Robust Value at Risk Prediction

Loriano Mancini
Swiss Finance Institute and Ecole Polytechnique Fédérale de Lausanne

Fabio Trojani
Swiss Finance Institute and University of Lugano

ABSTRACT
This paper proposes a robust semiparametric bootstrap method to estimate predictive distributions of GARCH–type models. The method is based on a robust estimation of parametric GARCH models and a robustified resampling scheme for GARCH residuals that controls bootstrap instability due to outlying observations. A Monte Carlo simulation shows that our robust method provides more accurate Value at Risk (VaR) forecasts than classical methods, often by a large extent, especially for several days ahead horizons and/or in presence of outlying observations. An empirical application confirms the simulation results. The robust procedure outperforms in backtesting several other VaR prediction methods, such as RiskMetrics, CAViaR, historical simulation, and classical filtered historical simulation methods. We show empirically that robust estimation reduces tail estimation risk, providing more accurate and more stable VaR prediction intervals over time. (*JEL*: C14, C15, C23, C59)

KEYWORDS: backtesting, breakdown point, *M*-estimator, extreme value theory

Large portfolios of traded assets held by many financial firms have made the measurement of market risk, that is, the risk of losses on the trading book due to adverse market movements, a primary concern for regulators, and risk managers.

The Basel Committee (1996) requires that financial firms hold a certain amount of capital against market risk. This capital is called Value at Risk (VaR) and must be sufficient to cover losses on the trading book over a 10-day holding period 99% of the times. In practice, VaR is computed for several holding periods and confidence levels, such as 95% confidence level and horizon of 1 day.

From a statistical viewpoint, VaR is the quantile of the profit and loss (P&L) distribution of a portfolio over a certain holding period. Hence, a key issue in implementing VaR and related risk measures is to obtain accurate estimates for the tails of conditional P&L distributions.

Semiparametric methods, commonly called filtered historical simulation (FHS) methods, have been found to provide rather accurate estimates of P&L distributions; see, for example, Pritsker (1997), Hull and White (1998), Diebold, Schuermann, and Stroughair (1998), Barone-Adesi, Giannopoulos, and Vosper (1999), McNeil and Frey (2000), Pritsker (2001), and Kuester, Mittnik, and Paolella (2006). In FHS methods, parametric GARCH-type models are typically fitted to historical returns using pseudo maximum likelihood (PML). Then GARCH residuals are resampled using bootstrap methods. The FHS methods allow for time-varying conditional moments of returns (via GARCH-type models) and non-parametric structures in conditional distribution of returns (because innovation distributions are estimated nonparametrically). The last feature is crucial in applications and avoids too simplistic assumptions on return conditional distributions, such as normality; see, for example, JP Morgan's RiskMetrics (1995).

In this paper, we propose a general robust semiparametric bootstrap method to estimate predictive distributions of asset returns in GARCH-type volatility models. The method allows for general parametric specifications of time-varying conditional mean and volatility of asset returns. As an application, we use the proposed robust method to predict VaR over different forecasting horizons. Our approach achieves robustness in two steps. In the first step, we estimate a parametric GARCH-type model using the optimal bounded influence estimator in Mancini, Ronchetti, and Trojani (2005). In the second step, we fit generalized Pareto distribution (GPD) using the robust estimator in Dupuis (1999) and Juárez and Schucany (2004) to the tails of GARCH residuals distribution and we resample from this distribution. In order to ensure robustness of the whole procedure, both robustification steps are necessary.

PML estimators with Gaussian pseudo densities are often used to fit GARCH-type models and they feature a number of convenient theoretical properties. For instance, they achieve maximal efficiency (being maximum likelihood) when returns are indeed conditionally Gaussian. Moreover, even under non-Gaussian returns, they imply consistent estimation of the parameters in the conditional mean and variance functions, provided the latter are correctly specified. Finally, in some cases, Gaussian PML estimators coincide with weighted nonlinear least squares estimators and can achieve the semiparametric efficiency bound for estimating a correctly specified conditional moment function.

In the more general situation of (i) potentially misspecified conditional mean and variance functions and (ii) non Gaussian returns, these convenient properties of PML can break down. In this case, PML implicitly estimates a pseudo true value for the conditional GARCH dynamics, which minimizes the Kullback–Leibler discrepancy between the unknown conditional density of returns and a parametric family of Gaussian pseudo likelihoods. A major issue is that even if the relevant degree of misspecification might be small, in the sense that the true conditional moments of the data-generating process might be only slightly different from those specified under the parametric GARCH assumption, there is no guarantee that the conditional moment function implied by the theoretical pseudo true value will again be only slightly different from the true one; see, for example, Sakata and White (1998) and Mancini, Ronchetti, and Trojani (2005) for some concrete evidence on this point. Similarly, under such circumstances, there is no guarantee that PML can produce estimates of the pseudo true conditional mean and variance functions that are comparably efficient as in the case of a correct specification of the latter. These features of Gaussian PML are strongly related to the functional structure of these estimators, which implies excess sensitivity of pseudo true values and PML asymptotic covariance matrices with respect to the underlying data distribution. Such a sensitivity or nonrobustness of these estimators can be problematic because even a moderate misspecification of the assumed parametric model can lead to either strongly biased or inefficient results, for example, in terms of the implied mean square error (MSE) for the estimated conditional moments.

In this paper, we apply a class of robust *M*-estimators for GARCH-type models proposed in Mancini, Ronchetti, and Trojani (2005). These estimators ensure smoothness of the implied pseudo true value with respect to the underlying data distribution using a set of robustified moment conditions defined by a bounded estimating function. The robust estimating function is obtained from the Gaussian PML score function by a downweighting procedure. This procedure bounds the potential damaging effects of data points generating a large sensitivity of theoretical PML pseudo true values with respect to the underlying data distribution. Our estimators provide robust, that is, smooth, estimation results in an abstract nonparametric neighborhood of a fixed reference model, which we take to be a GARCH-type model with Gaussian errors. In general, a trade off between robustness (with respect to model misspecification) and efficiency (under a correct model specification) emerges when choosing between robust and PML estimators, and the researcher has to decide on this trade off. It is an empirical question whether the quantitative effects of a small misspecification of GARCH-type models, for example, in the form of an incorrect variance dynamics, can lean the trade off in favor of using robust methods in the estimation of GARCH-models for predicting VaR. Sakata and White (1998) and Mancini, Ronchetti, and Trojani (2005), among others, show that such a favorable trade off exists when estimating conditional variance dynamics already under a moderate model misspecification. In our Monte Carlo simulations, we find that robust GARCH estimators combined with a robust extreme value estimator for the conditional tails of returns produce lower

MSEs in forecasting the true VaR already under very small misspecifications of the assumed parametric model.

The issue of the impact of a model misspecification on the statistical properties of an estimator is even more pronounced when fitting innovation tail distributions, as required for VaR predictions. For instance, generalized Pareto distribution (GPD) is typically used in extreme value theory to model tail distributions of returns above a given threshold. Theoretically, the GPD is an asymptotic approximation of the tail, which improves in the limit as the threshold goes to infinity (or to the end point of the distribution). In practice, however, the threshold is fixed and the error implied by the GPD approximation for the true tail distribution can have nontrivial consequences for the properties of PML estimators based on a GPD pseudo density. This is so because the GPD PML score leads to an unbounded estimating function that implies nonsmooth pseudo true values with respect to variations of the underlying data distribution. In this context, the robust GPD estimator in our approach is a natural estimator because it is based on a bounded estimating function that explicitly downweights the damaging effects of a potential misspecification of the tail specified by the GPD, thus providing more accurate estimators when such deviations are indeed present in the data. We confirm this intuition in several Monte Carlo simulations, showing that robust GPD estimator provides more accurate quantile estimates (e.g., in terms of a lower MSE) under a number of realistic specifications of the tail. These estimates are also much less sensitive to the choice of threshold levels than classical methods. The latter is a further desirable property of the robust estimator because in applications the selection of the threshold level is a difficult task.

Another important issue is the nonrobustness of several resampling procedures used to compute VaR estimates at several days ahead horizons. It has been recognized that a few large observations are sufficient to cause the break down of quantile estimates based on nonparametric residual bootstrap; see, among others, Singh (1998), Gagliardini, Trojani, and Urga (2005), Davidson and Flachaire (2007), and Camponovo, Scaillet, and Trojani (2009, 2010). We find that standard nonparametric bootstrap procedures for VaR computation have a very low breakdown point (BP), meaning that VaR forecasts can be heavily affected already by a few large observations, especially when longer forecast horizons of, for example, 10 days are considered. Using the robust GPD estimator in our approach, we are also able to develop a resampling procedure for VaR forecasting that controls for the instability generated by outlying observations in estimated GARCH residual distributions; see also Cowell and Victoria-Feser (1996).

We perform an extensive Monte Carlo simulation and show that our robust method provides more accurate VaR predictions than classical methods under conditionally non-Gaussian data, in particular, for several days ahead horizons. When the GARCH model used to predict VaR is not exactly the same as the true data-generating process, robust VaR predictions have mean square prediction errors several times smaller than those of classical procedures. In nearly all Monte Carlo experiments, our robust procedure has the lowest mean square prediction errors,

often by a large extent. In contrast to classical methods, our procedure never fails validation tests at 10% confidence level.

The simulation evidence is confirmed by the real data application. We backtest VaR prediction methods using about 20 years of S&P 500, Dollar-Yen, Microsoft, and Boeing historical returns. We compare our method to several alternative VaR prediction procedures, such as (i) historical simulation, (ii) RiskMetrics, (iii) semi-parametric GARCH model of Engle and Gonzalez-Rivera (1991), (iv) CAViaR model of Engle and Manganelli (2004), and for 10-day ahead VaR predictions, (v) GARCH model applied directly to 10-day asset returns. Only our robust method passes all validation tests at 10% confidence level. Moreover, we find that the reduction in tail estimation risk of our robust procedure provides more accurate and more stable VaR prediction intervals over time. For instance, in the case of S&P 500 and Boeing, robust VaR prediction intervals are nearly 20% narrower and 50% less volatile than classical ones. Given the higher accuracy of robust VaR predictions documented in violation tests, the stability over time of robust VaR profiles is a feature that allows financial firms to adapt properly risky positions to VaR limits more smoothly and thus more efficiently.

Section 1 introduces classical and robust semiparametric bootstrap methods for VaR predictions. Section 2 presents Monte Carlo evidence on VaR predictions under different forms of conditional nonnormal returns. Section 3 presents the real data application and backtesting for four financial time series. Section 4 concludes.

1 SETTING

1.1 Return Dynamics and Measures of Market Risk

Let $\mathcal{Y} := \{Y_t\}_{t \in \mathbb{Z}}$ be a strictly stationary time series process on probability space $(\mathbb{R}^\infty, \mathcal{F}, \mathbb{P}_*)$, modeling the daily rate of return on a financial asset with price P_t at time t, that is, $Y_t := P_t/P_{t-1} - 1$. We assume that distribution $\mathbb{P}_*$ can be "approximated" by a parametric reference model $\mathbb{P}_{\theta_0}$ in the parametric family $\mathcal{P} := \{\mathbb{P}_\theta, \theta \in \Theta \subseteq \mathbb{R}^p\}$. Even if $\mathbb{P}_*$ might not be a member of $\mathcal{P}$, so that the parametric family is misspecified, we assume that $\mathbb{P}_*$ belongs to a nonparametric neighborhood of $\mathbb{P}_{\theta_0}$, denoted by $\mathcal{U}(\mathbb{P}_{\theta_0})$. The neighborhood is assumed to be small, in the sense that the distribution distance between $\mathbb{P}_*$ and $\mathbb{P}_{\theta_0}$ is assumed to be moderate.

Remark 1.1. Neighborhood $\mathcal{U}(\mathbb{P}_{\theta_0})$ can be defined using different metrics between distributions, and it represents a proximity of similar models used to provide an approximate statistical description of $\mathbb{P}_*$. Robust estimators are designed to provide a smooth statistical behavior over neighborhood $\mathcal{U}(\mathbb{P}_{\theta_0})$ in order to make the estimator's properties not excessively dependent on which specific direction of misspecification of $\mathbb{P}_{\theta_0}$ might indeed be present in the data. In applications, the specific size of neighborhood $\mathcal{U}(\mathbb{P}_{\theta_0})$ is in most cases only implicitly fixed by the

degree of robustness imposed on the estimator used. Intuitively, the more robust an estimator the broader the implicit neighborhood of potential model misspecifications that are considered as possible relevant data-generating processes in the robust approach.

Under $\mathbb{P}_{\theta_0}$, we assume that process $\mathcal{Y}$ satisfies the dynamic model

$$Y_t = \mu_t(\theta_0) + \sigma_t(\theta_0)Z_t, \tag{1}$$

where $\mu_t(\theta_0)$ and $\sigma_t^2(\theta_0)$ parameterize the conditional mean and conditional variance of Y_t, given information $\mathcal{F}_{t-1}$ up to time $t-1$. Under $\mathbb{P}_{\theta_0}$, innovations Z are a strong white noise, that is, $Z_t \sim \text{IIN}(0,1)$. Given $Y_1^m := \{Y_1, \ldots, Y_m\}$, denote by $\mathbb{P}_*^m$ ($\mathbb{P}_{\theta_0}^m$) the m-dimensional marginal distribution of Y_1^m under $\mathbb{P}_*$ ($\mathbb{P}_{\theta_0}$). $F_{t,t+h}$ is the conditional distribution function of h days returns $Y_{t,t+h} := P_{t+h}/P_t - 1$ under $\mathbb{P}_*$, given information $\mathcal{F}_t$. For $0 < \alpha < 1$ and horizon h days, let $y_{t,t+h}^\alpha$ denote the α-quantile of $F_{t,t+h}$, that is, $y_{t,t+h}^\alpha := \inf\{y \in \mathbb{R} : F_{t,t+h}(y) \geqslant \alpha\}$. For an asset with market price P_t, the VaR at time t, confidence level α, and horizon h days, $\text{VaR}_{t,t+h}^\alpha$, is defined by[1]

$$\alpha = \mathbb{P}_*(P_{t+h} - P_t < -\text{VaR}_{t,t+h}^\alpha | \mathcal{F}_t). \tag{2}$$

Hence, $-\text{VaR}_{t,t+h}^\alpha = P_t\, y_{t,t+h}^\alpha$ is the α-quantile of the conditional P&L distribution under $\mathbb{P}_*$ over the next h days, given $\mathcal{F}_t$.[2] Another measure of market risk is the expected shortfall (ES, Artzner et al. 1999), $S_{t,t+h}^\alpha := E_*[Y_{t,t+h}|Y_{t,t+h} < y_{t,t+h}^\alpha, \mathcal{F}_t]$, where $E_*[\cdot]$ denotes expectation with respect to $\mathbb{P}_*$. For horizon $h = 1$ day,

$$y_{t,t+1}^\alpha = \mu_{t+1}(\theta_0) + \sigma_{t+1}(\theta_0)z_\alpha, \quad S_{t,t+1}^\alpha = \mu_{t+1}(\theta_0) + \sigma_{t+1}(\theta_0)E_*[Z|Z < z_\alpha],$$

where z_α is the α-quantile of the distribution of Z. Estimation of VaR or ES can be obtained by estimating model (1) and tail distribution of residuals Z. For longer horizons $h \geqslant 2$, estimating model (1) is only the starting point to compute market risk measures. Joint conditional distributions of $\{Z_{t+1}, \ldots, Z_{t+h}\}$, $\{\mu_{t+1}, \ldots, \mu_{t+h}\}$, and $\{\sigma_{t+1}, \ldots, \sigma_{t+h}\}$ have to be estimated, a task that is considerably more difficult. FHS methods estimate $F_{t,t+h}$ using semiparametric bootstrap of model (1) over horizon $[t, t+h]$. Our goal is to develop robust semiparametric bootstrap methods for estimating $F_{t,t+h}$.

1.2 Estimation of GARCH-type Models

The parameters of model (1) are usually estimated by PML; see White (1982), Gouriéroux, Monfort, and Trognon (1984), and Bollerslev and Wooldridge (1992).

---

[1]For brevity, Equation (2) assumes a continuous P&L distribution.
[2]See, for example, Duffie and Pan (1997) and Gouriéroux, Laurent, and Scaillet (2000) for a general discussion on conditional VaR, and Aït-Sahalia and Lo (2000) for an economic interpretation of VaR.

The functional pseudo maximum likelihood estimator (PMLE) $a(\cdot)$ is defined by

$$E_*[s(Y_1^m; a(\mathbb{P}_*^m))] = 0,$$

where the Gaussian score function

$$s(Y_1^m; \theta) := \frac{1}{\sigma_m^2(\theta)} \frac{\partial \mu_m(\theta)}{\partial \theta} \varepsilon_m(\theta) + \frac{1}{2\sigma_m^2(\theta)} \frac{\partial \sigma_m^2(\theta)}{\partial \theta} \left(\frac{\varepsilon_m^2(\theta)}{\sigma_m^2(\theta)} - 1 \right) \qquad (3)$$

and $\varepsilon_m(\theta) := \sigma_m(\theta) Z_m$. PMLE has a number of convenient and useful properties. First, when true distribution, $\mathbb{P}_*$, coincides with $\mathbb{P}_{\theta_0}$, PMLE is indeed the Maximum Likelihood estimator. Second, if the conditional mean and variance functions $\mu_m(\theta_0)$ and $\sigma_m^2(\theta_0)$ are correctly specified, they are consistently estimated even under non-Gaussianity of the errors distribution. However, an important issue is that even when the degree of misspecification might be small, for example, with true conditional moments that might be only slightly different from those under the parametric model, there is no guarantee that the pseudo true conditional moments will remain only slightly different from the true ones. This feature of Gaussian PML is determined by the functional structure of these estimators, which implies excess sensitivity of pseudo true values and PML asymptotic covariance matrices with respect to the underlying data distribution. Such a sensitivity or nonrobustness of these estimators is due to the fact that the estimating function of Gaussian PMLE is unbounded.

In this paper, we apply a class of robust M-estimators for GARCH-type models proposed in Mancini, Ronchetti, and Trojani (2005). These estimators ensure smoothness of the implied pseudo true value with respect to the underlying data distribution using a set of robustified moment conditions implied by a bounded estimating function. The robust estimating function is obtained from the Gaussian PML score function by a downweighting procedure that bounds the potential damaging effects of data points generating a too large sensitivity of Gaussian PML pseudo true values.

Mancini, Ronchetti, and Trojani (2005) robust estimator is efficient and computationally feasible for highly nonlinear models and compares favorably with other robust estimators, such as robust Generalized Method of Moments estimators (Ronchetti and Trojani 2001) or robust Efficient Method of Moments estimators (Ortelli and Trojani 2005) for time series. It is defined as follows. Let

$$\psi_c(s(Y_1^m; \theta)) := A(\theta)(s(Y_1^m; \theta) - \tau(Y_1^{m-1}; \theta))w(Y_1^m; \theta), \qquad (4)$$

where $s(Y_1^m; \theta)$ is the PML score function defined in Equation (3), and

$$w(Y_1^m; \theta) := \min(1, c\|A(\theta)(s(Y_1^m; \theta) - \tau(Y_1^{m-1}; \theta))\|^{-1}).$$

The robust functional M-estimator $\bar{a}(\cdot)$ of θ is defined by

$$E_*[\psi_c(s(Y_1^m; \bar{a}(\mathbb{P}_*^m)))] = 0, \qquad (5)$$

where nonsingular matrix $A(\theta) \in \mathbb{R}^p \times \mathbb{R}^p$ and $\mathcal{F}_{m-1}$-measurable random vector $\tau(Y_1^{m-1};\theta) \in \mathbb{R}^p$ are determined by the implicit equations

$$E_{\theta_0}[\psi_c(s(Y_1^m;\theta_0))\,\psi_c(s(Y_1^m;\theta_0))^\top] = I, \tag{6}$$

$$E_{\theta_0}[\psi_c(s(Y_1^m;\theta_0))|\mathcal{F}_{m-1}] = 0. \tag{7}$$

The estimating function ψ_c is a truncated version of PML score function (3) because by construction $\|\psi_c(s(Y_1^m;\theta))\| \leqslant c$. The constant c controls for the degree of robustness. When $c = \infty$, the robust estimator $\bar{a}$ in Equation (5) is indeed the PMLE. Section 1.5 discusses how to select c.[3]

To obtain VaR predictions, conditional mean and conditional volatility in model (1) have to be specified; see, for example, Ghysels, Harvey, and Renault (1996). Several GARCH-type models have been proposed in the financial literature. Our robust method can accommodate general specifications for $\mu_t(\theta_0)$ and $\sigma_t(\theta_0)$ that imply different estimating functions but do not change the overall procedure. In our simulations and empirical applications, we adopt a fairly flexible model, namely an autoregressive AR(1) model for the conditional mean $\mu_t(\theta_0)$ and an asymmetric GARCH(1,1) model for the conditional variance $\sigma_t^2(\theta_0)$ (Glosten, Jagannathan, and Runkle 1993),

$$\mu_t(\theta_0) = \rho_0 + \rho_1 Y_{t-1}, \tag{8}$$

$$\sigma_t^2(\theta_0) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2(\theta_0) + \alpha_2 \sigma_{t-1}^2(\theta_0) + \alpha_3 \varepsilon_{t-1}^2(\theta_0) I_{t-1}(\theta_0), \tag{9}$$

where $\alpha_0, \alpha_1, \alpha_2 > 0$, $|\rho_1| < 1$, $\alpha_1 + \alpha_2 + \alpha_3/2 < 1$, $I_{t-1}(\theta_0) = 1$ when $\varepsilon_{t-1}(\theta_0) < 0$ and 0 otherwise. The AR(1) model for μ_t captures potential autocorrelations in daily returns, for instance, due to nonsynchronous transactions in different index components. The parameter $\alpha_3 > 0$ accounts for the leverage effect,[4] that

---

[3]Formally, optimality results in Mancini, Ronchetti, and Trojani (2005) hold for ARCH- but not GARCH-type models. As in Sakata and White (1998), however, we can expect that our robust estimator performs well also under GARCH models with sufficient memory decay. To investigate this point, in the estimating function (4), we approximate the GARCH volatility by an ARCH model. For example in the GARCH(1,1) model,

$$\sigma_t^2(\theta) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2(\theta) + \alpha_2 \sigma_{t-1}^2(\theta) = \sum_{j=0}^{+\infty} \alpha_2^j (\alpha_0 + \alpha_1 \varepsilon_{t-1-j}^2(\theta))$$

$$= \sum_{j=0}^{l-1} \alpha_2^j (\alpha_0 + \alpha_1 \varepsilon_{t-1-j}^2(\theta)) + \alpha_2^l \sigma_{t-l}^2(\theta) =: \sigma_t^2(\theta)_{\text{trunc}} + \alpha_2^l \sigma_{t-l}^2(\theta) \approx \sigma_t^2(\theta)_{\text{trunc}}.$$

For sufficiently large lag l, $\alpha_2^l \sigma_{t-l}^2(\theta) \approx 0$, and the bias of the robust estimator $\bar{a}_{\text{trunc}}$ based on $\sigma_t^2(\theta)_{\text{trunc}}$ is expected to be negligible. Monte Carlo simulation in Section 2.1 studies this issue.

[4]The terminology leverage effect was introduced by Black (1976) who suggested that a large negative return increases the financial and operating leverage and rises equity return volatility; see also Christie (1982). Campbell and Hentschel (1992) suggested an alternative explanation based on the market risk premium and volatility feedback effects; see also Bekaert and Wu (2000). Following common practice, we shall use the terminology leverage effect when referring to the asymmetric reaction of volatility to positive and negative return innovations.

is, negative shocks ($\varepsilon_{t-1}(\theta_0) < 0$) raise future volatility more than positive shocks ($\varepsilon_{t-1}(\theta_0) \geqslant 0$) of the same absolute magnitude. Compared to symmetric GARCH models ($\alpha_3 = 0$), asymmetric GARCH models are better able to fit volatility dynamics of equity and index returns; see, for example, Engle and Ng (1993) and Rosenberg and Engle (2002). To our knowledge, robust estimators of asymmetric GARCH models have not yet been applied in the statistics and econometrics literature.

1.3 Bootstrap Methods of GARCH-type Processes

We study bootstrap methods for GARCH-type processes consisting of two steps. In the first step, we fit model (8)–(9) to historical returns, $y_1, \ldots, y_T$, using either PML or optimal robust estimators, obtaining parameter estimates $\hat{\theta}$. In the second step, to estimate $F_{T,T+h}$, we apply various bootstrap procedures to estimated scaled residuals,

$$\hat{z}_t = \frac{y_t - \mu_t(\hat{\theta})}{\sigma_t(\hat{\theta})}, \quad t = 1, \ldots, T.$$

We denote by $\mathbb{P}_T$ the empirical distribution of estimated scaled residuals $\hat{z}_1, \ldots, \hat{z}_T$.

1.3.1 Nonparametric residual bootstrap and VaR estimation.   Nonparametric residual bootstrap relies on the empirical distribution of GARCH residuals, $\mathbb{P}_T$. Estimation of 1-day ahead VaR forecast, $\hat{y}^\alpha_{T,T+1}$, is easily obtained using the empirical quantile $\hat{z}_\alpha$ of $\mathbb{P}_T$, yielding $\hat{y}^\alpha_{T,T+1} = \mu_{T+1}(\hat{\theta}) + \sigma_{T+1}(\hat{\theta})\,\hat{z}_\alpha$. Estimation of VaR measures for horizons $h \geqslant 2$ days is more involved and obtained by simulation as follows. Select randomly a GARCH innovation from $\mathbb{P}_T$, say, $z_1^\star$, update μ_{T+1} and σ_{T+1}, draw a second innovation, $z_2^\star$, update μ_{T+2} and σ_{T+2}, and so on up to $T + h$. The h-day simulated return is $y^\star_{T,T+h} := \prod_{j=1}^h (1 + y^\star_{T+j}) - 1 = p^\star_{T+h}/p_T - 1$. Repeat the procedure, say, $B = 10000$ times, to obtain an estimate of $F_{T,T+h}$ as the bootstrap distribution $F^\star_{T,T+h}$ of h-day simulated returns $\{y^{\star(b)}_{T,T+h}\}_{b=1}^B$. The h days ahead VaR forecast at level α is given by the empirical α-quantile of distribution $F^\star_{T,T+h}$. The bootstrap method provides an estimate of the entire predictive distribution $F_{T,T+h}$. Other risk measures, such as ES, can be readily computed.

Estimated innovations, $\hat{z}_1, \ldots, \hat{z}_T$, might well include some large observations that can distort VaR predictions. Via the resampling procedure, each of these observations can enter several times in a simulated sample path return, affecting VaR forecasts adversely by making them very volatile. To investigate this issue, we compute the breakdown point (BP) of $\hat{y}^\alpha_{T,T+h}$ based on bootstrap distribution $F^\star_{T,T+h}$. Intuitively, the BP represents the largest amount of outliers in the data that is tolerated by the VaR forecasting procedure. Formally, the BP, b_α, is the smallest fraction of scaled innovations in the original sample that need to go to $-\infty$ in

Table 1 Breakdown point (BP). For different time horizons, h in days, each entry represents the minimal percentage of outliers in estimated scaled innovations $\hat{z}_1, \ldots, \hat{z}_T$ that is sufficient to cause break down of VaR estimates based on FHS method

h (%)	1	2	3	4	5	6	7	8	9	10
$\alpha = 5$	5.00	2.53	1.70	1.27	1.02	0.85	0.73	0.64	0.57	0.51
$\alpha = 1$	1.00	0.50	0.33	0.25	0.20	0.17	0.14	0.13	0.11	0.10

order to force $\hat{y}^\alpha_{T,T+h}$ to go to $-\infty$; see also Singh (1998). Without loss of generality, we assume outliers causing the break down of $\hat{y}^\alpha_{T,T+h}$ to be in the lower tail of the empirical distribution $\mathbb{P}_T$. The BP is given by

$$b_\alpha = 1 - (1 - \alpha)^{1/h}. \tag{10}$$

To understand Equation (10), let η denote the fraction of outliers, that is, estimated innovations that can be potentially very large, in the original sample, $\hat{z}_1, \ldots, \hat{z}_T$. Then

$$\mathbb{P}_T(y^\star_{T,T+h} \text{ has at least one outlier}) = 1 - (1 - \eta)^h.$$

By definition, $\hat{y}^\alpha_{T,T+h}$ is the α-quantile of $F^\star_{T,T+h}$. Therefore, $\hat{y}^\alpha_{T,T+h}$ breaks down when a sufficiently large proportion of simulated returns $y^\star_{T,T+h}$ is corrupted and precisely when

$$\mathbb{P}_T(y^\star_{T,T+h} \text{ has at least one outlier}) \geqslant \alpha.$$

The probability on the left-hand side gives the fraction of corrupted $y^\star_{T,T+h}$ in the simulation. When this fraction is larger than α, $\hat{y}^\alpha_{T,T+h}$ breaks down. Therefore,

$$b_\alpha = \arg\min_\eta \{1 - (1 - \eta)^h \geqslant \alpha\}$$

implying Equation (10).

The BP $b_\alpha \to 0$ when $h \to +\infty$. That is, for a longer horizon h fewer outliers are sufficient to carry $\hat{y}^\alpha_{T,T+h}$ to $-\infty$. For $h = 1$ day, $b_\alpha = \alpha$ as the α-quantile is estimated by the corresponding empirical quantile of $\mathbb{P}_T$. Table 1 presents numerical values of b_α for different horizons h. The low BP for long horizons, such as 0.10% for VaR at 1% level and 10-day horizon, suggests inaccurate VaR forecasts based on nonparametric residual bootstrap, already under a moderate number of large estimated innovations. Monte Carlo simulation in Section 2.4 confirms this conjecture.

1.3.2 Semiparametric residual bootstrap and VaR estimation.
Semiparametric residual bootstrap with extreme value theory (EVT) relies on a different

estimator of the innovation distribution F. Instead of using the empirical distribution $\mathbb{P}_T$, tails of F are estimated semiparametrically. To estimate the upper tail of F, fix a high threshold $\overline{u}$, such as the 90th percentile of $\{\hat{z}_j\}_{j=1}^T$. Then for any $k > \overline{u}$,

$$\mathbb{P}_*(Z_t > k) = \mathbb{P}_*(Z_t > k | Z_t > \overline{u}) \mathbb{P}_*(Z_t > \overline{u})$$

$$= \mathbb{P}_*(Z_t - \overline{u} > k - \overline{u} | Z_t > \overline{u}) \mathbb{P}_*(Z_t > \overline{u}). \tag{11}$$

In Equation (11), $\mathbb{P}_*(Z_t > \overline{u})$ is easily estimated nonparametrically by $\sum_{j=1}^T \mathbf{1}\{\hat{z}_j > \overline{u}\}/T$, where $\mathbf{1}\{z > \overline{u}\} = 1$ when $z > \overline{u}$ and 0 otherwise. Excess distribution $F_{\overline{u}}(k - \overline{u}) := 1 - \mathbb{P}_*(Z_t - \overline{u} > k - \overline{u} | Z_t > \overline{u})$ above threshold $\overline{u}$ is typically approximated by a GPD, $G_{\xi,\beta}$,

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi}, \xi \neq 0, \\ 1 - \exp(-x/\beta), \qquad \xi = 0, \end{cases}$$

whose support is $[0, +\infty)$ for $\xi \geqslant 0$, and $[0, -\beta/\xi]$ for $\xi < 0$; see Embrechts, Klüppelberg, and Mikosch (1997). To estimate lower tail of F, fix a low threshold $\underline{u}$, such as the 10th percentile of $\{\hat{z}_j\}_{j=1}^T$ and apply for every $k < \underline{u}$ the above procedure to excess losses $x = -(k - \underline{u})$ using $G_{\xi,\beta(-\underline{u})}$.

Given GPD parameter estimates $\hat{\xi}^{(1)}, \hat{\beta}^{(1)}$, and $\hat{\xi}^{(2)}, \hat{\beta}^{(2)}$ for lower and upper tails of F, respectively, innovations $z_1^\star, \ldots, z_h^\star$ are sampled from $\mathbb{P}_T$ as follows. For $j = 1, \ldots, h$:

- If $z_j^\star < \underline{u}$, sample a GPD($\hat{\xi}^{(1)}$, $\hat{\beta}^{(1)}$) distributed excess loss x_1 and return $\underline{u} - x_1$.
- If $z_j^\star > \overline{u}$, sample a GPD($\hat{\xi}^{(2)}$, $\hat{\beta}^{(2)}$) distributed excess gain x_2 and return $\overline{u} + x_2$.
- If $\underline{u} \leqslant z_j^\star \leqslant \overline{u}$, return scaled residual $z_j^\star$ itself.

Then semiparametric bootstrap methods rely on the same simulation procedure as in Section 1.3.1 to estimate the predictive distribution $F_{T,T+h}$.

1.4 Tail Estimation

The GPD parameters, $\zeta := (\xi, \beta)^\top$, are usually estimated by PML. The PMLE, $q(\cdot)$, is defined by

$$E_{\mathbb{G}_*}[s_{\text{gpd}}(X; q(\mathbb{G}_*))] = 0,$$

where $\mathbb{G}_*$ is the true tail distribution, and $s_{\text{gpd}}(x; \zeta)$ is the GPD score function,

$$s_{\text{gpd}}(x; \zeta) = \begin{pmatrix} \xi^{-2} \log(1 + \xi x/\beta) - (1 + 1/\xi)(1 + \xi x/\beta)^{-1} x/\beta \\ -\beta^{-1} + (1 + 1/\xi)\beta^{-1}(1 + \xi x/\beta)^{-1} \xi x/\beta \end{pmatrix}. \tag{12}$$

The approximation of the excess distribution, $F_{\bar{u}}$, by a GPD is motivated by the limit result (see Balkema and de Haan 1974; Pickands 1975),

$$\lim_{\bar{u} \to \bar{x}} \sup_{0 \leqslant x < \bar{x} - \bar{u}} |F_{\bar{u}}(x) - G_{\zeta, \beta(\bar{u})}(x)| = 0, \tag{13}$$

where $\bar{x}$ is the (finite or infinite) right end point of F and $\beta(\bar{u})$ is a positive measurable function. Equation (13) implies that GPD describes the tail exactly only in the limit when the threshold $\bar{u}$ approaches the right end point $\bar{x}$. However, in finite samples, the threshold is fixed and the GPD is only an approximation for the true tail distribution.

When the true tail of the GARCH residuals is not a GPD distribution, PML estimator (12) estimates a pseudo true value that minimizes the Kullback–Leibler discrepancy between the true tail and the parametric GPD tail. Since the estimating function of GPD, PML is unbounded, this estimator is not robust and can imply large variations in both pseudo true values and asymptotic variances of the estimator, even when the actual distance between a large part of the true and the parametric tail is small. This feature can generate quite dramatic increases in the MSE of estimated VaRs relative to the case with no misspecification of the tail.

To illustrate this important point, we perform the following simple Monte Carlo experiment. We simulate 2000 observations from a Student-t_5 and fix the threshold at the 0.90 empirical quantile to estimate the tail distribution. Left graph in Figure 1 shows estimation results. The classical EVT method clearly overestimates the whole tail distribution, which is highly affected by a few relatively large observations that indeed do not fit well within the chosen parametric model. The robust EVT estimator introduced in Equations (14) and (15) below produces a
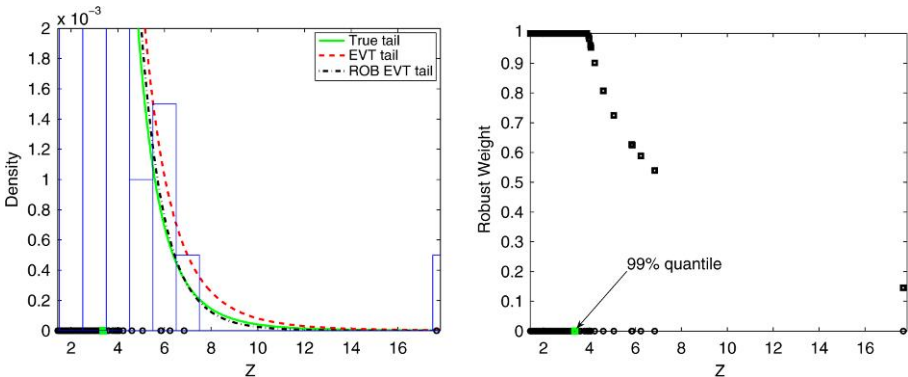

Figure 1 Left graph: estimated right tail of Student-t_5 distribution based on 200 observations above the 0.90 empirical quantile of an i.i.d. random sample of 2000 observations; histogram represents empirical density. Right graph: robust weights in Equation (15). Circles on the x-axis represent simulated Student-t_5 observations above the 0.90 empirical quantile.

much better GPD estimate for the true tail distribution. This is illustrated also by the right graph in Figure 1, which presents the robust weight of each observation implied by this estimator. These weights automatically downweight the observations that are less well captured by the GPD tail. Intuitively, they identify some observations that are too influential in the score function (12) of the classical GPD estimator when compared to the other tail observations. Given that $s_{\mathrm{gpd}}(x;\zeta)$ is unbounded in x, these observations have a strong impact on the classical EVT estimator, inflating the overall tail estimation.

The robust EVT estimator is defined as follows. Given positive constant $c_{\mathrm{gpd}} \geqslant \sqrt{2}$, robust estimator of GPD parameters, $\overline{q}$, is defined by (Dupuis 1999)

$$E_{\mathbf{G}_*}[\psi_c(s_{\mathrm{gpd}}(X;\overline{q}(\mathbf{G}_*)))] = 0, \tag{14}$$

where $s_{\mathrm{gpd}}(x;\zeta)$ is the GPD score function (12) and

$$\psi_c(s_{\mathrm{gpd}}(X;\zeta)) := A(\zeta)(s_{\mathrm{gpd}}(X;\zeta) - \tau(\zeta))w(X;\zeta),$$
$$w(X;\zeta) := \min(1, c_{\mathrm{gpd}}\|A(\zeta)(s_{\mathrm{gpd}}(X;\zeta) - \tau(\zeta))\|^{-1}). \tag{15}$$

Matrix $A(\zeta)$ and vector $\tau(\zeta)$ are solutions of the equations

$$E_{\zeta_0}[\psi_c(s_{\mathrm{gpd}}(X;\zeta_0))\psi_c(s_{\mathrm{gpd}}(X;\zeta_0))^\top] = I$$
$$E_{\zeta_0}[\psi_c(s_{\mathrm{gpd}}(X;\zeta_0))] = 0.$$

Figure 1 shows that the robust estimator provides a nearly perfect tail estimation. This result is achieved by automatically downweighting only a few outlying observations, using the weighting function $w(X;\zeta)$ in Equation (15); see right graph in Figure 1. Only observations above the 0.99 quantile are downweighted, but the larger the tail observation, the lower the robust weight.[5]

The previous discussion is further supported by the following Monte Carlo simulation. We generate 1000 samples of 2000 i.i.d. observations each from a Student-t_5 distribution, resembling model residual distributions. Using different threshold levels, we estimate the 0.99 quantile of the t_5-distribution applying (i) the empirical quantile (HS), (ii) the Hill (1975) estimator, (iii) the classical EVT, and (iv) the robust EVT method.

The simulation allows us to study the precision of quantile estimates and the sensitivity of classical and robust procedures with respect to threshold levels. The choice of threshold level plays a key role in EVT applications because it determines the trade-off between variance and bias of GPD parameter estimates.[6]

---

[5]Several authors have emphasized the instability of PML estimates of GPD when a moderate number of influential points is present in the sample; see, for instance, Cowell and Victoria-Feser (1996) and Juárez and Schucany (2004).

[6]A too high threshold results in too few exceedances and hence high-variance estimators. A too low threshold induces biased estimates as the approximation implied by limit result in Equation (13) can imply large errors.
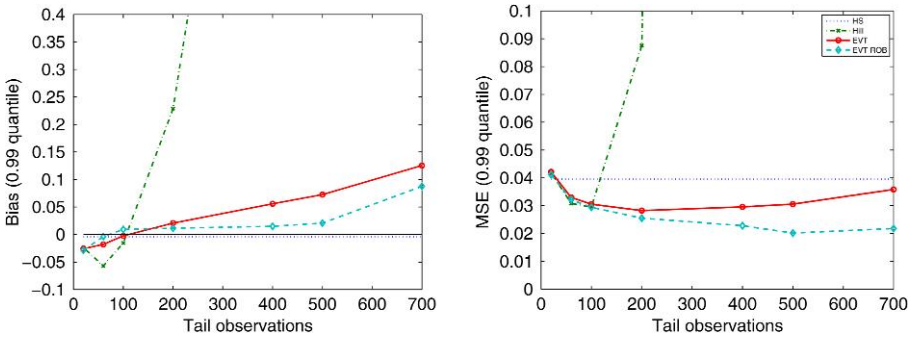
Figure 2 Estimated bias and MSE for various estimators of the 0.99 quantile of a Student-t_5 distribution based on an i.i.d. sample of 2000 data points using different threshold levels, that is, numbers of tail observations.

Figure 2 shows bias and MSE of estimated 0.99 quantile for the four tail estimators, as a function of the chosen threshold level (number of observations in the tail). By definition, the empirical quantile does not depend on thresholds but is generally inaccurate. The Hill estimator is the most sensitive to the threshold level, making its empirical application rather delicate.

The robust EVT method has the lowest MSE for most threshold levels and outperforms classical EVT method consistently. For example, fixing the threshold at 0.90 quantile, that is, using 200 tail observations, classical EVT quantile estimates have MSE 11% larger than robust estimates. For lower thresholds, that is, using more tail observations, classical EVT estimates deteriorate rapidly, while robust EVT estimates are even more accurate in terms of MSE. Both in terms of bias and MSE, accuracy of robust EVT estimates is least sensitive to threshold levels among tail estimators. This is certainly a desirable property of robust EVT method because it is difficult to select thresholds optimally in empirical applications.[7]

In the following Monte Carlo experiments and empirical applications, we take the empirical 10th and 90th quantiles of model residual distributions as threshold levels for estimating lower and higher quantiles, that is, using 200 tail observations. Such a threshold choice is the one suggested by McNeil and Frey (2000) for the classical EVT method and achieves a minimal MSE for this method in the Monte Carlo simulation of Figure 2. The lowest MSE of the robust EVT method is achieved at a lower threshold level (see Figure 2). Therefore, the VaR forecasting performance of robust EVT could be in principle further improved by considering different choices of the threshold level. We do not investigate this issue in more detail in the sequel.

---

[7]See McNeil and Frey (2000) and Gonzalo and Olmo (2004) for further evidence on the choice of the threshold level.

1.5 Choice of Robustness Tuning Constants

The tuning constants c and c_{gpd} in the estimating Equations (5) and (14) control for the degree of robustness of GARCH and GPD estimators, respectively. Following Mancini, Ronchetti, and Trojani (2005), we set such constants to achieve a given asymptotic efficiency under parametric reference models $\mathbb{P}_{\theta_0}$ and $\mathbb{G}_{\zeta_0}$. The relative efficiency of the robust estimator $\bar{a}$ is measured as trace$(V(s;\hat{\theta}_n))/$trace$(V(\psi_c;\hat{\bar{\theta}}_n))$, where $V(s;\hat{\theta}_n)$ and $V(\psi_c;\hat{\bar{\theta}}_n)$ are the asymptotic covariance matrices of the PML and robust estimators, respectively. Relative efficiencies of robust estimators are presented in Mancini and Trojani (2010). For instance, the choice $c = 11$ implies approximately 98% asymptotic relative efficiency. The relative efficiency of $\bar{q}$ is computed analogously.

2 MONTE CARLO SIMULATION

	PML GARCH	Robust GARCH
Empirical dist.	fhs	fhs rob
PML GPD	evt	—
Robust GPD	—	evt rob

The panel above summarizes the four VaR prediction methods studied here. For brevity, the method based on nonparametric residual bootstrap is called fhs. When GARCH dynamics are estimated using the robust estimator (5), we call this method fhs rob; evt rob uses robust estimators both for GARCH dynamics and GPD tail estimations; evt uses PML estimators at both stages. The simulation design allows to evaluate the contribution of each robustification step to the accuracy of VaR predictions. Comparing fhs and fhs rob, VaR predictions allows to assess the potential improvement of VaR forecasts due to robust instead of PML estimation of the GARCH model. In Section 2.4, we compare VaR forecasts using true GARCH parameters. In that setting comparing evt and evt rob allows to assess the potential improvement of VaR forecasts due to robust instead of PML estimation of tail distributions.

We compute out-of-sample VaR forecasts at 1% and 5% confidence levels and horizons $h = 1$ day and $h = 10$ days, under an AR(1), asymmetric GARCH(1,1) model for daily returns. We simulate the following dynamics for $\mathcal{Y} := \{Y_t\}_{t\in\mathbb{Z}}$:

1. *Student-t_5 innovation model.* In this experiment, innovation in model (1) is given by

$$Z_t = ((\nu - 2)/\nu)^{1/2}\, T_\nu, \tag{16}$$

where random variable T_ν has a Student-t distribution with $\nu = 5$ degrees of freedom. Hence, $Z_t \sim$ i.i.d.$(0,1)$ and model (1) is dynamically correctly specified.

2. *Laplace innovation model.* Innovation in model (1) is given by

$$Z_t = 2^{-1/2} L, \qquad (17)$$

where random variable L has a Laplace (or double exponential) distribution. Such a distribution has a symmetric convex density and displays fatter tails than the t_5-distribution. Also, in this experiment, $Z_t \sim$ i.i.d.$(0, 1)$, and model (1) is dynamically correctly specified.

3. *Replace innovative model.* In this model, $\mathcal{Y} := \{Y_t\}_{t \in \mathbb{Z}}$ is generated as follows:

$$Y_t = \begin{cases} \rho_0 + \rho_1 Y_{t-1} + \varepsilon_t, & \text{with probability } 1 - \kappa, \\ \check{Y}_t, & \text{with probability } \kappa, \end{cases} \qquad (18)$$

where $\check{Y}_t \sim N(0, \varrho^2)$, $\varepsilon_t \sim N(0, \sigma_t^2)$ and σ_t^2 is given by Equation (9). At time t, there is a probability κ that observation Y_t is not generated by the GARCH dynamic. The possible "shock," $\check{Y}_t$, will affect future realizations of the process mainly by "inflating" the conditional variance on subsequent days. In this experiment, model (1) is "slightly" misspecified as the dynamic Equations (8) and (9) are not satisfied for every t. We set $\kappa = 0.2\%$ and $\varrho = 10$. The probability of contamination, κ, is very low and implies (on average) four contaminated observations out of 2000 observations. The choice for ϱ allows us to compare the accuracy of the different VaR estimators under very infrequent, but dramatic, (symmetric) shocks. Such shocks could occur over short time periods in real data, as, for instance, in daily equity returns.

We set the AR(1), asymmetric GARCH(1,1) model parameters to $\rho_0 = \rho_1 = 0.01$, $\alpha_0 = 0.03$, $\alpha_1 = 0.02$, $\alpha_2 = 0.8$, and $\alpha_3 = 0.2$. This parameter choice reflects somehow parameter estimates typically obtained for daily percentage index or exchange rate returns; see, for instance, Bollerslev et al. (1994). At the reference model $\mathbb{P}_{\theta_0}$, annualized volatility of Y_t is about 12%. The robust GARCH estimators have tuning constants $c = 11$. The robust GPD estimator has $c_{\mathrm{gpd}} = 8$. The sample size $T = 2000$. Each model is simulated 1000 times. For each simulated sample path, we use the VaR prediction methods (fhs, fhs rob, evt, and evt rob) to compute VaR forecasts. In the financial industry, virtually only out-of-sample VaR forecasts are required, and in-sample measurements of VaR are far less important. In our simulations and empirical applications, all VaR forecasts are out-of-sample ones.

2.1 GARCH Dynamics Estimation

Bias and MSE of PML and robust estimators for the AR(1), asymmetric GARCH(1,1) model (8)–(9) are reported in Mancini and Trojani (2010). Estimation results for the robust estimator $\bar{a}_{\mathrm{trunc}}$ (with $l = 30$ lags) discussed in Footnote 3 are also reported in the appendix. Under reference model $\mathbb{P}_{\theta_0}$, i.e., when GARCH residuals are Gaussian, PMLE (that is indeed MLE) is only slightly more efficient than the robust estimators, $\bar{a}$ and $\bar{a}_{\mathrm{trunc}}$. In all other experiments, both

robust estimators always outperform classical PML estimator in terms of MSEs, especially under the replace innovative model. The overall performances of the two robust estimators $\bar{a}$ and $\bar{a}_{\text{trunc}}$ are very close but $\bar{a}$ has somewhat lower MSEs. The last finding supports the application of $\bar{a}$ for estimating GARCH-type models.

2.2 VaR Violation

Standard analysis of VaR prediction methods is based on violation tests. In the ith simulation, a violation occurs when the actual loss is larger than the predicted VaR, that is, $I(i) := \mathbf{1}\{y_{T,T+h}(i) < \hat{y}^{\alpha}_{T,T+h}(i)\} = 1$ and 0 otherwise. Under the null hypothesis that VaR is correctly estimated, the test statistic $\sum_{i=1}^{1000} I(i)$ is binomially distributed as the 1000 simulations are independently drawn for both horizons $h = 1$ day and 10 days. For $\alpha = 0.05$ and 0.01, the expected number of violations are 50 and 10, and two-side confidence intervals at 95% level are $[37, 64]$ and $[4, 17]$, respectively. Table 2 shows number of violations for fhs, fhs rob, evt, and evt rob. All methods exhibit numbers of violations within such confidence intervals, but it is known that violation tests have typically low power. Table 2 also hints some differences among VaR prediction methods. In the first two Monte Carlo experiments (Student-t_5 and Laplace innovations), only evt rob never exhibits p-values below 0.10, even though estimated GARCH models are correctly specified. These results suggest that evt rob can outperform other approaches even in setups relatively favorable to classical methods, but this phenomenon is not clearly detected by violation tests. Next section studies the precision of VaR forecasts, which is a key issue for measuring market risk.

Table 2 Number of violations under different simulation models. In parentheses two-side p-values for null hypothesis H_0: number of violations equals to the expected number of violations, that is, 50 and 10 for 5% and 1% confidence levels, respectively

	$h = 1$ day ahead VaR forecasts			$h = 10$ days ahead VaR forecasts		
	Student-t_5	Laplace	Replace innovations	Student-t_5	Laplace	Replace innovations
fhs 5%	42 (0.25)	41 (0.19)	45 (0.47)	62 (0.08)	62 (0.08)	44 (0.38)
fhs rob 5%	44 (0.38)	43 (0.31)	47 (0.66)	61 (0.11)	61 (0.11)	43 (0.31)
evt 5%	41 (0.19)	43 (0.31)	47 (0.66)	62 (0.08)	62 (0.08)	44 (0.38)
evt rob 5%	43 (0.31)	44 (0.38)	50 (1.00)	60 (0.15)	61 (0.11)	42 (0.25)
fhs 1%	14 (0.20)	15 (0.11)	10 (1.00)	15 (0.11)	13 (0.34)	8 (0.53)
fhs rob 1%	15 (0.11)	15 (0.11)	7 (0.34)	16 (0.06)	14 (0.20)	5 (0.11)
evt 1%	14 (0.20)	15 (0.11)	8 (0.53)	13 (0.34)	13 (0.34)	8 (0.53)
evt rob 1%	13 (0.34)	14 (0.20)	7 (0.34)	12 (0.53)	13 (0.34)	7 (0.34)

2.3 Accuracy of VaR Prediction

Left panel in Table 3 shows bias and MSE of 1-day ahead VaR predictions. In all Monte Carlo experiments, robust versions of FHS and EVT methods have smaller MSEs than corresponding classical versions. The reduction in MSE is small in the Laplace innovation model but reaches about 80% in the contaminated replace innovative model. In almost all cases, evt rob has the lowest MSE, often by several times. To gauge economic differences among VaR prediction methods, we can compare nominal and effective coverage of predicted VaR. Under the Student-t_5 model, for example, a \$100 value portfolio with $\mu_t = 0.01$ and $\sigma_t = \sqrt{0.375}$ has a daily VaR at 1% of \$2.05. If the true VaR is underestimated by \$0.12, which is approximately one root MSE in the Student-t_5 simulation (i.e., $0.12 \approx \sqrt{0.015}$), the predicted VaR of \$1.93 cannot attain the perfect coverage of 1% but is violated with a sufficiently close probability of 1.2%. Hence, under this model, all VaR prediction methods give economically sensible VaR predictions. Similar conclusions hold for the Laplace model. However, under replace innovative model (18) underestimating the true VaR by \$0.53 or \$0.27, that is, one root MSE of evt and evt rob methods, respectively, implies substantially different situations. In the evt case, predicted VaR at 1% is indeed violated with a probability of 6.8%, while in the evt rob case only with a probability of 2.8%, and this difference is economically sizable. To further understand the magnitude of MSEs, we can standardize them by true

Table 3 Bias and MSE of classical and robust, FHS and EVT VaR prediction methods for $h = 1$ day, 10-day ahead and confidence levels $\alpha = 5\%$, 1% under different simulation models

	$h = 1$ day ahead VaR forecasts				$h = 10$ days ahead VaR forecasts			
	Student-t_5 innovations							
	VaR 5%		VaR 1%		VaR 5%		VaR 1%	
	bias	MSE	bias	MSE	bias	MSE	bias	MSE
fhs	−0.0023	0.0043	−0.0041	0.0165	−0.0044	0.0831	−0.0134	0.3574
fhs rob	−0.0038	0.0038	−0.0075	0.0144	0.0148	0.0831	0.0454	0.3950
evt	−0.0042	0.0040	0.0046	0.0136	−0.0139	0.0835	0.0222	0.3368
evt rob	−0.0067	0.0035	0.0014	0.0119	−0.0123	0.0821	−0.0235	0.2877
	Laplace innovations							
fhs	−0.0036	0.0047	−0.0079	0.0161	−0.0132	0.0809	−0.0265	0.3357
fhs rob	−0.0046	0.0046	−0.0110	0.0159	−0.0005	0.0805	0.0079	0.3463
evt	−0.0088	0.0041	−0.0078	0.0128	−0.0135	0.0812	0.0476	0.3187
evt rob	−0.0137	0.0040	−0.0187	0.0126	−0.0257	0.0827	−0.0347	0.2864
	Replace innovative model							
fhs	0.0053	0.1283	0.0483	0.2481	0.0208	1.4550	0.2790	6.1700
fhs rob	0.0032	0.0241	0.0119	0.0535	0.0928	1.4371	0.5971	14.9516
evt	−0.0189	0.1306	0.1612	0.2786	0.0310	1.4466	−0.1933	3.9456
evt rob	−0.0231	0.0240	0.0818	0.0710	−0.0629	0.4249	−0.6181	1.5884

unconditional variances. In the first two experiments, unconditional daily variance of percentage returns is 0.375. Hence, an MSE of 0.015 for VaR at 1% level amounts to only 4% of the unconditional variance. Under replace innovative model, MSEs of evt and evt rob are 49% and 12% of the unconditional variance (that is equal to 0.574), respectively, suggesting that evt rob provides much more accurate VaR predictions.

Right panel in Table 3 shows the accuracy of VaR predictions at $h = 10$ days ahead horizon. In the first two experiments, the dynamic model (1) is correctly specified and all VaR prediction methods tend to perform similarly in predicting VaR at 5% level, although evt rob outperforms all other methods in predicting VaR at 1% level. In the third experiment, the dynamic model (1) is slightly misspecified and both FHS methods perform very poorly, with fhs rob having the largest MSE for VaR predictions at 1% level. At first sight, the last finding might appear puzzling given the higher accuracy of robust GARCH estimates; see Mancini and Trojani (2010). This result is explained by the low BP of VaR predictions based on nonparametric residual bootstrap.[8] This point is discussed in Section 2.4 below. In terms of MSE, evt rob largely outperforms all other methods. For example, under replace innovative model, the ratio of MSE of VaR forecasts at 1% level over 10-day unconditional variance is 69% for evt and 28% for evt rob, confirming that evt rob provides economically large improvements in VaR predictions.

2.4 Bootstrap Breakdown Point and Quantile Estimates Accuracy

To disentangle the contribution of nonparametric, PML and robust EVT tail estimation to VaR predictions, we repeat the previous Monte Carlo simulation using true GARCH parameters. We also investigate theoretical predictions of Equation (10) on BPs of bootstrap quantiles.

We estimate 5% and 1% quantiles (i.e., VaR) of 10-day ahead return distribution. As GARCH parameters are not estimated, classical and robust FHS methods coincide, and we call them "resampling" in this section. We consider different ways of implementing semiparametric bootstrap methods using EVT. We make an additional distinction depending on whether the quantile of simulated 10-day ahead distribution is estimated nonparametrically or using a GPD (PML or robust) estimator. This distinction highlights the additional contribution of parametric GPD over nonparametric tail estimations in producing accurate VaR forecasts. We compute VaR predictions using the following five methods:

1. Resampling (i.e., FHS method).
2. EVT applied to both daily returns and simulated 10-day ahead returns.

---

[8] To raise the BP of nonparametric bootstrap quantiles, Singh (1998) suggests to winsorize the data before bootstrapping. We winsorized innovations at 0.5% and 1% levels, respectively, and then we computed 10-day ahead VaR predictions using fhs and fhs rob. MSEs of the winsorized VaR predictions did decrease but only by a small amount and the results are not reported.

3. Robust EVT applied to both daily returns and simulated 10-day ahead returns.
4. EVT applied to daily returns, and empirical quantile of 10-day ahead return distributions.
5. Robust EVT applied to daily returns, and empirical quantile of 10-day ahead return distributions.

Table 4 reports simulation results for the five methods under the previous Monte Carlo experiments. All MSEs of VaR forecasts in Table 4 are lower than those in Table 3 as variability deriving from estimation of GARCH parameters is absent now. In the first two experiments (Student-t_5 and Laplace innovations), the data-generating processes do not produce "outliers." For 5% quantiles, resampling procedures and robust EVT perform well, but classical EVT method is the least precise. For 1% quantiles, robust EVT methods have uniformly higher accuracy. Therefore, the misspecification of the GPD tail in these Monte Carlo experiments produces a quite favorable trade-off for using our robust methods in estimating the true tail of GARCH residuals.

Table 4 Bias and MSE of quantile estimates for resampling (first row); EVT applied to daily and 10-day ahead returns (second row); robust EVT applied to daily and 10-day ahead returns (third row); EVT applied to daily returns and empirical quantile of 10-day ahead returns (fourth row); robust EVT applied to daily returns and empirical quantile of 10-day ahead returns (fifth row), under different simulation models

| | Student-t_5 innovations | | | |
| | Quantile 5% | | Quantile 1% | |
	Bias	MSE	Bias	MSE
Resampling	0.0025	0.0566	0.0032	0.2223
Daily evt, 10-day evt	−0.0145	0.0733	0.0193	0.2784
Daily evt rob, 10-day evt rob	0.0135	0.0599	0.0300	0.1913
Daily evt, 10-day emp. quant.	−0.0047	0.0696	−0.0276	0.2706
Daily evt rob, 10-day emp. quant.	0.0096	0.0554	0.0181	0.2209
	Laplace innovations			
Resampling	0.0012	0.0594	−0.0024	0.2007
Daily evt, 10 days evt	−0.0354	0.1143	−0.0015	0.4107
Daily evt rob, 10 days evt rob	0.0043	0.0624	0.0355	0.1889
Daily evt, 10 days emp. quant.	−0.0280	0.1100	−0.0305	0.4169
Daily evt rob, 10 days emp. quant.	−0.0012	0.0603	0.0293	0.2002
	Replace innovative model			
Resampling	−0.0073	0.1283	0.3037	7.3803
Daily evt, 10 days evt	−0.0143	0.1633	−0.2485	1.2285
Daily evt rob, 10 days evt rob	−0.0044	0.1547	−0.3853	0.8931
Daily evt, 10 days emp. quant.	0.0052	0.1588	−0.3086	1.1703
Daily evt rob, 10 days emp. quant.	−0.0167	0.1329	−0.3985	0.9217

Under the replace innovative model, resampling method breaks down in the estimation of 1% quantile, whereas it produces accurate results in estimating 5% quantile. From Table 1, the BP of VaR at 5% level corresponds to 0.51% outliers in the data, whereas the BP of VaR at 1% level is 0.10%. Hence, as predicted by Equation (10), $\kappa = 0.20\%$ of outliers in the data breaks down VaR predictions at 1%, but not at 5% using nonparametric residual bootstrap.[9] For quantile at 1% level, the ratio of MSE over 10-day unconditional variance is 129% for the resampling method (FHS), and only 16% for the robust EVT method applied to both daily and 10-day returns, confirming that robust EVT provides economically large improvements in predicting VaR. Overall, robust EVT tail estimation is particulary important when forecasting VaR at low confidence levels and/or data are contaminated by outliers.

To further understand the impact of PML and robust GARCH estimates on VaR predictions, we repeated the Monte Carlo simulation, but for comparison and to regularize the bootstrap procedure, we always fitted the tails of daily innovation distributions (and 10-day ahead returns) with the classical EVT method. Unreported results confirm that VaR predictions based on robust GARCH estimates are still more accurate than VaR predictions based on PML GARCH estimates especially for the 10-day horizon and in the presence of outlying observations. This finding suggests that robust GARCH estimates contributes significantly to the accuracy of VaR forecasts.

3   REAL DATA ESTIMATION AND BACKTESTING

We backtest VaR prediction methods on four historical series of daily rate of returns: S&P 500 index price from December 1988 to July 2003, Dollar-Yen exchange rate from January 1986 to January 2005, Microsoft share price from March 1986 to January 2005, and Boeing share price from January 1980 to January 2005. The data are downloaded from Datastream. Denote by $y_1, \ldots, y_N$ the historical series of returns, where, for example, $N = 4500$. To backtest, for example, evt rob method we proceed as follows. We use $n = 2000$ returns, that is, about eight years of daily data, to estimate the AR(1), asymmetric GARCH(1,1) model with the robust estimator Equation (5) and tuning constant $c = 8$. Return innovation distribution is estimated using the filtered return innovations, $\hat{z}_1, \ldots, \hat{z}_n$, and the robust EVT approach discussed in Section 1.4, with $c_{\text{gpd}} = 6$ for robust GPD estimator Equation (14).[10] For day $T = n$, out-of-sample VaR forecasts, $\hat{y}_{T,T+h}^{\alpha}$, are computed at horizons $h = 1$ day, 10 days, and confidence levels $\alpha = 1\%$, 5%, using the semiparametric residual bootstrap discussed in Section 1.3. Left tail of simulated 10-day

---

[9]This finding also appears in the right panel of Table 3.

[10]Considering the "noisier" nature of real data, as opposed to simulated data, and the different characteristics of financial time series (indexes, stocks, and exchange rates) used in backtesting, we take a somewhat more conservative viewpoint setting the robustness tuning constants c and c_{gpd} to lower levels than in the Monte Carlo study.

ahead return distribution is fitted using robust GPD estimator to calculate VaRs. The VaR prediction is calculated for each day $T \in \mathcal{T} = \{n, n + 1, \ldots, N - h\}$ using a moving time window of n historical returns for filtering returns and estimating distributions. GARCH estimates, however, are updated only every 500 days. The other three VaR prediction methods are similarly backtested: fhs and fhs rob rely on nonparametric rather than semiparametric residual bootstrap; evt uses PML rather than robust GARCH estimation.

For comparison, we also include (i) Historical Simulation HS; (ii) RiskMetrics (1995) (RM); (iii) Engle and Gonzalez-Rivera (1991) semiparametric GARCH model (EGR); (iv) Engle and Manganelli (2004) CAViaR model;[11] and for 10-day ahead VaR predictions, (v) GARCH model applied directly to 10-day asset returns. The HS and RM methods are popular in financial industry. The EGR approach provides flexible and efficient estimates of semiparametric GARCH models. The CAViaR model offers a challenging benchmark for VaR predictions, which estimates VaR directly using quantile regressions.[12]

3.1 Data and GARCH Estimation

Table 5 shows summary statistics for the daily rate of returns. The different characteristics of assets make the backtesting exercise particularly interesting. For example, Dollar-Yen exchange rates have large skewness and Microsoft returns large kurtosis. PML and robust estimates of AR(1), asymmetric GARCH(1,1) models for the different financial assets are collected in Mancini and Trojani (2010). In several occasions and especially for the volatility parameters, the two estimates are rather different. Next sections show how these estimates induce different VaR forecasts.

3.2 Backtesting VaR Prediction

To assess the forecasting performance of the VaR prediction methods, we adopt the testing framework proposed by Christoffersen (1998). This framework consists of three tests and has become a standard setting for evaluating out-of-sample forecasts. We refer the reader to Christoffersen (2003, Chapter 8) for an in-depth description of the tests; a short description is also available in Mancini and Trojani (2010). The test of unconditional coverage checks whether or not the overall number of violations is statistically acceptable. The test of independence aims at

---

[11] Engle and Manganelli (2004) find that empirically Asymmetric Slope and Indirect GARCH CAViaR models tend to outperform other CAViaR specifications. Given the asymmetric impact of positive and negative returns on volatility (and possibly on quantiles) documented in our sample, we use the asymmetric slope CAViaR model in our empirical analysis. The Matlab code for the CAViaR model is freely available at Simone Manganelli's webpage, http://www.simonemanganelli.org.

[12] Koenker and Bassett (1978) introduce quantile regression methods; see also Foresi and Peracchi (1995) and Peracchi (2002). From a robustness perspective, drawbacks of quantile regression is its behavior under heteroscedasticity and the nonrobustness to leverage points; see Koenker and Bassett (1982).

Table 5 Summary statistics for the daily rate of returns in percentage: S&P 500 index price from 12/88 to 7/03; Dollar-Yen exchange rate from 1/86 to 1/05; Microsoft share price from 3/86 to 1/05; and Boeing share price from 1/80 to 1/05

	Sample size	Mean	Std.	Skew.	Kurt.	Min	Max
S&P 500	3799	0.034	1.029	−0.163	7.137	−7.113	5.573
Dollar-Yen	4969	0.016	0.688	0.496	7.596	−3.505	6.795
Microsoft	4918	0.146	2.565	−0.228	12.223	−31.111	19.552
Boeing	6535	0.055	1.938	−0.052	8.388	−17.625	15.347

verifying possible clusterings of violations over time. The test of conditional coverage checks in which respect the time series of VaR violations does not satisfy the correct conditional coverage.

Tables 6, 7, and 8 show number of violations and p-values of unconditional, independence and conditional coverage tests for 1-day ahead VaR forecasts.[13] Only evt rob passes all violation tests with p-values above 0.10. Nearly all other methods fail both unconditional and conditional coverage tests for the S&P 500 backtesting. For instance, in the conditional coverage test and VaR predictions at 5%, fhs has a p-value of 0.046, fhs rob of 0.076, EGR of 0.035, evt and CAViaR of 0.012, and HS below 0.001. Generally, HS and RM methods do not work well, especially for VaR predictions at 1% level with several p-values below 0.05. VaR prediction method based on semiparametric EGR model performs similarly to fhs method. The CAViaR model fails violation tests for the S&P 500 backtesting with most p-values below 0.05.

These empirical findings confirm the simulation results and document the accuracy of VaR predictions based on our robust approach. We now turn the attention to the time-series properties of VaR forecasts. Temporal profiles of VaR predictions have economic relevance because asset allocations need to satisfy VaR constraints, and VaRs determine reserve amounts to cover market risk. Asset allocations and reserve amounts cannot change heavily from one day to the next, otherwise the financial firm can incur in a variety of costs, such as transaction costs or financial losses due to liquidation of risky assets at stressed prices in high-volatile periods to reduce risk exposures. Left panel in Table 10 summarizes the time-series properties of VaR forecasts for Dollar-Yen backtesting. The corresponding statistics for S&P 500, Microsoft, and Boeing are similar and collected in Mancini and Trojani (2010). In Table 10, "VaR" denotes average VaR forecasts, Δ average daily changes in VaR predictions $\{\hat{y}^{\alpha}_{T+1,T+1+h} - \hat{y}^{\alpha}_{T,T+h}\}_{T \in \mathcal{T}}$, Δ^2 corresponding empirical second moment, and $|\Delta|\%$ average absolute relative changes in percentage. The last three statistics describe daily changes of VaR forecasts. In nearly all backtested time series and VaR confidence levels, evt rob has the lowest

---

[13]See also Kuester, Mittnik, and Paolella (2006) for a recent comparison of FHS methods.

Table 6 Number of violations of classical and robust, FHS and EVT, and HS, RM, EGR, CAViaR, $h = 1$ day ahead VaR predictions

	fhs 5%	fhs rob 5%	evt 5%	evt rob 5%	Expected 5%	fhs 1%	fhs rob 1%	evt 1%	evt rob 1%	Expected 1%
S&P 500	108	106	109	104	90	19	21	18	19	18
Dollar-Yen	147	149	152	153	149	32	33	35	33	30
Microsoft	155	153	156	152	146	39	38	39	38	29
Boeing	211	234	217	244	227	47	49	41	46	45

	HS 5%	RM 5%	EGR 5%	CAViaR 5%	Expected 5%	HS 1%	RM 1%	EGR 1%	CAViaR 1%	Expected 1%
S&P 500	186	96	109	109	90	49	34	19	28	18
Dollar-Yen	149	151	149	148	149	31	44	32	35	30
Microsoft	153	117	155	142	146	40	37	39	37	29
Boeing	265	206	211	237	227	66	61	47	47	45

Table 7 Likelihood ratio p-values for violations of classical and robust, FHS and EVT, $h = 1$ day ahead VaR predictions

	Unconditional coverage							
	fhs 5%	fhs rob 5%	evt 5%	evt rob 5%	fhs 1%	fhs rob 1%	evt 1%	evt rob 1%
S&P 500	0.059	0.092	0.046	0.139	0.814	0.489	1.000	0.814
Dollar-Yen	0.899	0.966	0.769	0.706	0.675	0.550	0.342	0.550
Microsoft	0.447	0.552	0.398	0.610	0.083	0.117	0.083	0.117
Boeing	0.276	0.305	0.501	0.120	0.808	0.592	0.508	0.838
	Independence							
S&P 500	0.108	0.129	0.028	0.153	0.524	0.481	0.546	0.524
Dollar-Yen	0.781	0.569	0.420	0.681	0.404	0.389	0.433	0.389
Microsoft	0.781	0.993	0.900	0.726	0.550	0.522	0.550	0.522
Boeing	0.531	0.059	0.420	0.148	0.515	0.560	0.388	0.450
	Conditional coverage							
S&P 500	0.046	0.076	0.012	0.121	0.794	0.614	0.834	0.794
Dollar-Yen	0.955	0.849	0.692	0.856	0.646	0.577	0.468	0.577
Microsoft	0.720	0.838	0.695	0.826	0.185	0.239	0.185	0.239
Boeing	0.454	0.099	0.576	0.105	0.785	0.731	0.553	0.736

values of Δ^2 and $|\Delta|\%$, implying smoothest VaR profiles over time. For instance, for VaR at 5% level in the Dollar-Yen backtesting, $|\Delta|\%$ for evt is 13% larger than those for evt rob. Using robust VaR predictions, the financial firm can adjust portfolio risk exposures to VaR limits more smoothly and thus more efficiently.

The empirical analysis of 10-day ahead VaR forecasts confirms and further strengthens the previous findings. Table 9 shows number of violations of 10-day ahead VaR forecasts and robust Newey and West (1987) two-side p-values for the null hypothesis that the given method predicts VaR correctly.[14] The lowest p-value for evt rob is 0.28. All other FHS methods are too conservative in predicting VaR at 5% level for the Boeing backtesting, with p-values below 0.07. VaR prediction method denoted by h-ret applies fhs method directly to nonoverlapping 10-day returns. Hence h-ret avoids the resampling procedure to simulate daily returns up to 10-day horizon and relies on the empirical quantile of estimated 10-day return innovations to predict VaR. CAViaR model is fitted to nonoverlapping 10-day returns as well. Both methods do not work well. They suffer the inefficient use of available information, discarding nine out of ten observations when computing nonoverlapping 10-day returns.[15] RiskMetrics uses the suggested

---

[14]Robust standard errors are computed using Newey–West covariance matrix with $h - 1$ lags.

[15]In principle, EVT methods could be applied directly to 10-day returns as well, but then the issue of limited sample size would be even more sever. To achieve 200 data points as in our previous applications, 80 years of daily returns would be required.

Table 8 Likelihood ratio p-values for violations of HS, RM, EGR, and CAViaR, $h = 1$ day ahead VaR predictions

	Unconditional coverage							
	HS 5%	RM 5%	EGR 5%	CAViaR 5%	HS 1%	RM 1%	EGR 1%	CAViaR 1%
S&P 500	0.000	0.521	0.046	0.046	0.000	0.001	0.814	0.028
Dollar-Yen	0.966	0.834	0.966	0.966	0.812	0.014	0.675	0.342
Microsoft	0.552	0.011	0.447	0.736	0.057	0.163	0.083	0.163
Boeing	0.011	0.150	0.276	0.490	0.004	0.027	0.808	0.808
	Independence							
S&P 500	0.239	0.954	0.098	0.028	0.580	0.674	0.524	0.347
Dollar-Yen	0.023	0.904	0.842	0.542	0.419	0.685	0.404	0.361
Microsoft	0.164	0.545	0.781	0.423	0.018	0.090	0.550	0.330
Boeing	0.017	0.583	0.531	0.855	0.089	0.263	0.515	0.099
	Conditional coverage							
S&P 500	0.000	0.812	0.035	0.012	0.000	0.003	0.794	0.058
Dollar-Yen	0.076	0.971	0.980	0.830	0.701	0.045	0.646	0.419
Microsoft	0.318	0.033	0.720	0.685	0.010	0.090	0.185	0.235
Boeing	0.002	0.306	0.454	0.775	0.004	0.046	0.785	0.249

$\sqrt{h}$-rule to scale daily volatility to 10-day horizon but fails Microsoft backtesting with a p-value of 0.03. EGR fails Boeing backtesting with a p-value of 0.06.

In nearly all backtested time series, evt rob VaR forecasts are the most stable over time in terms of squared and absolute relative changes, Δ^2 and $|\Delta|\%$. For example, for VaR at 5% level, Δ^2 for evt is 9% higher than that for evt rob in the Dollar-Yen backtesting; see right panel in Table 10. To save space descriptive statistics of HS, RM, EGR, CAViaR, and h-ret temporal VaR profiles are not reported but collected in Mancini and Trojani (2010). The h-ret and CAViaR methods have the most volatile temporal VaR profiles among all considered VaR prediction methods. EGR performs similarly as fhs method.

3.3 Tail Estimation Risk

An important source of variability of VaR forecasts can be reestimation of tail distributions that occur every day and induces the so-called tail estimation risk. Our Monte Carlo experiments suggest that under a number of realistic dynamic specifications, the VaR estimation risk implied by our robust procedure is lower than the one of classical approaches. To measure such estimation risk exactly in real-data applications, we would need to compare predicted VaR and true VaR, but the latter is obviously unknown. A feasible approach to quantify estimation risk is to provide empirical prediction intervals for the VaR forecast itself. When a VaR prediction method is unbiased and behaves properly in terms of VaR

Table 9 Number of violations of classical and robust, FHS and EVT, and HS, RM, EGR, CAViaR, and h-ret, $h = 10$ days ahead VaR predictions. In parentheses Newey–West robust two-side p-values for null hypothesis H_0: number of violations equals to the expected number of violations (Expt.)

	fhs 5%	fhs rob 5%	evt 5%	evt rob 5%	Expt. 5%	fhs 1%	fhs rob 1%	evt 1%	evt rob 1%	Expt. 1%
S&P 500	75 (0.42)	75 (0.42)	76 (0.46)	75 (0.42)	90	14 (0.49)	14 (0.51)	13 (0.38)	16 (0.75)	18
Dollar-Yen	135 (0.64)	147 (0.97)	148 (1.00)	157 (0.76)	148	25 (0.67)	26 (0.74)	20 (0.38)	24 (0.61)	30
Microsoft	151 (0.84)	150 (0.87)	152 (0.82)	155 (0.74)	146	27 (0.84)	25 (0.69)	29 (0.99)	30 (0.93)	29
Boeing	164 (0.06)	161 (0.05)	166 (0.07)	190 (0.28)	226	57 (0.59)	58 (0.57)	59 (0.53)	63 (0.43)	45

	HS 5%	RM 5%	EGR 5%	CAViaR 5%	h-ret 5%	HS 1%	RM 1%	EGR 1%	CAViaR 1%	h-ret 1%
S&P 500	209 (0.00)	81 (0.68)	80 (0.62)	116 (0.20)	136 (0.06)	56 (0.02)	29 (0.32)	14 (0.49)	44 (0.02)	39 (0.06)
Dollar-Yen	161 (0.65)	134 (0.64)	132 (0.58)	207 (0.09)	166 (0.54)	35 (0.69)	30 (0.98)	24 (0.60)	42 (0.29)	35 (0.66)
Microsoft	178 (0.34)	97 (0.03)	148 (0.93)	177 (0.30)	166 (0.53)	54 (0.15)	20 (0.32)	27 (0.84)	67 (0.01)	54 (0.15)
Boeing	277 (0.22)	206 (0.57)	164 (0.06)	207 (0.59)	223 (0.92)	68 (0.30)	76 (0.19)	60 (0.51)	90 (0.07)	67 (0.31)

Table 10 Averages of VaR forecasts and daily changes in VaR forecasts, Δ, for the Dollar-Yen exchange rate from 1/86 to 1/05. $|\Delta|\%$ are Δ in absolute values divided by VaR forecasts and in percentage

	$h = 1$ day ahead VaR forecasts				$h = 10$ days ahead VaR forecasts			
	VaR	$\Delta \times 10^4$	$\Delta^2 \times 100$	$|\Delta|\%$	VaR	$\Delta \times 10^4$	$\Delta^2 \times 100$	$|\Delta|\%$
fhs 5%	1.037	−0.556	0.463	3.518	3.364	−0.861	3.979	3.291
fhs rob 5%	1.032	−0.694	0.431	3.144	3.300	−0.835	3.595	3.106
evt 5%	1.024	−0.498	0.450	3.541	3.305	−1.088	3.825	3.186
evt rob 5%	1.021	−0.554	0.413	3.128	3.262	−1.163	3.306	2.935
fhs 1%	1.653	−0.953	1.103	3.302	4.992	−2.263	9.882	3.616
fhs rob 1%	1.650	−1.189	1.075	3.065	4.880	−2.516	9.023	3.458
evt 1%	1.637	−1.351	1.111	3.320	5.012	−2.646	9.379	3.448
evt rob 1%	1.642	−1.390	1.082	3.059	4.960	−3.271	8.146	3.204

violations, the narrower the prediction interval, the lower the tail estimation risk. Christoffersen and Gonçalves (2005) propose a resampling technique to measure estimation risk, and we follow their methodology here. As any other resampling techniques, the procedure is computationally demanding. To keep computations feasible, we limit the analysis to $h = 10$ days ahead VaR predictions based on fhs and evt rob. For both methods and for each day $T \in \mathcal{T}$, we obtain $S = 199$ VaR predictions $\{\hat{y}_{T,T+h}^{\alpha\,(s)}\}_{s=1}^{S}$ at confidence levels $\alpha = 5\%$ and 1%, that is, each day we repeat S times the forecasting procedures of classical FHS and robust EVT methods. The robust EVT method is particularly demanding because on each day and for each one of the 199 random samples, GPD distributions are fitted using the robust estimator to both tails of the GARCH residual distribution and to the left tail of simulated 10-day ahead return distribution. For each day $T \in \mathcal{T}$, we compute the prediction interval at 80% confidence level for the VaR forecast $\hat{y}_{T,T+h}^{\alpha}$,

$$[Q_{0.1}(\{\hat{y}_{T,T+h}^{\alpha\,(s)}\}_{s=1}^{S}),\ Q_{0.9}(\{\hat{y}_{T,T+h}^{\alpha\,(s)}\}_{s=1}^{S})],$$

where $Q_x(\cdot)$ is the x-quantile of the empirical distribution of $\{\hat{y}_{T,T+h}^{\alpha\,(s)}\}_{s=1}^{S}$. Other confidence levels are certainly conceivable but the results based on the 80% level are likely to be representative of the findings based on other confidence levels.

Table 11 shows average absolute and relative widths of 10-day ahead VaR prediction intervals in our backtesting period. For all backtested assets, evt rob has narrower prediction intervals than fhs, both in absolute and in relative terms. Therefore, evt rob provides more accurate and reliable VaR predictions than fhs. For example, in S&P 500 and Boeing backtesting and for VaR forecasts at 1% level, classical FHS relative prediction intervals are 14% and 21% larger than those for robust EVT. In Table 11, var% denotes variance of daily changes in prediction intervals. In all but one case, daily changes of evt rob have smaller variances

Table 11 Prediction intervals that account for tail estimation risk in 10-day ahead VaR forecasts. For each day T in the backtesting period, we obtain $S = 199$ 10-day ahead VaR forecasts using a resampling technique applied to classical FHS and robust EVT methods. Then we compute prediction intervals at 80% confidence level for 10-day ahead VaR forecasts. Absolute is the average width of the 80% prediction intervals (90% minus 10% percentile) of VaR forecasts; Relative% is the average width of the 80% prediction intervals as a percentage of the VaR forecasts; var% is the corresponding variance

	Prediction interval VaR 5%			Prediction interval VaR 1%		
	Absolute	Relative%	VaR%	Absolute	Relative%	VaR%
	S&P 500					
fhs	0.283	4.474	0.143	0.692	6.550	0.449
evt rob	0.257	4.115	0.097	0.589	5.727	0.278
	Dollar-Yen					
fhs	0.121	3.548	0.076	0.250	4.909	0.167
evt rob	0.112	3.344	0.064	0.240	4.708	0.170
	Microsoft					
fhs	0.393	4.056	0.101	0.807	5.387	0.202
evt rob	0.359	3.766	0.079	0.753	4.938	0.195
	Boeing					
fhs	0.346	3.932	0.103	0.832	5.969	0.581
evt rob	0.308	3.550	0.082	0.668	4.934	0.315

than daily changes of fhs prediction intervals. In S&P 500 and Boeing backtesting and VaR forecasts at 1% level, such variances for evt rob are nearly 50% those of fhs. Overall, our robust procedure appears to control tail estimation risk in a better way than classical procedures and this induces more stable VaR profiles over time.

4 Conclusion

We propose a general robust semiparametric bootstrap method to estimate predictive distributions of GARCH-type models. Our approach is based on a robust estimation of parametric GARCH-type models and a robustified resampling method for GARCH residuals, which controls the bootstrap instability due to influential observations. In the latter, a robust extreme value estimator is used to fit innovation tail distributions above some threshold levels. A Monte Carlo study shows that the robust extreme value estimator provides more accurate quantile estimates and is less sensitive to the choice of threshold levels than classical estimators.

Our robust procedure offers improvements in accuracy of VaR predictions, especially for several days ahead horizons and/or in the presence of outlying observations. In nearly all Monte Carlo experiments, our robust procedure has lower mean square prediction errors than classical methods, often by a large extent. Only our method passes all validation tests at usual significance levels. Theoretical predictions of bootstrap breakdown points are confirmed by simulations and nonrobust bootstrap procedures break down approximately at the calculated BP.

The simulation evidence is confirmed by the real data application. We backtest several VaR prediction methods using about 20 years of S&P 500, Dollar-Yen, Microsoft, and Boeing daily returns. Only our robust procedure passes all validation tests at usual significance levels and outperforms several other VaR prediction methods, such as RiskMetrics, CAViaR, HS, and classical FHS methods. Overall, robust VaR profiles are more accurate and more stable over time than classical forecasts. Given the accuracy of our robust method, the stability of VaR profiles is a desirable feature because it allows financial firms to adapt risky positions to VaR limits more smoothly and thus more efficiently. We show empirically that our robust procedure controls for tail estimation risk better than classical methods and this induces more accurate and more stable over time VaR prediction intervals.

Robust semiparametric bootstrap methods have applications beyond risk management. For example, Engle and Gonzalez-Rivera (1991) propose an iterative procedure to estimate semiparametric GARCH models computing the likelihood function using nonparametric estimation of innovation distribution. Applying our robust procedure extends the estimation of such GARCH models to the robust setting. Another application of our method can be computing fund performance measures using the robust bootstrap procedure.

REFERENCES

Aït-Sahalia, Y., and A. Lo. 2000. Nonparametric Risk Management and Implied Risk Aversion. *Journal of Econometrics* 94: 9–51.

Artzner, P., F. Delbaen, J. Eber, and D. Heath. 1999. Coherent Measures of Risk. *Mathematical Finance* 9: 203–228.

Balkema, A., and L. de Haan. 1974. Residual Life Time at Great Age. *Annals of Probability* 2: 792–804.

Barone-Adesi, G., K. Giannopoulos, and L. Vosper. 1999. VaR without Correlations for Nonlinear Portfolios. *Journal of Futures Markets* 19: 583–602.

Basel Committee 1996. *Overview of the Amendment to the Capital Accord to Incorporate Market Risks*. Basel Committee on Banking Supervision.

Bekaert, G., and G. Wu. 2000. Asymmetric Volatility and Risk in Equity Markets. *Review of Financial Studies* 13: 1–42.

Black, F. 1976. "Studies of Stock Market Volatility Changes." In *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistic Section*, pp. 177–181.

Bollerslev, T. P., and J. M. Wooldridge. 1992. Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances. *Econometric Reviews* 11: 143–172.

Bollerslev, T. P., R. F. Engle, and D. B. Nelson. 1994. "ARCH models," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. McFadden. Vol. IV. North-Holland, Amsterdam, pp. 2959–3038.

Campbell, J. Y., and L. Hentschel. 1992. No News Is Good News: An Asymmetric Model of Changing Volatility in Stock Returns. *Journal of Financial Economics* 31: 281–318.

Camponovo, L., O. Scaillet, and F. Trojani. 2009. "Robust Subsampling." Working paper, University of Lugano.

———. 2010. "Robust Resampling Methods for Time Series." Working paper, University of Lugano.

Christie, A. 1982. The Stochastic Behavior of Common Stock Variances: Value, Leverage and Interest Rate Effects. *Journal of Financial Economics* 10: 407–432.

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

———. 2003. *Elements of Financial Risk Management*. San Diego, CA: Academic Press.

Christoffersen, P. F., and S. Gonçalves. 2005. Estimation Risk in Financial Risk Management. *Journal of Risk* 1–29.

Cowell, F. A., and M.-P. Victoria-Feser. 1996. Robustness Properties of Inequality Measures. *Econometrica* 64: 77–101.

Davidson, R., and E. Flachaire. 2007. Asymptotic and Bootstrap Inference for Inequality and Poverty Measures. *Journal of Econometrics* 141: 141–166.

Diebold, F. X., T. Schuermann, and J. Stroughair. 1998. "Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management." In A.-P. N. Refenes, J. D. Moody and A. N. Burgess (eds.), *Advances in Computational Finance*. Amsterdam, The Netherlands: Kluwer Academic Publishers, pp. 3–12.

Duffie, D., and J. Pan. 1997. An Overview of Value at Risk. *Journal of Derivatives* 4: 7–49.

Dupuis, D. J. 1999. Exceedances Over High Thresholds: A Guide to Threshold Selection. *Extremes* 1: 251–261.

Embrechts, P., C. Klüppelberg, and T. Mikosch. 1997. *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer.

Engle, R. F., and G. Gonzalez-Rivera. 1991. Semiparametric ARCH Models. *Journal of Business and Economic Statistics* 9: 345–359.

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

Engle, R. F., and V. Ng. 1993. Measuring and Testing the Impact of News on Volatility. *Journal of Finance* 48: 1749–1778.

Foresi, S., and F. Peracchi. 1995. The Conditional Distribution of Excess Returns: An Empirical Analysis. *Journal of the American Statistical Association* 90:451–466.

Gagliardini, P., F. Trojani, and G. Urga. 2005. Robust GMM Tests for Structural Breaks. *Journal of Econometrics* 129: 139–182.

Ghysels, E., A. C. Harvey, and E. Renault. 1996. "Stochastic Volatility." In G. S. Maddala and C. R. Rao (eds.), *Handbook of Statistics*. Amsterdam, The Netherlands: North-Holland, pp. 119–191.

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. *Journal of Finance* 48: 1779–1801.

Gonzalo, J., and J. Olmo. 2004. Which Extreme Values are Really Extreme?. *Journal of Financial Econometrics* 2: 349–369.

Gouriéroux, C., A. Monfort, and A. Trognon. 1984. Pseudo Maximum Likelihood Methods: Theory. *Econometrica* 52: 681–700.

Gouriéroux, C., J. P. Laurent, and O. Scaillet. 2000. Sensitivity Analysis of Values at Risk. *Journal of Empirical Finance* 7: 225–245.

Hill, B. 1975. A Simple General Approach to Inference about the Tail of a Distribution. *Annals of Statistics* 3: 1163–1174.

Hull, J., and A. White. 1998. Incorporating Volatility Updating into the Historical Simulation Method for VaR. *Journal of Risk* 1: 5–19.

Juárez, S. F., and W. R. Schucany. 2004. Robust and Efficient Estimation for the Generalized Pareto Distribution. *Extremes* 7: 237–251.

Koenker, R. W., and G. Bassett. 1978. Regression Quantiles. *Econometrica* 46: 33–50.

——. 1982. Robust Tests for Heteroskedasticity Based on Regression Quantiles. *Econometrica* 50: 43–61.

Kuester, K., S. Mittnik, and M. S. Paolella. 2006. Value-at-Risk Prediction: A Comparison of Alternative Strategies. *Journal of Financial Econometrics* 4: 53–89.

Mancini, L., E. Ronchetti, and F. Trojani. 2005. Optimal Conditionally Unbiased Bounded-Influence Inference in Dynamic Location and Scale Models. *Journal of the American Statistical Association* 100: 628–641.

Mancini, L., and F. Trojani. 2010. "Robust Value at Risk Prediction: Appendix." Technical report. Available at http://ssrn.com/abstract=1674761.

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

Newey, W. K., and K. D. West. 1987. A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55: 703–708.

Ortelli, C., and F. Trojani. 2005. Robust Efficient Method of Moments. *Journal of Econometrics* 128: 69–97.

Peracchi, F. 2002. On Estimating Conditional Quantiles and Distribution Functions. *Computational Statistics and Data Analysis* 38: 433–447.

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

Pritsker, M. 1997. Towards Assessing the Magnitude of Value at Risk Errors Due to Errors in the Correlation Matrix. *Financial Engineering News* 2: 14–16.

———. 2001. *"The Hidden Dangers of Historical Simulation." Working paper, Finance and Economics discussion series 2001–27*. Washington, DC: Board of Governors of the Federal Reserve System.

RiskMetrics. 1995. *RiskMetrics Technical Document*, 3rd edn. J.P. Morgan.

Ronchetti, E., and F. Trojani. 2001. Robust Inference with GMM Estimators. *Journal of Econometrics* 101: 37–69.

Rosenberg, J. V., and R. F. Engle. 2002. Empirical Pricing Kernels. *Journal of Financial Economics* 64: 341–372.

Sakata, S., and H. White. 1998. High Breakdown Point Conditional Dispersion Estimation with Application to S&P 500 Daily Returns Volatility. *Econometrica* 66: 529–567.

Singh, K. 1998. Breakdown Theory for Bootstrap Quantiles. *Annals of Statistics* 26: 1719–1732.

White, H. 1982. Maximum Likelihood Estimation of Misspecified Models. *Econometrica* 50: 1–25.