Controlling Tail Risk Measures with Estimation Error^{*}

Tetsuya Kaji¹ and Hyungjune Kang²

¹University of Chicago ²Two Sigma Investments

June 12, 2025

Abstract

Controlling tail risk is a key ingredient in modern risk management and financial regulation. Tail risk, such as Value-at-Risk and expected shortfall, often aims to control the probability of an undesirable event. However, estimation error inherent in risk management may invalidate the guarantees put forward by the risk measures. This paper provides a general way to incorporate estimation risk into tail risk management problems. We start with the assumption that a valid confidence interval is available for the risk measure, and provide an *observable* risk estimate with which the true *unobservable* risk probability is bounded. We give a recommended choice of the tuning parameter that makes the risk estimate consistent. An empirical application to Value-at-Risk and expected shortfall in a stylized investment problem demonstrates that the proposed risk estimates with the guarantees are only modestly larger than the naive estimates without the guarantees.

Keywords: tail risk, risk measure, estimation risk, value-at-risk, expected shortfall, banking regulation

JEL Codes: D81, G28, G32, C58.

^{*}We thank Anna Mikusheva, Isaiah Andrews, Victor Chernozhukov, Deborah Lucas, and Alex Belloni for helpful comments, and August Shen for research assistance. This work is supported by the Richard N. Rosett Faculty Fellowship and the Liew Family Faculty Fellowship at the University of Chicago Booth School of Business. This paper is based on the authors' dissertation chapters.

1 Introduction

Controlling tail risk is a key ingredient in modern risk management and financial regulation. As such, there has been a substantial amount of research on how to measure the tail risk, how to assess the quality of the measures, and how to use the measures in various contexts. For example, the axiomatic foundations and theoretical properties of (tail) risk measures such as Value-at-Risk (VaR), expected shortfall (ES), and their generalizations are studied by Cont et al. (2013), Kou et al. (2013), Kou and Peng (2016), Wang and Zitikis (2021), and Liu and Wang (2021), and reviewed in He et al. (2022). The backtest procedures for assessment and comparison of risk measures are considered by Nolde and Ziegel (2017), Du and Escanciano (2017), and Du et al. (2024). A capital adequacy test for banking regulation based on VaR are studied by He and Peng (2018).

In practice, true risk measures are not observed, and we need to estimate them to be used in practice. McNeil et al. (2005, pp. 40–41) warned the danger of a naive use of estimated risk, pointing out that it is subject to the problems of estimation error, model risk, and market liquidity. The problem of estimation error is that the sampling error of the estimated risk may distort the desired characteristics of the risk measure. The problem of model risk arises when the postulated model is largely misspecified and its estimates are misleading or useless. The problem of market liquidity comes from the fact that liquidation incurs transaction costs and an asset changes its price in the process of liquidation; thus, the asset worth \$1 may not convert to \$1 in hand.

The literature developed various methods to account for these issues. For model risk, Stahl (1997) argued that multiplying 3 to VaR hedges against the model uncertainty using Chebyshev's inequality;¹ Leippold and Vanini (2002) showed that for ES, the multiplier for model risk can be lowered to 1.5; see also Hendricks and Hirtle (1997), Lopez (1998), and Novak (2010). Blanchet and Murthy (2019) provided a general method to bound the risk measures to account for model risk using optimal transport. Zhu and Fukushima (2009) and Li (2018) discussed derivation and the use of worst-case risk measures when we only have partial information for the loss distribution. For the liquidity problem, many institutions use "liquidity-adjusted" VaR that takes into account realistic holding periods; Gârleanu and Pedersen (2007) analyzed the feedback effect of such practice; Adrian and Brunnermeier (2016) pro-

¹Stahl's (1997) justification was retrospective, after Basel II employed the multiplier of 3.

posed a measure of systemic risk that subsumes liquidity risk; Elsinger et al. (2006) and Capponi and Rubtsov (2022) analyzed models to capture systemic tail risks in respective contexts. Embrechts et al. (2015) introduced a notion of robustness to deal with model uncertainty in risk aggregation. An optimal liquidation strategy is studied by Min et al. (2022).

For estimation risk, various portfolio optimization problems are known to be vulnerable to estimation error (Michaud, 1989; Chopra and Ziemba, 1993; Lima et al., 2011), and methods to integrate estimation error into the portfolio optimization problems are proposed (Ceria and Stubbs, 2006; Michaud and Michaud, 2008). Du and Escanciano (2017) and Nolde and Ziegel (2017) developed backtesting procedures to assess and compare the accuracy of different risk estimates. Cont et al. (2010) considered a "risk measurement procedure" that encompasses both the measurement of risk and the choice of risk measure, and studied the robustness and sensitivity properties. Casellina et al. (2023) used Monte Carlo simulation of the asymptotic single risk factor model to quantify the bias of VaR estimates for banking regulations. There is also vast literature on estimation and inference for risk measures (Embrechts et al., 1997; Fan and Gu, 2003; Gao and Song, 2008; Chun et al., 2012; Jin et al., 2003; Jiang and Kou, 2021; Beutner et al., 2024, among many others).

However, a generic way to recover the desired risk guarantee with estimated risk seems missing. For example, the regulator may require banks to compute not only the 99% VaR but also its confidence interval, but how should the banks use that information to uphold the "guarantee" that the probability of insolvency is capped by 1%? This paper proposes a general method to incorporate estimation error into tail risk control problems that comes with a guarantee of this type. Under the assumption that a valid confidence interval is available for the risk measure, we provide a simple method to bound the true *unobservable* risk probability using an *observable* risk estimate. Our method applies to risk measures that aim to control the probability of undesirable events, and we show that VaR and ES—arguably the two most popular regulatory risk measures—are contained as special cases.

The main idea of the paper is best illustrated in the specific case of VaR. In particular, the VaR is defined as the maximum loss that can be incurred with a user-specified probability 1 - p. For example, if we let X be the random variable that represents the asset value of an institution, the VaR is given by the smallest value of

C that satisfies $P(-X > C) \le p^2$ The idea of "controlling VaR" is to maintain the capital reserve of C so that the institution stays solvent with probability at least 1-p when X realizes. We denote such C as VaR_p and call it the VaR of coverage $1-p^3$. We call the event -X > C the risk event, and its probability the risk probability. The "risk guarantee" put forward by VaR is, therefore, that the probability of this risk event $-X > VaR_p$ is bound to be at most p.

The problem of estimation error is that, when the VaR is replaced by its estimate $\widehat{\text{VaR}}_p$, we no longer have $P(-X > \widehat{\text{VaR}}_p) \leq p$ as a theoretical guarantee. In fact, we demonstrate in Section 2 that the actual coverage is distorted, that is, the left-hand side (LHS) is strictly greater than p when we use a naive VaR estimate.⁴ This distortion can be aggravated when we consider a very small value of p, e.g., when we wish to control the chance of a catastrophic event such as financial crisis.

Our idea is to "allot" the risk probability allowance p for two undesirable events separately: (1) the loss exceeds the VaR and (2) the estimated VaR underestimates the true VaR. Let p = q + r be a user-specified allocation of p, and suppose that a onesided (1-r)-confidence interval for VaR_q is available, that is, $P(\text{VaR}_q > \text{VaR}_{q,r}) \leq r$ holds for some observable quantity $VaR_{a,r}$. Then, by the Bonferroni inequality, we have $P(-X > \widetilde{\operatorname{VaR}}_{q,r}) \leq P(-X > \operatorname{VaR}_q) + P(\operatorname{VaR}_q > \widetilde{\operatorname{VaR}}_{q,r}) \leq q + r = p.$ This is because, when the event $-X > VaR_{q,r}$ takes place, we have that either (1) $-X > \operatorname{VaR}_q$ or (2) $\operatorname{VaR}_q > \operatorname{VaR}_{q,r}$ must take place. Note that the LHS does not depend on the unknown, VaR_{q} , but only on the observable, $VaR_{q,r}$. Therefore, by holding the capital worth $VaR_{q,r}$, we can make sure that the actual risk probability is capped by p, including the sampling variation of $\operatorname{VaR}_{q,r}$. A legitimate concern here is that the Bonferroni inequality is known to be a "loose" bound. To address this point, we demonstrate in the empirical application in Section 4 that $\operatorname{VaR}_{q,r}$ is only modestly larger than a naive estimate \widehat{VaR}_p , roughly by 10–50%. Also, we give a practical recommendation of the allocation (q, r) that is not conservative asymptotically so that no allowance is wasted when we have a large dataset to estimate VaR precisely. We show that the above technique extends to other risk measures that aim to control the probability of various unwanted events.

²The C that attains the equality P(-X > C) = p may not exist if the distribution of X is discrete.

³This 1 - p is often called the "confidence level" in the literature, but because we deal with the "confidence interval" of the risk measure, we adopt a different term to forestall needless confusion.

 $^{^{4}}$ Casellina et al. (2023) also observed such systematic bias in risk probability.

This paper adds to the long history of concerns on estimation error in risk control (Jorion, 1996; Hendricks, 1996; Pritsker, 1997, 2006; Barone-Adesi and Giannopoulos, 2001; Berkowitz and O'Brien, 2002; Aussenegg and Miazhynskaia, 2006; Thiele, 2019). Chen (2008) noted that the effective sample size for ES at confidence level 1 - p is the actual sample size times p^2 , stating that the estimate's high volatility is a common challenge for statistical inference. Caccioli et al. (2018) noted that because of high dimensionality of institutional portfolios and the lack of long-run stationarity, portfolio optimization is plagued by (relatively) small sample sizes. Given the statistical difficulty inherent in tail risks, Deo and Murthy (2020) proposed an importance sampling method to consistently approximate ES-based objectives for heavy-tailed distributions. The estimation risk and model risk are considered in a unified optimization framework in portfolio management and related contexts (El Ghaoui et al., 2003; Chen et al., 2007; Natarajan et al., 2008, 2009).

This paper is organized as follows. Section 2 presents a simulation exercise that demonstrates that the estimation error acts adversely to risk control. Section 3 defines the class of tail risk measures and develops a simple method to control the true risk probability using an observable quantity. Section 4 applies our method to an empirical application of controlling VaR and ES in a stylized investment problem and compares our risk estimates against naive estimates. Section 5 concludes. Appendix A provides additional simulation results to supplement Section 2.

2 Failure to Control Tail Risk due to Estimation Error

The defining feature of VaR is that the probability of a loss exceeding the VaR is bounded at a desired level, $P(-X > \text{VaR}_p) \leq p$. We show, however, that if we substitute the VaR with an estimator, the actual risk probability can be larger than the proclaimed level p, building on a generalized autoregressive conditional heteroskedasticity (GARCH) model that has gained strong empirical support for capturing volatility in financial returns (Bollerslev et al., 1992; Engle and Patton, 2007).⁵

The setup of the simulation is as follows. First, we draw T = 200 observations

⁵Simulation of i.i.d. returns exhibits the same problem. See Kaji (2018) for details.



(a) Nonparametric estimator of VaR for p = 0.05. The MSE is 0.0339.





(b) Coverage of estimator in (a). The *p*-value for H_0 : Avg = 0.05 is 6.48×10^{-10} .



(c) Nonparametric estimator of VaR for p = 0.01. The MSE is 0.0786.

(d) Coverage of estimator in (c). The *p*-value for H_0 : Avg = 0.01 is 5.86×10^{-34} .

Figure 1: Histograms of the estimated VaR and its coverage on the simulation of GARCH(1,1). The red lines on the left figures are the true $\text{VaR}_p/\sigma_{T+1}$, and those on the right are the intended risk probability p.

 $\{X_1,\ldots,X_T\}$ from a GARCH(1,1) model,

$$\begin{cases} X_t = \sigma_t z_t, \\ \sigma_t^2 = \omega + \alpha z_{t-1}^2 + \beta \sigma_{t-1}^2 \end{cases}$$

where the true parameters are $\omega = 0.001$, $\alpha = 0.05$, and $\beta = 0.9$, and z_t follows independent standard normal distribution. We first estimate (ω, α, β) by maximum likelihood estimation (MLE) and fit innovations $\{\hat{z}_1, \ldots, \hat{z}_T\}$ and volatility $\hat{\sigma}_{T+1}$. We then estimate VaR_p for $p \in \{0.05, 0.01\}$ for the next observation X_{T+1} by $\widehat{\text{VaR}}_p = -\hat{\sigma}_{T+1}\kappa_p$, where κ_p is given by the empirical *p*-quantile of $\{\hat{z}_1, \ldots, \hat{z}_T\}$. Since $z_T \sim N(0, 1)$, we can calculate the exact risk probability conditional on the history, $P(-X_{T+1} > \widehat{\text{VaR}}_p \mid \mathcal{F}_T)$.⁶ We repeat this exercise for S = 1,000 times with newly drawn observations and compute the unconditional risk probability $P(-X_{T+1} > \widehat{\text{VaR}}_p)$.

Figure 1 presents the histograms of the estimated risk and the risk probability for

⁶Here, \mathcal{F}_T is the σ -field that contains all information up to time T, including the VaR estimate.

VaR of coverage 95% (p = 0.05) and 99% (p = 0.01). Figure 1a shows the histogram of $\widehat{\text{VaR}}_p$ for X_{T+1} normalized by the volatility σ_{T+1} . Since the variance of X_{T+1} differs in each iteration, this normalization makes the estimates comparable across iterations. The red line indicates the true value of $\operatorname{VaR}_p/\sigma_{T+1}$.

Figure 1b is the histogram of the risk probability conditional on history. The red line indicates the intended risk probability p, and the blue dotted line is the average of the conditional risk probability over iterations, estimating the actual unconditional risk probability. We may interpret 0.0539 as meaning that if we maintain the capital reserve of $\widehat{\text{VaR}}_p$, the risk event takes place about 5.32% of the time, more often than the intended 5%. This frequency is subject to the "sampling error" of the simulation, so we conduct a *t*-test. By design, each draw of the conditional risk probability is independent and identically distributed. We can then easily test the hypothesis that the true unconditional risk probability equals the intended level,

$$H_0: P(-X_{T+1} > \widehat{\operatorname{VaR}}_p) = p.$$

The *p*-value is 9.04×10^{-11} as given in the caption of Figure 1b, indicating that the unconditional risk probability exceeds the intended level with statistical significance.

Note that the distortion rate is substantially worse for p = 0.01. In Figure 1d, the risk event occurs 1.39% of the time as opposed to the intended 1% (which is also statistically significant), so the rate of distortion is $(0.0139 - 0.01)/0.01 \approx 38.8\%$. This much larger than the 6.4% for p = 0.05. This is especially problematic when we want to minimize the probability of a catastrophic event such as financial crisis.⁷

In Appendix A, we use parametric and semiparametric estimators of VaR and find that better estimation of VaR does not necessarily lead to less distortion of risk probability and that distortion can take place even when the estimator overestimates VaR on average.⁸

⁷For regulatory capital calculations within the Basel framework, 99.9% coverage (p = 0.001) is typically used for VaR.

⁸Tsafack and Cataldo (2021) also reported that even an unbiased estimator of VaR is likely to produce systematic overviolation (coverage distortion).

3 Main Results

This section presents the main results of the paper. Section 3.1 defines the class of risk measures we consider. Section 3.2 provides a method to control the true risk probability and formally proves the guarantee. Section 3.3 gives a practical recommendation of the tuning parameters (q, r).

3.1 Definition

The risk measures to which our method applies are the ones that come with probability guarantees. We start with reformulating VaR and then generalize it. Let (Ω, \mathcal{F}, P) be the probability space we consider.

The key feature of VaR is that it provides a guarantee on the probability of a *risk* event, $\{-X > C\}$. Recall that VaR_p is formally defined as the smallest value of C such that

$$P(-X > C) \le p.$$

Equivalently, it can be cast as the smallest value of C such that

$$\sup_{\substack{E \in \mathcal{F}}} \left\{ P(E) : \inf_{\substack{\omega \in E}} -X(\omega) - C > 0 \right\} \le p.$$

the maximum prob-
ability of an event in which the smallest is at
loss exceeds C most p

The variables in this definition are visualized in Figure 2a. The middle component,

$$\inf_{\omega \in E} -X(\omega) - C_{\varepsilon}$$

is the key riskiness metric embedded in VaR. In other words, VaR defines the risk of an event E to be such that, even the best-case loss in that event is greater than the capital reserve C. Holding the capital worth C, therefore, corresponds to keeping the maximum probability of such an event at or below p.

Other risk measures can be obtained by replacing this riskiness component. Acerbi et al. (2001) defined the ES of coverage 1 - p as

$$\operatorname{ES}_p(X) \coloneqq \frac{1}{p} \int_0^p Q(\alpha) d\alpha$$



(a) VaR regards the risk of event E as the smallest loss exceeding reserve C.



(b) ES regards the risk of event E as the expected value exceeding reserve C.

Figure 2: Illustration of the definitions of VaR and ES as tail risk measures.

where Q is the quantile function of X. This can be cast as the smallest value of C such that

$$\sup_{E \in \mathcal{F}} \left\{ P(E) : \mathbb{E}[-X \mid E] - C > 0 \right\} \le p,$$

the maximum prob-
ability of an event loss exceeds C most p

as illustrated in Figure 2b. Thus, ES defines the riskiness of X by $\mathbb{E}[-X \mid E] - C$ when the event E takes place and the capital reserve is C.⁹

Let $\rho(X, C \mid E)$ be an arbitrary metric of *riskiness* of asset value X under event E that is nonincreasing in reserve C. The tail risk measure is formally defined as follows.

Definition (tail risk measure). The *tail risk measure of coverage* 1-p is the infimum of C such that

$$\sup_{E \in \mathcal{F}} \left\{ P(E) : \varrho(X, C \mid E) > 0 \right\} \le p.$$

If there is no event E that attains $\rho(X, C \mid E) > 0$ for a fixed C, we may understand the supremum in that case to be 0.

Intuitively, the tail risk measure controls the probability of the undesirable event $\rho > 0$. The supremum over \mathcal{F} represents a search for the maximum probability that it can happen. The value C is the "capital reserve" with which we can avoid the risk event with probability at least 1 - p. This is a direct control of the risk probability compared to ones relying on the Markov bound (Roy, 1952; Goovaerts et al., 2003).

The choice of ρ defines a risk measure. It is straightforward to extend it to various

⁹If we adopt $\mathbb{E}[-X \mid X \in X(E)] - C$, we obtain what is called the tail conditional expectation by Acerbi (2002). Here, X(E) is the set of values that X may take under the event E.

tail risks. For example, the "two-sided" expected shortfall may be obtained via

$$\varrho(X, C \mid E) = \mathbb{E}[|X| \mid E] - C,$$

which is useful when we want to use a common risk measure for longing and shorting the asset, as in shorting the straddle option strategy. Or, the downside tail conditional variance can be obtained by

$$\varrho(X, C \mid E) = \mathbb{E}[(\min\{X - \mathbb{E}[X], 0\})^2 \mid E] - C,$$

which is essentially what was discussed in Valdez (2005). We may also obtain the "utility-based" shortfall risk,

$$\varrho(X, C \mid E) = \mathbb{E}[\ell(-X - C) \mid E],$$

for a nondecreasing loss function $\ell : \mathbb{R} \to \mathbb{R}$. This allows one to, e.g., incorporate loss aversion into the risk measure.

Our definition shares a similar spirit with the tail risk measure analyzed by Liu and Wang (2021), although there are some notable differences. For $p \in (0, 1)$, they defined a (1 - p)-tail risk measure $\rho(X)$ to be one that depends only on the tail distribution of X, that is, $\rho(X) = \rho(Y)$ if X and Y have a common quantile function on the interval (0, p]. Our definition, on the other hand, regards any event as a risk event so long as $\rho > 0$, resulting in the supremum over the entire set of events \mathcal{F} . For VaR and ES, the "maximizing events" turn out to be identical to the tail events considered in Liu and Wang (2021), while we leave other possibilities such as for the two-sided expected shortfall defined above. Allowing the risk event to be arbitrary may be convenient when we consider risks of multiple assets since the "tail events" of different assets may occur in different situations.

Having said that, for the purpose of this paper, there is nothing that prohibits us from maximizing over a subset of \mathcal{F}^{10} . If we take the supremum over $\{\{X < c\} : c \in \mathbb{R}\}$, for example, our definition looks much more in line with that of Liu and Wang (2021). However, even with this modification, the two definitions do not perfectly coincide. A key property that characterizes ours is that the risk measure is monotonic in p. Let c_p be the tail risk measure of coverage 1 - p. Then, we have

¹⁰Theorem 2 below continues to hold if we replace \mathcal{F} with an arbitrary subset of it.

 $c_q \ge c_p$ for q < p by construction.

Lemma 1 (Monotonicity in *p*). The tail risk measure c_p is nonincreasing in *p*. Proof. For q < p, we have

$$\sup_{E \in \mathcal{F}} \left\{ P(E) : \varrho(X, c_q \mid E) > 0 \right\} \le q < p.$$

Since c_p is the infimum that bounds the LHS by p, we must have $c_q \ge c_p$.

This monotonicity is essential in our framework since it ensures that c_q serves as a "more conservative" alternative to c_p . Meanwhile, some examples discussed in Liu and Wang (2021) such as the tail standard deviation are not necessarily monotonic in p.¹¹

A risk measure is often discussed in connection with the dual theory of choice (Yaari, 1987) or the Choquet expected utility theory (Bassett et al., 2004). There, various risk measures are expressed as a weighted integral of the quantile function Q of X, i.e.,

$$\rho(X) = \int_0^1 -Q(u) dg(u)$$

for nondecreasing g such that g(0) = 0 and g(1) = 1. For example, the VaR of coverage 1 - p emerges with $g(u) = \mathbb{1}\{u \ge p\}$, and the ES of coverage 1 - p with $g(u) = 0 \lor \frac{u}{p} \land 1$. This type of risk measure is generally known as the *distortion risk* measure. If g is concave, it is called the *spectral risk measure* (Acerbi, 2002, Definition 3.1) and is known to be coherent (Balbás et al., 2009, p. 390). Bassett et al. (2004) associated concavity of g with pessimism and convexity of g with optimism. If $g \equiv 1$ on the interval [p, 1], it can be viewed that the distorted probability focuses exclusively on the tail event (0, p].

To bring the distortion risk measure into our framework, we distinguish the part that extracts the tail event and the part that represents the risk preference. Let $g: [0,1] \rightarrow [0,1]$ be a nondecreasing function with g(0) = 0 and g(1) = 1, and let $\mathbb{E}_{g}[X \mid E]$ be the distorted conditional expectation such that

$$\mathbb{E}_g[X \mid E] = \int_0^1 Q_{X|E}(u) dg(u),$$

¹¹Note that unlike the "downside tail conditional variance" defined above, the tail standard deviation considered in Liu and Wang (2021) involves centering at the tail conditional expectation, which shatters monotonicity.

where $Q_{X|E}$ is the quantile function of X conditional on event E. With this, we can define the distorted tail risk measure via

$$\varrho(X, C \mid E) = \mathbb{E}_g[-X \mid E] - C.$$

Finally, we note that there are well-known risk measures that are not tail risk measures. A simple example is variance, which concerns the entire distribution of X. For the same reason, the expectile (Newey and Powell, 1987) and entropic VaR (Ahmadi-Javid, 2012) fail to be tail risk measures. Both of them are closely related to VaR and ES and are defined for each risk level $p \in (0, 1)$. However, their definitions involve the entire distribution of X at every value of p, so they do not concern a particular tail event. For expectiles, p does not even represent probability.

3.2 Tail Risk Estimates with Guarantees

This section provides a method to control the actual risk probability with an observable risk estimate. The essential assumption is that a (one-sided) confidence interval for the tail risk measure c_q is available, that is, we have access to an observable quantity $\tilde{c}_{q,r}$ that satisfies

$$P(c_q > \tilde{c}_{q,r}) \le r$$

for an arbitrary confidence level 1 - r.

To lay out how the Bonferroni bound can be generalized to arbitrary tail risk measures, let us illustrate it for the case of ES. Note that the risk probability of ES can be written as

$$\sup_{E \in \mathcal{F}} P(E, \mathbb{E}[-X \mid E] > \mathrm{ES}_p).$$

Since " $\mathbb{E}[-X \mid E] > \mathrm{ES}_p$ " is a nonrandom statement given E, the probability is either P(E) when it holds or 0 otherwise. If we substitute the confidence interval into the ES, we can apply the Bonferroni bound as

$$P(E, \mathbb{E}[-X \mid E] > \widetilde{\mathrm{ES}}_{q,r}) \le P(E, \mathbb{E}[-X \mid E] > \mathrm{ES}_q) + P(E, \mathrm{ES}_q > \widetilde{\mathrm{ES}}_{q,r}),$$

since violation of both inequalities on the right implies violation of the one on the

left. By taking the supremum of both sides with respect to E, we find

$$\sup_{E \in \mathcal{F}} P(E, \mathbb{E}[-X \mid E] > \widetilde{\mathrm{ES}}_{q,r}) \leq \sup_{E \in \mathcal{F}} \left[P(E, \mathbb{E}[-X \mid E] > \mathrm{ES}_q) + P(E, \mathrm{ES}_q > \widetilde{\mathrm{ES}}_{q,r}) \right]$$
$$\leq \sup_{E \in \mathcal{F}} P(E, \mathbb{E}[-X \mid E] > \mathrm{ES}_q) + P(\mathrm{ES}_q > \widetilde{\mathrm{ES}}_{q,r})$$
$$\leq q + r.$$

Thus, the risk probability including the sampling error of $\widetilde{\text{ES}}_{q,r}$ is bounded by q + r. The general statement is as follows.

Theorem 2 (Tail risk control with estimation error). Let c_q be the tail risk measure of coverage 1-q, and $\tilde{c}_{q,r}$ the (1-r)-confidence bound of c_q , that is, $P(c_q > \tilde{c}_{q,r}) \leq r$. Then, for p = q + r, we have

$$\sup_{E \in \mathcal{F}} P(E, \, \varrho(X, \tilde{c}_{q,r} \mid E) > 0) \le p.$$

Proof. Fix an event $E \in \mathcal{F}$. By the Bonferroni inequality, we have

$$P(E, \varrho(X, \tilde{c}_{q,r} \mid E) > 0) \le P(E, \varrho(X, c_q \mid E) > 0) + P(E, c_q > \tilde{c}_{q,r}).$$

Taking the supremum of both sides with respect to E, we find

$$\sup_{E \in \mathcal{F}} P(E, \ \varrho(X, \tilde{c}_{q,r} \mid E) > 0) \leq \sup_{E \in \mathcal{F}} \{ P(E, \ \varrho(X, c_q \mid E) > 0) + P(E, \ c_q > \tilde{c}_{q,r}) \}$$
$$\leq \sup_{E \in \mathcal{F}} \{ P(E) : \varrho(X, c_q \mid E) > 0 \} + P(c_q > \tilde{c}_{q,r})$$
$$\leq q + r = p.$$

This completes the proof.

Note that Theorem 2 does not make use of asymptotic arguments. Therefore, it is valid in finite samples as long as the confidence interval is. It is however the case that many existing inference methods are justified as asymptotic approximations. In that case, the risk statement given in Theorem 2 should also be understood in asymptotic terms.

While the theorem is valid for an arbitrary choice of (q, r), we cannot "hunt" for the smallest $\tilde{c}_{q,r}$ after observing the data. This changes the distribution of $\tilde{c}_{q,r}$ for randomness introduced by the hunting, and the probability bound will no longer be valid. Such is an example of data dredging. In the next section, we discuss a practical choice of (q, r) that does not depend on the data.

The theorem assumes a one-sided confidence interval, but we note that the upper bound of any two-sided confidence interval is by construction a valid one-sided confidence interval at the same confidence level, although it will be more conservative than necessary. Also, our results require only a confidence interval but not an estimator, so the inference methods that do not require an estimator can also be used.

When the risk measure is subadditive, it is straightforward to tweak Theorem 2 for the sum of assets X + Y. Subadditivity has gained its popularity as one of the desiderata of risk measures (Artzner et al., 1999; Acerbi, 2002; Dowd and Blake, 2006). It is well known that ES is subadditive (McNeil et al., 2005, Proposition 6.9; Embrechts and Wang, 2015), while VaR is subadditive under additional assumptions (Gourieroux et al., 2000, Section 2.4; Ibragimov, 2005, p. 25; McNeil et al., 2005, Theorem 6.8; Garcia et al., 2007, Proposition 3.1). If a tail risk measure is subadditive, it means that for the smallest values $c_{p,X}$, $c_{p,Y}$, and $c_{p,X+Y}$ such that

$$\sup_{E \in \mathcal{F}} \left\{ P(E) : \varrho(X, c_{p,X} \mid E) > 0 \right\} \le p,$$
$$\sup_{E \in \mathcal{F}} \left\{ P(E) : \varrho(Y, c_{p,Y} \mid E) > 0 \right\} \le p,$$
$$\sup_{E \in \mathcal{F}} \left\{ P(E) : \varrho(X + Y, c_{p,X+Y} \mid E) > 0 \right\} \le p,$$

we have $c_{p,X+Y} \leq c_{p,X} + c_{p,Y}$. Since ρ is nonincreasing in C, we also have

$$\sup_{E \in \mathcal{F}} \left\{ P(E) : \varrho(X + Y, c_{p,X} + c_{p,Y} \mid E) > 0 \right\} \le p.$$

Therefore, if a joint confidence set for $(c_{p,X}, c_{p,Y})$ is available, we may use the (1-r)confidence bound for $c_{p,X} + c_{p,Y}$ as the capital reserve for X + Y. If we only have access
to the marginal confidence bounds for $c_{p,X}$ and $c_{p,Y}$, as is the case in Section 4, we
may use the sum of (1-r/2)-confidence bounds for $c_{p,X}$ and $c_{p,Y}$ as the capital reserve
for X + Y. This is another application of the Bonferroni bound. Finally, we note
that when there is enough computational resource, we can avoid overconservatism
introduced by the use of subadditivity and Bonferroni bound altogether by directly
estimating $c_{p,X+Y}$ and using its (1-r)-confidence bound.

Finally, we note that estimation and inference for risk measures are a widely studied area, ranging from parametric to nonparametric methods. Embrechts et al. (1997) and Bali and Theodossiou (2008) considered estimation and inference of VaR and ES based on extreme value theory.¹² Chen and Tang (2005) and Scaillet (2004) discussed nonparametric estimation and inference of VaR and ES. Linton and Xiao (2013) and Hill (2015) considered nonparametric estimation and inference of ES when X may not have a variance. Belomestny and Krätschmer (2012) established asymptotic normality of plug-in estimators of law-invariant coherent risk measures. Gao and Song (2008) derived asymptotic distribution of VaR and ES estimators in the GARCH model estimated by the filtered historical simulation method. There is also large literature on estimation and inference of VaR and ES conditional on covariates (Chernozhukov and Umantsev, 2001; Cai and Wang, 2008; Kato, 2012; Chernozhukov and Fernández-Val, 2011; Chun et al., 2012; Chernozhukov et al., 2017; Martins-Filho et al., 2018; Candila et al., 2023) as well as regression-based nested simulation methods to estimate risk measures in response to risk factors (Broadie et al., 2011, 2015).

3.3 Practical Choice of r

As the sample size increases, estimation of the risk measure is expected to be more precise. This provokes a thought that we can spend less allowance on r and let q be closer to p. In this section, we discuss the choice of r as a function of the sample size T, assuming that the confidence interval is constructed with a (sub)polynomially-tailed distribution.

Suppose, first, that \hat{c}_q is a normally-distributed unbiased estimator with known variance σ_q^2 , that is,

$$\sqrt{T}(\hat{c}_q - c_q) \sim N(0, \sigma_q^2).$$

The one-sided (1 - r)-confidence bound for c_q is then given by

$$\hat{c}_q + \frac{\sigma_q}{\sqrt{T}}\kappa_r,$$

where κ_r is the (1-r)th quantile of the standard normal distribution. This suggests that the confidence bound shrinks at rate $1/\sqrt{T}$ for fixed r and blows up as $r \to 0$ for fixed T. In reasonable situations, we can expect that σ_q is continuous at q = p,

¹²See also Nolde and Zhou (2021) for a concise review.

so it is enough to consider a shrinking sequence $r \to 0$ such that κ_r / \sqrt{T} goes to zero.

Let ϕ and Φ be the pdf and cdf of a standard normal distribution. Then, κ_r can be expressed as

$$\kappa_r = \Phi^{-1}(1-r).$$

We know, by the property of the Mills ratio, that for $x \ge 1/\sqrt{2\pi}$, we have

$$1 - \Phi(x) = \Phi(-x) < \frac{1}{x}\phi(x) \le e^{-x^2/2}.$$

Therefore, for $0 < r \leq e^{-1/4\pi}$, we have

$$\kappa_r = \Phi^{-1}(1-r) = -\Phi^{-1}(r) < \sqrt{-2\log r}.$$

Thus, we can let $r \to 0$ at the speed at which $\kappa_r/\sqrt{T} < \sqrt{-2\log r}/\sqrt{T} \to 0$, i.e., $r \gg e^{-T/2}$. Therefore, r can indeed go very fast to zero; e.g., any rational function in T converges slower than $e^{-T/2}$. For the sample sizes of order hundreds to thousands, r = 1/T would be a good choice.

In other cases, the confidence interval may arise from a distribution that is more heavy-tailed than the normal distribution, such as the *t*-distribution. Suppose that κ_r is bounded by a polynomial function,

$$\kappa_r \lesssim r^{-\frac{1}{\nu}},$$

for some $\nu > 0$ when r is close to zero. This is the case when, e.g., κ_r is the quantile of a t-distribution with ν degrees of freedom. Then, we need that $r^{-1/\nu}/\sqrt{T}$ go to zero, i.e., $r \gg T^{-\nu/2}$. In other words, if $\nu > 2$ (which corresponds to having a finite variance), the choice r = 1/T is justified.

We summarize that our recommended choice is $r = T^{-1}$ for applications where the estimator is root-*n* consistent and the asymptotic distribution has a finite variance. This covers most estimators referenced at the end of Section 3.2. One exception is Linton and Xiao (2013), who considered estimation of ES when the underlying time series has an infinite variance. They observed that the convergence rate is strictly slower than \sqrt{T} and the limiting distribution is a stable law. In that case, we will need to find our own choice of r by going through similar considerations. Or, we may use the trimming method proposed by Hill (2015) to bring it back to an asymptotically normal setup.

Practicality of the above argument may be diminished when the standard error is expected to be relatively large. Another possible method is sample-splitting.

- 1. Split the sample into two subsamples $\mathcal{T}_1 = \{1, \ldots, \tau\}$ and $\mathcal{T}_2 = \{\tau + 1, \ldots, T\}$.
- 2. Estimate $\tilde{c}_{q,r}$ as a function of (q,r) using \mathcal{T}_1 .
- 3. Pick (q^*, r^*) that minimizes $\tilde{c}_{q,r}$.
- 4. Estimate \tilde{c}_{q^*,r^*} using \mathcal{T}_2 .

The subsample \mathcal{T}_1 may be replaced with simulation draws if we have a reasonable base model. Various simulation methods are developed in the literature (Glasserman et al., 2000; Jin et al., 2003; Lesnevski et al., 2007; Hong and Liu, 2009; Fuh et al., 2011; Jiang and Kou, 2021).

4 Empirical Application to VaR and ES

In this section, we apply our method to real data drawing on an investment problem. The purpose is to demonstrate that our proposed quantity is practically not too large compared to a naive estimator, despite relying on a Bonferroni bound.

Consider the problem of controlling VaR and ES of a portfolio with three assets: the stock of the Bank of America Corp. (BAC), the stock of Morgan Stanley (MS), and the index fund for the Dow Jones Industrial Average (DJI). The choice of assets is due to the ease of access to the data, but it can be any assets, e.g., on a bank's balance sheet. We use the daily adjusted close values from February 23, 1993 to December 31, 2017. The price data are retrieved from Yahoo! Finance. Figure 3 shows the adjusted close values of these assets.

Denoting by Y_t the daily close value, the daily return is calculated as $X_t = (Y_t - Y_{t-1})/Y_{t-1}$. Figure 4 shows the daily returns of the three assets. The daily portfolio return is given by

$$X = w_1 X_{\text{BAC}} + w_2 X_{\text{MS}} + w_3 X_{\text{DJI}},$$

where X_{TIC} is the daily return of the stock of ticker symbol TIC, and (w_1, w_2, w_3) are the weights. Our goal is to estimate VaR and ES of coverage 1 - p of the return X of the entire portfolio without losing probability guarantees.



Figure 3: Daily adjusted close values of BAC, MS, and DJI from Feb. 23, 1993 to Dec. 31, 2017. BAC and MS are on the left *y*-axis and DJI on the right.



Figure 4: Daily returns of BAC, MS, and DJI from Feb. 23, 1993 to Dec. 31, 2017. We see stochastic trends in their volatility.

We model each daily return as a GARCH(1,1) process,

$$\begin{cases} X_t = \mu + \sigma_t z_t, \\ \sigma_t^2 = \omega + \alpha z_{t-1}^2 + \beta \sigma_{t-1}^2 \end{cases}$$

where $\{z_t\}$ are i.i.d. random variables and $\omega, \alpha, \beta \ge 0$ are GARCH parameters.¹³ We use Gao and Song (2008) to estimate VaR and ES for individual assets.

Let r = 1/T and q = p - 3r. The reason why q is more conservative than p - r is as discussed at the end of Section 3.2. The asymptotic distribution given in Gao and Song (2008, Theorems 3.2 and 3.3) is of the marginal distribution of each estimator. To ignore the correlation of risk estimates, we need to allot r = 1/T to the confidence bound of each asset separately. If there are many assets, this would lead to a very conservative bound. An alternative is to use bootstrap methods such as Beutner et al. (2024) and estimate the joint distribution of risk measures across assets.

 $^{^{13}}X_t$ is stationarity if $\alpha + \beta < 1$ and i.i.d. if $\alpha = \beta = 0$.



Figure 5: Estimated VaR and ES of daily returns for (p, q, r) = (0.05, p - 3/T, 1/T). The top figures use the data from Nov. 1, 2016 to Dec. 31, 2017 (T = 293), and the bottom from Feb. 23, 1993 to Dec. 31, 2017 (T = 6,260).

Our estimation procedure goes as follows.

- 1. Estimate the GARCH parameters $(\mu, \omega, \alpha, \beta)$ by quasi-MLE assuming normality of z_t , and fit $\hat{\sigma}_{T+1}$ and $\{\hat{z}_1, \ldots, \hat{z}_{t-1}\}$.
- 2. Let κ_q be the empirical q-quantile of \hat{z}_t and compute the q-trimmed average of \hat{z}_t by $\eta_q = \sum_{t=1}^T \hat{z}_t \mathbb{1}\{\hat{z}_t < \kappa_q\} / \sum_{t=1}^T \mathbb{1}\{\hat{z}_t < \kappa_q\}.$
- 3. Obtain the point estimates of VaR_q and ES_q by $\widehat{\text{VaR}}_q = \hat{\sigma}_{T+1}\kappa_q$ and $\widehat{\text{ES}}_q = \hat{\sigma}_{T+1}\eta_q$.
- 4. Estimate $\operatorname{Var}(\widehat{\operatorname{VaR}}_q)$ and $\operatorname{Var}(\widehat{\operatorname{ES}}_q)$ using analytical formulae provided in Gao and Song (2008, Theorems 3.2 and 3.3).

5. Construct the one-sided (1 - r)-confidence bounds $\operatorname{VaR}_{q,r}$ and $\operatorname{ES}_{q,r}$, which are our proposed risk estimates.

Figure 5 shows the estimates of VaR and ES, which are also summarized in Table 1. Figures 5a and 5b use the data from November 1, 2016 to December 31, 2017, consisting of T = 293 daily returns; Figures 5c and 5d from February 23, 1993 to December 31, 2017, totaling T = 6,260 observations. The top three bars in Figure 5a show the "naive" estimates of VaR of coverage 1 - p, which have no guarantee on the actual risk probability as discussed in Section 2. The next three bars are the interim estimates of VaR of coverage 1 - q. The last three bars are the upper bounds of the one-sided (1 - r)-confidence intervals of VaR_q; they satisfy the risk guarantee that the probability of the next loss going above is less than p.

The numbers in parentheses in Table 1 are the multipliers relative to the "naive" estimates; in columns (1–3) they are ratios relative to column (1), and in columns (4–6) ratios relative to column (4); for example, in column (3), $\widetilde{\text{VaR}}_{q,r}/\widehat{\text{VaR}}_p$, and in column (6), $\widetilde{\text{ES}}_{q,r}/\widehat{\text{ES}}_p$. The VaR has overall low multipliers, meaning that estimation error can be addressed without being too conservative. The ES has bigger multipliers, but the magnitudes are still comparable to the multipliers needed for model risk.¹⁴ Whenever we can construct a confidence interval, we advocate the use of our method since the multipliers heavily depend on the sample size. When statistical inference is not available, using a fixed multiplier of 2 may be practically reasonable to address estimation error of VaR and ES.¹⁵

With the above results, we may control the risk of the entire portfolio. Assuming that VaR is subadditive (Gourieroux et al., 2000, Section 2.4; Ibragimov, 2005, p. 25; McNeil et al., 2005, Theorem 6.8; Garcia et al., 2007, Proposition 3.1), we can bound the VaR of the entire portfolio of coverage 1 - p by

$$w_1 2.05\% + w_2 2.01\% + w_3 0.65\%,$$

where (2.05%, 2.01%, 0.65%) are the VaR_{q,r} from Figure 5a and (w_1, w_2, w_3) are weights of the portfolio on BAC, MS, and DJI. Note that even without the assumption of subadditivity, we can directly compute the bound by applying the same exercise to

¹⁴The recommended multiplier for model risk is 3 for VaR (Stahl, 1997) and 1.5 for ES (Leippold and Vanini, 2002).

¹⁵Note that our estimates used a conservative choice of q = p - 3r to get away with correlation estimation, which may be unnecessary in some applications.

	(1)	(2)	(3)	(4)	(5)	(6)
	$\widehat{\mathrm{VaR}}_p$ [%]	$\widehat{\operatorname{VaR}}_q$ [%]	$\widetilde{\mathrm{VaR}}_{q,r} \ [\%]$	$\widehat{\mathrm{ES}}_p$ [%]	$\widehat{\mathrm{ES}}_q$ [%]	$\widetilde{\mathrm{ES}}_{q,r}$ [%]
Nov. 1, 2	016 to Dec.	31, 2017 (T	= 293)			
BAC	1.649	1.762	2.047	2.772	3.046	4.612
	(1.000)	(1.069)	(1.241)	(1.000)	(1.099)	(1.664)
MS	1.369	1.466	2.009	2.584	2.806	4.054
	(1.000)	(1.071)	(1.468)	(1.000)	(1.086)	(1.569)
DJI	0.494	0.520	0.646	0.842	0.917	1.322
	(1.000)	(1.052)	(1.308)	(1.000)	(1.089)	(1.570)
Feb. 24, 1	1993 to Dec.	31, 2017 (7	7 = 6,260)			
BAC	1.800	1.801	1.927	2.596	2.604	2.917
	(1.000)	(1.001)	(1.071)	(1.000)	(1.003)	(1.124)
MS	1.612	1.616	1.848	2.255	2.261	2.634
	(1.000)	(1.002)	(1.146)	(1.000)	(1.003)	(1.168)
DJI	0.792	0.796	0.886	1.134	1.137	1.291
	(1.000)	(1.004)	(1.119)	(1.000)	(1.003)	(1.138)

Table 1: Estimates of 95% VaR and ES in the percentage units. The numbers in parentheses are the multipliers relative to $\widehat{\text{VaR}}_p$ and $\widehat{\text{ES}}_p$.

the transformed historical data $\{w_1X_{BAC} + w_2X_{MS} + w_3X_{DJI}\}$. The advantage of this is that it will not be conservative. If we use subadditivity, it eliminates the need to re-estimate the risk as we consider other weights, but at the expense of being more conservative than necessary.

Since ES is subadditive by construction, we can bound the ES of our portfolio of coverage 95% by

$$w_1 4.61\% + w_2 4.05\% + w_3 1.32\%,$$

where the numbers come from the $\widetilde{\text{ES}}_{q,r}$ in Figure 5b.

We may also allow for short positions (negative weights). Observe that the risk associated with shorting X_t is equal to the risk of longing $-X_t$. Then, we can apply the same method with the risk estimates on the other tail.

5 Conclusion

Many risk measures are motivated by a certain guarantee on the probability of a ruin. Estimation error involved in risk assessment is one of many factors that impair the intended risk guarantee (Section 2). We addressed this issue and proposed a method to recover the intended risk probability guarantee with an observable risk estimate, namely the confidence interval of the risk measure.

We characterized the class of risk measures to which our method can be applied, and named them the *tail risk measure* (Section 3.1). We showed that the class contains VaR and ES, and discussed that it can be extended to various other tail risks (Section 3.2). Our method of risk control is based on the Bonferroni inequality and hence is robust against arbitrary correlation between the risk variable X and the risk estimate. We also provided a recommendation of the tuning parameters (q, r) so that our risk estimate is consistent (Section 3.3).

In the empirical application, we applied our method to the VaR and ES estimation for an arbitrary portfolio of three assets (Section 4). We found in our setup that our proposed risk estimates on VaR are generally larger than the naive estimates by 10–50%, and those on ES by 20–70%. These are modest inflation compared to the multipliers for model risk. This demonstrated that our method produces risk estimates that are practically not too conservative.

Appendices

A Failure to Control Tail Risk with Other Estimators

In Section 2, we showed that the estimation error of a nonparametric estimator of VaR distorts the risk probability. To see how sensitive this is to the choice of the estimator, we examine canonical parametric and semiparametric estimators of VaR.

The setup is the same as before, but now the κ_p in $Va\dot{R}_p = -\hat{\sigma}_{T+1}\kappa_p$ is given by one of the following estimators.

- 1. Parametric. The p-quantile of a standard normal distribution.
- 2. Semiparametric. Weissman's (1978) estimator of the *p*-quantile of $\{\hat{z}_1, \ldots, \hat{z}_T\}$,

$$\kappa_p = \hat{z}_{(k)} - \left(\hat{z}_{(k)} - \frac{1}{k}\sum_{i=1}^k \hat{z}_{(i)}\right) \log\left(\frac{k}{pT}\right) \quad \text{for} \quad k = 10.$$

The results are summarized in Figure 6. Being correctly specified, the parametric estimator achieves the smallest mean squared error (MSE). For p = 0.05, the MSE of the parametric estimator is 0.0207, while those of the semi- and nonparametric estimators are 0.0355 and 0.0339 (as given in the captions of Figures 6a, 6e, and 1a). Thus, in the context of this simulation, the parametric estimator is the "best" among the three. However, this relationship does not carry over to the risk probability. The distortion of the risk probability is 5.32% - 5.00% = 0.32% for the parametric estimator, which is worse than 0.15% of the semiparametric estimator. In other words, higher precision of VaR estimation should not be taken as evidence of less distortion of the risk probability.

Also, it is interesting to note that while the semiparametric estimator *overesti*mates the true VaR for p = 0.05 on average (that is, the blue dotted line is above the red line in Figure 6e), the risk probability is still *above* the intended level (that is, the blue dotted line is above the red line in Figure 6f).

As was for the nonparametric estimator, the observation that the distortion rates are worse for p = 0.01 than for p = 0.05 carries over to the parametric and semiparametric estimators.

References

- ACERBI, C. (2002): "Spectral Measures of Risk: A Coherent Representation of Subjective Risk Aversion," *Journal of Banking and Finance*, 26, 1505–1518.
- ACERBI, C., C. NORDIO, AND C. SIRTORI (2001): "Expected Shortfall as a Tool for Financial Risk Management," ArXiv:cond-mat/0102304.
- ADRIAN, T. AND M. K. BRUNNERMEIER (2016): "CoVaR," American Economic Review, 106, 1705–1741.
- AHMADI-JAVID, A. (2012): "Entropic Value-at-Risk: A New Coherent Risk Measure," Journal of Optimization Theory and Applications, 155, 1105–1123.
- ARTZNER, P., F. DELBAEN, J.-M. EBER, AND D. HEATH (1999): "Coherent Measures of Risk," *Mathematical Finance*, 9, 203–228.



(a) Parametric estimator of VaR for p = 0.05. The MSE is 0.0207.



(c) Parametric estimator of VaR for p = 0.01. The MSE is 0.0414.



(e) Semiparametric estimator of VaR for p = 0.05. The MSE is 0.0355.



(g) Semiparametric estimator of VaR for p = 0.01. The MSE is 0.0749.





(b) Coverage of estimator in (a). The *p*-value for H_0 : Avg = 0.05 is 9.04×10^{-11} .



(d) Coverage of estimator in (c). The *p*-value for H_0 : Avg = 0.01 is 5.02×10^{-19} .



(f) Coverage of estimator in (e). The *p*-value for H_0 : Avg = 0.05 is 0.017.



(h) Coverage of estimator in (g). The *p*-value for H_0 : Avg = 0.01 is 4.15×10^{-57} .

- AUSSENEGG, W. AND T. MIAZHYNSKAIA (2006): "Uncertainty in Value-at-Risk Estimates under Parametric and Non-parametric Modeling," *Financial Markets* and Portfolio Management, 20, 243–264.
- BALBÁS, A., J. GARRIDO, AND S. MAYORAL (2009): "Properties of Distortion Risk Measures," *Methodology and Computing in Applied Probability*, 11, 385–399.
- BALI, T. G. AND P. THEODOSSIOU (2008): "Risk Measurement Performance of Alternative Distribution Functions," *Journal of Risk and Insurance*, 75, 411–437.
- BARONE-ADESI, G. AND K. GIANNOPOULOS (2001): "Non-parametric VaR Techniques. Myths and Realities," *Economic Notes*, 30, 167–181.
- BASSETT, GILBERT W., J., R. KOENKER, AND G. KORDAS (2004): "Pessimistic Portfolio Allocation and Choquet Expected Utility," *Journal of Financial Econometrics*, 2, 477–492.
- BELOMESTNY, D. AND V. KRÄTSCHMER (2012): "Central Limit Theorems for Law-Invariant Coherent Risk Measures," *Journal of Applied Probability*, 49, 1–21.
- BERKOWITZ, J. AND J. O'BRIEN (2002): "How Accurate Are Value-at-Risk Models at Commercial Banks?" *Journal of Finance*, 57, 1093–1111.
- BEUTNER, E., A. HEINEMANN, AND S. SMEEKES (2024): "A Residual Bootstrap for Conditional Value-at-Risk," *Journal of Econometrics*, 238, 105554.
- BLANCHET, J. AND K. MURTHY (2019): "Quantifying Distributional Model Risk via Optimal Transport," *Mathematics of Operations Research*, 44, 565–600.
- BOLLERSLEV, T., R. Y. CHOU, AND K. F. KRONER (1992): "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics*, 52, 5–59.
- BROADIE, M., Y. DU, AND C. C. MOALLEMI (2011): "Efficient Risk Estimation via Nested Sequential Simulation," *Management Science*, 57, 1172–1194.
- (2015): "Risk Estimation via Regression," *Operations Research*, 63, 1077–1097.

- CACCIOLI, F., I. KONDOR, AND G. PAPP (2018): "Portfolio Optimization under Expected Shortfall: Contour Maps of Estimation Error," *Quantitative Finance*, 18, 1295–1313.
- CAI, Z. AND X. WANG (2008): "Nonparametric Estimation of Conditional VaR and Expected Shortfall," *Journal of Econometrics*, 147, 120–130.
- CANDILA, V., G. M. GALLO, AND L. PETRELLA (2023): "Mixed-Frequency Quantile Regressions to Forecast Value-at-Risk and Expected Shortfall," *Annals of Operations Research*, 1–34.
- CAPPONI, A. AND A. RUBTSOV (2022): "Systemic Risk-Driven Portfolio Selection," Operations Research, 70, 1598–1612.
- CASELLINA, S., S. LANDINI, AND M. UBERTI (2023): "Credit Risk Measures and the Estimation Error in the ASRF Model under the Basel II IRB Approach," *Communications in Nonlinear Science and Numerical Simulation*, 118, 106977.
- CERIA, S. AND R. A. STUBBS (2006): "Incorporating Estimation Errors into Portfolio Selection: Robust Portfolio Construction," *Journal of Asset Management*, 7, 109–127.
- CHEN, S. X. (2008): "Nonparametric Estimation of Expected Shortfall," *Journal of Financial Econometrics*, 6, 87–107.
- CHEN, S. X. AND C. Y. TANG (2005): "Nonparametric Inference of Value-at-Risk for Dependent Financial Returns," *Journal of Financial Econometrics*, 3, 227–255.
- CHEN, X., M. SIM, AND P. SUN (2007): "A Robust Optimization Perspective on Stochastic Programming," *Operations Research*, 55, 1058–1071.
- CHERNOZHUKOV, V. AND I. FERNÁNDEZ-VAL (2011): "Inference for Extremal Conditional Quantile Models, with an Application to Market and Birthweight Risks," *Review of Economic Studies*, 78, 559–589.
- CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, AND T. KAJI (2017): "Extremal Quantile Regression," in *Handbook of Quantile Regression*, ed. by R. Koenker, V. Chernozhukov, X. He, and L. Peng, Chapman and Hall/CRC, chap. 18, 333–362.

- CHERNOZHUKOV, V. AND L. UMANTSEV (2001): "Conditional Value-at-Risk: Aspects of Modeling and Estimation," *Empirical Economics*, 26, 271–292.
- CHOPRA, V. K. AND W. T. ZIEMBA (1993): "The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice," *Journal of Portfolio Optimization*, 19, 6–11.
- CHUN, S. Y., A. SHAPIRO, AND S. URYASEV (2012): "Conditional Value-at-Risk and Average Value-at-Risk: Estimation and Asymptotics," *Operations Research*, 60, 739–756.
- CONT, R., R. DEGUEST, AND X. D. HE (2013): "Loss-Based Risk Measures," Statistics and Risk Modeling, 30, 133–167.
- CONT, R., R. DEGUEST, AND G. SCANDOLO (2010): "Robustness and Sensitivity Analysis of Risk Measurement Procedures," *Quantitative Finance*, 10, 593–606.
- DEO, A. AND K. MURTHY (2020): "Optimizing Tail Risks Using an Importance Sampling Based Extrapolation for Heavy-Tailed Objectives," in 2020 59th IEEE Conference on Decision and Control (CDC), 1070–1077.
- DOWD, K. AND D. BLAKE (2006): "After VaR: The Theory, Estimation, and Insurance Applications of Quantile-Based Risk," *Journal of Risk and Insurance*, 73, 193–229.
- DU, Z. AND J. C. ESCANCIANO (2017): "Backtesting Expected Shortfall: Accounting for Tail Risk," *Management Science*, 63, 940–958.
- DU, Z., P. PEI, X. WANG, AND T. YANG (2024): "Powerful Backtests for Historical Simulation Expected Shortfall Models," *Journal of Business and Economic Statistics*, 42, 864–874.
- EL GHAOUI, L., M. OKS, AND F. OUSTRY (2003): "Worst-Case Value-At-Risk and Robust Portfolio Optimization: A Conic Programming Approach," *Operations Research*, 51, 543–556.
- ELSINGER, H., A. LEHAR, AND M. SUMMER (2006): "Risk Assessment for Banking Systems," *Management Science*, 52, 1301–1314.

- EMBRECHTS, P., C. KLÜPPELBERG, AND T. MIKOSCH (1997): Modelling Extremal Events for Insurance and Finance, Berlin: Springer.
- EMBRECHTS, P., B. WANG, AND R. WANG (2015): "Aggregation-Robustness and Model Uncertainty of Regulatory Risk Measures," *Finance and Stochastics*, 19, 763–790.
- EMBRECHTS, P. AND R. WANG (2015): "Seven Proofs for the Subadditivity of Expected Shortfall," *Dependence Modeling*, 3, 126–140.
- ENGLE, R. F. AND A. J. PATTON (2007): "What good is a volatility model?" in Forecasting Volatility in the Financial Markets, ed. by J. Knight and S. Satchell, Oxford: Butterworth-Heinemann, Quantitative Finance, chap. 2, 47–63, third ed.
- FAN, J. AND J. GU (2003): "Semiparametric Estimation of Value at Risk," *Econo*metrics Journal, 6, 261–290.
- FUH, C.-D., I. HU, Y.-H. HSU, AND R.-H. WANG (2011): "Efficient Simulation of Value at Risk with Heavy-Tailed Risk Factors," *Operations Research*, 59, 1395– 1406.
- GAO, F. AND F. SONG (2008): "Estimation Risk in GARCH VaR and ES Estimates," *Econometric Theory*, 24, 1404–1424.
- GARCIA, R., E. RENAULT, AND G. TSAFACK (2007): "Proper Conditioning for Coherent VaR in Portfolio Management," *Management Science*, 53, 483–494.
- GÂRLEANU, N. AND L. H. PEDERSEN (2007): "Liquidity and Risk Management," American Economic Review, 97, 193–197.
- GLASSERMAN, P., P. HEIDELBERGER, AND P. SHAHABUDDIN (2000): "Variance Reduction Techniques for Estimating Value-at-Risk," *Management Science*, 46, 1349–1364.
- GOOVAERTS, M. J., R. KAAS, J. DHAENE, AND Q. TANG (2003): "A Unified Approach to Generate Risk Measures," *ASTIN Bulletin*, 33, 173–191.
- GOURIEROUX, C., J.-P. LAURENT, AND O. SCAILLET (2000): "Sensitivity Analysis of Values at Risk," *Journal of Empirical Finance*, 7, 225–245.

- HE, X. D., S. KOU, AND X. PENG (2022): "Risk Measures: Robustness, Elicitability, and Backtesting," Annual Review of Statistics and Its Application, 9, 141–166.
- HE, X. D. AND X. PENG (2018): "Surplus-Invariant, Law-Invariant, and Conic Acceptance Sets Must Be the Sets Induced by Value at Risk," *Operations Research*, 66, 1268–1275.
- HENDRICKS, D. (1996): "Evaluation of Value-at-Risk Models Using Historical Data," FRBNY Economic Policy Review, 2, 39–70.
- HENDRICKS, D. AND B. HIRTLE (1997): "Bank Capital Requirements for Market Risk: The Internal Models Approach," *FRBNY Economic Policy Review*, 1–12.
- HILL, J. B. (2015): "Expected Shortfall Estimation and Gaussian Inference for Infinite Variance Time Series," *Journal of Financial Econometrics*, 13, 1–44.
- HONG, L. J. AND G. LIU (2009): "Simulating Sensitivities of Conditional Value at Risk," *Management Science*, 55, 281–293.
- IBRAGIMOV, R. (2005): "New Majorization Theory In Economics And Martingale Convergence Results In Econometrics," Ph.D. thesis, Yale University.
- JIANG, W. AND S. KOU (2021): "Simulating Risk Measures via Asymptotic Expansions for Relative Errors," *Mathematical Finance*, 31, 907–942.
- JIN, X., M. C. FU, AND X. XIONG (2003): "Probabilistic Error Bounds for Simulation Quantile Estimators," *Management Science*, 14, 230–246.
- JORION, P. (1996): "Risk²: Measuring the Risk in Value at Risk," *Financial Analysts Journal*, 52, 47–56.
- KAJI, T. (2018): "Essays on Asymptotic Methods in Econometrics," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.
- KATO, K. (2012): "Weighted Nadaraya–Watson Estimation of Conditional Expected Shortfall," Journal of Financial Econometrics, 10, 265–291.
- KOU, S. AND X. PENG (2016): "On the Measurement of Economic Tail Risk," *Operations Research*, 64, 1056–1072.

- KOU, S., X. PENG, AND C. C. HEYDE (2013): "External Risk Measures and Basel Accords," *Mathematics of Operations Research*, 38, 393–417.
- LEIPPOLD, M. AND P. VANINI (2002): "Half as Many Cheers—The Multiplier Reviewed," Wilmott Magazine, 2, 104–107.
- LESNEVSKI, V., B. L. NELSON, AND J. STAUM (2007): "Simulation of Coherent Risk Measures Based on Generalized Scenarios," *Management Science*, 53, 1756– 1769.
- LI, J. Y.-M. (2018): "Closed-Form Solutions for Worst-Case Law Invariant Risk Measures with Application to Robust Portfolio Optimization," Operations Research, 66, 1533–1541.
- LIMA, A. E. B., J. G. SHANTHIKUMARB, AND G.-Y. VAHN (2011): "Conditional Value-at-Risk in Portfolio Optimization: Coherent but Fragile," *Operations Re*search Letters, 39, 163–171.
- LINTON, O. AND Z. XIAO (2013): "Estimation of and Inference about the Expected Shortfall for Time Series with Infinite Variance," *Econometric Theory*, 29, 771–807.
- LIU, F. AND R. WANG (2021): "A Theory for Measures of Tail Risk," *Mathematics* of Operations Research, 46, 1109–1128.
- LOPEZ, J. A. (1998): "Regulatory Evaluation of Value-at-Risk Models," Journal of Risk, 1, 37–63.
- MARTINS-FILHO, C., F. YAO, AND M. TORERO (2018): "Nonparametric Estimation of Conditional Value-at-Risk and Expected Shortfall based on Extreme Value Theory," *Econometric Theory*, 34, 23–67.
- MCNEIL, A. J., R. FREY, AND P. EMBRECHTS (2005): Quantitative Risk Management: Concepts, Techniques and Tools, Princeton: Princeton University Press.
- MICHAUD, R. AND R. MICHAUD (2008): "Estimation Error and Portfolio Optimization: A Resampling Solution," *Journal of Investment Management*, 6, 8–28.
- MICHAUD, R. O. (1989): "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?" *Financial Analysts Journal*, 45, 31–42.

- MIN, S., C. C. MOALLEMI, AND C. MAGLARAS (2022): "Risk-Sensitive Optimal Execution via a Conditional Value-at-Risk Objective," ArXiv:2201.11962.
- NATARAJAN, K., D. PACHAMANOVA, AND M. SIM (2008): "Incorporating Asymmetric Distributional Information in Robust Value-at-Risk Optimization," *Man*agement Science, 54, 573–585.
- (2009): "Constructing Risk Measures from Uncertainty Sets," *Operations Research*, 57, 1129–1141.
- NEWEY, W. K. AND J. L. POWELL (1987): "Asymmetric Least Squares Estimation and Testing," *Econometrica*, 55, 819–847.
- NOLDE, N. AND C. ZHOU (2021): "Extreme Value Analysis for Financial Risk Management," Annual Review of Statistics and Its Application, 8, 217–240.
- NOLDE, N. AND J. F. ZIEGEL (2017): "Elicitability and Backtesting: Perspectives for Banking Regulation," Annals of Applied Statistics, 11, 1833–1874.
- NOVAK, S. Y. (2010): "A Remark Concerning Value-at-Risk," International Journal of Theoretical and Applied Finance, 13, 507–515.
- PRITSKER, M. (1997): "Evaluating Value at Risk Methodologies: Accuracy versus Computational Time," *Journal of Financial Services Research*, 12, 201–242.
- (2006): "The Hidden Dangers of Historical Simulation," Journal of Banking and Finance, 30, 561–582.
- ROY, A. D. (1952): "Safety First and the Holding of Assets," *Econometrica*, 20, 431–449.
- SCAILLET, O. (2004): "Nonparametric Estimation and Sensitivity Analysis of Expected Shortfall," *Mathematical Finance*, 14, 115–129.
- STAHL, G. (1997): "Three Cheers," Risk, 10, 67–69.
- THIELE, S. (2019): "Detecting Underestimates of Risk in VaR Models," *Journal of Banking and Finance*, 101, 12–20.
- TSAFACK, G. AND J. CATALDO (2021): "Backtesting and Estimation Error: Valueat-Risk Overviolation Rate," *Empirical Economics*, 61, 1351–1396.

- VALDEZ, E. A. (2005): "Tail Conditional Variance for Elliptically Contoured Distributions," *Belgian Actuarial Bulletin*, 5, 26–36.
- WANG, R. AND R. ZITIKIS (2021): "An Axiomatic Foundation for the Expected Shortfall," *Management Science*, 67, 1413–1429.
- WEISSMAN, I. (1978): "Estimation of Parameters and Large Quantiles Based on the k Largest Observations," Journal of the American Statistical Association, 73, 812–815.
- YAARI, M. E. (1987): "The Dual Theory of Choice under Risk," *Econometrica*, 55, 95–115.
- ZHU, S. AND M. FUKUSHIMA (2009): "Worst-Case Conditional Value-at-Risk with Application to Robust Portfolio Management," *Operations Research*, 57, 1155– 1168.