Disparity, Shortfall, and Twice-Endogenous HARA Utility*

M. Ryan Haley†, M. Kevin McGee‡and Todd B. Walker§

December 4, 2010

Abstract

We derive a mapping between the shortfall-minimizing portfolio selection based on higher-order entropy measures and expected utility theory. We show that the family of HARA utility functions has a minimum-divergence, shortfall-based representation. This facilitates an interpretation in which the risk aversion parameters and the type of risk aversion arise endogenously. We provide a numerical example illustrating this interpretation.

Keywords: Entropy, Measure Change, Cressie-Read, Endogenous Utility, Shortfall

JEL Classification Numbers: C44, G11, D03

*We would like to thank Essie Maasoumi, Chuck Whiteman and an anonymous referee for helpful comments.
†University of Wisconsin–Oshkosh; haley@uwosh.edu
‡University of Wisconsin–Oshkosh; mcgee@uwosh.edu
§Indiana University; walkertb@indiana.edu

1 Introduction

The purpose of this paper is to expose an interesting relationship between Hyperbolic Absolute Risk Aversion (HARA) expected utility, disparity minimization, and shortfall. Specifically, we show that the family of HARA expected utility functions has a minimum-divergence, shortfall-based representation, which means that HARA expected utility can be understood through the simple notion that the decision maker seeks the allocation of wealth to assets that minimizes the probability of realizing an outcome below some pre-determined target or benchmark level. This result bridges the intuitively appealing notion of shortfall minimization, first espoused by Roy (1952), with the now-familiar expected utility idea in a broad way.

The general spirit of this paper is to emphasize the close relationship between expected utility theory and information theory. Of course, this paper is not the first to discover such a relationship exists. Maasoumi (1986) makes an explicit connection between a class of inequality measures (including the well-known Theil index) and expected utility. As in this paper, he appeals to information theoretic arguments in constructing this mapping by showing that higher-order utility functions can be interpreted as ideal functionals based on relative entropies.[1] Our paper is most closely related to that of Stutzer (2000,2003). Specifically, we extend the endogenous utility arguments of Stutzer showing that his findings are special cases of a much more expansive relationship between shortfall, disparity, and conventional expected utility.

2 Theory

This section lays out the theory in which we derive the mapping between shortfall and the HARA class of expected utility functions within a general portfolio selection environment. The following section provides a numerical example.

2.1 Shortfall and Exponential Tilting

Suppose an investor desires the portfolio that minimizes the probability of realizing a return below some self-selected (or imposed) target or benchmark rate of return, denoted as d. Falling below this level results in a shortfall, with d being the point of shortfall. Let the portion of initial wealth W_0 allocated to asset j be denoted as w_j, and collect them in the vector $\mathbf{w} = (w_1, ..., w_J)$, where J indicates the number of admissible assets. Further assume that the returns of each asset are random variables, denoted as R_j. The shortfall probability is given by $\Pr[R(\mathbf{w}) \leq d]$ where $R(\mathbf{w}) = \sum_{j=1}^{J} w_j R_j$. Assuming a sample of T observations of returns are available, then the shortfall-minimizing investor's problem may be written as

$$\min_{\mathbf{w}} \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}_{(-\infty, \mathbf{d}]}[\mathbf{R_t}(\mathbf{w})] \tag{1}$$

where $\mathbf{1}$ denotes the indicator function. While direct computation of this expression delivers the true (in-sample) shortfall-minimizing portfolio, non-differentiability of (1) has forced

[1] Several recent papers have also derived relationships between classical projection problems with finance applications and expected utility theory [*e.g.*, Kramkov and Schachermayer (1999), Goll and Rüschendorf (2002), Bellini and Frittelli (2002), Föllmer (2008) Almeida and Garcia (2008,2009)]. The distinguishing feature of our paper is the focus on higher-order entropy measures, shortfall, and expected utility.

researchers to consider approximations to the actual shortfall probability [see Haley (2003) for an extended discussion of alternative approaches to approximate (1)].

One such approximation is given by Stutzer (2000, 2003) and Haley and Whiteman (2008), who rank portfolios by measuring the Kullback-Leibler (KL) disparity between the unweighted portfolio return distribution and the d-meaned "tilted" counterpart. Recall the KL divergence (or cross-entropy) from "real world" probabilities u_t to tilted probabilities π_t is defined as $\sum_{t=1}^{T} \pi_t \log(\pi_t/u_t)$. The objective is to chose a weighting scheme $\mathbf{w}$ such that the unweighted portfolio distribution is farthest, in the KL sense, away from its twisted counterpart; in practical terms, it is hoped that the portfolio that requires the most reweighting will be least likely to induce a shortfall return.

In *iid* samples, the objective function is a d-mean constrained minimum-disparity optimization problem:

$$\max_{\theta,\mathbf{w}} \sum_{t=1}^{T} \pi_t(\theta,\mathbf{w}) \log\left[\frac{\pi_t(\theta,\mathbf{w})}{u_t}\right] - \theta\Big[\sum_{t=1}^{T} \pi_t(\theta,\mathbf{w}) R_t(\mathbf{w}) - d\Big],$$

where

$$\pi_t(\theta,\mathbf{w}) = \frac{\exp[\theta R_t(\mathbf{w})]}{\sum_{t=1}^{T} \exp[\theta R_t(\mathbf{w})]},$$

and θ is a Lagrange multiplier.

Stutzer (2000,2003) provides an alternative interpretation of the parameter θ. He uses the Gärtner-Ellis Large Deviations Theorem [see, Bucklew (1990)] to calculate the decay rate of the shortfall $\Pr[R(\mathbf{w}) \le d] \to 0$ as $T \to \infty$. For a log portfolio return process with random log return, $\log R_t(\mathbf{w})$, the time average of the partial sums' log moment generating function is given by

$$\lim_{T\to\infty} \frac{1}{T} \log E[\exp(\theta \sum_{t=1}^{T} \log R_t(\mathbf{w}))] = \lim_{T\to\infty} \frac{1}{T} \log E\left[\left(\frac{W_T}{W_0}\right)^{\theta}\right] \tag{2}$$

where the equality follows from $W_T = W_0 \prod_{t=1}^{T} R_t(\mathbf{w})$. The Legendre-Fenchel transform of (2) gives the decay rate as $D_{\mathbf{w}}(\log d) \equiv \max_\theta \theta \log d - \phi(\theta)$. Stutzer shows the large T approximation of the maximum decay rate can be written as

$$-\exp(-\max_{\mathbf{w}} D_{\mathbf{w}}(\log d)) \approx \max_{\mathbf{w}} E\left[-\left(\frac{W_T}{W_0 d^T}\right)^{\theta_{\max(\mathbf{w})}}\right]. \tag{3}$$

Notice that (3) takes the form of power utility, where the argument is the ratio of invested wealth to the "benchmark" level of wealth as determined by the shortfall point. The utility specification is well justified in that the function limits to zero as its argument limits to infinity, is strictly concave, and has a constant degree of relative risk aversion $(1 - \theta_{\max(\mathbf{w})})$. The parameter θ is written as $\theta_{\max(\mathbf{w})}$ to emphasize that it is endogenously chosen as part of the optimization procedure (the Lagrange multiplier); hence, Stutzer's terminology "endogenous utility."

The notion of endogenous utility raises an interesting question: Does a broader relationship exist between shortfall-minimization, higher-order entropy, and expected utility theory?

In the next section, we show that posing the shortfall minimization portfolio selection problem in terms of a higher-order entropy family permits the mapping directly into the family of HARA expected utility functions. Therefore, HARA expected utility functions do in fact have a shortfall-based analog. This relationship also implies that HARA expected utility functions have endogenous expected utility analogs. We now show this formally.

2.2 Shortfall and Higher-Order Entropy Measures

Developing a generalized alternative, which nests the KL-based rules, is the focus of this section. Our approach involves higher-order entropy measures. In particular, we focus on the Cressie-Read (CR) divergence family; see Cressie and Read (1984) or Baggerly (1998). We focus on the Cressie-Read family because many members of this family are well known and it delivers both increasing and decreasing risk aversion, as we show below. However, other families exist as well [see, for example, Granger *et al.* (2004)] and there is no reason to believe that our connections to expected utility theory are unique to the Cressie-Read family.

The CR divergence between the observed (τ_t) and tilted ($\hat{\tau}_t$) measure is defined by

$$CR_t(\hat{\tau}_t, \tau_t; \lambda) = \frac{2}{\lambda(1+\lambda)}\tau_t\left[\left(\frac{\tau_t}{\hat{\tau}_t}\right)^{\lambda} - 1\right],$$

for fixed scalar parameter $\lambda \in \Re$. We use the CR divergence because it generalizes many well-known divergence measures. For example, $\lambda = -2$ yields the Neyman-modified χ^2 divergence, $\lambda = 1$ gives Pearson's χ^2, and $\lambda = -1/2$ is the Freeman-Tukey measure. Two limiting distributions which are also encountered frequently are the empirical likelihood measure ($\lambda \to 0$) and the KL measure ($\lambda \to -1$).

Embedding the portfolio selection rule (*i.e.*, an optimization over $\mathbf{w}$) into the CR function culminates in definition 1.

Definition 1. *Let the relevant measure of disparity be governed by the CR power divergence. Then, for portfolio return $R_t(\mathbf{w})$, benchmark return d, Lagrange multipliers θ and ϕ, initial measure $u = u_t = 1/T$ for all t, and tilted weights $\hat{\tau}_1, ...\hat{\tau}_T$, the CR optimal portfolio is determined by*

$$\max_{\mathbf{w},\theta,\phi} \ \min_{\hat{\tau}} \sum_{t=1}^{T} \frac{u_t}{\lambda(\lambda+1)}\left[\left(\frac{\hat{\tau}_t}{u_t}\right)^{-\lambda} - 1\right] + \theta\left[\sum_{t=1}^{T}\hat{\tau}_t R_t(\mathbf{w}) - d\right] + \phi\left(\sum_{t=1}^{T}\hat{\tau}_t - 1\right)$$

subject to the usual wealth exhaustion constraint. This can be simplified by solving the interior minimization problem, which gives

$$\hat{\tau}_t(\mathbf{w},\theta,\phi) = \frac{u_t}{\big((\lambda+1)\{\theta[R_t(\mathbf{w}) - d] + \phi\}\big)^{\frac{1}{\lambda+1}}}.$$

Back-substituting gives

$$\max_{\mathbf{w},\theta,\phi} \ \sum_{t=1}^{T} \frac{u_t\big(\{\theta[R_t(\mathbf{w}) - d] + \phi\}(\lambda+1)\big)^{\frac{\lambda}{\lambda+1}}}{\lambda} - \phi - \frac{1}{\lambda(\lambda+1)}$$

subject to wealth exhaustion.

Proposition 1. *The CR portfolio has a unique solution.*

Proof. See appendix. □

The inner-workings of all members of the CR family of rules parallel the intuition of the KL-based rules: find the portfolio with the largest of the CR-minimum disparities. As before, the portfolio that requires the *most* re-weighting to achieve the tilted mean restriction is the portfolio that, intuitively, is least likely to deliver a return below the target rate d.

Our core contribution rests in exposing the relationship between our general CR-based family of minimum disparity portfolio selection rules to the widely used HARA utility family. To see the connection, rewrite the CR objective function from Definition 1 as

$$\psi(\theta, \phi) = \frac{1}{\lambda T}\sum_{t=1}^{T}[(\lambda+1)(\theta r_t + \phi)]^{\frac{\lambda}{\lambda+1}} - \phi$$

where $r_t \equiv R_t(\mathbf{w}) - d$. Now let $\beta \equiv \lambda/(\lambda+1)$, which implies that $1/\lambda = (1-\beta)/\beta$ and $(\lambda+1) = 1/(1-\beta)$. Also, let $\eta \equiv (\lambda+1)\phi$, so $\phi = \eta(1-\beta)$. Using these notational substitutions yields

$$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1-\beta}{\beta}\right)\left(\frac{\theta r_t}{1-\beta} + \eta\right)^{\beta} - \eta(1-\beta);$$

i.e., maximizing $\psi(\cdot)$ is equivalent to maximizing a time-averaged HARA utility function, and thus provides a generalization of (3).

The HARA parameter β is plainly pinned down by the choice for CR parameter λ. The other two HARA parameters, θ and η are pinned down by the CR parameters θ and ϕ. While λ must be set exogenously (a point which we discuss further below), θ and ϕ are decision variables within the CR objective function, and are thus endogenous in the same sense as in Stutzer (2000,2003). These values are likewise identified by the user's choice for d. The CR formulation exposes how the HARA parameters θ and η are actually comprised of the CR Lagrange multipliers on the twisted mean restriction, the constraint that the twisted weights must sum to one, and the choice of the tilting measure (encapsulated by λ).

The endogenous analog to HARA utility is "twice endogenous" in the sense that for various values of θ and ϕ that may be achieved by the optimization, CRRA or CARA may obtain. For example, if $\phi = 0$ arose endogenously, then η would also equal zero, and the HARA expression would reduce to the familiar CRRA utility model; this would be true for any values of θ and λ. Thus, the type of risk aversion arises endogenously, as do the risk aversion parameter values.

Regarding λ and the type of absolute risk aversion: the Arrow-Pratt measure of absolute risk aversion is

$$\frac{\theta}{(\lambda+1)(\theta r_t + \phi)},$$

the derivative of which is

$$-(\lambda+1)\left[\frac{\theta}{(\lambda+1)(\theta r_t + \phi)}\right]^2.$$

Hence we have increasing (constant) (decreasing) absolute risk aversion if λ is less than (equal to) (greater than) -1; table 1 summarizes some of these relationships. Note that the case where $\lambda = -1$ corresponds to the endogenous negative exponential utility model as derived in Stutzer (2000).

Table 1: Five CR Rules and their HARA Analogs

λ	Disparity Name	β	Utility Name	Type of Absolute Risk Aversion
1	Pearson's χ^2	0.5	—	decreasing
0	Empirical Likelihood	0	—	decreasing
-0.5	Freeman-Tukey	-1	—	decreasing
-1	Kullback-Leibler	$\beta \to \infty$	Exponential	constant
-2	Euclidean	2	Quadratic	increasing

3 Numerical Example

We now illustrate the CR portfolio formation approach with a stock portfolio selection problem [see Haley and McGee (2010) for a more thorough treatment of the numerical issues associated with related formulations of the CR objective function]. To facilitate comparison with existing work, we use the same monthly stock returns that appear in Haley and Whiteman (2008), themselves a subset of the returns used in Stutzer (2000). The data, summarized in Table 2, consist of monthly stock returns from 21 stocks from January 1977 through December 1996 (240 total observations).

A natural question is how to evaluate which member of the CR family is "best." The obvious ways include comparisons based on moments of the resulting optimal portfolios and the in-sample shortfall probability. However, there is an additional way to evaluate candidate members of the CR family in light of their divergence-based representations, which extends from the Relative Numerical Efficiency (RNE) value proposed by Geweke (1989). We provide only a brief overview of how RNE can be used to evaluate the reliability of the portfolio selection process. We direct the interested reader to Haley and Whiteman (2008), who develop this methodology in detail for the KL case.

Because the calculation of the CR portfolio involves tilting from one distribution to another, concepts from importance sampling can be used to gather additional insight into the portfolio ranking process. Like our case here, importance sampling involves evaluating the usefulness of various "importance densities" (or kernels, where applicable) in estimating moments from a different density, often call the "target" density. How well these target moments are estimated depends critically on the importance density. If the importance density is poor, the moment estimates will be likewise poor. Haley and Whiteman (2008), using Geweke (1989), adapt this intuition to the KL-based portfolio selection process, which we, in turn, adapt here to the CR family of portfolio selection rules. The result is a way to evaluate the reliability of the portfolio selection process. In this case, the "importance density" is the empirical distribution of portfolio returns induced by a given vector of asset weights $\mathbf{w}$. The "target density" here is the tilted distribution, meaned at d. The moment of interest, therefore, is the twisted mean. The RNE indicates whether or not the twisted mean is being estimated reliably in the portfolio selection process. This RNE value, generally appearing in the unit interval, should be as close to one as possible. Geweke (1989) states that the effective sample size being used to estimate the target moment, the twisted mean in our case, is equal to the original sample size T times the RNE value.

When using the CR family, the researcher must specify d and λ prior to conducting any analysis. Selecting d is often straightforward especially for the end-user of portfolio rules

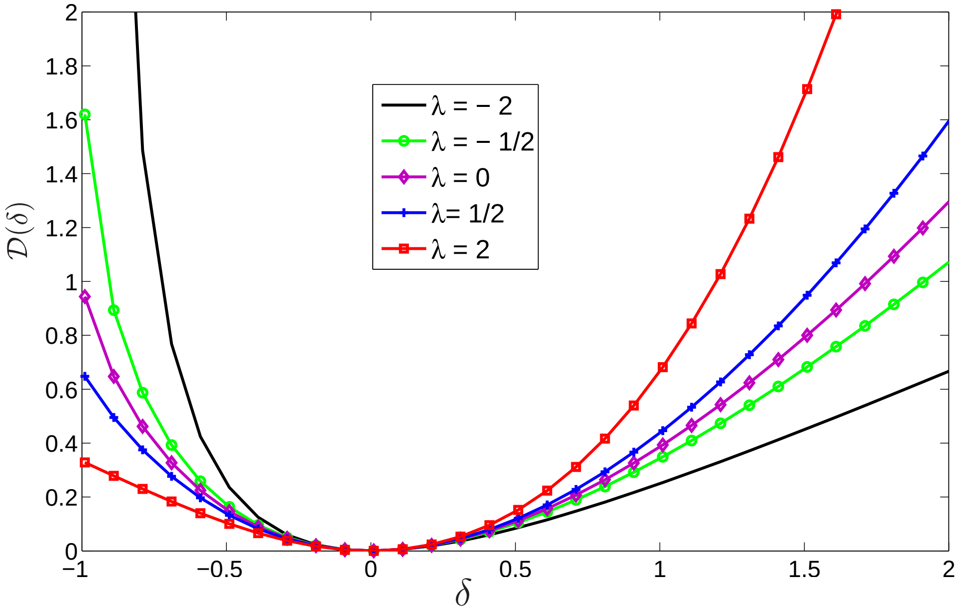

Figure 1: Disparity measures for various λ

(*e.g.*, financial planners, fund managers). Benchmark or target rates are readily available to investors, and have long been considered within modern portfolio theory. The Sharpe portfolio [see, for example, Sharpe (1994)] builds its excess return by subtracting out the risk-free rate; the term "differential return" or "information ratio" is used if something other than the risk-free rate serves as the benchmark. Mutual fund managers today face considerable criticism if they fail to meet or beat the sector-, cap-, or style-specific index that is most comparable to the stated objectives of their fund.

While the choice for d is typically straightforward, choosing a value for λ is more difficult. This partially amounts to specifying a specific HARA expected utility function (table 1), but is *less restrictive* than traditional HARA implementation because the remaining parameters are identified using the endogenous Lagrange multipliers. Intuition for how λ alters the portfolio choice can be seen by factoring the CR function according to Basu and Lindsay (1994):

$$\begin{aligned}\sum_{t=1}^{T} CR_t(\hat{\tau}_t, \tau_t; \lambda) &= \frac{2}{\lambda(\lambda+1)} \sum_{t=1}^{T} \left[\tau_t \left\{ \left(\frac{\tau_t}{\hat{\tau}_t} \right)^{\lambda} - 1 \right\} + \lambda(\hat{\tau}_t - \tau_t) \right] \\ &= 2 \sum_{t=1}^{T} \hat{\tau}_t \mathcal{D}(\delta_t; \lambda)\end{aligned}$$

where

$$\mathcal{D}(\delta; \lambda) = \frac{(\delta+1)^{\lambda+1} - (\delta+1)}{\lambda(\lambda+1)} - \frac{\delta}{\lambda+1}, \quad \delta_t = \left(\frac{\tau_t}{\hat{\tau}_t} - 1 \right).$$

Thus, the CR divergence may be interpreted as a weighted function ($\mathcal{D}$) of disparity measures (δ) between the actual and tilted probability measures. The function $\mathcal{D}(\cdot)$ is non-negative, defined on $[-1, \infty)$ and equals zero if and only if the disparity between the two measures is also zero (*i.e.*, $\tau_t = \hat{\tau}_t \ \forall \ t$). In figure 1 we plot the CR disparity measure for $\lambda = [-2, -0.5, 0, 0.5, 2]$.[2] The figure shows the asymmetric nature of the alternative measure changes. This is, of course, related to the different types of risk aversion documented in table 1.[3] As λ goes from negative to positive values, the risk aversion changes from increasing ($\lambda = -2$) to constant ($\lambda = -1$) to decreasing ($\lambda > -0.5$). Note that for positive (negative) values of δ, positive (negative) λ lead to higher values for $\mathcal{D}$. Thus CR divergence measures with positive (negative) λ restrict the degree to which the actual (tilted) probability can exceed the tilted (actual) probability. In other words, an investor with increasing risk aversion will penalize deviations from the actual probability distribution more so than an investor with decreasing risk aversion.

Table 2: Summary Statistics and Optimal Portfolios for $d = 0$

Ticker	Min	Max	Mean	Std.	Skew	EL	KL	EU
ACK	-0.294	0.33	0.014	0.087	0.087	-0.053	-0.127	-0.130
ALK	-0.323	0.519	0.015	0.105	0.557	-0.069	-0.039	-0.041
AMD	-0.477	0.487	0.024	0.163	0.261	-0.025	0	-0.01
APD	-0.306	0.215	0.013	0.076	-0.021	0.036	-0.02	-0.028
AT	-0.177	0.171	0.016	0.06	-0.216	-0.062	0.065	0.066
BAX	-0.162	0.203	0.011	0.073	-0.029	-0.028	-0.162	-0.1754
BK	-0.25	0.287	0.018	0.082	0.154	-0.004	0.037	0.043
BMY	-0.139	0.177	0.015	0.058	0.056	0.166	0.181	0.19
BUD	-0.155	0.258	0.017	0.063	0.324	0.182	0.192	0.197
CC	-0.437	0.415	0.035	0.126	0.115	0.126	0.111	0.11
CPL	-0.137	0.188	0.013	0.05	0.011	0.176	0.071	0.069
CRS	-0.309	0.273	0.013	0.078	-0.127	0.044	0.091	0.095
DCN	-0.256	0.333	0.012	0.08	0.169	-0.147	-0.101	-0.099
DG	-0.265	0.347	0.028	0.121	0.376	0.068	0.07	0.075
GD	-0.277	0.34	0.019	0.091	0.36	0.098	0.035	0.036
KMB	-0.146	0.322	0.017	0.062	0.866	0.114	0.223	0.243
KU	-0.107	0.139	0.012	0.04	-0.14	0.141	0.355	0.367
LTD	-0.431	0.409	0.026	0.129	0.072	-0.012	-0.056	-0.058
MUR	-0.26	0.274	0.015	0.09	0.055	0.142	0.042	0.045
NCC	-0.244	0.216	0.014	0.064	-0.161	0.054	0.041	0.039
ORU	-0.102	0.235	0.011	0.042	0.494	0.052	-0.008	-0.034
Mean						0.019	0.019	0.019
Std						0.04	0.037	0.035
Skew						-0.002	-0.204	-0.206
Pr$<d$						0.308	0.313	0.315
RNE						0.388	0.421	0.445

Table 2 presents the results of the portfolio formation for $d = 0$. We examine the Empirical Likelihood (EL), Kullback-Leibler (KL), and Euclidean (EU) divergence measures,

[2] The measure becomes $(1+\delta)\log(\delta+1) - \delta$ when $\lambda \to 0$.

[3] Attaching an expected utility interpretation to the shortfall problem has an alternative interpretation of "disciplined GMM" [see Maasoumi (1994)].

which correspond to decreasing, constant and increasing absolute risk aversion, respectively.[4] The portfolio return based on the EL divergence measure (decreasing risk aversion) generates the highest standard deviation, whereas the EU divergence measure (increasing risk aversion) generates the smallest. This is consistent with the intuition from expected utility theory: investors with increasing risk aversion prefer less volatility. The tradeoff for an increase in standard deviation is a more favorable skewness statistic. Both the KL and EU measures generate portfolio returns which carry a less favorable skewness statistic than the EL portfolio rule. The EL measure also yields the smallest probability of a shortfall return. As discussed above, the RNE calculations shed additional light on these results. The EL divergence carries with it the lowest RNE measure, while the EU divergence delivers the highest. Therefore, while the EL divergence may appear to be the "best" in terms of the shortfall criterion (Pr$< d$) and skewness, it carries with it a somewhat lower, and less desirable, RNE value.

4 Conclusion

We have proposed a new family of disparity-based shortfall minimizing portfolio selection rules, which we have related to the familiar HARA family of expected utility functions. In this capacity, our work extends the endogenous utility interpretation from the KL case found in Stutzer (2000,2003) to the entire CR family. This permits the HARA family to be interpreted as a minimum disparity estimation problem built on the hypothesis that investors seek to minimize the probability of realizing a return below some pre-determined target or benchmark rate. This application of disparity minimization forms an interesting bridge between the seemingly simplistic notion of shortfall minimization to the formally structured expected utility approach to decision under uncertainty.

[4]The Eucliean divergence-based portfolio has an interesting relationship to Roy's (1952) Safety First rule and the Sharpe portfolio in the no-shorting case; see Haley and McGee (2006) for details.

Appendix

Collected herein is a detailed proof of the primary proposition. We construct the proof as five subproofs, after first setting out a foundation and introducing a more compact notation.

Proof of Proposition 1

Recall the objective function:

$$\max_{\mathbf{w},\theta,\phi} \; \min_{\hat{\tau}} \sum_{t=1}^{T} \frac{u}{\lambda(\lambda+1)} \left[\left(\frac{\hat{\tau}_t}{u} \right)^{-\lambda} - 1 \right] + \theta \left(\sum_{t=1}^{T} \hat{\tau}_t r_t \right) + \phi \left(\sum_{t=1}^{T} \hat{\tau}_t - 1 \right) \tag{A1}$$

where $r_t \equiv R_t(\mathbf{w}) - d$ (*i.e.*, the return net of the target d) and where $u = u_t = 1/T \; \forall \; t$. For notational parsimony we have suppressed r_t's dependence on $\mathbf{w}$.

Taking first-order conditions of the interior minimization gives

$$\mathcal{L}_{\hat{\tau}_t} = -\frac{1}{\lambda+1} \left(\frac{\hat{\tau}_t}{u} \right)^{-(\lambda+1)} + \theta r_t + \phi = 0,$$

which implies that

$$\hat{\tau}_t(\mathbf{w},\theta,\phi) = \frac{u}{[(\lambda+1)(\theta r_t + \phi)]^{\frac{1}{\lambda+1}}}. \tag{A2}$$

Back-substituting gives

$$\max_{\mathbf{w},\theta,\phi} \sum_{t=1}^{T} \frac{u[(\theta r_t + \phi)(\lambda+1)]^{\frac{\lambda}{\lambda+1}}}{\lambda} - \phi - \frac{1}{\lambda(\lambda+1)}$$

subject to wealth exhaustion or, equivalently,

$$\max_{\mathbf{w},\theta,\phi} \sum_{t=1}^{T} \frac{u[(\theta r_t + \phi)(\lambda+1)]^{\frac{\lambda}{\lambda+1}}}{\lambda} - \phi - \nu \left(\sum_{j=1}^{J} w_j - 1 \right)$$

where ν is the Lagrange multiplier for the (explicit) wealth exhaustion constraint.[5]

Because $\theta r_t = \theta \sum_j w_j r_{jt}$ we can define $\gamma_j \equiv \theta w_j$ such that $\theta r_t = \sum_j \gamma_j r_{jt}$ and $\sum_j \gamma_j = \theta$, which means that the wealth exhaustion constraint can be folded into the objective function, thus reducing the maximization problem to

$$\max_{\phi,\gamma} \; \psi(\phi,\gamma) \equiv \sum_{t=1}^{T} \frac{u[(\sum_j \gamma_j r_{jt} + \phi)(\lambda+1)]^{\frac{\lambda}{\lambda+1}}}{\lambda} - \phi. \tag{A3}$$

The first-orders for the resulting maximization problem are

$$\psi_\phi = \sum_{t=1}^{T} u[(\lambda+1)(\theta r_t + \phi)]^{\frac{-1}{\lambda+1}} - 1 = 0 \tag{A4a}$$

$$\psi_{\gamma_j} = \sum_{t=1}^{T} u r_{jt}[(\lambda+1)(\theta r_t + \phi)]^{\frac{-1}{\lambda+1}} = 0 \tag{A4b}$$

Equation (A4a) implies that $\sum_t \hat{\tau}_t = 1$, while (A4b) states that $\sum_t \hat{\tau}_t r_{jt} = 0$; *i.e.*, the tilting weights will be chosen so that the twisted distributions of *each* of the assets will be meaned at d.

To assess the second-order conditions of (A3) let

$$m_t \equiv -u[(\lambda+1)(\theta r_t + \phi)]^{\frac{-(\lambda+2)}{\lambda+1}},$$

and write the Hessian as

$$\begin{bmatrix} \sum_t m_t & \sum_t m_t r_{1t} & \sum m_t r_{2t} & \dots & \sum_t m_t r_{Jt} \\ \sum_t m_t r_{1t} & \sum_t m_t r_{1t}^2 & \sum m_t r_{1t} r_{2t} & \dots & \sum_t m_t r_{1t} r_{Jt} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \sum_t m_t r_{Jt} & \sum_t m_t r_{1t} r_{Jt} & \sum m_t r_{2t} r_{Jt} & \dots & \sum_t m_t r_{Jt}^2 \end{bmatrix}.$$

[5] Assuming an interior solution, the second-order condition for the interior minimization produces a diagonal Hessian with positive elements, indicating it to be positive definite.

The Hessian principal minors are all weighted sums of squares:

$$H_1 = \sum_{t=1}^{T} m_t$$
$$H_2 = \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} m_t m_s (r_{1t} - r_{1s})^2$$
$$H_3 = \sum_{t=1}^{T-2} \sum_{s=t+1}^{T-1} \sum_{q=s+1}^{T} m_t m_s m_q [r_{2t}(r_{1s} - r_{1q}) + r_{2s}(r_{1q} - r_{1t}) + r_{2q}(r_{1t} - r_{1s})]^2$$

and so forth. If all the m_ts are negative then the Hessian is negative definite. However, from the definition of m_t, m_t is negative if $(\lambda+1)(\theta r_t + \phi)$ is positive, which is also a necessary condition for $\hat{\tau}_t$ to be positive. Since at least some of the r_{jt} must be negative for the portfolio decision to be nontrivial, we must have for

$$\begin{aligned} \lambda > -1, \qquad (\theta r_t + \phi) > 0 \\ \lambda < -1, \qquad (\theta r_t + \phi) < 0 \end{aligned} \tag{A5}$$

If (A5) is met, any solution we have will be unique. We verify these using the following three cases.[6]

[i] The "upper" case: $\lambda > 0$, which implies that $0 < \frac{\lambda}{\lambda+1} < 1$. To explore the impact of choosing the weight γ_j for any asset j, suppose that ϕ and the other $\gamma_{k \neq j}$s have been chosen, and denote $\phi_t^* = (\phi + \sum_{k \neq j} \gamma_k r_{jt})$ for each time period t. Also denote as r_{jmax} and r_{jmin} the rates of return for asset j farthest above and below their respective ϕ_t^*. From (A3), $\psi(\phi, \gamma)$ will be real valued if all the $(\sum_j \gamma_j r_{jt} + \phi)$ are nonnegative, which in turn requires

$$\frac{-\phi_t^*}{r_{jmax}} \leq \gamma_j \leq \frac{-\phi_s^*}{r_{jmin}}. \tag{A6}$$

If (A6) is satisfied for all J assets, then $(\theta r_t + \phi)$ will be positive, which by (A5) assures concavity. Within the range (A6), all of the $(\gamma_j r_{jt} + \phi_t^*)$ are either monotonically increasing or decreasing concave functions of γ_j, depending on the sign of r_{jt}. At the two boundaries $(\gamma_j r_{jmin} + \phi_t^*)/\lambda$ or $(\gamma_j r_{jmax} + \phi_s^*)/\lambda$ will equal zero. Since presumably the majority of the r_{jt}s are positive, we might expect $\psi(\phi, \gamma)$ to be increasing over most of the range. However, since from (A4b) the slope of $\psi(\phi, \gamma)$ approaches positive and negative infinity at the lower and upper limits of the range, we must have an interior maximum.

□

[ii] The "central" case: $-1 < \lambda < 0$, which implies that $\frac{\lambda}{\lambda+1} < 0$. Using the same notation as the previous case, note that since the exponent in (A3) will be negative, $\psi(\phi, \gamma)$ will be discontinuous at every value of $\gamma_j = -\phi_t^*/r_{jt}$. The only range for which all the values of $(\gamma_j r_{jt} + \phi_t^*)$ will be positive will be between the points of discontinuity for r_{jmin} and r_{jmax}. Then a sufficient condition for all the $\hat{\tau}_t$s to be positive and, by (A5), for all $\psi(\cdot)$ to be concave is

$$\frac{-\phi_t^*}{r_{jmax}} < \gamma_j < \frac{-\phi_s^*}{r_{jmin}} \tag{A7}$$

for each asset.

As γ_j approaches either boundary, either $(\gamma_j r_{jmin} + \phi_t^*)/\lambda$ or $(\gamma_j r_{jmax} + \phi_s^*)/\lambda$ will approach negative infinity. Otherwise the $(\gamma_j r_{jt} + \phi_t^*)$ are monotonically increasing or decreasing functions of γ_j, depending on the sign of r_{jt}, and the sum $\psi(\phi, \gamma)$ will have a unique interior maximum.

□

[iii] The "lower" case: $\lambda < -1$, which implies that $\frac{\lambda}{\lambda+1} > 1$. Again using the same notation, since $(\lambda+1)$ is negative, our sufficient condition for concavity requires $(\theta r_t + \phi) \leq 0$, or each $(\gamma_j r_{jt} + \phi_t^*) \leq 0$, giving the range for γ_j:

$$\frac{-\phi_t^*}{r_{jmin}} \leq \gamma_j \leq \frac{-\phi_s^*}{r_{jmax}}.$$

At the boundaries, ψ_{γ_j} in (A4b) is zero for r_{jmin} or r_{jmax}; since its slope for the other r_{jt} will be positive or negative depending on the sign of r_{jt}, and the magnitudes of those slopes depend on the particular r_{jt} values, there is no guarantee that the summed slope will be positive at the lower boundary or negative at the upper boundary. In short, we cannot be sure an interior maximum exists. However, for values of λ that satisfy $\lambda = -2k/(2k+1)$, so $\lambda/(\lambda+1) = -2k$, where k is any positive integer, then $[(\lambda+1)(\gamma_j r_{jt} + \phi_t^*)]^{-2k}/\lambda$ is everywhere concave, and hence $\psi(\phi, \gamma)$ is everywhere concave, and a unique maximum exists.

[6] For completeness, we include cases four and five, though they have already been discovered elsewhere, as noted.

$\square$

[iv] Case four (Kullback-Leibler): $\lambda = -1$; see Haley and Whiteman (2008) for additional discussion. The following results establish the concavity of GSF in θ and $\mathbf{w}$. The first result (regarding θ) is standard in the theory of convex conjugates (see, for example, Rockafellar, 1970), while the second (regarding $\mathbf{w}$) is specific to portfolio analysis. We include these proofs for completeness.

For portfolio return $R(\mathbf{w})$, portfolio weights $\mathbf{w}$, target rate d, and parameter $\theta \in \Re^-$, GSF is concave in θ and $\mathbf{w}$. To see this, let $\lambda(\theta, \mathbf{w}) \equiv \log\big(E\{\exp[\theta R(\mathbf{w})]\}\big)$. For $\alpha, \beta \in [0,1]$ such that $\alpha + \beta = 1$ and $\theta_1, \theta_2 \in \Re^-$,

$$E\{\exp[(\alpha\theta_1 + \beta\theta_2)R(\mathbf{w})]\} = E\{\exp[\alpha R(\mathbf{w})\theta_1]\exp[\beta R(\mathbf{w})\theta_2]\},$$

then Hölder's inequality implies that

$$E\{\exp[\alpha R(\mathbf{w})\theta_1]\exp[\beta R(\mathbf{w})\theta_2]\} \le \big(E\{\exp[R(\mathbf{w})\theta_1]\}\big)^{\alpha}\big(E\{\exp[R(\mathbf{w})\theta_2]\}\big)^{\beta}.$$

The monotonicity of $\log(\cdot)$ then implies that

$$\lambda(\alpha\theta_1 + \beta\theta_2, \mathbf{w}) \le \alpha\lambda(\theta_1, \mathbf{w}) + \beta\lambda(\theta_2, \mathbf{w}),$$

which immediately implies that $\lambda(\theta, \mathbf{w})$ is convex in θ and that $[d\theta - \lambda(\theta, \mathbf{w})]$ is concave in θ.

Regarding $\mathbf{w}$, for $\alpha, \beta \in [0,1]$ such that $\alpha + \beta = 1$ and $\mathbf{w}_a, \mathbf{w}_b \in \Re^N$,

$$E\{\exp[\theta R(\alpha\mathbf{w}_a + \beta\mathbf{w}_b)]\} = E\{\exp[\alpha\theta R(\mathbf{w}_a)]\exp[\beta\theta R(\mathbf{w}_b)]\},$$

which uses the fact that $R(\alpha\mathbf{w}_a + \beta\mathbf{w}_b)$ is equal to $[\alpha R(\mathbf{w}_a) + \beta R(\mathbf{w}_b)]$. Then Hölder's inequality implies that

$$E\{\exp[\alpha\theta R(\mathbf{w}_a)]\exp[\beta\theta R(\mathbf{w}_b)]\} \le \big\{E[\exp[\theta R(\mathbf{w}_a)]\big\}^{\alpha}\big\{E[\exp[\theta R(\mathbf{w}_b)]\big\}^{\beta}.$$

The monotonicity of $\log(\cdot)$ then implies that

$$\lambda(\theta, \alpha\mathbf{w}_a + \beta\mathbf{w}_b) \le \alpha\lambda(\theta, \mathbf{w}_a) + \beta\lambda(\theta, \mathbf{w}_b),$$

which immediately implies that $\lambda(\theta, \mathbf{w})$ is convex in $\mathbf{w}$ and that $[d\theta - \lambda(\theta, \mathbf{w})]$ is concave in $\mathbf{w}$.

$\square$

[v] Case five (Empirical Likelihood): $\lambda = 0$. This proof appears in Haley and McGee (2009), but we include it here for completeness. Under EL divergence the twisting loss function that needs to be minimized is

$$\sum_{t=1}^{T} u\log(u/\rho_t).$$

Therefore, we seek the portfolio weights w_i and the twisted probabilities ρ_t that solve:

$$\max_{\theta, w_i}\left[\min_{\rho_t,\theta,\phi}\sum_{t=1}^{T} u\log(u/\rho_t) + \theta\sum_{t=1}^{T}\rho_t r_t + \phi\Big(\sum_{t=1}^{T}\rho_t - 1\Big)\right] \tag{A8}$$

subject to wealth exhaustion. Beginning with the interior minimization problem, the first-order conditions are

$$\begin{aligned}
\mathcal{L}_{\rho_t} &= -(u/\rho_t) + \theta r_t + \phi = 0,\\
\mathcal{L}_{\theta} &= \sum_{t=1}^{T}\rho_t r_t = 0,\\
\mathcal{L}_{\phi} &= \sum_{t=1}^{T}\rho_t - 1 = 0.
\end{aligned}$$

The conditions with respect to the ρ_t can be rewritten $\theta\rho_t r_t + \phi\rho_t = u$. Summing over the T periods gives

$$\theta\sum_{t=1}^{T}\rho_t r_t + \phi\sum_{t=1}^{T}\rho_t = \sum_{t=1}^{T} u = 1. \tag{A9}$$

The other two first-order conditions imply $\sum_{t=1}^{T}\rho_t r_t = 0$ and $\sum_{t=1}^{T}\rho_t = 1$, so (A9) reduces to $\phi = 1$, giving

$$\rho_t = u/(1 + \theta r_t). \tag{A10}$$

The second-order conditions for the interior minimization problem produce the bordered Hessian

$$|\bar{H}| = \begin{vmatrix} 0 & 0 & 1 & ... & 1 \\ 0 & 0 & r_1 & ... & r_T \\ 1 & r_1 & u/\rho_1^2 & ... & 0 \\ \vdots & \vdots & \dots & \ddots & \vdots \\ 1 & r_T & 0 & \dots & u/\rho_T^2 \end{vmatrix} = \sum_{t=1}^{T} \sum_{s=t+1}^{T} (r_t - r_s)^2 \prod_{n \neq s,t} \frac{u}{\rho_n^2},$$

which is strictly positive for all the principal minors, ensuring a unique minimum.

Substituting (A10) for ρ_t into (A8), the outer maximization problem becomes

$$\max_{\theta, w_i} \sum_{t=1}^{T} u \log \left(1 + \theta r_t\right) \tag{A11}$$

The derivative with respect to θ again produces $E(r_t) = 0$, or $E(R_t) = d$. To characterize the solving of θ, note first from (A10) that for all the ρ_t to be positive, we must have all $(1 + \theta r_t) > 0$. This condition is also required by (A11), for the logarithm to be defined. Denoting r_{min} and r_{max} as the most extreme values of r_t (for any given set of weights w_i), where $r_{min} < 0 < r_{max}$, then $(1 + \theta r_t) > 0$ for all t requires

$$-\frac{1}{r_{max}} < \theta < -\frac{1}{r_{min}},$$

that is, θ must fall in an interval bounded around zero.[7]

To further limit the value of θ, let $g(\theta)$ be the partial derivative of (A11) with respect to θ:

$$g(\theta) = \sum_{t=1}^{T} \frac{r_t}{1 + \theta r_t}$$

with the solution for θ at $g(\theta) = 0$. Clearly, $g'(\theta) = -\sum_{t=1}^{T} r_t^2/(1 + \theta r_t)^2$ is negative. Since $g(0) = \sum_{t=1}^{T} r_t = (\mu - d) > 0$, θ has a unique positive solution:

$$0 < \theta < -\frac{1}{r_{min}}.$$

Rather than minimize (A11) with respect to the w_i directly, note that $\theta r_t = \sum_{t=1}^{T} \theta w_i r_{it} \equiv \sum_{t=1}^{T} \gamma_i r_{it}$, where $\gamma_i \equiv \theta w_i$ and $\sum_{t=1}^{T} \gamma_i = \theta$. Replacing the $N+1$ choice variables θ and the w_i in (A11) with the N variables γ_i makes the wealth exhaustion constraint redundant, so (A11) reduces to

$$\max_{\gamma_i} = \sum_{t=1}^{T} u \log \Big(1 + \sum_{n=1}^{N} \gamma_i r_{it}\Big).$$

The first-order conditions are

$$\mathcal{L}_{\gamma_i} = \sum_{t=1}^{T} \frac{u r_{it}}{1 + \theta r_t} = 0,$$

which again requires that under the twisted probabilities, the mean return for each of the N assets equals d. The Hessian of the second partial derivatives is

$$|H| = \begin{vmatrix} -\sum_{t=1}^{T} \frac{u r_{1t}^2}{(1+\theta r_t)^2} & -\sum_{t=1}^{T} \frac{u r_{1t} r_{2t}}{(1+\theta r_t)^2} & ... & -\sum_{t=1}^{T} \frac{u r_{1t} r_{Nt}}{(1+\theta r_t)^2} \\ -\sum_{t=1}^{T} \frac{u r_{1t} r_{2t}}{(1+\theta r_t)^2} & -\sum_{t=1}^{T} \frac{u r_{2t}^2}{(1+\theta r_t)^2} & ... & -\sum_{t=1}^{T} \frac{u r_{2t} r_{Nt}}{(1+\theta r_t)^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\sum_{t=1}^{T} \frac{u r_{1t} r_{Nt}}{(1+\theta r_t)^2} & -\sum_{t=1}^{T} \frac{u r_{2t} r_{Nt}}{(1+\theta r_t)^2} & ... & -\sum_{t=1}^{T} \frac{u r_{Nt}^2}{(1+\theta r_t)^2} \end{vmatrix}.$$

The Hessian's principal minors are weighted sums of squares, with weights of alternating sign. Thus the Hessian is negative definite and the solution is a unique maximum.

□

[7] This fact is useful when implementing this rule using numerical optimization techniques.

References

[1] Almeida, C. and R. Garcia. 2008. "Empirical Likelihood Estimators for Stochastic Discount Factors," EDHEC: working paper.

[2] Almeida, C. and R. Garcia. 2009. "Assessing Misspecifications in Asset Pricing Models with Nonlinear Projections of Pricing Kernels," EDHEC: working paper.

[3] Baggerly, K. 1998. "Empirical Likelihood as a Goodness-of-Fit Measure," *Biometrika* 85: 535.

[4] Basu, A. and C. Lindsay. 1994. "Minimum Disparity Estimation for Continuous Models: Efficiency, Distributions and Robustness," *Annals of the Institute of Statistical Mathematics*, 46(4), 683705.

[5] Bellini, F. and M. Frittelli. 2002. "On the Existence of Minimax Martingale Measures," *Mathematical Finance* 12: 1–21.

[6] Bucklew, J. 1990. *Large Deviations Techniques in Decision, Simulation, and Estimation*, New York: Wiley.

[7] Cressie, N. and T. Read. 1984. "Multinomial Goodness-of-Fit Tests," *Journal of the Royal Statistical Assocation. Series B (Methodological)* 46: 440–464.

[8] Föllmer, H. 2008. "Financial Uncertainty, Risk Measures, and Robust Preferences," In: "Aspects of Mathematical Finance" (Ed. M. Yor), 3-15, Springer: Berlin.

[9] Geweke, J. 1989. "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica* 57: 1317–1339.

[10] Granger, C., Maasoumi, E., and J. Racine. 2004. "A Dependence Metric for Possible Nonlinear Processes," *Journal of Time Series Analysis* 25: 649–669.

[11] Goll, T. and L. Rüschendorf. 2002. "Minimax and Minimal Distance Martingale Measures and Their Relationship to Portfolio Optimization," *Finance and Stochastics* 5: 557–581.

[12] Haley, M. 2003. *On the Allocation of Crops and Stocks*, PhD Dissertation, University of Iowa.

[13] Haley, M. and M. McGee. 2006. "Tilting Safety First and the Sharpe Portfolio," *Finance Research Letters*, 3: 173–180.

[14] Haley, M. and C. Whiteman. 2008. "Generalized Safety First and a New Twist on Portfolio Performance," *Econometric Reviews* 27: 457–483.

[15] Haley, M. and M. McGee. 2010. "KLICing There and Back Again: Portfolio Selection Using the Empirical Likelihood Divergence and Hellinger Distance," *Journal of Empirical Finance*, (forthcoming).

[16] Kramkov, D. and W. Schachermayer. 1999. "The Asymptotic Elasticity of Utility Functions and Optimal Investment in Incomplete Markets," *Annals of Applied Probability* 9: 904–950.

[17] Maasoumi, E. 1986. "The Measurement and Decomposition of Multi-Dimensional Inequality," *Econometrica* 54: 991–998.

[18] Maasoumi, E. 1994. "A Compendium to Information Theory in Economics and Econometrics," *Econometric Reviews* 12: 137–181.

[19] Rockafellar, R. 1970. *Convex Analysys*, Princeton: Princeton University Press.

[20] Roy, A. 1952. "Safety First and the Holding of Assets," *Econometrica* 20: 431–449.

[21] Sharpe, W. 1994. "The Sharpe Ratio," *Journal of Portfolio Management* Fall: 49–58.

[22] Stutzer, M. 2000. "A Portfolio Performance Index," *Financial Analyst Journal* May/June: 52–61.

[23] Stutzer, M. 2003. "Portfolio Choice With Endogenous Utility: A Large Deviations Approach," *Journal of Econometrics* 116: 365–386.