
	(1.01)	(0.11)	(0.71)	(0.08)	(1.41)	(0.14)
$\log RK_{t-1}$	5.96***	0.90***	5.13***	0.80***	6.52***	1.04***
	(0.99)	(0.11)	(0.69)	(0.08)	(1.26)	(0.14)

Notes: We consider the specification $\text{logit}(\phi_t) = \varphi_0 + \varphi_1 \log RV_{t-1}$. *,**,***Significance at the 5%, 1%, and 0.1% levels, respectively.

Table 6 Dynamic GP model with other HF measures

	2000–2004			2005–2009			2010–2014		
RM_{t-1}	κ_0	κ_1	ξ	κ_0	κ_1	ξ	κ_0	κ_1	ξ
$\log RV_{t-1}$	−2.27	0.31*	0.02	−0.95	0.42***	0.00	1.24	0.68***	−0.17
	(1.19)	(0.14)	(0.09)	(0.54)	(0.07)	(0.08)	(1.33)	(0.15)	(0.16)
$\log RVss_{t-1}$	−2.18	0.31*	0.03	−0.85	0.42***	−0.01	1.64	0.71***	−0.27
	(1.29)	(0.15)	(0.09)	(0.57)	(0.07)	(0.08)	(1.13)	(0.13)	(0.15)
$\log RV10_{t-1}$	−2.54*	0.28*	0.04	−0.89	0.43***	0.01	0.87	0.64***	−0.21
	(1.15)	(0.14)	(0.09)	(0.54)	(0.06)	(0.08)	(1.23)	(0.15)	(0.17)
$\log RV10ss_{t-1}$	−2.26	0.30*	0.03	−0.96	0.41***	0.00	1.63	0.71***	−0.26
	(1.26)	(0.15)	(0.09)	(0.58)	(0.07)	(0.08)	(1.04)	(0.13)	(0.16)
$\log RK_{t-1}$	−2.28	0.31*	0.03	−0.83	0.43***	−0.01	1.24	0.68***	−0.23
	(1.27)	(0.15)	(0.09)	(0.54)	(0.07)	(0.08)	(1.14)	(0.13)	(0.16)

Notes: The parameter ξ is constant, while ν_t varies according to $\log \nu_t = \kappa_0 + \kappa_1 \log RV_{t-1}$. *,**,***Significance at the 5%, 1%, and 0.1% levels, respectively.

model $\mathcal{M}_0$ and a sequence of observations y, the deviance is defined as $D(y, \mathcal{M}_0) = -2(\log \Pr(y|\widehat{\theta}_0) - \log \Pr(y|\widehat{\theta}_s))$, where $\widehat{\theta}_0$ denotes the fitted values of the parameters in the model $\mathcal{M}_0$ and $\widehat{\theta}_s$ denotes the fitted parameters for the saturated model, that is a model with a parameter for every observation. Two measures of deviance are particularly important in a Logit model: the *null deviance* which represents the deviance for a model with only the intercept, and the *model deviance* representing the deviance of the fitted model. To evaluate the contribution of the predictors, one can subtract the model deviance from the null deviance, that is, $D_{\text{null}} - D_{\text{fitted}} = -2\left(\log \Pr\left(y|\widehat{\theta}_n\right) - \log \Pr\left(y|\widehat{\theta}_0\right)\right)$ and assess the difference on a χ^2-distribution with degrees of freedom equal to the difference in the number of estimated parameters. If the model deviance is significantly smaller

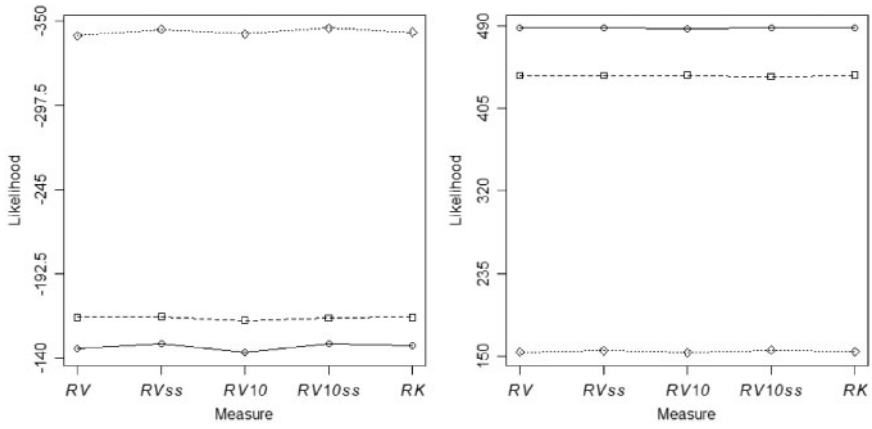

Figure 3 Likelihood for several realized measures. Logit likelihood (left panel) and dynamic GP likelihood (right panel). The three sub-samples are: 2000–2004 (solid), 2005–2009 (dotted), and 2010–2014 (dashed).

than the null deviance then one can conclude that the predictors significantly improve the model fit.

Another commonly used goodness-of-fit test in Logit regression is the Hosmer–Lemeshow test (Hosmer and Lemeshow, 2004). In our case, it tests the null hypothesis of equality between the observed frequency of exceedances and that expected from the fitted model. For each observation y_i in the sample, the predicted probability π_i of exceeding the threshold is computed. Then, the y_i's are split into G groups of size N_g according to the rank of their predicted probabilities, with $g \in \{1, \ldots, G\}$. Finally, for each group g, the average predicted probability $\pi_g = (N_g)^{-1} \sum_{i_g=1}^{N_g} y_{i_g}$ and the expected number of exceedances $E_g = N_g \pi_g (1 - \pi_g)$ are computed. The test statistic under H_0 is a Pearson χ^2-statistic of the form $H = \sum_{g=1}^{G} \frac{(O_g - E_g)^2}{E_g}$ where O_g is the number of observed exceedances in the gth group. The test statistic follows asymptotically a χ^2-distribution with $G-2$ degrees of freedom. As there are no specific rules to choose the number of groups, we set $G = 10$ as it is usual in this literature, confident that our large sample size leads to enough observations in every decile.

Table 7 reports the p-values for both the deviance χ^2 and the Hosmer–Lemeshow tests. In the first test, the null hypothesis of equal explanatory power between the null and the fitted model is rejected for all the specifications, across the three windows. At the same time, the null hypothesis of the Hosmer–Lemeshow is rejected on only one occasion, suggesting that in general the observed exceedance rate and that implied by the model do not differ significantly. These results indicate a good fit of the Logit model across the different windows.

To assess the goodness of fit of the dynamic GP distribution as a model for the size of the excesses, we perform a graphical validation as in Coles (2001). When data are assumed to be identically distributed, goodness of fit can be evaluated by means of a qq-plot or formal testing as in Choulakian and Stephens (2001). However, the lack of homogeneity among observations means that some modifications are required. Diagnostic procedures

Table 7 Diagnostics for logit model

	χ^2-test			Hosmer–Lemeshow test		
Model	2000 – 2004	2005 – 2009	2010 – 2014	2000 – 2004	2005 – 2009	2010 – 2014
I	0.000***	0.000***	0.000***	0.319	0.627	0.589
II	0.000***	0.000***	0.000***	0.009**	0.485	0.974
III	0.000***	0.000***	0.000***	0.391	0.764	0.521
IV	0.000***	0.000***	0.000***	0.777	0.618	0.383

Notes: *P*-values for the χ^2-test on the difference between the deviance residuals and the null residuals and for the Hosmer–Lemeshow test on the null of equality between the expected and observed frequency of exceedances. *,**,***Significance at the 5%, 1%, and 0.1% levels, respectively.

are applied to a standardized version of the data, conditional on the fitted parameter values. Consider a threshold u leading to k exceedances $y_1, \ldots, y_k$, where the estimated model is $Y_j - u \sim \text{GP}(\widehat{\nu}_j, \widehat{\xi}_j)$ with $j \in 1, \ldots, k$ and $\widehat{\nu}_j$ and $\widehat{\xi}_j$ are, respectively, the estimated scale and shape parameters of the GP distribution at the time when the exceedance occurs. Transforming the observations $Y_j - \widehat{u}$ to standard exponentially distributed variables $\tilde{Y}_j = \frac{1}{\widehat{\xi}_j}\log\{1 + \widehat{\xi}_j(\frac{Y_j - \widehat{u}}{\widehat{\sigma}_j})\}$ and denoting the ordered values of the observed $\tilde{Y}_j$'s as $\tilde{y}_{(1)}, \ldots, \tilde{y}_{(k)}$, a quantile plot is obtained using the pairs $\{(\tilde{y}_{(i)}, -\log(1 - i/(k+1))); i = 1, \ldots, k\}$.

The latter plot appears in Figure 4 for our dynamic GP models. The fits are generally satisfactory for all the specifications across the three windows considered, confirming the adequacy of the chosen thresholds.

We check whether the assumption of conditional independence between threshold exceedances, required for the estimation of the Logit model, is sensible. Figure 5 reports the autocorrelation of the Pearson residuals, $\rho_t = (I_t - \widehat{\phi}_t)/\sqrt{\widehat{\phi}_t(1 - \widehat{\phi}_t)}$, for the Logit model with log RV_{t-1} as covariate, and a model with constant probability equivalent to the unconditional exceedance rate, $\phi = n_u/n$, where n_u is the number of observations exceeding the threshold u. If the dynamic model for the exceedance rate accounts for the serial dependence, the corresponding Pearson residuals show zero autocorrelation while the Pearson residuals from the constant model should still be autocorrelated. Figure 5 shows this precisely, that is, conditional modeling based on RV completely filters out the dependence in the exceeding events.

We verify that the exceedance sizes are conditionally independent. Figure 6 shows the autocorrelation of the residuals R_j from the dynamic GP model with log RV_{t-1} as covariate and an analogous model with constant parameters. The plots show that the dependence between exceedance sizes is strong only in the second sub-sample, while the conditional modeling based on RV perfectly cleans this sequence. The autocorrelation in the third panel exhibits higher variability due to the small number of observations ($\approx$40) over the required higher threshold level (97th quantile).

Finally, we check the validity of the assumption of conditional independence between the rate and the magnitude of the exceedances. This assumption was required to perform separate maximization of the joint likelihood in Equation (8). We adopt the following

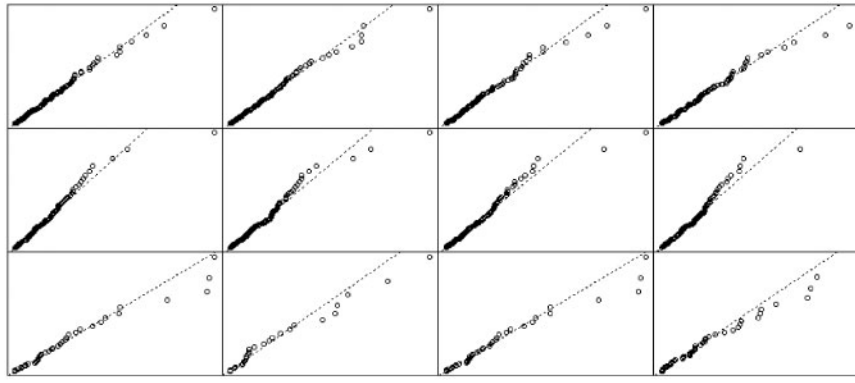

Figure 4 QQ-plot for dynamic GP distribution. From top to bottom, time intervals 2000–2004, 2005–2009, and 2010–2014, respectively. From left to right, Specifications I, II, III, and IV, respectively. Each panel has the theoretical quantiles of a unit-rate exponential distribution on the *x*-axis, and the empirical quantiles of the excesses transformed to exponential on the *y*-axis.

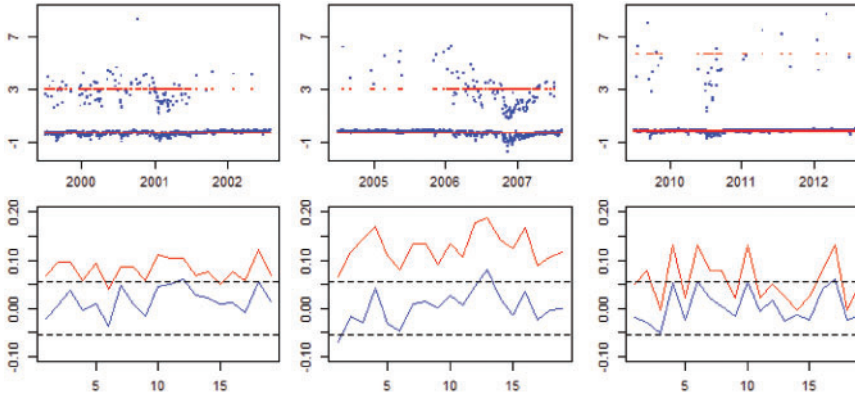

Figure 5 Logit residuals. Pearson residuals (*upper panel*) and corresponding autocorrelation (*lower panel*) for the Logit model (dark (blue)) and for a constant probability (pale (red)). Dashed lines in the lower panel correspond to the Bartlett confidence bands. Columns show results for 2000–2004, 2005–2009, and 2010–2014, respectively.

strategy: we obtain an estimate of the exceedance probability conditional on the lagged log RV_{t-1}; we obtain estimates of the conditional probability of the size of the exceedances based on the dynamic GP model with log RV_{t-1}, and on the model with constant parameters; we then plot the filtered probability from the Logit model against the probability from the dynamic and constant GP models, respectively. Figure 7 reports the results for each subsample. The upper panel shows that the exceedance probability and the distribution of the size of the exceedances as obtained from the constant GP model are positively and significantly related. In contrast, the lower panel shows that when we model the size of the exceedances with the dynamic GP model, this dependence disappears.

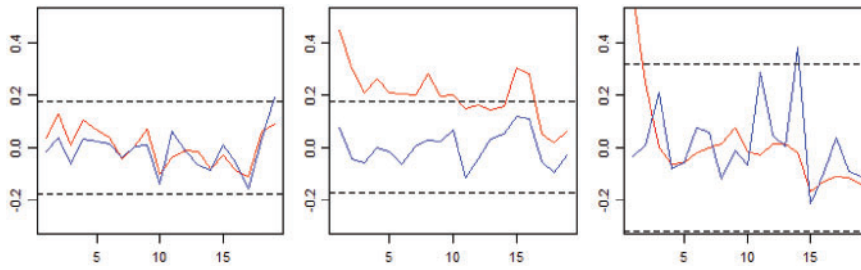

Figure 6 GP residuals. Autocorrelation of residuals from the dynamic GP model (dark (blue)) and the constant GP model (pale (red)). Dashed lines correspond to the Bartlett confidence bands. Columns show results for 2000–2004, 2005–2009, and 2010–2014, respectively.

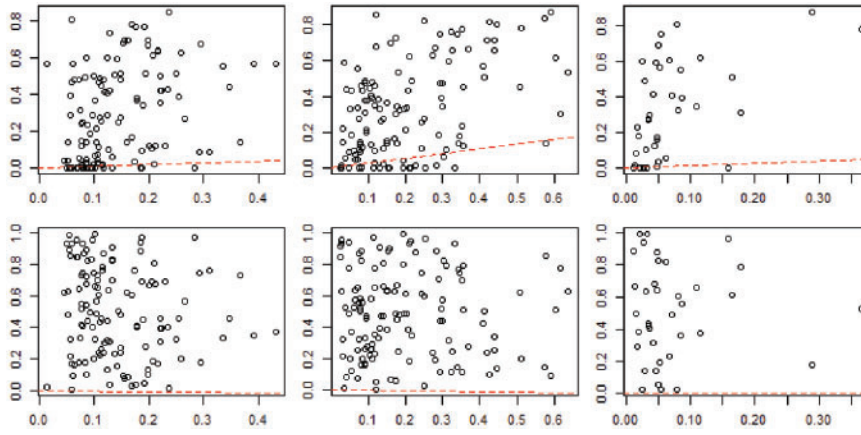

Figure 7 Cross residuals. Scatterplot of the probability of the size of exceedances from the GP model (*x*-axis) and the filtered exceedance probability from the Logit model (*y*-axis). The results for the constant GP model and the dynamic GP model are in the upper and lower panel, respectively. Regression lines are dashed with the following coefficients (standard errors): 0.09 (0.02), 0.27 (0.04), 013 (0.03) (*Upper*); 0.03 (0.03), −0.02 (0.05), 0.02 (0.03) (*Lower*). Columns show results for 2000–2004, 2005–2009, and 2010–2014, respectively.

4.5 The Value of HF Data

The previous sections have shown that RPOT models fit the data reasonably well, confirming that HF-based measures are informative with respect to threshold exceedances. An interesting question is whether HF data add information beyond that carried by LF data. It is important to note that our framework generalizes the past dynamic EVT models in finance, in the sense that LF realized measures, such as the squared returns can be easily included. Chavez-Demoulin, Davison, and McNeil (2005) and Chavez-Demoulin, Embrechts, and Sardy (2014) consider past daily exceedances to learn the behavior of the future exceedances, but this is equivalent to using the daily semi-variance as realized measure in our framework.

Table 8 Fitted logit models with HF and LF covariates

		2000–2004			2005–2009			2010–2014		
LM_{t-1}	Model	φ	φ_{LF}	φ_{HF}	φ	φ_{LF}	φ_{HF}	φ	φ_{LF}	φ_{HF}
I_{t-1}	LF.	−2.27***	0.62*		−2.26***	0.60*		−3.52***	1.06	
		(0.10)	(0.27)		(0.10)	(0.26)		(0.17)	(0.63)	
	HF.	5.60***	−0.11	0.86***	5.71***	−0.63*	0.85***	6.29***	−0.68	1.02***
		(1.03)	(0.29)	(0.11)	(0.76)	(0.30)	(0.08)	(1.27)	(0.76)	(0.14)
$\log R2_{t-1}$	LF.	−1.36	0.08		−0.15	0.20***		−2.05**	0.14*	
		(0.52)	(0.05)		(0.47)	(0.05)		(0.80)	(0.07)	
	HF.	5.41***	−0.05	0.89***	4.99***	−0.06	0.85***	5.72***	−0.11	1.09***
		(0.96)	(0.04)	(0.12)	(0.69)	(0.04)	(0.09)	(1.14)	(0.05)	(0.15)
$\log DR_{t-1}$	LF.	3.16***	0.57***		3.38***	0.59***		4.30***	0.79***	
		(0.85)	(0.09)		(0.61)	(0.07)		(1.17)	(0.12)	
	HF.	5.44***	−0.06	0.91***	4.96***	−0.60**	1.41***	5.55***	−0.12	1.08***
		(0.96)	(0.17)	(0.20)	(0.69)	(0.18)	(0.22)	(1.14)	(0.31)	(0.33)
$\log h_{t-1}$	LF.	7.60***	1.11***		5.62***	0.88***		6.33***	1.05***	
		(1.12)	(0.13)		(0.70)	(0.08)		(1.33)	(0.15)	
	HF.	7.94***	0.78***	0.36*	5.74***	0.49**	0.39*	6.97***	0.47*	0.64***
		(1.14)	(0.22)	(0.18)	(0.71)	(0.18)	(0.17)	(1.32)	(0.24)	(0.20)

Notes: LF. $\text{logit}(\phi_t) = \varphi + \varphi_{LF} LM_{t-1}$
HF. $\text{logit}(\phi_t) = \varphi + \varphi_{LF} LM_{t-1} + \varphi_{HF} \log RV_{t-1}$
for LM_{t-1} as in first column.
*,**,***Significance at the 5%, 1%, and 0.1% levels, respectively.

To verify whether the HF-based measures add information to that conveyed by LF-based measures, we now consider models for the exceedance rate and the size of excesses that use both as covariates.

Let $R2_t$ be the squared returns, DR_t be the daily range, and h_t be the GARCH(1, 1) filtered variance on day t. Table 8 shows the results for the Logit model with each lagged variable I_{t-1}, $\log R2_{t-1}$, $\log DR_{t-1}$, and $\log h_{t-1}$ as a covariate. The realized measures are significant in the LF models across the different periods, however, adding RV can reduce the explanatory power of the LF measures. The coefficient φ_{HF} is strongly significant and similar in magnitude to that observed in Table 3. Figure 8 reports the value of the maximized likelihood for the different models in the three sub-samples. The plot shows that including HF measures always adds to the likelihood, with the size of the contribution being greater for I_{t-1} and smaller for $\log h_{t-1}$.

Let W_t be the excess size at time $t \in \mathcal{E}$, Chavez-Demoulin, Davison, and McNeil (2005) assumes that exceedances are Markov with $W_t|W_{t-1}$ distributed as a GP distribution with ν_t depending on W_{t-1}. In Table 9, we consider W_{t-1}, $\log R2_{t-1}$, $\log DR_{t-1}$, and $\log h_{t-1}$ as covariates in the dynamic GP model. We see that the LF coefficient κ_{LF} tends to be significant when LF realized measures are considered alone, but adding RV strongly reduces their explanatory power and the HF coefficient κ_{HF} is mostly significant. Note that over the period 2010–2014, the ξ parameter for the model with $\log h_{t-1}$ is negative and statistically significant. This can happen in a finite sample when the true tail is exponentially decaying. Figure 9 reports the likelihood value at the maximum for the different dynamic GP models. Again, the HF measure always contributes.

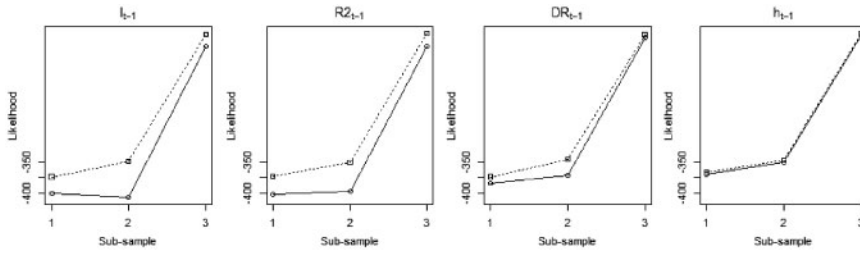

Figure 8 Logit likelihood with HF and LF covariates. Models with LF measures (solid line). Models with both HF and LF measures (dotted line).

Overall, this analysis shows that HF data convey information beyond that provided by LF data on the behavior of the extremes, confirming the merits of the RPOT approach.

4.6 The Impact of Illiquidity

We investigate whether illiquidity can be an explanatory factor of the dynamics of the extreme negative returns. The available data allow us to consider two proxies of illiquidity[7]: first, following Brennan and Subrahmanyam (1995) we consider the inverse of the daily volume, $\text{IV}_t = 1/\text{VOL}_t$, with VOL_t the transaction volume on day t; second, akin to the ILLIQ measure of Amihud (2002), we consider the ratio of the daily absolute returns to the daily volume of transactions, $\text{ILLIQ}_t = |r_t|/\text{VOL}_t$. Both these measures are based on volume, and finer measures of illiquidity could possibly be obtained from microstructure data on transactions and quotes. Given the limited dataset at our disposal, we do not pursue this path, but consider it an interesting topic for future research.

Tables 10–11 report the estimates of the Logit and dynamic GP models with illiquidity covariates. Illiquidity shows a significant effect on both the probability of an exceedance and the size of the exceedance in the second and third sub-samples only when using the inverse of daily volume as illiquidity proxy. The sign of this effect is negative, in agreement with the fact that there exists a positive risk premium associated to illiquidity (Amihud, 2002). The significance of these parameters vanishes however when we add the RV, with the latter presenting again a positive and strongly significant effect on the negative extremes.

4.7 Adding Autoregressive Terms

We include autoregressive terms, adapting our modeling approach to the GAS framework of Creal, Koopman, and Lucas (2013). We combine the Logit model for the exceedance probability and the GP model for the size of the exceedances in a global censored model. Let y_t be the negative return time series and u a fixed threshold. Defining the sequence of exceedances $z_t = y_t - u$, we can write the cumulative distribution function of the model as,

$$H_t(z_t|\nu_t, \phi_t, \xi) = \begin{cases} 1 - \phi_t & z_t = 0 \\ 1 - \phi_t \left(1 + \frac{\xi}{\nu_t} z_t\right)^{-1/\xi} & z_t \geq 0. \end{cases}$$

7 Note that the Ox-Realized Library does not provide volume of transaction. Records for this quantity were downloaded free of charge from Yahoo Finance.

Table 9 Fitted dynamic GP models with HF and LF covariates

		2000–2004				2005–2009				2010–2014			
LM_{t-1}	Model	κ	κ_{LF}	κ_{HF}	ξ	κ	κ_{LF}	κ_{HF}	ξ	κ	κ_{LF}	κ_{HF}	ξ
W_{t-1}	LF.	−4.99***	6.21		0.02	−4.73***	25.84***		−0.02	−5.01***	1.31		0.17
		(0.16)	(11.29)		(0.09)	(0.13)	(5.25)		(0.08)	(0.22)	(10.87)		(1.33)
	HF.	−2.28	−0.54	0.31**	0.02	−1.89*	14.68*	0.32***	−0.14*	0.08	0.03	0.55**	−0.18
		(1.19)	(10.35)	(0.14)	(0.09)	(0.62)	(6.16)	(0.07)	(0.06)	(0.19)	(1.74)	(0.20)	(0.17)
$\log R2_{t-1}$	LF.	−4.29***	0.06*		−0.01	−3.66***	0.08**		0.10	−5.83***	−0.08		0.16
		(0.33)	(0.03)		(0.08)	(0.36)	(0.03)		(0.08)	(1.06)	(0.09)		(0.17)
	HF.	−2.26*	0.03	0.27*	0.02	−1.01*	−0.05	0.47***	0.01	0.28	−0.15*	0.76***	−0.17
		(1.15)	(0.03)	(0.14)	(0.09)	(0.54)	(0.04)	(0.07)	(0.08)	(1.19)	(0.07)	(0.17)	(0.13)
$\log DR_{t-1}$	LF.	−2.93**	0.23*		0.04	−1.21*	0.36***		−0.01	−0.25	0.49***		−0.19
		(1.01)	(0.12)		(0.09)	(0.57)	(0.06)		(0.08)	(1.15)	(0.13)		(0.17)
	HF.	−2.27	0.02	0.29	0.03	−1.07	−0.15	0.57**	0.01	1.32	−0.20	0.90*	−0.15
		(1.20)	(0.19)	(0.22)	(0.09)	(0.57)	(0.16)	(0.17)	(0.08)	(1.31)	(0.39)	(0.45)	(0.17)
$\log h_{t-1}$	LF.	−1.16	0.45**		0.08	−0.61	0.46***		−0.13*	4.23***	1.01***		−0.54**
		(1.42)	(0.17)		(0.09)	(0.59)	(0.08)		(0.06)	(0.94)	(0.08)		(0.16)
	HF.	−0.89	0.26	0.21	0.04	−0.55	0.41**	0.06	−0.12*	4.56***	0.95***	0.10	−0.55**
		(1.45)	(0.25)	(0.19)	(0.09)	(0.60)	(0.13)	(0.11)	(0.06)	(1.15)	(0.15)	(0.17)	(0.17)

Notes: The parameter ξ is constant while the parameter ν_t is allowed to vary according to

LF. $\log \nu_t = \kappa + \kappa_{LF} LM_{t-1}$

HF. $\log \nu_t = \kappa + \kappa_{LF} LM_{t-1} + \kappa_{HF} \log RV_{t-1}$

for LM_{t-1} as in first column. *,**,***Significance at the 5%, 1%, and 0.1% levels, respectively.

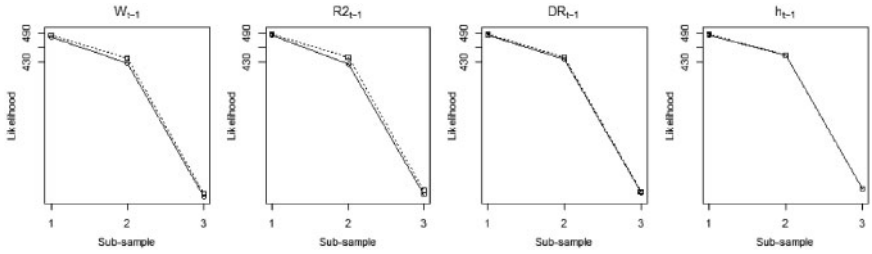

Figure 9 Dynamic GP likelihood with HF and LF covariates. Models with LF measures (solid line). Models with both HF and LF measures (dotted line).

Letting $\sigma_t = \log \nu_t$ and $p_t = \log\left(\frac{\phi_t}{1-\phi_t}\right)$, we consider the following dynamics,

$$\begin{aligned} \sigma_t &= \omega_1 + \beta_1 \sigma_{t-1} + \alpha_1 s_{1,t-1} + \gamma_1 \log \mathrm{RV}_{t-1}, \\ p_t &= \omega_2 + \beta_2 p_{t-1} + \alpha_2 s_{2,t-1} + \gamma_2 \log \mathrm{RV}_{t-1}, \end{aligned} \tag{16}$$

where $s_t = S_t \nabla_t$ is a score-based update suggested in Creal, Koopman, and Lucas (2013). In particular, $\nabla_t = \left(\frac{\partial \log h_t}{\partial \sigma_t}, \frac{\partial \log h_t}{\partial p_t}\right)'$, where h_t is the density of H_t and S_t is a scaling matrix, that is, the inverse Hessian or its square root. Details on the implementation of this model can be found in the Supplementary Material.

Table 12 displays the estimated parameters of the model in Equation (16) for the S&P 500. The results show that the persistence parameters β_1 and β_2 are strongly significant, even when the RV is added. In contrast, the coefficients related to the score-based updates, α_1 and α_2, are significant in the baseline model (GAS), but not when the RV is included. The coefficients of the RV are positive and significant. In sum, the inclusion of an autoregressive term along with a HF-based measure may be worthwhile.

5 Out-of-Sample Forecasts

In Section 4, we have shown that HF data contribute significantly toward explaining the behavior of daily extreme returns. However, in-sample fit does not guarantee a satisfactory out-of-sample forecast performance. In this section, we investigate whether HF data also lead to good out-of-sample forecasts. To this end, we perform an out-of-sample analysis of the 1-day-ahead VaR and ES forecasts defined, respectively, in Equations (13) and (14). We consider Logit-type RPOT models with one lagged realized measure as covariate in both the probability of exceedance and the size of the exceedances. We consider RV and BV as HF measures and the LF measures used in Section 4.5: $R2$ and the I&W combination à la Chavez-Demoulin, Davison, and McNeil (2005).

We apply a rolling-window scheme to obtain a time series of VaR and ES predictions at level $\alpha = 0.01$. Let n be the size of the available sample and s be the length of the window. We have two sequences of forecasts: $\{\mathrm{VaR}_t^\alpha\}_{t=s+1}^n$ and $\{\mathrm{ES}_t^\alpha\}_{t=s+1}^n$ of length $m = n - s$, where each prediction is obtained considering the observations $l_{t-s}, \ldots, l_{t-1}$. We produce $m = 1744$ predictions when considering a window of size $s = 2000$. The threshold level is fixed at the 90th quantile of the unconditional loss distribution obtained from the s observations.

Table 10 Logit models with illiquidity measures

	2000–2004			2005–2009			2010–2014		
	φ_0	φ_1	φ_2	φ_0	φ_1	φ_2	φ_0	φ_1	φ_2
(i)	−9.57	−0.35		−52.10***	−2.26***		−76.42***	−3.29***	
	(10.50)	(0.50)		(5.29)	(0.24)		(14.02)	(0.63)	
(ii)	1.73	0.15		−0.02	0.08		−0.17	0.12	
	(2.62)	(0.10)		(2.34)	(0.09)		(4.01)	(0.15)	
(iii)	19.64*	0.65	0.90***	−19.98	−1.02**	0.51***	−18.50	−1.03	0.82***
	(9.79)	(0.45)	(0.11)	(8.69)	(0.35)	(0.12)	(19.58)	(0.83)	(0.18)
(iv)	3.76	−0.08	0.87***	1.08	−0.16*	0.83***	0.33	−0.23	1.07***
	(2.19)	(0.09)	(0.11)	(2.03)	(0.08)	(0.08)	(2.48)	(0.11)	(0.14)

Notes: The following specifications are considered:
(i) $\text{logit}(\phi_t) = \varphi_0 + \varphi_1 \log \text{IV}_{t-1}$
(ii) $\text{logit}(\phi_t) = \varphi_0 + \varphi_1 \log \text{ILL}_{t-1}$
(iii) $\text{logit}(\phi_t) = \varphi_0 + \varphi_1 \log \text{IV}_{t-1} + \varphi_2 \log \text{RV}_{t-1}$
(iv) $\text{logit}(\phi_t) = \varphi_0 + \varphi_1 \log \text{ILL}_{t-1} + \varphi_2 \log \text{RV}_{t-1}$.

Table 11 Dynamic GP models with illiquidity measures

	2000–2004				2005–2009				2010–2014			
	κ_0	κ_1	κ_2	ξ	κ_0	κ_1	κ_2	ξ	κ_0	κ_1	κ_2	ξ
I	−14.47**	−0.45		0.04	−30.18***	−1.15***		0.11	−19.69	−0.66		0.19
	(5.33)	(0.25)		(0.08)	(6.11)	(0.27)		(0.10)	(24.89)	(1.12)		(0.17)
II	−2.04	0.11		−0.01	−2.72	0.06		0.13	−9.19	−0.15		0.17
	(1.64)	(0.06)		(0.08)	(2.15)	(0.08)	(0.09)		(5.34)	(0.19)		(0.17)
III	−1.62	0.03	0.32*	0.04	3.73	0.19	0.46***	−0.02	−12.54	−0.59	0.62***	−0.21
	(8.04)	(0.34)	(0.16)	(0.09)	(11.10)	(0.46)	(0.12)	(0.07)	(13.67)	(0.59)	(0.15)	(0.22)
IV	−0.89	0.06	0.28*	0.02	−2.85	−0.08	0.44***	0.01	−5.76*	−0.27	0.72***	−0.19
	(1.86)	(0.06)	(0.14)	(0.09)	(2.05)	(0.08)	(0.07)	(0.08)	(2.45)	(0.10)	(0.15)	(0.13)

Notes: The following specifications are considered:
I $\log \nu_t = \kappa_0 + \kappa_1 \log \text{IV}_{t-1}$
II $\log \nu_t = \kappa_0 + \kappa_1 \log \text{ILL}_{t-1}$
III $\log \nu_t = \kappa_0 + \kappa_1 \log \text{IV}_{t-1} + \kappa_2 \log \text{RV}_{t-1}$
IV $\log \nu_t = \kappa_0 + \kappa_1 \log \text{ILL}_{t-1} + \kappa_2 \log \text{RV}_{t-1}$.

We evaluate the performance of the RPOT approach by performing a battery of tests. We consider the binary indicator of VaR failure $\left\{H_{t+1} = I\left(l_{t+1} > \widehat{\text{VaR}}^{\alpha}_{t+1|t}\right)\right\}$, where $I(\cdot)$ is the indicator function. Commonly used tests are the Unconditional Coverage (UC), Independence (IND), and Conditional Coverage (CC) suggested by Christoffersen (1998) and the Dynamic Quantile (DQ) test suggested by Engle and Manganelli (2004). We also perform a Diebold–Mariano (DM) test on the null of equal predictive accuracy against the model based on the I&W. To evaluate the performance for the ES, we test the hypothesis that conditional upon exceeding the 99th quantile of the loss distribution, the difference between the actual return and the predicted ES has mean zero. In particular, we perform a

Table 12 RPOT models with autoregressive terms

	ω_1	ω_2	α_1	α_2	β_1	β_2	ξ	γ_1	γ_2
GAS	−0.096*	−0.016**	0.019***	0.021***	0.98***	0.992***	−0.02		
	(0.052)	(0.006)	(0.005)	(0.003)	(0.108)	(0.029)	(0.03)		
RPOT-GAS	−0.048	0.991*	0.012	0.011	0.725***	0.839***	−0.097**	0.140***	0.144*
	(0.139)	(0.512)	(0.008)	(0.011)	(0.051)	(0.084)	(0.035)	(0.031)	(0.074)

Notes: Estimates of the dynamic model in Equation (16) without (GAS) and with (RPOT-GAS) the RV as external regressors. Standard errors are in parenthesis. *,**,***Significance at the 5%, 1%, and 0.1% levels, respectively.

Table 13 Tests on conditional risk measures

Measure	Violation	UC	IND	CC	DQ	BOOT
RV	0.97	0.91	0.56	0.83	0.99	0.38
BV	1.03	0.89	0.54	0.81	0.99	0.54
R2	1.60	**0.02**	0.08	**0.01**	**0.01**	0.09
I&W	1.55	**0.03**	0.35	0.07	**0.01**	**0.03**

Notes: Percentage of violations (Violation); the *p*-values for the unconditional coverage (UC), the independence assumption (IND), the conditional coverage (CC), and the DQ for the VaR predictions at level $\alpha = 0.01$, and the bootstrap test (BOOT) for the ES predictions at level $\alpha = 0.01$. Rejection at the 5% significance level appears in bold.

one-sided test with the alternative that the mean is greater than zero using a bootstrap that makes no assumption about the distribution of the differences (McNeil and Frey, 2000).

The *p*-values in Table 13 show that the models based on HF data perform well, while those based on LF data can be problematic. The *p*-values from the DM tests against the model based on I&W confirm the superior performance of the RPOT: 0.02 (RV), 0.01 (BV), and 0.23 (*R*2). We conclude that the forecasting ability of RPOT models with HF measures is superior to that of models based on LF measures.

To confirm the ability of the RPOT to produce good forecasts, we also provide an analysis using a window size of 1000 observations instead of 2000. Results in the Supplementary Material show that the RPOT outperforms the model based on LF measures in this case as well.

6 Conclusions

The availability of HF data has lead to breakthroughs in the financial econometrics literature, and models that exploit this source of information are superseding standard econometric models. In this paper, we propose a novel HF extreme value approach where realized measures are used to model the time-varying behavior of extreme returns. In-sample fit of these models shows that measures of variation built from HF data add information on the extremes, beyond that conveyed by LF data, illiquidity measures, and score-based innovations. Moreover, out-of-sample forecasts of standard risk measures obtained with HF covariates are superior to those obtained with LF covariates.

We are working on refinements and extensions of the RPOT with the aim of establishing a complete framework where HF data are used within extreme value models. For example, adding parametric or non-parametric smoothing components may enhance the stability of the extreme value models. Furthermore, the intuition of this paper can be used to extend the time-varying threshold model of Wang, Li, and He (2012) to financial returns. Finally, the development of multivariate models where the realized covariance is used as a source of information would allow for the modeling of the joint occurrence of extreme events. Such a framework could be used to provide a much needed extreme value perspective measure of contagion effects among assets.

Supplementary Data

Supplementary data are available at *Journal of Financial Econometrics* online.

Appendix

Proof of Proposition 1. Let $y_t = (z, \mathbf{RM}_t)$ and $\theta = (\boldsymbol{\kappa}, \xi) \in \Theta \subset \mathbb{R}^{p+1} \times (0, 1]$, we have

$$\log\ g_t(y_t;\theta) = -\mathbf{RM}_t'\boldsymbol{\kappa} - \left(1+\frac{1}{\xi}\right)\log\left(1+\frac{\xi}{\exp\left(\mathbf{RM}_t'\boldsymbol{\kappa}\right)}z\right),$$

with $z \geq 0$. The shape parameter ξ must be lower than 1 for z to have finite first moment. We use y_t to stress that the likelihood properties depend on both the random variables z and $\mathbf{RM}_t$. To prove consistency, we verify the conditions of Theorem 2.5 of Newey and McFadden (1994).

The first requirement is identifiability. The second requirement is $\theta_0 \in \Theta$ and compactness of Θ. These two requirements hold by Assumption (i).

The third requirement is continuity. Since the density of the GP and the exponential function used to link the GP scale parameter to the covariates are both continuous, the requirement is satisfied.

The last requirement is that $\mathbb{E}\left[\sup_{\theta\in\Theta}|\log g_t(\mathbf{y}_t;\boldsymbol{\theta})|\right] < \infty$. We need to show that there exists a function $d_t(\mathbf{y}_t)$ such that $|\log g_t(\mathbf{y}_t;\boldsymbol{\theta})| \leq d_t(\mathbf{y}_t) \forall \boldsymbol{\theta} \in \Theta$ and $\mathbb{E}\left(d_t(\mathbf{y}_t)\right) < \infty$. We have that,

$$\begin{aligned}
|\log\ g_t(\mathbf{y}_t;\boldsymbol{\theta})| &= |-\mathbf{RM}_t'\boldsymbol{\kappa} - \left(\frac{1}{\xi}+1\right)\log\left(1+\xi\ \exp\left(-\mathbf{RM}_t'\boldsymbol{\kappa}\right)z_t\right)| \\
&\leq |\mathbf{RM}_t'\boldsymbol{\kappa}| + \left|\frac{1}{\xi}+1\right||\log\ \left(1+\xi\ \exp\ \left(-\mathbf{RM}_t'\boldsymbol{\kappa}\right)z_t\right)| \\
&\leq \|\mathbf{RM}_t\|\|\boldsymbol{\kappa}\| + \left|\frac{1}{\xi}+1\right||\log\ \left(1+\xi\ \exp\ \left(-\mathbf{RM}_t'\boldsymbol{\kappa}\right)z_t\right)|.
\end{aligned}$$

By Assumption (i), $\boldsymbol{\theta}$ belongs to a compact set, therefore there exist positive constants bounding $\|\boldsymbol{\kappa}\|$ and $|\frac{1}{\xi}+1|$, respectively. Since $\xi\ \exp\left(-\mathbf{RM}_t'\boldsymbol{\kappa}\right)z_t > 0$, we have that the term $|\log\ \left(1+\xi\ \exp\ \left(-\mathbf{RM}_t'\boldsymbol{\kappa}\right)z_t\right)| < 1+\xi\ \exp\ \left(-\mathbf{RM}_t'\boldsymbol{\kappa}\right)z_t$. Besides, $\exp\ \left(-\mathbf{RM}_t'\boldsymbol{\kappa}\right) \leq \exp\ (\|\mathbf{RM}_t\|\|\kappa\|)$, so we let

$$d_t(\mathbf{y}_t) = C||\mathbf{RM}_t|| + C(1+\exp\ (C||\mathbf{RM}_t||)z_t),$$

with $C>0$ a large enough positive constant. We can then compute the expected value of $d_t(\mathbf{y}_t)$. Let $\mathbb{E}_z\big(d_t(\mathbf{y}_t)\big)$ denote the conditional expectation with respect to z_t. For C large enough, we have

$$\begin{aligned}\mathbb{E}_z\big(d_t(\mathbf{y}_t)\big) &= \int_0^{\infty} (C||\mathbf{RM}_t||+C(1+\exp(C||\mathbf{RM}_t||)z_t))\exp\left(-\mathbf{RM}_t'\boldsymbol{\kappa}\right)\left(1+\xi\ \exp\left(-\mathbf{RM}_t'\boldsymbol{\kappa}\right)z_t\right)^{-1/\xi-1}\mathrm{d}z\\ &= C||\mathbf{RM}_t||+C+C\ \exp(C||\mathbf{RM}_t||)\int_0^{\infty} z_t\ \exp\left(-\mathbf{RM}_t'\boldsymbol{\kappa}\right)\left(1+\xi\ \exp\left(-\mathbf{RM}_t'\boldsymbol{\kappa}\right)z_t\right)^{-1/\xi-1}\mathrm{d}z\\ &= C||\mathbf{RM}_t||+C+C\ \exp(C||\mathbf{RM}_t||)\frac{\exp\left(\mathbf{RM}_t'\boldsymbol{\kappa}\right)}{1-\xi}\\ &\leq C||\mathbf{RM}_t||+C+2C\frac{\exp(C||\mathbf{RM}_t||)}{1-\xi},\end{aligned}$$

so that $\mathbb{E}\big(d_t(\mathbf{y}_t)\big)$ is finite by Assumption (iii).

Asymptotic normality follows from Theorem 3.3 of Newey and McFadden. Condition 3.3(i) holds by Assumption (i). Condition 3.3(ii) follows from the differentiability of the GP distribution and of the exponential function. Then, we have that

$$J=\mathbb{E}\big(\nabla_\theta\ \log\ g_t(\mathbf{y}_t;\boldsymbol{\theta}_0)\nabla_\theta\ \log\ g_t(\mathbf{y}_t;\boldsymbol{\theta}_0)'\big)=\frac{1}{1+2\xi_0}\begin{pmatrix}\mathbb{E}(\mathbf{RM}_t\mathbf{RM}_t') & \dfrac{1}{1+\xi_0}\\ \bullet & \dfrac{2}{1+\xi_0}\end{pmatrix},$$

and condition 3.3(iv) is satisfied by Assumption (ii). Finally, Conditions 3.3(iii)–(v) follow from arguments analogous to those used in the proof of the dominance condition for consistency. Details on these computations can be found in the Supplementary Material. □

References

Amihud, Y. 2002. Illiquidity and Stock Returns: Cross-section and Time-Series Effects. *Journal of Financial Markets* 5: 31–56.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys. 2001. The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association* 96: 42–55.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys. 2003. Modeling and Forecasting Realized Volatility. *Econometrica* 71: 579–625.

Andersen, T. G., T. Bollerslev, and F. X. Diebold. 2007. Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. *Review of Economics and Statistics* 89: 701–720.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys. 2009. "Parametric and Nonparametric Volatility Measurement." In Y. Ait-Sahalia and L. Hansen (eds), *Handbook of Financial Econometrics*, Vol. 1. p. 67–138. North Holland, Amsterdam.

Bandi, F. M., and R. Renò. 2012. Time-Varying Leverage Effects. *Journal of Econometrics* 169: 94–113.

Barndorff-Nielsen, O. E., and N. Shephard. 2004. Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics* 2: 1–37.

Barndorff-Nielsen, O. E., and N. Shephard. 2007. "Variation, Jumps and High Frequency Data in Financial Econometrics." In R. Blundell, T. Persson, and W. Newey (eds), *Advanced in Economics and Econometrics: Theory and Applications, Ninth World Congress*, Econometric Society Monograph, p. 328–72. Cambridge, UK: Cambridge University Press.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard. 2008. Designing Realized Kernels to Measure the Ex Post Variation of Equity Prices in the Presence of Noise. *Econometrica* 76: 1481–1536.

Beirlant, J., and Y. Goegebeur. 2004. Local Polynomial Maximum Likelihood Estimation for Pareto-Type Distributions. *Journal of Multivariate Analysis* 89: 97–118.

Bollerslev, T., and V. Todorov. 2011. Estimation of Jump Tails. *Econometrica* 79: 1727–1783.

Bollerslev, T., and V. Todorov. 2014. Time-Varying Jump Tails. *Journal of Econometrics* 183: 168–180.

Brennan, M. J., and A. Subrahmanyam. 1995. Investment Analysis and Price Formation in Securities Markets. *Journal of Financial Economics* 38: 361–381.

Chavez-Demoulin, V., and A. C. Davison. 2005. Generalized Additive Modelling of Sample Extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 54: 207–222.

Chavez-Demoulin, V., A. Davison, and A. J. McNeil. 2005. Estimating Value-at-Risk: A Point Process Approach. *Quantitative Finance* 5: 227–234.

Chavez-Demoulin, V., P. Embrechts, and S. Sardy. 2014. Extreme-Quantile Tracking for Financial Time Series. *Journal of Econometrics* 181: 44–52.

Choulakian, V., and M. Stephens. 2001. Goodness-of-Fit Tests for the Generalized Pareto Distribution. *Technometrics* 43: 478–484.

Christoffersen, P. F. 1998. Evaluating Interval Forecasts. *International Economic Review* 39: 841–862.

Cohen, J. P. 1982. Convergence Rates for the Ultimate and Penultimate Approximations in Extreme-Value Theory. *Advances in Applied Probability* 14: 833–854.

Coles, S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. London, UK: Springer.

Corsi, F. 2009. A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics* 7: 174–196.

Creal, D., S. J. Koopman, and A. Lucas. 2013. Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics* 28: 777–795.

Danielsson, J., and C. G. de Vries. 1997. Tail Index and Quantile Estimation with Very High Frequency Data. *Journal of Empirical Finance* 4: 241–257.

Davison, A. C., and D. V. Hinkley. 1997. *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.

Davison, A. C., and R. L. Smith. 1990. Models for Exceedances over High Thresholds. *Journal of the Royal Statistical Society: Series B* 52: 393–442.

De Lira Salvatierra, I. A., and A. J. Patton. 2015. Dynamic Copula Models and High Frequency Data. *Journal of Empirical Finance* 30: 120–135.

Diebold, F. X., T. Schuermann, and J. D. Stroughair. 2000. Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management. *The Journal of Risk Finance* 1: 30–35.

Engle, R. F., and S. Manganelli. 2004. Caviar: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business & Economic Statistics* 22: 367–381.

Hall, P., and N. Tajvidi. 2000. Nonparametric Analysis of Temporal Trend When Fitting Parametric Models to Extreme-Value Data. *Statistical Science* 15: 153–167.

Hansen, B. E. 1994. Autoregressive Conditional Density Estimation. *International Economic Review* 35: 705–730.

Hansen, P. R., and Z. Huang. 2016. Exponential GARCH Modeling with Realized Measures of Volatility. *Journal of Business & Economic Statistics* 34: 269–287.

Hansen, P. R., Z. Huang, and H. H. Shek. 2012. Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility. *Journal of Applied Econometrics* 27: 877–906.

Hansen, P. R., A. Lunde, and V. Voev. 2014. Realized Beta GARCH: A Multivariate GARCH Model with Realized Measures of Volatility. *Journal of Applied Econometrics* 29: 774–799.

Harvey, C. R., and A. Siddique. 1999. Autoregressive Conditional Skewness. *Journal of Financial and Quantitative Analysis* 34: 465–487.

Heber, G., A. Lunde, N. Shephard, and K. Sheppard. 2009. *Oxford-Man Institute Realized Library, version 0.2*. Oxford-Man Institute, University of Oxford.

Hosking, J. R., and J. R. Wallis. 1987. Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics* 29: 339–349.

Hosmer, D. W., and S. Lemeshow. 2004. *Applied Logistic Regression*. New York, NY: John Wiley & Sons.

Kelly, B., and H. Jiang. 2014. Tail Risk and Asset Prices. *Review of Financial Studies* 27: 2841–2871.

Longin, F. 2000. From Value at Risk to Stress Testing: The Extreme Value Approach. *Journal of Banking & Finance* 24: 1097–1130.

Massacci, D. 2017. Tail Risk Dynamics in Stock Returns: Links to the Macroeconomy and Global Markets Connectedness. *Management Science* 63: 3072–3089.

McNeil, A. J., and R. Frey. 2000. Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach. *Journal of Empirical Finance* 7: 271–300.

Mikosch, T., and C. Stărică. 2004. Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects. *Review of Economics and Statistics* 86: 378–390.

Newey, W. K., and D. McFadden. 1994. "Large sample estimation and hypothesis testing." In R. F. Engle and D. L. McFadden (eds), *Handbook of Econometrics*, Vol. 4, p. 2111–2245. Amsterdam: Elsevier Science.

Noureldin, D., N. Shephard, and K. Sheppard. 2012. Multivariate High-Frequency-Based Volatility (HEAVY) Models. *Journal of Applied Econometrics* 27: 907–933.

Oh, D. H., and A. J. Patton. 2016. High Dimension Copula-Based Distributions with Mixed Frequency Data. *Journal of Econometrics* 193: 349–366.

Pickands, J. 1975. Statistical Inference Using Extreme Order Statistics. *Annals of Statistics* 3: 119–131.

Shephard, N., and K. Sheppard. 2010. Realising the Future: Forecasting with High-Frequency-Based Volatility (HEAVY) Models. *Journal of Applied Econometrics* 25: 197–231.

Smith, R. L. 1985. Maximum Likelihood Estimation in a Class of Nonregular Cases. *Biometrika* 72: 67–90.

Wang, H. J., D. Li, and X. He. 2012. Estimation of High Conditional Quantiles for Heavy-Tailed Distributions. *Journal of the American Statistical Association* 107: 1453–1464.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia. 2005. A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data. *Journal of the American Statistical Association* 100: 1394–1411.

Zhang, X., and S. Bernd. 2016. "Tail Risk in Government Bond Markets and ECB Unconventional Policies." Working paper.