PESSIMISTIC PORTFOLIO ALLOCATION AND CHOQUET EXPECTED UTILITY

GILBERT W. BASSETT, JR., ROGER KOENKER, AND GREGORY KORDAS

Abstract. Recent developments in the theory of choice under uncertainty and risk yield a pessimistic decision theory that replaces the classical expected utility criterion with a Choquet expectation that accentuates the likelihood of the least favorable outcomes. A parallel theory has recently emerged in the literature on risk assessment. It is shown that a general form of pessimistic portfolio optimization based on the Choquet approach may be formulated as a problem of linear quantile regression.

1. Introduction

The cornerstone of the modern theory of decision making under risk is expected utility maximization as elaborated by Bernoulli (1738), Ramsey (1931), de Finetti (1937), von Neumann and Morgenstern (1944), Savage (1954), Fishburn (1982) and many subsequent authors. And yet there have been persistent doubts expressed about the adequacy of the expected utility framework to encompass the full range of preferences over risky alternatives. Many variations on the the expected utility theme have been proposed, of these perhaps the most successful has been the family of non-additive, or dual, or rank-dependent formulations of Schmeidler (1989), Yaari (1987), and Quiggin (1982). The familiar Lebesgue integral of the expected utility computation is replaced by a Choquet integral, thereby permitting, for example, the probability weights associated with the least favorable outcomes to be accentuated and thereby yielding a pessimistic decision criterion.

Our aim in this paper is to provide an elementary exposition of one variant of the Choquet expected utility theory as it applies to decisions under risk, to link this theory with some recent developments in the literature on risk assessment, and finally to describe its application to the problem of portfolio choice. The problem of portfolio allocation is central to economic decision theory and thus in our view constitutes an acid test of any new criterion designed to evaluate risky alternatives.

Key words and phrases. Expected Utility, Choquet Capacity, Quantile Regression.

Version June 21, 2004. Corresponding author: Roger Koenker, Department of Economics, University of Illinois, Champaign, Illinois, 61820; (email: rkoenker@uiuc.edu). This research was partially supported grant SES-02-40781. The authors would like to thank Quanshui Zhao, Steve Portnoy and two anonymous referees for helpful comments on this work.

By offering a general approach to portfolio allocation for pessimistic Choquet preferences we hope to encourage a critical reexamination of the role of attitudes toward risk in this important setting. In contrast to conventional mean-variance portfolio analysis implemented by solving least squares problems, pessimistic portfolios will be constructed by solving quantile regression problems. Our results complement recent work in the financial econometrics literature on estimation and inference in conditional quantile models, as e.g. in Engle and Manganelli (2001), Bouyé and Salmon (2002), Christoffersen and Diebold (2003), and Linton and Whang (2003).

2. Choquet Expected Utility

Consider the problem of choosing between two real-valued random variables, X and Y, having distribution functions, F and G, respectively. Comparing expected utilities, we prefer X to Y if

$$E_F u(X) = \int_{-\infty}^{\infty} u(x) dF(x) > \int_{-\infty}^{\infty} u(x) dG(x) = E_G u(Y)$$

In this framework, the utility function, u, bears the entire burden of representing the decision makers' attitudes toward risk. Changing variables, we have the equivalent formulation that X is preferred to Y if

$$E_F u(X) = \int_0^1 u(F^{-1}(t)) dt > \int_0^1 u(G^{-1}(t)) dt = E_G u(Y), \tag{2.1}$$

in effect, we have transformed the original integrand into "equally likely" slices on the interval $[0, 1]$, now integrable with respect to Lebesgue measure. This formulation of expected utility in terms of the quantile functions $F^{-1}(t)$ of X and $G^{-1}(t)$ of Y is particularly revealing since it offers a sharp contrast to the Choquet theory.

In its simplest form Choquet expected utility introduces the possibility that preferences may require a distortion of the original probability assessments, rather than integrating dt we are permitted to integrate, $d\nu(t)$, with respect to some other probability measure defined on the interval $[0, 1]$. Preferences are thus represented by a pair of functions (u, ν): u transforms (monetary) outcomes into utility terms, while ν transforms probability assessments.[1] Now X is preferred to Y if

$$E_{\nu,F} u(X) = \int_0^1 u(F^{-1}(t)) d\nu(t) > \int_0^1 u(G^{-1}(t)) d\nu(t) = E_{\nu,G} u(Y). \tag{2.2}$$

Since our change of variables has ordered the events, $u(F^{-1}(t))$ and $u(G^{-1}(t))$ are also ordered according to increasing desirability. We assume throughout that u is monotone increasing. The distortion function ν acts to inflate or deflate the probabilities *according to the rank ordering of the outcomes*. The distortion may systematically

[1]Whether this latter transformation is a matter of "perception" or "attitude" or some other category of psychological experience has been a matter of some debate. We prefer Yaari's assertion that the theory describes "how perceived risk is processed into choice."

increase the implicit likelihood of the least favorable events, for example. But, unlike the original form of the prospect theory of Kahneman and Tversky (1979), it cannot systematically distort small probabilities *per se*. In contrast, cumulative prospect theory, as developed by Tversky and Kahneman (1992) and Wakker and Tversky (1993), is quite closely aligned with the Choquet approach.

Admittedly, there is something mildly schizophrenic about decision makers who accept at face value probabilities represented by the distribution functions F and G, and then intentionally distort these probabilities in the process of turning them into actions. More general forms of Choquet expected utility, following Schmeidler (1989), dispense entirely with initial probabilities. Monotone set functions, or Choquet capacities, are employed to assign nonadditive probabilities directly to events, enabling the theory to address distinctions between risk, uncertainty and ambiguity that have played a prominent role in the recent literature. By restricting attention to settings in which the Choquet capacity is expressed as the composition of an increasing function $\nu : [0,1] \to [0,1]$ with $\nu(0) = 0$ and $\nu(1) = 1$, and a probability measure we sacrifice some of the power and elegance of the general theory, but in return we obtain a very concise representation of preferences in terms of a pair of simple and easily interpretable functions (u, ν). Following Yaari (1987), we will focus our attention on the function ν, first interpreting it in terms of the possible optimism or pessimism of the decision maker, then briefly describing its axiomatic underpinnings.[2] A brief exposition of the interrelationship between various forms of the Choquet expectation is provided in an Appendix.

2.1. **Pessimism.** Choquet expected utility leads naturally to an interpretation of the distorting probability measure ν as a reflection of optimism or pessimism. If the distortion function, ν, is concave, then the least favorable events receive increased weight and the most favorable events are discounted reflecting pessimism. To see this, it may be helpful to consider the case that ν is absolutely continuous as well as concave and therefore has a decreasing density with respect to Lebesgue measure. Thus, instead of the uniform weighting implicit in the expected utility criterion, the Choquet criterion accentuates the weight of the least favorable events and reduces the weight assigned to the most favorable events. When ν is convex the situation is reversed, optimism prevails, and the Choquet integral exaggerates the likelihood of the more favorable events and downplays the likelihood of the worst outcomes. As noted already by Quiggin (1982) and Schmeidler (1989), distortion functions that are initially concave and then convex offer an alternative to the classical Friedman and

[2]Wakker (1990) has shown that Schmeidler's general theory of Choquet preferences coincides with the rank dependent, or anticipated utility, formulations of Quiggin and Yaari for decision making under risk, that is when choices are over given probability distributions.

Savage (1948) rationale for the widespread, but still puzzling, observation that many people seem to be willing to buy insurance and gamble at (nearly) the same time.[3]

The simplest distortions are those given by the one parameter family, $\nu_\alpha(t) = \min\{t/\alpha, 1\}$ for $\alpha \in [0,1]$. In this case, we have,

$$E_{\nu_\alpha} u(X) = \alpha^{-1} \int_0^\alpha u(F^{-1}(t))dt,$$

and we see that – relative to the usual computation of expected utility – the probabilities of the α least favorable outcomes are inflated and the $1-\alpha$ proportion of most favorable outcomes are discounted entirely. This family will play an important role as a building block for more general distortion functions in the sequel.

2.2. **Axiomatics.** Recall that if $\mathcal{A}$ denotes a mixture space of "acts" consisting of functions mapping states of nature, S to a set of probability distributions on outcomes, then the von Neumann Morgenstern axioms impose the following condition on the binary preference relation ($\succ$) over $\mathcal{A}$:

(A) (Independence) For all $f, g, h \in \mathcal{A}$ and $\alpha \in (0,1)$, then $f \succ g$ implies $\alpha f + (1-\alpha)h \succ \alpha g + (1-\alpha)h$,

Schmeidler's seminal 1989 paper argues that the independence axiom "may be too powerful to be acceptable", and he proposes a weaker form of independence:

(A') (Comonotonic Independence) For all pairwise comonotonic acts $f, g, h \in \mathcal{A}$ and $\alpha \in (0,1)$, then $f \succ g$ implies $\alpha f + (1-\alpha)h \succ \alpha g + (1-\alpha)h$,

Comonotonicity, a concept introduced by Schmeidler, plays a pivotal role.

Definition 1. *Two acts f and g in $\mathcal{A}$ are comonotonic if for no s and t in S, $f(s) \succ f(t)$ and $g(t) \succ g(s)$.*

By restricting the applicability of the independence axiom only to comonotonic acts, the force of the axiom is considerably weakened. Schmeidler (1986, 1989) shows that the effect of replacing (A) by (A') is precisely to enlarge the scope of preferences from those representable by a von Neumann-Morgenstern (Savage) utility function, u, to those representable by a pair, (u, ν), consisting of a utility function and a Choquet capacity.

To clarify the role of comonotonicity it is useful to consider the following reformulation of the definition in terms of random variables.

[3]Whether optimism and pessimism will eventually find a legitimate place in formal decision theory is likely to remain somewhat controversial – it seems advisable to recall that Savage (1954, p. 68) did not seem to think that they could be acommodated into his view of subjective expected utility. But given the long-standing dissatisfaction with expected utility theory it is surely worthwhile to continue to explore the implications of alternative schemes particularly when they are sufficiently tractable to yield concrete competing predictions about behavior for problems of real economic significance.

Definition 2. *Two random variables $X, Y : \Omega \to \mathbb{R}$ are comonotonic if there exists a third random variable $Z : \Omega \to \mathbb{R}$ and increasing functions f and g such that $X = f(Z)$ and $Y = g(Z)$.*

In the language of classical rank correlation, comonotonic random variables are perfectly concordant, i.e. have rank correlation 1. The classical Fréchet bounds for the joint distribution function, H, of two random variables, X and Y, with univariate marginals F and G is given by

$$\max\{0, F(x) + G(y) - 1\} \leq H(x, y) \leq \min\{F(x), G(y)\}.$$

Comonotonic X and Y achieve the upper bound.

From our point of view a crucial property of comonotonic random variables is the behavior of the quantile functions of their sums. For comonotonic random variables X, Y, we have

$$F^{-1}_{X+Y}(u) = F^{-1}_X(u) + F^{-1}_Y(u).$$

This is because, by comonotonicity we have a $U \sim U[0, 1]$ such that $Z = h(U) = F^{-1}_X(U) + F^{-1}_Y(U)$ where h is left continuous and increasing, so by monotone invariance of the quantile function we have, $F^{-1}_{h(U)} = h \circ F^{-1}_U = F^{-1}_X + F^{-1}_Y$.

This property will play an essential role in our formulation of pessimistic portfolio theory, a topic that we will introduce with a brief review of some closely related recent work in the theory of risk assessment. It is worth emphasizing that focusing attention on comonotonic risks is not intended to suggest that such pairings arise commonly in the natural history of finance, but only that the independence axiom – when restricted to such pairings – is more compelling than when it is claimed to apply in full generality.

3. Measuring Risk

Motivated by regulatory concerns in the financial sector there has been considerable recent interest in the question of how to measure portfolio risk. An influential paper in this literature is Artzner, Delbaen, Eber, and Heath (1999), which provides an axiomatic foundation for "coherent" risk measures.

Definition 3. *(Artzner et al) For real valued random variables $X \in \mathcal{X}$ on $(\Omega, \mathcal{A})$ a mapping $\varrho : \mathcal{X} \to \mathbb{R}$ is called a coherent risk measure if it is:*

(i) *Monotone: $X, Y \in \mathcal{X}$, with $X \leq Y \Rightarrow \varrho(X) \geq \varrho(Y)$.*
(ii) *Subadditive: $X, Y, X + Y \in \mathcal{X}$, $\Rightarrow \varrho(X + Y) \leq \varrho(X) + \varrho(Y)$.*
(iii) *Linearly Homogeneous: For all $\lambda \geq 0$ and $X \in \mathcal{X}$, $\varrho(\lambda X) = \lambda\varrho(X)$.*
(iv) *Translation Invariant: For all $\lambda \in \mathbb{R}$ and $X \in \mathcal{X}$, $\varrho(\lambda + X) = \varrho(X) - \lambda$.*

These requirements rule out many of the conventional measures of risk traditionally used in finance. In particular, measures based on second moments are ruled out by the monotonicity requirement, and quantile based measures like the value-at-risk are

ruled out by subadditivity. A measure of risk that *is* coherent and that has gained considerable recent prominence in the wake of these findings is,

$$\varrho_{\nu_\alpha}(X) = -\int_0^1 F^{-1}(t)d\nu(t) = -\alpha^{-1}\int_0^\alpha F^{-1}(t)dt,$$

where $\nu(t) = \min\{t/\alpha, 1\}$ as in Section 2. Variants of $\varrho_{\nu_\alpha}(X)$ have been suggested under a variety of names, including expected shortfall (Acerbi and Tasche (2002)), conditional value at risk (Rockafellar and Uryasev (2000)), tail conditional expectation (Artzner, Delbaen, Eber, and Heath (1999)). For the sake of brevity we will call $\varrho_{\nu_\alpha}(X)$ the α-risk of the random prospect X. Clearly, α-risk is simply the negative Choquet ν_α expected return.

Having defined α-risk in this way, it is natural to consider the criteria: $\varrho_{\nu_\alpha}(X) - \lambda\mu(X)$, or $\mu(X) - \lambda\varrho_{\nu_\alpha}(X)$. Minimizing the former criterion may be viewed as minimizing risk subject to a constraint on mean return; maximizing the latter criterion may be viewed as maximizing return subject to a constraint on α-risk. Several authors, including Denneberg (1990), Rockafellar and Uryasev (2000), Jaschke and Küchler (2001), and Ruszczyski and Vanderbei (2003), have suggested criteria of this form as alternatives to the classical Markowitz criteria in which α-risk is replaced by the standard deviation of the random variable X. Since $\mu(X) = \int F_X^{-1}(t)dt = -\varrho_{\nu_1}(X)$ these criteria are special cases of the following more general class.

Definition 4. *A risk measure ϱ is pessimistic if, for some probability measure φ on $[0, 1]$,*

$$\varrho(X) = \int_0^1 \varrho_{\nu_\alpha}(X)d\varphi(\alpha).$$

To see why such risk measures should be viewed as pessimistic, note that by the Fubini Theorem we can write,

$$\varrho(X) = -\int_0^1 \alpha^{-1}\int_0^\alpha F^{-1}(t)dtd\varphi(\alpha) = -\int_0^1 F^{-1}(t)\int_t^1 \alpha^{-1}d\varphi(\alpha)dt.$$

In the simplest case, we can take φ as a finite sum of (Dirac) point masses, say $d\varphi = \sum_{i=0}^m \varphi_i\delta_{\tau_i}$ with $\varphi_i \geq 0$, $\sum \varphi_i = 1$, and $0 = \tau_0 < \tau_1 < ... < \tau_m \leq 1$. Noting that

$$\int_t^1 \alpha^{-1}\delta_\tau(\alpha)d\alpha = \tau^{-1}I(t < \tau),$$

we can write

$$\varrho(X) = \int_0^1 \varrho_{\nu_\alpha}(X)d\varphi(\alpha) = -\varphi_0 F^{-1}(0) - \int_0^1 F^{-1}(t)\gamma(t)dt \tag{3.1}$$

where $\gamma(t) = \sum_{i=1}^m \varphi_i\tau_i^{-1}I(t < \tau_i)$. Positivity of the point masses, φ_i, assures that the resulting density weights are decreasing, so the resulting distortion in probabilities

acts to accentuate the implicit likelihood of the least favorable outcomes, and depress the likelihood of the most favorable ones. Such preferences are clearly "pessimistic".[4]

We may interpret the α-risks, $\varrho_\alpha(\cdot)$ as the extreme points of the convex set of coherent risk measures, and this leads to a nice characterization result. Following Kusuoka (2001) we will impose some additional regularity conditions on ϱ:

Definition 5. *Let $\mathcal{L}^\infty$ denote the space of all bounded real-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{P}$ non-atomic. A map $\varrho : \mathcal{L}^\infty \to \mathbb{R}$ is a regular risk measure if:*

(i.) *ϱ is law invariant, i.e. $\varrho(X) = \varrho(Y)$ if $X, Y \in \mathcal{L}^\infty$ have the same probability law.*

(ii.) *ϱ is comonotonically additive, i.e. $X, Y \in \mathcal{L}^\infty$ comonotone implies that $\varrho(X + Y) = \varrho(X) + \varrho(Y)$.*

The first condition requires only that ϱ depend upon observable properties of its argument. Random variables with the same distribution functions should have the same risk. The second condition refines slightly the subadditivity property: subadditivity becomes additivity when X and Y are comonotone. We can now succinctly reformulate the characterization result of Kusuoka (2001) and Tasche (2002).

Theorem 1. *A regular risk measure is coherent if and only if it is pessimistic.*

This result justifies our earlier comment that the elementary α-risks are fundamental building blocks: any concave distortion function, ν, can be approximated by a piecewise linear concave function. These linear approximants are nothing but the weighted sums of Dirac's introduced above. They correspond to risk functions of the form (3.1) with $\gamma(t) = \sum_{i=1}^m \varphi_i \tau_i^{-1} I(t < \tau_i)$ piecewise constant and decreasing. Note that as a special case such ν's are entitled to place positive weight φ_0 on the least favorable outcome, so they include the extreme case of maximin preferences. Note also that pessimistic risk measures respect second order stochastic dominance, see e.g. De Giorgi (2004).

Example: A simple, yet intriguing, class of pessimistic ν functions is the one-parameter family,

$$\nu_\theta(t) = 1 - (1-t)^\theta \qquad \theta \geq 1.$$

[4] A referee has suggested that there may be interesting connections of this theory to the "safety first" principle of Roy (1952) according to which the investor attempts to minimize the probability that a portfolio's return falls below a prespecified level, subject perhaps to a constraint on mean return. While measures of risk based on such a "probability of ruin" seem attractive and have received considerable attention, they do not fall easily within the purview of our analysis. They share some of the same difficulties as value-at-risk since they ignore the magnitude of the disasters lurking within the permitted probability of ruin. Similarly, weakening the Artzner *et. al.* axioms by replacing (ii.) and (iii.) by convexity has some appeal. But given linear homogeneity, a risk measure is convex if and only if it is subadditive, so one would have to abandon linear homogeneity to move in this direction, and this would require us to abandon the Choquet characterization of our risk measures.

Members of the family range from the perfectly equanimous at $\theta = 1$, to the profoundly pessimistic as $\theta \to \infty$. Note that, changing variables, so $t = F(u)$, we have,

$$E_{\nu_\theta} X = \int_0^1 F^{-1}(t) d\nu(t) = \int_{-\infty}^{\infty} u dG(u),$$

where $G(u) = 1 - (1 - F(u))^\theta$. When θ is a positive integer, say m, this Choquet expectation has a nice interpretation: $G(u)$ is the distribution function of the minimum of m independent realizations of the random variable X. The pessimist imagines not a single occurence of X, but m such realizations among which he will get the worst. That eternal maxim of the cautious aunt and misanthropic uncle, "expect the worst, and you won't be disappointed," is transformed into a precise calculus. In the process the parameter, θ, becomes a natural measure of the degree of pessimism. ■

4. Pessimistic Portfolio Allocation

We have seen that decision making based on minimizing a "coherent" measure of risk as defined by Artzner, Delbaen, Eber, and Heath (1999) is equivalent to Choquet expected utility maximization using a linear form of the utility function and a concave distortion function ν. But the question remains: Does the Choquet approach to decision making under risk lead to tractable methods for analyzing real decision problems of practical importance? An affirmative answer to this question is provided in this section.

4.1. On α-risk as Quantile Regression.

Empirical strategies for minimizing α-risk lead immediately to the methods of quantile regression. Let

$$\rho_\alpha(u) = u(\alpha - I(u < 0))$$

denote the piecewise linear (quantile) loss function and consider the problem,

$$\min_{\xi \in \mathbb{R}} E\rho_\alpha(X - \xi).$$

It is well known, *e.g.* Ferguson (1967), that any ξ solving this problem is an αth quantile of the random variable X. Evaluating at a minimizer, ξ_α we find that minimizing the usual α-quantile objective function is equivalent to evaluating the sum of expected return and the α-risk of X, and then multiplying by α.

Theorem 2. *Let X be a real-valued random variable with $EX = \mu < \infty$, then*

$$\min_{\xi \in \mathbb{R}} E\rho_\alpha(X - \xi) = \alpha(\mu + \varrho_{\nu_\alpha}(X)).$$

Proof: Noting that,

$$E\rho_\alpha(X - \xi) = \alpha(\mu - \xi) - \int_{-\infty}^{\xi} (x - \xi) dF_X(x),$$

is minimized when $\xi_\alpha = F_X^{-1}(\alpha)$, we have,

$$E\rho_\alpha(X - \xi_\alpha) = \alpha\mu - \int_0^\alpha F^{-1}(t)dt = \alpha\mu + \alpha\varrho_{\nu_\alpha}(X).$$

■

This result provides a critical link to the algorithmic formulation developed in the next subsection.

4.2. **How to be Pessimistic.** Given a random sample $\{x_i : i = 1, ..., n\}$ on X, an empirical analogue of α-risk can thus be formulated as

$$\hat{\varrho}_{\nu_\alpha}(x) = (n\alpha)^{-1} \min_{\xi \in \mathbb{R}} \sum_{i=1}^{n} \rho_\alpha(x_i - \xi) - \hat{\mu}_n \tag{4.1}$$

where $\hat{\mu}_n$ denotes an estimator of $EX = \mu$, presumably, $\bar{x}_n$. The advantage of the optimization formulation becomes apparent as soon as we begin to consider *portfolios*.

Let $Y = X^\top \pi$ denote a portfolio of assets comprised of $X = (X_1, ..., X_p)^\top$ with portfolio weights π. Suppose now that we observe a random sample $\{x_i = (x_{i1}, ..., x_{ip}) : i = 1, ..., n\}$ from the joint distribution of asset returns, and we wish to consider portfolios minimizing the Lagrangean,

$$\min_\pi \varrho_{\nu_\alpha}(Y) - \lambda\mu(Y). \tag{4.2}$$

This is evidently equivalent to simply minimizing $\varrho_{\nu_\alpha}(Y)$ subject to a constraint on mean return. Alternatively, we could consider maximizing mean return subject to a constraint on α-risk; this equivalent formulation might be interpreted as imposing a regulatory constraint on a risk neutral agent. Since $\mu(Y) = -\varrho_{\nu_1}(Y)$, the mean is also just another α-risk, so the problem (4.2) may be viewed as a special discrete example of the pessimistic risk measures of Definition 4. More general forms are introduced in Section 4.4.

We will impose the additional constraint that the portfolio weights π sum to one, and reformulate the problem as,

$$\min_\pi \varrho_{\nu_\alpha}(X^\top \pi) \; s.t. \; \mu(X^\top \pi) = \mu_0, \; 1^\top \pi = 1.$$

Taking the first asset as numeraire we can write the sample analogue of this problem as

$$\min_{(\beta,\xi) \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\alpha(x_{i1} - \sum_{j=2}^{p} (x_{i1} - x_{ij})\beta_j - \xi) \; s.t. \; \bar{x}^\top \pi(\beta) = \mu_0, \tag{4.3}$$

where $\pi(\beta) = (1 - \sum_{j=2}^p \beta_j, \beta^\top)^\top$. Note that the term $\hat{\mu}_n$ in (4.1) can be neglected since the mean is constrained to take the value μ_0, and the factor $(n\alpha)^{-1}$ can be absorbed into the Lagrange multiplier of the mean return constraint.

This is a linear quantile regression problem of the type considered by Koenker and Bassett (1978). Instead of solving for a scalar quantity representing the αth sample

quantile, we are solving for p coefficients of a linear function estimating the αth conditional quantile function of the numeraire return given the other returns. A very non-technical introduction to quantile regression is provided by Koenker and Hallock (2001), and further details may be found in Koenker (2005).

The problem (4.3) can be solved very efficiently by classical simplex methods, and even when n and p are large by interior point methods. It is easy to verify that the solution is invariant to the choice of the numeraire asset. At the solution, $\hat{\xi}$ is an αth sample quantile of the chosen portfolio's (in-sample) returns distribution. The required mean return constraint implicitly corresponds to a particular choice of λ in the original specification (4.2). Note that we have not (yet) imposed any further constraints on the portfolio weights β, but given the linear programming form of the problem (4.3) it is straightforward to do so. Software to compute solutions to the problems (5.2-4) subject to arbitrary linear inequality constraints, and written for the public domain dialect, R, of the S language of Chambers (1998), is available from the authors on request. The algorithms used are based on the interior point, log-barrier methods described in Portnoy and Koenker (1997).

The problem posed in (4.3) is (almost) a conventional quantile regression problem. The only minor idiosyncrasy is the mean return constraint, but it is easy to impose this constraint by simply adding a single pseudo observation to the sample consisting of response $\kappa(\bar{x}_1 - \mu_0)$ and design row $\kappa(0, \bar{x}_1 - \bar{x}_2, ..., \bar{x}_1 - \bar{x}_p)^\top$. The zero element corresponds to the intercept column of the design matrix. Given κ the problem is a completely standard linear quantile regression. For sufficiently large κ we are assured that the constraint will be satisfied. Varying μ_0 we would obtain an empirical α-risk frontier analogous to the Markowitz mean-variance frontier.

4.3. **An Example.** Consider forming portfolios from four independently distributed assets with marginal returns densities illustrated in Figure 1. Asset 1 is normally distributed with mean 0.05 and standard deviation 0.02. Asset 2 has a reversed χ^2_3 density with location and scale chosen so that its mean and variance are identical to those of asset 1. Similarly, assets 3 and 4 are constructed to have identical means (0.09) and standard deviations (0.07). But for the second pair of densities the χ^2_3 asset is skewed to the right, as usual, not to the left. The example is obviously chosen with malice aforethought. Asset 2 is unattractive relative to its twin, asset 1, having worse performance in both tails. And asset 3 is unattractive relative to asset 4 in both tails. But from a mean-variance viewpoint the two pairs of assets are indistinguishable.

To compare portfolio choices we generate a sample of $n = 1000$ observations from the returns distribution of the four assets. We then estimate optimal mean variance portfolio weights by solving

$$\min_{(\beta,\xi)\in\mathbb{R}^p} \sum_{i=1}^{n} (x_{i1} - \sum_{j=2}^{p} (x_{i1} - x_{ij})\beta_j - \xi)^2 \; s.t. \; \bar{x}^\top \pi(\beta) = \mu_0, \tag{4.4}$$

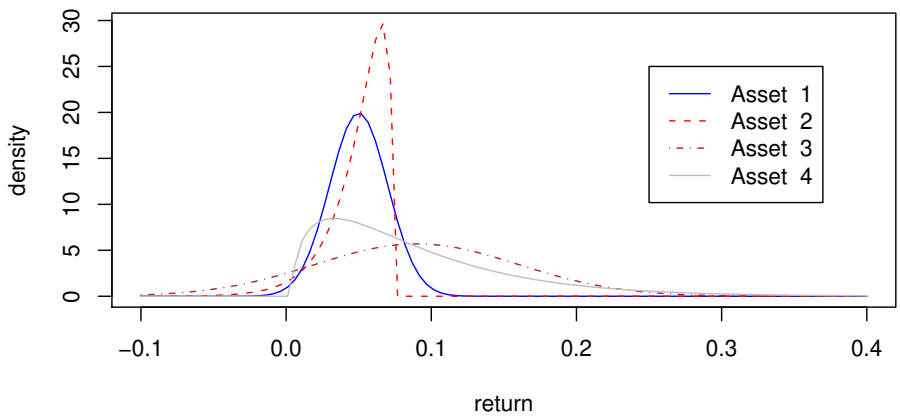

Figure 1. Four Asset Densities: Note that assets 1 and 2 have identical mean and variance of returns, likewise assets 3 and 4 have the same mean and variance. However, due to asymmetry of the returns distribution asset 1 has better performance in both tails that asset 2, and asset 4 has better performance than asset 3 in both tails.

where $\pi(\beta) = (1 - \sum_{j=2}^{p} \beta_j, \beta^\top)^\top$. We choose $\mu_0 = 0.07$, and obtain $\pi(\hat{\beta}) = (0.271, 0.229, 0.252, 0.248)$. We then estimate optimal α-risk portfolio weights for $\alpha = .1$ by solving (4.3) again with required mean return of $\mu_0 = 0.07$, obtaining $\pi(\hat{\beta}) = (0.303, 0.197, 0.123, 0.377)$. As expected, we allocate larger proportions to our "attractive" assets 1 and 4, and less to assets 2 and 3, compared to the mean variance portfolio.

To evaluate our two portfolio allocations we generate a new sample of $n = 100,000$ observations for the four assets, and we plot estimated densities of the two portfolios returns in Figure 2. It is apparent that the α-risk portfolio has better performance than the mean-variance (σ-risk) portfolio *in both tails*. It is safer in the lower tail, and also more successful in the upper tail. Improved performance in the lower tail should be expected from the Choquet portfolio since it is explicitly designed to maximize the expected return from the α least favorable outcomes. Improved performance in the upper tail may be attributed to the inability of the variance to distinguish upside from downside risk, and thus its desire to avoid both. The Choquet portfolio exhibits no such tendency. Risk is entirely associated with the behavior in the lower tail, and the expected return component of the objective function is able to exploit whatever opportunities exist in the upper tail.

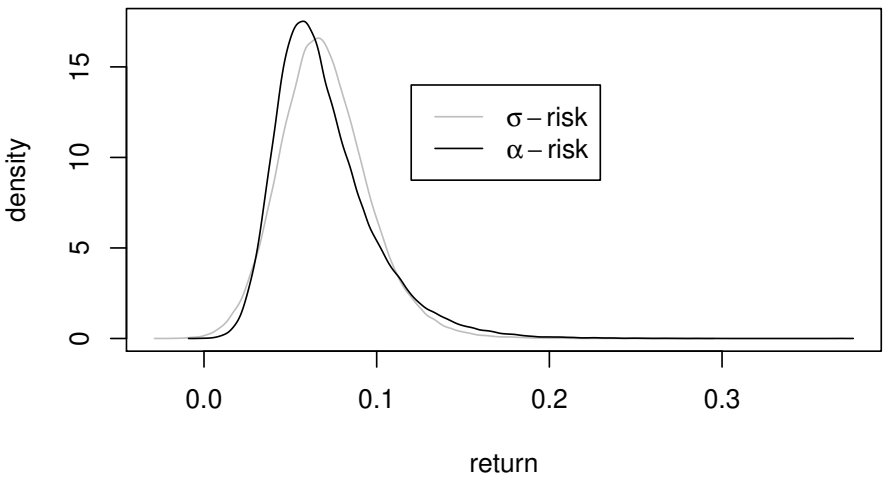

FIGURE 2. Markowitz (σ-risk) vs. Choquet (α-risk) portfolio returns densities: Portfolios are constrained to achieve the same in-sample mean return; note that the Choquet portfolio has better performance than the Markowitz portfolio in both the lower and upper tails.

The foregoing comparison is strongly conditioned by the requirement that the means of the two portfolios are identical. Our second comparison keeps the same σ-risk portfolio, but chooses the α-risk portfolio to maximize expected return subject to achieving the same α-risk as that achieved by the σ-risk portfolio. This yields a new α-risk portfolio with weights $\pi(\hat{\beta}) = (0.211, 0.119, 0.144, 0.526)$ that places even more weight on asset 4 and thereby achieves a mean return (in the initial sample) of 0.768. We plot the estimated returns densities of the two portfolios based on the $n = 100,000$ sample in Figure 3. Matching the α-risk in the lower tail (with $\alpha = 0.1$) we are able to achieve considerably better (68 basis points!) performance. Of course the variance and standard deviation of the new α-risk portfolio is larger than for its σ-risk counterpart, but this is attributable to its better performance in the upper tail, a result that underscores the highly unsatisfactory nature of second moment information as a measure of risk in asymmetric settings.

4.4. **Portfolio Allocation for General Pessimistic Preferences.** Although the α-risks provide a convenient one-parameter family of coherent risk measures, they are obviously rather simplistic. As we have already suggested, it is natural to consider

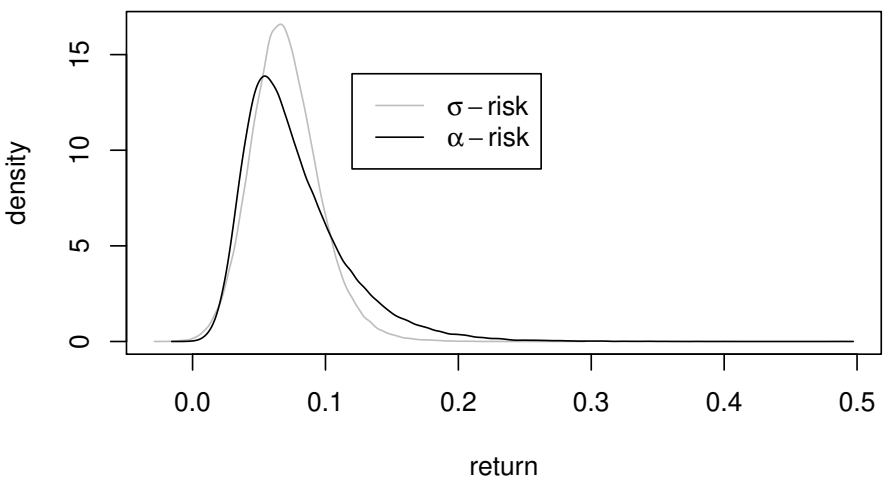

FIGURE 3. Markowitz (σ-risk) vs. Choquet (α-risk) portfolio returns densities: Portfolios are constrained to achieve the same in-sample α-risk; so performance in the lower tail is virtually identical, but the Choquet portfolio now has much better performance than the Markowitz portfolio in the upper tail.

weighted averages of α-risks:

$$\varrho_\nu(X) = \sum_{k=1}^{m} \nu_k \varrho_{\nu_{\alpha_k}}(X),$$

where the weights, $\nu_k : k = 1, ..., m$, are positive and sum to one. These general risk criteria can also be easily implemented empirically extending the formulation in (4.3)

$$\min_{(\beta,\xi)\in\mathbb{R}^{p+m}} \sum_{k=1}^{m}\sum_{i=1}^{n} \nu_k \rho_\alpha(x_{i1} - \sum_{j=2}^{p}(x_{i1} - x_{ij})\beta_j - \xi_k) \; s.t. \; \bar{x}^\top \pi(\beta) = \mu_0. \tag{4.5}$$

The only new wrinkle is the appearance of m distinct intercept parameters representing the m estimated quantiles of the returns distribution of the chosen portfolio. In effect we have simply stacked m distinct quantile regression problems on top of one another and introduced a distinct intercept parameter for each of them, while constraining the portfolio weights to be the same for each quantile. Since the ν_k are all positive, they may be passed inside the ρ_α function to rescale the argument. The asymptotic behavior of such constrained quantile regression estimators is discussed in Koenker (1984). Since any pessimistic distortion function can be represented by

a piecewise linear concave function generated as a weighted average of α-risks this approach offers a general solution to the problem of estimating pessimistic portfolio weights.

5. Conclusion

We have shown that a general form of pessimistic Choquet preferences under risk with linear utility leads to optimal portfolio allocation problems that can be solved by quantile regression methods. These portfolios may also be interpreted as maximizing expected return subject to a "coherent" measure of risk in the sense of Artzner, Delbaen, Eber, and Heath (1999). By providing a general method for computing pessimistic portfolios based on Choquet expected returns we hope to counter the impression occassionally seen in the literature that alternatives to the dominant expected utility view are either too vague or too esoteric to be applicable in problems of practical significance. There is a growing body of scholarship suggesting that pessimistic attitudes toward risk offer a valuable complement to conventional views of risk aversion based on expected utility maximization; portfolio allocation would seem to provide a valuable testing ground for further evaluation of these issues.

Appendix A. Choquet Capacities and the Choquet Integral

Since our formulation of Choquet expected utility in terms of quantile functions is somewhat less conventional than equivalent expressions formulated in terms of distribution functions, it is perhaps worthwhile to briefly describe the connection between the two formulations.

Let X be a real-valued random variable on Ω with an associated system of of sets, or "events", $\mathcal{A}$. A Choquet capacity, or non-additive measure, μ on $\mathcal{A}$, is a monotone set function (for $A, B \in \mathcal{A}$, $A \subseteq B$ implies that $\mu(A) \leq \mu(B)$), normalized so that $\mu(\emptyset) = 0$ and $\mu(\Omega) = 1$. The Choquet expectation of X, with respect to a capacity μ, is given by,

$$E_\mu X = \int X d\mu = \int_{-\infty}^{0} (\mu(\{X > x\}) - 1)dx + \int_0^{\infty} \mu(\{X > x\})dx$$

In this form, with μ defined directly on events, there is no necessity for any intermediate evaluation of probabilities. The Choquet expectation can accommodate various notions of ambiguity, or uncertainty, arising, for example, in the well-known examples of Ellsberg (1961). If, on the other hand, X is assumed to possess a distribution function F, and therefore a survival function, $\bar{F}(x) = 1 - F(x)$ we may write, $\mu(\{X > x\}) = \Gamma(\bar{F}(x))$, for some monotone function Γ such that $\Gamma(0) = 0$ and $\Gamma(1) = 1$. The Choquet expectation becomes,

$$E_\mu X = \int_{-\infty}^{0} (\Gamma(\bar{F}(x)) - 1)dx + \int_0^{\infty} \Gamma(\bar{F}(x))dx.$$

In the special case that μ is Lebesgue measure, Γ is the identity and we have the well-known expression,

$$EX = -\int_{-\infty}^{0} F(x)dx + \int_0^{\infty} (1 - F(x))dx.$$

If Γ is absolutely continuous with density γ, we can integrate by parts to obtain,

$$E_\mu X = \int_{-\infty}^{\infty} x d\Gamma(\bar{F}(x)) = -\int_{-\infty}^{\infty} x\gamma(\bar{F}(x))dF(x).$$

Finally, changing variables, $x \to F^{-1}(t)$, we have

$$\begin{aligned} E_\mu X &= -\int_0^1 F^{-1}(t)\gamma(1-t)dt \\ &= \int_0^1 F^{-1}(1-s)\gamma(s)ds \\ &= \int_0^1 F^{-1}(1-s)d\Gamma(s). \end{aligned}$$

The last expression reveals another slightly mysterious aspect of the conventions used in defining the Choquet capacity. The distortion Γ as defined above integrates the quantile function backwards. Thus following the conventions of Section 2, if we would like to define a distortion ν such that

$$\int_0^1 F^{-1}(s)d\nu(s) = \int_0^1 F^{-1}(1-s)d\Gamma(s),$$

then clearly we must have $\nu(s) = -\Gamma(1-s)$. Consequently, if Γ is convex, then ν will be concave, and vice versa. It is for this reason that pessimism, and uncertainty aversion, are often identified with convex capacities, as for example in Schmeidler (1989).

References

Acerbi, C., and D. Tasche (2002): "Expected Shortfall: A Natural coherent alternative to Value at Risk," *Economic Notes*, 31, 379–388.

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

Bernoulli, D. (1738): "Specimen theoriae novae de mensura sortis," *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 5, 175–192, translated by L. Sommer in *Econometrica*, 22, 23-36.

Bouyé, and M. Salmon (2002): "Dynamic Copula Quantile Regressions and Tail Area Dynamic Dependence in Forex Markets," preprint.

Chambers, J. M. (1998): *Programming with Data: A Guide to the S Language*. Springer.

Christoffersen, P., and F. Diebold (2003): "Financial asset returns, direction-of-change forecasting, and volatility dynamics," preprint.

de Finetti, B. (1937): "La prévision ses lois logiques, ses sources subjectives," *Annals de l'Instutute Henri Poincaré*, 7, 1–68, translated by H.E. Kyburg in *Studies in Subjective Probability*, H.E. Kyburg and H.E. Smokler, (eds.), Wiley, 1964.

De Giorgi, E. (2004): "Reward-Risk Portfolio Selection and Stochastic Dominance," *Journal of Banking and Finance*, forthcoming.

Denneberg, D. (1990): "Premium Calculation: Why standard deviation should be replaced by absolute deviation," *ASTIN Bulletin*, 20, 181–190.

Engle, R., and S. Manganelli (2001): "CAViaR: Conditional autoregressive value at risk by regression quantiles," preprint.

Ferguson, T. S. (1967): *Mathematical Statistics: A Decision Theoretic Approach*. iAcademic Press.

Fishburn, P. (1982): *Foundations of Expected Utility Theory*. D. Reidel.

Friedman, M., and L. Savage (1948): "The utility analysis of choices involving risk," *J. Political Economy*, 56, 279–304.

Jaschke, S., and U. Küchler (2001): "Coherent Risk Measure and Good Deal Bounds," *Finance and Stochastics*, 5, 181–200.

KAHNEMAN, D., AND A. TVERSKY (1979): "Prospect theory: An analysis of decision under risk," *Econometrica*, 47, 263–291.

KOENKER, R. (1984): "A Note on L-estimators for Linear Models," *Stat and Prob Letters*, 2, 323–325.

——— (2005): *Quantile Regression*. Cambridge U. Press.

KOENKER, R., AND G. BASSETT (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

KOENKER, R., AND K. HALLOCK (2001): "Quantile Regression," *Journal of Economic Perspectives*, 15, 143–156.

KUSUOKA, S. (2001): "On Law Invariant Coherent Risk Measures," *Advances in Math. Econ.*, 3, 83–95.

LINTON, O., AND Y.-J. WHANG (2003): "A quantilogram approach to evaluating directional predictability," preprint.

PORTNOY, S., AND R. KOENKER (1997): "The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-error Versus Absolute-error Estimators," *Statistical Science*, 12, 299–300.

QUIGGIN, J. (1982): "A Theory of Anticipated Utility," *J. of Econ. Behavior and Organization*, 3, 225–243.

RAMSEY, F. (1931): "Truth and Probability," in *The Foundations of Mathematics and Other Logical Essays*. Harcourt Brace.

ROCKAFELLAR, R., AND S. URYASEV (2000): "Optimization of conditional VaR," *J. of Risk*, 2, 21–41.

ROY, A. (1952): "Safety First and the Holding of Assets," *Econometrica*, 20, 434–449.

RUSZCZYSKI, A., AND R. VANDERBEI (2003): "Frontiers of stochastically nondominated portfolios," *Econometrica*, 7, 1287–1298.

SAVAGE, L. (1954): *Foundations of Statistics*. Wiley.

SCHMEIDLER, D. (1986): "Integral representation without additivity," *Proceedings of Am Math Society*, 97, 255–261.

——— (1989): "Subjective Probability and Expected Utility without Additivity," *Econometrica*, 57, 571–587.

TASCHE, D. (2002): "Expected Shortfall and Beyond," *Journal of Banking and Finance*, 26, 1519–1533.

TVERSKY, A., AND D. KAHNEMAN (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *J. Risk and Uncertainty*, 5, 297–323.

VON NEUMANN, J., AND O. MORGENSTERN (1944): *Theory of Games and Economic Behavior*. Princeton.

WAKKER, P. (1990): "Under Stochastic Dominance Choquet-Expected Utility and Anticipated Utility are Identical," *Theory and Decision*, 29, 119–132.

WAKKER, P., AND A. TVERSKY (1993): "An Axiomatization of Cumulative Prospect Theory," *J. Risk and Uncertainty*, 7, 147–176.

YAARI, M. (1987): "The Dual Theory of Choice under Risk," *Econometrica*, 55, 95–115.

DEPARTMENT OF FINANCE, UNIVERSITY OF ILLINOIS AT CHICAGO, 60607
E-mail address: gib@uic.edu

DEPARTMENT OF ECONOMICS, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, 61820
E-mail address: rkoenker@uiuc.edu

DEPARTMENT OF ECONOMICS, UNIVERSITY OF PENNSYLVANIA, 19104-6297
E-mail address: kordas@econ.upenn.edu