Estimation of Tail Risk Measures for Heteroskedastic Financial Time Series: A Extreme Value Approach With Covariates*

Robert James   Henry Leung   Jessica Leung   Artem Prokhorov

July 22, 2021

Abstract

In this study, we explore how economic and financial covariates can be incorporated into the two-stage GARCH-EVT risk model of McNeil and Frey (2000) and whether or not this additional information improves Value-at-Risk and Expected Shortfall forecasts. We specify the extreme value distribution scale parameter in the second stage of the model as a log-linear function of covariates. However, the wealth of available data means that risk managers face a variable selection problem. Accordingly, we incorporate ℓ_1 regularization when estimating distribution parameters. Using an extensive set of performance criteria we demonstrate that our ℓ_1-regularized GARCH-EVT risk model produces competitive risk forecasts, particularly during periods financial distress.

*Helpful comments from participants of the 41st International Symposium on Forecasting and the International Conference on Econometrics and Business Analytics are gratefully acknowledged. The use of the University of Sydney's high performance computing cluster, Artemis, is acknowledged.

1 Introduction

Financial risk is typically summarized by the Value-at-Risk (VaR) and the Expected Shortfall (ES), measures that specify the amount of capital required to be held as a buffer against potential future losses on a financial asset or investment portfolio. These tail risk measures underpin international prudential regulations governing the financial industry. An extensive literature has developed many parametric, semi-parametric and non-parametric risk models to forecast the VaR and the ES. The performance of these risk models is underpinned by their ability to accurately model the tails of the conditional loss distribution. However, many existing risk models ignore the wealth of high frequency financial and economic data currently available and instead use only the historical returns of the financial assets or investment portfolio.

It is our intuition that economic and financial variables can provide additional information about the likelihood and severity of future losses. That is, models for the conditional loss distribution can be improved by incorporating covariates derived from the wealth of data which is now available. For example, the Federal Reserve Bank of St.Lous's FRED database records over 550,000 economic and financial time series. Moreover, the Trade-and-Quote (TAQ) database records every intraday trade and quote for some 8000 U.S. equities[1]. This high frequency data could be used to construct measures of trading activity and demand/supply patterns which may be relevant for understanding idiosyncratic downside risk. Of course, not all the information in these databases is relevant for predicting future losses, but it is reasonable to assume that at least some covariates could be useful[2].

In this study, we explore how economic and financial covariates can be integrated into an established financial risk forecasting model and whether or not this information improves VaR and ES forecasts. Traditionally, if a risk manager wanted to include covariates into a model then a theoretically relevant set of covariates would first need to be identified. In reality, identifying this set is no longer straightforward, due to the vast amount of economic and financial data that is now available. As such, the risk manager faces a variable selection problem. Since the conditional loss distribution is time-varying, this selection problem is further compounded by the possibility that the predictive importance of covariates changes over time.

Including the entire universe of potentially relevant covariates is not a valid solution for two reasons. First, the model will almost surely over-fit the data and produce poor VaR and ES forecasts. Second, when the set of potentially relevant covariates is large, parameter estimation techniques used in traditional models may break down as degrees of freedom are exhausted. For example, risk models that use extreme value distributions estimate parameters using only a small number of upper order statistics, typically 10% or less of the total sample size n. In high dimensions, such as cases where $p > 0.1n$, standard maximum likelihood estimation for extreme value distribution parameters is not possible since the first-order condition is rank deficient.

We hypothesize that there is a large set of relevant covariates to forecast future tail risk, but that only a handful of these covariates are statistically informative at any given time. In other words, we hypothesize the that the best risk model is sparse. We incorporate our hypothesis into the popular semi-parametric two-stage GARCH-EVT model, introduced by McNeil and Frey (2000). In the first stage of this model, conditional heteroskedasticity and auto-correlation are removed from the raw log-returns using a GARCH model. The standardized residuals from this GARCH model are assumed to form an approximately i.i.d time series. In the second stage, a generalized Pareto extreme value distribution is used to approximate the upper tail of the negated standardized residuals.

Our principle contributions are as follows. First, we use the GJR-GARCH model instead of the standard GARCH(1,1) model in the first stage of the model. Consistent with Trapin (2018), we find that the

[1] This equates to approximately 750 gigabytes of new data every month.

[2] For individual banks, internal trading desks record thousands of transactions per day while research divisions routinely produce proprietary forecasts for economic indicators and individual equities. This information could be used to enhance internal risk models.

GJR-GARCH model does a better job of removing dependence among the most extreme returns. Second, we re-parameterize the generalized Pareto distribution scale parameter as a log-linear function of economic and financial covariates. The covariates we study are idiosyncratic measures of the market microstructure and trading activity, measures of economic uncertainty, indices that track the U.S equity market volatility and broad measures of U.S. financial market stress. Third, ℓ_1-norm regularization is added as a secondary objective to the likelihood function of the generalized Pareto distribution. Two hyper-parameters to separately control the degree of variable selection and coefficient shrinkage induced ℓ_1-norm penalty, in the spirit of the so-called relaxed Lasso (Least Absolute Shrinkage and Selection Operator) method (Meinshausen, 2007; Hastie et al., 2020). Our revised bi-criterion likelihood function jointly performs parameter estimation and variable selection. We refer to our augmented model as the ℓ_1-regularized GARCH-EVT risk model. Finally, using this regularized forecasting model we study when our chosen economic and financial covariates improves tail risk forecasts and how each covariates is related to downside risk.

The rolling 1-day ahead VaR forecasts from the ℓ_1-regularized GARCH-EVT model are competitive with forecasts from benchmark risk models over a 2075-day period of normal market volatility. In particular, 1-day ahead VaR forecasts from the ℓ_1-regularized GARCH-EVT model typically outperform benchmark models at the $\alpha = 0.99$ level, the level of confidence that is prescribed by regulators. Moreover, the ℓ_1-regularized GARCH-EVT model outperforms VaR forecasts from the traditional GARCH-EVT model at the $\alpha = 0.975$ and $\alpha = 0.99$ levels.

During periods of financial market distress, we find that both the 1-day ahead VaR forecasts from the ℓ_1-regularized GARCH-EVT model outperform all benchmark models at the $\alpha = 0.975$ and $\alpha = 0.99$ levels. The 1-day ahead ES forecasts from the ℓ_1-regularized GARCH-EVT model also outperform forecasts produced by all other benchmark models at the $\alpha = 0.975$ level, the level of most interest to regulators. By studying differences in the value of consistent scoring functions for the VaR and ES, together with the behaviour of the active coefficient set, we demonstrate that information from our selected set of economic and financial covariates is particularly valuable during the Global Financial Crisis and around the "Black Monday" stock market crash on the 8th August 2011. Therefore, the ℓ_1-regularized GARCH-EVT model generates accurate risk forecasts, during times of financial market distress, when accurate forecasts are needed the most.

To demonstrate the need for variable selection mechanisms in financial risk forecasting models, we show that a naive GARCH-EVT model that includes all covariates without ℓ_1 regularization produces inadequate VaR and ES forecasts. On average, the ℓ_1-regularized GARCH-EVT model uses a constant scale parameter 44% of days in our backtest. When covariates are selected by the ℓ_1-regularized GARCH-EVT model, the average size of the active set is six. The majority of covariates we use have a mean predictive duration less than two days. Together, these statistics support our hypothesis that the best financial risk forecasting model uses a sparse set of covariates and that the predictive importance of covariates is time-varying.

Our analysis suggests that daily sell volume, total trade volume in the after-hours market and trade/quote based intraday volatility are generally the most important covariates for financial risk forecasting within the GARCH-EVT framework. Each of these covariates are positively related to the likelihood and severity of observing large negative returns. These covariates could be used to design effective risk forecasting models and more realistic stress testing scenarios.

Our study is related to the literature that develops novel forecasting methods for financial risk management and is therefore important for improving financial stability. We use a number of widely cited risk forecasting approaches as benchmarks in our experiments: the traditional GARCH-EVT model of McNeil and Frey (2000), the Conditional Auto-regressive Value-at-Risk model of Engle and Manganelli (2004), the conditional-Autoregressive Expectiles model of Taylor (2008) and the Hawkes Peaks-Over-Threshold (Hawkes POT) model of Chavez-Demoulin et al. (2005) and Chavez-Demoulin and McGill (2012). Rather than developing a new model from the ground up, we contribute to the literature by incorporating novel covariates and a contemporary statistical technique (a two hyper-parameter variant of ℓ_1 regularization) into the well-established GARCH-EVT model. Our approach extends this seminal model for use in the big data environments that are now prevalent. In doing so, we improving in tail risk forecasting accuracy, particularly during financial distress.

While economic and financial variables have been studied in models of market-wide systematic risk (see, e.g., Tobias and Brunnermeier, 2016), their role in idiosyncratic risk forecasting model for individual equities or investment portfolios has only recently been explored. Herrera and Clements (2020) develop a point process risk model that mutually incorporates a small number of covariates into the tail distribution of returns. However, the study of Hambuckers et al. (2018), who considered the relationship between quarterly economic variables and the severity of bank operational losses at a single bank using ℓ_1-regularized generalized Pareto regressions, is more closely related to ours. We extend this emergent set of literature by analyzing a more diverse set of financial covariates, in particular covariates that are derived from high-frequency millisecond level intraday transaction data, for a comparatively large sample of U.S equities using a more flexible variant of ℓ_1 regularization. Moreover, we extend a well-established risk forecasting model, as this approach is of considerable importance to the industry.

Finally, our study is related to an developing set of literature studying sparse models for financial applications. For example, Freyberger et al. (2020) and Chinco et al. (2019) use Lasso regression to investigate the determinants of return predictability within the cross-section of stocks. We extend this literature by studying a contemporary two hyper-parameter variant of ℓ_1 regularization, within a maximum likelihood framework for financial risk management applications. Consistent with the overarching conclusion from this emerging literature, we also find that sparse models can be effective for data-driven financial applications. Moreover consistent with (Easley et al., 2021), we also find that covariates which describe the market microstructre and trading process are useful for financial forecasting applications.

The remainder of this study is structured as follows. Section 2 formally defines the VaR and ES measures and briefly introduces the statistical theory associated with tail modelling using the generalized Pareto distribution. Section 3 introduces the traditional GARCH-EVT risk model and discusses the GJR-GARCH extension. The framework for the ℓ_1-regularized GRACH-EVT risk model is discussed in 4. Section 5 describes our empirical sample and covariates. The performance criteria used to assess the equality of VaR and ES forecasts is described in Section 6. Forecasting performance of the ℓ_1 regularized GRACH-EVT risk model and benchmark risk models is assessed in Section 7. Section 8 explores when covariates are most valuable and Section 9 analyzes the estimated coefficients. Finally, Section 10 concludes the study.

2 Risk Measures and Traditional Extreme Value Theory

To formally define the VaR and ES risk measures we use in this study, let P_t be the daily value of a financial asset or investment portfolio and $\{Y_{t,h} = -(\log(P_{t+h}) - \log(P_t)), t = 1, \dots, T\}$ be the time series of negative log returns for investment horizon h. We denote the conditional distribution of negative log returns at time t as $F_{Y_{t,h}}$. When discussing forecasts of risk measures conditional on the information available up to time t, we will include the time and investment horizon h subscripts, otherwise when discussing general properties of risk measures, we will omit these subscripts for brevity.

First, the Value-at-Risk (VaR) at time t for investment horizon h and confidence level α, is the smallest value for which the probability that the realized loss exceeds some value y is $1-\alpha$. Formally, the VaR is,

$$VaR_{t,h}^{(\alpha)} = \inf\{y : F_{Y_{t,h}}(y) > 1 - \alpha\} \tag{1}$$

In other words, the VaR is the α-quantile of the conditional loss distribution. Typically, when dealing with negative log returns, α is close to 1, so that VaR estimation methods involve modeling the upper tail of $F_{Y_{t,h}}$. The VaR is the cornerstone measure in the field of financial risk management and is also widely used in other disciplines, such as environmental engineering and health care quality control. However, the VaR fails to provide any information about losses that occur with a probability smaller than $1-\alpha$ (i.e losses beyond the α-quantile of the loss distribution). Moreover, the VaR is often criticized for failing to satisfy the axiom of sub-additivity (McNeil et al., 2015).

The second quantity of central importance in financial risk management is the Expected Shortfall (ES), also known within the literature as the Conditional Tail expectation or the Superquantile. The ES at time t, investment horizon h and confidence level α is defined as the conditional expectation of losses which exceed the $VaR_{t,h}^{(\alpha)}$ level. For continuous loss distributions, as is the case when working with negative log returns, the ES is,

$$ES_{t,h}^{(\alpha)} = E[Y|Y \geq VaR_{t,h}^{(\alpha)}] \tag{2}$$

Hence, the ES is sensitive to the size of all losses in the tail of the conditional loss distribution. Equivalently, the $ES^{(\alpha)}$ is the average of all VaRs in the tail beyond the $VaR^{(\alpha)}$ level (Acerbi and Tasche, 2002).

The ES is a coherent risk measure. As such, the Basel Committee on banking supervision has outlined a shift in risk methodologies from VaR towards ES (Basel Committee on Banking Supervision, 2016). However, the performance of ES forecasts for a financial institution's trading book is still evaluated by backtesting the VaR forecasts produced by the same risk model under the upcoming Basel III framework. Hence, both the VaR and the ES are important measures of risk.

To forecast the VaR and ES, a model for the upper tail of the conditional loss distribution is required. The so-called "Peaks-Over-Threshold" (POT) method from the field of extreme value theory offers one particularly elegant way to model the tails of an underlying loss distribution. Importantly, this approach alleviates the need to make potentially inappropriate parametric assumptions over the entire support of conditional loss distribution and is particularly effective for modeling distributions with heavy tails. The POT method is a natural solution for financial risk management problems where inference about heavy tailed losses is required.

To introduce the statistical foundations of the POT method, assume we have independent and identically distributed (i.i.d.) random variables Y, whose common distribution function is denoted by F. We will return to the case of non i.i.d. financial return data in Section 3. Let u be a high threshold value so that the set $Z = \{Y - u|Y > u\}$ is the excess loss above u. The number of exceedances above the threshold is $N_u = |Z|$, the cardinality of Z. The Pickands-Balkema-de Haan theorem states that the distribution of the excess loss converges to the generalized Pareto family of distributions for a suitably large threshold value (Pickands III et al., 1975; Balkema and De Haan, 1974),

$$\lim_{u \to Y^+} \sup_{0<z<(Y^+-u)} |Pr(Z < z) - G_{\xi,\sigma}(z)| = 0 \tag{3}$$

where Y^+ is the largest value in the sample space of Y and $G_{\xi,\sigma}$ is the generalized Pareto distribution function with shape and scale parameters ξ and σ respectively. The generalized Pareto distribution function is,

$$G_{\xi,\sigma}(z) = \begin{cases} 1 - \left(1 + \dfrac{\xi z}{\sigma}\right)^{-1/\xi}, & \text{for } \xi \neq 0,\ z > 0 \\ 1 - \exp\left(-\dfrac{z}{\sigma}\right), & \text{for } \xi = 0,\ z > 0 \end{cases} \tag{4}$$

When the conditional loss distribution is heavy tailed the shape parameter of the generalized Pareto distribution is greater than zero.

Assuming that the tail of F begins at u, then for some $\eta > u$ the following equality holds,

$$F(\eta) = 1 - (1 - F(u))(1 - F_u(\eta - u)), \quad \eta > u \tag{5}$$

where F_u represents the distribution of the excess loss above u. The POT method constructs an estimate for $F(\eta)$. The term $(1 - F(u))$ is approximated non-parametrically by the proportion of observations in the tail region above u, that is N_u/T. The distribution of the excess loss, $(F_u(\eta - u))$, is generalized Pareto by the Pickands-Balkema-de Haan theorem. Hence, the POT tail estimator is,

$$\hat{F}(\eta) = \hat{F}_Z(z) = 1 - \frac{N_u}{T}\left(1 + \frac{\xi z}{\sigma}\right)^{-1/\xi} \tag{6}$$

where $z = \eta - u$. Assuming that $\alpha > F(u)$, we may invert $\hat{F}_Z$ to estimate the tail quantiles and hence the $VaR^{(\alpha)}$ as,

$$\hat{F}_Z^{-1}(\alpha) = VaR^{(\alpha)} = u + \frac{\hat{\sigma}}{\hat{\xi}}\left[\left(\frac{T}{N_u}(1-\alpha)\right)^{-\hat{\xi}} - 1\right], \quad \alpha > F(u) \tag{7}$$

If $\xi < 1$, then an expression for the ES can be derived. For some $u' > u$, the excess loss $\{Y - u' | Y > u'\}$, also follows a generalized Pareto distribution with an identical shape parameter ξ but new scale parameter given by $\sigma + \xi(u' - u)$ (Coles, 2001). Using the expression for the mean of the generalized Pareto distribution, the $ES^{(\alpha)}$ can be written as,

$$\begin{aligned} ES^{(\alpha)} &= E[Z | Z > \hat{F}_Z^{-1}(\alpha)] \\ &= \frac{\hat{F}_Z^{-1}(\alpha) + \sigma - \xi u}{1 - \xi}, \quad \xi < 1,\ \alpha > F(u) \end{aligned} \tag{8}$$

Maximum Likelihood can be used to estimate the shape and scale parameters of the generalized Pareto distribution. The log-likelihood function is (Hosking and Wallis, 1987),

$$\ell(\xi, \sigma) = -N_u \ \log(\sigma) - \left(1 + \frac{1}{\xi}\right)\sum_{i=1}^{N_u} \log\left(1 + \frac{\xi z_i}{\sigma}\right)_+ \tag{9}$$

The choice of the initial threshold value is important in applied extreme value analysis and represents a trade-off between bias and variance. If the threshold is set too low then estimates of tail risk measures will be biased, since the asymptotic foundations of extreme value theory are likely to be violated. On the contrary, if the threshold is set too high, only a small number of observations are used to estimate the parameters of the generalized Pareto distribution. Both graphical methods (see, e.g., Davison and Smith, 1990), and statistical methods (see, e.g., Bader et al., 2018; Langousis et al., 2016), have been proposed to select a value for the initial threshold. In this study, since our primary contribution is to develop statistical advancements to existing risk models, we set u such that 10% of the sample are used as threshold exceedances at any given time. This value works well in finance and insurance applications (Chavez-Demoulin et al., 2014; Chavez-Demoulin and Embrechts, 2004).

3 Traditional GARCH-EVT Model

To apply the POT method introduced in Section 2, the observed losses must be approximately i.i.d. However, it is well known that the returns of financial assets exhibit conditional heteroskedasticity (Engle, 1982; Bollerslev, 1986). In other words, periods of high/low volatility tend to cluster together, causing clusters of extreme returns. These clusters violate the assumption that observations are independent. As such, a naive application of extreme value theory to the raw time series of negative log returns tends to produce poor estimates of the VaR and ES (see, e.g., Chavez-Demoulin et al., 2014).

To address the statistical issues associated with raw financial returns, McNeil and Frey (2000) develop the two-step GARCH-EVT risk model that first removes any serial dependence and conditional heteroskedasticity in raw log-returns using a GARCH model. The residuals from this GARCH model form an approximately i.i.d. series. In the second stage, POT extreme value theory is used to construct a model for the upper tail distribution of the negated standardized GARCH residuals.

We prefer to work with the 2-step GARCH-EVT model over alternative financial risk forecasting models for three reasons. First, the performance of this model has remained competitive for financial risk management applications since its inception (Nieto and Ruiz, 2016). Moreover, the model has even been shown to work well in situations where the GARCH dynamics are theoretically invalid or misspecified (Jalal and Rockinger, 2008). Second, the traditional GARCH-EVT method can easily be extended to incorporate more complex models for the mean and variance dynamics of log-returns and contemporary advancements in applied extreme value theory. Third, GARCH type models are well-established within the industry and the generalized Pareto extreme value distribution is used most often by risk management professionals to model the tails of the conditional loss distribution (Basel Committee on Banking Supervision, 2006).

Formally, in the first stage, McNeil and Frey (2000) model the mean and variance dynamics of log-returns using the following AR(1)-GARCH(1,1) process,

$$Y_t = \mu + \phi_1 Y_{t-1} + \epsilon_t \tag{10}$$

$$\upsilon_t^2 = \omega + \gamma_1 \epsilon_{t-1}^2 + \psi_1 \upsilon_{t-1}^2 \tag{11}$$

$$\epsilon_t = \upsilon_t e_t \tag{12}$$

where $e_t \sim N(0,1)$. In the second stage, the parameters of the generalized Pareto distribution are estimated via Maximum Likelihood using the set of standardized residuals that exceed the chosen threshold value. Using the 1-day ahead forecasts for the conditional mean and conditional variance from the first-stage GARCH model, denoted by $\hat{\mu}_{t+1}$ and $\hat{\upsilon}_{t+1}^2$ respectively, together with the expression for the inverse of the POT tail estimator (7), the GARCH-EVT model computes the 1-day ahead VaR forecast at confidence level α as,

$$\widehat{VaR}_{t+1}^{(\alpha)} = \hat{\mu}_{t+1} + \hat{\upsilon}_{t+1} \hat{F}_{Z_t}^{-1}(\alpha) \tag{13}$$

Similarly, using the expression for the conditional mean of the POT tail estimator (8), the GARCH-EVT model computes the 1-day ahead ES forecast at confidence level α as,

$$\widehat{ES}_{t+1}^{(\alpha)} = \hat{\mu}_{t+1} + \hat{\upsilon}_{t+1} E[Z | Z > \hat{F}_{Z_t}^{-1}(\alpha)] \tag{14}$$

3.1 Removing Dependence Among Extreme Returns

Since the generalized Pareto distribution parameters in the POT tail estimator are estimated using a sample of i.i.d. upper order statistics, it is important that dependence among the most extreme negative returns is removed in the first step of the GARCH-EVT risk model. Accordingly, our first modification to the traditional GARCH-EVT model is to use an asymmetric volatility model, namely the GJR-GARCH(1,1) model (Glosten et al., 1993), in place of the standard GARCH(1,1) model. This decision is supported by the recent findings of Trapin (2018), who suggest that volatility models endowed with a so-called "leverage" component are required to properly remove dependence among extreme returns[3]. The AR(1)-GJR-GARCH(1,1) model incorporates the leverage effect via the following econometric specification,

$$Y_t = \mu + \phi_1 Y_{t-1} + \epsilon_t \tag{15}$$

$$\upsilon_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \gamma_1 \epsilon_{t-1}^2 \mathbb{1}(\epsilon_{t-1} < 0) + \psi_1 \upsilon_{t-1}^2 \tag{16}$$

$$\epsilon_t = \upsilon_t e_t \tag{17}$$

[3]Alternative leveraged volatility models include the leverage HAR model and the leveraged HEAVY model. However, since the GJR-GARCH model was shown to perform at least as good as the more complex leveraged HEAVY model in the empirical study of Trapin (2018) and is a simple extension of the GARCH(1,1) model we prefer this approach.

where $\mathbb{1}(\cdot)$ is the indicator function. We demonstrate empirically that the GJR-GARCH model properly removes dependence among extreme returns in Section 5.

4 A Tail Estimator With Covariates

The POT tail estimator that is applied to the standardized residuals in the second step of the GARCH-EVT model uses time-invariant shape and scale parameters. These parameters are estimated using only information from a set of upper order statistics of the standardized residuals. We argue that economic and financial covariates may provide additional important information about the likelihood and severity of losses, thereby improving forecasts for the VaR and ES. Based upon this intuition we would like to incorporate covariates into the model for the upper tail of the standardized residuals, that is the POT tail estimator. One natural approach is to re-parameterize the shape and/or scale of the generalized Pareto distribution as a function of the relevant covariates.

Let $\mathbf{x}_t = (x_{t1}, \ldots, x_{tp})$ denote the p-dimensional vector of economic and financial covariates observed at time t. We assume that the data has been standardized to zero mean and unit variance. It is common to allow only the generalized Pareto distribution scale parameter to be time-varying (see, e.g., Davison and Smith, 1990; Coles, 2001). A constant shape parameter is expected to help control the variance of VaR and ES forecasts. In view of the constraint that $\sigma_t > 0$, we use the following log-linear model for the Generalized Pareto distribution scale parameter,

$$\sigma_t = \exp(\nu + \mathbf{x}_t^T \boldsymbol{\beta}) \tag{18}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is the coefficient vector.

Substitution of (18) into the equations outlined Section 2, yields a time-varying POT tail estimator whose behaviour now depends on a potentially large vector of economic and financial covariatres. The log-linear parameterization of σ_t belongs to the family of parametric Generalized Additive Models of Location, Scale and Shape (GAMLSS), introduced by Rigby and Stasinopoulos (2005). The revised log-likelihood function for the generalized Pareto distribution with time-varying scale parameter is,

$$\tilde{\ell}(\xi, \nu, \boldsymbol{\beta}) = -N_u \sum_{i=1}^{N_u} \nu + \boldsymbol{x}_i^T \boldsymbol{\beta} - (1 + \frac{1}{\xi}) \log \left(1 + \frac{\xi z_i}{\exp(\nu + \boldsymbol{x}_i^T \boldsymbol{\beta})} \right)_+ \tag{19}$$

4.1 Incorporating ℓ_1 Regularization

So far, our discussion regarding the time-varying scale parameter for the generalized Pareto distribution assumes that that the econometrician knows the optimal set of relevant covariates, and that this optimal set remains constant through time. Yet, the scale of modern economic and financial databases means that the number of potentially relevant covariates is now likely to be exceptionally large. Therefore, we argue that *ex-ante* knowledge about the optimal set of covariates and their time-varying behavior is no longer feasible.

One solution could be to include the entire universe of potentially relevant covariates. This approach is not practical for two reasons. First, the model will almost surely over-fit the training data since neither the log-linear parameterization in (19) nor the corresponding maximum likelihood objective function guarantee that any coefficient estimate will be zero. We demonstrate this circumstance empirically in Section 7. Second, in cases where $p > N_u$, that is the number of covariates exceeds the number of upper order statistics used for parameter estimation, standard maximum likelihood estimates cannot be obtained as the first-order condition is rank deficient. Traditional step-wise variable selection approaches are not viable for financial risk management, since the conditional loss distribution is time-varying and forecasts need to be updated at a high frequency. Hence, risk managers face a variable selection problem when incorporating covariates into risk models.

With the aforementioned arguments in mind, we hypothesize that the best risk model uses only a small number of potential covariates at any point in time, and that the statistical importance of covariates is time-varying. That is, we hypothesize that the best risk model is sparse. As such, a modeling framework that automatically performs both parameter estimation and variable selection is required. The ℓ_1 regularization applied to the coefficient vector $\boldsymbol{\beta}$ performs these tasks. In the context of linear regression, ℓ_1 regularization corresponds to the well known Lasso estimator Tibshirani (1996).The ℓ_1-regularized maximum likelihood parameter estimates for the generalized Pareto distribution with a time-varying scale parameter are then given by,

$$(\hat{\xi}^{(\ell_1)}, \hat{\nu}^{(\ell_1)}, \hat{\boldsymbol{\beta}}^{(\ell_1)}) = \arg\min_{\xi,\nu,\boldsymbol{\beta}} -\tilde{\ell}(\xi, \nu, \boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|_1 \tag{20}$$

where $\lambda \geq 0$ is a hyper-parameter that defines the weight on the ℓ_1-norm penalty. If $\lambda > 0$ then the ℓ_1-norm penalty has two effects on the solution, namely variable selection and coefficient shrinkage. First, some coefficient estimates are shrunk to exactly zero, effectively removing the corresponding covariates from the model.For a sufficiently large λ, all coefficient estimates are shrunk to zero and the regular time-invariant scale parameter is recovered. Second, all non-zero coefficient estimates in the active set, that is the set of non-zero coefficients $\mathcal{A}^{(\ell_1)} = \{\beta_j : \beta_j \neq 0, j = 1, \ldots, p\}$, are shrunk towards zero relative to their non-regularized counterparts.

Controlling the degree of both variable selection and coefficient shrinkage using a single hyper-parameter leaves room for improvement. Consequently, multiple extensions to the original Lasso estimator of the have been proposed. We use the form of an extension proposed by Meinshausen (2007), known as the relaxed Lasso, that introduces a second hyper-parameter $\phi \in (0, 1]$, to explicitly control the degree of coefficient shrinkage separately to variable selection. The analogous relaxed Lasso solution in our setting is given by,

$$(\hat{\xi}^{(relaxed)}, \hat{\nu}^{(relaxed)}, \hat{\boldsymbol{\beta}}^{(relaxed)}) = \arg\min_{\xi,\nu,\boldsymbol{\beta}\in\mathcal{A}} -\ell(\xi, \nu, \boldsymbol{\beta}) + \phi\lambda\|\boldsymbol{\beta}\|_1 \tag{21}$$

where $\boldsymbol{\beta} \in \mathcal{A}$ corresponds to requiring that $\beta_j = 0$ if $\beta_j \notin \mathcal{A}^{(relaxed)}$ for all $j = 1, \ldots, p$.

When $\phi < 1$, the degree of coefficient shrinkage is less than what is imposed by standard ℓ_1 regularization. The relaxed Lasso is particularly flexible, since it offers a large continuum of solutions and faster convergence rates when the number of noisy variables increases (Meinshausen, 2007). This is because traditional Lasso regularization requires large values for λ in the presence of many noise variables, which inadvertently imposes a large degree of shrinkage on all coefficient estimates. As such, it has been shown that the relaxed lasso tends to perform well in both high and low signal-to-noise ratio settings (Hastie et al., 2020).

Rather than solving Equation (21) directly, we implement a simplified version used by Hastie et al. (2020). Let $\hat{\boldsymbol{\beta}}_{\mathcal{A}^{(\ell_1)}}$ be the full p-dimensional solution obtained from maximizing the standard non-regularized log-likelihood function (19), retaining only the coefficients that would be selected by the corresponding ℓ_1 solution $\hat{\boldsymbol{\beta}}^{(\ell_1)}$ and replacing all remaining coefficients with zeros. The simplified solution is then,

$$\hat{\boldsymbol{\beta}}^{(relaxed)} = \phi\hat{\boldsymbol{\beta}}^{(\ell_1)} + (1 - \phi)\hat{\boldsymbol{\beta}}_{\mathcal{A}^{(\ell_1)}} \tag{22}$$

We refer to the GJR-GARCH-EVT risk model that uses $(\hat{\xi}^{(\ell_1)}, \hat{\nu}^{(\ell_1)})$ from (20) and $\hat{\boldsymbol{\beta}}^{(relaxed)}$ from (22) as the ℓ_1-regularized GARCH-EVT model. The simplified outlined above estimator eases the computational burden of selecting the optimal hyper-parameters (λ, ϕ).

Since we are working with log-return data that are naturally heavy tailed, we impose the inequality constraint that $0 < \xi < 1$ when maximizing the likelihood. This inequality constraint also ensures the mean of the generalized Pareto distribution is finite so that the ES can always be estimated. As starting values for the numerical minimization routines, we set $\beta_1 = \cdots = \beta_p = 0.1$ and take the best results for ξ obtained using the generalized Pareto distribution parameter estimation methods of Hosking (1990), Zhang and Stephens (2009) and Zhang (2010).

4.2 Hyper-parameter Tuning

The ℓ_1 regularized maximum likelihood parameter estimates are obtained using a sample of upper order statistics from the standardized residual series. Since this sample of order statistics are assumed to be approximately i.i.d. we use 5-fold cross validation[4] to select the value of the hyper-parameters (λ, ϕ). We search for the optimal value of λ using 100 logarithmic spaced values between 0.1 and 20 and for the value of ϕ using 100 equally spaced values between 0.01 and 1.

Cross-validation necessitates the definition of a loss function used to compare each candidate set of hyper-parameters. Our goal is to select the generalized Pareto distribution that provides the best estimate for the upper tail of standardized GJR-GARCH residuals. This distribution should then produce the most accurate VaR and ES forecasts across a wide range of confidence levels. With this goal in mind, we compute the root-mean-squared error loss between the estimated generalized Pareto CDF and empirical CDF over all standardized residuals in the holdout fold.

5 Data

We backtest the VaR and ES forecasts from our ℓ_1-regularized GARCH-EVT model using a sample of 25 U.S. equities in the S&P100 index. This sample includes stocks from the financial, technology, healthcare and energy sectors. The ticker symbols for each equity we study are listed in Table 1. For each equity we obtain stock price data from Compustat over a sample period of 4085 trading days from 2003-09-10 to 2019-11-29.

Table 1 reports descriptive statistics for the 1-day negative log-returns of each equity we study. We also compute two measures that characterize the behavior of extreme returns, namely the tail index and the extremal index. The tail index quantifies heavy-tailed behavior, with Larger values for the tail index indicate heavier upper tails. We use the non-parametric Hill estimator (Hill, 1975), applied to the standardized residuals from a GJR-GARCH(1,1) model. A sample of 408 upper order statistics is used to estimate the tail index for each equity in the sample, which corresponds to a threshold value set at the 90^{th} quantile of the standardized residual series. All equities we study exhibit considerable excess kurtosis and all estimates of the tail index are comfortably above zero. Hence, the returns exhibit heavy tails and the constraint $\xi > 0$, used in the optimization routines is well justified.

The extremal index, denoted by $\theta(\cdot) \in [0, 1]$, measures the the extent to which extreme returns tend to cluster together. When $\theta(\cdot) = 1$, extreme observations are independent. When $\theta(\cdot) < 1$, extreme returns exhibit some degree of serial dependence[5]. We compute $\theta(\cdot)$ using the "intervals" estimator of Ferro and Segers (2003) for the raw negative log return series, the standardized residuals from a GARCH(1,1) model and the standardized residuals from a GJR-GARCH(1,1) model using a threshold set at the 90^{th} quantile for each series.

The extremal index of raw negative log returns is appreciably below 1 for all equities, suggesting strong dependence among the largest negative returns. However, the extremal index of standardized GJR-GARCH residuals are all close to or equal to 1. The value of the extremal index for the standardized residuals from the GARCH(1,1) model are always smaller than those from the GJR-GARCH(1,1) model for all equities we study. As such, Table 1 demonstrates that the GJR-GARCH model does an adequate job of removing dependence among the largest negative log returns[6].

[4]We experimented with an information criterion based approach but found that 5-fold cross-validation unanimously selected models which produced substantially more accurate VaR and ES forecasts.

[5]More specifically, Ledford and Tawn (2003) show that $\theta(\cdot) < 1$ implies asymptotic dependence of extreme observations for at least 1 lag.

[6]The 95% confidence intervals computed using the bootstrap method of Ferro and Segers (2003) for the extremal index of the standardized GJR-GARCH residual series all contain 1.

To study the forecasting performance of risk models under different market conditions, we construct two backtests. The first backtest uses a rolling 2000-day training sample beginning on 2003-09-10 to compute VaR and ES forecasts for an out-of-sample period of 2075 days, ending on 2019-11-29. We refer to this first backtest sample as the normal backtest, since it covers a period of normal market volatility, with only a handful of daily returns exceeding 5% in absolute value. It is also important to backtest risk models over periods of financial distress, as this is when accurate risk forecasts are particularly important. As such, the second backtest uses a rolling 1000-day training sample, beginning on 2003-09-10 to generate VaR and ES forecasts for and an out-of-sample period of 1084 days, ending on 2011-12-31. This second backtest includes the entire Global Financial Crisis period, allowing us to study how each model performs during a period of prolonged financial market distress. Moreover, the financial crisis backtest covers the largely unanticipated "Black Monday" crisis, where U.S. debt was downgraded.

5.1 Description of the Economic and Financial Covariates

We use a set of 24 covariates, computed at a daily frequency, in the log-linear model for the generalized Pareto distribution scale parameter. Table 2 provides a brief description of each covariate together with information about the data source. We study four broad classes of covariates.

The first class of variables are idiosyncratic measures of trading activity and the market microstructure. These covariates are computed from millisecond-level intraday trade and quote data. Multiple studies document a relationship between stock returns and measures of intraday transaction activity (Amihud, 2002; Brennan et al., 2012). Moreover, a growing set of literature acknowledges the bid-ask spread and the quoted depth, derived from high frequency limit order book records, contains information about idiosyncratic volatility, as these variables reflect demand/supply pressures and the risk management behaviour of intermediaries (Foucault et al., 2007; Budish et al., 2015). As such, it is conceivable that covariates derived from intraday transaction data may be informative about short-term variation in extreme returns.

To the best of our knowledge, the literature has not widely considered the role of variables computed from high-frequency intraday transaction data for risk management applications. The Trade and Quote (TAQ) database is used to compute each measure of intraday trading activity for each equity in our sample. The TAQ data for each equity is first filtered using the rules outlined by Holden and Jacobsen (2014). Trade direction is inferred using the Lee and Ready algorithm (Lee and Ready, 1991). Since our focus is on predicting future downside risk, we only include covariates related to selling activity and liquidity measures for the bid side of the order book.

Individual equities are now associated with numerous alternative investment products, such as options and futures derivatives contracts. Options contracts are inherently forward-looking and therefore provide idiosyncratic information about the future expectations for a stock (Pan and Poteshman, 2006). Moreover, it is known that informed traders transact in both stock and option markets, so information from both markets has the potential to be relevant for understanding future stock returns (Chakravarty et al., 2004). As such, we also include the daily put option open interest as a covariate.

The second class of covariates are measures of economic uncertainty. Pastor and Veronesi (2012) construct a model where shocks about economic policy uncertainty affect the price of financial assets. Moreover, if economic uncertainty affects future investment decisions, then this information should be related to excess returns (Brogaard and Detzel, 2015). Among others, empirical support for the relationship between economic uncertainty and asset returns is given by Antonakakis et al. (2013) and Brogaard and Detzel (2015). To measure economic uncertainty we use the Economic Policy Uncertainty (EPU) index and the Monetary Policy Uncertainty (MPU) index of Baker et al. (2016). Generally speaking, both indices are derived using the the frequency of news articles that mention words associated with economic, political and monetary uncertainty.

The third class of covariates we consider are measures of U.S. equity market volatility. To capture future expectations about volatility, we use the Chicago Board Options-Exchange VXO index. This index captures the market's 30-day expectation of future volatility for the S&P100. Data for each index is obtained from the Federal Reserve Bank of St.Louis's Economic Research Database (FRED). Our second measure of volatility for the U.S. market is the news based volatility index of Baker et al. (2019). The final set of covariates are broad measures of U.S. financial market stress and funding market liquidity. We use the TED spread, the spread between an index of high yield non-investment grade bonds and treasury yields and smoothed U.S. rescission probabilities computed from a Markov switching model. Data for each of these financial market stress measures are obtained from FRED.

6 Evaluating VaR and ES Forecasts

The performance of a financial risk forecasting model is assessed by comparing the VaR and ES forecasts, with the out-of-sample realizations of the negative log return. We use multiple statistical tests to assess adequate levels of risk coverage together with consistent scoring functions to evaluate the performance of VaR and ES forecasts. Each evaluation criteria is discussed briefly below.

6.1 VaR Forecast Quality Measures

A VaR violation is defined as the case where $Y_{t,h} > VaR_{t,h}^{(\alpha)}$. Using this definition, the VaR violation indicator sequence for a prescribed level of confidence α and forecast horizon h is,

$$I^{(\alpha)} = \{\mathbb{1}(Y_{t,h} > VaR_{t,h}^{(\alpha)}), t = 1, \dots, T\} \tag{23}$$

Tests of unconditional coverage assess whether the observed number of VaR violations matches the expected number of violations for the prescribed level of confidence over the backtest. The first unconditional coverage test, introduced by Christoffersen (1998), tests the null hypothesis that $E[I_t^{(\alpha)}] = \alpha$, against the 2-tailed alternative using a likelihood ratio test. The test statistic is,

$$LR_{uc} = -2\log[((1-\alpha)^{n_0}\alpha^{n_1})/(1-\hat{\kappa})^{n_0}\hat{\kappa}^{n_1}] \sim \chi^2_{(1)} \tag{24}$$

where $n_1 = \sum_{t=1}^{T} I_t^{(\alpha)}$, $n_0 = T - n_1$ and $\hat{\kappa} = n_1/(n_0 + n_1)$.

From the perspective of a prudential regulator an incorrect level of unconditional coverage because $E[I_t^{(\alpha)}] < \alpha$ is preferable to the alternative that $E[I_t^{(\alpha)}] > \alpha$. This is because the former scenario ensures that downside risk is minimized, albeit by holding an overly conservative amount of capital to cover potential future losses. Hence, it is important to identify specific cases where a risk model systematically underestimates the VaR. Our second unconditional coverage test performs a one-tailed test against the alternative hypothesis that $E[I_t^{(\alpha)}] > \alpha$ using the approach introduced by Ziggel et al. (2014). The test statistic is,

$$MCS_{uc} = \sum_{t=1}^{T} I_t^{(\alpha)} + \epsilon \tag{25}$$

where $\epsilon \sim N(0, 0.001)$ and is used to break ties between test values. Monte-Carlo simulation is used to obtain the distribution of the test statistic under the null hypothesis.

The following two statistical tests assess whether VaR violations are independent and hence whether or not a risk model provides a correct level of coverage when conditioned on time. Let π_{ij} and n_{ij} denote the probability and number of observations corresponding to a transition for the VaR violation indicator sequence from value i on day $t-1$ to value j on day t. Christoffersen (1998) tests the null hypothesis that

$\pi_{01} = \pi_{11}$, that is that past VaR violations do not contain information about current and future violations. The transition probabilities of interest are estimated as,

$$\hat{\pi}_{01} = \frac{n_{01}}{n_{00} + n_{01}}, \quad \hat{\pi}_{11} = \frac{n_{11}}{n_{10} + n_{11}} \tag{26}$$

Defining $\pi = (n_{01} + n_{11})/(n_{00} + n_{01} + n_{10} + n_{11})$, the likelihood ratio test statistic is,

$$LR_{ind} = -2\log[((1-\hat{\pi}_{01})^{n_{00}}\hat{\pi}_{01}^{n_{01}}(1-\hat{\pi}_{11})^{n_{10}}\hat{\pi}_{11}^{n_{11}})/((1-\pi)^{n_{00}+n_{10}}\pi^{n_{01}+n_{11}})] \sim \chi^2_{(1)} \tag{27}$$

We define a matrix $\mathbf{A}$ to hold 4-lags of the VaR violation indicator sequence and 4 lags of the negative log return at each out of sample forecast day over the backtest. Further, let the de-meaned VaR violation indicator sequence be defined as $\tilde{I} = \{I^{(\alpha)} - \alpha\}$. The Dynamic Quantile test, introduced by Engle and Manganelli (2004), evaluates the null hypothesis that $\tilde{I}$ is linearly unrelated to the information in $\mathbf{A}$ using the following test statistic,

$$DQ = \frac{\tilde{I}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\tilde{I}}{r\alpha(1-\alpha)} \sim \chi^2_{(r)} \tag{28}$$

where $r = \text{rank}(\mathbf{A})$.

Finally, the unconditional coverage and independence tests of Christoffersen (1998) can be combined to assess the so-called conditional coverage of a risk model. The likelihood ratio test statistic for the conditional coverage test is,

$$LR_{cc} = (LR_{uc} + LR_{ind}) \sim \chi^2_{(2)} \tag{29}$$

6.2 ES Forecast Quality Measures

We implement three unconditional risk level coverage tests for the ES forecasts that make no parametric assumptions about the tails of conditional loss distribution. First, we use the unconditional converge test introduced by McNeil and Frey (2000), who define an exceedance residual sequence as,

$$\left\{\frac{y_{t+h} - ES^{(\alpha)}_{t+h}}{VaR^{(\alpha)}_{t+h}}\right\}, \quad y_{t+h} > VaR_{t+h}$$

Since the $ES^{(\alpha)}$ estimates the mean of the truncated tail distribution above the corresponding $VaR^{(\alpha)}$ level, McNeil and Frey (2000) test the null hypothesis that the exceedance residual sequence has a mean of zero, against the alternative hypothesis that the the mean is greater than zero. This alternative hypothesis corresponds to the situation where the ES is systematically underestimated. To account for the presence of any dependence between consecutive ES forecasts we use the circular block bootstrap (Jalal and Rockinger, 2008). The optimal block length is selected following Patton et al. (2009).

The second test of the unconditional coverage for ES forecasts was introduced by Acerbi and Szekely (2014) and is derived from re-writing the expected shortfall as the following unconditional expectation,

$$\begin{aligned} ES^{(\alpha)}_t =& E\left[\frac{y_t I^{(\alpha)}_t}{\alpha}\right] \\ 0 =& E\left[\frac{y_t I^{(\alpha)}_t}{\alpha ES^{(\alpha)}_t}\right] - 1 \end{aligned}$$

which suggests the following test statistic,

$$AS_{uc} = \frac{1}{T\alpha}\sum_{t=1}^{T}\frac{y_t I^{(\alpha)}_t}{ES^{(\alpha)}_t} - 1 \tag{30}$$

The null hypothesis states that $AS_{uc} = 0$, and hence ES forecasts correctly estimate the mean of the truncated tail distribution. The alternative hypothesis states that $AS_{uc} > 0$, that is, the ES is systematically underestimated. Acerbi and Szekely (2014) recommend using a Student's-t distribution with 3 degrees of freedom to approximate the distribution of the test statistic.

Under the upcoming Basel III regulations, backtesting the ES is still based upon an assessment of the VaR forecasts produced by the same forecasting model (Basel Committee on Banking Supervision, 2016). Acerbi and Tasche (2002), showed that the ES is equivalent to the average of all VaR estimates at and beyond the α-quantile of the conditional loss distribution. Using this equivalence, Kratz et al. (2018) propose to backtest ES forecasts by simultaneously testing the unconditional coverage rates of multiple VaR forecasts. Specifically, the authors approximate the ES using K different VaR estimates at equally spaced and increasing levels of confidence, that is,

$$ES^{(\alpha)} \approx \frac{1}{K}\sum_{i=1}^{K} VaR^{(v_i)}, \quad v_i = \alpha + \frac{i-1}{K}(1-\alpha) \tag{31}$$

Let $V_t = \sum_{i=1}^{K} I_t^{(v_i)}$, be the count of different VaR levels violated at time t, and define $O_i = \sum_{t=1}^{T} \mathbb{1}(V_t = i), i = 1, \ldots, K$. Then, the vector $(O_1, \ldots, O_K)$ follows a Multinomial distribution. To test the null hypothesis of correct unconditional coverage at all VaR levels simultaneously, against the alternative hypothesis of incorrect unconditional coverage for at least one VaR level. Kratz et al. (2018) propose the following test statistic,

$$S_K = c\sum_{i=0}^{K} \frac{(O_{i+1} - T(v_{i+1} - vj))^2}{T(v_{i+1} - v_i)} \sim \chi^2_w \tag{32}$$

where $w = cE(S_K)$, $c = E(S_K)/var(S_K)$ and

$$E(S_K) = K, \quad var(S_K) = 2K - \frac{K^2 + 4K + 1}{T} + \frac{1}{T}\sum_{i=1}^{K} \frac{1}{v_{i+1} - v_i}$$

We set $K = 4$ as recommended by Kratz et al. (2018).

6.3 Scoring Functions

In addition to assessing the quality of VaR and ES forecasts using tests of risk coverage and independence, Nolde et al. (2017) advocate for the use of consistent scoring functions to compare forecasts. In expectation, a consistent scoring function is minimized by issuing the correct forecast of a given risk measure. Broadly speaking, a consistent scoring function ensures that, on average, a misspecified risk model will not outperform a correctly specified risk model, facilitating an unbiased comparison of forecasting models [7].

Model comparisons based upon a scoring function necessitate that the risk measure being forecast is elicitable (Gneiting, 2011). The VaR is elicitable. By itself, the ES is not elicitable but is instead jointly elicitable with the VaR for the same level of confidence. That is the 2-dimensional vector $(VaR^{(\alpha)}, ES^{(\alpha)})$ is elicitable. For the VaR we use well known asymmetric piecewise linear scoring function (Gneiting, 2011),

$$S(VaR^{(\alpha)}_{t,h}, y_{t,h}, \alpha) = (\mathbb{1}(VaR^{(\alpha)}_{t,h} \geq y_{t,h}) - \alpha)(VaR^{(\alpha)}_{t,h} - y_{t,h}) \tag{33}$$

We use the so-called (1/2)-homogeneous scoring function outlined by Nolde et al. (2017) to jointly evaluate

[7] For a more detailed discussion of consistent scoring functions see Nolde et al. (2017) and Gneiting (2011).

the VaR and ES,

$$S(VaR_{t,h}^{(\alpha)}, ES_{t,h}^{(\alpha)}, y_{t,h}, \alpha) = \mathbb{1}(y_{t,h} > VaR_{t,h}^{(\alpha)}) \frac{y - VaR_{t,h}^{(\alpha)}}{2\sqrt{ES_{t,h}^{(\alpha)}}} + (1-\alpha) \frac{VaR_{t,h}^{(\alpha)} + ES_{t,h}^{(\alpha)}}{2\sqrt{ES_{t,h}^{(\alpha)}}} \tag{34}$$

The value of these scoring functions are averaged over the out-of-sample forecast period and scaled by $1/(1-\alpha)$ to produce a single summary value.

7 Out-of-Sample Forecast Performance

For each of the two backtests, we compute VaR and ES forecasts on a rolling basis, so that all information in the training sample up to and including time t is used to generate a VaR and ES forecast for time $t+h$. We consider VaR forecasts for $\alpha \in \{0.975, 0.99, 0.995\}$ and ES forecasts for $\alpha \in \{0.975, 0.99\}$, because these are the most important levels of confidence in practice[8]. For each aforementioned level of confidence, we consider VaR and ES forecasts for $h \in \{1, 10\}$. For each day in the backtests, we re-estimate the optimal values of the two hyper-parameters for the ℓ_1-regularized GARCH-EVT model (λ, ϕ), using 5-fold cross-validation applied to the current rolling training sample.

The principle benchmark forecasting model is the traditional GARCH-EVT model, implemented exactly as proposed by McNeil and Frey (2000). To demonstrate our hypothesized need for variable selection and shrinkage, the second benchmark risk model is a non-regularized GARCH-EVT model that includes all covariates but does not use ℓ_1 regularization. Since GARCH type risk models are exceptionally popular in the literature and in practice, our third benchmark risk model is a GJR-GARCH(1,1) model with a students-t distribution for the errors. Finally, we implement three popular risk models from prior literature, namely, the Hawkes POT model (Chavez-Demoulin et al., 2005; Chavez-Demoulin and McGill, 2012), the CaViaR model (Engle and Manganelli, 2004), and the CARE model (Taylor, 2008). Details regarding the implementation of these three models are discussed in Appendix A

Each of the forecast results tables presented in this section report the number of times each null hypothesis was rejected at the 5% level for the statistical tests used to evaluate VaR/ES forecasts criteria discussed in Section 6.1 and 6.2. We also report the total number of rejections across all statistical tests as a summary measure of forecast quality. The final column of each table reports the mean value of the VaR and the joint VaR/ES scoring function scaled by $1/(1-\alpha)$. For all forecast performance criteria reported in the tables smaller values indicate better forecasting performance.

The total rejection count and/or the average value of the scoring functions are larger, and in most cases considerably larger, for VaR and ES forecasts produced by the non-regularized GARCH-EVT model, compared to the ℓ_1-regularized model. This result demonstrates that variable selection mechanisms iare essential to produce reasonable risk forecasts when covariates are incorporated into the GARCH-EVT model. On average, we find that the active set of the ℓ_1-regularized GARCH-EVT model is empty on approximately 44% of days in both the normal and financial crisis backtests. That is, almost half of the time the best GARCH-EVT risk model uses a constant scale parameter. The average size of the active set in any given equity at any given time is 5 in the normal backtest and 6 in the financial crisis backtest. Hence, only a handfull of our covariates are selected by the ℓ_1-regularized GARCH-EVT model at any point in time. Together, these statistics support our hypotheses that the best risk model is sparse and that the predictive importance of covariates is time-varying.

[8] The multinomial test of unconditional coverage for ES forecasts also requires us to compute VaR forecasts for $\alpha \in \{0.9812, 0.9925, 0.9875, 0.9936, 0.9975\}$. These VaR forecasts are used solely to backtest the $ES^{(0.975)}$ and $ES^{(0.99)}$ forecasts.

7.1 1-Day Ahead Forecasts

Table 3 reports the 1-day ahead VaR forecast results over the normal backtest. The 1-day ahead VaR forecasts from the ℓ_1-regularized GARCH-EVT model are associated with the smallest number of total rejections at the $\alpha = 0.99$ level of confidence, which is the prescribed level of confidence set by regulators. Moreover, the ℓ_1-regularized GARCH-EVT model produces 1-day ahead VaR forecasts that are associated with a lower total number of rejections for the $\alpha = 0.975$ and the $\alpha = 0.99$ levels, when compared to the traditional GARCH-EVT model. In aggregate across all three levels of confidence, the 1-day ahead VaR forecasts from the ℓ_1-regularized GARCH-EVT model are associated with 28 rejections, the smallest value for all benchmark models we consider.

The 1-day ahead VaR forecasts from the traditional GARCH-EVT model are associated with a total of 46 rejections. However, the mean value of the VaR scoring function for the ℓ_1-regularized GARCH-EVT model is marginally larger then that of the traditional GARCH-EVT model. The MCS_{uc} statistic shows that neither of the GARCH-EVT risk models systematically underestimate risk over the normal backtest. Taken together, Table 3 demonstrates that the ℓ_1-regularized GARCH-EVT model is an exceptionally competitive extension of the GARCH-EVT model during periods of normal market volatility.

With the exception of the non-regularized GARCH-EVT model, the Hawkes POT model is associated with the largest total number of rejections across all levels of confidence. The VaR forecasts from the Hawkes POT model generally have acceptable levels of unconditional coverage and do not systematically underestimate the VaR according to the MCS_{uc} statistic. However, VaR violations from the Hawkes POT model often violate the tests of independence, suggesting that the Hawkes POT model does not provide correct levels of coverage when conditioned on time. The GJR-GARCH model with student's-t errors is always associated with the smallest value of the scoring function and performs well in terms of the total number of rejections at the $\alpha = 0.975$ and $\alpha = 0.99$ levels.

Table 4 reports the 1-day ahead ES forecast results over the normal backtest. The 1-day ahead ES forecasts produced by the ℓ_1-regularized GARCH-EVT risk model marginally underperform the 1-day ahead ES forecasts from the traditional GARCH-EVT risk in terms of the total number of rejections at either level of confidence. The higher number of rejections for the bootstrap test of McNeil and Frey (2000), suggests that the ℓ_1-regularized GARCH-EVT risk model underestimates the ES more often than the traditional GARCH-EVT risk model during periods of normal market volatility. However, the mean value of joint VaR and ES scoring function for the ℓ_1-regularized GARCH-EVT model and the traditional GARCH-EVT model are approximately equal at the $\alpha = 0.975$ level.

The GJR-GARCH model generate 1-day ahead ES forecasts that are associated with the smallest number of total rejections at both levels of confidence. The mean value of the joint VaR and ES scoring function for the ℓ_1-regularized GARCH-EVT model and the traditional GARCH-EVT model are smaller than competing models across both levels of confidence. Interestingly, despite the small number of total rejections, the value of the average scoring function for the GJR-GARCH model is considerably higher than the average scoring function values for all other risk models. This result may suggest that the GJR-GARCH model tends to overestimate the ES during periods of normal market volatility.

Table 5 reports the 1-day ahead VaR forecast results over the financial crisis backtest. We find that the ℓ_1-regularized GARCH-EVT risk model outperforms competing benchmark models in terms of the number of total rejections for the $\alpha = 0.975$ and $\alpha = 0.99$ levels. In aggregate, across the three levels of confidence, 1-day ahead VaR forecasts from the ℓ_1-regularized GARCH-EVT model are associated with only 37 rejections while the 1-day ahead VaR forecasts from the next best model, the traditional GARCH-EVT model, are associated with 51 rejections. The mean value of the VaR scoring function for the ℓ_1-regularized GARCH-EVT model is lower than that of the traditional GARCH-EVT model at all three levels of confidence.

The 1-day ahead VaR forecasts from the CaViaR, CARE, Hawkes POT and GJR-GARCH Student's-t model are inadequate over the financial crisis backtest. The CaViaR and CARE models produce the largest number of rejections across all levels of confidence and do not provide correct levels of unconditional coverage in the majority of equities. The Hawkes POT model again provides acceptable levels of unconditional risk coverage but fails to satisfy tests for the independence of VaR violations. The GJR-GARCH model with students-t errors is associated with a higher aggregate number of rejections and underestimates the 1-day ahead VaR more often than either of the GARCH-EVT models. Together, these results demonstrate that risk models which use extreme value distributions are particularly effective during periods of financial distress.

Table 6 reports the 1-day ahead ES forecast results over the financial crisis backtest. We find that the 1-day ahead ES forecasts from the ℓ_1-regularized GARCH-EVT model are competitive to those from the traditional GARCH-EVT model. At the $\alpha = 0.975$ level, the prescribed level of confidence set by regulators, the ℓ_1-regularized GARCH-EVT model marginally outperforms the traditional GARCH-EVT model in terms of both the total rejection count, and the mean value of the joint VaR and ES scoring function. The 1-day ahead ES forecasts from both the ℓ_1-regularized GARCH-EVT model and the traditional GARCH EVT based risk models perform best over the financial crisis backtest.

The 1-day ahead ES forecasts from the CARE model and Hawkes POT model are associated with the largest number of rejections in the financial crisis backtest. Hence, we find that conditional auto-regressive risk models do not accurately forecast risk during periods of financial market stress. Interestingly, across both backtests and both levels of confidence, we only observe a single rejection for the AC_{UC} test. This result suggests that the direct test for unconditional coverage test introduced by Acerbi and Szekely (2014) may not adequately distinguish between effective risk models.

In summary, we demonstrate that financial and economic covariates have the potential to improve short term financial risk forecasts if incorporated via variable selection mechanisms, such as ℓ_1 regularization. The addition of covariates routinely improves forecasts of the 1-day ahead VaR from the the traditional GARCH-EVT model. These improvements in VaR forecasts are particularly evident during periods of financial market stress and at the $\alpha = 0.99$ level of confidence. The ES forecasts issued by the ℓ_1-regularized GARCH-EVT model are competitive with those produced by the traditional GARCH-EVT model over the financial crisis backtest, particularly at the $\alpha = 0.975$ level that is prescribed by regulators.

7.2 10-Day Ahead Forecasts

To generate 10-day ahead VaR and ES forecasts, we rely on the widely used "square-root-of-time" scaling rule[9]. Table A.0.3 reports the 10-day ahead VaR forecast results over the normal backtest and the financial crisis backtest. We only focus on the results for unconditional coverage tests when studying 10-day ahead risk forecasts since all risk models reject tests of independence and conditional coverage due to the long term nature of the forecast.

In aggregate, we find that the 10-day ahead VaR forecasts from the ℓ_1-regularized GARCH-EVT model are associated with a higher number of total rejections and higher value of the mean VaR scoring function when compared to the traditional GARCH-EVT model over the normal backtest at all three levels of confidence. Over the financial crisis backtest, the ℓ_1-regularized GARCH-EVT model marginally outperforms the traditional GARCH-EVT model at the $\alpha = 0.99$ level.

Table 9 reports the 10-day ahead ES forecast results over the normal backtest and the financial crisis backtest. We find that the 1-day ahead ES forecasts produced by the ℓ_1-regularized GARCH-EVT model do not outperform those from the traditional GARCH-EVT model over the normal backtest. Over the financial crisis backtest, the 1-day ahead ES forecasts from the ℓ_1-regularized GARCH-EVT model are associated with a smaller number of rejections than those from the traditional GARCH-EVT model at the $\alpha = 0.975$ level. The mean value of the joint VaR and ES scoring functions for the ℓ_1-regularized GARCH-EVT model

[9] We also examine the performance of 10-day ahead risk forecasts generated using the simulation approach of McNeil and Frey (2000) for the GARCH type models but find that this approach tends to produce inferior forecasts.

and the traditional GARCH-EVT model are approximately equivalent and are smaller than those of all other benchmark models.

In summary, the ℓ_1-regularized GARCH-EVT model produces 10-day ahead VaR and ES forecasts that generally do not outperform the forecasts from the traditional GARCH-EVT model. As such, we conclude that information from our set of economic and financial information is most valuable for short term 1-day ahead risk forecasting.

8 When are Covariates Most Valuable?

We have shown that ℓ_1-regularized GARCH-EVT model can produce competitive 1-day ahead VaR and ES forecasts, particularly over the financial crisis backtest. As such, we now investigate precisely when economic and financial covariates are most valuable for financial risk forecasting. We also study how the size and predictive duration of the active coefficient set changes during the periods where covariates are particularly useful. This analysis facilitates a deeper understanding of variable selection and shrinkage mechanisms for financial risk forecasting.

Figure 1 plots the difference between the VaR scoring function and the joint VaR and ES scoring function for 1-day ahead forecasts from the ℓ_1-regularized GARCH-EVT model and the traditional GARCH-EVT model. We study the $\alpha = 0.99$ level for VaR forecasts and the $\alpha = 0.975$ level for ES forecasts as these are the most important levels of confidence in practice. A negative difference implies that forecasts from the ℓ_1-regularized GARCH-EVT model are more accurate than those from the traditional GARCH-EVT model. Two notable financial market crisis events are shaded. The first crisis event is the Global Financial Crisis. The second crisis event is a 10-day period centered around the "Black Monday" market crash on 8th August, 2011 when U.S. debt was downgraded from AAA to AA+. We find that all of the largest negative differences in the VaR scoring function and the joint VaR and ES scoring function occur within the two shaded market crisis events. The mean difference in both scoring function values is negative in the shaded regions. Figure 1 therefore suggests that the ℓ_1-regularized GARCH-EVT model uses exogenous economic and financial information to generate more accurate risk forecasts during financial crisis events, when accurate financial risk forecasts are most valuable.

If covariates are particularly valuable during periods of financial distress, as indicated above, then it is informative to study temporal dynamics in the behaviour of the active coefficient set that is selected by the ℓ_1-regularized GARCH-EVT model during these periods. Accordingly, we estimate two generalized linear models. The first model studies the size of the active coefficient set using Poisson regression. The second model studies the predictive duration of the typical covariate using Weibull regression. The predictive duration of a covariate is defined as the number of consecutive days that a covariate remains in the active set at each out-of-sample forecast day[10]. Predictive durations are averaged over all selected covariates at each date.

In both regression models we include three independent variables. The first independent variable is an indicator that takes the value of 1 for all observations between 1st July, 2008 and 30th July, 2009, the global financial crisis period, and 0 otherwise. This indicator variable identifies a period of prolonged financial market distress. The second independent variable of interest is an indicator that takes the value of 1 for all observations between 28th July, 2011 and 18th August, 2011, a 10-day period centered around the "Black Monday" market crash and zero otherwise. This indicator variable identifies a short-term market crash

[10]For example, if sell volume is used to make a forecast at date 2008-06-01 but is not used to make a risk forecast at date 2008-06-02 then we say that sell volume had a predictive duration of 1 for the date 2008-06-01. On the other hand, if sell volume is used to make a risk forecast at date 2008-06-01 and is used to make a risk forecast at date 2008-06-02, 2008-06-03 and 2008-06-04, then we say that sell volume had a predictive duration of 4 for the date 2008-06-01. The predictive duration is effectively a survival time, where the death event occurs when the covariate drops out of the active set of the ℓ_1-regularized GARCH-EVT model at a future date.

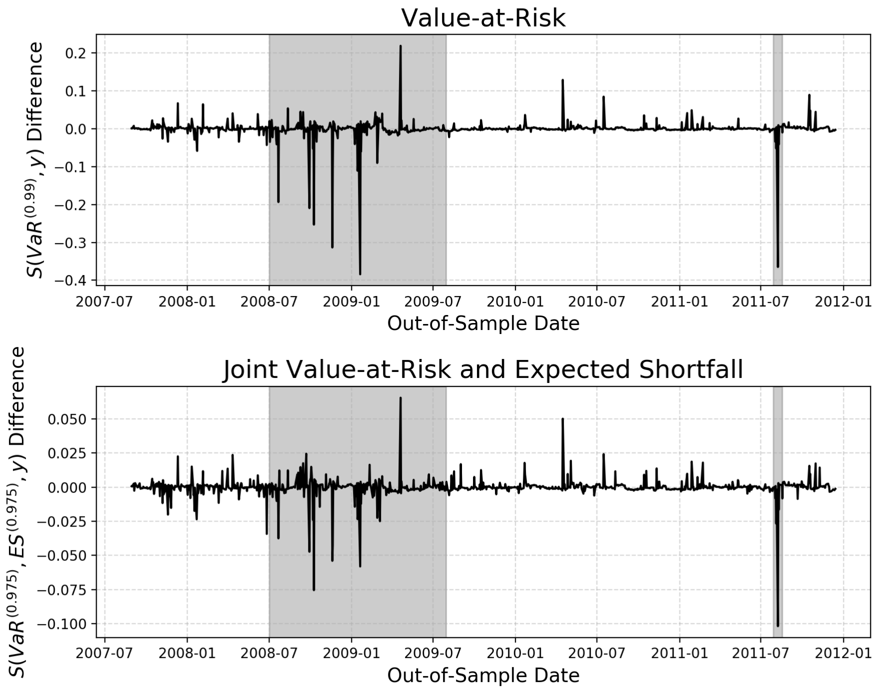

Figure 1: Time Series Difference between scoring functions for 1-day ahead forecasts for the ℓ_1-regularized GARCH-EVT model and the traditional GARCH-EVT model. The first grey shaded region covers the Global Financial Crisis period. The second grey shaded region covers a 10-day period before and after the "Black Monday" crash on 8th August, 2011.

event. The third independent variable of interest is an indicator variable that takes the value of 1 if the 1-day ahead variance forecast from the GJR-GARCH(1,1) model is above the median 1-day ahead variance forecast over the entire financial crisis backtest and zero otherwise. We include all independent variables simultaneously in each regression model.

Table 10 reports the coefficient estimates from the two aforementioned regression models. We find that both the size of the active coefficient set and the predictive duration of the typical covariate are larger during the global financial crisis period. That is, the ℓ_1-regularized GARCH-EVT model uses more covariates over a longer consecutive forecasting period when financial markets are in distress. This result provides further evidence that the set of economic and financial variables we consider in this study are most valuable during times of prolonged financial distress. Interestingly, the ℓ_1-regularized GARCH-EVT model uses fewer covariates and for a shorter predictive duration during the period surrounding the "Black Monday" market crash event. These differences in the behaviour of the active coefficient set further highlight the importance of implementing adaptive variable selection mechanisms into financial risk forecasting models that use covariates.

9 Analysis of Estimated the Coefficients

The ℓ_1-regularized GARCH-EVT model allows us to draw inference about the relationship between each covariate and the likelihood and severity of short-term downside risk. These relationship can be used to inform the construction of future risk models and to design more comprehensive stress testing scenarios. Table 11 reports the mean regularized coefficient estimate, mean selection frequency and mean predictive duration for each covariate we study over the normal backtest and the financial crisis backtest. We find that mean selection frequency and the mean predictive duration of covariates is almost always higher during the financial crisis backtest. This finding is consistent with the conclusion from Section 8, that our chosen set of covariates are particularly useful for financial risk forecasting during periods of financial market distress.

Sell volume has the largest positive coefficient estimate, highest selection frequency and longest predictive duration across both backtests. As such, sell volume is the single most important covariate we study. We find that higher daily sell volume is associated with a higher value of the generalized Pareto distribution scale parameter and therefore a higher likelihood of observing extreme future losses. When sell volume is selected by the ℓ_1-regularized GARCH-EVT model the variable typically remains in the active coefficient set for the following three days. Total trade volume after market hours market is also selected comparatively more often than other covariates and is positively associated with the likelihood and severity of short-term downside risk. Hence, we observe that measures of daily transaction activity are particularly useful for financial risk forecasting.

Both quote and trade based measures of high frequency intraday volatility are positively associated with the generalized Pareto distribution scale parameter and are selected comparatively often. The mean coefficient estimates for both of these covariates are considerably larger over the financial crisis backtest. Interestingly, monetary policy uncertainty is considered to be more important for risk forecasting during the financial crisis backtest while economic policy uncertainty is selected more often over the normal backtest.

With the exception of sell volume and the two measures of intraday volatility, all other covariates have a reasonably similar mean selection frequency and predictive duration. Importantly, we find that the mean predictive duration of most covariates is below 2, suggesting that the importance of most economic and financial covariates we study is highly time-varying and depends on prevailing market conditions. This finding support our hypothesis that knowledge surrounding the optimal covariate set is difficult to establish *ex-ante*.

10 Conclusion

Existing financial risk forecasting models typically use the historical returns of the financial asset or investment portfolio, failing to consider the wealth of economic and financial data that is being recorded at a high frequency. In this study, we augment the second step of the traditional GARCH-EVT risk model, first introduced by McNeil and Frey (2000), to incorporate economic and financial covariates. The generalized Pareto distribution scale is re-parameterized as a log-linear function of covariates. However, we hypothesize that the best risk model only uses a small number of all possible covariates at any given time, and that the statistical importance of covariates is time-varying. Accordingly, we incorporate a two hyper-parameter variant of ℓ_1 regularization into the maximum likelihood objective function of the generalized Pareto distribution. This revised financial risk forecasting model, which we refer to as the ℓ_1-regularized GARCH-EVT model, now jointly performs parameter estimation together with variable selection and adaptive coefficient shrinkage.

During periods of normal market volatility, we find that the VaR forecasts from the ℓ_1-regularized GARCH-EVT model are exceptionally competitive with those generated from the traditional GARCH-EVT model as well as other popular benchmark risk model from the literature. We further demonstrate that 1-day ahead VaR and ES forecasts from the ℓ_1-regularized GARCH-EVT model can be substantially more accurate during periods of financial market distress, particularly at the levels of confidence prescribed by regulators. The largest gains in forecast accuracy over the traditional GARCH-EVT model occur during the Global Financial Crisis period and around the "Black Monday" market crash on 8th August, 2011. Hence, the ℓ_1-regularized GARCH-EVT model generates accurate risk forecasts when they are needed the most.

To demonstrate the need for variable selection mechanisms, we first show that a GARCH-EVT model without ℓ_1 regularization produces inadequate risk forecasts. We then illustrate that the active set of covariates used by the ℓ_1-regularized GARCH-EVT model is sparse and that covariates in this set typically have short predictive durations. Amongst the covariates we study, only daily sell volume could be considered to be a systematically important predictor of future financial risk. However, the relationships we identify between our chosen set of covariates and the likelihood and severity of downside risk may inform the design of future financial forecasting models and realistic scenarios for stress testing.

Rather than designing a novel risk model from the ground up, this study contributes to the risk management literature by augmenting the popular GARCH-EVT model for use in modern big data environments. The augmented risk model exploits the richness of economic and financial data to leverage improvements in forecast accuracy, particularly when financial markets are distressed. Recently, a growing branch of literature studies the relationship between exogenous covariates and extreme losses (Hambuckers et al., 2018; Embrechts et al., 2018). We contribute by examining a unique set of covariates, many of which are derived from millisecond level intraday data. Furthermore, we use a regularized forecasting model to demonstrate how and when certain covariates are related to extreme downside risk. Our results are of particular interest to prudential regulators, but also practitioners given the relative simplicity and naturality of our regularized forecasting model.

A Benchmark Risk Models

Below we briefly describe the background and implementation of the Hawkes POT, CaViaR, and CARE risk models that are used as benchmarks in the backtests. All models are implemented using Python 3.7.

A.0.1 Hawkes POT

Chavez-Demoulin et al. (2005) and Chavez-Demoulin and McGill (2012) develop a point process risk model, known as the Hawkes POT model. Negative log returns that exceed the threshold u are considered to be realizations of a marked point process, characterized by the exceedance time T_i, and the exceedance size Z_i. The history of the event process $(t_i, z_i), i \in \mathbb{Z}$, is denoted by $\mathcal{H}_t$. To account for conditional heteroskedasticity and serial dependence among extreme negative log returns the marked point process is assumed to belong to the general class of linearly self-exciting Hawkes process. The conditional intensity of a marked point process is,

$$\lambda(t, y|\mathcal{H}_t) = \lambda_g(t|\mathcal{H}_t) f(z|\mathcal{H}_t) \tag{35}$$

where $\lambda_g(t|\mathcal{H}_t)$ is the conditional intensity of the ground process which measures the occurrence rate of negative log returns exceeding the threshold and $f(z|\mathcal{H}_t)$ is the density of the marks conditional on the history of the process. The conditional intensity of the ground process is parameterized as self-exciting,

$$\lambda_g(t|\mathcal{H}_t) = \mu + \phi \sum_{j:t_j<t} \exp(-\gamma(t - t_j)) \tag{36}$$

The distribution of the marks is assumed to be generalized Pareto. However, the scale parameter of the Generalized Pareto distribution is modelled as a linear function of time using the same self-exciting dynamics as in (36), that is,

$$\sigma(t|\mathcal{H}_t) = \beta_0 + \beta_1 \sum_{j:t_t<t} \exp(\kappa(t - t_j)) \tag{37}$$

The generalized Pareto distribution for the marks then has the conditional probability density function,

$$f(z|\mathcal{H}_t) = g_{\xi,\sigma(t|\mathcal{H}_t)}(z) = \frac{1}{\sigma(t|\mathcal{H}_t)}\left(1 + \frac{\xi z}{\sigma(t|\mathcal{H}_t)}\right)^{-(1/\xi+1)} \tag{38}$$

Via these dynamics, clusters of extreme returns cause successively larger increases in the generalized Pareto distribution scale parameter and the conditional intensity of the ground process, before decaying back to their respective baseline rates.

Parameter estimation is performed by maximizing the log-likelihood function of the Hawkes POT model, given in Herrera et al. (2017) for example, subject to the constraint that $\mu, \phi, \beta_0, \beta_1, \gamma, \kappa, \xi > 0$. We use the same set of starting values in the numerical optimization routine as Herrera et al. (2017). The 1-day ahead VaR and ES forecasts from the Hawkes POT model can be obtained by substituting (36) and (37) into the expressions from Section 2, that is,

$$V\hat{a}R^{(\alpha)}_{t+1} = u + \frac{\hat{\sigma}_t(t|\mathcal{H}_t)}{\hat{\xi}}\left[\left(\frac{(1-\alpha)}{\hat{\lambda}_g(t|\mathcal{H}_t)}\right)^{-\hat{\xi}} - 1\right] \tag{39}$$

$$\hat{ES}^{(\alpha)}_{t+1} = \frac{V\hat{a}R^{(\alpha)}_{t+1} + \hat{\sigma}_t(t|\mathcal{H}_t) - \hat{\xi}u}{1-\hat{\xi}} \tag{40}$$

A.0.2 CaViaR

The conditional autoregressive Value-at-Risk (CaViaR) model of Engle and Manganelli (2004), uses quantile regression to estimate the VaR. We focus on the symmetric absolute value CaViaR model (CaViaR henceforth) since this is the most widely used within the literature. The CaViaR model is,

$$Q^{(\alpha)}_t = \beta_0 + \beta_1 Q^{(\alpha)}_{t-1} + \beta_2|\tilde{y}_{t-1}| \tag{41}$$

where $Q^{(\alpha)}_t$ is the α-quantile of the underlying loss distribution and $\tilde{y}$ is the de-meaned log return series. The parameters of the CaViaR model are the solution to,

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = \arg\min_{\beta_0,\beta_1,\beta_2} \sum_{t=1}^{T} (\alpha - \mathbb{1}(\tilde{y}_t < Q^{\alpha}_t))(\tilde{y}_t - Q^{\alpha}_t) \tag{42}$$

To set the starting parameters for the numerical minimization routine we follow the approach taken by Taylor (2008). In the first moving window of the backtest we use 10^5 different parameter vectors of uniformly distributed random numbers in the half-open interval [0,1) as starting values to multiple different solutions. The 10 vectors of starting values that produced the smallest value of the quantile regression objective function (42), are retained. These solutions obtained from the 10 best vectors of starting values are then re-used as the initial values in the numerical minimization routine and we select the single parameter vector which minimizes the quantile regression objective function as the final set of starting parameters. For all subsequent moving windows in the backtest we use the optimal parameter values from the previous moving window as starting values. The CaViaR model can only be used to estimate the VaR.

A.0.3 CARE

Taylor (2008) propose the conditional autoregressive expectile (CARE) model which uses the conditional τ-expectile as an estimator of the conditional α-quantile. Importantly, this model estimates both the VaR and ES within a conditional autoregressive framework. The τ-expectile is defined as the value m which minimizes,

$$E[|\tau - \mathbb{1}(y < m)|(y - m)^2]$$

The following autoregressive symmetric absolute value model is proposed to describe the dynamics of the conditional τ-expectile, denoted $e_t^{(\tau)}$,

$$e_t^{(\tau)} = \beta_0 + \beta_1 e_{t-1}^{(\tau)} + \beta_2 |\tilde{y}_{t-1}| \tag{43}$$

Parameters of this conditional expectile model are estimated using asymmetric least squares,

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = \arg \min_{\beta_0, \beta_1, \beta_2} \sum_{t=1}^{T} |\tau - \mathbb{1}(y_t < e_t^{(\tau)})|(y_t - e_t^{(\tau)})^2 \tag{44}$$

Since, for each τ-expectile there exists a corresponding α-quantile (Efron, 1991), it is possible to use this expectile based model to estimate the VaR. Taylor (2008) propose to find the τ-expectile which is equivalent to the α-quantile in terms of the proportion of in-sample observations are below the τ-expectile, and use this to estimate the $VaR^{(\alpha)}$ via the conditional symmetric absolute value model given in (43).

Taylor (2008) derive the following relationship[11] between the τ-expectile and the ES,

$$ES^{(\alpha)} = \left(1 + \frac{\tau}{(1 - 2\tau)\alpha}\right) e_t^{(\tau)} \tag{45}$$

This expression suggests that the ES can be estimated using the following auto-regressive symmetric absolute value model

$$ES_t^{(\alpha)} = \gamma_0 + \gamma_1 ES_{t-1,h}^{(\alpha)} + \gamma_2 |\tilde{y}_{t-1}| \tag{46}$$

where $\gamma_1 = \beta_1$ and $\gamma_i = (1 + \frac{\tau}{(1 - 2\tau)\alpha})\beta_i$ for $i \in \{0, 2\}$. The procedure used to find the starting values for the asymmetric least squares objective is the same as that used for the CaViaR model from Section A.0.2.

[11] See Taylor (2008) for a detailed derivation.

Table 1: Summary Statistics

The table presents summary statistics for each equity in the sample. The column θ *(Log Returns)* reports the value of the extremal index for the negative log-return series. The column θ *(GARCH)* reports the value of the extremal index for the standardized GARCH residual series. The column θ *(GJR-GARCH Residuals)* reports the value of the extremal index for the standardized GJR-GARCH residual series. The extremal index is computed using the "intervals" estimator of Ferro and Segers (2003) using a threshold set at the 90^{th} quantile value. The tail index is computed using the Hill estimator (Hill, 1975) for a sample of 408 upper order statistics from the standardized GJR-GARCH residual series.

Ticker	Std	Min	Max	Skewness	Excess Kurtosis	Tail Index	θ (Log Returns)	θ (GARCH)	θ (GJR-GARCH)
AAPL	2.08	-13.019	19.747	0.123	5.699	0.385	0.545	0.8	0.939
BAC	2.998	-30.21	34.206	0.374	28.508	0.393	0.238	0.912	1
C	3.139	-45.632	49.47	0.577	45.776	0.387	0.329	0.882	0.97
CSCO	1.799	-14.799	17.686	0.386	10.926	0.418	0.582	0.76	0.953
CVX	1.579	-18.941	13.341	-0.05	12.72	0.364	0.604	0.944	1
DIS	1.585	-14.818	10.231	-0.359	9.848	0.388	0.658	0.957	1
GE	1.869	-17.984	13.684	-0.039	10.507	0.386	0.343	0.889	0.923
HD	1.549	-13.161	8.579	-0.306	5.495	0.348	0.564	0.999	1
IBM	1.326	-10.899	8.662	0.258	6.492	0.424	0.688	0.888	0.947
INTC	1.801	-11.199	13.221	0.261	5.261	0.393	0.685	0.906	0.938
JNJ	1.016	-11.537	10.578	0.082	12.004	0.381	0.569	0.793	0.916
JPM	2.28	-22.392	23.228	-0.34	19.724	0.362	0.384	0.875	0.979
KO	1.079	-12.997	9.068	-0.121	13.401	0.384	0.517	0.864	0.926
MRK	1.624	-12.251	31.171	2.052	42.609	0.406	0.646	0.825	1
MSFT	1.612	-17.063	12.458	0	10.4	0.394	0.609	0.832	0.929
ORCL	1.705	-12.283	12.393	0.076	6.382	0.41	0.656	0.817	0.97
PEP	1.026	-8.204	12.705	0.321	10.576	0.394	0.627	0.95	0.997
PFE	1.382	-9.687	11.817	0.333	7.909	0.391	0.535	1	1
PG	1.062	-9.726	8.226	0.124	7.616	0.415	0.688	0.899	0.931
SLB	2.129	-13.902	20.339	0.457	7.869	0.326	0.492	0.909	0.928
UNH	1.915	-29.829	20.624	-0.39	26.216	0.385	0.577	0.9	0.962
VZ	1.285	-13.656	8.413	-0.175	7.665	0.385	0.78	0.901	0.982
WFC	2.44	-28.341	27.21	-0.929	29.705	0.387	0.224	0.971	1
WMT	1.202	-10.502	10.74	-0.049	10.678	0.405	0.693	0.868	0.906
XOM	1.449	-15.863	15.027	0.013	13.446	0.397	0.638	0.926	0.956

Table 2: Description of Economic and Financial Covariates

The table provides a short description of each covariate included in the ℓ_1-regularized GARCH-EVT risk model. The data source for each covariate is also provided.

Name	Description	Source
sell_vol	Sell volume	Millisecond TAQ
total_vol_a	Trade volume after market hours	Millisecond TAQ
quoted_spread	Time-weighted Quoted Spread	Millisecond TAQ
best_bid_depth	Time-weighted best bid depth	Millisecond TAQ
effective_spread	Share-weighted Effective Spread	Millisecond TAQ
ivol_trade	Trade-based intraday volatility	Millisecond TAQ
ivol_quote	NBBO Quote-based intraday volatility	Millisecond TAQ
abs_volume_imbalance	Absolute trade imbalance	Millisecond TAQ
signed_volume_imbalance	Signed trade imbalance	Millisecond TAQ
kyles_lambda	Kyles Lambda (price impact coefficient)	Millisecond TAQ
herfindahl_index	Herfindahl Index	Millisecond TAQ
var_ratio_1	Variance Ratio (15- second/3*5-second)	Millisecond TAQ
var_ratio_2	Variance Ratio (1-min/4*15- second)	Millisecond TAQ
var_ratio_3	Variance Ratio (5-min/5*1- min)	Millisecond TAQ
var_ratio_4	Variance Ratio (15-min/3*5- min)	Millisecond TAQ
var_ratio_5	Variance Ratio (30-min/2*15- min)	Millisecond TAQ
open_interest	Put option open interest	OptionMetrics
EPU	Economic policy uncertianty index	https://www.policyuncertainty.com
EMV	News based market volatility index	https://www.policyuncertainty.com
MPU	Monetary policy uncertianty index	https://www.policyuncertainty.com
TED_spread	TED spread	FRED
US_high_yield_spread	ICE BofA High Yield bond spread	FRED
recssion_prob	Smoothed recession probabilities	FRED
CBOE_vol_idx	CBOE Volatility Index	FRED

Table 3: Normal Backtest 1-day Ahead VaR Forecast Results

The table reports the number of times the null hypothesis was rejected at the 5% level for each of the statistical tests used to evaluate 1-day ahead VaR forecasts over the normal backtest. Definitions of each statistical test are provided in Section 6.1. The column *Total* is the sum of rejection counts across all statistical tests. The final column reports the mean value of the VaR scoring function (33), scaled by $1/(1-\alpha)$. Smaller numbers indicate better forecasts.

Panel A: $\alpha = 0.975$							
Risk Model	LR_{UC}	MCS_{UC}	LR_{Ind}	LR_{CC}	DQ	Total	$S(VaR, y)$
ℓ_1-Regularized GARCH-EVT	6	0	2	7	5	20	3.7213
Non-Regularized GARCH-EVT	1	5	2	3	9	20	4.1439
Traditonal GARCH-EVT	5	0	5	9	7	26	3.7190
Hawkes POT	7	0	11	9	11	38	3.7578
Symmetric Absolute Value CaViaR	1	2	4	4	10	21	3.7196
Symmetric Absolute Value CARE	2	1	7	7	12	29	3.7494
GJR-GARCH Student's t	3	0	2	3	5	13	3.6733
Panel B: $\alpha = 0.99$							
Risk Model	LR_{UC}	MCS_{UC}	LR_{Ind}	LR_{CC}	DQ	Total	$S(VaR, y)$
ℓ_1-Regularized GARCH-EVT	0	0	1	1	1	3	4.8313
Non-Regularized GARCH-EVT	11	16	2	11	15	55	5.6862
Traditonal GARCH-EVT	1	0	1	2	2	6	4.7890
Hawkes POT	1	0	5	1	7	14	4.8935
Symmetric Absolute Value CaViaR	1	4	2	3	6	16	4.7900
Symmetric Absolute Value CARE	2	1	2	3	5	13	4.8148
GJR-GARCH Student's t	0	0	2	1	2	5	4.7374
Panel C: $\alpha = 0.995$							
Risk Model	LR_{UC}	MCS_{UC}	LR_{Ind}	LR_{CC}	DQ	Total	$S(VaR, y)$
ℓ_1-Regularized GARCH-EVT	0	2	1	0	3	6	5.8219
Non-Regularized GARCH-EVT	21	23	0	18	21	83	7.2204
Traditonal GARCH-EVT	0	0	1	0	3	4	5.7456
Hawkes POT	0	1	6	1	12	20	5.9231
Symmetric Absolute Value CaViaR	1	3	2	0	9	15	5.7901
Symmetric Absolute Value CARE	0	1	4	1	9	15	5.8311
GJR-GARCH Student's t	4	4	1	3	6	18	5.6989

Table 4: Normal Backtest 1-day Ahead ES Forecast Results

The table reports the number of times the null hypothesis was rejected at the 5% level for each of the statistical tests used to evaluate 1-day ahead ES forecasts over the normal backtest. Definitions of each statistical test are provided in Section 6.2. The column *Total* is the sum of rejection counts across all statistical tests. The final column reports the avermean value of the joint VaR and ES scoring function (34), scaled by $1/(1-\alpha)$. Smaller numbers indicate better forecasts.

Panel A: $\alpha = 0.975$					
Model	McNeil (2000) Bootstrap	AC_{UC}	S_4	Total	$S(VaR, ES, y)$
ℓ_1-Regularized GARCH-EVT	7	0	3	10	1.9049
Non-Regularized GARCH-EVT	7	0	19	26	1.9896
Traditonal GARCH-EVT	5	0	1	6	1.9044
Hawkes POT	25	0	1	26	1.9264
CARE	8	0	0	8	1.9144
GJR-GARCH Student's t	0	0	3	3	2.1567
Panel B: $\alpha = 0.99$					
Model	McNeil (2000) Bootstrap	AC_{UC}	S_4	Total	$S(VaR, ES, y)$
ℓ_1-Regularized GARCH-EVT	5	0	1	6	2.1761
Non-Regularized GARCH-EVT	2	0	22	24	2.3437
Traditonal GARCH-EVT	3	0	1	4	2.1662
Hawkes POT	23	0	0	23	2.2014
CARE	7	0	2	9	2.1751
GJR-GARCH Student's t	0	0	3	3	2.4345

Table 5: Financial Crisis Backtest 1-day Ahead VaR Forecast Results
The table reports the number of times the null hypothesis was rejected at the 5% level for each of the statistical tests used to evaluate 1-day ahead VaR forecasts over the financial crisis backtest. Definitions of each statistical test are provided in Section 6.1. The column *Total* is the sum of rejection counts across all statistical tests. The final column reports the mean value of the VaR scoring function (33), scaled by $1/(1-\alpha)$. Smaller numbers indicate better forecasts.

Panel A: $\alpha = 0.975$

Risk Model	LR_{UC}	MCS_{UC}	LR_{Ind}	LR_{CC}	DQ	Total	$S(VaR, y)$
ℓ_1-Regularized GARCH-EVT	2	7	1	2	8	20	6.0743
Non-Regularized GARCH-EVT	17	21	0	15	18	71	7.0333
Traditonal GARCH-EVT	4	8	0	3	14	29	6.1332
Hawkes POT	0	3	3	2	19	27	6.4064
Symmetric Absolute Value CaViaR	13	14	4	8	24	63	6.5886
Symmetric Absolute Value CARE	13	15	4	9	24	65	6.6516
GJR-GARCH Student's t	5	7	0	5	13	30	6.0465

Panel B: $\alpha = 0.99$

Risk Model	LR_{UC}	MCS_{UC}	LR_{Ind}	LR_{CC}	DQ	Total	$S(VaR, y)$
ℓ_1-Regularized GARCH-EVT	0	0	0	0	6	6	7.7020
Non-Regularized GARCH-EVT	22	22	0	22	22	88	9.7388
Traditonal GARCH-EVT	0	0	1	1	11	13	7.7828
Hawkes POT	1	4	3	3	19	30	8.2523
Symmetric Absolute Value CaViaR	14	14	0	10	25	63	8.7993
Symmetric Absolute Value CARE	18	21	3	16	24	82	9.2573
GJR-GARCH Student's t	4	5	0	2	9	20	7.6263

Panel C: $\alpha = 0.995$

Risk Model	LR_{UC}	MCS_{UC}	LR_{Ind}	LR_{CC}	DQ	Total	$S(VaR, y)$
ℓ_1-Regularized GARCH-EVT	2	4	0	0	6	12	9.2315
Non-Regularized GARCH-EVT	23	24	0	23	24	94	12.4246
Traditonal GARCH-EVT	0	1	0	0	8	9	9.3634
Hawkes POT	1	4	2	1	15	23	9.8433
Symmetric Absolute Value CaViaR	12	16	1	10	21	60	11.2410
Symmetric Absolute Value CARE	18	19	2	17	22	78	12.1329
GJR-GARCH Student's t	3	5	2	2	10	22	9.0829

Table 6: Financial Crisis Backtest 1-day Ahead ES Forecast Results
The table reports the number of times the null hypothesis was rejected at the 5% level for each of the statistical tests used to evaluate 1-day ahead ES forecasts over the financial crisis backtest. Definitions of each statistical test are provided in Section 6.2. The column *Total* is the sum of rejection counts across all statistical tests. The final column reports the mean value of the joint VaR and ES scoring function (34), scaled by $1/(1-\alpha)$. Smaller numbers indicate better forecasts.

Panel A: $\alpha = 0.975$					
Model	McNeil (2000) Bootstrap	AC_{UC}	S_4	Total	$S(VaR, ES, y)$
ℓ_1-Regularized GARCH-EVT	0	0	2	2	2.3764
Non-Regularized GARCH-EVT	1	0	23	24	2.5231
Traditonal GARCH-EVT	0	0	3	3	2.3864
Hawkes POT	22	0	1	23	2.4717
CARE	11	0	14	25	2.5229
GJR-GARCH Student's t	0	0	5	5	2.6930
Panel B: $\alpha = 0.99$					
Model	McNeil (2000) Bootstrap	AC_{UC}	S_4	Total	$S(VaR, ES, y)$
ℓ_1-Regularized GARCH-EVT	2	0	2	4	2.6823
Non-Regularized GARCH-EVT	0	0	24	24	2.9887
Traditonal GARCH-EVT	2	0	1	3	2.6871
Hawkes POT	20	0	3	23	2.8025
CARE	14	1	21	36	3.1095
GJR-GARCH Student's t	0	0	4	4	3.0125

Table 9: 10-Day Ahead ES Forecast Results

The table reports the number of times the null hypothesis was rejected at the 5% level for each of the statistical tests used to evaluate 10-day ahead ES forecasts over the normal backtest and the financial crisis backtest. Definitions of each statistical test are provided in Section 6.2. The column *Total* is the sum of rejection counts across all statistical tests. The final column reports the mean value of the joint VaR and ES scoring function (34), scaled by $1/(1-\alpha)$. Smaller numbers indicate better forecasts.

	(i) Standard Backtest					(i) Financial Crisis Backtest				
Panel A: $\alpha = 0.975$										
Model	McNeil (2000) Bootstrap	AC_{UC}	S_4	Total	$S(VaR, ES, y)$	McNeil (2000) Bootstrap	AC_{UC}	S_4	Total	$S(VaR, ES, y)$
ℓ_1-Regularized GARCH-EVT	4	0	11	15	3.3570	7	0	11	18	4.4442
Relaxedℓ_1-Regularized GARCH-EVT	5	0	16	21	3.5420	12	2	21	35	4.7539
Traditonal GARCH-EVT	2	0	9	11	3.3273	7	0	13	20	4.4408
Hawkes POT	22	0	8	30	3.3867	21	0	15	36	4.5140
CARE	4	0	12	16	3.3358	12	0	15	27	4.5323
GJR-GARCH Student's t	0	0	11	11	3.8165	0	0	14	14	4.9159
Panel B: $\alpha = 0.99$										
Model	McNeil (2000) Bootstrap	AC_{UC}	S_4	Total	$S(VaR, ES, y)$	McNeil (2000) Bootstrap	AC_{UC}	S_4	Total	$S(VaR, ES, y)$
ℓ_1-Regularized GARCH-EVT	5	0	8	13	3.7601	7	2	11	20	5.0348
Non-Regularized GARCH-EVT	3	0	18	21	4.1293	0	7	22	29	5.8196
Traditonal GARCH-EVT	6	0	8	14	3.7029	4	2	12	18	5.0080
Hawkes POT	16	0	10	26	3.7807	19	3	13	35	5.1495
CARE	4	0	7	11	3.7091	11	7	19	37	5.5635
GJR-GARCH Student's t	0	0	7	7	4.2543	0	0	12	12	5.5196

Table 10: Analysis of the Active Coefficient Set

The table reports the coefficient estimates from regressions studying the size of the active coefficient set and the predictive duration of the typical selected covariate over the financial crisis backtest. Specification (i) is a Poisson regression. Specification (ii) is a Weibull regression. Independent variables are described in Section 8. *,** and *** represent statistical significance at the 10%, 5% and 1% levels, respectively.

	(i) $\lvert\mathcal{A}_t\rvert$	(ii) Hazard ratio
Crisis Period	0.0395***	1.158***
	-0.0087	-0.0299
BM Period	-0.0611**	0.791***
	-0.0244	-0.0549
Volatility $> 50^{th}$ percentile	-0.0250***	0.976
	-0.0064	-0.0189
χ^2	166.7***	564.3***
Obs	27,075	15,097

Table 11: Regularized Coefficient Statistics

The table reports the mean regularized coefficient estimate, the mean selection frequency and the mean predictive duration for each covariate across all stocks in the standard backtest and financial crisis backtest. Coefficient estimates are multiplied by 100.

	(i) Normal Backtest				(ii) Financial Crisis Backtest		
	Coefficient Estimate	Selection Frequency	Duration		Coefficient Estimate	Selection Frequency	Duration
sell_vol	3.7552	0.4469	3.0120	sell_vol	3.2554	0.4220	3.0411
ivol_quote	0.8150	0.2037	1.7833	ivol_trade	1.3546	0.2545	1.7032
ivol_trade	0.8588	0.1917	1.4365	ivol_quote	1.1741	0.2065	1.5240
total_vol_a	0.5826	0.1753	1.3437	total_vol_a	0.9069	0.1989	1.3911
EPU	-0.0347	0.1738	1.3335	MPU	0.4330	0.1947	1.3600
var_ratio_5	0.0770	0.1734	1.3206	effective_spread	-0.3609	0.1945	1.3395
US_high_yield_spread	-1.3443	0.1732	1.3157	var_ratio_2	0.1139	0.1929	1.3555
CBOE_vol_idx	1.5767	0.1732	1.3176	var_ratio_3	-0.0461	0.1927	1.3644
recssion_prob	-0.0873	0.1731	1.3110	CBOE_vol_idx	3.1602	0.1921	1.3383
put_open_interest	-0.9003	0.1729	1.3192	kyles_lambda	-0.4393	0.1920	1.3517
best_bid_depth	-0.3558	0.1729	1.3084	EMV	-1.8776	0.1917	1.3452
kyles_lambda	0.0880	0.1728	1.3312	var_ratio_4	0.1248	0.1917	1.3491
var_ratio_2	-0.0368	0.1728	1.3148	EPU	-0.1330	0.1916	1.3524
EMV	-1.7409	0.1727	1.3127	var_ratio_5	0.0107	0.1911	1.3439
MPU	0.1292	0.1727	1.3240	herfindahl_index	0.4714	0.1909	1.3539
TED_spread	-0.0481	0.1727	1.3130	recssion_prob	0.3414	0.1905	1.3438
effective_spread	-0.4291	0.1726	1.3168	put_open_interest	-0.1293	0.1903	1.3635
var_ratio_4	0.0764	0.1725	1.3114	US_high_yield_spread	-2.5578	0.1900	1.3398
quoted_spread	-1.2747	0.1724	1.3291	TED_spread	-1.3663	0.1898	1.3300
var_ratio_1	0.0122	0.1723	1.3283	quoted_spread	-1.6572	0.1896	1.3415
var_ratio_3	-0.1963	0.1722	1.3141	var_ratio_1	0.2047	0.1891	1.3489
herfindahl_index	0.9273	0.1719	1.3173	best_bid_depth	0.1060	0.1890	1.3445
abs_vol_imbal	-0.4625	0.1712	1.3193	abs_vol_imbal	-0.2834	0.1883	1.3558
signed_vol_imbal	0.1577	0.1631	1.3071	signed_vol_imbal	-0.1384	0.1816	1.3444

References

Acerbi, C. and B. Szekely (2014). Back-testing expected shortfall. *Risk 27*(11), 76–81.

Acerbi, C. and D. Tasche (2002). On the coherence of expected shortfall. *Journal of Banking & Finance 26*(7), 1487–1503.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets 5*(1), 31–56.

Antonakakis, N., I. Chatziantoniou, and G. Filis (2013). Dynamic co-movements of stock market returns, implied volatility and policy uncertainty. *Economics Letters 120*(1), 87–92.

Bader, B., J. Yan, X. Zhang, et al. (2018). Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *The Annals of Applied Statistics 12*(1), 310–329.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. *The quarterly journal of economics 131*(4), 1593–1636.

Baker, S. R., N. Bloom, S. J. Davis, and K. J. Kost (2019). Policy news and stock market volatility. Technical report, National Bureau of Economic Research.

Balkema, A. A. and L. De Haan (1974). Residual life time at great age. *The Annals of probability*, 792–804.

Basel Committee on Banking Supervision (2006). Observed range of practice in key elements of advanced measurement approaches (AMA).

Basel Committee on Banking Supervision (2016). Minimum capital requirements for market risk. Bank for International Settlements.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics 31*(3), 307–327.

Brennan, M. J., T. Chordia, A. Subrahmanyam, and Q. Tong (2012). Sell-order liquidity and the cross-section of expected stock returns. *Journal of Financial Economics 105*(3), 523–541.

Brogaard, J. and A. Detzel (2015). The asset-pricing implications of government economic policy uncertainty. *Management Science 61*(1), 3–18.

Budish, E., P. Cramton, and J. Shim (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics 130*(4), 1547–1621.

Chakravarty, S., H. Gulen, and S. Mayhew (2004). Informed trading in stock and option markets. *The Journal of Finance 59*(3), 1235–1257.

Chavez-Demoulin, V., A. C. Davison, and A. J. McNeil (2005). Estimating value-at-risk: a point process approach. *Quantitative Finance 5*(2), 227–234.

Chavez-Demoulin, V. and P. Embrechts (2004). Smooth extremal models in finance and insurance. *Journal of Risk and Insurance 71*(2), 183–199.

Chavez-Demoulin, V., P. Embrechts, and S. Sardy (2014). Extreme-quantile tracking for financial time series. *Journal of Econometrics 181*(1), 44–52.

Chavez-Demoulin, V. and J. McGill (2012). High-frequency financial data modeling using Hawkes processes. *Journal of Banking & Finance 36*(12), 3415–3426.

Chinco, A., A. D. Clark-Joseph, and M. Ye (2019). Sparse signals in the cross-section of returns. *The Journal of Finance 74*(1), 449–492.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review 39*(4), 841–862.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*, Volume 208. Springer.

Davison, A. C. and R. L. Smith (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological) 52*(3), 393–425.

Easley, D., M. López de Prado, M. O'Hara, and Z. Zhang (2021). Microstructure in the machine age. *The Review of Financial Studies 34*(7), 3316–3363.

Efron, B. (1991). Regression percentiles using asymmetric squared error loss. *Statistica Sinica 1*(1), 93–125.

Embrechts, P., K. Mizgier, and X. Chen (2018). Modeling operational risk depending on covariates: an empirical investigation. *Journal of Operational Risk 13*(3), 17–46.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society 50*(4), 987–1007.

Engle, R. F. and S. Manganelli (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *22*(4), 367–381.

Ferro, C. A. and J. Segers (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*(2), 545–556.

Foucault, T., S. Moinas, and E. Theissen (2007). Does anonymity matter in electronic limit order markets? *The Review of Financial Studies 20*(5), 1707–1747.

Freyberger, J., A. Neuhierl, and M. Weber (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies 33*(5), 2326–2377.

Glosten, L. R., R. Jagannathan, and D. 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*(5), 1779–1801.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association 106*(494), 746–762.

Hambuckers, J., A. Groll, and T. Kneib (2018). Understanding the economic determinants of the severity of operational losses: A regularized generalized Pareto regression approach. *Journal of Applied Econometrics 33*(6), 898–935.

Hastie, T., R. Tibshirani, R. Tibshirani, et al. (2020). Best subset, forward stepwise or lasso? analysis and recommendations based on extensive comparisons. *Statistical Science 35*(4), 579–592.

Herrera, R. and A. Clements (2020). A marked point process model for intraday financial returns: modeling extreme risk. *Empirical Economics 58*(4), 1575–1601.

Herrera, R., A. Rodriguez, and G. Pino (2017). Modeling and forecasting extreme commodity prices: A Markov-Switching based extreme value model. *Energy Economics 63*, 129–143.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The annals of statistics 3*(5), 1163–1174.

Holden, C. W. and S. Jacobsen (2014). Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. *The Journal of Finance 69*(4), 1747–1785.

Hosking, J. R. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society: Series B (Methodological) 52*(1), 105–124.

Hosking, J. R. and J. R. Wallis (1987). Parameter and quantile estimation for the generalized pareto distribution. *Technometrics 29*(3), 339–349.

Jalal, A. and M. Rockinger (2008). Predicting tail-related risk measures: The consequences of using GARCH filters for non-GARCH data. *Journal of Empirical Finance 15*(5), 868–877.

Kratz, M., Y. H. Lok, and A. J. McNeil (2018). Multinomial VaR backtests: A simple implicit approach to backtesting expected shortfall. *Journal of Banking & Finance 88*, 393–407.

Langousis, A., A. Mamalakis, M. Puliga, and R. Deidda (2016). Threshold detection for the generalized Pareto distribution: Review of representative methods and application to the NOAA NCDC daily rainfall database. *Water Resources Research 52*(4), 2659–2681.

Ledford, A. W. and J. A. Tawn (2003). Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*(2), 521–543.

Lee, C. M. and M. J. Ready (1991). Inferring trade direction from intraday data. *The Journal of Finance 46*(2), 733–746.

McNeil, A. J. and R. Frey (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an execheme value approach. *Journal of empirical finance 7*(3-4), 271–300.

McNeil, A. J., R. Frey, and P. Embrechts (2015). *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press.

Meinshausen, N. (2007). Relaxed lasso. *Computational Statistics & Data Analysis 52*(1), 374–393.

Nieto, M. R. and E. Ruiz (2016). Frontiers in var forecasting and backtesting. *International Journal of Forecasting 32*(2), 475–501.

Nolde, N., J. F. Ziegel, et al. (2017). Elicitability and backtesting: Perspectives for banking regulation. *The annals of applied statistics 11*(4), 1833–1874.

Pan, J. and A. M. Poteshman (2006). The information in option volume for future stock prices. *The Review of Financial Studies 19*(3), 871–908.

Pastor, L. and P. Veronesi (2012). Uncertainty about government policy and stock prices. *The journal of Finance 67*(4), 1219–1264.

Patton, A., D. N. Politis, and H. White (2009). Correction to "automatic block-length selection for the dependent bootstrap" by d. politis and h. white. *Econometric Reviews 28*(4), 372–375.

Pickands III, J. et al. (1975). Statistical inference using extreme order statistics. *the Annals of Statistics 3*(1), 119–131.

Rigby, R. A. and D. M. Stasinopoulos (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 54*(3), 507–554.

Taylor, J. W. (2008). Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics 6*(2), 231–252.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological) 58*(1), 267–288.

Tobias, A. and M. K. Brunnermeier (2016). Covar. *The American Economic Review 106*(7), 1705.

Trapin, L. (2018). Can volatility models explain extreme events? *Journal of Financial Econometrics 16*(2), 297–315.

Zhang, J. (2010). Improving on estimation for the generalized pareto distribution. *Technometrics 52*(3), 335–339.

Zhang, J. and M. A. Stephens (2009). A new and efficient estimation method for the generalized pareto distribution. *Technometrics 51*(3), 316–325.

Ziggel, D., T. Berens, G. N. Weiß, and D. Wied (2014). A new set of improved value-at-risk backtests. *Journal of Banking & Finance 48*, 29–41.