Asymptotic Theory for the Empirical Haezendonck Risk Measure

Jae Youn Ahn[a,*], Nariankadu D. Shyamalkumar[a,**,1]

[a]*Dept. of Statistics and Act. Sci., The University of Iowa, Schaeffer Hall, Iowa City, IA 52242, United States*

Abstract

Haezendonck risk measures is a recently introduced class of risk measures which includes, as its minimal member, the Tail Value-at-Risk (T-VaR) - T-VaR arguably the most popular risk measure in global insurance regulation. In applications often one has to estimate the risk measure given a random sample from an *unknown* distribution. The distribution could either be truly unknown or could be the distribution of a complex function of economic and idiosyncratic variables with the complexity of the function rendering indeterminable its distribution. Hence statistical procedures for the estimation of Haezendonck risk measures is a key requirement for its use in practice. A natural estimator of the Haezendonck risk measure is the Haezendonck risk measure of the empirical distribution, but its statistical properties have not yet been explored in detail. The main goal of this article is to both establish the strong consistency of this estimator and to derive weak convergence limits for this estimator. We also conduct a simulation study to lend insight into the sample sizes required for these asymptotic limits to take hold.

Key words: Orlicz Premium, Tail value-at-Risk (T-VaR), Conditional Tail Expectation (CTE), Empirical CTE
Classification: 10.050: IM10, 10.130: IM 30

[*]Corresponding Author
[**]Principal Corresponding Author
Email addresses: jaeyoun-ahn@uiowa.edu (Jae Youn Ahn), shyamal-kumar@uiowa.edu (Nariankadu D. Shyamalkumar)
[1]Dept. of Statistics and Act. Sci., The University of Iowa, Schaeffer Hall, Iowa City, IA 52242, United States. Tel: 319 335 1980; Fax: 319 335 3017

1. Introduction

Haezendonck risk measures is a class of risk measures which was recently introduced in Goovaerts et al. (2004). It is based on the premium calculation principle induced by an Orlicz norm, as presented in Haezendonck and Goovaerts (1982) (see also Goovaerts et al. (2003)), in the sense that it is the translation-equivariant minimal Orlicz risk measure. This class of risk measures was further studied in Bellini and Gianin (2008a), and they present an alternate formulation of these risk measures which makes them *coherent* in the sense of Artzner et al. (1997, 1999). It is worth mention that the Haezendonck risk measures preserve convex order; see Goovaerts et al. (2004), Bellini and Gianin (2008a), Nam et al. (2011), and Ahn and Shyamalkumar (2011b) for further properties of these risk measures. The most prominent member of this class, and in fact its minimal member, is the Tail Value-at-Risk (T-VaR) - T-VaR arguably the most popular risk measure in global insurance regulation.

In applications often one has to estimate the risk measure given a random sample from an *unknown* distribution. The distribution could either be truly unknown or could be the distribution of a complex function of economic and idiosyncratic variables with the complexity of the function rendering indeterminable its distribution. Hence statistical procedures for the estimation of Haezendonck risk measures is a key requirement for its use in practice. While the above references study properties of the Haezendonck risk measures, only Bellini and Gianin (2008b) deals with its statistical estimation. Note that the natural nonparametric estimator for the Haezendonck risk measure is its empirical analog, the Haezendonck risk measure of the empirical distribution. In Bellini and Gianin (2008b) the authors conduct a simulation study of this estimation procedure, and also use it to estimate the efficient frontier when the risk is measured by a Haezendonck risk measure. While this study suggests, in some cases, a normal asymptotic limit for this estimator, neither consistency nor weak convergence of this estimator has been established. This then is the main goal of this article; we provide a strong consistency and a weak convergence result for this non-parametric estimator with the latter also covering situations with a non-normal limit. The difficulty in establishing asymptotic results arises in good part from the lack of a convenient closed form expression for the Haezendonck risk measure of the empirical distribution function. After some definitions and establishing the notation we provide an example which demonstrates this inherent nature of our problem, and another that demonstrates that non-normal asymptotic weak limits occur even in non-pathological situations.

A non-negative, strictly increasing, convex function $\Psi(\cdot)$ on $\mathbb{R}^+$ with $\Psi(0) = 0$ and $\Psi(1) = 1$ is called a normalized Young function (see Rao and Ren (1991) for details). In the following we will work with the extensions of such functions to the whole of $\mathbb{R}$ satisfying $\Psi(x) = 0$ for $x < 0$. For convenience we will refer to such extensions simply as Young function. The class of Haezendonck risk measures is indexed by the class of Young function, and for each Haezendonck risk measure there exits a class of random variables for which it is well defined. A subset of this class of random variables is denoted by $\mathbb{X}_\Psi$, and is defined by

$$\mathbb{X}_\Psi := \left\{ X \middle| \Pr\left(X \leq 0\right) = 1 \text{ or } \exists s_\infty \geq 0 \text{ such that } \mathbb{E}\left(\Psi\left(\frac{X}{s}\right)\right) < \infty \Leftrightarrow s > s_\infty \right\}. \tag{1}$$

In Bellini and Gianin (2008a), for convenience, the random variables were restricted to L^∞, the space of essentially bounded random variables, a subset of $\mathbb{X}_\Psi$; we allow s_∞ to be greater than 0 in (1), unlike in Goovaerts et al. (2004), to accommodate situations like those in Example 5 where $\Psi(\cdot)$ is exponential and X is an exponential random variable.

The Orlicz premium principle corresponding to $\Psi(\cdot)$, and at level $\alpha \in [0, 1)$, is denoted by $H^\alpha_\Psi(\cdot)$, and for $X \in \mathbb{X}_\Psi$, $H_\Psi^\alpha(X)$ is defined as the unique solution of the equation

$$\mathbb{E}\left(\Psi\left(\frac{X}{H_\Psi^\alpha(X)}\right)\right) = 1-\alpha,$$

with $H_\Psi^\alpha(0) := 0$ (see Haezendonck and Goovaerts (1982), Goovaerts et al. (2004), and Bellini and Gianin (2008a)). For $X \in \mathbb{X}_\Psi$, following Bellini and Gianin (2008a) and Goovaerts et al. (2004), we define the Haezendonck risk measure at level $\alpha \in [0,1)$, denoted by $\pi_\Psi^\alpha(X)$, as

$$\pi_\Psi^\alpha(X) := \inf_{x\in\mathbb{R}} \left(H_\Psi^\alpha(X-x)+x\right). \tag{2}$$

For $X \in L^\infty$, Proposition 16 of Bellini and Gianin (2008a) shows that the above infimum is attained for $\alpha \in (0,1)$; their argument is easily extended to $\mathbb{X}_\Psi$. Moreover, examples exists where this infimum is not attained when $\alpha = 0$. Along the lines of Example 15 of Bellini and Gianin (2008a), one such example is when $\Psi(\cdot)$ and $F(\cdot)$ are defined as

$$\Psi(x) = \begin{cases} 0, & x<0; \\ x^{2k}, & \text{otherwise}; \end{cases}, \quad \text{where } k \geq 1, \text{ and } F(x) = \begin{cases} 0, & x<-1; \\ \frac{1}{2}, & -1 \leq x < 1; \\ 1, & \text{otherwise}; \end{cases}$$

For this reason, and also since for risk management purposes it is only the high values of α that are of interest, in the following we will restrict our attention to $\alpha \in (0,1)$ when working with Haezendonck risk measures .

For convenience we define $\pi_\Psi^\alpha(X;\cdot)$ as

$$\pi_\Psi^\alpha(X;x) := \left(H_\Psi^\alpha(X-x)+x\right), \quad x \in \mathbb{R}. \tag{3}$$

and denote by $\mathcal{I}_\Psi^\alpha(F)$ the set of minimizers of $\pi_\Psi^\alpha(F,\cdot)$. In Bellini and Gianin (2008a) it is shown that $\pi_\Psi^\alpha(X;\cdot)$ is convex function for $X \in L^\infty$, and we note that this result too can be easily extended to $X \in \mathbb{X}_\Psi$. This extension in particular implies that $\mathcal{I}_\Psi^\alpha(F)$ is a closed interval.

From now on $X, X_1, X_2, \ldots$ will denote a sequence of identically distributed random variables on our underlying probability space $(\Omega, \mathcal{F}, P)$ with $X \in \mathbb{X}_\Psi$. Also, $F(\cdot)$ we will denote their common distribution function. Note that by the definition of the Haezendonck risk measure, we could define $H_\Psi^\alpha(F)$, $\pi_\Psi^\alpha(F)$ and $\pi_\Psi^\alpha(F;\cdot)$ to equal $H_\Psi^\alpha(X)$, $\pi_\Psi^\alpha(X)$ and $\pi_\Psi^\alpha(X;\cdot)$, respectively. By $F_n(\cdot)$ we will denote the empirical distribution function of the random sample of size n consisting of $X_1, \ldots, X_n$, *i.e.*

$$F_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}_{(-\infty,\mathrm{x}]}\left(\mathrm{X_i}\right), \quad \mathrm{x} \in \mathbb{R}.$$

We denote by $\mathbb{E}_n\left(g(Y)\right)$ the expectation of $g(Y)$ with $Y \sim F_n$. As for such Y we have $Y \in L^\infty$ $(\subseteq \mathbb{X}_\Psi)$, $H_\Psi^\alpha(F_n)$ and $\pi_\Psi^\alpha(F_n)$ are both well defined, and moreover are easily seen to be a random variable defined on $(\Omega, \mathcal{F}, P)$. Note that $H_\Psi^\alpha(F_n)$ and $\pi_\Psi^\alpha(F_n)$ are natural (plug-in type) non-parametric estimators for $H_\Psi^\alpha(F)$ and $\pi_\Psi^\alpha(F)$, respectively; we refer to them as the empirical Orlicz premium and the empirical Haezendonck risk measure, respectively. Also, for a sequence of random variables $\{Z_i\}_{i\geq 1}$, by $Z_n \xrightarrow{\mathrm{d}} Z$ we denote the convergence in distribution or weak convergence of $\{Z_i\}_{i\geq 1}$ to Z, as n tends to infinity. For $x \in \mathbb{R}$, $(x)_+$

equals x for non-negative x, and equals zero for negative x.

Illustrative Example 1. An example of an Young function for which the Haezendonck risk measure of the empirical distribution function is only implicitly defined is given by

$$\Psi(x) = \begin{cases} 0, & x < 0; \\ \frac{\exp\{\beta x\}-1}{\exp\{\beta\}-1}, & \text{otherwise}; \end{cases}, \quad \text{for } \beta > 0. \tag{4}$$

This Young function was considered in Bellini and Gianin (2008a), and from definition it is easy to see that

$$H_\Psi^\alpha(Y-y) = \frac{\beta}{M_y^{-1}\left(1+(1-\alpha)(\exp\{\beta\}-1)\right)}$$

where $M_y(\cdot)$ is the moment generating function of $(Y-y)_+$ for $Y \sim F_n$. Hence, $\pi_\Psi^\alpha(F_n)$ equals

$$\inf_{y\in\mathbb{R}} y + \frac{\beta}{M_y^{-1}\left(1+(1-\alpha)(\exp\{\beta\}-1)\right)},$$

an optimal value that clearly lacks a closed form expression. In Example 5 below we derive the asymptotic normal distribution for $\pi_\Psi^\alpha(F_n)$ when F is an exponential distribution. □

Illustrative Example 2. Now we present a rather simple example to demonstrate that non-normal limits for the empirical Haezendonck risk measure arise quite naturally. Let F be a Bernoulli distribution, $\Psi(\cdot)$ be defined by

$$\Psi(x) = \begin{cases} 0, & x < 0; \\ x, & 0 \le x \le 1; \\ 2x-1, & \text{otherwise}; \end{cases}, \tag{5}$$

and $\alpha = 50\%$. This piecewise linear Young function, non-differentiable at 1, was mentioned in Bellini and Gianin (2008b) as an example of a non-differentiable Young function which fails to satisfy the conditions of their weak convergence result for the Orlicz premium. For $Y \sim F_n$, some straightforward calculations lead to the following form for $H_\Psi^\alpha(Y-x)+x$: For $F_n(0) \ge 1/2$ we have

$$x + H_\Psi^\alpha(Y-x) = \begin{cases} 2(1-F_n(0)) - x, & x < (1-2F_n(0)); \\ \frac{4(1-F_n(0))}{3-2F_n(0)} - \left(\frac{1}{3-2F_n(0)}\right)x, & (1-2F_n(0)) \le x < 0; \\ \frac{4(1-F_n(0))}{3-2F_n(0)} + \left(\frac{2F_n(0)-1}{3-2F_n(0)}\right)x, & 0 \le x \le 1; \\ x, & x > 1; \end{cases}, \tag{6}$$

and for $F_n(0) < 1/2$ we have

$$x + H_\Psi^\alpha(Y-x) = \begin{cases} 2(1-F_n(0)) - x, & x < 0; \\ 2(1-F_n(0)) + (2F_n(0)-1)\,x, & 0 \le x \le 1; \\ x, & x > 1; \end{cases}. \tag{7}$$

From (6) and (7) it follows that

$$\pi_\Psi^\alpha(F_n) = \begin{cases} 1, & F_n(0) \le 1/2; \\ 1 - \left(\frac{2F_n(0)-1}{3-2F_n(0)}\right), & F_n(0) > 1/2; \end{cases}. \tag{8}$$

From (8), and observing that both $F_n(\cdot)$ and $F(\cdot)$ are Bernoulli distributions, we have using the strong law of large numbers that with probability one, for large n

$$\sqrt{n}\left(\pi_\Psi^\alpha(F_n) - \pi_\Psi^\alpha(F)\right) = \begin{cases} 0, & F(0) < 1/2; \\ -\sqrt{n}\left(\frac{F_n(0)-1/2}{3/2-F_n(0)}\right)_+, & F(0) = 1/2; \\ \frac{\sqrt{n}(F(0)-F_n(0))}{(3/2-F(0))(3/2-F_n(0))}, & F(0) > 1/2; \end{cases}.$$

This with Slutsky's lemma, and the ordinary central limit theorem implies that

$$\sqrt{n}\left(\pi_\Psi^\alpha(F_n) - \pi_\Psi^\alpha(F)\right) \xrightarrow{\text{d}} \begin{cases} 0, & F(0) < 1/2; \\ -\left[\frac{1}{2}\right](Z)_+, & F(0) = 1/2; \\ \left[\frac{\sqrt{F(0)(1-F(0))}}{(3/2-F(0))^2}\right] Z, & F(0) > 1/2; \end{cases}, \tag{9}$$

where Z is a standard normal random variable. Hence a non-normal limit results in this simple Bernoulli example with a piecewise linear Young function whenever $F(0) \le 1/2$. In fact, in Example 3 below we show that the above result follows as an application of our weak convergence result for the empirical Haezendonck risk measure. □

In the next section we establish the strong consistency of the empirical Haezendonck risk measure. In the following section we prove a weak convergence result for this estimator, and provide some examples to illustrate the asymptotic behavior of this estimator both under and without our assumptions. Also, in a parametric example we compare the performance of this non-parametric estimator with the parametric Maximum Likelihood Estimator (MLE). In section 4 we report the results from a simulation study to lend insight into the sample sizes required for the asymptotic limits to take hold. Section 5 provides a brief discussion of our results.

2. Strong Consistency

The main goal of this section is to establish the consistency of the empirical Haezendonck risk measure as an estimator for the Haezendonck risk measure. To motivate the formulation of our result we consider the case of the T-VaR which is the Haezendonck risk measure with the identity as the Young function. To do this we define the α-level VaR of $F(\cdot)$, denoted by $q_\alpha(F)$, by

$$q_\alpha(F) := \inf\{x \in \mathbb{R} : F(x) \ge \alpha\}. \tag{10}$$

Also, following Rockafellar and Uryasev (2002) we define the α-level upper VaR of of $F(\cdot)$, denoted by $q_\alpha^+(F)$, by

$$q_\alpha^+(F) := \inf\{x \in \mathbb{R} : F(x) > \alpha\}. \tag{11}$$

Observe that $q_\alpha(F) \le q_\alpha^+(F)$ with equality if and only if $F(\cdot)$ is strictly increasing at $q_\alpha(F)$. We note that in the case of the α-level T-VaR $\mathcal{I}_\Psi^\alpha(F)$ equals $[q_\alpha(F), q_\alpha^+(F)]$, the set of α-level quantiles; and in the case that $q_\alpha(F)$ is a point of strict increase of $F(\cdot)$, $\mathcal{I}_\Psi^\alpha(F) = \{q_\alpha(F)\}$. Partly motivated by this example, of recent, $\mathcal{I}_\Psi^\alpha(F)$ is gaining in interest. Hence the consistency, with respect to a suitable distance, of $\mathcal{I}_\Psi^\alpha(F_n)$ is also of interest. A natural candidate is the asymmetric distance $d(\cdot,\cdot)$ defined by

$$d(A,B) := \sup_{a\in A}\inf_{b\in B}|a-b|, \tag{12}$$

for any two subsets A and B of $\mathbb{R}$. A symmetrized version of this distance yields the Hausdorff metric given by

$$d_H(A,B) := \max\left(\sup_{a\in A}\inf_{b\in B}|a-b|, \sup_{b\in B}\inf_{a\in A}|a-b|\right) \tag{13}$$

The following example shows the *convergence* of $\mathcal{I}_\Psi^\alpha(F_n)$ to $\mathcal{I}_\Psi^\alpha(F)$ may hold only in the asymmetric distance $d(\cdot,\cdot)$ and not in the Hausdorff metric.

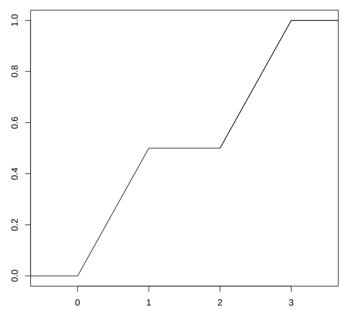


Figure 1: Distribution for which $d_H\left(\mathcal{I}_\Psi^\alpha(F_n), \mathcal{I}_\Psi^\alpha(F)\right) \nrightarrow 0$

Example 1
In this example $F(\cdot)$ is taken to be the equal mixture of $U(0,1)$ and $U(2,3)$; Figure 1 contains a graph of this distribution. For $\alpha = 50\%$ it is easily checked that $q_\alpha(F) = 1$ and $q_\alpha^+(F) = 2$. Hence for the case of the T-VaR risk measure at the 50%-level, *i.e.* with $\Psi(x) = x$ for $x \geq 0$, $\mathcal{I}_\Psi^\alpha(F) = [1,2]$. For a random sample from F with sample size n it is easily checked that

$$\mathcal{I}_\Psi^\alpha(F_n) = \begin{cases} \{X_{(n+1)/2:n}\}, & n \text{ odd;} \\ [X_{n/2:n}, X_{(n/2+1):n}], & \text{otherwise;} \end{cases}.$$

Hence it is clear that for this F

$$\limsup_{n\to\infty} d_H\left(\mathcal{I}_\Psi^\alpha(F_n), \mathcal{I}_\Psi^\alpha(F)\right) = 1, \quad \text{a.s. } P.$$

Nevertheless, it is also easy to show that

$$\lim_{n\to\infty} d\left(\mathcal{I}_\Psi^\alpha(F_n), \mathcal{I}_\Psi^\alpha(F)\right) = 0, \quad \text{a.s. } P.$$

□

The above example demonstrates that while expecting convergence of $\mathcal{I}_\Psi^\alpha(F_n)$ to $\mathcal{I}_\Psi^\alpha(F)$ in the Hausdorff metric is unrealistic, the asymmetric distance $d(\cdot,\cdot)$ between $\mathcal{I}_\Psi^\alpha(F_n)$ and $\mathcal{I}_\Psi^\alpha(F)$ may converge to zero. The following theorem establishes this latter statement, and using it proves the consistency the empirical Haezendonck risk measure.

Theorem 1. *For $F \in \mathbb{X}_\Psi$, and $\alpha \in (0,1)$ we have with probability one,*

$$\lim_{n\to\infty} d\left(\mathcal{I}_\Psi^\alpha(F_n), \mathcal{I}_\Psi^\alpha(F)\right) = 0, \quad \textit{and} \quad \lim_{n\to\infty} \pi_\Psi^\alpha(F_n) = \pi_\Psi^\alpha(F).$$

Proof. We start by observing that from the consistency of the Orlicz norm (see Bellini and Gianin (2008b)) we have the almost sure pointwise convergence of $\pi_\Psi^\alpha(F_n,\cdot)$ to $\pi_\Psi^\alpha(F,\cdot)$. Moreover, since $\pi_\Psi^\alpha(F_n,\cdot)$ is convex, Theorem 10.8 of Rockafellar (1997) implies that $\pi_\Psi^\alpha(F_n,\cdot)$ converges to $\pi_\Psi^\alpha(F,\cdot)$ uniformly on compacts with probability one.

As observed earlier, $\mathcal{I}_\Psi^\alpha(F)$ is a closed interval. Moreover, using the fact that $X \in \mathbb{X}_\Psi$ and an argument similar to that of Proposition 16 of Bellini and Gianin (2008a) it is easily shown that $\mathcal{I}_\Psi^\alpha(F)$ is compact as well. Hence we denote $\mathcal{I}_\Psi^\alpha(F)$ by $[x_l^*, x_u^*]$, $-\infty < x_l^* \leq x_u^* < \infty$.

It suffices for the first assertion to show that with probability one,

$$x_l^* \leq \liminf_{n\to\infty} \left(\inf \mathcal{I}_\Psi^\alpha(F_n)\right) \leq \limsup_{n\to\infty} \left(\sup \mathcal{I}_\Psi^\alpha(F_n)\right) \leq x_u^*. \tag{14}$$

This follows from almost sure pointwise convergence of $\pi_\Psi^\alpha(F_n,\cdot)$ and the fact that for any $\epsilon > 0$ we have

$$\pi_\Psi^\alpha(F, x_l^*) = \pi_\Psi^\alpha(F, x_u^*) < \min\left(\pi_\Psi^\alpha(F, x_u^* + \epsilon), \pi_\Psi^\alpha(F, x_l^* - \epsilon)\right).$$

The second assertion now follows from the first assertion and almost sure convergence on compacts of $\pi_\Psi^\alpha(F_n,\cdot)$. □

We note that in the case that $\mathcal{I}_\Psi^\alpha(F)$ is a single point, from Theorem 1 we have that with probability one,

$$\lim_{n\to\infty} d_H\left(\mathcal{I}_\Psi^\alpha(F_n), \mathcal{I}_\Psi^\alpha(F)\right) = 0.$$

3. Weak Convergence Results

Our approach towards deriving the weak convergence limits of the empirical Haezendonck risk measure involves viewing Haezendonck risk measures as the optimal value of certain convex programming problems, and then applying the functional delta method. For details of such an approach we refer to Shapiro (1991), and Chapter 6 of Rubinstein and Shapiro (1993). Now we present an outline of this section. First, we discuss the assumptions that we need for our weak convergence result for the empirical Haezendonck risk measure. Second, we present the details of our approach before stating and proving this weak limit result. Third, we present some examples to demonstrate the generality of our result, and present an example in which $F(\cdot)$ is embedded in a parametric family of distributions which allows for comparison of the empirical Haezendonck risk measure with the efficient parametric MLE. We end this section with a weak convergence result for the Orlicz premium using the approach mentioned above; we present this result as it is a more general version of the one stated in Bellini and Gianin (2008b).

For the weak convergence theorem for $\pi_\Psi^\alpha(F_n)$, apart from requiring that $F(\cdot)$ belongs to $\mathbb{X}_\Psi$, we further require that $F(\cdot)$ satisfies the following conditions that will be referred to as Assumption C:

Assumption C

C1. $\pi_\Psi^\alpha(F)$ is strictly less than $\operatorname{ess\,sup}(F)$

C2. $F \in \mathbb{X}_\Psi$ and furthermore satisfies

$$\mathbb{E}\left(\left[\Psi\left(\frac{X}{\delta_l}\right)\right]^2\right) < \infty \tag{15}$$

where $\delta_l > 0$ is such that

$$\delta_l < \inf\left\{\pi_\Psi^\alpha(F,x) - x | \pi_\Psi^\alpha(F,x) = \pi_\Psi^\alpha(F)\right\}.$$

When $\operatorname{ess\,sup}(F) = \infty$, which is usually the case in risk management applications, Assumption C1 essentially requires a finite $\pi_\Psi^\alpha(F)$. Hence, in this case Assumption C1 is not restrictive. Later in this section we discuss Assumption C1 in the case when $\pi_\Psi^\alpha(F) = \operatorname{ess\,sup}(F) < \infty$. Assumption C2 is analogous to the requirement of finite second moment for the ordinary central limit theorem, and in this sense it is an appropriate requirement for the $\sqrt{n}$ rate of convergence. Moreover, in the case of T-VaR it is easily seen- to be a necessary requirement as well. It is worth mention that while our weak convergence result addresses situations where we have weak convergence at the $\sqrt{n}$ rate, there exist examples of $F(\cdot)$ with sufficiently fat tails such that $\pi_\Psi^\alpha(F_n)$ converges to $\pi_\Psi^\alpha(F)$ at much slower rates. The following example is one such.

Example 2

Let $\alpha = 0.5$, and $\Psi(\cdot)$ be given by

$$\Psi(x) = \begin{cases} 0, & x < 0; \\ x, & \text{otherwise}; \end{cases}.$$

Recall that this definition corresponds to the T-VaR risk measure at the 50% level. Let F be the symmetric (about zero) distribution such that

$$\Pr\left(|X| > x\right) = \begin{cases} 1 - \frac{x}{2}, & 0 \le x < 1; \\ \frac{1}{2x^\beta}, & \text{otherwise}; \end{cases},$$

where $1 < \beta < 2$. It is easy to check that for these specifications

$$\mathcal{M}_\Psi^\alpha(F) = \left\{\left(0, \frac{3\beta - 1}{4(\beta - 1)}\right)\right\}, \quad \text{and} \quad \pi_\Psi^\alpha(F) = \frac{3\beta - 1}{4(\beta - 1)}.$$

For the sake of expositional ease let the sample size be $2n + 1$, for some $n \ge 1$. Then the empirical T-VaR at the 50% level is given by

$$\begin{aligned}\left(\frac{1}{2n+1}\right)\left[X_{(n+1):(2n+1)} + 2\sum_{i=n+2}^{2n+1} X_{i:(2n+1)}\right] &= \left(\frac{1}{2n+1}\right)\left[X_{(n+1):(2n+1)} - \left(\frac{1}{n}\right)\sum_{i=n+2}^{2n+1} X_{i:(2n+1)}\right] \\ &\quad + \frac{1}{n}\sum_{i=n+2}^{2n+1} X_{i:(2n+1)} \\ &= O\left(\frac{1}{n}\right) + \frac{1}{n}\sum_{i=n+2}^{2n+1} X_{i:(2n+1)},\end{aligned}$$

where the last step follows from the ordinary strong law of large numbers. Let N denote the random variable defined by

$$N := \sum_{i=1}^{2n+1} I_{X_i \geq 0},$$

which is clearly distributed as $\mathrm{Bin}(2n+1, 1/2)$. We note that

$$\left| \sum_{i=n+2}^{2n+1} X_{i:(2n+1)} - \sum_{i=1}^{2n+1} X_i I_{X_i \geq 0} \right| \leq \begin{cases} \sum_{i=n+2}^{2n+1-N} X_{i:(2n+1)}, & N < n; \\ 0, & N = n; \\ \sum_{i=2n+2-N}^{n+1} X_{i:(2n+1)}, & \text{otherwise}; \end{cases} .$$

Using this bound, the fact that $N - n = O(\sqrt{n})$, and that $\lim_{n\to\infty} \Pr\left(|X_{(n+1):(2n+1)}| > 1\right) = 0$, we have

$$\left| \sum_{i=n+2}^{2n+1} X_{i:(2n+1)} - \sum_{i=1}^{2n+1} X_i I_{X_i \geq 0} \right| = O(\sqrt{n}) \tag{16}$$

It is easy to show (for example using Theorem 7.7 of Durrett (2005)) that

$$n^{1-\frac{1}{\beta}} \left(\frac{1}{n} \sum_{i=1}^{2n+1} X_i I_{X_i \geq 0} - \frac{3\beta - 1}{4(\beta - 1)} \right) \xrightarrow{\mathrm{d}} Z, \tag{17}$$

for some non-degenerate random variable Z. Combining (16) with (17) and using the fact that $1 < \beta < 2$ we have

$$n^{1-\frac{1}{\beta}} \left(\pi_\Psi^\alpha(F_n) - \pi_\Psi^\alpha(F) \right) = n^{1-\frac{1}{\beta}} \left(\frac{1}{n} \sum_{i=n+2}^{2n+1} X_{i:(2n+1)} - \frac{3\beta - 1}{4(\beta - 1)} \right) \xrightarrow{\mathrm{d}} Z.$$

□

In the rest of this section we will suppose that $F(\cdot)$ and $\Psi(\cdot)$ satisfy Assumption C, unless mentioned otherwise. The key idea behind our approach to establishing the weak convergence of $\pi_\Psi^\alpha(F_n)$ is to formulate $\pi_\Psi^\alpha(F)$ as the optimal value of a convex programming problem. To this end note that $H_\Psi^\alpha(X)$ is easily seen to satisfy

$$H_\Psi^\alpha(X) = \inf \left\{ s \middle| \mathbb{E}\left(\Psi\left(\frac{X}{s} \right) \right) \leq 1 - \alpha \right\}.$$

Using this observation we note that $\pi_\Psi^\alpha(F)$ is the optimal value of the mathematical programming problem given by

$$\begin{aligned} &\text{minimize} && \theta_1 + \theta_2, && (\theta_1, \theta_2) \in \mathbb{R} \times (0, \infty) \\ &\text{subject to} && \mathbb{E}\left(\Psi\left(\frac{X - \theta_1}{\theta_2} \right) \right) - (1 - \alpha) \leq 0. \end{aligned}$$

For convenience we denote the coordinates of a vector $\tilde{\theta}$ in $\mathbb{R}^2$ by θ_1 and θ_2, *i.e.* $\tilde{\theta} = (\theta_1, \theta_2)$. Now we note that the function $\chi(\cdot)$ from $\mathbb{R} \times (0, \infty)$ to $\mathbb{R}_+$ defined by

$$\chi(\tilde{\theta}) := \mathbb{E}\left(\Psi\left(\frac{X - \theta_1}{\theta_2} \right) \right) - (1 - \alpha),$$

is easily seen to satisfy

$$\chi(p\tilde{\theta} + (1-p)\tilde{\theta}') \leq \max\{\chi(\tilde{\theta}), \chi(\tilde{\theta}')\}, \quad \tilde{\theta}, \tilde{\theta}' \in \mathbb{R} \times (0, \infty),\ p \in [0,1].$$

This implies that $\chi(\cdot)$ is a quasi-convex function; we refer to section 3.4 of Boyd and Vandenberghe (2004) for a discussion of quasi-convexity. We note that examples exist where $\chi(\cdot)$ is *not* convex. Nevertheless, quasi-convexity preserves convexity of sub-level sets, and hence

$$\left\{\tilde{\theta} | \chi(\tilde{\theta}) \leq 0\right\}$$

is a convex set. This implies, trivially, the existence of a *convex* function $\eta(\cdot)$ with a sub-level set coinciding with that of $\chi(\cdot)$ given above. But we seek a *nice* such function, and one such candidate is defined by

$$\eta(\tilde{\theta}) := \theta_2 \left[\mathbb{E}\left(\Psi\left(\frac{X - \theta_1}{\theta_2}\right)\right) - (1-\alpha)\right], \quad \tilde{\theta} \in \mathbb{R} \times (0, \infty).$$

This candidate is motivated by the fact that the perspective function of a convex function is convex (see for example section 3.2.6 of Boyd and Vandenberghe (2004)). Hence, $\pi_\Psi^\alpha(F)$ is the optimal value of the convex programming problem ($\mathbf{P}_\infty$) given by

$$\begin{array}{lll} \text{minimize} & \theta_1 + \theta_2, & (\theta_1, \theta_2) \in \mathbb{R} \times (0, \infty) \\ \text{subject to} & \theta_2 \left[\mathbb{E}\left(\Psi\left(\frac{X - \theta_1}{\theta_2}\right)\right) - (1-\alpha)\right] \leq 0. & \end{array} \qquad (\mathbf{P}_\infty)$$

Let $\mathcal{M}_\Psi^\alpha(F)$ denote the set of minimizers of the programming problem ($\mathbf{P}_\infty$). Recall that $H_\Psi^\alpha(X - z)$ as a function in z is convex; this is easily checked to imply that $\mathcal{M}_\Psi^\alpha(F)$ is a line segment in $\mathbb{R} \times (0, \infty)$. We note that Assumption C further implies that $\mathcal{M}_\Psi^\alpha(F)$ is a closed set contained in an open rectangle $(x_l, x_u) \times (\delta_l, \delta_u)$, for some $-\infty < x_l < x_u < \infty$, and $0 < \delta_l < \delta_u < \infty$. In the case $\pi_\Psi^\alpha(F)$ is finite and equals $\operatorname{ess\,sup}(F)$, $(\operatorname{ess\,sup}(F), 0)$ is clearly an optimal solution, and in some cases (as in Example 3 below) it can moreover be the unique optimal solution. This in general clearly creates problems for the above representation for $\pi_\Psi^\alpha(F)$. Nevertheless, in specific cases (such as piecewise linear $\Psi(\cdot)$) this does not cause any issues; Example 3 below is one such example. But for expositional ease, we state our weak convergence result by excluding this case, which as mentioned before is unlikely to occur in risk management applications.

The above representation for $\pi_\Psi^\alpha(F)$ implies that $\pi_\Psi^\alpha(F_n)$ is the optimal value of the convex programming problem ($\hat{\mathbf{P}}_n$) given by

$$\begin{array}{lll} \text{minimize} & \theta_1 + \theta_2, & (\theta_1, \theta_2) \in \mathbb{R} \times (0, \infty) \\ \text{subject to} & \mathbb{E}_n\left(\psi\left(\tilde{\theta}, Y\right)\right) \leq 0. & \end{array} \qquad (\hat{\mathbf{P}}_n)$$

where

$$\psi(\tilde{\theta}, x) := \theta_2 \left(\Psi\left(\frac{x - \theta_1}{\theta_2}\right) - (1-\alpha)\right), \quad \tilde{\theta} \in \mathbb{R} \times (0, \infty).$$

The above representations for $\pi_\Psi^\alpha(F)$ (and $\pi_\Psi^\alpha(F_n)$) in turn allow us to use the functional delta method as explained briefly below. Note that the convex programming problem ($\mathbf{P}_\infty$) is fully specified by the constraint function $\eta(\cdot)$, which is replaced in ($\hat{\mathbf{P}}_n$) by its sample analog. Hence if the optimal value is an appropriately differentiable functional of the constraint function, and the constraint function in $(\hat{\mathbf{P}}_n)$, $\mathbb{E}_n(\psi(\cdot, Y))$, converges at the $\sqrt{n}$ rate to $\eta(\cdot)$, then $\pi_\Psi^\alpha(F_n)$ converges to $\pi_\Psi^\alpha(F)$ at the $\sqrt{n}$ rate as well. Moreover, the weak limit of $\mathbb{E}_n(\psi(\cdot, Y))$ and the differential then determine the weak limit of $\pi_\Psi^\alpha(F_n)$. This is the approach that is used to prove the following theorem, our weak convergence result for $\pi_\Psi^\alpha(F_n)$.

Theorem 2. *For $\alpha \in (0,1)$, and $\Psi(\cdot)$ and $F(\cdot)$ satisfying Assumption C we have*

$$\sqrt{n}\left(\pi_\Psi^\alpha(F_n) - \pi_\Psi^\alpha(F)\right) \xrightarrow{\text{d}} \min_{\tilde{\theta}\in\mathcal{M}_\Psi^\alpha(F)} \max_{\lambda\in\Lambda} \lambda Z\left(\tilde{\theta}\right), \tag{18}$$

where Λ is the set of Lagrange multipliers for the convex programming problem $(\mathbf{P}_\infty)$, and $Z(\cdot)$ is a mean zero Gaussian process on $S(F)$ with covariance given by

$$\operatorname{Cov}\left(Z\left(\tilde{\theta}\right), Z\left(\tilde{\theta}'\right)\right) = \operatorname{Cov}\left(\psi\left(\tilde{\theta}, X\right), \psi\left(\tilde{\theta}', X\right)\right), \quad \tilde{\theta} \in S(F), \tag{19}$$

where $S(F) := [x_l, x_u] \times [\delta_l, \delta_u]$.

Proof. Consistency of $\mathcal{M}_\Psi^\alpha(F_n)$ as stated in Theorem 1 implies that with probability one, for large n, $\mathcal{M}_\Psi^\alpha(F_n) \subseteq (x_l, x_u)\times(\delta_l, \delta_u)$. This implies that with probability one, for large n, the programming problem $(\hat{\mathbf{P}}_n)$ will yield the same optimal value as the convex programming problem $(\hat{\mathbf{P}}_n^*)$ given by

$$\begin{aligned} &\text{minimize} && \theta_1 + \theta_2, && (\theta_1, \theta_2) \in S(F) \\ &\text{subject to} && \theta_2\left[\mathbb{E}_n\left(\psi\left(\tilde{\theta}, Y\right)\right) - (1-\alpha)\right] \le 0. \end{aligned} \tag{$\hat{\mathbf{P}}_n^*$}$$

This implies that the weak convergence result for the optimal value of $(\hat{\mathbf{P}}_n^*)$ coincides with that of $(\hat{\mathbf{P}}_n)$. Henceforth, for simplicity, we will denote the optimal value of $(\hat{\mathbf{P}}_n^*)$ also by $\pi_\Psi^\alpha(F_n)$.

We note that the objective function in $\left(\hat{\mathbf{P}}_n^*\right)$ is deterministic and, in particular, coincides with that of $(\mathbf{P}_\infty)$. Also note that (15), Lemma 1, and Theorem A.3 of King and Rockafellar (1990) together imply that

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \psi(\cdot, X_i) - \mathbb{E}(\psi(\cdot, X))\right) \xrightarrow{\text{d}} Z(\cdot) \tag{20}$$

on the space $C(S(F))$ of bounded continuous real valued functions on $S(F)$, where $Z(\cdot)$ is a mean zero Gaussian process on $S(F)$ with covariance as given in (19).

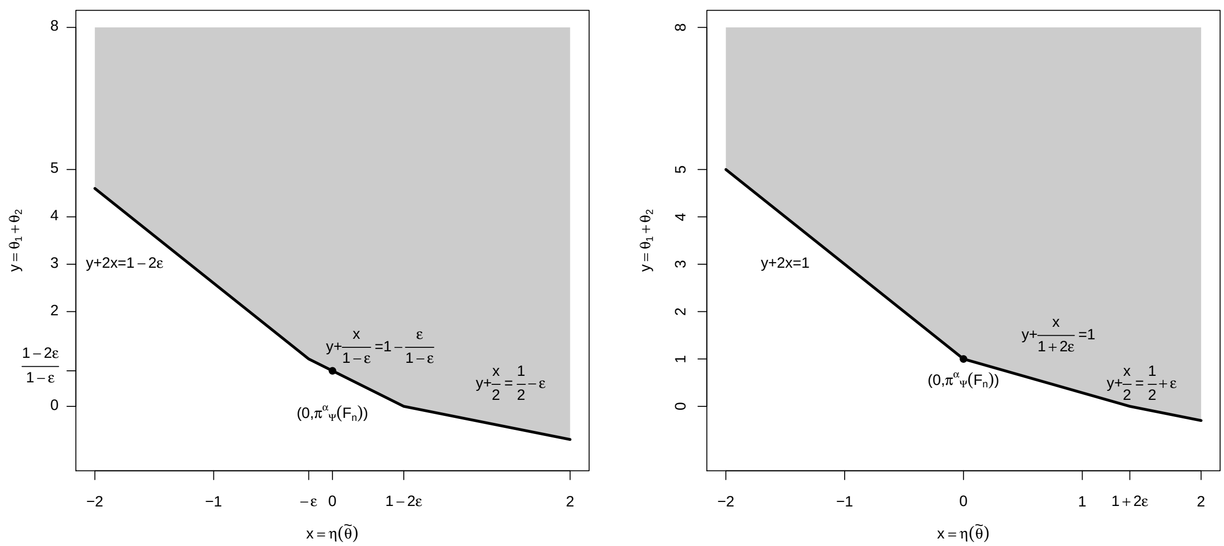
The rest of the argument rests on the above discussed formulation of $\pi_\Psi^\alpha(F_n)$ and $\pi_\Psi^\alpha(F)$ as optimal values of convex programming problems $\left(\hat{\mathbf{P}}_n^*\right)$ and $(\mathbf{P}_\infty)$, respectively, and the delta-method for convex programming problems as stated in Theorem 3.5 of Shapiro (1991). All that remains to be checked then is that the Slater condition is satisfied for $(\mathbf{P}_\infty)$, but that follows from (15) as a simple application of the dominated convergence theorem. □

We now provide examples to demonstrate the generality of Theorem 2. The following example is a continuation of *Illustrative Example 2* of the introduction that involves a Young function which is not differentiable at a point in $(0, \infty)$, and an $F(\cdot)$ such that neither $\mathcal{M}_\Psi^\alpha(F)$ nor Λ are singletons; in particular this results in the limiting distribution being non-normal. For this example we will find it convenient to define $\mathcal{A} \subseteq \mathbb{R}^2$ as

$$\mathcal{A} := \{(u,t) | \exists\, \tilde{\theta} \in \mathbb{R}\times(0,\infty), \eta(\tilde{\theta}) \le u, \theta_1 + \theta_2 \le t\}.$$

We note that $\mathcal{A}$ is a variation of the epigraph (see Boyd and Vandenberghe (2004)) for the convex programming problem $\mathbf{P}_\infty$ with the property that

$$\pi_\Psi^\alpha(F) = \inf\{t|(0,t) \in \mathcal{A}\}.$$



(a) $F(0) = 1/2 + \epsilon, 0 \leq \epsilon \leq 1/2$ (b) $F(0) = 1/2 - \epsilon, 0 \leq \epsilon \leq 1/2$

Figure 2: Epigraph of the Convex Programming Problem associated with the Computation of $\pi_\Psi^\alpha(F)$

Example 3

We continue here with *Illustrative Example 2* of the introduction with $F(\cdot)$ denoting the Bernoulli distribution, α equals 50%, and $\Psi(\cdot)$ as defined in (5). For this Young function it follows that

$$\eta(\tilde{\theta}) = \mathbb{E}\left((X-\theta_1)_+\right) + \mathbb{E}\left((X-\theta_1-\theta_2)_+\right) - (1-\alpha)\theta_2. \tag{21}$$

Figure 2 plots $\mathcal{A}$ (the shaded region) using the above and somewhat tedious calculations. From Figure 2 it follows that

$$\pi_\Psi^\alpha(F) = \begin{cases} 1, & F(0) \leq 1/2; \\ 1 - \frac{\epsilon}{1-\epsilon}, & F(0) = 1/2 + \epsilon > 1/2; \end{cases} \quad \text{and that} \quad \Lambda = \begin{cases} \left[\frac{1}{1+2\epsilon}, 2\right], & F(0) = 1/2 - \epsilon < 1/2; \\ [1,2], & F(0) = 1/2; \\ \frac{1}{1-\epsilon}, & F(0) = 1/2 + \epsilon > 1/2; \end{cases}.$$

Moreover, from (21) it follows that

$$\mathcal{M}^{\alpha}_{\Psi}(F) = \begin{cases} \{(1,0)\}, & F(0) = 1/2 - \epsilon < 1/2; \\ \{\tilde{\theta} |\ 0 \le \theta_1 \le 1;\ \theta_2 = 1 - \theta_1\}, & F(0) = 1/2; \\ \left\{ \left(0, 1 - \frac{\epsilon}{1-\epsilon}\right) \right\}, & F(0) = 1/2 + \epsilon > 1/2; \end{cases}.$$

Since $F(\cdot)$ and $F_n(\cdot)$ are both Bernoulli distributions, we note that we had essentially derived in (8) an expression for $\pi^{\alpha}_{\Psi}(F)$ directly from the definition of the Haezendonck risk measure.

Even though Assumption C is violated here, due to $\pi^{\alpha}_{\Psi}(F)$ being equal to $\operatorname{ess\,sup}(F)$, the representation of $\eta(\cdot)$ in (21) allows one to nevertheless easily establish a version of Lemma 1 so that the proof of Theorem 2 goes through. It is straightforward to check that the process $Z(\cdot)$ of Theorem 2 satisfies,

$$\operatorname{Cov}(Z(\theta), Z(\theta')) = F(0)(1 - F(0))(2 - 2\theta_1 - \theta_2)(2 - 2\theta_1' - \theta_2'), \quad \forall \tilde{\theta}, \tilde{\theta}' \in \mathbb{R}^+ \times \mathbb{R}^+$$

In particular, this implies that

$$\min_{\tilde{\theta} \in \mathcal{M}^{\alpha}_{\Psi}(F)} \max_{\lambda \in \Lambda} \lambda Z\left(\tilde{\theta}\right) \stackrel{\mathrm{d}}{=} \sqrt{F(0)(1 - F(0))} \min_{\tilde{\theta} \in \mathcal{M}^{\alpha}_{\Psi}(F)} \max_{\lambda \in \Lambda} \lambda (2 - 2\theta_1 - \theta_2) Z \tag{22}$$

$$\stackrel{\mathrm{d}}{=} \begin{cases} 0, & F(0) < 1/2; \\ -\left[\frac{1}{2}\right](Z)_+, & F(0) = 1/2; \\ \frac{\sqrt{F(0)(1-F(0))}}{(3/2 - F(0))^2} Z, & F(0) > 1/2; \end{cases}. \tag{23}$$

where Z is a standard normal random variable, which agrees with (9) of the introduction, where it was derived using first principles. □

We now present a corollary for the case of $F(\cdot)$ without mass points and a Young function which is differentiable on $(0, \infty)$; in such situations it significantly simplifies application of Theorem 2.

Corollary 1. *For $\Psi(\cdot)$ differentiable on $(0, \infty)$, $\alpha \in (0, 1)$, $F(\cdot)$ a continuous function, and both $\Psi(\cdot)$ and $F(\cdot)$ satisfying Assumption C we have*

$$\sqrt{n}\left(\pi^{\alpha}_{\Psi}(F_n) - \pi^{\alpha}_{\Psi}(F)\right) \stackrel{\mathrm{d}}{\longrightarrow} \left[\mathbb{E}\left(\Psi'\left(\frac{X - \theta_1^*}{\theta_2^*}\right)\right)\right]^{-1} \min_{\tilde{\theta} \in \mathcal{M}^{\alpha}_{\Psi}(F)} Z\left(\tilde{\theta}\right), \tag{24}$$

where $\tilde{\theta}^*$ is any member of $\mathcal{M}^{\alpha}_{\Psi}(F)$, and $Z(\cdot)$ is as given in Theorem 2.

Proof. Clearly the objective function of the convex programming problem ($\mathbf{P}_{\infty}$) is differentiable; differentiability of the constraint $\eta(\cdot)$ follows from differentiability of $\Psi(\cdot)$ and (15). As the Slater condition is satisfied for ($\mathbf{P}_{\infty}$) (as observed in the proof of Theorem 2) we have the Karusch-Kuhn-Tucker conditions given by

$$\eta(\tilde{\theta}) = 0, \quad \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} \mathbb{E}\left(\Psi'\left(\frac{X - \theta_1}{\theta_2}\right)\right) \\ \mathbb{E}\left(\Psi'\left(\frac{X - \theta_1}{\theta_2}\right)\left[\frac{X - \theta_1}{\theta_2}\right]\right) \end{pmatrix} \tag{25}$$

are both necessary and sufficient conditions for optimality. The rest follows from observing that the set of Lagrange multipliers is invariant of the choice of optimal solution from $\mathcal{M}^{\alpha}_{\Psi}(F)$. □

The following example demonstrates that a non-normal limit can arise even in situations covered by Corollary 1 where Λ is a singleton. This of course happens when $\mathcal{M}_{\Psi}^{\alpha}(F)$ is not a singleton.

Example 4
In this example we continue to use the same definition of $\Psi(\cdot)$ and α as in Example 1 resulting in the 50%-level T-VaR as our choice of Haezendonck risk measure. The sampling distribution $F(\cdot)$ is the equal mixture of $U(0,1)$ and $U(2,3)$ as in Example 1, and is plotted in Figure 1. Observe that the median $q_\alpha(F) = 1$, and that the T-VaR at the 50%-level is 2.5. In fact the asymptotic distribution of the empirical T-VaR at the 50%-level for this choice of $F(\cdot)$ was derived in Ahn and Shyamalkumar (2011a) (using first principles) to be that of

$$\frac{Z_1}{\sqrt{6}} + 3^{I(Z_2<0)}\left(\frac{Z_2}{2}\right), \tag{26}$$

where Z_1, Z_2 are i.i.d. standard normal random variables. Moreover, in Ahn and Shyamalkumar (2011a) it is shown that the heuristics of influence function fails in this example. In the following we will derive the limiting distribution of the empirical T-VaR estimator using Corollary 1.

We begin by observing that easy calculations yield

$$H_{\Psi}^{\alpha}(X-\theta_1) = \begin{cases} 3-2\theta_1, & \theta_1 < 0; \\ \frac{\theta_1^2}{2} - 2\theta_1 + 3, & 0 \leq \theta_1 \leq 1; \\ \frac{5}{2} - \theta_1, & 1 < \theta_1 \leq 2; \\ \frac{(3-\theta_1)^2}{2}, & 2 < \theta_1 \leq 3; \\ 0, & \text{otherwise}; \end{cases},$$

which in turn implies that

$$\mathcal{M}_{\Psi}^{\alpha}(F) = \{(\theta_1, 5/2-\theta_1) : 1 \leq \theta_1 \leq 2\}.$$

It is also easily checked that the Lagrange multiplier λ equals 2, and that the process $Z(\cdot)$ of Corollary 1 satisfies,

$$\mathrm{Cov}(Z(\theta), Z(\theta')) = \frac{(5/2-\theta_1)(5/2-\theta_1')}{4} + \frac{1}{24}, \quad \forall \tilde{\theta}, \tilde{\theta}' \in \mathcal{M}_{\Psi}^{\alpha}(F).$$

An equivalent representation of $Z(\cdot)$ on $\mathcal{M}_{\Psi}^{\alpha}(F)$ is the following:

$$Z(\theta) = \frac{(5/2-\theta_1)}{2} Z_1 + \frac{1}{2\sqrt{6}} Z_2, \quad \forall \theta \in \mathcal{M}_{\Psi}^{\alpha}(F).$$

Using this representation it is easy to see that

$$\lambda \min_{\tilde{\theta} \in \mathcal{M}_{\Psi}^{\alpha}(F)} Z\left(\tilde{\theta}\right) \stackrel{\mathrm{d}}{=} \begin{cases} \frac{3}{2} Z_1 + \frac{1}{\sqrt{6}} Z_2, & Z_1 \leq 0; \\ \frac{1}{2} Z_1 + \frac{1}{\sqrt{6}} Z_2, & \text{otherwise}; \end{cases}$$

□

The goal of the final example of this section is to compare the performance of $\pi_{\Psi}^{\alpha}(F_n)$, when $F(\cdot)$ is restricted to a parametric family of distributions, with the maximum likelihood estimator (MLE) to develop some sense of the tradeoff made when using a non-parametric estimator. Also, unlike in the previous examples both $\mathcal{M}_{\Psi}^{\alpha}$ and Λ are singletons. This implies a normal weak limit, thus facilitating a comparison in terms of the asymptotic standard deviations.

Example 5
Let $F(\cdot)$ be the exponential distribution with hazard rate μ, $\Psi(\cdot)$ be as defined in (4), and $\alpha \in (0,1)$. For these choices it can be shown that $H_\Psi^\alpha(\cdot)$ is given by

$$H_\Psi^\alpha(X-\theta_1) = \left[\frac{\beta}{\mu}\right]\left(1 + \frac{\exp\{-\mu\theta_1\}}{(\exp\{\beta\}-1)(1-\alpha)}\right), \quad \theta_1 \geq 0.$$

Since $H_\Psi^\alpha(\cdot)$ has a nice closed form we can directly minimize $H_\Psi^\alpha(X-\theta_1)+\theta_1$. This minimization yields the following expression for the optimal $\tilde{\theta^*}$:

$$\tilde{\theta^*} = \left(\frac{1}{\mu}\log\left(\frac{\beta}{(\exp\{\beta\}-1)(1-\alpha)}\right), \frac{\beta+1}{\mu}\right).$$

It is noteworthy that θ_2^* does not depend on α, and moreover its expression above implies that δ_l of Assumption C is less than θ_2^*. As

$$\mathbb{E}\left(\left[\Psi\left(\frac{X}{\delta_l}\right)\right]^2\right) < \infty \iff \frac{2\beta}{\delta_l} < \mu$$

we have that Assumption C is satisfied only in the case that $\beta < 1$. In this case, applying Corollary 1 we have

$$\sqrt{n}\left(\pi_\Psi^\alpha(F_n) - \pi_\Psi^\alpha(F)\right) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0,\sigma^2\right).$$

where σ is given by

$$\sigma^2 := \frac{\left(\frac{2\beta}{1-\beta} - (\exp\{\beta\}-1)(1-\alpha)\right)}{(\exp\{\beta\}-1)(1-\alpha)\mu^2}.$$

Figure 3 plots the ratio of the asymptotic standard deviation of the parametric estimator to that of the non-parametric estimator. While it is expected that this ratio would be non-increasing as a function of α (as the non-parametric estimator essentially uses $(1-\alpha)$ part of the sample), it is noteworthy that the ratio in this example was significantly away from zero even for $\alpha = 95\%$. □

We end this section by providing a more general version of a weak convergence result for the empirical Orlicz premium than that provided in Bellini and Gianin (2008b). The idea is similar to that used above, and involves viewing $H_\Psi^\alpha(F)$ as the solution to the convex programming problem $(\mathbf{PH}_\infty)$ given by,

$$\begin{aligned} &\text{minimize} && \theta, && \theta \in (0,\infty) \\ &\text{subject to} && \theta\left[\mathbb{E}\left(\Psi\left(\frac{X}{\theta}\right)\right) - (1-\alpha)\right] \leq 0. \end{aligned} \qquad (\mathbf{PH}_\infty)$$

The above implies that $H_\Psi^\alpha(F_n)$ is the optimal value of the convex programming problem given by

$$\begin{aligned} &\text{minimize} && \theta, && \theta \in (0,\infty) \\ &\text{subject to} && \mathbb{E}_n\left(\gamma\left(\theta, Y\right)\right) \leq 0. \end{aligned}$$

where

$$\gamma(\theta, x) := \theta\left(\Psi\left(\frac{x}{\theta}\right) - (1-\alpha)\right), \quad \theta \in (0,\infty).$$

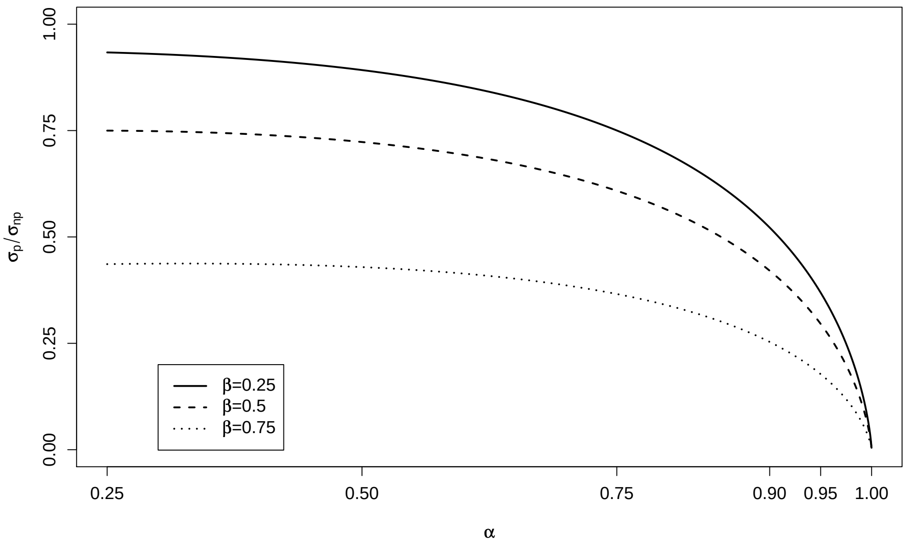


Figure 3: Ratio of Asymptotic Standard Deviations of Parametric and Non-Parametric Estimators

Now we state the result without proof as the proof is similar to that of Theorem 2.

Theorem 3. *Let X be a random variable in $\mathbb{X}_\Psi$ with $\Pr(X > 0) > 0$, $\alpha \in [0, 1)$, and let $\delta_* > 0$ be such that $\delta_* < H_\Psi^\alpha(F)$. Then for $F(\cdot)$ satisfying*

$$\mathbb{E}\left(\left[\Psi\left(\frac{X}{\delta_*}\right)\right]^2\right) < \infty,$$

we have

$$\sqrt{n}\left(H_\Psi^\alpha(F_n) - H_\Psi^\alpha(F)\right) \xrightarrow{\mathrm{d}} \max_{\lambda \in \Lambda'} \lambda V, \tag{27}$$

where Λ' is the set of Lagrange multipliers for the convex programming problem $(\mathbf{PH}_\infty)$, and V is a mean zero Gaussian random variable with variance given by $\mathbb{V}ar\left(\gamma\left(H_\Psi^\alpha(F), X\right)\right)$. Moreover, under the further assumption that $\Psi(\cdot)$ is differentiable on $(0, \infty)$ we have

$$\sqrt{n}\left(H_\Psi^\alpha(F_n) - H_\Psi^\alpha(F)\right) \xrightarrow{\mathrm{d}} \left[\mathbb{E}\left(\Psi\left(\frac{X}{H_\Psi^\alpha(F)}\right)\left[\frac{X}{H_\Psi^\alpha(F)}\right]\right)\right]^{-1} V.$$

4. Simulation Study

In this section we report on a simulation study conducted to lend insight into the effect of the sampling distribution $F(\cdot)$, Young function $\Psi(\cdot)$, and the level α on the sample sizes required for the above derived asymptotic limits to take hold. All of the programs were written to be run on the R software environment for statistical computing and graphics (see Team (2008)). The programs were run parallel on 40 processors over 10 nodes of a 22 node Beowulf cluster using the snow R package (Tierney et al. (2008, 2009)). The algorithm we used to compute the empirical Haezendonck risk measure $\pi_\Psi^\alpha(F_n)$ is presented in Ahn and Shyamalkumar (2011b) - this algorithm tested to be many folds faster than using the R optimization package `alabama` (Varadhan and Grothendieck (2011)), and significantly faster than using `convexmin` on Matlab (see Ahn and Shyamalkumar (2011b) for detailed comparisons).

We use a Gaussian kernel density estimator, via the function `density` on R, to estimate the densities of the empirical Haezendonck risk measures. The bandwidth used equals 0.9 times the minimum of the standard deviation and the interquartile range divided by 1.34 times the sample size to the negative one-fifth power (*i.e.* Silverman's 'rule of thumb', see Silverman (1986)) unless the quartiles coincide, in which case a positive value is used.

4.1. *Effect of the Sampling Distribution and the Level*

In this sub-section we will work with $\Psi(\cdot)$ defined given by

$$\Psi(x) = \begin{cases} 0, & x < 0; \\ \frac{x^2+x}{2}, & \text{otherwise}; \end{cases},$$

and distributions $G_i(\cdot)$, $i = 1, 2, 3$ defined by

$$G_1(x) = \begin{cases} 0, & x < 0; \\ x, & 0 \le x \le 1; \\ 1, & \text{otherwise}; \end{cases}, \quad G_2(x) = \begin{cases} 0, & x < 0; \\ 1 - (1+\sigma x)^{-\beta}, & x \ge 0; \end{cases},$$

and

$$G_3(x) = \begin{cases} 0, & x < 0; \\ \Phi\left(\frac{\log(x)-\mu}{\sigma}\right), & x \ge 0; \end{cases},$$

where $\mu \in \mathbb{R}$, $\sigma > 0$, $\beta > 4$, and $\Phi(\cdot)$ is the standard normal distribution function. Note that $G_1(\cdot)$, $G_2(\cdot)$, and $G_3(\cdot)$ are distribution functions corresponding to the uniform distribution on $(0, 1)$, the two-parameter Pareto family and the lognormal distribution, respectively.

For each of the stated three distributions we will show below that there is a unique optimal $\tilde{\theta}^*$. For $i = 1, 2, 3$, we define Y_i and λ_i by

$$Y_i := \frac{1}{\theta_2^*} \max\left(Z_i - \theta_1^*, 0\right), \quad \text{and} \quad \lambda_i := \left[\mathbb{E}\left(Y_i^2 + \frac{Y_i}{2}\right)\right]^{-1} = \left[\mathbb{E}\left(Y_i + \frac{I(Y_i > 0)}{2}\right)\right]^{-1},$$

where $Z_i \sim G_i$. Also, we define σ_i by

$$\sigma_i^2 := \left[\frac{\lambda_i}{2}\right]^2 \mathbb{V}ar\left(Y_i^2 + Y_i\right), \quad i = 1, 2, 3.$$

Now using Corollary 1 we have that the asymptotic limit of the empirical Haezendonck risk measure under $G_i(\cdot)$ is given by $\mathrm{N}\left(\pi_\Psi^\alpha(G_i), \sigma_i^2\right)$, for $i = 1, 2, 3$. In the following we will derive the unique optimal $\tilde{\theta}^*$ for each of the three distributions. The uniqueness part in all of the cases follows from θ_1^* belonging to the interior of the supports of the distributions.

For $G_1(\cdot)$, it can then be shown that for $\tilde{\theta} \in (0,1) \times (0,\infty)$ we have

$$\mathbb{E}\left(\Psi'\left(\frac{Z_1-\theta_1}{\theta_2}\right)\right) = \frac{(1-\theta_1)(1-\theta_1+\theta_2)}{2\theta_2},$$

and

$$\mathbb{E}\left(\Psi'\left(\frac{Z_1-\theta_1}{\theta_2}\right)\left[\frac{Z_1-\theta_1}{\theta_2}\right]\right) = \frac{(1-\theta_1)^3}{3\theta_2^2} + \frac{(1-\theta_1)^2}{4\theta_2}.$$

Now using the Karusch-Kuhn-Tucker conditions in (25) we have that any optimal $\tilde{\theta}^*$ satisfies

$$\frac{1-\theta_1^*}{\theta_2^*} = \frac{12}{\sqrt{105}-3}.$$

Also, from (25) we have

$$\mathbb{E}\left(\Psi\left(\frac{Z_1-\theta_1^*}{\theta_2^*}\right)\right) = \frac{(1-\theta_1^*)^3}{6(\theta_2^*)^2} + \frac{(1-\theta_1^*)^2}{4\theta_2^*} = 1-\alpha.$$

Using the above two equations we get the following closed form expression for $\tilde{\theta}^*$ as a function of α,

$$\tilde{\theta}^* = \left(1 - \frac{(\sqrt{105}-3)^2(1-\alpha)}{3(\sqrt{105}+5)}, \frac{(\sqrt{105}-3)^3(1-\alpha)}{36(\sqrt{105}+5)}\right).$$

In the case of $G_2(\cdot)$ the procedure to find the unique optimal $\tilde{\theta}^*$ is very similar to that for $G_1(\cdot)$; hence we directly provide an expression for it. The unique optimal $\tilde{\theta}^*$ is given by

$$\tilde{\theta}^* = \left(q_{p_\alpha}(F_2), \frac{3(1-p_\alpha)(1+\sigma q_{p_\alpha}(F_2))}{\sigma(\beta-1)\left(4(1-\alpha)-(1-p_\alpha)\right)}\right),$$

where p_α is given by

$$1 - \frac{32(\beta-1)(1-\alpha)}{8(\beta-1)+3(\beta-2)\left(1+\sqrt{1+16\left(\frac{\beta-1}{\beta-2}\right)}\right)}.$$

In the case of $G_3(\cdot)$, unlike the earlier two distributions, the optimal $\tilde{\theta}^*$ does not have a closed form expression. Nevertheless, the Karusch-Kuhn-Tucker conditions in (25) imply that the optimal θ_1^* satisfies

$$\left(\frac{\mathbb{E}\left((Z_3-\theta_1^*)_+^2\right)}{9\left[\mathbb{E}\left((Z_3-\theta_1^*)_+\right)\right]^2}\right)\left[4(1-\alpha)-\Pr\left(Z_3>\theta_1^*\right)\right]^2 - \left(\frac{1}{3}\right)\left[2(1-\alpha)+\Pr\left(Z_3>\theta_1^*\right)\right] = 0, \qquad (28)$$

where

$$\Pr\left(Z_3>\theta_1^*\right) = \Phi\left(\frac{\log(\theta_1^*)-\mu}{\sigma}\right),$$

and

$$\mathbb{E}\left((Z_3-\theta_1^*)_+^m\right) = \sum_{i=0}^{m}\binom{m}{i}\exp\left\{\mu i + \frac{(\sigma i)^2}{2}\right\}\Phi\left(\sigma i - \left[\frac{\log(\theta_1^*)-\mu}{\sigma}\right]\right)(-\theta_1^*)^{m-i}.$$

Sampling Distribution	$\alpha = 95\%$				$\alpha = 99\%$			
	θ_1^*	$\Pr(X > \theta_1^*)$	$\pi_\Psi^\alpha(F)$	σ	θ_1^*	$\Pr(X > \theta_1^*)$	$\pi_\Psi^\alpha(F)$	σ
Uniform	0.9426	5.741%	0.9773	0.1157	0.9885	1.148%	0.9955	0.05256
Pareto	27.03	7.477%	54.63	197.4	50.73	1.495%	86.82	579.0
Lognormal	4.102	7.905%	9.978	53.64	8.641	1.552%	17.20	158.7

Table 1: Estimation of the Haezendonck Risk Measure with $\Psi(x) = \frac{x^2+x}{2} I(x > 0)$: Varying Sampling Distributions

Now θ_1^* can be solved for using (28) and numerical algorithms like the Newton-Raphson method (on R we used the function `uniroot`), and then θ_2^* can be found by using the expression

$$\theta_2^* = \frac{3\mathbb{E}\left((Z_3 - \theta_1^*)_+\right)}{4(1-\alpha) - \Pr\left(Z_3 > \theta_1^*\right)},$$

which is derived from (25) as well.

In the following, the parameters σ and β of the Pareto distribution $G_2(\cdot)$ are chosen to be 0.02 and 6, respectively; the parameters μ and σ of the lognormal distribution $G_3(\cdot)$ are chosen to be 0 and 1, respectively. Table 1 contains the values of the optimal θ_1^*, Haezendonck risk measure and asymptotic standard error for the empirical Haezendonck risk measure for each of the above sampling distributions and for both the 95% and 99% levels. The table clearly exhibits the effect of heaviness of tails on all of the reported values.

We simulated $100,000$ sets of random samples for each combination of sampling distribution, level, and sample size. Figure 4 contains the plot of the estimated densities of the standardized (using $\pi_\Psi^\alpha(F)$ and the asymptotic standard deviation) empirical Haezendonck risk measure. As expected, Figure 4 confirms that a higher α value requires a larger sample size for the asymptotic limit to take hold. In a sense $n * \Pr(X > \theta_1^*)$ is the *effective* sample size as it is essentially only the observations beyond θ_1^* that determine $\pi_\Psi^\alpha(F_n)$. From this point of view, Figure 4 suggests that in the case of the uniform distribution normality takes hold rather quickly, while in the Pareto and lognormal cases a moderate sample size is required for normality to take hold. On the other hand it is noteworthy that in Figure 4 normality seems to take hold in a similar fashion for both the Pareto and the lognormal distributions, whereas the former distribution has a much heavier tail than the latter.

4.2. *Effect of Young Function*

In this subsection our sampling distribution is the exponential distribution with unit mean, and we work with three Young functions $\Psi_i(\cdot)$, $i = 1, 2, 3$, defined by

$$\Psi_1(x) = \begin{cases} 0, & x < 0; \\ \frac{(\exp\{x/2\}-1)}{\exp\{1/2\}-1}, & x \geq 0; \end{cases}, \quad \Psi_2(x) = \begin{cases} 0, & x < 0; \\ \frac{x^2+x}{2}, & x \geq 0; \end{cases}, \quad \text{and} \quad \Psi_3(x) = \begin{cases} 0, & x < 0; \\ x, & x \geq 0; \end{cases}.$$

Note that $\Psi_1(\cdot)$ is more convex than $\Psi_2(\cdot)$, and both of these are more convex than $\Psi_3(\cdot)$.

Similar to the development in the previous sub-section we first identify the unique optimal $\tilde{\theta}^*$. This has been done for $\Psi_1(\cdot)$ in Example 5; as $\Psi_3(\cdot)$ corresponds to the case of T-VaR, θ_1^* is the α-level quantile

Young Function	$\alpha = 95\%$				$\alpha = 99\%$			
	θ_1^*	$\Pr(X > \theta_1^*)$	$\pi_\Psi^\alpha(F)$	σ	θ_1^*	$\Pr(X > \theta_1^*)$	$\pi_\Psi^\alpha(F)$	σ
$\frac{(\exp\{x/2\}-1)I(x>0)}{\exp\{1/2\}-1}$	2.735	6.487%	4.235	7.788	4.345	1.297%	5.845	17.530
$\frac{(x^2+x)I(x>0)}{2}$	2.681	6.847%	4.243	7.337	4.291	1.369%	5.852	16.527
$xI(x > 0)$	2.996	5%	3.996	6.245	4.605	1%	5.605	14.107

Table 2: Estimation of Haezendonck Risk Measure for Exponential Distribution: Different Young Functions

given by $-\log(1-\alpha)$ and θ_2^* is 1 by the memoryless property. It is worth mention that in the case of the T-VaR the asymptotic distribution is given in Manistre and Hancock (2005) and Brazauskas et al. (2008). For $\Psi_2(\cdot)$, the procedure to find the optimal $\tilde{\theta}^*$ is similar to that of the previous subsection, and is given by

$$\tilde{\theta}^* = \left(-\log\left[(1-\alpha)\left(\sqrt{153}-11\right)\right], \frac{\sqrt{153}-11}{5-\sqrt{17}}\right).$$

Table 2 contains the values of the optimal θ_1^*, the Haezendonck risk measure and the asymptotic standard error for the empirical Haezendonck risk measure for each of the above Young functions and for both the 95% and 99% levels. It is noteworthy that only the asymptotic standard error that shows a direct relationship with the convexity of the Young function; in other words, the the more convex the Young function the larger the asymptotic standard error.

As earlier, we simulated $100,000$ sets of random samples for each sample size. Figure 5 contains the plot of the estimated densities of the standardized (using $\pi_\Psi^\alpha(F)$ and the asymptotic standard deviation) empirical Haezendonck risk measure at the 95%-level. Figure 5 suggests that a more convex Young function requires a larger sample size for the asymptotic limit to take hold.

5. Discussions and Future Work

In this article we have establish a weak convergence result for the empirical Haezendonck risk measure which allows for any Young function, and essentially all distributions with suitable upper tail. In a sense, comparing it with the case of the ordinary central limit theorem, the conditions for this result are near ideal. Also, we establish strong consistency for this estimator, as well as for the optimal solution for the minimization problem defining this estimator.

The importance of the weak convergence result in applications is accentuated by the fact that even in non-pathological situations the empirical Haezendonck risk measure can have a non-normal limiting distribution. Also, even in the case where the limiting distribution is normal, as reported in Bellini and Gianin (2008b), simply testing for normality using for example the Jarque-Bera statistic can be misleading. For example, such testing resulted in rejecting normality, with a p-value less than 1%, for both Example 5 (with $\beta = 1$) and the uniform example discussed in section 5, where as normality does indeed hold true in the latter case.

As observed in Bellini and Gianin (2008b), computing the empirical Haezendonck risk measures can be a challenging computational task. This results in the estimate being, as in other similar problems, subject to both sampling error as well as computational error. Hence it is worth mention that in Ahn and Shyamalkumar (2011b) we provide a computational algorithm for the empirical Haezendonck risk measure such that the approximation to the empirical Haezendonck risk measure computed using this algorithm is guaranteed to have the same asymptotic behavior as the empirical Haezendonck risk measure.

As mentioned above, in the case that there is a unique x^* satisfying $\pi_\Psi^\alpha(F, x^*) = \pi_\Psi^\alpha(F)$, this optimal solution can be viewed as an analog of the quantile. Due to this viewpoint it is of interest to derive a weak convergence result for the sample analog of x^*. The approach taken in this paper, we believe, is a promising way towards establishing such a result.

The weak convergence result is useful in practice as it paves a way for construction of confidence intervals. For interval estimation of the Haezendonck risk measure one would further need a consistent estimator of the asymptotic standard error. While an obvious candidate exists in the case of a differentiable $\Psi(\cdot)$ and a continuous $F(\cdot)$, investigation of the performance of the interval estimator resulting from its use would be beneficial.

References

Ahn, J. and Shyamalkumar, N. (2011a). Large sample behavior of the CTE and VaR estimators under importance sampling. *North American Actuarial Journal (to appear).*

Ahn, J. and Shyamalkumar, N. (2011b). On computation of the Haezendonck risk measure: An algorithm and its complexity. Technical Report 409, The University of Iowa - Dept. of Statistics and Actuarial Science.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1997). Thinking coherently. *RISK*, 10(11):68–71.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Math. Finance*, 9(3):203–228.

Bellini, F. and Gianin, E. (2008a). On Haezendonck risk measures. *Journal of Banking & Finance*, 32(6):986–994.

Bellini, F. and Gianin, E. (2008b). Optimal portfolios with Haezendonck risk measures. *Statistics & Decisions*, 26:89–108.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge.

Brazauskas, V., Jones, B., Puri, M., and Zitikis, R. (2008). Estimating conditional tail expectation with actuarial applications in view. *Journal of Statistical Planning and Inference*, 138(11):3590–3604.

Durrett, R. (2005). *Probability: Theory and Examples*. Duxbury Press, Belmont, CA, third edition.

Goovaerts, M., Kaas, R., Dhaene, J., and Tang, Q. (2004). Some new classes of consistent risk measures. *Insurance: Mathematics and Economics*, 34(3):505–516.

Goovaerts, M. J., Kaas, R., Dhaene, J., and Tang, Q. (2003). A unified approach to generate risk measures. *Astin Bull.*, 33(2):173–191.

Haezendonck, J. and Goovaerts, M. (1982). A new premium calculation principle based on Orlicz norms. *Insurance Math. Econom.*, 1(1):41–53.

King, A. J. and Rockafellar, R. T. (1990). Asymptotic theory for solutions in generalized m-estimation and stochastic programming. Working paper WP-90-76, International Institute for Applied System Analysis, Austria.

Manistre, B. J. and Hancock, G. H. (2005). Variance of the CTE estimator. *N. Am. Actuar. J.*, 9(2):129–156.

Nam, H., Tang, Q., and Yang, F. (2011). Characterization of upper comonotonicity via tail convex order. *Insurance: Mathematics and Economics*, 48(3):368–373.

Rao, M. M. and Ren, Z. D. (1991). *Theory of Orlicz spaces*, volume 146 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker Inc., New York.

Rockafellar, R. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471.

Rockafellar, R. T. (1997). *Convex analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ. Reprint of the 1970 original, Princeton Paperbacks.

Rubinstein, R. Y. and Shapiro, A. (1993). *Discrete event systems*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Ltd., Chichester. Sensitivity analysis and stochastic optimization by the score function method.

Shapiro, A. (1991). Asymptotic analysis of stochastic programs. *Ann. Oper. Res.*, 30(1-4):169–186. Stochastic programming, Part I (Ann Arbor, MI, 1989).

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.

Team, R. D. C. (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Tierney, L., Rossini, A. J., Li, N., and Sevcikova, H. (2008). *snow: Simple Network of Workstations*. R package version 0.3-3.

Tierney, L., Rossini, A. J., Li, N., and Sevcikova, H. (2009). Snow: A parallel computing framework for the R system. *Int. J. Parallel Prog.*, 37:78–90.

Varadhan, R. and Grothendieck, G. (2011). *alabama: Constrained nonlinear optimization*. R package version 2011.3-1.

Nariankadu D. Shyamalkumar
Dept. of Statistics and Actuarial Science
The University of Iowa
241 Schaeffer Hall, Iowa City, IA 52242

A. Appendix

Lemma 1. *For $\tilde{\theta}, \tilde{\theta}' \in S(F)$, and $0 < \epsilon < 1$ we have*

$$\left|\psi(\tilde{\theta}, x) - \psi(\tilde{\theta}', x)\right| \leq \left[\frac{4\delta_u}{\epsilon\delta_l}\right] \max\left[\epsilon\partial_+\Psi(1), \Psi\left(\frac{-x_l}{\epsilon(1-\epsilon)\delta_l}\right), \Psi\left(\frac{x}{\delta_l(1-\epsilon)^2}\right)\right] \left\|\tilde{\theta} - \tilde{\theta}'\right\| \tag{29}$$

Proof. Without loss of generality let $y_1 \leq y_2$, where

$$y_1 := \frac{x - \theta_1'}{\theta_2'}, \quad \text{and} \quad y_2 := \frac{x - \theta_1}{\theta_2}.$$

Note that the bound above is trivially satisfied for $y_2 \leq 0$; hence for the rest of this proof we will assume without loss of generality that $y_2 > 0$. As

$$|y_1 - y_2| \leq \left(\frac{\sqrt{2}\max(1, y_2)}{\delta_l}\right) \left\|\tilde{\theta} - \tilde{\theta}'\right\|,$$

we have by convexity of $\Psi(\cdot)$ that

$$\begin{aligned}
|\Psi(y_2) - \Psi(y_1)| &\leq (y_2 - y_1)\partial_+\Psi(y_2) \\
&\leq \left(\frac{\sqrt{2}\max(\partial_+\Psi(1), y_2\partial_+\Psi(y_2))}{\delta_l}\right) \left\|\tilde{\theta} - \tilde{\theta}'\right\| \\
&\leq \left[\frac{\sqrt{2}}{\epsilon\delta_l}\right] \max\left[\epsilon\partial_+\Psi(1), \Psi\left(\frac{y_2}{1-\epsilon}\right)\right] \left\|\tilde{\theta} - \tilde{\theta}'\right\|
\end{aligned}$$

Also, some algebra yields

$$\Psi\left(\frac{y_2}{1-\epsilon}\right) \leq \Psi\left(\frac{\max(x/(1-\epsilon), -x_l/\epsilon)}{\delta_l(1-\epsilon)}\right) \leq \max\left[\Psi\left(\frac{x}{\delta_l(1-\epsilon)^2}\right), \Psi\left(\frac{-x_l}{\epsilon(1-\epsilon)\delta_l}\right)\right],$$

using which we have

$$|\Psi(y_2) - \Psi(y_1)| \leq \left[\frac{\sqrt{2}}{\epsilon\delta_l}\right] \max\left[\epsilon\partial_+\Psi(1), \Psi\left(\frac{-x_l}{\epsilon(1-\epsilon)\delta_l}\right), \Psi\left(\frac{x}{\delta_l(1-\epsilon)^2}\right)\right] \left\|\tilde{\theta} - \tilde{\theta}'\right\|$$

Now using the fact that $\partial_+\Psi(1) \geq 1$, and that

$$\left|\psi(\tilde{\theta}, x) - \psi(\tilde{\theta}', x)\right| \leq [1 - \alpha + \Psi(y_2)]\,|\theta_2 - \theta_2'| + \delta_u\,|\Psi(y_2) - \Psi(y_1)|$$

we have (29).

□

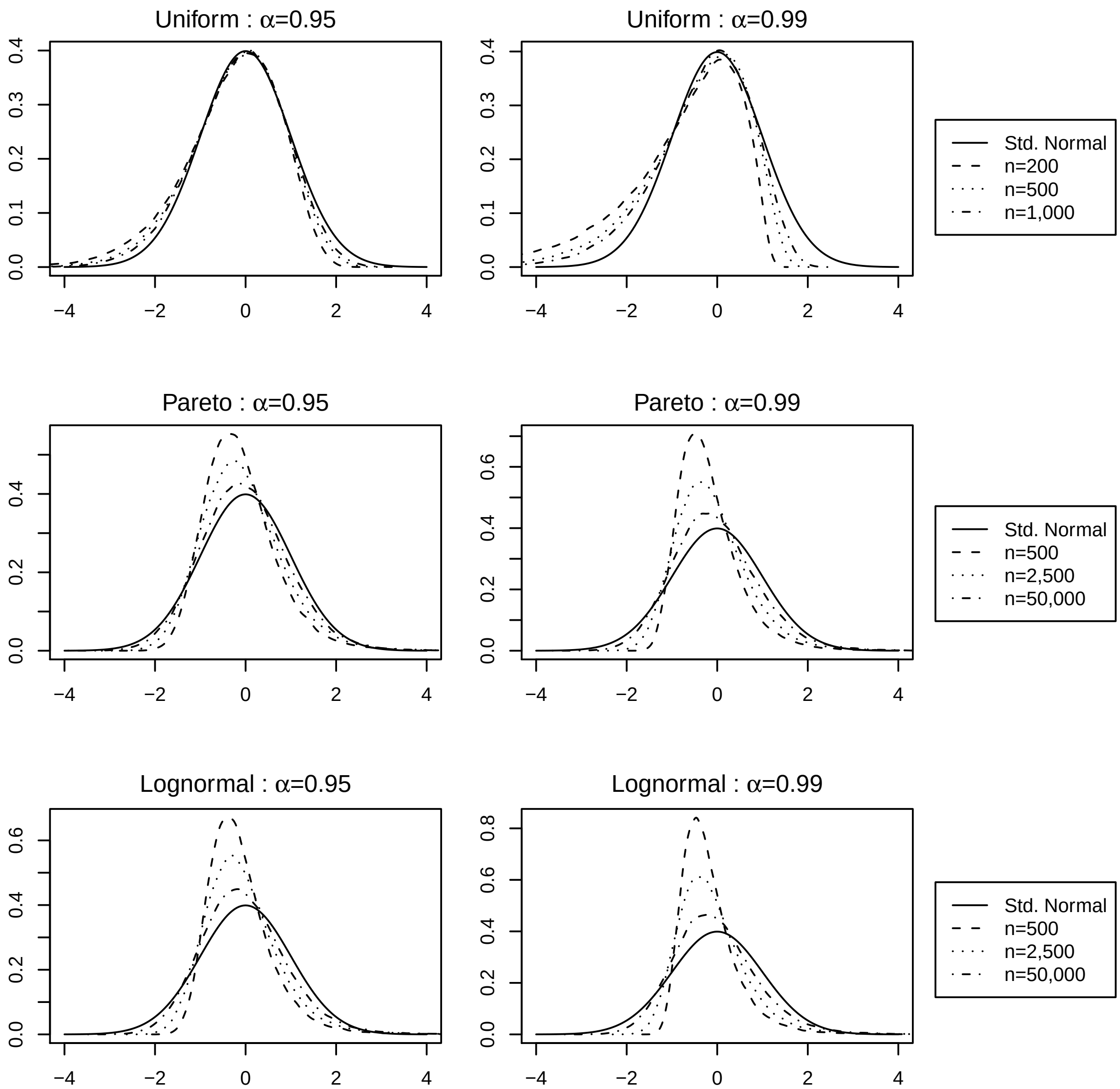


Figure 4: Estimated Densities of Empirical Haezendonck Risk Measures - Varying Levels and Sampling Distributions

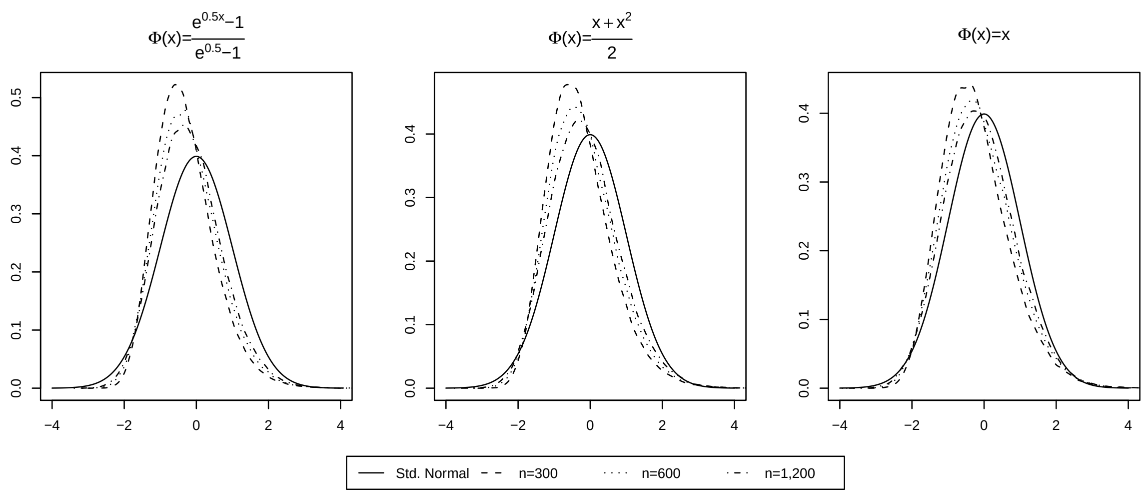


Figure 5: Estimated Densities of Empirical Haezendonck Risk Measures - Varying Young Function