The Uncertainty in Extreme Risk Forecasts from Covariate-Augmented Volatility Models

Yannick Hoga*

June 21, 2019

*Faculty of Economics and Business Administration, University of Duisburg-Essen, Universitätsstraße 12, D–45117 Essen, Germany, tel. +49 201 1834365, yannick.hoga@vwl.uni-due.de. The author would like to thank Christoph Hanck for his detailed comments. Full responsibility is taken for all remaining errors. Support of DFG (HO 6305/1-1) is gratefully acknowledged.

Abstract

For a GARCH-type volatility model with covariates, we derive asymptotically valid forecast intervals for risk measures, such as the Value-at-Risk and Expected Shortfall. To forecast these, we use estimators from extreme value theory. In the volatility model, we allow for the inclusion of exogenous variables, e.g., volatility indices or high-frequency volatility measures. Our framework for the volatility model captures leverage effects, thus allowing for sufficient flexibility in applications. In simulations, we find coverage of the forecast intervals to be adequate. Finally, we investigate if using covariate information from volatility indices or high-frequency data improves risk measure forecasts for real data. While—in our framework—volatility indices appear to be helpful in this regard, intra-day data are not.

Keywords: Extreme Value Theory, High-Frequency Volatility Measures, Risk Forecasts, Volatility Indices, Volatility Models

JEL classification: C13 (Estimation), C14 (Semiparametric and Nonparametric Methods), C53 (Forecasting and Prediction Methods)

1 Motivation

Risk forecasts are central to financial institutions and regulators. Indeed, under the Basel accords, banks have an incentive to issue prudent market risk forecasts that are neither too high nor too low. If risk is forecast to be too low over some period of time, the regulator requires risk capital to be increased under the Basel rules (Daníelsson, 2011, Sec. 8.1.1). If, on the other hand, risk forecasts are too high, the bank may internally decide to put more capital aside as a cushion against large losses. In both cases, the additional capital buffer can no longer earn any premiums. To avoid this, adequate risk forecasting models are of the utmost importance for both the profitability of individual financial institutions and the stability of the whole financial system.

Based on past log-losses $\varepsilon_1, \ldots, \varepsilon_n$ on some speculative asset, the aim in risk forecasting is to predict next period's risk inherent in ε_{n+1} given the current state of the market, which is captured by some information set $\mathcal{F}_n$. Here, $\mathcal{F}_n$ may not only contain past observations, but also covariate information. Many of the most popular risk forecasting methods rely on GARCH-type models. In this framework, log-losses ε_t are modelled as $\varepsilon_t = \sigma_t U_t$. Here, the volatility $\sigma_t > 0$ is assumed predictable from past information in $\mathcal{F}_{t-1}$. Furthermore, U_t is independent and identically distributed (i.i.d.) with zero mean and unit variance (abbreviated $U_t \overset{\text{i.i.d.}}{\sim} (0,1)$) and assumed to be independent from $\mathcal{F}_{t-1}$. Under the multiplicative structure of the model, the popular Value-at-Risk $q_{\alpha,n}$—defined as the α-quantile of the conditional distribution function (d.f.) $F_n(x) = P\left\{\varepsilon_{n+1} \leq x \mid \mathcal{F}_n\right\}$—is simply

$$q_{\alpha,n} = \sigma_{n+1} q_\alpha(U).$$

Here, $q_\alpha(U)$ denotes the α-quantile of a generic element U of the sequence $\{U_t\}$, and α is close to 1 to reflect large losses. Hence, to improve risk forecasts one can improve forecasts of volatility σ_{n+1} *and/or* improve estimates of the risk inherent in the innovations, measured here by $q_\alpha(U)$.

Volatility forecasts may be improved by incorporating additional information. For instance, Andersen *et al.* (2013) and Han and Kristensen (2014) list the following economic and financial variables that have been used as covariates in the literature: interest rates, bid-ask spreads, information flow, trading volumes, daily high–low ranges and numerous realized volatility measures; see also Shephard and Sheppard (2010) and Hansen *et al.* (2012) for the latter. On the other hand, Kuester *et al.* (2006) show that the estimation of $q_\alpha(U)$ may be improved using extreme value theory (EVT) by exploiting a Pareto-type tail shape for U_t. Such a Pareto-type assumption on the tail is often plausible in empirical applications as it is satisfied by, e.g., the popular Student t-distribution. The idea in EVT is to estimate a less extreme—and hence more easily estimated—quantile $q_\beta(U)$ ($\beta < \alpha$) first and then extrapolate to the desired $q_\alpha(U)$ using the Pareto-type tail. So while volatility forecasts may be improved exploiting *additional* data, estimates of $q_\alpha(U)$ may be improved by exploiting the *available* data more efficiently. In this paper, we consider both possibilities jointly.

In the applied literature, several authors have recently proposed risk forecasting methods that exploit both high-frequency information and EVT. For instance, Bee *et al.* (2016) filter the losses using a volatility model based on high-frequency (HF) measures and then apply the Peaks-over-Threshold method to the filtered residuals. Bee *et al.* (2019) suggest directly fitting a time-varying parameter Generalized Pareto distribution to the losses above some high threshold, where the parameters vary as functions of HF measures. Bee *et al.* (2018) propose a realized extreme quantile approach that also exploits HF information, but combines EVT with quantile regression. However, as the theoretical literature has not kept pace, little is known about the asymptotic properties of risk forecasts issued from models combining HF measures and EVT. Even for the popular HEAVY model of Shephard and Sheppard (2010) and the Realized GARCH model of Hansen *et al.* (2012), that both exploit HF measures, asymptotic theory for risk forecasts is lacking.

Thus, it is the first main aim of this paper to derive forecast intervals for EVT-enhanced risk forecasts issued from GARCH-type volatility models with covariates. The work to date by Chan *et al.* (2007) and Hoga (2019a+) only covers GARCH and ARMA–GARCH models, which—in contrast to our framework—do not allow for leverage effects, modelling of different powers of volatility, and inclusion of covariates. So despite the empirical relevance, there is a lack of available asymptotic theory. This may be explained by the scarce estimation theory available for covariate-augmented GARCH-type models.[1] Some exceptions include the work of Han and Kristensen (2014) for GARCH–X(1, 1) models and Francq and Thieu (2019) for APARCH–X(p, q) models. We build on the work of the latter, because of the flexible volatility dynamics. In particular, the APARCH–X model captures the well-known 'leverage effect', which allows a negative return to have a stronger impact on future volatility than a positive return of equal magnitude. As a special case, our framework covers the GJR–GARCH of Glosten *et al.* (1993) that Trapin (2018) finds particularly suitable to predict extreme events, which we are also interested in here.

The second contribution of this article is to investigate if covariates improve extreme risk measure forecasts in practice. While, as pointed out above, HF measures have been investigated in this context by Bee *et al.*, other exogenous variables have received less attention. For instance, Blair *et al.* (2001) find the Chicago Board Options Exchange's volatility index (VIX) to be more helpful for volatility forecasting than HF measures. Hence, we extend Blair *et al.*'s (2001) comparison of HF measures and volatility indices to the case of risk measure forecasting (instead of volatility forecasting) using EVT-based methods. Investigating risk measure forecasts rather than volatility forecasts is of interest because—under the current regulatory framework of the Basel Committee on Banking Supervision (2019)—risk measures like the Value-at-Risk (VaR) and Expected Shortfall play a more central role in risk management than volatility.

We consider forecasts of the following risk measures. First, we deal with extreme distortion risk measures (DRMs) introduced by El Methni and Stupfler (2017). This is a general class of quantile-based risk measures, covering the popular VaR and Expected Shortfall (see El Methni and Stupfler, 2017, Table 1). Second, we also study expectiles due to Newey and Powell (1987). Expectiles possess some appealing theoretical properties. For instance, Ziegel (2016) demonstrates that they are the only law-invariant risk measures that are both elicitable and coherent. Elicitability allows for sensible comparisons of different forecasts (Gneiting, 2011). A coherent risk measure satisfies the four axioms of translation invariance, subadditivity, positive homogeneity and monotonicity introduced by Artzner *et al.* (1999). Bellini and Di Bernardino (2017) and Daouia *et al.* (2018) provide excellent motivation for the use of expectiles in risk management.

The remainder of the article is organized as follows. Within our APARCH–X framework, Section 2 derives asymptotically valid interval forecasts for extreme risk. Monte Carlo simulations in Section 3 illustrate the good coverage of the forecast intervals. Section 4 explores the usefulness of covariate information—from volatility indices and high-frequency data—in risk forecasting. Finally, Section 5 concludes. Proofs are relegated to an Appendix.

[1] For instance, despite the popularity of HEAVY and Realized GARCH models, little is known even about the asymptotic properties of their parameter estimators.

2 Asymptotic Forecast Intervals

Subsection 2.1 introduces the APARCH–X model and the main assumptions that will be used throughout. Then, Subsection 2.2 presents the EVT-based risk forecasts issued from APARCH–X models. Finally, in Subsection 2.3 we derive the asymptotic forecast intervals for the extreme risk measures.

2.1 The APARCH–X Model

Let $\delta^\circ > 0$. The model assumed to be driving the log-losses ε_t in the following is given by

$$\varepsilon_t = \sigma_t U_t, \tag{1}$$

$$\sigma_t^{\delta^\circ} = \omega^\circ + \sum_{j=1}^{p} \left\{ \alpha_{+,j}^\circ (\varepsilon_{t-j})_+^{\delta^\circ} + \alpha_{-,j}^\circ (\varepsilon_{t-j})_-^{\delta^\circ} \right\} + \sum_{j=1}^{q} \beta_j^\circ \sigma_{t-j}^{\delta^\circ} + (\boldsymbol{\pi}^\circ)' \boldsymbol{x}_{t-1}, \tag{2}$$

where $x_+ = \max\{x, 0\}$, $x_- = \max\{-x, 0\}$, and $\boldsymbol{x}_t = (x_{1,t}, \ldots, x_{M,t})'$ is a vector of M exogenous covariates with coefficients $\boldsymbol{\pi}^\circ = (\pi_1^\circ, \ldots, \pi_M^\circ)'$. To enforce $\sigma_t^{\delta^\circ} > 0$, we assume positive covariates, $\omega^\circ > 0$, and non-negative coefficients $\alpha_{+,j}^\circ$, $\alpha_{-,j}^\circ$, β_j° and π_j°.

For $\boldsymbol{\pi}^\circ = \mathbf{0}$, when no covariates are present, this model nests several well-known GARCH variants. For general $\delta^\circ > 0$, the model reduces to the Asymmetric Power GARCH (APARCH) of Ding *et al.* (1993). Hence, Francq and Thieu (2019) call the above the APARCH–X(p, q) model. We obtain Zakoïan's (1994) TARCH model for $\delta^\circ = 1$ and the GJR–GARCH of Glosten *et al.* (1993) for $\delta^\circ = 2$. When $\delta^\circ = 2$ and $\alpha_{+,j}^\circ = \alpha_{-,j}^\circ$, (2) is the classic GARCH model of Bollerslev (1986).

In the remainder of the paper, we follow Francq and Thieu (2019) and assume δ° to be a fixed user-specified constant. Francq and Thieu (2019, Sec. 2.5) provide some theoretically-backed guidance on the choice of δ° in practice. Since δ° is fixed, a generic parameter vector from the parameter space $\boldsymbol{\Theta} \subset (0, \infty) \times [0, \infty)^{d-1}$ $(d = 2p + q + M + 1)$ will be denoted by

$$\boldsymbol{\theta} = (\omega, \alpha_{+,1}, \ldots, \alpha_{+,p}, \alpha_{-,1}, \ldots, \alpha_{-,p}, \beta_1, \ldots, \beta_q, \pi_1, \ldots, \pi_M)'.$$

We denote the true parameter vector by $\boldsymbol{\theta}^\circ = (\omega^\circ, \alpha_{+,1}^\circ, \ldots, \alpha_{+,p}^\circ, \alpha_{-,1}^\circ, \ldots, \alpha_{-,p}^\circ, \beta_1^\circ, \ldots, \beta_q^\circ, \pi_1^\circ, \ldots, \pi_M^\circ)'$.

The most comprehensive estimation theory for APARCH–X models has been developed by Francq and Thieu (2019). Denote by $\varepsilon_1, \ldots, \varepsilon_n$ a trajectory from the APARCH–X model, and let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ denote the exogenous variables. Then, for initial values $\varepsilon_{1-p} = \ldots = \varepsilon_0 = 0$ and $\widehat{h}_{1-q} = \ldots = \widehat{h}_0 = 0$ and $\boldsymbol{x}_0 = 0$, the Gaussian quasi-likelihood is

$$L_n(\boldsymbol{\theta}) = L_n(\boldsymbol{\theta}, \varepsilon_1, \ldots, \varepsilon_n, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi \widehat{h}_t^{2/\delta^\circ}}} \exp\left\{ -\frac{\varepsilon_t^2}{2\widehat{h}_t^{2/\delta^\circ}} \right\},$$

where

$$\widehat{h}_t = \widehat{h}_t(\boldsymbol{\theta}) = \omega + \sum_{j=1}^{p} \left\{ \alpha_{+,j}(\varepsilon_{t-j})_+^{\delta^\circ} + \alpha_{-,j}(\varepsilon_{t-j})_-^{\delta^\circ} \right\} + \sum_{j=1}^{q} \beta_j \widehat{h}_{t-j} + \boldsymbol{\pi}'\boldsymbol{x}_{t-1}, \qquad t \geq 1.$$

The quasi-maximum likelihood estimator (QMLE) of $\boldsymbol{\theta}^\circ$ is $\widehat{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} L_n(\boldsymbol{\theta})$. We now state the main assumptions that Francq and Thieu (2019) require to derive limit theory for $\widehat{\boldsymbol{\theta}}_n$.

(C1) $(U_t, \boldsymbol{x}_t')$ is a strictly stationary and ergodic process, and there exists $s > 0$ such that $\mathrm{E}\|\boldsymbol{x}_1\|^s < \infty$, where $\|\cdot\|$ denotes the Euclidean norm.

(C2) $U_t \overset{\text{i.i.d.}}{\sim} (0,1)$ with U_t independent from $\mathcal{F}_{t-1}$, where $\mathcal{F}_{t-1}$ is the σ-field generated by $\{\varepsilon_{t-i}, \boldsymbol{x}_{t-i}, i \geq 1\}$, i.e., $\mathcal{F}_{t-1} = \sigma(\varepsilon_{t-i}, \boldsymbol{x}_{t-i}, i \geq 1)$.

(C3) $\boldsymbol{\theta}^\circ \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is compact, and all components of $\boldsymbol{\theta}^\circ$ are strictly positive.

(C4) For all $i \geq 1$ and $t \in \mathbb{Z}$, the support of the distribution of η_{t-i} given $\mathcal{F}_{t,i} = \sigma(U_{t-j}, j > i, \boldsymbol{x}_{t-k}, k > 0)$ is not included in $[0, \infty)$ or in $(-\infty, 0]$ and contains at least three points.

(C5) The top Lyapunov exponent γ (Francq and Thieu, 2019, p. 41) satisfies $\gamma < 0$ and $\sum_{j=1}^{q} \beta_j < 1$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

(C6) If $q > 0$, $\mathcal{B}_{\boldsymbol{\theta}^\circ}(z) = 1 - \sum_{j=1}^{q} \beta_j^\circ z^j$ has no common roots with $\mathcal{A}_{\boldsymbol{\theta}^\circ,+}(z) = \sum_{j=1}^{p} \alpha_{+,j}^\circ z^j$ and $\mathcal{A}_{\boldsymbol{\theta}^\circ,-}(z) = \sum_{j=1}^{p} \alpha_{-,j}^\circ z^j$; $\mathcal{A}_{\boldsymbol{\theta}^\circ,+}(1) + \mathcal{A}_{\boldsymbol{\theta}^\circ,-}(1) \neq 0$ and $\alpha_{+,p}^\circ + \alpha_{-,p}^\circ + \beta_q^\circ \neq 0$ (with the notation $\alpha_{+,0}^\circ = \alpha_{-,0}^\circ = \beta_0^\circ = 1$).

(C7) If $\boldsymbol{c} \in \mathbb{R}^M \setminus \{\boldsymbol{0}\}$, then the distribution of $\boldsymbol{c}'\boldsymbol{x}_1$ is not degenerated.

Conditions **(C1)**–**(C7)** are sufficient for the almost sure consistency of the QMLE $\widehat{\boldsymbol{\theta}}_n$ (Francq and Thieu, 2019, Theorem 1). Note that we do not require Francq and Thieu's (2019) condition **A6**, since we assume the U_t to be i.i.d. (Francq and Thieu, 2019, Lemma 2). To derive the asymptotic normality of $\widehat{\boldsymbol{\theta}}_n$, the following further conditions are required.

(C8) $\mathcal{C} = \cup_{n=1}^{\infty} \left\{ \sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ) \ : \ \boldsymbol{\theta} \in \boldsymbol{\Theta} \right\} = \prod_{i=1}^{d} \mathcal{C}_i$, where $\mathcal{C}_i = [0, \infty)$ when the i-th component of $\boldsymbol{\theta}$ is equal to zero, and $\mathcal{C}_i = \mathbb{R}$ otherwise.

(C9) $\mathrm{E}\,|U_t|^4 < \infty$.

Under conditions **(C1)**–**(C9)**, Francq and Thieu (2019, Theorem 2) prove the asymptotic normality of the QMLE. For discussion of the technical conditions **(C1)**–**(C9)**, we refer to Francq and Thieu (2019).

2.2 Risk Measure Forecasting in APARCH–X Models

We now introduce risk measure forecasts from APARCH–X models that use extreme value theory. Suppose that today you have a sample $\varepsilon_1, \ldots, \varepsilon_n$ of log-losses and a sample of covariates $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. Then, $\mathcal{F}_n = \sigma(\varepsilon_n, \varepsilon_{n-1}, \ldots; \boldsymbol{x}_n, \boldsymbol{x}_{n-1}, \ldots)$—defined in **(C2)**—captures the current market conditions. The task in risk forecasting is to forecast risk inherent in tomorrow's realization ε_{n+1} given today's state of the market, embodied by $\mathcal{F}_n$. Formally, this risk (e.g., $\xi_{\alpha,n}$ or $\mathrm{DRM}_{\alpha,n}$ defined below) is a scalar derived from the conditional d.f. $F_n(\cdot) = P\left\{\varepsilon_{n+1} \leq \cdot \mid \mathcal{F}_n\right\}$.

First, we introduce the risk measures covered by our analysis, viz. expectiles and distortion risk measures. Following Newey and Powell (1987, p. 824), the *conditional* expectile is

$$\xi_{\alpha,n} := \operatorname*{arg\,min}_{x\in\mathbb{R}} \mathrm{E}[\eta_\alpha(\varepsilon_{n+1} - x) \mid \mathcal{F}_n], \tag{3}$$

where $\eta_\alpha(x) = |\alpha - I_{\{x\leq 0\}}| \cdot |x|^2$. Define $q_{\alpha,n}$ as the α-quantile of $F_n(\cdot)$. Then, the *conditional* extreme distortion risk measure (DRM) is given by

$$\mathrm{DRM}_{\alpha,n} := \int_0^1 q_{1-(1-\alpha)s,n} \,\mathrm{d}g(s), \tag{4}$$

where $g(\cdot)$ is a distortion function, i.e., a non-decreasing and right-continuous function with $g(0) = 0$ and $g(1) = 1$. This representation nests several well-known risk measures, such as the VaR at level α, $q_{\alpha,n}$ (for $g(x) = I_{\{x=1\}}$), and the Expected Shortfall at level α, $\mathrm{ES}_{\alpha,n} = \frac{1}{1-\alpha}\int_\alpha^1 q_{x,n}\,\mathrm{d}x$ (for $g(x) = x$).

Under the multiplicative structure of the APARCH–X model in (1) and (2), the expressions in (3) and (4) simplify due to the location–scale equivariance of DRMs and expectiles. Specifically, Hoga (2019+) shows that

$$\xi_{\alpha,n} = \sigma_{n+1}\xi_\alpha(U), \tag{5}$$

$$\mathrm{DRM}_{\alpha,n} = \sigma_{n+1}\,\mathrm{DRM}_\alpha(U), \tag{6}$$

where $\xi_\alpha(U) = \arg\min_x \mathrm{E}[\eta_\alpha(U_1 - x)]$ ($\mathrm{DRM}_\alpha(U) = \int_0^1 q_{1-(1-\alpha)s}(U)\,\mathrm{d}g(s)$) is the unconditional expectile (DRM) associated with U. Obviously, both $\xi_{\alpha,n}$ and $\mathrm{DRM}_{\alpha,n}$ are random variables, since volatility σ_{n+1} is stochastic.

Clearly, σ_{n+1} in (5) and (6) can be forecast easily via $\widehat{h}_{n+1}^{1/\delta^\circ}(\widehat{\boldsymbol{\theta}}_n)$. To estimate the remaining terms $\xi_\alpha(U)$ and $\mathrm{DRM}_\alpha(U)$, we require two sequences k_n and α_n (cf. **(C10)**) and an assumption on the tail decay of the U_t (cf. **(C11)**):

(C10) The sequence k_n tends to ∞ with $k_n/n^{1-\iota} \to 0$, as $n \to \infty$, for some $\iota > 0$. Further, $\alpha_n \to 1$ with $n(1-\alpha_n) \to c > 0$, as $n \to \infty$.

(C11) The $(1-1/x)$-quantile $U(x)$ of a generic element U of $\{U_t\}$ satisfies the following: There exist $\gamma > 0$, $\rho < 0$ and a function $A(\cdot)$ with $\lim_{t\to\infty} A(t) = 0$ and constant sign, such that

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^{\gamma}}{A(t)} = x^{\gamma}\frac{x^{\rho}-1}{\rho} \qquad \text{for all } x > 0.$$

Additionally, $A(\cdot)$ satisfies $\sqrt{k}A(n/k_n) \underset{(n\to\infty)}{\longrightarrow} 0$. The parameter γ is called the *extreme value index* and ρ is the *second-order parameter.*

Remark 1. (i) Following standard convention in EVT, we shall suppress the subindex of k_n in **(C10)** and simply write k. The sequence α_n in **(C10)** represents the probability level at which the risk measures are evaluated subsequently. This probability level is extreme in the sense that it converges to 1 as the sample size increases.

(ii) Assumption **(C11)** is popular in EVT and controls the speed of convergence to a Pareto tail in

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^{\gamma}. \tag{7}$$

Since (7) even holds without the limit for a Pareto distribution, the innovations may be said to possess Pareto-type tails. The larger the extreme value index γ, the heavier the tail. Furthermore, the smaller the second-order parameter $\rho < 0$ in **(C11)**, the faster $A(t)$ converges to 0, since $|A(\cdot)|$ is necessarily regularly varying with index ρ (de Haan and Ferreira, 2006, Theorem 2.3.3). So the more negative ρ, the faster the speed of convergence in (7) and, hence, the better the Pareto approximation.

A key ingredient for the estimators of $\xi_{\alpha_n} = \xi_{\alpha_n}(U)$ and $\mathrm{DRM}_{\alpha_n} = \mathrm{DRM}_{\alpha_n}(U)$ is the Weissman (1978) estimate of the quantile $q_{\alpha_n} = q_{\alpha_n}(U)$. Since the innovations U_t are unobserved, we work with the standardized residuals $\widehat{U}_t := \varepsilon_t/\widehat{h}_t^{1/\delta^\circ}(\widehat{\boldsymbol{\theta}}_n)$. Denote the order statistics by $\widehat{U}_{(1)} \geq \ldots \geq \widehat{U}_{(n)}$. The Weissman (1978) estimator $\widehat{q}_{\alpha_n}$ is based on the insight that q_{α_n} relates to the less extreme—and hence more easily estimated—$q_{1-k/n}$ ($(1-k/n) < \alpha_n$) as follows:

$$q_{\alpha_n} \approx \left(\frac{k}{n(1-\alpha_n)}\right)^{\gamma} q_{1-k/n} \approx \left(\frac{k}{n(1-\alpha_n)}\right)^{\widehat{\gamma}} \widehat{U}_{(k+1)} =: \widehat{q}_{\alpha_n}. \tag{8}$$

Here, we have used the approximation $U(tx)/U(t) \approx x^{\gamma}$ from (7) for $t = n/k$ and $x = k/(n[1-\alpha_n])$. Furthermore, $\widehat{\gamma}$ in (8) denotes the Hill (1975) estimator

$$\widehat{\gamma} = \frac{1}{k}\sum_{i=1}^{k} \log\left(\widehat{U}_{(i)}/\widehat{U}_{(k+1)}\right).$$

Under (7), expectiles and DRMs decay with similar speed as quantiles far out in the tail. More precisely, El Methni and Stupfler (2017, Lemma 3) and Bellini and Di Bernardino (2017, Proposition 2.3) show for extreme DRMs and expectiles, respectively, that under (7)

$$\frac{\mathrm{DRM}_{\alpha_n}}{q_{\alpha_n}} \underset{(n\to\infty)}{\longrightarrow} \int_0^1 s^{-\gamma}\,\mathrm{d}g(s), \qquad \text{if } \int_0^1 s^{-\gamma-\iota}\,\mathrm{d}g(s)<\infty \text{ for some } \iota>0,$$
$$\frac{\xi_{\alpha_n}}{q_{\alpha_n}} \underset{(n\to\infty)}{\longrightarrow} (\gamma^{-1}-1)^{-\gamma}, \qquad \text{if } \gamma\in(0,1).$$

These relations motivate the plug-in estimators

$$\widehat{\mathrm{DRM}}_{\alpha_n} := \int_0^1 s^{-\widehat{\gamma}}\,\mathrm{d}g(s)\widehat{q}_{\alpha_n}, \tag{9}$$
$$\widehat{\xi}_{\alpha_n} := \left(\widehat{\gamma}^{-1}-1\right)^{-\widehat{\gamma}}\widehat{q}_{\alpha_n}. \tag{10}$$

El Methni and Stupfler (2017) and Daouia *et al.* (2018) study these estimators for i.i.d. data. Finally, our forecasts of $\mathrm{DRM}_{\alpha_n,n}$ and $\xi_{\alpha_n,n}$ are given by

$$\widehat{\xi}_{\alpha_n,n} = \widehat{h}_{n+1}^{1/\delta^\circ}(\widehat{\boldsymbol{\theta}}_n)\widehat{\xi}_{\alpha_n},$$
$$\widehat{\mathrm{DRM}}_{\alpha_n,n} = \widehat{h}_{n+1}^{1/\delta^\circ}(\widehat{\boldsymbol{\theta}}_n)\widehat{\mathrm{DRM}}_{\alpha_n}.$$

2.3 Forecast Intervals for Risk Measures

Our main theoretical result is the following

Theorem 1. *Suppose Conditions* **(C1)**–**(C11)** *hold for the APARCH–X process* $\{\varepsilon_t\}$ *defined in* (1) *and* (2). *Then,*

(i) if $\int_0^1 s^{-\gamma-1/2-\nu}\,\mathrm{d}g(s)<\infty$ *for some* $\nu>0$ *and* $\delta\mapsto q_\delta(U)$ *is continuous and strictly increasing in a neighbourhood of 1,*

$$\frac{1}{\widehat{\gamma}}\frac{\sqrt{k}}{\log\{k/(n[1-\alpha_n])\}}\log\left(\frac{\widehat{\mathrm{DRM}}_{\alpha_n,n}}{\mathrm{DRM}_{\alpha_n,n}}\right) \underset{(n\to\infty)}{\overset{d}{\longrightarrow}} N(0,1); \tag{11}$$

(ii) if $\delta\mapsto q_\delta(U)$ *is strictly increasing,*

$$\frac{1}{\widehat{\gamma}}\frac{\sqrt{k}}{\log\{k/(n[1-\alpha_n])\}}\log\left(\frac{\widehat{\xi}_{\alpha_n,n}}{\xi_{\alpha_n,n}}\right) \underset{(n\to\infty)}{\overset{d}{\longrightarrow}} N(0,1). \tag{12}$$

The convergences in (11) and (12) imply the asymptotic $(1-\tau)$-confidence forecast interval for $z_{\alpha_n,n}$ ($z\in\{\mathrm{DRM},\xi\}$):

$$I_{1-\tau} := \left[\widehat{z}_{\alpha_n,n}\exp\left\{-\Phi^{-1}_{1-\frac{\tau}{2}}\widehat{\gamma}\frac{\log\{k/(n[1-\alpha_n])\}}{\sqrt{k}}\right\}, \widehat{z}_{\alpha_n,n}\exp\left\{\Phi^{-1}_{1-\frac{\tau}{2}}\widehat{\gamma}\frac{\log\{k/(n[1-\alpha_n])\}}{\sqrt{k}}\right\}\right],$$

where $\Phi^{-1}_{1-\frac{\tau}{2}}$ denotes the $(1-\tau/2)$-quantile of a standard normal random variable.

In practice, the choice of k is difficult since it is only specified asymptotically. Daníelsson *et al.* (2016) suggest to choose k, such that the largest distance between a fitted Pareto-type tail and the empirical quantiles is minimized. To introduce the method, consider the following two estimators of the $(1-j/n)$-quantile $U(n/j)$. First, $U(n/j)$ may simply be estimated by the empirical quantile $\widehat{U}_{(j+1)}$. Second, by similar arguments that led to (8),

$$U\left(\frac{n}{j}\right) \approx \widehat{U}_{(k+1)}\left(\frac{j}{k}\right)^{-\widehat{\gamma}} =: q(j,k). \tag{13}$$

Now, for an adequate choice of k the absolute deviation between the semi-parametric estimate $q(j,k)$ and the non-parametric estimate $\widehat{U}_{(j+1)}$ should be small for a range of j's. This motivates the data-adaptive choice

$$k^* := \operatorname*{arg\,min}_{k=k_{\min},\dots,k_{\max}} \left[\sup_{j=1,\dots,k_{\max}} \left|\widehat{U}_{(j+1)} - q(j,k)\right|\right], \tag{14}$$

where we choose $k_{\min} = \lfloor(\log n)^2\rfloor$ and $k_{\max} = \lfloor 4(\log n)^2\rfloor$.[2] As the method of Daníelsson *et al.* (2016) is based on quantiles, it is well-suited for the DRM and expectile estimators in (9) and (10) that are both based on quantile estimates. We investigate the choice k^* in the simulations and the empirical application.

3 Simulations

We now compare the finite-sample coverage of the asymptotic 90%-confidence forecast intervals $I_{0.9}$ for VaR, Expected Shortfall and expectiles of time series with length $n \in \{1000,\ 2000\}$. Also, we compute the bias and RMSE of the risk measure forecasts. We do so using the data-dependent choice of k in (14). All simulation results are based on 10,000 replications. We use *R version 3.5.2* (R Core Team, 2018).

We generate time series $\{\varepsilon_t\}_{t=-v+1,\dots,n}$ from the APARCH–X(1,1) model with $\delta^\circ = 1$ given by

$$\begin{aligned}\varepsilon_t &= \sigma_t U_t,\\ \sigma_t &= \omega^\circ + \alpha^\circ_{+,1}(\varepsilon_{t-1})_+ + \alpha^\circ_{-,1}(\varepsilon_{t-1})_- + \beta^\circ\sigma_{t-1} + \pi^\circ_1 x_{t-1},\end{aligned} \tag{15}$$

with $\boldsymbol{\theta}^\circ = (\omega^\circ, \alpha^\circ_{+,1}, \alpha^\circ_{-,1}, \beta^\circ_1, \pi^\circ_1)' = (0.046, 0.027, 0.092, 0.843, 0.089)'$ and

$$x_t = \exp(y_t), \qquad y_t = 0.8 \cdot y_{t-1} + e_t,$$

where $e_t \overset{\text{i.i.d.}}{\sim} N(0,1)$. This data-generating process is taken from Francq and Thieu (2019, Sec. 3.1).

However, unlike Francq and Thieu (2019), we assume U_t to be heavy-tailed. More precisely, we consider $U_t \sim t_\nu\sqrt{(\nu-2)/\nu}$ with $\nu = 5$. For this standardized t-distribution, the parameters in **(C11)** satisfy $\gamma = 1/\nu$ and $\rho = -2/\nu$ by Hua and Joe (2011, Example 3) and de Haan and Ferreira

[2]The choice of $k_{\min}$ and $k_{\max}$ is motivated by Chan *et al.* (2007), who use a fixed $k = \lfloor 1.5(\log n)^2\rfloor$.

(2006, Theorem 2.3.9). We also use $U_t \sim R_t B_t / \sqrt{\mathrm{E}[B_t^2]}$, where R_t is i.i.d. Rademacher (i.e., ± 1 with equal probability 1/2), independent from the i.i.d. Burr($a = 1, b = 5$)-distributed B_t. The Burr(a, b)-distribution has d.f.

$$1 - F(x) = \left(\frac{1}{1 + x^b} \right)^a, \qquad x,\ a,\ b > 0.$$

From Hua and Joe (2011, Example 2) and de Haan and Ferreira (2006, Theorem 2.3.9), we obtain for the Burr(a, b)-distribution that $\gamma = 1/(ab)$ and $\rho = -1/a$ in **(C11)**.

Note that $U_t \overset{\text{i.i.d.}}{\sim} (0, 1)$ for both the Student t and the Burr distribution. Also, $\gamma = 1/5$ for our parameter choices in both cases. This implies in particular that $\mathrm{E}\,|U_t|^4 < \infty$, as required by **(C9)**. We estimate the APARCH–X(1,1) model via the QMLE $\widehat{\boldsymbol{\theta}}_n$ using the `rugarch` package (Ghalanos, 2019).

We consider three risk measures. As special cases of distortion risk measures, we investigate the two most popular risk measures, viz. VaR and Expected Shortfall. The third risk measure we use are expectiles. Since volatility estimates $\widehat{h}_t^{1/\delta^\circ}(\widehat{\boldsymbol{\theta}}_n)$ may be imprecise for the first few t due to initialization effects (see also Hall and Yao, 2003), we discard the first $v = 10$ standardized residuals and only consider $\{\widehat{U}_t = \varepsilon_t / \widehat{h}_t^{1/\delta^\circ}(\widehat{\boldsymbol{\theta}}_n)\}_{t=1,\ldots,n}$ for estimating z_{α_n} ($z \in \{q,\ \mathrm{ES},\ \xi\}$).[3]

Table 1 displays the results. We draw the following conclusions:

1. Bias and RMSE: For all risk forecasts, bias and RMSE decrease in n. As expected, the RMSEs increase the larger α_n, i.e., the more extreme the risk measure. Further, for the Burr-distribution the bias is lower than for the t-distribution. This is because the second-order parameter for the Burr-distribution ($\rho = -1$) is smaller than the second-order parameter for the t_5-distribution ($\rho = -2/5$). Hence, the Pareto approximation is more accurate for the former distribution and less bias is incurred by extrapolation.

2. Coverage: Coverage of the 90%-confidence intervals improves the larger the sample size n. Furthermore, coverage improves the closer α_n is to 1. This reflects the fact that by **(C11)** $n(1-\alpha_n) \to c > 0$ is required for the asymptotic analysis. Overall, the interval forecasts, particularly those for ES, suffer from some undercoverage. This may be explained by the fact that the confidence intervals do not take into account the estimation uncertainty of the parameter vector $\boldsymbol{\theta}^\circ$, which can be estimated with $\sqrt{n}$-rate. Hence, asymptotically the $\sqrt{k}/\log\{k/(n[1-\alpha_n])\}$-rate of the estimators $\widehat{\xi}_{\alpha_n}$ and $\widehat{\mathrm{DRM}}_{\alpha_n}$ dominates; cf. the proof of Theorem 1. Nonetheless, in finite samples, the estimation effects of $\boldsymbol{\theta}^\circ$ certainly contribute to the estimation uncertainty of the risk forecasts.

[3] Recall that we generate time series $\{\varepsilon_t\}_{t=-v+1,\ldots,n}$. Hence, discarding the first v residuals, leaves us with $\{\widehat{U}_t\}_{t=1,\ldots,n}$.

n	U_i	k^*/n	z	$1-\alpha_n$	Bias	RMSE	Coverage	Length
1000	t	5.5%	q	1%	−0.03	3.30	71.6	2.78
				0.5%	0.32	4.35	82.8	4.91
				0.1%	2.61	9.61	89.8	14.0
			ES	1%	1.13	5.89	60.7	4.10
				0.5%	2.17	8.42	68.6	7.27
				0.1%	6.95	20.0	72.6	20.9
			ξ	1%	0.36	2.90	69.0	2.21
				0.5%	0.70	4.01	78.5	3.91
				0.1%	2.72	9.38	84.8	11.2
	Burr	7.5%	q	1%	0.02	2.91	70.6	1.62
				0.5%	0.05	3.36	77.6	2.56
				0.1%	0.20	5.20	85.4	5.90
			ES	1%	0.10	3.81	61.0	2.06
				0.5%	0.17	4.64	69.2	3.27
				0.1%	0.46	7.99	79.3	7.57
			ξ	1%	0.08	2.31	72.0	1.23
				0.5%	0.09	2.68	78.8	1.95
				0.1%	0.20	4.28	85.9	4.51
2000	t	3.5%	q	0.5%	0.05	3.73	74.7	3.18
				0.1%	1.35	7.46	88.2	9.42
				0.05%	2.50	10.5	89.5	13.8
			ES	0.5%	1.12	6.51	64.9	4.50
				0.1%	3.98	14.5	76.2	13.4
				0.05%	6.13	20.7	76.9	19.8
			ξ	0.5%	0.33	3.19	73.0	2.47
				0.1%	1.43	6.82	85.5	7.36
				0.05%	2.38	9.73	86.8	10.8
	Burr	5%	q	0.5%	0.01	2.44	73.4	1.79
				0.1%	0.04	3.83	83.3	4.40
				0.05%	0.08	4.73	85.6	6.03
			ES	0.5%	0.04	3.38	63.9	2.27
				0.1%	0.14	5.59	76.4	5.57
				0.05%	0.21	7.05	79.3	7.65
			ξ	0.5%	0.05	1.87	75.0	1.36
				0.1%	0.06	2.97	84.2	3.35
				0.05%	0.10	3.71	86.4	4.59

Table 1: Average values over all replications of k^*, bias, RMSE, coverage and interval lengths of asymptotic 90%-confidence intervals.

3. Interval length: The lengths of the interval forecasts decrease in n for a fixed probability level α_n. For fixed n, the length increases the closer α_n is to 1. This simply reflects that there is more

estimation uncertainty for more extreme risk forecasts.

In view of the coverage reported by Chan *et al.* (2007) for their GARCH-based VaR forecast intervals, we find the coverage frequencies in Table 1 to be satisfactory. The fact that undercoverage is most severe for ES may be explained as follows. Since ES is an average over quantiles above VaR, ES is more extreme than VaR evaluated at the same risk level. Hence, the estimation uncertainty in ES forecasts and ES forecast intervals is increased. Hoga (2019a+) finds considering interval forecasts based on self-normalization may improve coverage. Considering these is however beyond the scope of this paper.

4 The Use of Covariates in Applied Risk Forecasting

One main innovation of the asymptotic theory developed in Section 2 is that covariate information can be incorporated into EVT-based risk forecasts. So it is of interest to investigate whether including covariates in EVT-based forecasts leads to improvements. In Subsection 4.1, we explore if option-implied expectations of future volatility help in risk forecasting for returns on major US stock indices. Then, Subsection 4.2 examines this for high-frequency measures derived from intra-day data. Our choice of covariates is inspired by Blair *et al.* (2001) and Koopman *et al.* (2005).

However, unlike Blair *et al.* (2001) and Koopman *et al.* (2005) who compare *volatility* forecasts, we compare *risk* forecasts. As can be seen from (5) and (6), volatility forecasts are just one component of a risk forecast. The other component is an estimate of the risk in U_t. It is this part of the risk forecasting method that we additionally account for in our comparison. Since, as pointed out above, risk forecasts are of more direct interest than volatility forecasts, we argue that the evaluation of the former is more informative about the quality of the risk forecasting methodology than the evaluation of the latter. Additionally, and in line with a large part of the volatility forecasting literature, Blair *et al.* (2001) and Koopman *et al.* (2005) require a (inherently noisy) volatility proxy. In contrast, we do not require a proxy for volatility, as we rely on scoring functions to compare risk forecasts (Gneiting, 2011).

In this section, we compare point forecasts of risk. These point forecasts are consistent for the true risk under the data-generating processes (DGPs) covered by Theorem 1. We expect these DGPs to provide reasonable approximations to the true dynamic of the index returns. Ideally, we would have compared different forecast intervals (derived from the asymptotic normality result in Theorem 1) instead of point forecasts, as the former are more informative. However, to the best of our knowledge, there exists no suitable scoring (or also: loss) function to compare forecast intervals for the risk measures considered here; see also Askanazi *et al.* (2018) for more on the general difficulties of comparing interval predictions. Hence, we stick to the task of evaluating point forecasts for which well-known scoring functions exist.

4.1 Volatility Indices as Covariates

Roughly speaking, volatility forecasts can be produced from time series models and option-implied standard deviations. The former typically only rely on the asset's own price history, whereas the latter are based on the broader information set of market participants' expectations. In their review of volatility forecasting, Poon and Granger (2003) find that methods exploiting option-implied information often compare favourably with time series methods. APARCH–X models allow to combine both methods by using option-implied volatility as a covariate. Here, we investigate whether including the value of a volatility index (derived from option prices) as covariate improves risk forecasts for the underlying index.

More specifically, we consider log-losses $\{\varepsilon_1, \dots, \varepsilon_N\}$ on the S&P 500, NASDAQ, Russell 2000 and Dow Jones Industrial Average (DJIA) indices from 1/1/2004–12/31/2018.[4] As covariates, we employ the Chicago Board Options Exchange's (CBOE) volatility indices on the S&P 500, NASDAQ, Russell 2000 and DJIA, respectively.[5] We denote these univariate time series by $\{x_1, \dots, x_N\}$. For details on how these volatility indices are computed from option prices, we refer to the CBOE's website at *www.cboe.com*.

To explore whether inclusion of the volatility indices improves risk forecasts, we fit an APARCH–X(1,1) model with $\delta^\circ = 1$—as in (15)—with and without the constraint $\pi_1^\circ = 0$, i.e., with and without covariates.[6] We do so based on the sample $\{\varepsilon_1, \dots, \varepsilon_n\}$ ($n \in \{1010, 2010\}$) and proceed as described in Section 3, taking $v = 10$, to obtain risk forecasts $\widehat{z}_{\alpha_n,n}^{\mathrm{AP}}$ and $\widehat{z}_{\alpha_n,n}^{\mathrm{AP–X}}$ ($z \in \{q,\ \mathrm{ES},\ \xi\}$) from the APARCH and APARCH–X model, respectively. We repeat this procedure by rolling a moving window $\{\varepsilon_{j-n+1}, \dots, \varepsilon_j\}_{j=n,\dots,N-1}$ through the sample. This generates one set of risk forecasts $\{\widehat{z}_{\alpha_n,n}^{(j),\mathrm{AP–X}}\}_{j=n,\dots,N-1}$ exploiting option-implied information and another set $\{\widehat{z}_{\alpha_n,n}^{(j),\mathrm{AP}}\}_{j=n,\dots,N-1}$ that does not. In each case, $\widehat{z}_{\alpha_n,n}^{(j),\mathrm{AP–X}}$ and $\widehat{z}_{\alpha_n,n}^{(j),\mathrm{AP}}$ forecast the risk inherent in ε_{j+1} based on past observations $\varepsilon_{j-n+1}, \dots, \varepsilon_j$ (and covariates $x_{j-n+1}, \dots, x_j$).

To compare the two sets of forecasts, we use scoring functions (Gneiting, 2011). We compute the

[4] The data have been taken from *quotes.wsj.com* (ticker symbols: SPX, NDX, RUT and DJIA).

[5] The data have neen taken from *www.cboe.com/products/vix-index-volatility/volatility-indexes* (ticker symbols: VIX, VXN, RVX and VXD).

[6] Francq and Thieu (2019) find $\delta^\circ = 1$ to deliver the best description of S&P 500 volatility dynamics. For simplicity, we also use this choice for the other indices—NASDAQ, Russell 2000, and DJIA.

average scores based on the forecasts $\widehat{z}_{\alpha_n,n}^{(j),\mathrm{AP–X}}$, $\widehat{z}_{\alpha_n,n}^{(j),\mathrm{AP}}$ and verifying observations ε_{j+1}

$$\overline{S}_{z,\alpha_n}^{\mathrm{AP–X}} = \frac{1}{N-n}\sum_{j=n}^{N-1} S_{z,\alpha_n}(\widehat{z}_{\alpha_n,n}^{(j),\mathrm{AP–X}}, \varepsilon_{j+1}) \quad \text{and} \quad \overline{S}_{z,\alpha_n}^{\mathrm{AP}} = \frac{1}{N-n}\sum_{j=n}^{N-1} S_{z,\alpha_n}(\widehat{z}_{\alpha_n,n}^{(j),\mathrm{AP}}, \varepsilon_{j+1}) \tag{16}$$

based on the scoring functions

$$S_{z,\alpha}(x,y) = \begin{cases} (1-\alpha - I_{\{x<y\}})(x-y), & \text{if } z = q; \\ |1-\alpha - I_{\{x<y\}}|(x-y)^2, & \text{if } z = \xi. \end{cases}$$

Gneiting (2011) shows that these scoring functions are suitable in the sense of being strictly consistent. Recall that scoring functions are negatively oriented, so that lower scores are preferable. Since ES is not elicitable (Gneiting, 2011, Theorem 11), there exists no suitable scoring function. However, ES is jointly elicitable with VaR (Fissler and Ziegel, 2016). We use the scoring function

$$\begin{aligned} S_{\mathrm{ES},\alpha}(x_V, x_E, y) = \Big[1-\alpha - I_{\{x_V<y\}}\Big]\left[x_V - y\right] + \frac{1}{1-\alpha}G_2(x_E) I_{\{x_V<y\}}\left[y - x_V\right] \\ - G_2(x_E)[x_E - x_V] + \big[\mathcal{G}_2(x_E) - \mathcal{G}_2(y)\big] \end{aligned}$$

for the pair (VaR, ES), where $G_2(x) = e^x/[1+e^x]$ and $\mathcal{G}_2' = G_2$, i.e., $\mathcal{G}_2(x) = \log(1+e^x)$ (Nolde and Ziegel, 2017; Ziegel *et al.*, 2019). The average scores $\overline{S}_{\mathrm{ES},\alpha_n}^{\mathrm{AP–X}}$ and $\overline{S}_{\mathrm{ES},\alpha_n}^{\mathrm{AP}}$ for ES (jointly with VaR) are then calculated as in (16). As is common, all scoring functions are normalized to satisfy $S_{z,\alpha}(y,y) = 0$ ($z \in \{\mathrm{DRM}, \xi\}$) and $S_{\mathrm{ES},\alpha}(y,y,y) = 0$.

Table 2 compares the ratios $\overline{S}_{z,\alpha_n}^{\mathrm{AP}}/\overline{S}_{z,\alpha_n}^{\mathrm{AP–X}}$ ($z \in \{q,\ \mathrm{ES},\ \xi\}$) of the average scores for different sample sizes $n \in \{1010, 2010\}$ and probability levels $1-\alpha_n$. Since lower scores are better, score ratios above 1 indicate improved risk forecasts using covariate information. We use a standard Diebold and Mariano (1995) test to detect statistically significant differences in average scores. Consistent with Blair *et al.* (2001), we find that across almost all indices and risk measures, including the respective volatility indices as covariates in the model improves risk forecasts. The larger the sample size, the larger the improvements in terms of both score differences and statistical significance. This may be explained as follows. More observations lead to less noise in estimating the additional parameter π_1° in the APARCH–X model and, hence, the advantage of the additional covariate can emerge more clearly.

We also generate a third set of risk forecasts by proceeding as before, except that now we fit a benchmark GARCH(1,1) model to the returns. The corresponding average score is denoted by $\overline{S}_{z,\alpha_n}^{\mathrm{G}}$ and we report the score ratios $\overline{S}_{z,\alpha_n}^{\mathrm{G}}/\overline{S}_{z,\alpha_n}^{\mathrm{AP}}$ in Table 3. We do so for two reasons. First, we want to examine the benefit of modeling different powers and asymmetry—allowed in our framework, but not in that of Chan *et al.* (2007) and Hoga (2019a+)—in risk forecasting. In other words, we want to examine the advantages of the APARCH specification (without covariates) over a GARCH

z	n	$1-\alpha_n$	S&P 500	NASDAQ	Russell 2000	DJIA
q	1010	1%	1.069*	1.010	1.030*	1.037
		0.5%	1.089	0.976	1.053**	1.082
		0.1%	1.031***	0.999	1.053***	1.092
	2010	0.5%	1.156***	1.069***	1.087***	1.059**
		0.1%	1.168***	1.131***	1.134***	1.074***
		0.05%	1.183***	1.140***	1.146***	1.048***
ES	1010	1%	1.069*	1.010	1.030*	1.037
		0.5%	1.088	0.976	1.053**	1.080
		0.1%	1.032***	1.000	1.052***	1.090
	2010	0.5%	1.155***	1.069***	1.087***	1.059**
		0.1%	1.167***	1.130***	1.133***	1.073***
		0.05%	1.181***	1.138***	1.145***	1.047***
ξ	1010	1%	1.048	1.003	1.035	1.063
		0.5%	1.094	0.988	1.047*	1.103
		0.1%	1.083	0.926***	1.085***	1.089
	2010	0.5%	1.141***	1.055**	1.080**	1.094***
		0.1%	1.307***	1.163***	1.207***	1.106***
		0.05%	1.351***	1.203***	1.249***	1.122***

Table 2: Score ratios $\overline{S}^{\mathrm{AP}}_{z,\alpha_n}/\overline{S}^{\mathrm{AP\text{–}X}}_{z,\alpha_n}$ ($z \in \{q, \mathrm{ES}, \xi\}$) for forecasts based on n observations and volatility indices as covariates. Significantly different average scores at the 10%/5%/1%-level are indicated by a */**/***.

specification. Second, comparing $\overline{S}^{\mathrm{G}}_{z,\alpha_n}/\overline{S}^{\mathrm{AP}}_{z,\alpha_n}$ with the ratios $\overline{S}^{\mathrm{AP}}_{z,\alpha_n}/\overline{S}^{\mathrm{AP\text{–}X}}_{z,\alpha_n}$ gives us a better sense of the latter's magnitudes reported in Table 2. As expected, Table 3 shows that the leverage effect and modelling of volatility itself (instead of its square) quite generally lead to superior risk forecasts. Except perhaps for the Russell 2000, the incremental risk forecasting improvements of including volatility indices in the APARCH specification are roughly equally large as those of allowing for modelling different powers and asymmetry in the GARCH specification. Thus, including covariates may be as important for risk forecasting as allowing for the well-established stylized facts of asymmetry and strong autocorrelation in absolute returns.

4.2 Realized Volatility Measures as Covariates

For ease of comparison with the previous subsection, we again consider the S&P 500, NASDAQ, Russell 2000 and DJIA during the time period 1/1/2004–12/31/2018. As a covariate, we use the median realized variance (MedRV) based on 5-minute intra-day returns.[7] MedRV was proposed by Andersen *et al.* (2012) as a jump-robust measure of integrated variance. We take the square root of

[7]All data have been taken from the realized library of Heber *et al.* (2009), available at *realized.oxford-man.ox.ac.uk*.

z	n	$1-\alpha_n$	S&P 500	NASDAQ	Russell 2000	DJIA
q	1010	1%	1.075*	1.081**	1.036	1.075**
		0.5%	1.058	1.094***	1.017	1.033
		0.1%	1.088***	1.100***	1.015	1.019
	2010	1%	1.051	1.083	0.989	1.026
		0.5%	1.046	1.106**	0.940***	1.090
		0.1%	1.004	1.167	0.923***	1.111***
ES	1010	1%	1.077*	1.082**	1.037	1.076**
		0.5%	1.059	1.095***	1.017	1.034
		0.1%	1.090***	1.101***	1.015	1.019
	2010	0.5%	1.052	1.084	0.990	1.026
		0.1%	1.046	1.106**	0.939***	1.090
		0.05%	1.004	1.168	0.922***	1.111***
ξ	1010	1%	1.134**	1.134***	1.070**	1.126*
		0.5%	1.132*	1.156***	1.065*	1.121
		0.1%	1.115	1.173***	1.044	1.069
	2010	0.5%	1.098	1.182*	1.027	1.097*
		0.1%	1.015	1.150**	0.910**	1.102*
		0.05%	0.986	1.132**	0.851***	1.137***

Table 3: Score ratios $\overline{S}^{\mathrm{G}}_{z,\alpha_n}/\overline{S}^{\mathrm{AP}}_{z,\alpha_n}$ ($z \in \{q, \mathrm{ES}, \xi\}$) for forecasts based on n observations. Significantly different average scores at the 10%/5%/1%-level are indicated by a */**/***.

MedRV for it to be in the same unit of measurement as ε_{t-j} in (2) (recall that we take $\delta^\circ = 1$) and denote the resulting time series by $\{x_1, \dots, x_N\}$. Unlike in the previous subsection, we now use open-to-close log-returns $\{\varepsilon_1, \dots, \varepsilon_N\}$ for the indices, since realized volatility is only an intra-day measure that does not capture overnight information. Otherwise we use the same notation and proceed as before.

Table 4 again displays the average score ratios $\overline{S}^{\mathrm{AP}}_{z,\alpha_n}/\overline{S}^{\mathrm{AP-X}}_{z,\alpha_n}$ ($z \in \{q,\ \mathrm{ES},\ \xi\}$). This time the score ratios are mostly below 1, most dramatically so for the NASDAQ and the Russell 2000. This indicates that inclusion of the high-frequency measure MedRV does not help in producing more accurate risk forecasts in our APARCH–X framework. This result is robust to the choice of the high-frequency volatility measure. Using other measures, such as the simple realized variance or bipower variation, does not change this qualitative conclusion.

The results of Table 4 are somewhat at odds with Blair *et al.* (2001) and Koopman *et al.* (2005), who generally find intra-day data to be useful. Yet, it must be kept in mind that they compared volatility forecasts for a different index (the S&P 100) during a different time period using a slightly different model. In their GARCH-type models, Shephard and Sheppard (2010) and Hansen *et al.* (2012) also find high-frequency volatility measures to be useful for out-of-sample volatility prediction.

z	n	$1-\alpha_n$	S&P 500	NASDAQ	Russell 2000	DJIA
q	1010	1%	0.979	0.908**	0.918**	0.989
		0.5%	0.987	0.844**	0.881**	0.996
		0.1%	1.068***	0.774*	0.865	1.012
	2010	0.5%	0.993	0.786**	0.879*	1.020
		0.1%	0.974	0.598**	0.701*	1.063***
		0.05%	1.081***	0.531*	0.607*	1.079***
ES	1010	1%	0.979	0.907**	0.918**	0.989
		0.5%	0.987	0.844**	0.881**	0.996
		0.1%	1.068***	0.774*	0.865	1.012
	2010	0.5%	0.993	0.786**	0.879*	1.020
		0.1%	0.974	0.598**	0.700*	1.063***
		0.05%	1.081***	0.531*	0.607*	1.078***
ξ	1010	1%	1.012	0.968	0.978	1.025
		0.5%	1.018	0.943	0.959	1.022
		0.1%	1.064***	0.999	0.944	1.046
	2010	0.5%	1.013	0.870	0.923	1.035***
		0.1%	1.018	0.716**	0.844	1.100***
		0.05%	1.035	0.698*	0.802	1.144***

Table 4: Score ratios $\overline{S}^{\mathrm{AP}}_{z,\alpha_n}/\overline{S}^{\mathrm{AP\text{-}X}}_{z,\alpha_n}$ ($z \in \{q, \mathrm{ES}, \xi\}$) for forecasts based on n observations and MedRV as covariates. Significantly different average scores at the 10%/5%/1%-level are indicated by a */**/***.

The differences we find in usefulness of volatility indices and realized volatility measures may be explained as follows. Recall that in this section, we have forecast volatility using (2) with $\delta^{\circ} = 1$ and $p = q = 1$. It appears that the additional predictive content of (the square root of) MedRV—a high-frequency proxy of volatility—above and beyond that of (the square root of) squared returns—a low frequency proxy of volatility—is rather small. Although MedRV is a less noisy proxy of volatility than the squared return, its inclusion leads to more noise from the additional parameter that has to be estimated. This may be exacerbated by the collinearity of MedRV and the squared return, which adds to the estimation noise. This issue is clearly reduced if volatility indices are used as covariates instead. These do not provide more 'time series' information, but provide 'cross-sectional' information regarding market participant's expectations about future volatility. Thus, intuitively, volatility indices provide more incremental information above and beyond squared returns than simply another (albeit less noisy) proxy of volatility.

5 Conclusion

Practitioners routinely incorporate exogenous variables in volatility models to forecast risk. To allow them to assess the uncertainty in those forecasts, we derive asymptotic theory for EVT-based DRM and expectile forecasts in APARCH–X models. Our framework allows risk forecasts to be improved in two ways. First, volatility forecasts may be improved by allowing for leverage effects, modelling of powers of volatility, and incorporating covariates, such as volatility indices or realized volatility measures. Second, under a Pareto-type tail assumption, the risk incorporated in the innovations can be estimated more efficiently using EVT-based estimators. We derive asymptotic forecast intervals for quite general risk measures, which provide valuable additional information beyond the point forecast. In simulations, we generally find that the forecast intervals provide reasonable coverage, except perhaps for 'not too extreme' risk measures. Using data on major US stock indices, we find that inclusion of covariates in the volatility model may or may not improve risk forecasts. The expected future volatility derived from option prices seems to improve risk forecasts, whereas incorporating additional (past) information in the form of high-frequency measures does not. Thus, as a possible avenue for future research, it may be of interest to study HEAVY- or Realized GARCH-style models not with realized volatility measures, but with volatility indices as additional drivers of volatility dynamics.

References

Andersen TG, Bollerslev T, Christoffersen PF, Diebold FX. 2013. Financial risk measurement for financial risk management. In Constantinides GM, Harris M, Stulz RM (eds.) *Handbook of the Economics of Finance*, vol. 2, Amsterdam: Elsevier, 1 edn., pages 1127–1220.

Andersen TG, Dobrev D, Schaumburg E. 2012. Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics* **169**: 75–93.

Artzner P, Delbaen F, Eber JM, Heath D. 1999. Coherent measures of risk. *Mathematical Finance* **9**: 203–228.

Askanazi R, Diebold FX, Schorfheide F, Shin M. 2018. On the comparison of interval forecasts. *Journal of Time Series Analysis* **39**: 953–965.

Basel Committee on Banking Supervision. 2019. *Minimum Capital Requirements for Market Risk*. Basel: Bank for International Settlements (https://www.bis.org/bcbs/publ/d457.pdf).

Bee M, Dupuis DJ, Trapin L. 2016. Realizing the extremes: Estimation of tail-risk measures from a high-frequency perspective. *Journal of Empirical Finance* **36**: 86–99.

Bee M, Dupuis DJ, Trapin L. 2018. Realized extreme quantile: A joint model for conditional quantiles and measures of volatility with EVT. *Journal of Applied Econometrics* **33**: 398–415.

Bee M, Dupuis DJ, Trapin L. 2019. Realized peaks over threshold: A time-varying extreme value approach with high-frequency based measures. *Journal of Financial Econometrics* **17**: 254–283.

Bellini F, Di Bernardino E. 2017. Risk management with expectiles. *The European Journal of Finance* **23**: 487–506.

Blair BJ, Poon SH, Taylor SJ. 2001. Forecasting S&P 100 volatility: The incremental information content of implied volatilites and high-frequency index returns. *Journal of Econometrics* **105**: 5–26.

Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**: 307–327.

Chan NH, Deng SJ, Peng L, Xia Z. 2007. Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. *Journal of Econometrics* **137**: 556–576.

Daníelsson J. 2011. *Financial Risk Forecasting.* Chichester: Wiley.

Daníelsson J, de Haan L, Ergun LM, de Vries CG. 2016. Tail index estimation: Quantile driven threshold selection (Available at SSRN: *http://dx.doi.org/10.2139/ssrn.2717478*).

Daouia A, Girard S, Stupfler G. 2018. Estimation of tail risk based on extreme expectiles. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**: 263–292.

Davidson J. 1994. *Stochastic Limit Theory.* Oxford: Oxford University Press.

de Haan L, Ferreira A. 2006. *Extreme Value Theory.* New York: Springer.

Diebold FX, Mariano RS. 1995. Comparing predictive accuracy. *Journal of Business & Economic Statistics* **13**: 253–263.

Ding Z, Granger CWJ, Engle RF. 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* **1**: 83–106.

Einmahl JHJ, de Haan L, Zhou C. 2016. Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78**: 31–51.

El Methni J, Stupfler G. 2017. Extreme versions of Wang risk measures and their estimation for heavy-tailed distributions. *Statistica Sinica* **27**: 907–930.

Fissler T, Ziegel JF. 2016. Higher order elicitability and Osband's principle. *The Annals of Statistics* **44**: 1680–1707.

Francq C, Thieu LQ. 2019. QML inference for volatility models with covariates. *Econometric Theory* **35**: 37–72.

Ghalanos A. 2019. *rugarch: Univariate GARCH models*. R package version 1.4-1.

Glosten LR, Jagannathan R, Runkle DE. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance* **48**: 1779–1801.

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

Hall P, Yao Q. 2003. Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* **71**: 285–317.

Han H, Kristensen D. 2014. Asymptotic theory for the QMLE in GARCH-X models with stationary and nonstationary covariates. *Journal of Business & Economic Statistics* **32**: 416–429.

Hansen PR, Huang Z, Shek HH. 2012. Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics* **27**: 877–906.

Heber G, Lunde A, Shephard N, Sheppard KK. 2009. *Oxford-Man Institute's Realized Library (Version 0.3)*. University of Oxford: Oxford-Man Institute.

Hill B. 1975. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* **3**: 1163–1174.

Hoga Y. 2019+. General limit theory for forecasts of extreme distortion risk measures and expectiles. Revised & Resubmitted (2nd round) to *Journal of Financial Econometrics* .

Hoga Y. 2019a+. Confidence intervals for conditional tail risk measures in ARMA–GARCH models. Forthcoming in *Journal of Business & Economic Statistics* (DOI: 10.1080/07350015.2017.1401543) **XX**: 1–12.

Hsing T. 1991. On tail index estimation using dependent data. *The Annals of Statistics* **19**: 1547–1569.

Hua L, Joe H. 2011. Second order regular variation and conditional tail expectation of multiple risks. *Insurance: Mathematics and Economics* **49**: 537–546.

Kim L, Lee S. 2016. On the tail index inference for heavy-tailed GARCH-type innovations. *Annals of the Institute of Statistical Mathematics* **68**: 237–267.

Kim M, Lee S. 2009. Test for tail index change in stationary time series with Pareto-type marginal distribution. *Bernoulli* **15**: 325–356.

Kim M, Lee S. 2011. Change point test for tail index for dependent data. *Metrika* **74**: 297–311.

Koopman SJ, Jungbacker B, Hol E. 2005. Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance* **12**: 445–475.

Kuester K, Mittnik S, Paolella MS. 2006. Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics* **4**: 53–89.

Newey WK, Powell JL. 1987. Asymmetric least squares estimation and testing. *Econometrica* **55**: 819–847.

Nolde N, Ziegel JF. 2017. Elicitability and backtesting: Perspectives for banking regulation. *The Annals of Applied Statistics* **11**: 1833–1874.

Poon SH, Granger CWJ. 2003. Forecasting volatility in financial markets: A review. *Journal of Economic Literature* **41**: 478–539.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Shephard N, Sheppard K. 2010. Realising the future: Forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* **25**: 197–231.

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

Weissman I. 1978. Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association* **73**: 812–815.

Zakoïan JM. 1994. Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* **18**: 931–955.

Ziegel JF. 2016. Coherence and elicitability. *Mathematical Finance* **26**: 901–918.

Ziegel JF, Krüger F, Jordan A, Fasciati F. 2019. Robust forecast evaluation of expected shortfall. Forthcoming in *Journal of Financial Econometrics* (DOI: 10.1093/jjfinec/nby035) **XX**: 1–26.