Routledge
Taylor & Francis Group

Stress scenario selection by empirical likelihood

PAUL GLASSERMAN†, CHULMIN KANG‡ and WANMO KANG*§

†Columbia Business School, Columbia University, New York, NY, USA
‡National Institute for Mathematical Sciences, Daejeon, Republic of Korea
§Department of Mathematical Sciences, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

This paper develops a method for selecting and analysing stress scenarios for financial risk assessment, with particular emphasis on identifying sensible combinations of stresses to multiple factors. We focus primarily on reverse stress testing – finding the most likely scenarios leading to losses exceeding a given threshold. We approach this problem using a nonparametric empirical likelihood estimator of the conditional mean of the underlying market factors given large losses. We then scale confidence regions for the conditional mean by a coefficient that depends on the tails of the market factors to estimate the most likely loss scenarios. We provide rigorous justification for the confidence regions and the scaling procedure when the joint distribution of the market factors and portfolio loss is elliptically contoured. We explicitly characterize the impact of the heaviness of the tails of the distribution, contrasting a broad spectrum of cases including exponential tails and regularly varying tails. The key to this analysis lies in the asymptotics of the conditional variances and covariances in extremes. These results also lead to asymptotics for marginal expected shortfall and the corresponding variance, conditional on a market stress; we combine these results with empirical likelihood significance tests of systemic risk rankings based on marginal expected shortfall in stress scenarios.

1. Introduction

Stress testing has long been part of the risk management toolkit, but it has gained new prominence through the recent financial crisis. This is reflected, for example, in the impact of the Supervisory Capital Assessment Program conducted by US financial regulators in 2009 (Hirtle *et al.* 2009), the annual Comprehensive Capital Assessment Reviews led by the Federal Reserve since 2010 (see Board of Governors of the Federal Reserve System 2012), the corresponding stress tests undertaken by the European Banking Authority, the stress testing requirements in the Dodd-Frank Act, and greater use of stress testing for internal risk management reported in industry surveys (Hida 2011, Moody's 2011).

Stress testing seeks to evaluate losses in extreme yet plausible scenarios that may be underweighted in a probabilistic model of market movements and absent from a historical backtest. An important challenge in designing effective stress tests lies in selecting scenarios that are indeed both sufficiently extreme and sufficiently plausible to improve risk management. Recent work on stress testing methodology includes Alfaro and Drehmann (2009), Breuer *et al.* (2009), Financial Services Authority (2009), Flood and Korenko (2012), Kopeliovich *et al.* (2013), Koyluoglu (2006),

Quagliarello (2009), and Rebonato (2010). Borio *et al.* (2012) provide a critical review of current practice.

Our objective in this article is to develop a data-driven procedure to inform the selection of scenarios that are both extreme and plausible. Our primary focus is on *reverse* stress testing, which seeks to identify scenarios that result in losses exceeding a given magnitude for a particular portfolio or firm. Because many different combinations of movements of market factors can produce losses of similar magnitude, we formulate the goal of reverse stress testing more precisely as one of identifying the most likely scenario or scenarios among all such combinations. These scenarios are, by definition, of primary importance to a particular portfolio, whereas purely hypothetical scenarios often seem arbitrary, making their consequences difficult to interpret. With a single risk factor, it may be relatively clear in which direction and even by how much to stress the factor to get a plausible adverse outcome, but identifying a sensible *combination* of stresses to multiple factors requires further analysis. This is one of the main challenges in defining stress scenarios.

We view the selection of stress scenarios as an exploratory process. Reliance on a single scenario – even the most likely one – is potentially misleading, so our objective is to identify important regions of stress scenarios, where importance reflects both the likelihood of the outcome and the severity of

---

*Corresponding author. Email: wanmo.kang@kaist.edu

the resulting loss. The contours of these regions provide sets of extreme stress scenarios that are, in a sense, equally plausible.

We approach the problem of identifying reverse stress testing regions in two steps. First, we estimate confidence regions for the conditional mean of the underlying market factors given a portfolio loss exceeding a specified level. Then we scale the conditional mean and the confidence regions by a multiplier that depends on the tails of the market factors to correct for the ratio of the conditional mean to the most likely loss scenario.

For the first step, the estimation of the conditional mean, we use an empirical likelihood estimator, in the sense of Owen (2001). Empirical likelihood (EL) is a nonparametric estimation procedure through which we get confidence regions for the conditional mean. Importantly, the EL estimator does not rely on significant assumptions about the conditional distribution of the market factors in extremes. The shape of the resulting confidence regions is able to capture skewness and other features present in extreme outcomes.

For the second step in our procedure – scaling the conditional mean – we derive asymptotically exact scaling multipliers (Proposition 1) when the joint distribution of market factors and portfolio value falls within a broad family of elliptically contoured distributions. This family includes, among many other examples, the multivariate normal, multivariate Laplace, and multivariate *t* distributions. Indeed, these three examples illustrate the range of qualitatively different tail behaviour captured within our framework – the normal distribution has very light tails, the Laplace distribution has tails that decay exponentially, and the *t* distribution has regularly varying tails. Relatively heavy tails are an important feature of market data, particularly in considering extreme scenarios. We provide rigorous justification for the scaling factor we derive, which reflects the heaviness of the tails, and for the combination of the scaling factor and the EL estimator; this combination yields asymptotically valid confidence regions for the most likely scenario leading to losses exceeding a given magnitude. We illustrate the results on market data.

The theoretical support for our method relies on a large loss level and a large sample size. In practice, data on large losses are limited, and this presents an obstacle to any attempt at rigorous statistical analysis without strong distributional assumptions. This difficulty is intrinsic to stress testing. By generating confidence regions around stress scenarios, our approach provides information about uncertainty in extremes, which is generally absent from the stress scenario selection process.

As part of our analysis, we derive results for the conditional variances and covariances of the underlying market factors given an extreme move by one factor. These results connect the tail behaviour of market factors with conditional moments in extreme scenarios.

Using these limits, we derive asymptotics for marginal expected shortfall (MES) and a corresponding marginal variance of shortfall (MVS), which are conditional moments in a stress scenario. An MES is itself a stress-test measure – an expected loss given a stress event. We show explicitly how the MES and MVS are affected by the underlying tail behaviour; in particular, the MVS grows faster in the extremes with heavier tails. Large variance values suggest the potential for a high degree of variability in MES estimates. With this in mind, we apply EL confidence regions to test the significance of systemic risk rankings in Acharya *et al.* (2010). The tests suggest that the top 50 companies rank roughly equally, as measured by MES, and that the difference between this group and the 100th ranked company is highly significant.

We comment briefly on some other relevant work. In addition to the papers cited previously on stress testing methodology, recent research addresses foundational questions in stress testing, particularly Goldstein and Sapra (2012), Pritsker (2011), and Schuermann (2012). Acharya *et al.* (2013) use marginal shortfall measures to evaluate regulatory stress tests. Baysal and Staum (2008) apply empirical likelihood estimation for expected shortfall and value-at-risk and find that it has the highest coverage among the methods they compare. Their setting considers confidence regions for the outputs of risk measurement whereas our concern is with the inputs in the form of most likely scenarios. Peng *et al.* (2012) extend the approach of Baysal and Staum (2008).

The rest of this paper is organized as follows. Section 2 motivates and formulates the reverse stress testing problem. Section 3 introduces empirical likelihood estimation, and section 4 converts estimates of conditional means to estimates of most likely loss scenarios. Section 5 illustrates the method through equity and currency portfolios. Section 6 presents our results on conditional extreme moments and applies these to analyse marginal expected shortfall and the corresponding variance. Proofs are deferred to an appendix.

2. The reverse stress testing problem

In an ordinary stress test, one posits a stress scenario and then evaluates the losses suffered by a given portfolio in that scenario. This approach offers the potential to uncover vulnerabilities that might be missed by other measures of risk. However, the results of a stress test can be difficult to interpret: If the losses in a stress scenario are large, should the portfolio be changed or is the scenario too extreme to be of concern? This inevitable question cannot be addressed without a view on the likelihood of the underlying scenario.

A reverse stress test starts by positing an adverse outcome – typically a loss of a given magnitude – rather than a scenario and then asks what scenarios would lead to that outcome. For a portfolio exposed to multiple risk factors, many different combinations of stresses might result in similar losses, so interest centres on the most likely combinations. The most likely scenario (for a given outcome) is portfolio-specific and, by construction, directly relevant to the portfolio in question in a way that a standard stress test may not be. A reverse stress test might reveal, for example, that while a portfolio is adequately hedged against movements in individual market factors, it remains exposed to a plausible combination of shocks across multiple market factors. By identifying specific vulnerabilities, a reverse stress test points to steps that should be taken to reduce these vulnerabilities. Focusing on the most likely adverse scenarios identifies the most important risks and removes much of the arbitrariness of a standard stress test based on a hypothetical scenario.

Our objective is thus to estimate most likely loss scenarios. Because information in extremes is necessarily limited, we are also interested in estimating confidence regions around these scenarios. Confidence regions measure uncertainty around the estimated scenarios and suggest important regions of factor stresses that go beyond a single combination of stresses.

To formulate these ideas precisely, let $\mathbf{Z}$ be a random d-dimensional vector representing the changes in market factors relevant to a portfolio – rates, prices, and economic variables. (We use boldface letters for vectors and matrices.) Suppose $\mathbf{Z}$ has a probability density f on $\mathbb{R}^d$. For a given portfolio exposed to these market factors, let $(\mathbf{Z}, L)$ have the joint distribution of the factors and the portfolio loss L. Write $f(\mathbf{z}|L \geq \ell)$ for the conditional density of $\mathbf{Z}$ given $L \geq \ell$, assuming it exists. The generic problem of reverse stress testing, for a loss threshold ℓ, is to find the most likely scenario (or scenarios) given a loss greater than or equal to ℓ; in other words, to solve

$$\text{(RST)} \qquad \mathbf{z}^*(\ell) = \underset{\mathbf{z} \in \mathbb{R}^d}{\operatorname{argmax}}\, f(\mathbf{z}|L \geq \ell).$$

We refer to a solution $\mathbf{z}^*(\ell)$ of this problem as a most likely loss scenario or as a solution to the reverse stress test. This is called the 'design point' in De and Tamarchenko (2002) and Koyluoglu (2006), based on an analogy with structural reliability.

3. Empirical likelihood estimation of the conditional mean

To reduce reliance on specific distributional assumptions, we adopt a nonparametric approach. Our objective remains to find the solution to (RST), but as an intermediate step we first focus on estimating $\mathbb{E}[\mathbf{Z}|L \geq \ell]$, the conditional mean of the factors given a large loss. Ordinarily, the conditional mean overestimates the most likely loss scenario. To offset this effect, we will derive a scaling correction based on the tail decay of $\mathbf{Z}$.

But first we need to estimate the conditional mean. We assume we have observations $(\mathbf{z}_i, L_i)$, $i = 1, 2, \ldots$, of past scenarios $\mathbf{z}_i$ and corresponding losses L_i. From these, we discard all observations except those for which the loss is at least ℓ. Through appropriate re-indexing, we are left with n observations $(\mathbf{z}_1, L_1), \ldots, (\mathbf{z}_n, L_n)$, all of which have $L_i \geq \ell$.

Once we have culled those observations for which $L_i \geq \ell$, the original problem of estimating a conditional mean reduces to one of estimating an unconditional mean. For this problem, we apply Owen's (2001) empirical likelihood (EL) method. This method considers convex combinations of the observations as candidate estimates of the mean:

$$w_1\mathbf{z}_1 + w_2\mathbf{z}_2 + \cdots + w_n\mathbf{z}_n,\ \sum_{i=1}^{n} w_i = 1,\ w_i \geq 0,\ i = 1, \ldots, n,$$

The profile empirical likelihood associated with a candidate value $\mathbf{x}$ is

$$\mathcal{R}(\mathbf{x}) = \max \left\{ \prod_{i=1}^{n} n w_i : \sum_{i=1}^{n} w_i \mathbf{z}_i = \mathbf{x}, \right.$$
$$\left. \sum_{i=1}^{n} w_i = 1, w_i \geq 0, i = 1, \ldots, n \right\}. \quad (1)$$

The product inside the braces is the likelihood ratio of the probability vector $(w_1, \ldots, w_n)$ to the uniform distribution $(1/n, \ldots, 1/n)$; $\mathcal{R}(\mathbf{x})$ is larger when $\mathbf{x}$ is a more uniform convex combination of the weights on the observations, and it is maximized at the sample mean of the observations.

Suppose that the observations are i.i.d. with mean $\boldsymbol{\mu}_0$ ($= E[\mathbf{Z}|L \geq \ell]$), and suppose that the convex hull of the observations contains $\boldsymbol{\mu}_0$ with probability approaching 1 as the number of observations increases. Then Owen's (2001) Theorem 3.2 states that $-2\log \mathcal{R}(\boldsymbol{\mu}_0)$ has an asymptotic χ_d^2 distribution for large n. This provides the basis for EL confidence regions: Fix a confidence level $1 - \alpha$ and find the quantile x_α for which $\mathbb{P}(\chi_d^2 \geq x_\alpha) = \alpha$; the corresponding $1 - \alpha$ confidence region for $\boldsymbol{\mu}_0$ is the set

$$\mathcal{C}_{1-\alpha,n} = \left\{ \sum_{i=1}^{n} w_i \mathbf{z}_i : \prod_{i=1}^{n} n w_i \geq \exp(-x_\alpha/2), \right.$$
$$\left. \sum_{i=1}^{n} w_i = 1, w_i \geq 0, i = 1, \ldots, n \right\}. \quad (2)$$

As discussed in Owen (2001), the maximization problem defining the profile empirical likelihood is easy to solve by first reformulating it as

$$\max_{w_1, \ldots, w_n} \sum_{i=1}^{n} \log w_i \quad \text{subject to} \quad \sum_{i=1}^{n} w_i = 1,\ \sum_{i=1}^{n} w_i \mathbf{z}_i = \mathbf{x}.$$

The resulting confidence regions are appealing, if n is not too small, because they make minimal assumptions about the distribution of the underlying data and are able to capture skewness and other notable shape characteristics in the data.

4. From conditional mean to most likely loss scenario

4.1. Multivariate models: elliptically contoured distributions

Recall that our objective is to estimate the solution $\mathbf{z}^*(\ell)$ to the reverse stress testing problem (RST), and in the previous section we have estimated a conditional mean $\mathbb{E}[\mathbf{Z}|L \geq \ell]$, which we denote by $\bar{\mathbf{z}}(\ell)$. The next step is therefore to relate these quantities. We will do so under the assumption that the loss level ℓ is large and that the joint distribution of the market factors and the portfolio loss is elliptically contoured.

In more detail, a random vector $\mathbf{X}$ in $\mathbb{R}^d$ has a spherical distribution if it admits a representation

$$\mathbf{X} = R\mathbf{S}, \quad (3)$$

in which $\mathbf{S}$ is uniformly distributed on the $(d-1)$-dimensional unit sphere and R is a nonnegative random variable independent of $\mathbf{S}$. (This is one of several equivalent definitions; see p. 31 of Fang *et al.* (1990).) We have $R = \sqrt{\mathbf{X}^\top \mathbf{X}}$ in (3), so the distribution of R is uniquely determined by that of $\mathbf{X}$. A random vector $\mathbf{Y}$ has an elliptically contoured distribution if it admits a representation (p. 31 of Fang *et al.* (1990))

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{X}, \quad (4)$$

in which the vector $\boldsymbol{\mu}$ and matrix $\mathbf{A}$ are constants and $\mathbf{X}$ has a spherical distribution. The distribution of $\mathbf{Y}$ is centred at $\boldsymbol{\mu}$, and if the components of $\mathbf{Y}$ have finite second moments then $\mathbf{Y}$ has covariance matrix $\boldsymbol{\Sigma} = w\mathbf{A}\mathbf{A}^\top$, where $w = \mathbb{E}[R^2]/d$. We will

assume that $\mathbf{A}$ is $d \times d$ so $\mathbf{Y}$ has the same dimension as $\mathbf{X}$. If $\mathbf{A}$ is invertible, then $\mathbf{\Sigma} > 0$ (meaning that it is positive definite). If R has a density and $\mathbf{\Sigma} > 0$, then $\mathbf{Y}$ has a density, and this density is constant on the ellipses $\{\mathbf{y} \in \mathbb{R}^d : \mathbf{y}^\top \mathbf{\Sigma}^{-1} \mathbf{y} = a\}$, $a > 0$; see p. 35 of Fang *et al.* (1990).

The multivariate normal distribution $N(0, \mathbf{\Sigma})$ is clearly elliptically contoured and corresponds to taking $R^2 \sim \chi_d^2$, the chi-square distribution with d degrees of freedom. More generally, (4) includes all mixtures of normals of the form

$$\mathbf{Y} = \boldsymbol{\mu} + \sqrt{W} N(0, \mathbf{\Sigma}), \qquad (5)$$

with W an independent mixing random variable. For example, taking W exponentially distributed yields the multivariate Laplace distribution, and taking $W = \nu/\chi_\nu^2$ yields the multivariate t distribution with ν degrees of freedom.

We distinguish different categories of tail behaviour among elliptically contoured distributions based on the concept of regular variation, discussed in greater detail in the appendix and in, for example, Resnick (2007). Loosely speaking, a random variable is regularly varying with index $\nu > 0$ if its tail distribution decays like the power $x^{-\nu}$. A smaller index ν thus indicates a heavier tail.

As shown in the appendix, linear combinations of the components of $\mathbf{Y}$ or $\mathbf{X}$ are regularly varying with index ν if the radial random variable R has this property, in which case we will refer to $\mathbf{Y}$ or $\mathbf{X}$ as regularly varying with index ν. We use the notation $\mathsf{RV}(\nu)$ to refer to this class of distributions.

We will say that R is in the class $\mathsf{ERV}(\alpha, \nu)$, $\alpha, \nu > 0$, if $\exp(R^\alpha)$ is regularly varying with index ν. As discussed in the appendix, if $R \in \mathsf{ERV}(\alpha, \nu)$, then so is each coordinate of $\mathbf{X}$ in (3), and we write $\mathbf{X} \in \mathsf{ERV}(\alpha, \nu)$ for brevity. Among distributions in the classes $\mathsf{ERV}(\alpha, \nu)$, those with larger α have lighter tails, and among those with the same α, those with larger ν have lighter tails. The standard normal distribution has $\alpha = 2$ and $\nu = 1/2$; the Laplace distribution with $E[W] = 1/\lambda$ and $\mathbf{\Sigma} = I$ in (5) has $\alpha = 1$ and $\nu = \sqrt{2\lambda}$. These are both special cases of the d-dimensional spherical Kotz distribution (Fang *et al.* 1990, p. 69) with $\mathsf{ERV}(\alpha, \nu)$ density

$$f(\mathbf{x}) \propto \|\mathbf{x}\|^m \exp(-\nu \|\mathbf{x}\|^\alpha), \quad \nu, \alpha > 0, \ 2m + d > 0,$$

in which the roles of α and ν are evident. Another example is provided by the symmetric generalized hyperbolic distribution

$$f(\mathbf{x}) \propto \frac{K_{\lambda-d/2}(\sqrt{\psi(\chi + \|\mathbf{x}\|^2)})}{(\sqrt{\psi(\chi + \|\mathbf{x}\|^2)})^{d/2-\lambda}}, \quad \lambda \in \mathbb{R}, \ \psi, \chi > 0,$$

where K_r denotes a Bessel function of the third kind with parameter r. It can be shown that this distribution is $\mathsf{ERV}(1, \sqrt{\psi})$, and the boundary case $\psi = 0$, $\lambda < 0$ is $\mathsf{RV}(-2\lambda)$.

The categories of distributions $\mathsf{RV}(\nu)$ and $\mathsf{ERV}(\alpha, \nu)$ allow a great deal of flexibility in capturing tail behaviour of market factors, with the regularly varying case the most relevant in applications.† Among elliptically contoured distributions, these cases also have qualitatively different tail *dependence*, with the multivariate normal having no tail dependence (except when perfectly correlated), and the multivariate t exhibiting positive tail dependence even with negative correlation (except when perfectly negatively correlated); see Schmidt (2002). The mixture representation in (5) and the associated tail behaviour can

†However, see Heyde and Kou (2004) for a comparison of empirical estimates of t tails and Laplace tails.

be interpreted as the result of heteroskedasticity or stochastic volatility in a dynamic model.

For our theoretical results, we will assume that the joint distribution of $(\mathbf{Z}, L)$, the market factors and the portfolio loss has the representation in (4), and that the underlying spherical $\mathbf{X}$ belongs to one of the families $\mathsf{RV}(\nu)$ or $\mathsf{ERV}(\alpha, \nu)$. We do not require L to be a deterministic function of $\mathbf{Z}$. We may think of $\mathbf{Z}$ as recording the most important factors influencing the portfolio, and then the model assumes that the tail behaviour of the portfolio loss is consistent with that of the most important factors. We will always assume that the restriction of $\mathbf{\Sigma}$ to the $d \times d$ covariance matrix of $\mathbf{Z}$ is positive definite, so that none of factors is redundant. This is sufficient to ensure that the most likely loss scenario $\mathbf{z}^*(\ell)$ is well defined.

4.2. Estimation

We now turn to the problem of estimating the most likely loss scenario through the conditional mean, beginning with the following result.

PROPOSITION 1 *Suppose the distribution of $\mathbf{Y} = (\mathbf{Z}, L)$ is elliptical as in (4), with $\mathbf{X}$ either $\mathsf{ERV}(\alpha, \nu)$, for some $\alpha, \nu > 0$, or $\mathsf{RV}(\nu)$, with $\nu > 1$. Let $\mathbf{z}^*(\ell) \in \mathbb{R}^d$ be the most likely loss scenario and let $\bar{\mathbf{z}}(\ell) \in \mathbb{R}^d$ denote the conditional mean $\mathbb{E}[\mathbf{Z}|L \geq \ell]$. Then there exists a positive scalar sequence κ_ℓ such that*

$$\mathbf{z}^*(\ell) = \kappa_\ell \bar{\mathbf{z}}(\ell), \quad \text{and} \quad \kappa_\ell \to \kappa \ \text{as} \ \ell \to \infty, \qquad (6)$$

where

- $\kappa = 1$ *for all* $\mathsf{ERV}(\alpha, \nu)$ *distributions;*
- $\kappa = (\nu - 1)/\nu$ *for all* $\mathsf{RV}(\nu)$ *distributions,* $\nu > 1$.

Based on this result, we can estimate the most likely loss scenario $\mathbf{z}^*(\ell)$ by estimating the conditional mean $\bar{\mathbf{z}}(\ell)$ and then scaling the result as needed. In the lighter tailed cases, no scaling is needed; in the $\mathsf{RV}(\nu)$ case, we multiply the estimate of the conditional mean by $(\nu - 1)/\nu$ asymptotically to estimate the most likely loss scenario. Market data are often well approximated using a t_ν distribution with ν in the range of 5–7, corresponding to scale factors in the range of 0.80–0.86. In addition to scaling the point estimate, we would like, more importantly, to scale the confidence regions for $\bar{\mathbf{z}}(\ell)$ to get confidence regions for $\mathbf{z}^*(\ell)$. Such a procedure involves two limits, because Proposition 1 applies as $\ell \to \infty$ whereas the chi-square limit that underpins the EL method holds as the number of observations grows. For a combined result, we therefore need an array version of the EL limit theorem, building on Owen's (2001) Theorem 4.1.

In the following, we let $\mathbf{Z}_1(\ell), \mathbf{Z}_2(\ell), \dots, \mathbf{Z}_{n_\ell}(\ell)$ denote i.i.d. observations from the conditional distribution of $\mathbf{Z}$ given $L \geq \ell$, with $n_\ell \to \infty$. As before, let x_α be the quantile defined by $\mathbb{P}(\chi_d^2 \geq x_\alpha) = \alpha$. Write $\mathcal{R}_\ell(\mathbf{x})$ for the profile empirical likelihood in (1) with $n = n_\ell$. For a set $\mathcal{C} \subseteq \mathbb{R}^d$ and a constant κ, $\kappa\mathcal{C}$ denotes the set of points of the form $\kappa\mathbf{x}$ with $\mathbf{x} \in \mathcal{C}$.

THEOREM 1 *Suppose the distribution of $\mathbf{Y} = (\mathbf{Z}, L)$ is elliptical as in (4), with $\mathbf{X}$ either $\mathsf{ERV}(\alpha, \nu)$, for some $\alpha, \nu > 0$, or $\mathsf{RV}(\nu)$, with $\nu > 4$. Then*

$$-2 \log \mathcal{R}_\ell(\bar{\mathbf{z}}(\ell)) = -2 \log \mathcal{R}_\ell(\kappa_\ell^{-1} \mathbf{z}^*(\ell)) \rightarrow \chi_d^2$$

in distribution, and $\kappa_\ell \mathcal{C}_{1-\alpha, n_\ell}$ is an asymptotic $100(1-\alpha)\%$ confidence region for the most likely loss scenario $\mathbf{z}^*(\ell)$; i.e.

$$\mathbb{P}(\mathbf{z}^*(\ell) \in \kappa_\ell \mathcal{C}_{1-\alpha, n_\ell}) \rightarrow 1-\alpha,$$

as $\ell \rightarrow \infty$, where $\kappa_\ell \rightarrow \kappa$, with κ as in Proposition 1.

This result leads to the following procedure. As in section 3, we extract the large loss scenarios from the available data. Using these observations, we construct EL confidence regions (2) for the conditional mean $\bar{\mathbf{z}}(\ell)$. We then scale the confidence regions by the factor κ_ℓ to get confidence regions for the most likely loss scenario $\mathbf{z}^*(\ell)$. As a simplifying approximation, one could use the limiting value κ in place of κ_ℓ.

We have described this procedure through its application to historical data. The same approach could be used with simulated data. In some settings – stress testing an entire bank portfolio, for example – fully evaluating each scenario is extremely time-consuming. A simplified model could then be applied to simulated scenarios from which one would then estimate the most likely loss scenarios for a more extensive evaluation. However, building a simulation model requires estimating parameters that would not be needed in a purely data-driven approach.

The EL procedure is nonparametric. The asymptotic scaling factor κ is 'lightly' parametric in the sense that it is determined by the tail decay of the factors. It should be noted that the procedure provided by Theorem 1 does not involve estimation of $\boldsymbol{\Sigma}$, which can be particularly difficult in high dimensions.

4.3. Coverage

Theorem 1 provides asymptotic support for confidence regions as the sample size and loss level increase. In practice, data on large losses are limited. Our procedure can still be implemented with limited number of observations; the drawback is that the actual coverage of the estimated confidence regions may differ from the nominal coverage.

To test the performance of the confidence regions at finite sample sizes and loss levels, we use simulation. We generate points from a multivariate t distribution with uncorrelated marginals. The loss is given by the linear function $\mathbf{c}^\top \mathbf{Z}$ with $\mathbf{c} = (1, 0, \ldots, 0)^\top$, so the most likely scenario producing a loss of ℓ is $\mathbf{z}^*(\ell) = (\ell, 0, \ldots, 0)^\top$. (Because of the symmetry of the distribution, this choice entails no loss of generality; see section A.1 or Fang *et al.* 1990, p. 31.) To test performance at sample size n, we generate enough points to get n observations for which the loss is at least ℓ; we then construct the confidence region, scaled by κ_ℓ, and check if it contains $\mathbf{z}^*(\ell)$. We repeat this 1000 times and record the percentage of times the confidence region contains $\mathbf{z}^*(\ell)$ as the estimated coverage.

Table 1 shows the results at degrees of freedom $\nu = 5, 6$ and 7; dimensions $d = 2, 5$ and 10; sample sizes $n = 10, 50$ and 500; and loss levels at the 95th, 99th and 99.9th percentile of the t_ν distribution. The top half of the table uses a confidence level of 95%, and the bottom half uses 50%. We need at least $d + 1$ points in dimension d to get a confidence region with nonzero volume, so the entries with $n = d = 10$ are blank. As expected, the observed coverage approaches the nominal

coverage as the sample size increases. (We have included the case $n = 500$ to illustrate the convergence; in practice, one is unlikely to have this many observations in the extremes.) The most significant shortfalls in coverage occur in high dimensions with few points. The coverage is not very sensitive to the loss level ℓ.

5. Application to equity and currency scenarios

In this section, we provide simple examples to illustrate our method. We consider an equity portfolio and a currency portfolio.

5.1. An equity portfolio

For our first application, we consider a portfolio of world equity indices: the S&P 500, FTSE, DAX, Nikkei 225, Hang Seng and Bovespa. We consider weekly returns from 3 May, 1993 to 26 December, 2011, and monthly returns from 1 June, 1993 to 1 December, 2011. We select portfolio weights based on the market capitalization traded on each exchange, as listed in table 2. This gives us a linear loss $L = \mathbf{c}^\top \mathbf{Z}$, with $\mathbf{c} = [-0.5050; -0.1362; -0.0539; -0.1443; -0.1022; -0.0583]$. For purposes of illustration, we choose ℓ to correspond to a 1% loss level with weekly data and a 5% loss level with monthly data.

Before proceeding with the application of our method, it is worth pausing to ask how existing methods would design a stress test for this portfolio. To the best of our knowledge, current practice offers no standard approach, even in this rather simple setting. One might envision generating a stress scenario by, for example, shocking each index by some number of standard deviations. Such an approach would ignore the nature of the dependence between indices in extremes, and it would ignore the specific construction of the portfolio in question.

To apply our method, we model the returns on the equity indices using a multivariate t distribution. The density with parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu$ is given by

$$f(\mathbf{x}|\mu, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma(\frac{1}{2}(\nu + d))}{\Gamma(\frac{1}{2}\nu)(\pi \nu)^{d/2} |\boldsymbol{\Sigma}|^{1/2}}$$
$$\times \left(1 + \frac{(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})}{\nu}\right)^{-(\nu+d)/2},$$
$$\text{for} \quad \mathbf{x} \in \mathbb{R}^d.$$

The mean and variance of the distribution are given by

$$\mathbb{E}[X] = \boldsymbol{\mu}, \qquad \mathbb{V}(X) = \frac{\nu}{\nu - 2} \boldsymbol{\Sigma},$$

assuming $\nu > 2$. To estimate ν, we first estimate the sample mean and covariance and then maximize the likelihood over ν. We get $\hat{\nu} = 5.0$ with weekly data and $\hat{\nu} = 5.8$ with monthly data. We have used the assumption of a t distribution solely to estimate the tail index ν; one might alternatively estimate this index directly without introducing a specific distribution.

The results are illustrated in figures 1–4. As one might expect for a portfolio of long positions in the equity indices, the most likely loss scenario has all indices declining. Less obviously,

Table 1. Estimated coverage of the most likely loss scenario for dimension d, sample size $n \geq d + 1$, loss level ℓ and degrees of freedom v at confidence levels of 95 and 50%.

95% confidence	$v = 5$			$v = 6$			$v = 7$		
	$n = 10$	50	500	10	50	500	10	50	500
$d = 2, \quad \ell = F_v^{-1}(.95)$	73.4	90.0	94.8	75.7	93.1	93.6	74.8	91.4	95.0
$F_v^{-1}(.99)$	71.6	90.2	95.7	74.6	92.3	95.7	75.6	91.8	94.0
$F_v^{-1}(.999)$	72.4	91.2	95.2	72.3	92.4	96.3	77.6	93.4	94.2
$d = 5, \quad \ell = F_v^{-1}(.95)$	30.1	84.4	94.6	29.2	86.8	94.3	30.4	86.6	95.2
$F_v^{-1}(.99)$	26.3	85.9	94.2	28.6	89.1	93.7	28.5	87.2	93.9
$F_v^{-1}(.999)$	25.2	86.6	93.8	28.5	89.1	95.1	31.0	86.3	94.4
$d = 10, \quad \ell = F_v^{-1}(.95)$		69.6	93.8		71.8	94.2		73.6	93.6
$F_v^{-1}(.99)$		68.0	92.3		69.5	92.9		71.0	94.4
$F_v^{-1}(.999)$		68.2	94.1		73.5	93.8		74.3	94.0
50% confidence	$n = 10$	50	500	10	50	500	10	50	500
$d = 2, \quad \ell = F_v^{-1}(.95)$	35.4	45.0	48.0	35.8	47.8	48.8	36.8	47.0	46.4
$F_v^{-1}(.99)$	30.8	43.2	48.4	32.5	44.4	50.2	35.4	47.0	49.1
$F_v^{-1}(.999)$	33.6	45.4	51.5	35.9	46.2	50.4	35.4	50.6	52.0
$d = 5 \quad \ell = F_v^{-1}(.95)$	10.7	39.2	50.6	12.6	40.6	48.4	11.3	39.7	50.1
$F_v^{-1}(.99)$	9.4	37.8	46.5	10.7	40.8	46.6	11.5	41.8	51.2
$F_v^{-1}(.999)$	8.8	37.3	45.2	11.9	39.8	51.1	12.0	39.0	46.0
$d = 10, \quad \ell = F_v^{-1}(.95)$		22.7	46.5		26.3	51.4		26.6	47.6
$F_v^{-1}(.99)$		23.1	44.9		24.3	48.0		25.7	49.6
$F_v^{-1}(.999)$		23.6	46.6		28.0	47.0		26.7	49.7

Table 2. Market caps of exchanges at 2010, in USD millions, from www.world-exchanges.org/statistics.

Exchange	Market cap	Proportion(%)
NYSE Euronext	13,394,081.8	50.50
London SE Group	3,613,064.0	13.62
Deutsche Börse	1,429,719.1	5.39
Tokyo SE Group	3,827,774.2	14.43
Hong Kong Exchanges	2,711,316.2	10.22
BM&FBOVESPA	1,545,565.7	5.83

the declines in the various indices in the most likely scenario are similar, despite the wide disparity in the portfolio weights.

The confidence regions provide a more nuanced picture. For purposes of illustration, we show confidence regions for pairs of indices at a time, though having an automated method is particularly valuable in multiple dimensions where visualization is difficult. Figures 1 and 2 show results for weekly data. The circles show the observations, and the crosses show the extreme observations – those beyond the loss threshold. The contours show 99 and 50% confidence regions for the conditional mean and (after scaling) for the most likely loss scenario. The confidence regions are clearly shaped by the data, yet tempered compared to the most extreme points.

In figure 1, for example, we see a marked difference in the comovements of two pairs of indices. The left panel indicates a high degree of confidence in the FTSE and S&P 500 declining together whereas the right panel shows a much weaker link between the Hang Seng and the Nikkei in extremes. Figures 3 and 4 suggest somewhat weaker dependence in extremes at monthly frequencies than at weekly frequencies. We interpret the confidence regions as sets of additional scenarios that merit

exploration along with the most likely scenario, so the shapes in the figures indicate further combinations of index moves to consider. Indeed, in Glasserman *et al.* (2013) we propose a sampling algorithm motivated by the EL regions to generate additional scenarios by simulation.

The confidence regions in the figures take v as fixed, but they could be modified to reflect uncertainty in this parameter. The mapping $v \mapsto (v - 1)/v$ is monotonic, so a confidence interval $(\hat{v}_1, \hat{v}_2)$ for v implies a confidence interval for the scaling factor; one might therefore replace a single confidence region $\kappa_v \mathcal{C}$ with the convex hull of confidence regions $\kappa_{\hat{v}_1} \mathcal{C}$ and $\kappa_{\hat{v}_2} \mathcal{C}$. A full justification of this procedure is beyond the scope of Theorem 1. As a practical matter, we consider uncertainty about v less significant than the uncertainty about combinations of factor moves leading to large losses. The survey in Haas and Pigorsch (2013) provides extensive references to the literature on tail index estimation for financial data, most of which finds indices falling in a fairly narrow range for equities and currencies.

In the example of this section and the next, we take the distribution of $\mathbf{Z}$ to be the distribution of i.i.d. market returns.
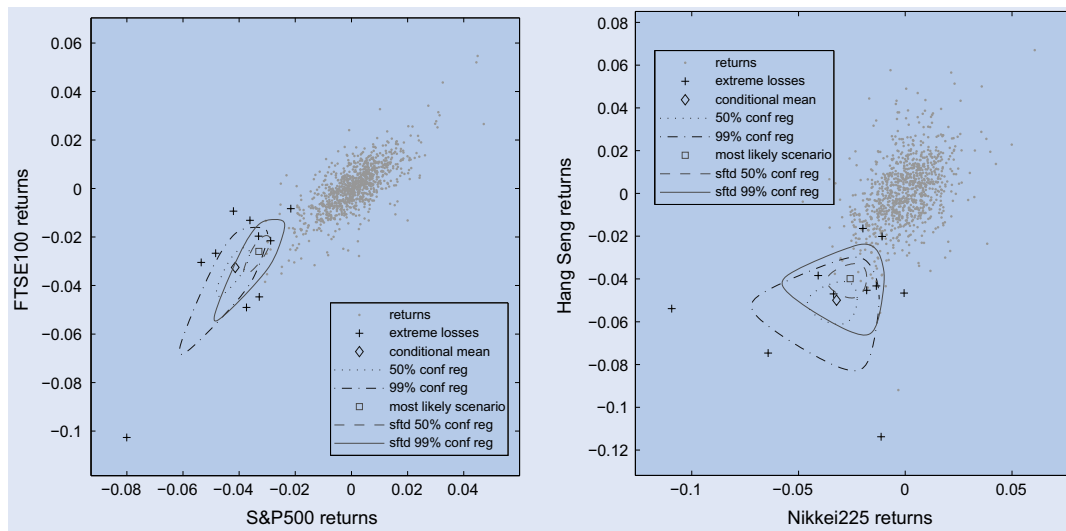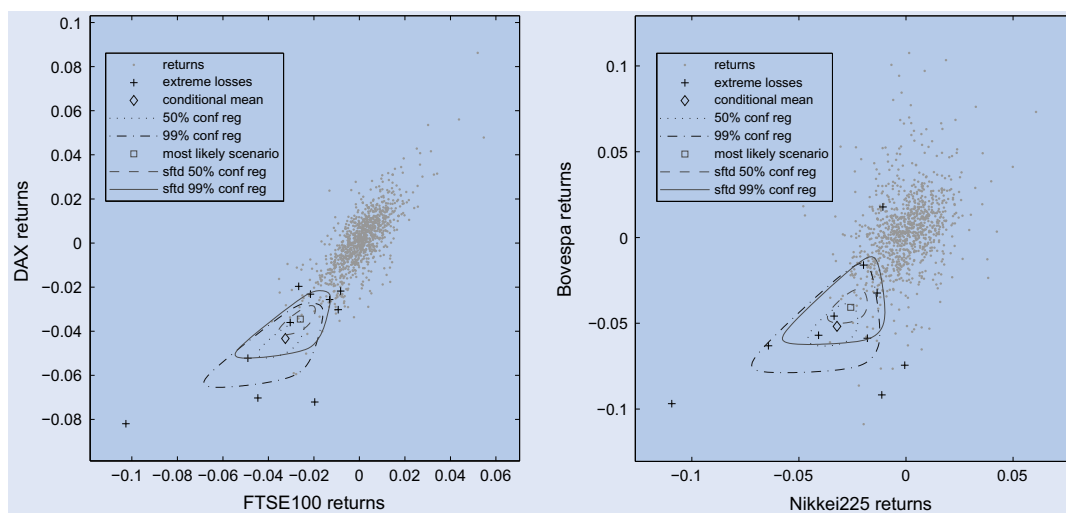
Figure 1.   Equity indices, weekly data.

Figure 2.   Equity indices, weekly data.

A more complex application is developed in Glasserman *et al.* (2013), based on the Federal Reserve's Comprehensive Capital Analysis and Review Board of Governors of the Federal Reserve System (2012), the supervisory stress test applied to the largest bank holding companies in the USA One of the extensions considered in Glasserman *et al.* (2013) is to treat **Z** as the distribution of innovations in a vector autoregressive model of financial and economic variables, thus allowing for serial dependence in these variables.

5.2. A currency portfolio

Next we consider a basket of currencies, half held in British pounds (GBP), the rest divided evenly among the Australian dollar (AUD), the euro (EUR), the Japanese yen (JPY) and the Swiss franc (CHF). We use monthly returns against the US dollar from February 2000 through December 2011. A maximum likelihood fit of the data to a multivariate t distribution

yields an estimate of $\hat{\nu} = 5.2$ to the degrees-of-freedom parameter. For the loss severity ℓ, we choose the loss threshold ℓ at the level of the worst 5% of losses in the sample period. Our estimated most likely loss scenario is

(AUD, EUR, JPY, CHF, GBP)
$= (-5.5907\%, -4.1142\%, 1.0402\%, -4.0246\%, -4.4338\%),$

the values on the right indicating one-month returns against the US dollar.

Figure 5 illustrates the results. The circles show the observations, and the crosses show the 5% most extreme observations – those beyond the loss threshold. The contours show 99 and 50% confidence regions for the conditional mean and (after scaling) for the most likely loss scenario. The squares indicate the point estimates.

In the left panel, we see that the confidence regions for the most likely loss reflect the skewness in the joint distribution of the EUR/USD and CHF/USD returns. The most likely loss
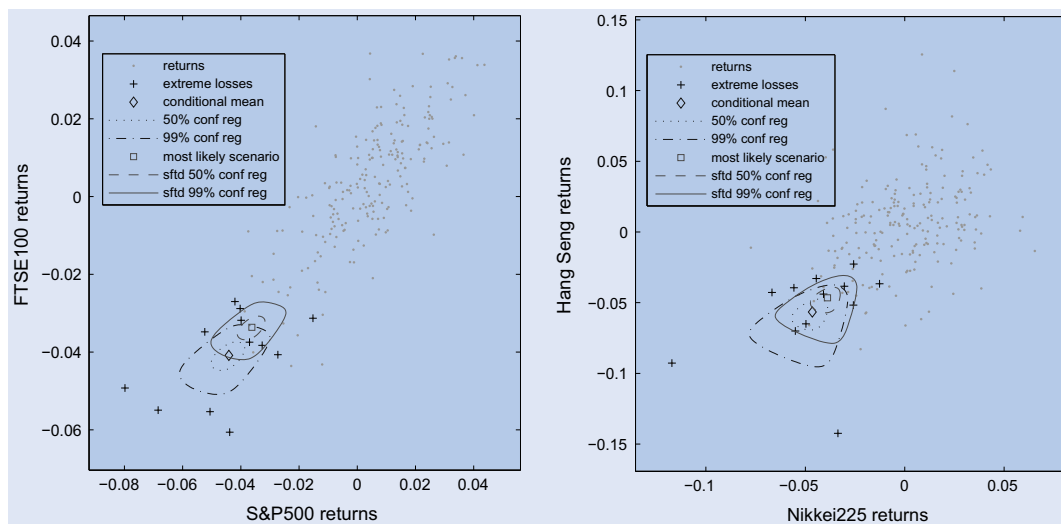
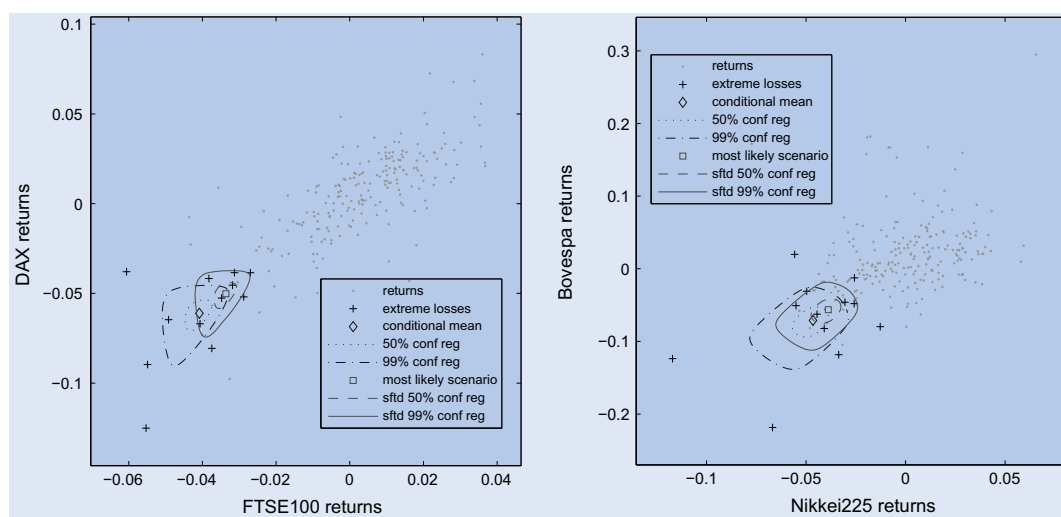Figure 3. Equity indices, monthly data.

Figure 4. Equity indices, monthly data.

scenario involves an increase in the JPY/USD rate, even though this increase would, by itself, generate a gain, not a loss, for the portfolio. This outcome is a nonobvious reflection of the joint distribution of the returns: the largest drops in the GBP (which makes up 50% of the portfolio) coincide with increases in the JPY/USD rate. However, the confidence regions in the right panel of figure 5 also indicate a wide range of outcomes for the JPY/USD rate when the GBP drops, suggesting that one should explore other scenarios in the large-loss region.

6. Conditional moments and marginal shortfall

Our approach to estimating the most likely loss scenario through the conditional mean is closely related to the problem of estimating marginal expected shortfall (MES), the conditional loss in a stress scenario. Indeed, the problem of estimating the conditional mean is formally equivalent to estimating MES.

MES was originally introduced as a mechanism for attributing a portfolio's overall loss to parts of the portfolio or to individual factors. (See, for example, the discussion in McNeil *et al.* (2005).) A subportfolio's MES is its expected loss conditional on the total portfolio loss exceeding some threshold. More recently, MES and MES-like measures have become part of the stress testing toolkit as measures of systemic risk; see, in particular, Acharya *et al.* (2010), Adrian and Brunnermeier (2009) and Huang *et al.* (2012). In this context, one looks at the expected loss suffered by a financial institution (rather than a subportfolio) conditional on a large shock to the system as a whole, just as in a conventional stress test. Indeed, Acharya *et al.* (2013) use their version of an MES measure to evaluate macroprudential stress tests and propose their method as an alternative to current regulatory stress testing. Huang *et al.* (2012) make a similar comparison between their method and an earlier round of bank stress tests. Oura and Schumacher (2012) include MES in their list of stress testing methodologies.
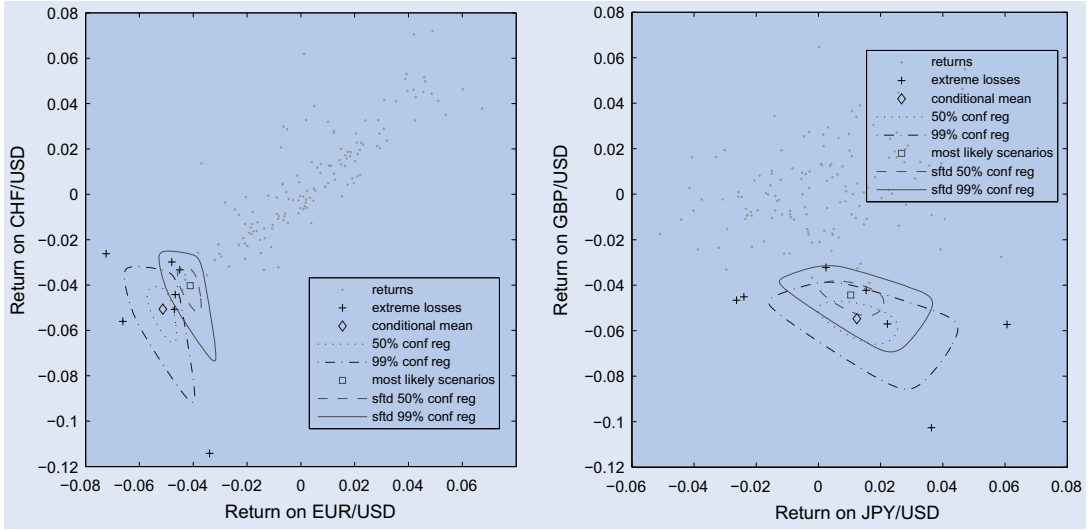
Figure 5.   Foreign currencies, monthly data.

Table 3.   Conditional means and variances of factors, given an extreme outcome of X_1.

Conditional moment	$\mathrm{ERV}(\alpha, \nu)$	$\mathrm{RV}(\nu), \nu > 2$
$\mathbb{E}[X_1 \mid X_1 \geq \ell]$	$\ell + o(\ell)$	$\frac{\nu}{\nu-1}\ell + o(\ell)$
$\mathbb{E}[X_1^2 \mid X_1 \geq \ell]$	$\ell^2 + o(\ell^2)$	$\frac{\nu}{\nu-2}\ell^2 + o(\ell^2)$
$\mathbb{V}(X_1 \mid X_1 \geq \ell)$	$\frac{1}{(\nu\alpha)^2}\ell^{-2(\alpha-1)} + o(\ell^{-2(\alpha-1)})$	$\frac{\nu}{(\nu-2)(\nu-1)^2}\ell^2 + o(\ell^2)$
$\mathbb{E}[X_2 \mid X_1 \geq \ell]$	0	0
$\mathbb{V}(X_2 \mid X_1 \geq \ell) = \mathbb{E}[X_2^2 \mid X_1 \geq \ell]$	$\frac{1}{(\nu\alpha)^2}\ell^{-(\alpha-2)} + o(\ell^{-(\alpha-2)})$	$\frac{\nu}{(\nu-2)(\nu-1)}\ell^2 + o(\ell^2)$

The analysis underlying Proposition 1 and Theorem 1 allows us to characterize how MES and a corresponding conditional variance behave across the families of distributions we consider. Moreover, the EL procedure provides a way to measure the precision of MES estimates used for systemic risk rankings.

6.1. Conditional moments in extremes

The key to this analysis (and to the proof of Theorem 1) is the calculation of conditional moments in extremes for the multivariate distributions we consider. In fact, it suffices (see the appendix) to consider a pair (X_1, X_2) with a spherical distribution. Table 3 summarizes the conditional means and variances of the factors, given an extreme outcome of one of the factors.

Moving from left to right in the table, we have heavier tails. As one might expect, all the conditional moments increase (for large ℓ) as we move from the ERV category to the RV category. Among the $\mathrm{ERV}(\alpha, \nu)$ distributions, the conditional moments are decreasing in α and ν, and among the $\mathrm{RV}(\nu)$ distributions they are decreasing in ν. In specific cases, more explicit results are sometimes available; for example, the last row is exactly 1 for the normal distribution, and all rows can be evaluated explicitly for the Laplace distribution. The results in the table suffice for our purposes, so we omit the details of specific cases.

6.2. Marginal shortfall

Using the asymptotic moments in table 3, we can analyse the MES conditional on a large loss in the portfolio. Let $\mathbf{Y}$ be elliptically contoured as in (4). We will condition on large values of $\mathbf{c}^\top \mathbf{Y}$ for a fixed vector $\mathbf{c}$. For each component Y_i of $\mathbf{Y}$, define the marginal expected shortfall and the corresponding variance by

$$MES_i = \mathbb{E}[Y_i \mid \mathbf{c}^\top \mathbf{Y} \geq \ell]$$
$$MVS_i = \mathbb{V}[Y_i \mid \mathbf{c}^\top \mathbf{Y} \geq \ell].$$

We analyse these quantities for large loss levels ℓ. The marginal shortfall contribution for the ith subportfolio or factor is c_i times the expression given here for MES_i.

To lighten notation, let $(\beta_1, \ldots, \beta_d) = \mathbf{c}^\top \mathbf{\Sigma} / (\mathbf{c}^\top \mathbf{\Sigma} \mathbf{c})$ and write, for each $i = 1, \ldots, d$,

$$Y_i = \mu_i + \beta_i \mathbf{c}^\top (\mathbf{Y} - \boldsymbol{\mu}) + \epsilon_i; \tag{7}$$

this representation defines ϵ_i and makes it uncorrelated with $\mathbf{c}^\top \mathbf{Y}$. Letting $\sigma_{\epsilon_i}^2$ denote the variance of ϵ_i, we get $\sigma_{\epsilon_i}^2 / w = \sigma_i^2 - \beta_i^2 \mathbf{c}^\top \mathbf{\Sigma} \mathbf{c}$, with $\sigma_i^2 = \mathbf{\Sigma}_{ii}$, and with w as defined following (4). Denote by $\mu_c = \mathbf{c}^\top \boldsymbol{\mu}$ the expected loss and by $\sigma_c^2 = w \mathbf{c}^\top \mathbf{\Sigma} \mathbf{c}$ its variance.

PROPOSITION 2 *Suppose that* $\mathbf{Y}$ *is elliptically distributed as in* (4). *As* $\ell \to \infty$, MES_i *and* MVS_i *behave as follows.*

Table 4. Estimates of MES and top 50 rankings based on daily returns from June 2006 through June 2007, as in Acharya *et al.* (2010). The last three columns show EL confidence levels for the significance of pairwise comparisons of rankings. NMX and WU traded for only part of the time period and are omitted from the comparison.

MES ranking, i	Name of company	Ticker	MES (%)	Confidence ($MES_i > MES_j$)		
				$j = i + 1$	$j = i + 10$	$j = 100$
1	INTERCONTINENTALEXCHANGE INC	ICE	3.36	0.47%	37.29%	98.88%
2	E TRADE FINANCIAL CORP	ETFC	3.29	4.57%	**	100.00%
3	BEAR STEARNS COMPANIES INC	BSC	3.15	1.33%	48.85%	100.00%
4	N Y S E EURONEXT	NYX	3.05	4.82%	39.18%	99.99%
5	C B RICHARD ELLIS GROUP INC	CBG	2.84	0.16%	82.10%	100.00%
6	LEHMAN BROTHERS HOLDINGS INC	LEH	2.83	5.10%	90.70%	100.00%
7	MORGAN STANLEY DEAN WITTER & CO	MS	2.72	1.57%	44.53%	100.00%
8	AMERIPRISE FINANCIAL INC	AMP	2.68	0.27%	53.85%	100.00%
9	GOLDMAN SACHS GROUP INC	GS	2.64	0.05%	67.56%	100.00%
10	MERRILL LYNCH & CO INC	MER	2.64	2.82%	40.74%	100.00%
11	SCHWAB CHARLES CORP NEW	SCHW	2.57	**	22.66%	99.97%
12	NYMEX HOLDINGS INC	NMX	2.47	**	**	**
13	C I T GROUP INC NEW	CIT	2.45	0.06%	**	100.00%
14	T D AMERITRADE HOLDING CORP	AMTD	2.43	7.68%	11.21%	99.99%
15	T ROWE PRICE GROUP INC	TROW	2.27	0.76%	12.06%	100.00%
16	EDWARDS A G INC	AGE	2.26	0.08%	19.52%	100.00%
17	FEDERAL NATIONAL MORTGAGE ASSN	FNM	2.25	0.43%	22.70%	100.00%
18	JANUS CAP GROUP INC	JNS	2.23	0.56%	20.89%	99.89%
19	FRANKLIN RESOURCES INC	BEN	2.20	0.96%	19.79%	99.94%
20	LEGG MASON INC	LM	2.19	0.34%	25.82%	99.88%
21	AMERICAN CAPITAL STRATEGIES LTD	ACAS	2.15	0.28%	38.14%	100.00%
22	STATE STREET CORP	STT	2.12	**	39.60%	100.00%
23	WESTERN UNION CO	WU	2.10	**	**	**
24	COUNTRYWIDE FINANCIAL CORP	CFC	2.09	0.11%	29.19%	96.14%
25	EATON VANCE CORP	EV	2.09	2.08%	51.50%	99.72%
26	S E I INVESTMENTS COMPANY	SEIC	2.00	0.72%	12.18%	99.86%
27	BERKLEY W R CORP	BER	1.95	0.08%	16.95%	99.85%
28	SOVEREIGN BANCORP INC	SOV	1.95	0.61%	37.21%	100.00%
29	JPMORGAN CHASE & CO	JPM	1.93	2.20%	45.10%	100.00%
30	BANK NEW YORK INC	BK	1.90	1.79%	63.02%	100.00%
31	M B I A INC	MBI	1.84	0.09%	17.72%	99.54%
32	BLACKROCK INC	BLK	1.83	0.53%	13.39%	99.58%
33	LEUCADIA NATIONAL CORP	LUK	1.80	0.00%	13.79%	99.89%
34	WASHINGTON MUTUAL INC	WM	1.80	2.20%	44.33%	99.95%
35	NORTHERN TRUST CORP	NTRS	1.75	0.39%	12.06%	100.00%
36	C B O T HOLDINGS INC	BOT	1.71	0.01%	5.91%	82.07%
37	PRINCIPAL FINANCIAL GROUP INC	PFG	1.71	4.21%	38.22%	99.97%
38	CITIGROUP INC	C	1.66	0.69%	18.75%	99.92%
39	LOEWS CORP	LTR	1.63	0.99%	11.10%	98.09%
40	GENWORTH FINANCIAL INC	GNW	1.59	0.23%	11.51%	99.77%
41	LINCOLN NATIONAL CORP IN	LNC	1.59	0.05%		100.00%
42	UNION PACIFIC CORP	UNP	1.58	0.86%		92.23%
43	AMERICAN EXPRESS CO	AXP	1.56	0.57%		99.94%
44	COMERICA INC	CMA	1.55	0.75%		99.92%
45	C I G N A CORP	CI	1.54	0.09%		76.81%
46	FIDELITY NATIONAL INFO SVCS INC	FIS	1.54	0.36%		99.65%
47	METLIFE INC	MET	1.52	0.61%		96.81%
48	PROGRESSIVE CORP OH	PGR	1.51	1.05%		99.50%
49	M & T BANK CORP	MTB	1.49	0.21%		98.13%
50	NATIONAL CITY CORP	NCC	1.48			91.21%
100	AMERICAN INTERNATIONAL GROUP INC	AIG	0.71			

(i) *If* **X** *is* ERV(α, v), *then*

$$MES_i = \mu_i + \beta_i(\ell - \mu_c) + o(\ell)$$
$$MVS_i = \frac{\sigma_{\epsilon_i}^2 \sigma_c^{\alpha-2}}{v^2 \alpha^2 w^{\alpha/2}}(\ell - \mu_c)^{-(\alpha-2)} + o(\ell^{-(\alpha-2)}).$$

(ii) *If* **X** *is* RV(v), $v > 2$, *then*

$$MES_i = \mu_i + \frac{v}{v-1}\beta_i(\ell - \mu_c) + o(\ell)$$
$$MVS_i = \left(\frac{\beta_i^2 v}{(v-1)^2(v-2)} - \frac{\sigma_{\epsilon_i}^2 v}{\sigma_c^2(v-1)(v-2)}\right)$$
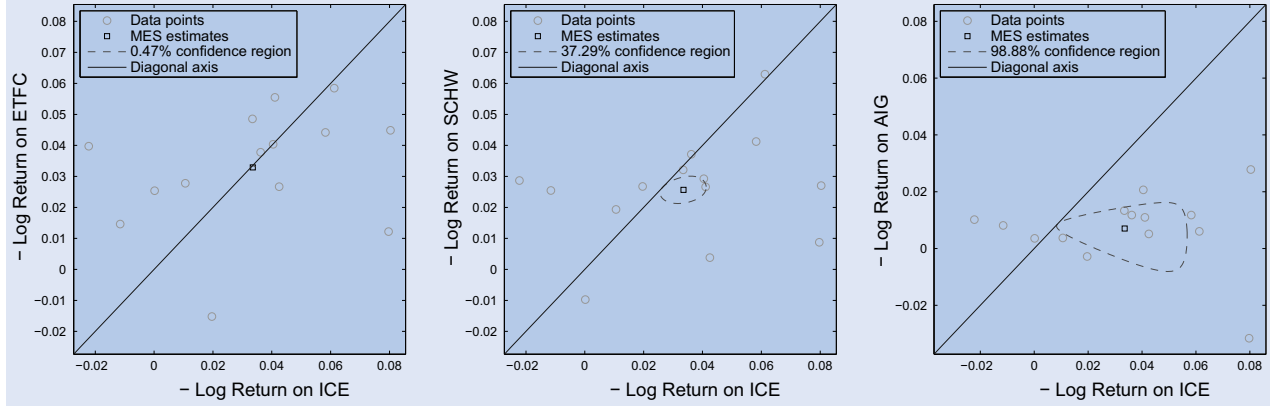$$\times (\ell - \mu_c)^2 + o(\ell^2).$$

Figure 6. The figures show the largest EL confidence regions for (MES_i, MES_j) below the 45° degree line for the cases $(i, j) = (1, 2)$, $(1, 11)$, and $(1, 100)$.

Proof The results follow from substituting (7) into the definitions of MES and MVS and then using table 3 to evaluate the conditional mean and the conditional variance, taking $W_1 = \sqrt{w}(\mathbf{c}^\top \mathbf{Y} - \mu_c)/\sigma_c$ and $W_2 = \sqrt{w}\epsilon_i/\sigma_{\epsilon_i}$. To see that (W_1, W_2) is in either $\mathsf{RV}(\nu)$ or $\mathsf{ERV}(\alpha, \nu)$, we recall that $\mathbf{Y} - \boldsymbol{\mu} = A\mathbf{X}$, where $\mathbf{X}$ has a spherical distribution in $\mathsf{RV}(\nu)$ or $\mathsf{ERV}(\alpha, \nu)$. So

$$W_1 = \frac{\sqrt{w}}{\sigma_c}\mathbf{c}^\top A\mathbf{X} = \left(\frac{\sqrt{w}}{\sigma_c}A^\top \mathbf{c}\right)^\top \mathbf{X},$$

$$W_2 = \frac{\sqrt{w}}{\sigma_{\epsilon_i}}(\mathbf{e}_i - \beta_i\mathbf{c})^\top A\mathbf{X} = \left(\frac{\sqrt{w}}{\sigma_{\epsilon_i}}A^\top(\mathbf{e}_i - \beta_i\mathbf{c})\right)^\top \mathbf{X}$$

By our choices of β_i, σ_c, and σ_{ϵ_i}, we have

$$(A^\top \mathbf{c}) \cdot (A^\top(\mathbf{e}_i - \beta_i\mathbf{c})) = \mathbf{c}^\top \boldsymbol{\Sigma}(\mathbf{e}_i - \beta_i\mathbf{c}) = 0,$$

$$\|\frac{\sqrt{w}}{\sigma_c}A^\top \mathbf{c}\|^2 = \frac{w}{\sigma_c^2}\mathbf{c}^\top \boldsymbol{\Sigma}\mathbf{c} = 1,$$

$$\|\frac{\sqrt{w}}{\sigma_{\epsilon_i}}(A^\top(\mathbf{e}_i - \beta_i\mathbf{c}))\|^2 = \frac{w}{\sigma_{\epsilon_i}}(\sigma_i^2 - \beta_i^2\mathbf{c}^\top \boldsymbol{\Sigma}\mathbf{c}) = 1.$$

That is, $\frac{\sqrt{w}}{\sigma_c}A^\top \mathbf{c}$ and $\frac{\sqrt{w}}{\sigma_{\epsilon_i}}A^\top(\mathbf{e}_i - \beta_i\mathbf{c})$ are orthonormal vectors. So there exists an orthogonal matrix $\mathbf{U}$ whose first two rows are $\left(\frac{\sqrt{w}}{\sigma_c}A^\top \mathbf{c}\right)^\top$ and $\left(\frac{\sqrt{w}}{\sigma_{\epsilon_i}}A^\top(\mathbf{e}_i - \beta_i\mathbf{c})\right)^\top$. Since $\mathbf{X}$ is spherical, $\mathbf{UX}$ has the same distribution as $\mathbf{X}$. By Propositions 3 and 6 in the appendix, (W_1, W_2) is a two-dimensional spherical random vector in $\mathsf{RV}(\nu)$ or $\mathsf{ERV}(\alpha, \nu)$. $\square$

Proposition 2 shows, as one might expect, that the MES is larger under heavier-tailed distributions. The result also highlights important differences in how the MVS depends on the loss level ℓ: the MVS is essentially constant with $\alpha = 2$ (as in the normal distribution), it grows linearly with ℓ for $\alpha = 1$ (as in the Laplace distribution), and it is quadratic in ℓ in the regularly varying case. A large MVS suggests that estimates of MES are likely to be imprecise, an issue we examine next.

6.3. EL significance of MES rankings

The EL method in section 3 can be used to estimate confidence regions for a full vector $(MES_1, \ldots, MES_d)$. Here we extend these ideas to measure the significance of MES rankings.

Acharya *et al.* (2010) use an MES measure as part of their analysis of systemic risk. Their MES for a company is the expected decline in the company's stock price conditional on a large decline in the whole market, as measured by a broad market index. This is a stress test in which the stress scenario is a decline in the market index. Acharya *et al.* (2010) rank firms by their MES as an indication of their systemic importance. Closely related measures are used in Adrian and Brunnermeier (2009) and Huang *et al.* (2012). Acharya *et al.* (2013) use their measure to evaluate macroprudential stress tests.

A ranking $MES_i > MES_j$ of firm i higher than firm j is equivalent to the point (MES_i, MES_j) lying below the 45° line in the plane. In practice, we estimate MES values from historical data and check if the point estimate falls in this halfspace. We can supplement the point estimate with an EL confidence region using the procedure in section 3. If a 95% confidence region is fully contained within the halfspace but a 99% confidence region is not, then the significance of the ranking is between 1 and 5%. Indeed, we can measure the significance of an estimated ranking by the smallest p for which the $(1 - p)$ confidence region is contained within the halfspace. The same idea can be applied to test the simultaneous significance of an ordering of three or more firms.

Following Appendix B of Acharya *et al.* (2010), we estimate MES values using daily stock returns for the 13 months from June 2006 through June 2007. We find the 5% of days with the largest declines in the CRSP value-weighted index and estimate the MES of each firm in Appendix B by averaging the firm's stock return over those days.† The resulting top 50 values and rankings, displayed in table 4, match those in Acharya *et al.* (2010). See Brownlees and Engle (2011) for a dynamic approach to MES estimation.

In table 4, we also report EL confidence levels (i.e. 1 minus significance levels) for pairwise comparisons between firms ranked consecutively, firms ranked ten apart, and between each firm and AIG, which is ranked 100th. None of the comparisons between consecutive firms or firms ranked 10 apart approaches conventional thresholds for statistical significance, corresponding to a confidence level of 90% or higher. There is too much conditional variability in the tails to draw reliable conclusions about the MES comparisons, as one might suspect from the MVS asymptotics in Proposition 2. Nearly all

---

†The worst days are, in decreasing order of severity, Feb 27, Mar 13, Jun 5, 2006, Jun 7, 2007, Jun 12, 2006, Nov 27, Jul 13, May 10, Jun 13, 2006, Jun 20, 2007, Mar 2, Mar 5, Sep 6, and May 24.

the comparisons with AIG are highly significant. The overall picture that emerges from the table is that the top 50 firms can be confidently ranked higher than the 100th, but all of the top 50 should be viewed as of roughly equal importance, as measured by MES.

The pairwise comparison for ICE with ETFC, SCHW and AIG is illustrated in figure 6. Each panel plots the negative log returns for the indicated stocks on the worst 5% of days for the index. Each panel also shows the largest EL confidence region contained below the 45° line. The corresponding confidence level (or, more precisely, the amount by which the confidence level falls short of 100%) measures the significance of each pairwise ordering. These are the values reported in table 4. The rightmost panel of figure 6 indicates substantial skewness in the extreme outcomes (possibly due in part to the small number of observations) and how this skewness is reflected in the confidence region.

7. Concluding remarks

We have developed a method for estimating the most likely scenario leading to large losses, which is the defining problem of reverse stress testing. Our method uses historical data and combines an empirical likelihood estimate with an asymptotic adjustment based on the heaviness of the tails of the underlying market factors. The adjustment has the same simple form for a wide class of multivariate distributions: it is determined by the index of regular variation in the heavy-tailed case, and no adjustment is needed with lighter tails. The same analysis allows us to examine the marginal expected shortfall and its associated variance, conditional on a stress scenario. Here, too, the behaviour is characterized by the heaviness of the tails. The empirical likelihood methodology then allows us to estimate the significance of rankings based on MES.

Acknowledgements

References

Acharya, V., Engle, R.F. and Pierret, D., Testing macroprudential stress tests: The risk of regulatory risk weights. Working Paper, Stern School of Business, New York University, New York, 2013.

Acharya, V., Pedersen, L.H., Philippon, T., Richardson, M., Measuring systemic risk. Working Paper, Stern School of Business, New York University, New York, 2010.

Adrian, T. and Brunnermeier, M., *CoVaR*, Staff Report No. 348, Federal Reserve Bank of New York, New York, 2009.

Alfaro, R.A. and Drehmann, M., Macro stress tests and crises: What can we learn? *BIS Quart. Rev.*, 2009, December, 29–41.

Baysal, R.E. and Staum, J., Empirical likelihood for value-at-risk and expected shortfall. *J. Risk*, 2008, **11** (1), 3–32.

Board of Governors of the Federal Reserve System Capital Plan Review 2013: Summary Instructions and Guidance, 9 November, 2012. Available online at: http://www.federalreserve. gov/bankinforeg/ccar.htm (accessed 2 June 2014).

Borio, C., Drehmann, M. and Tsatsaronis, K., Stress-testing macro stress testing: Does it live up to expectations? BIS Working Paper 369, Bank for International Settlements, Basel, Switzerland, 2012.

Breuer, T., Jandacka, M., Rheinberger, K. and Summer, M., How to find plausible, severe, and useful stress scenarios. *Int. J. Central Banking*, 2009, **5**, 205–224.

Brownlees, C. and Engle, R., Volatility, correlation and tails for systemic risk measurement. Working Paper, Stern School of Business, New York University, New York, 2011.

De, R. and Tamarchenko, T., VaR you can rely on. *Risk Mag.*, 2002, **15**, 77–81.

Fang, K.-T., Kotz, S. and Ng, K.-W., *Symmetric Multivariate and Related Distributions*, 1990 (Chapman and Hall: London).

Financial Services Authority Stress and Scenario Testing, Policy Statement 09/20, London, UK, 2009.

Flood, M. and Korenko, G., Systemic scenario selection. Working Paper 0006, Office of Financial Research, US Treasury Department, 2012. Available online at: http://www.treasury.gov/ initiatives/ofr/research (accessed 2 June 2014).

Glasserman, P., Kang, C. and Kang, W., Stress scenario selection by empirical likelihood. Working Paper 0007, in Office of Financial Research, US Treasury Department, 2013. Available online at: http://www.treasury.gov/initiatives/ofr/research (accessed 2 June 2014).

Goldstein, I. and Sapra, H., Should banks' stress test results be disclosed? An analysis of the costs and benefits. Working Paper, Wharton School, University of Pennsylvania, 2012.

Haas, M. and Pigorsch, C., Financial economics, fat-tailed distributions. In *Encyclopedia of Complexity and Systems Science* edited by R. Meyers, 2013 (Springer-Verlag: Berlin Heidelberg).

Heyde, C.C. and Kou, S.G., On the controversy over tailweight of distributions. *Oper. Res. Lett.*, 2004, **32**, 399–408.

Hida, E.T., *Deloitte Global Risk Management Survey*, 7th ed., 2011 (Deloitte: London).

Hirtle, B., Schuermann, T. and Stiroh, K., *Macroprudential Supervision of Financial Institutions: Lessons from the SCAP*, Staff Report 409, Federal Reserve Bank of New York, 2009.

Huang, X., Zhou, H. and Zhu, H., Systemic risk contributions. *J. Financ. Serv. Res.*, 2012, **42**, 55–83.

Kopeliovich, Y., Novosyolov, A., Satchkov, D., Schachter, B., Robust risk estimation and hedging: A reverse stress testing approach. Working Paper, EDHEC Business School, Nice, France, 2013.

Koyluoglu, H.U., Imagine ISO-loss surfaces. Working Paper, Oliver Wyman, New York, 2006.

McNeil, A., Frey, R. and Embrechts, P., *Quantitative Risk Management*, 2005 (Princeton University Press: Princeton, NJ).

Moody's, *Banking Industry Survey on Stress Testing*, 2011 (Moody's Analytics: New York).

Oura, H. and Schumacher, L., Macrofinancial stress testing – principles and practices. Working Paper 082212, International Monetary Fund, Washington, DC, 2012.

Owen, A.B., *Empirical Likelihood*, 2001 (Chapman & Hall/CRC Press: Boca Raton, FL).

Peng, L., Qi, Y., Wang, R. and Yang, J., Jackknife empirical likelihood method for some risk measures and related quantities. *Insurance Math. Econ.*, 2012, **51**, 142–150.

Pritsker, M., Enhanced stress testing and financial stability. Working Paper, Federal Reserve Bank of Boston, 2011.

Quagliariello, M., *Stress-testing the Banking System*, 2009 (Cambridge University Press: Cambridge).

Rebonato, R., *Coherent Stress Testing: A Bayesian Approach*, 2010 (Wiley: New York).

Resnick, S., *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, 2007 (Springer: New York).

Schmidt, R., Tail dependence for elliptically contoured distributions. *Math. Meth. Oper. Res.*, 2002, **55**, 301–327.

Schuermann, T., Stress testing banks, Wharton financial institutions. Working Paper 12–08, University of Pennsylvania, Philadelphia, PA, 2012.

Appendix A. Proofs

A.1. Main results

Most of the work in proving Proposition 1 and Theorem 1 lies in establishing the limits in table 3. Before detailing these limits, we show how they lead to the stated results. For Proposition 1 we need finite means, hence the condition $\nu > 1$; the variance limits in table 3 and Proposition 2 require finite variance, thus $\nu > 2$; and Theorem 1 requires finite fourth moments, hence $\nu > 4$.

Recall our standing assumption that the covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{Z}$ is positive definite. By relabelling L as the $(d+1)$st coordinate of $\tilde{\mathbf{Z}}$, we may rewrite $(\mathbf{Z}, L)$ simply as $\tilde{\mathbf{Z}} \in \mathbb{R}^{d+1}$. Conditioning on $L \geq \ell$ then reduces to conditioning on $\mathbf{c}_1^\top \tilde{\mathbf{Z}} \geq \ell$ with $\mathbf{c}_1^\top = (0, \ldots, 0, 1) \in \mathbb{R}^{d+1}$, provided the covariance matrix $\tilde{\boldsymbol{\Sigma}}$ of $(\mathbf{Z}, L)$ remains positive definite. If $\tilde{\boldsymbol{\Sigma}}$ fails to be positive definite, then there exists a vector $\mathbf{c}_2 \in \mathbb{R}^d$ such that $L = \mathbf{c}_2^\top \mathbf{Z} + h$, a.s., where h is a constant. Thus, for both cases, it suffices to consider conditioning on the form of $\hat{\mathbf{c}}^\top \hat{\mathbf{Z}} \geq \hat{\ell}$, for some $\hat{\ell}$ and some $\hat{\mathbf{c}} \neq \mathbf{0}$, with $\hat{\mathbf{Z}}$ ($=\tilde{\mathbf{Z}}$ or $\mathbf{Z}$) having a positive definite covariance matrix. We may therefore drop the hats.

Proof of Proposition 1   By replacing ℓ with $\ell - \mathbf{c}^\top \boldsymbol{\mu}$, we may take $\boldsymbol{\mu} = 0$ in the representation (4). Consider, then, $\mathbf{Z} = \mathbf{AX}$, with $\mathbf{X}$ spherical. Set $\hat{\mathbf{c}} = \mathbf{A}^\top \mathbf{c}$. Suppose we condition on $\mathbf{c}^\top \mathbf{Z} = \mathbf{c}^\top \mathbf{AX} = \hat{\mathbf{c}}^\top \mathbf{X}$. We can find an orthogonal matrix $\mathbf{U}$ such that $\hat{\mathbf{c}}^\top \mathbf{U} = \|\hat{\mathbf{c}}\| (1, 0, \ldots, 0)$, and by Theorem 2.5 of Fang *et al.* (1990), $\mathbf{U}^{-1}\mathbf{X}$ has the same distribution as $\mathbf{X}$. That is, $\mathbf{c}^\top \mathbf{Z}$ has the same distribution as

$$(\hat{\mathbf{c}}^\top \mathbf{U})(\mathbf{U}^{-1}\mathbf{X}) = \|\hat{\mathbf{c}}\| \hat{Z}_1$$

where $\hat{\mathbf{Z}} = \mathbf{U}^{-1}\mathbf{X}$. Hence, we can reduce an arbitrary linear combination to one of the form $\mathbf{c} = (1, 0, \ldots, 0)^\top$ with $\mathbf{Z}$ spherical.

In the setting of the proposition, this shows that we can find an invertible matrix $\mathbf{B} = \mathbf{U}^{-1}\mathbf{A}^{-1}$ such that $\mathbf{Bz}^*(\ell) = \ell \times (1, 0, \ldots, 0)^\top$ and $\mathbf{B}\bar{\mathbf{z}}(\ell) = \mathbb{E}[Z_1 | Z_1 \geq \ell] \times (1, 0, \ldots, 0)^\top$. By setting $\kappa_\ell = \ell / \mathbb{E}[Z_1 | Z_1 \geq \ell]$, we obtain equation (6). The first row of table 3 then gives the stated limits for κ_ℓ. $\square$

Owen (2001) provides a triangular array version of his EL theorem for data of the form $\mathbf{Z}_{1,n}, \ldots, \mathbf{Z}_{n,n}, n = 1, 2, \ldots$, in which variables with a shared second subscript are independent of each other and have a common mean. His result requires two conditions:

(i) For some $c > 0$, $\frac{\lambda_{m,n}}{\lambda_{M,n}} \geq c$, where $\lambda_{m,n}$ and $\lambda_{M,n}$ are the minimum and maximum eigenvalues of the covariance matrix associated with the n-th row of the array, respectively.

(ii) $\frac{1}{n^2} \sum_{i=1}^n \mathbb{E}\big[\|\mathbf{Z}_{i,n} - \boldsymbol{\mu}_n\|^4 \lambda_{M,n}^{-2}\big] \to 0$.

An additional convex hull condition required for the theorem is automatically satisfied by elliptical distributions.

Define

$$\mathbf{D}_n = \begin{cases} diag\{\ell_n^{\alpha-1}, \ell_n^{\alpha/2-1}, \ldots, \ell_n^{\alpha/2-1}\} & \text{for } \mathsf{ERV}(\alpha, \nu) \text{ case;} \\ diag\{\frac{1}{\ell_n}, \ldots, \frac{1}{\ell_n}\} & \text{for } \mathsf{RV}(\nu) \text{ case} \end{cases}$$

where $diag$ represents a diagonal matrix with specified elements. Let $\mathbf{Z}_1, \mathbf{Z}_2, \ldots$ be i.i.d. factor changes. Choose those satisfying $\mathbf{c}^\top \mathbf{Z}_j \geq \ell_n$, so that $\mathbf{Z}'_{1,n}, \ldots, \mathbf{Z}'_{n,n}$ are i.i.d. samples from the distribution of $\mathbf{Z}|\{\mathbf{c}^\top \mathbf{Z} \geq \ell_n\}$. Then, we can apply the EL theorem for triangular arrays to the scaled factors $\mathbf{Z}_{k,n} = \mathbf{D}_n \mathbf{B} \mathbf{Z}'_{k,n}$ where $\mathbf{B}$ is the matrix introduced in the proof of Proposition 1. Theorem 1 follows once we verify conditions (i) and (ii) for this array and apply Proposition 1. We will show that the off-diagonal conditional covariances all vanish in section A.2, so (i) will follow from limits of the conditional variances in table 3. For (ii), we need to extend these limits to higher moments and these computations are shown in sections A.4 and A.5.

We apply Theorem 1 to $\mathbf{D}_n \mathbf{B} \mathbf{Z}'_{k,n}$ for the estimation of $\mathbf{D}_n \mathbf{B} \mathbb{E}[\mathbf{Z} | \mathbf{c}^\top \mathbf{Z} \geq \ell_n]$. To estimate $\mathbb{E}[\mathbf{Z} | \mathbf{c}^\top \mathbf{Z} \geq \ell_n]$, we can apply the procedure to the original data $\mathbf{Z}'_{k,n}$ since the confidence region $C_{1-\alpha,n}$ consists of convex (so linear) combination of observed data.

A.2. Vanishing conditional covariances

Let $\mathbf{Z}$ have a spherical distribution. For all $1 \leq i < j$, by the invariance under orthogonal transformations, (Z_1, Z_i, Z_j) and $(Z_1, -Z_i, Z_j)$ share the same distribution. Hence $\mathbb{E}[Z_i | Z_1 \geq \ell] = \frac{1}{2}\big(\mathbb{E}[Z_i | Z_1 \geq \ell] + \mathbb{E}[-Z_i | Z_1 \geq \ell]\big) = 0$ and

$$\begin{aligned} &\mathbb{COV}(Z_i, Z_j | Z_1 \geq \ell) \\ &\quad = \mathbb{E}[Z_i Z_j | Z_1 \geq \ell] - \mathbb{E}[Z_i | Z_1 \geq \ell] \times \mathbb{E}[Z_j | Z_1 \geq \ell] \\ &\quad = \mathbb{E}[Z_i Z_j | Z_1 \geq \ell] \\ &\quad = \frac{1}{2}\big(\mathbb{E}[Z_i Z_j | Z_1 \geq \ell] + \mathbb{E}[-Z_i Z_j | Z_1 \geq \ell]\big) \\ &\quad = 0 . \end{aligned} \tag{8}$$

Hence, it is enough to consider the conditional variances to check the eigenvalue conditions (i) of conditional covariance matrices.

A.3. Tail integration representation

We make frequent use of the following representation result:

LEMMA 1 *Let X be a real-valued random variable and $f : \mathbb{R} \to \mathbb{R}$ be an absolutely continuous function. Then, for $\ell \in \mathbb{R}$, we have*

$$\mathbb{E}[f(X)\mathbb{1}_{\{X \geq \ell\}}] = f(\ell)\mathbb{P}(X \geq \ell) + \int_\ell^\infty f'(x)\mathbb{P}(X \geq x)dx.$$

Proof   Since f is absolutely continuous, $f(u) = f(\ell) + \int_\ell^u f'(x)dx$ for $\ell \geq u$. Therefore,

$$\begin{aligned} \mathbb{E}[f(X)\mathbb{1}_{\{X \geq \ell\}}] &= \mathbb{E}\Big[\Big(f(\ell) + \int_\ell^X f'(x)dx\Big)\mathbb{1}_{\{X \geq \ell\}}\Big] \\ &= f(\ell)\mathbb{P}(X \geq \ell) + \mathbb{E}\Big[\int_\ell^\infty f'(x)\mathbb{1}_{\{X \geq x\}}dx\Big] \\ &= f(\ell)\mathbb{P}(X \geq \ell) + \int_\ell^\infty f'(x)\mathbb{P}(X \geq x)dx. \end{aligned}$$

$\square$

A.4. Regularly varying tails

In this section, we consider a regularly varying spherically distributed random vector $\mathbf{X}$. We begin by recalling (see e.g. Resnick 2007, p. 20) that a positive function h is regularly varying (at ∞) with index $\rho \in \mathbb{R}$ if

$$\lim_{\ell \to \infty} \frac{h(\ell x)}{h(\ell)} = x^\rho,$$

for all $x > 0$. A random variable X is regularly varying with index ν ($\mathsf{RV}(\nu)$) if the function $\ell \mapsto \mathbb{P}(X \geq \ell)$ is regularly varying with index $-\nu$. Proposition 2.6 of Resnick (2007) shows that if h is regularly varying with index $-\nu < 0$, then for any $\epsilon > 0$ there exists $K > 0$ such that

$$(1 - \epsilon)x^{-\nu-\epsilon} < \frac{h(\ell x)}{h(\ell)} < (1 + \epsilon)x^{-\nu+\epsilon} \tag{9}$$

for all $\ell \geq K$ and $x \geq 1$.

Recall that we defined a spherical random vector $\mathbf{X} = R\mathbf{S}$ to be $\mathsf{RV}(\nu)$ if R is $\mathsf{RV}(\nu)$. This definition is justified by the following properties:

PROPOSITION 3 *Suppose that the spherically distributed random vector $\mathbf{X}$ is $\mathsf{RV}(\nu)$.*

(i) *For $1 \leq k \leq d$, $\mathbf{X}_k = (X_1, \cdots, X_k)^\top$ is a k-dimensional $\mathsf{RV}(\nu)$ spherical random vector.*
(ii) *For any $\mathbf{u} \in \mathbb{R}^d$, $\mathbf{u}^\top \mathbf{X}$ is $\mathsf{RV}(\nu)$.*

Proof   (i) See Proposition 3.1 of Schmidt (2002). (ii) By Theorem 2.4 of Fang *et al.* (1990), $\mathbf{u}^\top \mathbf{X} \stackrel{d}{=} \|\mathbf{u}\| X_1$. By (i), X_1 is $\mathsf{RV}(\nu)$, and then so is cX_1, for any $c > 0$. $\square$

PROPOSITION 4 *Suppose that the spherical random vector* **X** *is* RV(ν), $\nu > 0$. *For* $p \in [0, \nu)$, *we have*

$$\lim_{\ell \to \infty} \frac{1}{\ell^p} \mathbb{E}[X_1^p | X_1 \geq \ell] = \frac{\nu}{\nu - p}, \quad (10)$$

$$\lim_{\ell \to \infty} \frac{1}{\ell^p} \mathbb{E}[|X_2|^p | X_1 \geq \ell] = \frac{\nu}{\nu - p} \frac{\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{\nu-p+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu+1}{2}\right)}. \quad (11)$$

In particular, for integer $m \geq 1$ *with* $2m < \nu$,

$$\lim_{\ell \to \infty} \frac{1}{\ell^{2m}} \mathbb{E}[X_2^{2m} | X_1 \geq \ell] = \frac{\nu}{\nu - 2m} \prod_{k=1}^m \frac{2k-1}{\nu - (2k-1)}.$$

Proof First we shall prove (10). By Lemma 1,

$$\mathbb{E}[X_1^p \mathbb{1}_{\{X_1 \geq \ell\}}] = \ell^p \mathbb{P}(X_1 \geq \ell) + \int_\ell^\infty px^{p-1} \mathbb{P}(X_1 \geq x)dx$$
$$= \ell^p \mathbb{P}(X_1 \geq \ell) + \ell^p \int_1^\infty px^{p-1} \mathbb{P}(X_1 \geq \ell x)dx.$$

It follows that

$$\frac{1}{\ell^p} \mathbb{E}[X_1^p | X_1 \geq \ell] = \frac{\mathbb{E}[X_1^p \mathbb{1}_{\{X_1 \geq \ell\}}]}{\ell^p \mathbb{P}(X_1 \geq \ell)}$$
$$= 1 + \int_1^\infty px^{p-1} \frac{\mathbb{P}(X_1 \geq \ell x)}{\mathbb{P}(X_1 \geq \ell)}dx.$$

By (9) and the assumption $p < \nu$, the integrand of the above equation can be bounded by an integrable function. We apply the dominated convergence theorem to get

$$\lim_{\ell \to \infty} \frac{1}{\ell^p} \mathbb{E}[X_1^p | X_1 \geq \ell] = 1 + \int_1^\infty px^{p-\nu-1}dx = \frac{\nu}{\nu - p}.$$

Now, we turn to (11). By Proposition 3, the pair $(X_1, X_2)^\top$ extracted from **X** is itself a spherical random vector having regularly varying tail with index ν. Thus, it has a representation

$$(X_1, X_2) \overset{d}{=} (R_2 \cos \Theta, R_2 \sin \Theta),$$

where R_2 is a nonnegative RV(ν) random variable, and Θ is uniformly distributed on $(-\pi, \pi]$, independent of R_2. Then we apply Lemma 1 to obtain

$$\mathbb{E}[|X_2|^p \mathbb{1}_{\{X_1 \geq \ell\}}]$$
$$= \mathbb{E}[|R_2 \sin \Theta|^p \mathbb{1}_{\{R_2 \cos \Theta \geq \ell\}}]$$
$$= \mathbb{E}\left[|\sin \Theta|^p \mathbb{E}[R_2^p \mathbb{1}_{\{R_2 \cos \Theta \geq \ell\}}|\Theta] \mathbb{1}_{\{-\pi/2 \leq \Theta \leq \pi/2\}}\right]$$
$$= \frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} |\sin \theta|^p \mathbb{E}[R_2^p \mathbb{1}_{\{R_2 \geq \ell/\cos \theta\}}]d\theta$$
$$= \frac{1}{\pi} \int_0^{\pi/2} \sin^p \theta \frac{\ell^p}{\cos^p \theta} \mathbb{P}(R_2 \geq \ell/\cos \theta)d\theta$$
$$+ \frac{1}{\pi} \int_0^{\pi/2} \sin^p \theta \int_{\ell/\cos \theta}^\infty px^{p-1} \mathbb{P}(R_2 \geq x)dxd\theta$$
$$= \frac{\ell^p}{\pi} \int_0^{\pi/2} \tan^p \theta \, \mathbb{P}(R_2 \geq \ell/\cos \theta)d\theta$$
$$+ \frac{\ell^p}{\pi} \int_0^{\pi/2} \sin^p \theta \int_{1/\cos \theta}^\infty px^{p-1} \mathbb{P}(R_2 \geq \ell x)dxd\theta.$$

It follows that

$$\frac{\pi \mathbb{E}[|X_2|^p \mathbb{1}_{\{X_1 \geq \ell\}}]}{\ell^p \mathbb{P}(R_2 \geq \ell)}$$
$$= \int_0^{\pi/2} \tan^p \theta \frac{\mathbb{P}(R_2 \geq \ell/\cos \theta)}{\mathbb{P}(R_2 \geq \ell)}d\theta$$
$$+ \int_0^{\pi/2} \sin^p \theta \int_{1/\cos \theta}^\infty px^{p-1} \frac{\mathbb{P}(R_2 \geq \ell x)}{\mathbb{P}(R_2 \geq \ell)}dxd\theta.$$

Again by (9) and the assumption $p < \nu$, the integrands of the above equation can be bounded by integrable functions. Therefore, by the dominated convergence theorem,

$$\lim_{\ell \to \infty} \frac{\pi}{\ell^p \mathbb{P}(R_2 \geq \ell)} \mathbb{E}[|X_2|^p \mathbb{1}_{\{X_1 \geq \ell\}}]$$
$$= \int_0^{\pi/2} \tan^p \theta \cos^\nu \theta d\theta + \int_0^{\pi/2} \sin^p \theta \int_{1/\cos \theta}^\infty px^{p-\nu-1}dxd\theta$$
$$= \frac{\nu}{\nu - p} \int_0^{\pi/2} \sin^p \theta \cos^{\nu-p} \theta d\theta$$
$$= \frac{\nu}{2(\nu - p)} B\left(\frac{p+1}{2}, \frac{\nu-p+1}{2}\right),$$

with $B(\cdot, \cdot)$ the Beta function. Then

$$\lim_{\ell \to \infty} \frac{1}{\ell^p} \mathbb{E}[|X_2|^p | X_1 \geq \ell]$$
$$= \lim_{\ell \to \infty} \frac{\frac{\pi}{\ell^p \mathbb{P}(R_2 \geq \ell)} \mathbb{E}[|X_2|^p \mathbb{1}_{\{X_1 \geq \ell\}}]}{\frac{\pi}{\mathbb{P}(R_2 \geq \ell)} \mathbb{E}[\mathbb{1}_{\{X_1 \geq \ell\}}]}$$
$$= \frac{\frac{\nu}{2(\nu - p)} B\left(\frac{p+1}{2}, \frac{\nu-p+1}{2}\right)}{\frac{\nu}{2\nu} B\left(\frac{1}{2}, \frac{\nu+1}{2}\right)}$$
$$= \frac{\nu}{\nu - p} \frac{\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{\nu-p+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu+1}{2}\right)}.$$

$\square$

COROLLARY 1 *Suppose that the spherical random vector* **X** *is* RV(ν), $\nu > 0$.

(i) *If* $\nu > 2$, *then*

$$\lim_{\ell \to \infty} \mathbb{V}\left(\frac{X_1}{\ell} \mid X_1 \geq \ell\right) = \frac{\nu}{(\nu - 2)(\nu - 1)^2},$$
$$\lim_{\ell \to \infty} \mathbb{V}\left(\frac{X_2}{\ell} \mid X_1 \geq \ell\right) = \frac{\nu}{(\nu - 2)(\nu - 1)}.$$

(ii) *If* $\nu > 4$, *then*

$$\lim_{\ell \to \infty} \mathbb{E}\left[\left(\frac{X_1}{\ell} - \mathbb{E}\left[\frac{X_1}{\ell}|X_1 \geq \ell\right]\right)^4 \mid X_1 \geq \ell\right]$$
$$= \frac{3\nu(3\nu^2 + \nu + 2)}{(\nu - 1)^4(\nu - 2)(\nu - 3)(\nu - 4)},$$
$$\lim_{\ell \to \infty} \mathbb{E}\left[\left(\frac{X_2}{\ell} - \mathbb{E}\left[\frac{X_2}{\ell}|X_1 \geq \ell\right]\right)^4 \mid X_1 \geq \ell\right]$$
$$= \frac{3\nu}{(\nu - 4)(\nu - 1)(\nu - 3)}.$$

Proof For $m = 1, 2$ and $i = 1, 2$, we apply the binomial theorem to obtain

$$\lim_{\ell \to \infty} \mathbb{E}\left[\left(\frac{X_i}{\ell} - \mathbb{E}\left[\frac{X_i}{\ell}|X_1 \geq \ell\right]\right)^{2m} \mid X_1 \geq \ell\right]$$
$$= \lim_{\ell \to \infty} \sum_{k=0}^{2m} \binom{2m}{k} \frac{(-1)^{2m-k}}{\ell^k} \mathbb{E}[X_i^k | X_1 \geq \ell] \left\{\frac{1}{\ell} \mathbb{E}[X_i | X_1 \geq \ell]\right\}^{2m-k}.$$

Recall that $\mathbb{E}[X_2^{2k+1} | X_1 \geq \ell] = 0$ for $k \in \mathbb{N}$. Then proof is now easily completed by substituting (10) and (11) into this identity. $\square$

A.5. The ERV(α, ν) case

Recall that we define a nonnegative random variable R to be ERV(α, ν) if $\exp(R^\alpha)$ is RV(ν), in which case we also say that the spherical random vector $\mathbf{X} = R\mathbf{S}$ is ERV(α, ν).

We start with a simple lemma.

LEMMA 2 *Let $\alpha > 0$ and $x > 0$. Then*

$$\lim_{\ell \to \infty} \left\{ \ell^{\alpha-1}(x + \ell^\alpha)^{1/\alpha} - \ell^\alpha \right\} = \frac{x}{\alpha},$$

$$\lim_{\ell \to \infty} \left\{ \ell^{\alpha-2}(x + \ell^\alpha)^{2/\alpha} - \ell^\alpha \right\} = \frac{2x}{\alpha}.$$

Proof For $p > 0$ and $x > 0$,

$$(1 + x)^p - 1 = p \int_0^x (1+t)^{p-1} dt = px \int_0^1 (1+xt)^{p-1} dt.$$

Therefore,

$$\lim_{\ell \to \infty} \left\{ \ell^{\alpha-1}(x + \ell^\alpha)^{1/\alpha} - \ell^\alpha \right\} = \lim_{\ell \to \infty} \ell^\alpha \left\{ \left(1 + \frac{x}{\ell^\alpha}\right)^{1/\alpha} - 1 \right\}$$
$$= \frac{x}{\alpha} \lim_{\ell \to \infty} \int_0^1 \left(1 + \frac{x}{\ell^\alpha}t\right)^{1/\alpha-1} dt$$
$$= \frac{x}{\alpha}.$$

Similarly,

$$\lim_{\ell \to \infty} \left\{ \ell^{\alpha-2}(x + \ell^\alpha)^{2/\alpha} - \ell^\alpha \right\} = \frac{2x}{\alpha} \lim_{\ell \to \infty} \int_0^1 \left(1 + \frac{x}{\ell^\alpha}t\right)^{2/\alpha-1} dt$$
$$= \frac{2x}{\alpha}. \qquad \square$$

Lemma 3 characterizes the random variables in $\mathsf{ERV}(\alpha, \nu)$.

LEMMA 3 *Let X be a nonnegative random variable. Then, X is $\mathsf{ERV}(\alpha, \nu)$ if and only if*

$$\lim_{\ell \to \infty} \frac{\mathbb{P}(X^\alpha \geq \ell + x)}{\mathbb{P}(X^\alpha \geq \ell)} = e^{-\nu x}, \quad \forall x > 0.$$

Proof This is an immediate consequence of the definition. $\square$

PROPOSITION 5 *Suppose that the nonnegative random variable X is $\mathsf{ERV}(\alpha, \nu)$.*

(i) *For any $\epsilon > 0$ there exists $K > 0$ such that*

$$(1 - \epsilon)e^{-(\nu+\epsilon)x} < \frac{\mathbb{P}(X^\alpha \geq \ell + x)}{\mathbb{P}(X^\alpha \geq \ell)} < (1 + \epsilon)e^{-(\nu-\epsilon)x}$$

for all $\ell \geq K$ and $x \geq 0$.

(ii) *For any $x > 1$,*

$$\lim_{\ell \to \infty} \frac{\mathbb{P}(X \geq \ell x)}{\mathbb{P}(X \geq \ell)} = 0.$$

(iii) *For any $c > 0$, $cX \in \mathsf{ERV}(\alpha, \nu/c^\alpha)$.*

(iv) *If B is independent of X and $B^2 \sim \mathsf{Beta}(a, b)$, then BX is $\mathsf{ERV}(\alpha, \nu)$.*

Proof (i) follows from (9). (ii) Let $x > 1$ and take $0 < \epsilon < \nu$. Then, by (i), there exists $K > 0$ such that

$$(1 - \epsilon)e^{-(\nu+\epsilon)t} < \frac{\mathbb{P}(X^\alpha \geq \ell + t)}{\mathbb{P}(X^\alpha \geq \ell)} < (1 + \epsilon)e^{-(\nu-\epsilon)t}$$

holds for all $\ell \geq K$ and $t \geq 0$. Thus,

$$\limsup_{\ell \to \infty} \frac{\mathbb{P}(X \geq \ell x)}{\mathbb{P}(X \geq \ell)} \leq \limsup_{\ell \to \infty} \frac{\mathbb{P}(X^\alpha \geq \ell^\alpha + \ell^\alpha(x^\alpha - 1))}{\mathbb{P}(X^\alpha \geq \ell^\alpha)}$$
$$\leq \lim_{\ell \to \infty} (1 + \epsilon)e^{-(\nu-\epsilon)\ell^\alpha(x^\alpha-1)} = 0.$$

(iii) For any $x \geq 0$,

$$\lim_{\ell \to \infty} \frac{\mathbb{P}((cX)^\alpha \geq \ell + x)}{\mathbb{P}((cX)^\alpha \geq \ell)}$$
$$= \lim_{\ell \to \infty} \frac{\mathbb{P}(X^\alpha \geq \ell/c^\alpha + x/c^\alpha)}{\mathbb{P}(X^\alpha \geq \ell/c^\alpha)} = e^{-(\nu/c^\alpha)x}.$$

Thus, by Lemma 3, cX is in $\mathsf{ERV}(\alpha, \nu/c^\alpha)$. (iv) We define

$$F(u; a, b) = \mathbb{P}(B^2 \geq u)$$
$$= \frac{1}{B(a, b)} \int_u^1 t^{a-1}(1 - t)^{b-1} dt, \quad 0 \leq u \leq 1.$$

Then, for $r \geq \ell$,

$$\frac{d}{dr} F\left(\frac{\ell^2}{r^2}; a, b\right) = \frac{2\ell^{2a}}{B(a, b)} \frac{(r^2 - \ell^2)^{b-1}}{r^{2a+2b-1}}.$$

By Lemma 1, we have

$$\mathbb{P}(BX \geq \ell)$$
$$= \mathbb{E}\left[\mathbb{P}\left(B^2 \geq \frac{\ell^2}{X^2} \mid X\right)\mathbb{1}_{\{X \geq \ell\}}\right]$$
$$= \mathbb{E}\left[F\left(\frac{\ell^2}{X^2}; a, b\right)\mathbb{1}_{\{X \geq \ell\}}\right]$$
$$= \frac{2\ell^{2a}}{B(a, b)} \int_\ell^\infty \frac{(r^2 - \ell^2)^{b-1}}{r^{2a+2b-1}} \mathbb{P}(X \geq r) dr$$
$$= \frac{2\ell^{2a}}{B(a, b)\alpha} \int_0^\infty \frac{\{(x + \ell^\alpha)^{2/\alpha} - \ell^2\}^{b-1}}{(x + \ell^\alpha)^{(2a+2b-2)/\alpha+1}} \mathbb{P}(X^\alpha \geq x + \ell^\alpha) dx$$
$$= \frac{2\ell^{-\alpha b}}{B(a, b)\alpha} \int_0^\infty \frac{\{\ell^{\alpha-2}(x + \ell^\alpha)^{2/\alpha} - \ell^\alpha\}^{b-1}}{\left(1 + \frac{x}{\ell^\alpha}\right)^{(2a+2b-2)/\alpha+1}} \mathbb{P}(X^\alpha \geq x + \ell^\alpha) dx$$

Therefore, by Lemma 2,

$$\lim_{\ell \to \infty} \frac{\ell^{\alpha b}\mathbb{P}(BX \geq \ell)}{\mathbb{P}(X \geq \ell)} = \frac{2}{B(a, b)\alpha} \int_0^\infty \left(\frac{2x}{\alpha}\right)^{b-1} e^{-\nu x} dx$$
$$= \frac{2}{B(a, b)\alpha} \left(\frac{2}{\alpha}\right)^{b-1} \frac{\Gamma(b)}{\nu^b}.$$

Finally, we have

$$\lim_{\ell \to \infty} \frac{\mathbb{P}((BX)^\alpha \geq \ell^\alpha + x)}{\mathbb{P}((BX)^\alpha \geq \ell^\alpha)}$$
$$= \lim_{\ell \to \infty} \frac{(\ell^\alpha + x)^b \frac{\mathbb{P}(BX \geq (\ell^\alpha+x)^{1/\alpha})}{\mathbb{P}(X \geq (\ell^\alpha+x)^{1/\alpha})}}{\ell^{\alpha b} \frac{\mathbb{P}(BX \geq \ell)}{\mathbb{P}(X \geq \ell)}} \times \frac{\mathbb{P}(X^\alpha \geq \ell^\alpha + x)}{\mathbb{P}(X^\alpha \geq \ell^\alpha)}$$
$$= e^{-\nu x}.$$

Hence, by Lemma 3, BX is $\mathsf{ERV}(\alpha, \nu)$. $\square$

Proposition 6 states invariance properties of spherical distributions in $\mathsf{ERV}(\alpha, \nu)$

PROPOSITION 6 *Suppose that the spherical random vector $\mathbf{X}$ is in $\mathsf{ERV}(\alpha, \nu)$.*

(i) *For $1 \leq k \leq d$, $\mathbf{X}_k = (X_1, \cdots, X_k)^\top$ is a k-dimensional spherical random vector $\mathsf{ERV}(\alpha, \nu)$.*

(ii) *For any $\mathbf{u} \in \mathbb{R}^d$, the conditional distribution $(\mathbf{u}^\top \mathbf{X} \mid \mathbf{u}^\top \mathbf{X} \geq 0)$ is in $\mathsf{ERV}(\alpha, \nu/\|\mathbf{u}\|^\alpha)$.*

Proof (6) By Lemma 2.5 of Schmidt (2002), $\mathbf{X}_k$ has a representation

$$\mathbf{X}_k \stackrel{d}{=} BR\mathbf{S}^{(k)}$$

in which $R \stackrel{d}{=} \|\mathbf{X}\|$, $B^2 \sim \mathsf{Beta}\left(\frac{k}{2}, \frac{d-k}{2}\right)$, $\mathbf{S}^{(k)}$ is uniformly distributed on $\mathbb{S}^{k-1}$, and R, B, and $\mathbf{S}^{(k)}$ are mutually independent. Proposition 5 implies that $\|\mathbf{X}_k\| \stackrel{d}{=} BR$ is $\mathsf{ERV}(\alpha, \nu)$. Hence, $\mathbf{X}_k = (X_1, \cdots, X_k)^\top$ is a k-dimensional spherical random vector in $\mathsf{ERV}(\alpha, \nu)$. (6) Using the argument in section A.1, $\mathbf{u}^\top \mathbf{X} \stackrel{d}{=} \|\mathbf{u}\|X_1$. By Proposition 5 and (6), $\|\mathbf{u}^\top \mathbf{X}\| \stackrel{d}{=} \|\mathbf{u}\||X_1|$ is $\mathsf{ERV}(\alpha, \nu/\|\mathbf{u}\|^\alpha)$. Since $\mathbf{u}^\top \mathbf{X}$ is symmetric around zero, $\|\mathbf{u}^\top \mathbf{X}\| \stackrel{d}{=} (\mathbf{u}^\top \mathbf{X} \mid \mathbf{u}^\top \mathbf{X} \geq 0)$. Hence, $(\mathbf{u}^\top \mathbf{X} \mid \mathbf{u}^\top \mathbf{X} \geq 0)$ is $\mathsf{ERV}(\alpha, \nu/\|\mathbf{u}\|^\alpha)$. $\square$

We can now calculate the asymptotics of tail moments of $\mathsf{ERV}(\alpha, \nu)$ spherical distributions.

PROPOSITION 7 *Suppose that the spherical random vector* **X** *is in* ERV(α, ν). *For $p \geq 0$, we have*

$$\lim_{\ell \to \infty} \frac{1}{\ell^p} \mathbb{E}[X_1^p | X_1 \geq \ell] = 1, \tag{12}$$

$$\lim_{\ell \to \infty} \mathbb{E}[(\ell^{\alpha-1} X_1 - \ell^\alpha)^p | X_1 \geq \ell] = \frac{\Gamma(p+1)}{(\nu\alpha)^p} \tag{13}$$

$$\lim_{\ell \to \infty} \ell^{\frac{p(\alpha-2)}{2}} \mathbb{E}[|X_2|^p | X_1 \geq \ell] = \left(\frac{2}{\nu\alpha}\right)^{p/2} \frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{1}{2})}. \tag{14}$$

In particular, if $m \in \mathbb{N}$, then

$$\lim_{\ell \to \infty} \ell^{m(\alpha-2)} \mathbb{E}[X_2^{2m} | X_1 \geq \ell] = \frac{(2m-1)!!}{(\nu\alpha)^m}.$$

Proof First we shall prove (12). By Lemma 1 and Proposition 5(ii),

$$\lim_{\ell \to \infty} \frac{1}{\ell^p} \mathbb{E}[X_1^p | X_1 \geq \ell] = 1 + p \lim_{\ell \to \infty} \int_1^\infty x^{p-1} \frac{\mathbb{P}(X \geq \ell x)}{\mathbb{P}(X \geq \ell)} dx = 1.$$

Next, we consider (13). Again by Lemma 1,

$$\mathbb{E}[(\ell^{\alpha-1} X_1 - \ell^\alpha)^p \mathbb{1}_{\{X_1 \geq \ell\}}]$$
$$= p\ell^{\alpha-1} \int_\ell^\infty (\ell^{\alpha-1} x - \ell^\alpha)^{p-1} \mathbb{P}(X_1 \geq x) dx$$
$$= \frac{p}{\alpha} \int_0^\infty \frac{\{\ell^{\alpha-1}(x+\ell^\alpha)^{1/\alpha} - \ell^\alpha\}^{p-1}}{(1 + \frac{x}{\ell^\alpha})^{1-1/\alpha}} \mathbb{P}(X_1^\alpha \geq \ell^\alpha + x) dx$$

By Proposition 6, ($X_1 | X_1 \geq 0$) is ERV(α, ν), so we apply Lemma 3 to obtain

$$\lim_{\ell \to \infty} \mathbb{E}[(\ell^{\alpha-1} X_1 - \ell^\alpha)^p | X_1 \geq \ell]$$
$$= \lim_{\ell \to \infty} \frac{\mathbb{E}[(\ell^{\alpha-1} X_1 - \ell^\alpha)^p \mathbb{1}_{\{X_1 \geq \ell\}}]}{\mathbb{P}(X_1 \geq \ell)}$$
$$= \lim_{\ell \to \infty} \frac{p}{\alpha} \int_0^\infty \frac{\{\ell^{\alpha-1}(x+\ell^\alpha)^{1/\alpha} - \ell^\alpha\}^{p-1}}{(1 + \frac{x}{\ell^\alpha})^{1-1/\alpha}} \frac{\mathbb{P}(X_1^\alpha \geq \ell^\alpha + x)}{\mathbb{P}(X_1^\alpha \geq \ell^\alpha)} dx$$
$$= \frac{p}{\alpha} \int_0^\infty \left(\frac{x}{\alpha}\right)^{p-1} e^{-\nu x} dx$$
$$= \frac{p}{\alpha^p} \frac{\Gamma(p)}{\nu^p}$$
$$= \frac{\Gamma(p+1)}{(\nu\alpha)^p}.$$

Next, we turn to (14). By Proposition 6, the 2-dimensional spherical random vector $\mathbf{X}_2 = (X_1, X_2)^\top$ is ERV(α, ν), so it has the representation

$$(X_1, X_2) \overset{d}{=} (R_2 \cos \Theta, R_2 \sin \Theta),$$

where the radial variable R_2 is ERV(α, ν), and Θ is uniformly distributed on $(-\pi, \pi]$, independent of R_2. We define

$$f(x; \ell) = 2x^p \int_0^{\cos^{-1}(\frac{\ell}{x})} \sin^p \theta \, d\theta$$
$$= x^p \int_0^{1-(\frac{\ell}{x})^2} \left(\frac{t^{p-1}}{1-t}\right)^{1/2} dt, \quad x \geq \ell.$$

Then

$$f'(x; \ell) = \frac{p}{x} f(x; \ell) + \frac{2\ell}{x}(x^2 - \ell^2)^{(p-1)/2}.$$

Using Lemma 1,

$$\mathbb{E}[|X_2|^p \mathbb{1}_{\{X_1 \geq \ell\}}] = \mathbb{E}[R_2^p |\sin \Theta|^p \mathbb{1}_{\{R_2 \cos \Theta \geq \ell\}}]$$
$$= \mathbb{E}[R_2^p \mathbb{E}[|\sin \Theta|^p \mathbb{1}_{\{R_2 \cos \Theta \geq \ell\}} | R_2] \mathbb{1}_{\{R_2 \geq \ell\}}]$$
$$= \frac{1}{2\pi} \mathbb{E}[2R_2^p \left(\int_0^{\cos^{-1}(\frac{\ell}{R_2})} \sin^p \theta \, d\theta\right) \mathbb{1}_{\{R_2 \geq \ell\}}]$$
$$= \frac{1}{2\pi} \mathbb{E}[f(R_2; \ell) \mathbb{1}_{\{R_2 \geq \ell\}}]$$
$$= \frac{1}{2\pi} \int_\ell^\infty f'(x; \ell) \mathbb{P}(R_2 \geq x) dx$$
$$= \frac{1}{2\alpha\pi} \int_0^\infty \frac{f'((x+\ell^\alpha)^{1/\alpha}; \ell)}{(x+\ell^\alpha)^{1-1/\alpha}} \mathbb{P}(R_2^\alpha \geq \ell^\alpha + x) dx.$$

Here,

$$\frac{p}{(x+\ell^\alpha)^{1/\alpha}} f((x+\ell^\alpha)^{1/\alpha}; \ell)$$
$$= p(x+\ell^\alpha)^{(p-1)/\alpha} \int_0^{1-\frac{\ell^2}{(x+\ell^\alpha)^{2/\alpha}}} \left(\frac{t^{p-1}}{1-t^2}\right)^{1/2} dt$$
$$= p\ell^{-\frac{(p-1)(\alpha-2)}{2} - \alpha} \left(1 + \frac{x}{\ell^\alpha}\right)^{(p-1)/\alpha}$$
$$\times \int_0^{\ell^\alpha(1-\frac{\ell^2}{(x+\ell^\alpha)^{2/\alpha}})} \left(\frac{t^{p-1}}{1-\ell^{-\alpha}t}\right)^{1/2} dt.$$

By Lemma 2,

$$\lim_{\ell \to \infty} \ell^\alpha \left(1 - \frac{\ell^2}{(x+\ell^\alpha)^{2/\alpha}}\right)$$
$$= \lim_{\ell \to \infty} \left(\frac{\ell^\alpha}{x+\ell^\alpha}\right)^{2/\alpha} \{\ell^{\alpha-2}(x+\ell^\alpha)^{2/\alpha} - \ell^\alpha\}$$
$$= \frac{2x}{\alpha}.$$

It follows that

$$\lim_{\ell \to \infty} \frac{p\ell^{\frac{(p-1)(\alpha-2)}{2}}}{(x+\ell^\alpha)^{1/\alpha}} f((x+\ell^\alpha)^{1/\alpha}; \ell) = 0.$$

On the other hand,

$$\lim_{\ell \to \infty} \frac{\ell^{\frac{(p-1)(\alpha-2)}{2}+1}}{(x+\ell^\alpha)^{1/\alpha}} \{(x+\ell^\alpha)^{2/\alpha} - \ell^2\}^{(p-1)/2}$$
$$= \lim_{\ell \to \infty} \left(\frac{\ell^\alpha}{x+\ell^\alpha}\right)^{1/\alpha} \{\ell^{\alpha-2}(x+\ell^\alpha)^{2/\alpha} - \ell^\alpha\}^{(p-1)/2}$$
$$= \left(\frac{2x}{\alpha}\right)^{(p-1)/2}.$$

Therefore,

$$\lim_{\ell \to \infty} \ell^{\frac{(p-1)(\alpha-2)}{2}} f'((x+\ell^\alpha)^{1/\alpha}; \ell) = 2\left(\frac{2x}{\alpha}\right)^{(p-1)/2}.$$

Now we combine the identities above to obtain

$$\lim_{\ell \to \infty} \frac{\ell^{\alpha-1+\frac{(p-1)(\alpha-2)}{2}} \mathbb{E}[|X_2|^p \mathbb{1}_{\{X_1 \geq \ell\}}]}{\mathbb{P}(R_2^\alpha \geq \ell^\alpha)}$$
$$= \frac{1}{2\alpha\pi} \lim_{\ell \to \infty} \int_0^\infty \frac{\ell^{\frac{(p-1)(\alpha-2)}{2}} f'((x+\ell^\alpha)^{1/\alpha}; \ell)}{(1+\frac{x}{\ell^\alpha})^{1-1/\alpha}} \frac{\mathbb{P}(R_2^\alpha \geq \ell^\alpha + x)}{\mathbb{P}(R_2^\alpha \geq \ell^\alpha)} dx.$$
$$= \frac{1}{\alpha\pi} \int_0^\infty \left(\frac{2x}{\alpha}\right)^{\frac{p-1}{2}} e^{-\nu x} dx$$
$$= \frac{1}{\alpha\pi} \left(\frac{2}{\alpha}\right)^{\frac{p-1}{2}} \nu^{-\frac{p+1}{2}} \Gamma\left(\frac{p+1}{2}\right).$$

Hence

$$\lim_{\ell \to \infty} \ell^{\frac{p(\alpha-2)}{2}} \mathbb{E}\big[|X_2|^p \mid X_1 \geq \ell\big]$$

$$= \lim_{\ell \to \infty} \frac{\ell^{\alpha-1+\frac{(p-1)(\alpha-2)}{2}} \frac{\mathbb{E}[|X_2|^p \mathbb{1}_{\{X_1 \geq \ell\}}]}{\mathbb{P}(R_2^\alpha \geq \ell^\alpha)}}{\ell^{\alpha-1-\frac{\alpha-2}{2}} \frac{\mathbb{E}[\mathbb{1}_{\{X_1 \geq \ell\}}]}{\mathbb{P}(R_2^\alpha \geq \ell^\alpha)}}$$

$$= \left(\frac{2}{\nu\alpha}\right)^{p/2} \frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{1}{2})}.$$

$\square$

COROLLARY 2 *Suppose that that spherical random vector* $\mathbf{X}$ *is* ERV(α, ν). *Then*

$$\lim_{\ell \to \infty} \mathbb{V}\big(\ell^{\alpha-1} X_1 \mid X_1 \geq \ell\big) = \frac{1}{(\nu\alpha)^2},$$

$$\lim_{\ell \to \infty} \mathbb{V}\big(\ell^{\frac{\alpha-2}{2}} X_2 \mid X_1 \geq \ell\big) = \frac{1}{(\nu\alpha)^2},$$

$$\lim_{\ell \to \infty} \mathbb{E}\big[(\ell^{\alpha-1} X_1 - \mathbb{E}[\ell^{\alpha-1} X_1 | X_1 \geq \ell])^4 \mid X_1 \geq \ell\big] = \frac{9}{(\nu\alpha)^4},$$

$$\lim_{\ell \to \infty} \mathbb{E}\big[(\ell^{\frac{\alpha-2}{2}} X_2 - \mathbb{E}[\ell^{\frac{\alpha-2}{2}} X_2 | X_1 \geq \ell])^4 \mid X_1 \geq \ell\big] = \frac{3}{(\nu\alpha)^4}.$$

Proof For $n \in \mathbb{N}$, we apply the binomial theorem to obtain

$$\lim_{\ell \to \infty} \mathbb{E}\big[(\ell^{\alpha-1} X_1 - \mathbb{E}[\ell^{\alpha-1} X_1 | X_1 \geq \ell])^n \mid X_1 \geq \ell\big]$$

$$= \lim_{\ell \to \infty} \sum_{k=0}^{n} \binom{n}{k}(-1)^{n-k} \mathbb{E}\big[(\ell^{\alpha-1} X_1 - \ell^\alpha)^k | X_1 \geq \ell\big]$$

$$\times \big\{\mathbb{E}\big[\ell^{\alpha-1} X_1 - \ell^\alpha | X_1 \geq \ell\big]\big\}^{n-k}$$

$$= \sum_{k=0}^{n} \binom{n}{k}(-1)^{n-k} \frac{k!}{(\nu\alpha)^k}\left(\frac{1}{\nu\alpha}\right)^{n-k}$$

$$= \frac{n!}{(\nu\alpha)^n} \sum_{k=0}^{n} \frac{(-1)^k}{k!}$$

$$= \begin{cases} \frac{1}{(\nu\alpha)^2}, & n = 2, \\ \frac{9}{(\nu\alpha)^4}, & n = 4. \end{cases}$$

Recall that $\mathbb{E}[X_2 | X_1 \geq \ell] = 0$. For $m = 1, 2$,

$$\lim_{\ell \to \infty} \mathbb{E}\big[(\ell^{\frac{\alpha-2}{2}} X_2 - \mathbb{E}[\ell^{\frac{\alpha-2}{2}} X_2 | X_1 \geq \ell])^{2m} \mid X_1 \geq \ell\big]$$

$$= \lim_{\ell \to \infty} \ell^{m(\alpha-2)} \mathbb{E}[X_2^{2m} | X_1 \geq \ell] = \frac{(2m-1)!!}{(\nu\alpha)^m}$$

$$= \begin{cases} \frac{1}{(\nu\alpha)^2}, & m = 1, \\ \frac{3}{(\nu\alpha)^4}, & m = 2. \end{cases}$$

$\square$