

Computation and asymptotic properties of estimated coherent risk measures

D. J. Miller & M. Kim

Department of Economics, University of Missouri, USA

Abstract

Recent advances in mathematical finance established linkages among several key concepts related to coherence, distorted risk measures, and information theory. The purpose of this paper is to extend these theoretical results for empirical applications in computational finance. First, we use a concentrated (dual) entropy approach to derive a computational algorithm for estimating the parameters of a distorted probability model associated with a coherent risk measure for a given sample of observed data. Second, we derive the asymptotic sampling properties of the estimated model parameters, which may be used to conduct classical hypothesis tests or to form other statistical inferences based on the estimated coherent risk measure. Third, we note that researchers may also require an estimate of the net loss distribution, and we propose an information theoretic procedure for jointly estimating the net loss probability model and the distorted probability distribution associated with a particular coherent risk measure.

Keywords: coherence, distorted risk measure, entropy, extremum estimator, information theory.

1 Introduction

Risk measures based on the distribution of potential asset losses or returns are widely used in empirical finance, and prominent applications of these tools include determining insurance premia, option prices, margin deposits for hedged and speculative positions in futures markets, and capital reserve requirements for banks and other firms (e.g., see Wirth and Hardy [1]). To be specific, suppose an asset has risk or net loss represented by random variable X with cumulative distribution function (CDF) $F(x) = P(X \leq x)$. A risk measure is a mapping $\rho(X) : \mathbb{R} \rightarrow \mathbb{R}^+$



that represents the price or cost associated with holding the risky position. For example, the value-at-risk (VaR) with confidence level $1 - \alpha$ for an asset with a strictly continuous distribution is defined as

$$\rho_{VaR}(X) = F^{-1}(1 - \alpha) \quad (1)$$

In words, the VaR risk measure is the $100(1 - \alpha)\%$ quantile of the net loss distribution.

VaR is one of the most widely used risk measures because it is relatively easy to compute from estimates of the net loss distribution and is very convenient to interpret — the largest possible loss with lower tail probability $1 - \alpha$. Although VaR is commonly used in practice, several authors have noted that this risk measure is flawed because it does not properly account for added risks. Artzner, *et al.* [2] define an improved class of risk measures under the following conditions on random variables X and Y in class \mathcal{X} :

- A1 Monotonicity: if $F_X(z) \geq F_Y(z)$ for all $z \in \mathbb{R}$, then $\rho(X) \leq \rho(Y)$
- A2 Positive homogeneity (of degree one): $\rho(\lambda X) = \lambda \rho(X)$ for all $\lambda \in \mathbb{R}^+$
- A3 Linear invariance: $\rho(aX + b) = a\rho(X) + b$ for all $a, b \in \mathbb{R}$
- A4 Subadditivity: $\rho(X + Y) \leq \rho(X) + \rho(Y)$

If these conditions are satisfied for all random variables X and Y in class \mathcal{X} , then the risk measure is said to be *coherent* over the class (Definition 7, Artzner, *et al.* [2]).

The limitations associated with value-at-risk arise because VaR does not satisfy the subadditivity property (A4) and is not coherent. Consequently, researchers in the mathematical finance literature have developed coherent risk measures that are designed to satisfy all four conditions (A1–A4) and should improve on VaR. For example, Artzner, *et al.* [2] propose the conditional tail expectation (CTE) or Tail-VaR risk measure with confidence level $1 - \alpha$ defined as

$$\rho_{CTE}(X) = E[X \mid F(X) > 1 - \alpha] \quad (2)$$

For practical purposes, the CTE risk measure is based on roughly the same information required to compute VaR and should be relatively easy to use. However, VaR remains to be widely used in practice, and researchers in this field have worked to develop other coherent risk measures that are suitable for particular applications.

2 Coherence, distortion, and entropy

The conceptual challenges associated with deriving coherent risk measures have been eased by recent research outcomes that link coherence to other established concepts and methods in the fields of mathematical finance and statistical information theory. The key contributions in the recent literature are provided by McLeish and Reesor [7], who show that

- Distorted risk measures based on distortions or transformations of the net loss distribution are closely related to coherent risk measures



- Distorted risk measures may be derived from the net loss distribution under an information theoretic (entropy) optimization approach

Thus, we can use the information theoretic approach outlined by McLeish and Reesor [7] to derive coherent risk measures.

Following Definition 4 presented by McLeish and Reesor [7], a *distortion function* g is a right-continuous and non-decreasing function on the unit interval $[0, 1]$ with the end-point properties $g(0) = 0$ and $g(1) = 1$. Given g , the associated distorted cumulative distribution function F^* based on the original CDF for a net loss variable (F) may be formed as

$$F^*(x) = 1 - g(S(x)) \quad (3)$$

where $S(X) \equiv 1 - F(X)$. The distorted probability density function is

$$f^*(x) = f(x)g'(S(x)) \quad (4)$$

where $g'(z) = dg(z)/dz$. Finally, following Definition 6 in McLeish and Reesor [7], we can form a *distorted risk measure* as

$$\rho_g(X) = \int_0^\infty g(S(x))dx + \int_{-\infty}^0 (g(S(x)) - 1)dx = E_{F^*}(X) \quad (5)$$

where the expected value of X is computed as a Choquet integral with respect to the distorted probability distribution. Choquet integrals may be used to represent non-additive probability measures but reduce to the common Lebesgue integral under additive measures (see Denneberg [4] and Groes, et al. [13] for more details).

Distorted probability models and risk measures have been used by a large number of researchers in empirical finance, especially in insurance pricing problems (for example, see Wang [8] and Wirth and Hardy [14]). For purposes of this paper, the key property of distorted risk measures is provided by Theorem 5 in McLeish and Reesor [7], which claims that the following statements are equivalent:

- The distortion function g is concave.
- The distorted risk measure ρ_g is coherent over the class of all random variables.

Thus, we can use the distortion approach to derive coherent risk measures as long as the distortion function is concave.

McLeish and Reesor [7] also propose an information theoretic approach to derive distorted probability models, and their approach is based on the classical maximum entropy and minimum cross-entropy procedures that Jaynes [9, 10] originally developed for discrete probability distributions. The minimum cross-entropy method may also be extended for use with continuous probability distributions, and the details are outlined in Appendix 3. A of the book by Golan, et al. [5]). In either case, the basic purpose of the method is to recover a probability distribution that satisfies some moment conditions (e.g., mean, variance). In general, a probability distribution cannot be uniquely characterized by a finite number of moments, so there are multiple candidate distributions for any particular

problem. Accordingly, Jaynes recommends choosing the distribution that satisfies these conditions and is closest to some target or reference distribution, and the “closeness” of the distributions is measured with the Kullback–Leibler cross–entropy functional (see Kullback [6]). One of the key contributions provided by McLeish and Reesor [7] is the set of moment conditions, $E[G_k(S(x))] = \mu_k$, required to derive five of the most commonly used distorted probability models by the minimum cross–entropy method (see their Table 1).

Specifically, the information theoretic approach for deriving the distorted probability model is to minimize the cross–entropy objective function

$$I(f^*, f) = \int \ln\left(\frac{f^*(x)}{f(x)}\right) f^*(x) dx \quad (6)$$

subject to the known constraints

$$\int G_k(S(x)) f^*(x) dx = \mu_k \quad \text{and} \quad \int f^*(x) dx = 1 \quad (7)$$

McLeish and Reesor [7] note that the entropy problem may be solved for f^* by the method of Lagrange multipliers (i.e., calculus of variations), and the Lagrange equation is

$$\begin{aligned} \mathcal{L}(f^*, \lambda) = I(f^*, f) + \sum_{k=1}^K \lambda_k \left[\mu_k - \int G_k(S(x)) f^*(x) dx \right] \\ + \gamma \left[1 - \int f^*(x) dx \right] \end{aligned} \quad (8)$$

where $\lambda = (\lambda_1, \dots, \lambda_K)'$, λ_k is the Lagrange multiplier for the k^{th} moment constraint, and γ is the Lagrange multiplier on the additivity constraint in (7). Although McLeish and Reesor [7] do not provide the minimum cross–entropy solution for the distorted probability model

$$f^*(x) = \frac{f(x) \exp(\sum_{k=1}^K \lambda_k G_k(S(x)))}{\int f(x) \exp(\sum_{k=1}^K \lambda_k G_k(S(x))) dx} \quad (9)$$

we can use this expression to derive

$$g'(S(x)) = \frac{f^*(x)}{f(x)} = \frac{\exp(\sum_{k=1}^K \lambda_k G_k(S(x)))}{\int f(x) \exp(\sum_{k=1}^K \lambda_k G_k(S(x))) dx} \quad (10)$$

By integrating (10), we can derive the implicit solution for the distortion function provided by McLeish and Reesor [7] in their Equation (10)

$$g(S(x)) = \int_0^{S(x)} \exp\left(\sum_{k=1}^K \lambda_k G_k(S(z)) - \psi(\lambda)\right) dz \quad (11)$$

where $\psi(\lambda) \equiv \ln(\int f(x) \exp(\sum_{k=1}^K \lambda_k G_k(S(x))) dx)$ is the natural log of the normalizing constant from (9). Finally, McLeish and Reesor [7] present their

Theorem 4, which claims that the distortion function is concave if and only if

$$\sum_{k=1}^K \lambda_k G_k(S(z)) \quad (12)$$

is non-increasing in z . Under this condition, a distorted risk measure derived by the information theoretic approach will also be a coherent risk measure. The relationship between the concavity of the distortion function g and coherence (especially subadditivity) is also examined by Wang and Dhaene [11] and by Wirth and Hardy [14].

3 Concentrated entropy approach to computation

The theoretical developments outlined in the preceding sections have greatly expanded the conceptual range of risk measures available to researchers in mathematical finance. However, much of this effort has been focused on the analytical properties of these tools, and researchers in empirical finance need information about the computational and statistical properties of these tools in order to apply them in practice. In this section, we outline a computational algorithm based on a natural extension of the information theoretic approach developed by McLeish and Reesor [7], and we use this approach to derive the asymptotic sampling properties of the estimated risk measures in the following section.

To begin, suppose we have the sample moments for a given minimum cross-entropy estimation problem, and we denote these sample estimates as $\hat{\mu}_1, \dots, \hat{\mu}_K$. The key computational challenge to forming an empirical version of the distortion function in Equation (8) is to derive the optimal value of the Lagrange multipliers, $\hat{\lambda}_1, \dots, \hat{\lambda}_K$. Although it may be feasible to solve the problem in primal form as stated in Equations (6) and (7), the number of software packages that can solve such constrained numerical optimization problems is quite limited. Fortunately, we can convert this estimation problem to an unconstrained maximization problem that can be solved by a much wider range of numerical optimization software.

The key step is to form a concentrated or dual optimization problem by substituting the implicit solution for the distorted density function (9) back into the Lagrange expression for the primal version of the problem, (8). After simplifying terms, the concentrated Lagrange equation is

$$\mathcal{M}(\lambda) = \sum_{k=1}^K \lambda_k \hat{\mu}_k - \ln \left[\int f(x) \exp \left(\sum_{k=1}^K \lambda_k G_k(S(x)) \right) dx \right] \quad (13)$$

which is a function of the K -vector of Lagrange multipliers. If the minimum cross-entropy problem has a unique interior solution, we can use the saddle-point properties of the Lagrange equation to show that the concentrated objective function \mathcal{M} is a strictly concave function of λ such that

$$\mathcal{M}(\hat{\lambda}) > \mathcal{M}(\lambda) \quad (14)$$



for all $\lambda \neq \hat{\lambda}$. Thus, we can compute the optimal values of the Lagrange multipliers by maximizing \mathcal{M} with respect to λ without constraints. Further, we note that the gradient of \mathcal{M}

$$\frac{\partial \mathcal{M}}{\partial \lambda_k} = \hat{\mu}_k - \int G_k(S(x)) f^*(x) dx \quad \text{for } k = 1, \dots, K \quad (15)$$

is simply the set of moment conditions for the primal minimum cross-entropy problem from (7), and this gradient information may be used to compute the Lagrange multipliers under the Newton–Raphson or related numerical optimization algorithms. For example, the maximum entropy algorithm proposed by Ormoneit and White [12] may be extended to the minimum cross-entropy version of the estimation problem.

4 Asymptotic properties of the model parameter estimator

Although the authors in the preceding literature do not focus on the interpretation of the Lagrange multipliers, these values represent the marginal change in distortion (uncertainty) between the distorted distribution (F^*) and the reference distribution (F) given a small increase in the associated moment values. In empirical research, the estimated values of the Lagrange multipliers may be very useful for model diagnostic and specification purposes. For example, we may wish to test the statistical significance of one or more of the moment conditions in a candidate probability model, and a model specification test for the j^{th} moment may be conducted under the null hypothesis $H_0 : \lambda_j = 0$.

To begin, we assume the K -vector of sample moments $\hat{\mu} = (\hat{\mu}_1, \dots, \hat{\mu}_K)'$ imposed in the minimum cross-entropy problem (i.e., in the first constraint in (7)) are based on samples of size n and are

- Consistent such that $\hat{\mu} \xrightarrow{P} \hat{\mu}_0$ as $n \rightarrow \infty$ where $\hat{\mu}_0$ is the set of true population moments under the distorted probability measure
- Asymptotically normal such that $\sqrt{n}(\hat{\mu} - \hat{\mu}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Omega})$ as $n \rightarrow \infty$

Following the results on extremum estimation presented in Theorems 2.7 and 3.1 by Newey and McFadden [3], we can show that the K -vector of Lagrange multipliers $\hat{\lambda}$ is consistent such that $\hat{\lambda} \xrightarrow{P} \lambda_0$ as $n \rightarrow \infty$ where λ_0 is the set of true Lagrange multipliers under the distorted probability model. We can also derive the limiting normal distribution of the Lagrange multiplier estimators as

$$\sqrt{n}(\hat{\lambda} - \lambda_0) \xrightarrow{d} N(\mathbf{0}, \psi^{-1} \mathbf{\Omega} \psi^{-1}) \quad (16)$$

where

$$\psi = E \left[\frac{\partial^2 \mathcal{M}}{\partial \lambda \partial \lambda'} \bigg|_{\lambda = \lambda_0} \right] = -E_{F^*}[(\mathbf{G}(S(X)) - \hat{\mu})(\mathbf{G}(S(X)) - \hat{\mu})'] \quad (17)$$

is the negative variance matrix of $\mathbf{G}(S(X)) = (G_1(S(X)), \dots, G_K(S(X)))'$ under the distorted probability model F^* .

To estimate the asymptotic variance matrix of $\hat{\lambda}$, we could directly evaluate the elements of ψ

$$\hat{\Psi}_{kk} = - \int (G_k(S(x)) - \hat{\mu}_k)^2 f^*(x) dx \quad (18)$$

$$\hat{\Psi}_{jk} = - \int (G_j(S(x)) - \hat{\mu}_j)(G_k(S(x)) - \hat{\mu}_k) f^*(x) dx \quad (19)$$

by numerical integration under the distorted distribution f^* . Alternatively, we can rely on the special structure of the distorted density function f^* to estimate ψ by importance sampling from the reference distribution f . Given m random draws from f (denoted x_1, \dots, x_m), we can form the variance estimates as

$$\hat{\Psi}_{kk} = -m^{-1} \sum_{i=1}^m (G_k(S(x_i)) - \hat{\mu}_k)^2 w_i \quad (20)$$

$$\hat{\Psi}_{jk} = -m^{-1} \sum_{i=1}^m (G_j(S(x_i)) - \hat{\mu}_j)(G_k(S(x_i)) - \hat{\mu}_k) w_i \quad (21)$$

where

$$w_i \equiv \frac{\exp(\sum_{k=1}^K \hat{\lambda}_k G_k(S(x_i)))}{m^{-1} \sum_{i=1}^m \exp(\sum_{k=1}^K \hat{\lambda}_k G_k(S(x_i)))} \quad (22)$$

Under random sampling from f , the simulated variance components in (20) and (21) converge in probability to the estimates in (18) and (19) as $m \rightarrow \infty$.

We can also use importance sampling to estimate the coherent risk measure ρ_g for a particular data set. Although we could substitute (11) into (5) and use direct numerical integration to compute $\hat{\rho}_g(X)$, the importance sampling estimator

$$\hat{\rho}_g(X) = m^{-1} \sum_{i=1}^m x_i w_i \quad (23)$$

is consistent under random sampling from f as $m \rightarrow \infty$. For example, suppose we have a net loss variable X that is distributed as a standard lognormal random variable with mean 1.649 and variance 4.671. The VaR with $\alpha = 0.05$ from (1) is $\rho_{VaR} = 5.18$, and the CTE at the same level of significance from (2) is $\rho_{CTE} = 8.55$. Using these actual values as benchmarks, we can evaluate the performance of the importance sampling estimators of VaR and CTE based on $m = 5,000$ replications and various sample sizes (n). The simulation results are presented in Table 1, and the sample bias and estimated standard errors decline as n increases, as expected.

5 Joint estimation of the reference and distorted models

In practice, we may also have limited information about an appropriate specification of the reference distribution f for the net loss outcomes. In this section,



Table 1: Importance sampling results, VaR and CTE.

Sample size n	$\hat{\rho}_{VaR}$	Std. error	$\hat{\rho}_{CTE}$	Std. error
250	5.0020	0.6529	8.3020	1.5030
500	5.1518	0.4792	8.5591	1.1368
1000	5.1673	0.3411	8.5339	0.7936
2500	5.1740	0.2209	8.5520	0.5087

we show how the minimum cross-entropy approach may be extended to jointly estimate f and f^* from available data. In this case, we assume that we have a reference distribution $q(x)$ that represents our initial (pre-sample) estimate of f , and we use some sample moments related to the original data distribution (F) to update this estimate and form f . In particular, we can minimize the cross-entropy between f^* and q under the following Kullback–Leibler objective function

$$I(f^*, q) = \int \ln\left(\frac{f^*(x)}{q(x)}\right) f^*(x) dx \quad (24)$$

subject to an expanded set of moment conditions

$$\int G_k(S(x)) f^*(x) dx = \mu_k \quad \text{and} \quad \int H_j(x) f^*(x) dx = \theta_j \quad (25)$$

for $k = 1, \dots, K$ and $j = 1, \dots, L$ plus the additivity constraint in (7). Here, we use L additional moment conditions of the form $E_{F^*}[H_j(x)] = \theta_j$ to describe the properties of f and f^* .

Following the discussion in Section 2, we can show that the minimum cross-entropy solution for the distorted probability model is

$$f^*(x) = \frac{q(x) \exp(\sum_{j=1}^L \delta_j H_j(x) + \sum_{k=1}^K \lambda_k G_k(S(x)))}{\int q(x) \exp(\sum_{j=1}^L \delta_j H_j(x) + \sum_{k=1}^K \lambda_k G_k(S(x))) dx} \quad (26)$$

where δ_j is the Lagrange multiplier for the j^{th} moment constraint imposed on f^* in (25). The implicit form for the reference distribution is

$$f(x) = \frac{q(x) \exp(\sum_{j=1}^L \delta_j H_j(x))}{\int q(x) \exp(\sum_{j=1}^L \delta_j H_j(x)) dx} \quad (27)$$

and the derivative of the distortion function is

$$g'(S(x)) = \frac{\exp(\sum_{k=1}^K \lambda_k G_k(S(x))) \int q(x) \exp(\sum_{j=1}^L \delta_j H_j(x)) dx}{\int q(x) \exp(\sum_{j=1}^L \delta_j H_j(x) + \sum_{k=1}^K \lambda_k G_k(S(x))) dx} \quad (28)$$

such that $f^*(x)/f(x) = g'(S(x))$. Finally, the distortion function g may be recovered from (28) by integration as in (11).

To derive the large-sample properties of the model parameter estimators, we let $\beta = (\delta_1, \dots, \delta_L, \lambda_1, \dots, \lambda_K)'$ and $k = (\theta_1, \dots, \theta_L, \mu_1, \dots, \mu_K)'$, and we then assume that $\hat{k} \xrightarrow{P} k_0$ and

$$\sqrt{n}(\hat{k} - k_0) \xrightarrow{d} N(0, \Phi) \quad (29)$$

as $n \rightarrow \infty$. The results for extremum estimators provided by Newey and McFadden [3] may be used to show that the vector of Lagrange multiplier estimators is consistent such that $\hat{\beta} \xrightarrow{P} \beta_0$ as $n \rightarrow \infty$, and the limiting distribution of the estimators is

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \Gamma^{-1} \Phi \Gamma^{-1}) \quad (30)$$

where

$$\Gamma = -E_{F^*}[(A(X) - k)(A(X) - k)'] \quad (31)$$

and $A = (H_1(X), \dots, H_L(X), G_1(S(X)), \dots, G_K(S(X)))'$. As we show in the preceding section, we can use importance sampling or other convenient numerical integration tools to estimate the components of the asymptotic variance matrix for $\hat{\beta}$. Then, we can use the estimated asymptotic variance matrix to conduct classical hypothesis tests or form other inferences related to the fitted probability model and estimates of the coherent risk measure.

6 Conclusions

In summary, the proposed methods allow us to apply the recent theory related to coherence, distorted risk measures, and information theory to problems in empirical finance. The key innovations in this paper include an unconstrained computation algorithm and the statistical properties of the coherent risk estimators, and we use a few brief but familiar examples to demonstrate these claims.

References

- [1] Wirth, J. and Hardy, M., A synthesis of risk measures for capital adequacy. *Insurance: Mathematics and Economics*, **25**, pp. 337–47, 1999.
- [2] Artzner, P., Delbaen, F., Eber, J.M. and Heath, D., Coherent measures of risk. *Mathematical Finance*, **3**, pp. 203–28, 1999.
- [3] Newey, W. and McFadden, D., Large sample estimation and hypothesis testing (Chapter 36). *Handbook of Econometrics, Volume IV*, eds. R. Engle and D. McFadden, Elsevier Science: New York, pp. 2111–2245, 1994.
- [4] Denneberg, D., *Non-Additive Measure and Integral*. Kluwer Academic Publishers: Dordrecht, the Netherlands, 1994.
- [5] Golan, A., Judge, G. and Miller, D., *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. John Wiley and Sons: Chichester, UK, 1996.
- [6] Kullback, S., *Information Theory and Statistics*. John Wiley and Sons: New York, 1959.



- [7] McLeish, D. and Reesor, R.M., Risk, entropy, and the transformation of distributions. *North American Actuarial Journal*, **7**, pp. 128–44, 2003.
- [8] Wang, S., A class of distortion operators for pricing financial and insurance risks. *The Journal of Risk and Insurance*, **67**, pp. 15–36, 2000.
- [9] Jaynes, E.T., Information theory and statistical mechanics, I. *Physical Review*, **106**, pp. 620–30, 1957.
- [10] Jaynes, E.T., Information theory and statistical mechanics, II. *Physical Review*, **108**, pp. 171–90, 1957.
- [11] Wang, S. and Dhaene, J., Comonotonicity, correlation order, and premium principles. *Insurance: Mathematics and Economics*, **22**, pp. 235–42, 1998.
- [12] Ormoneit, D. and White, H., An efficient algorithm to compute maximum entropy densities. *Econometric Reviews*, **18**, pp. 127–40, 1999.
- [13] Groes, E., Jacobsen, H., Sloth, B. and Traneas, T., Axiomatic characterizations of the Choquet integral. *Economic Theory*, **12**, pp. 441–48, 1998.
- [14] Wirth, J. and Hardy, M., Distortion risk measures: Coherence and stochastic dominance, 2000. working paper, Heriot–Watt University.