Robust Portfolio Selection with generalized Preferences: A methodology for Private Banking

Michel Verlaine,
GRID UMR 8534 CNRS /ENSAM
E-mail: verlaine@grid.ensam.estp.fr
michel.verlaine@eco.etat.lu;

Abstract

This paper is a first attempt to develop a methodology, consistent with non-linear probability weighting, to construct portfolios for Private Banking customers. Empirical evidence suggests that decision makers transform probability kernels in a non-linear way (Kahneman and Tversky (1992), Prelec (1998)). This has led to the concept of probability weighting function. Standard finance theory, and notably portfolio theory, has not yet dealt with such behavior. Part of the reason is that such behavior is considered non rational. It is first shown how to estimate the preferences from individual costumers data. The paper, then shows that probability weighting may be very rational. Indeed, it may be derived from the maximization of a measure of uncertainty (entropy) given observed data. It is then shown how a dynamic continuous time portfolio choice problem can be easily solved. It is assumed that probability weighting stems from robust behavior in face of parametric uncertainty. With this interpretation, risk behavior determines the functional form of the utility function, whereas probability weighting stems from an information treatment recipe. In fact, our decision maker chooses parameters that maximize a maximum entropy likelihood function. The decision maker internalizes the fact that the observed likelihood is one of many possible. Finally, the *subjectively* robust parameters can then be used to derive the optimal policy rule of a dynamic continuous time portfolio model.

Keywords: Robustness, Portfolio Selection, Behavior under risk and uncertainty.

JEL classification: G11, D81

1 Introduction

The standard asset allocation literature presumes that individuals maximize utility of terminal wealth given probabilities. The latter are of two kinds: an *objective* probability distribution as estimated by an econometrician or a subjective probability distribution that is deduced from rationality axioms (Savage (1972)). This approach is at variance with empirical evidence as documented f.i. by Kahneman and Tversky (1979) and (1992). The representation most consistent with the data seems to be a a utility function defined on gains (or losses) compared to a subjective benchmark and a non-additive measure (capacity). It would hence seem logical to focus on asset allocation with this kind of preferences.

This has been questioned by Campbell and Viciera (2002) on the grounds that this representation should not be used in a normative way. In fact, standard Bayesian rationality is inconsistent with a non-additive measure. In that sense individuals are irrational and the theorists should help them take rational decisions. Our view is different for two reasons. Firstly, standard rationality axioms have been questioned by decision theorists. Moreover, those authors advocate Choquet capacities as a subjective way to represent uncertainty (see f.i. Schmeidler (1989)).Choquet capacities can be axiomatized in case of pessimism or optimism under uncertainty (see inter alia Zhang (2002)). Secondly, financial decisions under uncertainty are generally implemented by using statistical distributions of variables of interest. These distributions are considered as "objective", in line with the classical statistics paradigm. However, as has been forcefully argued by Jaynes (1984), this is based on imaginary sample spaces. The least ad hoc distribution being the one that maximizes uncertainty, notably Shannon-entropy, given the data set (see f.i. Jaynes (1988) and Jaynes (1982)). As argued by Verlaine (2003) a decision maker using the maximum entropy method will transform the sample probability kernel in a non-linear way consistent with the function suggested by Prelec (1998).

Over the last years Andersen et al. (2003), Hansen et al. (2004) as well as Maenhoudt (2002) have addressed robustness issues in intertemporal consumption-portfolio models. All these papers consider a benchmark stochastic process that is contaminated in the sense that the decision maker (DM) considers the set of all other locally absolutely continuous processes as possible. In fact, the latter are difficult to detect from finite data sets. The decision maker takes a robust decision in the sense that he chooses the policy function that maximizes the value function of the less favorable trajectory. Nature is malevolent. The set of perturbations is determined by a penalty (penalty control problem) or a constraint (constraint control problem) on the discounted relative

entropy. It can be shown that both methods lead to the same optimal decisions.

In this paper, robustness is modelled via an inverse s-shaped transformation of the data. This is consistent with empirical evidence as documented inter alia by Kahneman and Tversky (1992) and Prelec (1998). We assume that probability weighting stems from robust behavior in face of parametric uncertainty. In this interpretation risk behavior determines the functional form, whereas probability weighting stems from an information treatment recipe. We argue that violation of expected utility axioms, notably the independence axiom, implies violation of the axioms underpinning the likelihood principle. Our DM is robust in the sense that he maximizes a measure of uncertainty (entropy) under the constraint of observed moments, notably a function of a measure of information. The decision maker chooses parameters that maximize a maximum entropy likelihood function. The decision maker internalizes the fact that the observed likelihood is one of many possible. Finally, for most of the plausible utility functions, the portfolio can be optimized by optimizing the standard Hamilton-Jacoby-Bellman process with the robust parameters.

2 Robustness in the existing literature

In this section, we briefly explain how the existing literature (Andersen et al. (2003), Hansen et al. (2004) and Meanhoudt (2002)) deals with the robustness issue. It is assumed that there is a benchmark control problem:

$$J(x_0) = \max_{c \in C} E\left[\int_0^\infty \exp\left(-\delta t\right) U\left(c_t, x_t\right) dt \right] \tag{1}$$

s.t. the following state process

$$dx_t = \mu\left(c_t, x_t\right) dt + \sigma\left(c_t, x_t\right) dB_t$$

where x_0 is a given initial condition and B is a standard Brownian motion.

A set of perturbations $q \in Q$ to the measure q^0 over continuous functions of time induced by the Brownian motion B is defined. The following discounted relative entropy

$$R(q) = \delta \int_0^\infty \exp\left(-\delta t\right) \left(\int \log\left(\frac{dq_t}{dq_t^0}\right) dq_t \right) dt \tag{2}$$

is used to restrict the set of perturbations q to q^0. If we leave the dependence on B implicit and define the utility process $v_t\left(c\right) = U\left(c_t, x_t\right)$ the control problem can be modelled in the following two ways.

Penalty control problem

$$V(\theta) = \max_{c \in C} \min_{q \in Q} \int_0^\infty \exp(-\delta t)\left(\int v_t(c)dq_t\right)dt + \theta R(q) \qquad (3)$$

Constraint control problem

$$K(\eta) = \max_{c \in C} \min_{q \in Q(\eta)} \int_0^\infty \exp(-\delta t)\left(\int v_t(c)dq_t\right)dt \qquad (4)$$

where $Q(\eta) = \{q \in Q : R(q) \leq \eta\}$.

Actually, for each optimal solution to $K(\eta)$ we can find a θ such that the optimal solution is the same. The set $Q(\eta)$ is viewed as set of multiple priors that are locally absolutely continuous with q^0 and hence difficult to detect with finite data sets. The parameter θ is the preference for robustness. The model is then solved by adding that parameter. However, the DM opts for max-min behavior which is inconsistent with empirical evidence. We would like to model this problem with inverse s-shaped behavior as documented by empirical evidence. The next two sections introduce the kind of preferences used in this paper.

3 Capacities and Choquet Expected Utility

As already pointed out in the introduction, individuals do not accord with expected utility. Kahneman and Tversky's (1979) and (1992) studies show that individuals first rank outcomes and then apply a non-linear PWF. However, their model is just a subcase of the general Choquet Expected Utility (CEU) model which we summarize below.

Let Ω be a finite set of states of nature. Let X be a set of monetary outcomes that includes the neutral outcome 0, all other elements are interpreted as gains or losses. An uncertain prospect f (act) is a function from Ω into X. To define the cumulative functional, outcomes of each prospect are arranged in increasing order. The utility that is expected by the individual depends on the ranking of the outcomes and the function is called (CEU). A prospect f is represented as a sequence of pairs (x_i, A_i), which yields x_i if A_i occurs, where $x_i > x_j$ iff $i > j$, and (A_i) are the events of an algebra of Ω. Now, let's use positive subscripts to denote positive outcomes and negative subscripts to denote negative outcomes. The positive part of f, denoted f^+, is obtained by letting $f^+(s) = f(s)$ if $f(s) > 0$, and $f^+(s) = 0$ if $f(s) \leq 0$. As in expected utility theory, we assign to each prospect a value $V(f)$ s.t. $V(f) \geq V(g)$ iff $f \geq g$. Now, we call a capacity, a nonadditive set function that generalizes the standard notion of probability. More exactly, a capacity v is a function that assigns to each $A \subset \Omega$ a number $v(A)$ satisfying $v(\phi) = 0, v(\Omega) = 1$, and

4

$v(A) \geq v(B)$ whenever $B \subset A$. CPT asserts that there exists a strictly increasing utility function $u : X \to R$, satisfying $u(x_o) = u(0) = 0$, and capacities v^+ and v^-, such that for $f = (x_i, A_i)$, $-m \leq i < n$,

$$V(f) = V(f^+) + V(f^-) \tag{5}$$

$$V(f^+) = \sum_{i=0}^{n} \pi_i^+ u(x_i), V(f^-) = \sum_{i=-m}^{0} \pi_i^- u(x_i), \tag{6}$$

where the decision weights are defined by :

$$\pi_n^+ = v^+(A_n), \qquad \pi_{-m}^- = v^-(A_{-m}), \tag{7}$$

$$\pi_i^+ = v^+(A_i \cup ... \cup A_n) - v^+(A_{i+1} \cup ... \cup A_n), \qquad 0 \leq i \leq n-1 \tag{8}$$

$$\pi_i^- = v^-(A_{-m} \cup ... \cup A_i) - v^-(A_{-m} \cup ... \cup A_{i-1}), \qquad 1-m \leq i \leq 0. \tag{9}$$

Letting $\pi_i = \pi_i^+$ if $i \geq 0$ and $\pi_i = \pi_i^-$ if $i < 0$, $V(f)$ reduces to:

$$V(f) = \sum_{i=-m}^{n} \pi_i u(x_i). \tag{10}$$

The decision weight associated with an outcome can be interpreted as the marginal contribution of the respective event, defined in terms of capacities. The measure of an event thus depends on it's rank and it is a generalization of expected utility called (CEU). An individual maximizing (13) is called a Choquet Expected Utility (CEU) maximizer. Under risk, where an objective cumulative distribution function exists, it may be interpreted as maximizing expected utility with respect to the transformed cumulative distribution function $v \circ F$ where F is the objective cumulative distribution function. This model is called Rank Dependent Expected Utility (RDEU). As we argue below, such an objective distribution never exists. WE consider a situation Gayant (2001) calls "imprecise risk", a situation between risk and uncertainty, where probabilities are fuzzy. v is the PWF and it's shape seems to be rather robust across experiments (see Tversky and Wakker (1995) and Prelec (1998)). Moreover, Prelec (1998) derives the functional consistent with axioms.

4 Observed behavior of probability weighting

In the standard CEU model, $u(x_0)$ is considered to be the minimal utility that the individual will get. He then adds the increments in utility weighted by his beliefs on the future states. $u(x_0)$ is not necessarily the minimal utility, but may be the reference point with respect to which the individual weights gains and losses. Moreover, the model may exhibit slightly different since the capacities are different on gains and losses. Actually, CEU is a special case where the capacity for losses is the *dual* of the capacity for gains, i.e. $v^-(A) = 1 - v^+(\Omega - A)$. Tversky and Kahneman (1992) also considered a special case of CPT where $v^+ = v^-$. This property is based on preference conditions called *reflection*. Now, we assume a kind of continuity of the capacity, namely that the latter satisfies *solvability* (see Gilboa (1987)).

Definition: The capacity v satisfies *solvability* if $\forall A \subset C$ and $v(A) \leq p \leq v(C)$ $\exists B$ s.t. $v(B) = p$ and $A \subset B \subset C$.

Notice that there are two possibilities to define risk (or uncertainty) aversion: via the curvature of the utility function or via the functional form of the capacities. It is generally accepted that risk behavior is given by the functional form of the utility function, whereas the uncertainty (or ambiguity) behavior is given by the functional form of the capacities.

Now, let's rewrite $V(f^+)$ in a different way:

$$V(f^+) = u(x_0) + \sum_{i=1}^{n} [u(x_i) - u(x_{i-1})] \, v(\cup_{j=i}^{n} A_j) \qquad (11)$$

It has been suggested that uncertainty aversion can be analyzed with a CEU model. In fact, Schmeidler (1989) gives the following definition of uncertainty aversion. Uncertainty aversion is equivalent to the convexity of the capacity, that is:

$$v(A \cup B) + v(A \cap B) \geq v(A) + v(B) \qquad (12)$$

for all measurable events A and B. We have additivity when the right hand side and the left hand side are equal, and the individual is then uncertainty neutral. Now, experimental evidence (Tversky and Wakker (1995), Prelec (1998)) shows that the capacity v is sub-additive meaning that there exist events E and $E\prime$ s.t.

$$v(B) + v(A) \geq v(A \cup B) \; whenever \; v(A \cup B) \leq v(S - E) \qquad (13)$$

And

$$1 - v(S - B) \geq v(A \cup B) - v(A) \ whenever \ v(A) \geq v(E') \qquad (14)$$

The first condition is called lower subadditivity whereas the second is called upper subadditivity. The certainty effect is characterized by upper subadditivity, since the increase of a high probability to certainty has more effect than the same increase in probability for a medium probability. This kind of behavior has also been documented under risk. As will be explained in Section 6, our DM faces "imprecise risk" in the sense that his knowledge of the probability distribution is imprecise. In our interpretation the DM transforms an imprecise probability distribution.

Empirical evidence (Prelec (1998), Wu and Gonzalez (1996)) suggests that the probability weighting function has the following properties:

asymmetric - $v(p) = p$ approximately at $p = 1/3$, where p is the probability to get at least a certain amount.

s-shaped - concave on an initial interval and convex beyond

reflective - assigns the same weight to a given loss-probability as to a given gain-probability.

If the preference relation satisfies a certain number of axioms given in the appendix A1, the preference relation can be represented by a sign- and rank-dependent utility function. Prelec (1998) provides a set of axioms needed to deduce a functional form of the weighting function that respects the stylized facts. The most general specification for this weighting function is given by the following form:

$$v(p) = \gamma \exp\left(-\beta\left(-\ln p\right)^{\alpha}\right) \qquad (15)$$

However, if the preference relationship satisfies certain axioms, notably diagonal concavity, sub-proportionality and compound invariance (see Prelec (1998)), then the specification takes the following less complex form:

$$v(p) = \exp(-(-\ln p)^{\alpha}) \qquad (16)$$

where α indicates the degree of non-linearity of the function. One typically observes risk-seeking for small-probability gains and large-probability losses, and risk-aversion for small-probability losses and large-probability gains. A linear value function would be sufficient to explain observed behavior. However, if the value function is not linear, then the probability non-linearity must dominate the value non-linearity to explain empirical patterns. Such a dominance arises when the above probability weighting function is combined with a power function of the following form:

$$v(x) = x^{\sigma^+} if \quad x \geq 0 \quad , \quad -(-x)^{\sigma^-} if \quad x < 0 \tag{17}$$

Concerning the weighting function, the values of the parameters of v can be slightly different across studies, the functional form is quite robust, see Prelec (1998). However, some authors have proposed other functional forms deduced from their theories. For instance, Grant and Kajii (1998) discuss disappointment aversion (see Gul (1991)) in the rank dependent expected utility framework. They show that disappointment aversion can be modelled with $v(p) = p^\alpha$.

Now, section 4 shows how Prelec's probability weighting function can be estimated from customer data.

5 Estimation of the probability weighting function

We now briefly explain how the preference functional of costumers can be estimated. The idea is to use the methodology used to test for probability weighting as a tool for estimating investor preferences. If we know the customers' preferences we are able to select their optimal portfolio strategies with the methodology develop in the next sections. So we consider the most general preferences, hence the title "generalized preferences".

In the previous section, we saw that the most general form of $v(p)$ is the following:

$$v(p) = \gamma \exp\left(-\beta \left(-\ln p\right)^\alpha\right) \tag{18}$$

where p is the cumulative distribution function. The certainty equivalent is given by $u(c) \sim u(prospect)$. Assume that the utility function is x^δ. This leads to the following relationship:

$$\frac{c}{x} = \exp\left\{-\frac{\beta}{\delta} \left(-\ln p\right)^\alpha\right\} \tag{19}$$

Think of x as the return of a portfolio. We can then ask the value of the certainty equivalent c for different combinations of returns and probabilities. Formally, we ask him what c were if the prospect were at least a given return with a given probability. Since β is generally equal to 1, as in (19), we can transform the above relationship in the following way,

$$-\ln\left(-\ln\left(\frac{c}{x}\right)\right) = \ln \delta + \alpha \left(-\ln\left(-\ln p\right)\right) \tag{20}$$

it can be estimated by OLS. We are thus able to recover the weighting function for the CEU model. Notice, that we are able to estimate the

concavity of the utility function at the same time as the weighting function. In that sense, the utility function is consistent with the weighting function, which would not necessarily be the case if we estimated them in two steps.

6 The Statistical Decision Problem of the Investor

We model the problem of a decision maker (henceforth DM) observing the multivariate process of variables X (asset prices but also state variables). He tries to select a portfolio consistent with his preferences. Standard expected utility theory would advocate estimation of a statistical model (predictive model for returns), use that prediction as the objective probability and finally optimize the portfolio consistent with the utility function. This viewpoint is in line with the classical statistics approach in the sense that probabilities are "objective". We argue in a more Bayesian flavor, that our DM cares (and should care) about his precise decision problem and is well aware of his "imperfect view of the world". This means that the estimator as well as the portfolio weights are deduced (jointly) from preferences. Now, our DM internalizes the fact that his predictive model is (almost surely) flawed and knows that the density of returns could be different from the estimated one. We assume that this is due to parametric uncertainty and that it is as if nature drew another parameter θ from the parameter set Θ (see Blackwell and Girshick (1965)).

This is consistent with the Anscombe-Aumann (1963) way to model uncertainty. Individuals believe that there is a multivariate Data Generating Process (DGP) $g(R|X,\theta)$. Elements θ (vectors of parameters) from Θ are the states of the world. P is the set of all simple distributions on the set of return outcomes R (of the portfolio). Φ is the set of all functions from Θ to P. More exactly, elements of Φ have the following form : $(\omega_1, ..., \omega_n, F(R|X,\theta))$ where $F(R|X,\theta)$ denote the estimated joint density of returns of the assets and ω_i the weight of the asset in the portfolio. Acts are thus portfolio weights as well as statistical models. X is fixed and observed by the DM. Now, elements of Φ are denoted by φ. The standard expected utility DM would maximize $\int u(\varphi(\theta,\omega))\pi(\theta|X)d\theta$, where $\pi(\theta|X)$ is a probability measure over the events of the parameter states Θ given the observed sample X. We will discuss this issue in the next section.

We argue that this would be reasonable if the DM had a perfect knowledge of $(\Theta, \Re)$. However, it is very difficult, even for an expert in statistics, to grasp the underlying process. For instance, regime changes are difficult to detect and regime parameters difficult to estimate. We argue that for some regions the DM has good knowledge to form prob-

abilistic beliefs about the events, while for some yet unseen region the DM has a blurred view and has even difficulty to define the events. In fact, given a finite sample size he has difficulties to get a clear grasp of some parameter values. As described in section 2, the DM maker knows the set $Q(\eta)$ of multiple priors inferred from a limited data set. Verlaine (2003) argues, in line with Jaynes (1957) and (1982), that it is rational for such a DM to deduce the posterior by maximizing entropy and that this leads to the functional form of π^* suggested by Prelec (1998).

The individual hence maximizes the following function : $\int u(\varphi(\theta, \omega))\pi^*(\theta|X)$ where $\pi^*(.)$ is a (as v described in section 4) transformation of $F(R|X, \theta)$, .which is a function of θ for a fixed data set X. Now we have to devise an econometric technique to estimate the parameters (amounts to estimating a forecasting function $F(.)$) as well as the portfolio weights.

7 Likelihood principle, weak conditionality and independence

As we saw in the last section a standard Bayesian (DM) would maximize $\int u(\varphi(\theta, \omega))\pi(\theta|X)d\theta$, where $\pi(\theta|X)$ is a probability measure over the events of the parameter states Θ given the observed sample X. As is well known from Bayes Theorem:

$$\pi(\theta|X) = \frac{f(X|\theta)\pi(\theta)}{\int_\Theta f(X|\theta)\pi(\theta)\,d\theta} \propto f(X|\theta)\pi(\theta) \tag{21}$$

where $f(X|\theta)$ is the density of X for a given θ. The latter density is given by the likelihood function $l(\theta)$ and this leads to the following expression:

$$\pi(\theta|X) \propto l(\theta)\pi(\theta) \tag{22}$$

$f(X|\theta)$ is now considered a function of θ. Instead of maximizing the likelihood function $l(\theta)$ to recover θ, the Bayesian maximizes the generalized likelihood function $l(\theta)\pi(\theta)$.

However, this presupposes that the Likelihood Principle is not violated. Recall that classical statistics presumes randomization over the sample space. Now, the Likelihood Principle stems from two axioms: sufficiency and weak conditionality. While sufficiency will not be questioned here, weak conditionality seems less unquestionable. The idea behind weak conditionality is that our DM faced with a realization of the sample (while other samples could have occurred) doesn't care about the other possible sample realizations and infers the parameter value by considering only the sample he has. All relevant experimental information is contained in the likelihood function for observed X. The following

description is given in Berger (1985).

The Weak Conditionality Principle. *Suppose one can perform either of two experiments E_1 and E_2, both pertaining to θ, and that the actual experiment conducted is the mixed experiment of first choosing $J = 1$ or 2 with probability $\frac{1}{2}$ each (independent of θ), and then performing experiment E_J. Then the actual information about θ obtained from the overall mixed experiment should depend only on the experiment E_j that is actually performed.*

As was pointed out earlier experimental evidence suggests that the independence axiom is systematically violated. Moreover, violations of the independence axiom are responsible for the non linear probability weighting. As shown below, violations of the independence axiom imply violations of weak conditionality.

More formally, let Ω be a state space with elements s and Y be the set of distributions over Ω with finite supports

$$\left\{ Y = y : \Omega \to [0,1] \,|\, y(s) \neq 0 \, for \, finitely \, many \ sin\Omega and \sum_{s \in \Omega} y(s) = 1 \right\}$$
(23)

Convex combinations in Y are performed point-wise, i.e. for f and g in Y and α in $[0,1]$, $\alpha f + (1-\alpha)g = h$ means $\alpha f(s) + (1-\alpha)g(s) = h(s)$ for $s \in \Omega$.

We can now define the independence axiom.

Definition: The individual respects the independence axiom iff for all f, g and h in Y and for all α in $(0,1)$: $f \succ g$ implies $\alpha f + (1-\alpha)h \succ \alpha g + (1-\alpha)h$.

Now, rewrite the problem with $\theta \in \Theta$ (parameters are the states) and the likelihood functions $f(\theta)$, $g(\theta)$ and $h(\theta)$ of experiments f, g and h. We can thus rewrite the whole axiom with the likelihood notation and parametric uncertainty. The independence axiom can be interpreted in the following way: Ex-ante there is a probability α to draw experiment f (a sample in the frequency interpretation) and $(1-\alpha)$ to draw h. This is preferred to the mixed experiment with g. One of the experiments (samples) will be drawn and the weak conditionality principle implies that we should not care (for parametric inference) about what could have happened had the other experiment (sample) occurred. Now, if experiment f is preferred to experiment g, then the DM should change his preference if those experiments are mixed with another experiment (ex-ante).

Since the weak conditionality principle is conceptually the same thing as the independence axiom, violations of the latter imply violations of

weak conditionality. In a sense the DM is not consequentialist and doesn't snip off the part of the decision tree that didn't occur. This kind of behavior is consistent with non-linear probability weighting and we discuss parametric inference for such a decision maker.

8 Robust Bayesian posterior and estimation of the forecasting function.

Above, we saw that the posterior $\pi(\theta|X)$ is calculated using the likelihood function. From the weak conditionality principle we use this realized likelihood function to infer θ, without regard of the fact that another value of the likelihood could have obtained. According to experimental evidence DM's do not reason that way. We thus advocate a methodology consistent with violations of weak conditionality and behavior in terms of robustness. This technique can then be used to infer "robust" $\theta's$. The idea is based on the maximum entropy approach to determine probabilities on the parameter space.

Jaynes (1957) argues that a way to treat ill-posed statistics problems, is to consider observed functions , typically moments, of the possible events, as part of the available information. He suggests to consider the probability distribution that presumes less given the data. This is done by maximizing uncertainty given the realized values of the above mentioned functions. Uncertainty is measured with Shannon's information theoretic entropy measure. In fact, even though not the only measure of uncertainty, Shannon's measure is consistent with additivity of independent risks. More details can be found in a companion paper (Verlaine (2003)). Actually, the following entropy optimization postulate given in Kapur and Kesavan (1992), generalizes the principle.

Each probability distribution is an entropy optimization distribution; i.e., it can be obtained by maximizing an appropriate entropy measure or by minimizing a cross-entropy measure with respect to an appropriate a priori distribution, subject to its satisfying appropriate constraints.

Now, we consider a DM that faces an observed likelihood function but knows that he could have drawn another sample and hence another likelihood. As argued , a rigorous way to be robust, given partial information, is to maximize entropy under the constraint of partial information (Jaynes (1982)). We apply this method adapted to our situation, but the general developments are given in appendix A3. We can now prove the following proposition:

Proposition 1 *A DM maximizing uncertainty, given a sample cumulative probability to get a certain θ, reasons with a non-additive prior on Θ. Moreover, the latter has the same functional form as the density*

consistent with the probability weighting function in equation (11).

Proof: The general problem of a DM maximizing uncertainty subject to constraints given by the sample is:

$$En\left(\pi\right) = -\int \pi\left(\theta|X\right) \log\left(\frac{\pi\left(\theta|X\right)}{\pi_0\left(\theta\right)}\right) d\theta \tag{24}$$

s.t.

$$E^\pi\left[g_k\left(\theta\right)\right] = \int_\Theta g_k\left(\theta\right)\pi\left(\theta|X\right)d\theta = \mu_k \qquad k = 1, ..., m \tag{25}$$

where $g_k(\theta)$ are functions of the parameters and π_0 is a noninformative prior. The solution (see Berger (1985) is given by

$$\pi^*\left(\theta|X\right) = \frac{\pi_{0(\theta)}\exp\left[\sum_{k=1}^m \lambda_k g_k\left(\theta\right)\right]}{\int_\Theta \pi_0\left(\theta\right)\exp\left[\sum_{k=1}^m \lambda_k g_k\left(\theta\right)\right]d\theta} \tag{26}$$

Now let's consider the following constraint $g_k(\theta) = \left(-\log cdfn\left(l(X|\theta)\right)\right)^\alpha$ which is a function of θ for a fixed data set, hence $l(\theta)$ the likelihood function. Thus, $g_k(\theta) = \left(-\log cdfn(l(\theta))\right)^\alpha$ Notice the relationship of the function with the content of information. In fact, if you rewrite it in the following way: $g_k\left(\theta\right) = \left(\log\frac{1}{cdfn(l(\theta))}\right)^\alpha$, it is a measure of information content of the distribution with respect to certainty. Now if the DM maximizes the above mentioned program under this constraint, we get the following :

$$\pi^*\left(\theta|X\right) = \frac{\pi_{0(\theta)}\exp\left[\lambda_1\left(-\log cdfn(l(\theta))^\alpha\right]}{\int_\Theta \pi_0\left(\theta\right)\exp\left[\lambda_1\left(-\log cdfn(l(\theta))^\alpha\right]d\theta} \tag{27}$$

Considering a value for the denominator δ (after having integrated), we recover the density in appendix A2 (equation 27), with $\lambda_1 = -1$ and $\pi_0(\theta) = \alpha\frac{l(\theta)}{cdfn(l(\theta))}\left(-\ln(cdfn(l(\theta)))\right)^{\alpha-1}d\theta.$ $\square$
The robust posterior is the following

$$\pi^*(\theta|X) \propto \alpha\frac{l(\theta)}{cdfn(l(\theta))}\left(-\ln(cdfn(l(\theta)))^{\alpha-1}\exp\left(-\left(-\ln(cdfn(l(\theta)))\right)^\alpha\right)\ d\theta \tag{28}$$

and the parameter can be estimated in the following way:

$$\theta^* = \arg\max_{\theta\in\Theta}\pi^*(\theta|X) \tag{29}$$

This is consistent with our interpretation of parametric uncertainty. Parametric uncertainty is taken into account in the estimation process.

Now, recall from section 5 that in a first step, our DM takes into account parametric uncertainty (as described above) and then maximizes Expected Utility as if he knew the parameter. As has been shown by Levy and Markowitz (1979) and Kroll, Levy and Markowitz (1984) show that Mean-variance portfolio optimization is an operational way to select portfolios from an infinite number of possible portfolios, and this for various non-quadratic (expected !) utility functions. However, as shown by Choi (2001) this result is not robust to non-linear probability weighting as is the case in rank dependent expected utility models. Since probability weighting is treated in the first step, we now advocate Mean variance optimization with robust estimators as described above. Our DM hence acts consistently with Choi's findings.

9 The setup of the dynamic portfolio problem

We consider a standard dynamic portfolio problem in continuous time with a changing investment opportunity set. The financial market consists of the following three processes for the risky asset P_t , risk-less asset B_t and state process S_t driving the dynamics of the investment opportunity set.

$$\frac{dP_t}{P_t} = \mu_p(S,t)dt + \sigma_p(S,t)dZ_{pt} \tag{30}$$

$$\frac{dB_t}{B_t} = r(S,t)dt \tag{31}$$

$$dS_t = \mu_s(S,t)dt + \sigma_s(S,t)dZ_{st} \tag{32}$$

where dZ_{pt} and dZ_{st} are Wiener processes with $dZ_{pt}dZ_{st} = \rho_{ps}(S,t)dt$. The drifts, volatilities and correlations may be functions of S and t but for the following development we simply write μ_p and μ_s.

If the investors preferences defined over consumption are time-separable, we can formulate his optimal portfolio and consumption problem in the following way:

$$\max_{C,\alpha} E_0 \left[\int_0^\infty U(C,t)dt \right] \tag{33}$$

subject to the continuous-time inter-temporal budget constraint

$$dW_t = \left[(\alpha_t (\mu_p - r) + r) W_t - C_t \right] dt + \alpha_t W_t \sigma_p dZ_{pt} \tag{34}$$

with the obvious constraints that consumption $C_t > 0$ and wealth $W_t > 0$. α_t denotes the fraction of wealth invested in the risky asset.

Consider the following value function $J(W, S, t)$. The Bellman principle implies:

$$0 = \max_{C,\alpha} \left\{ U(C, t) + \frac{1}{dt} E_t \left[dJ(W, S, t) \right] \right\} \tag{35}$$

Now, by Itô's Lemma we have the following dynamic for the value function:

$$dJ(W, S, t) = J_W dW + J_S dS + \left(\frac{\partial J}{\partial t} \right) dt + \frac{1}{2} J_{WW} (dW)^2 + J_{SW} dW dS + \frac{1}{2} J_{SS}(dS)^2 \tag{36}$$

From the last two equations, we derive the following Hamilton-Jacobi-Bellman equation:

$$0 = \max_{C,\alpha}\{U(C, t) + J_W[(\alpha(\mu_p - r) + r)W - C] + J_S \mu_S + \frac{\partial J}{\partial t}$$

$$+ \frac{1}{2} J_{WW} \alpha^2 W^2 \sigma_P^2 + J_{WS} \alpha W \sigma_p \sigma_s \rho_{ps} + \frac{1}{2} J_{ss} \sigma_s^2\}$$

Now, we impose the following transversality condition:

$$\lim_{t \to \infty} E_0 \left[J(W, S, t) \right] = 0 \tag{37}$$

This condition is imposed to make sure that no investment strategy leads to infinite utility.

By deriving the Hamilton-Jacobi-Bellman equation with respect to C and α, we get the following policy (optimal rule).

$$U_C = J_W$$

$$\alpha = \frac{1}{-J_{WW} W / J_W} \left(\frac{\mu_p - r}{\sigma_p^2} \right) - \frac{J_{Ws}}{J_{WW} W} \left(\frac{\sigma_S}{\sigma_p} \rho_{ps} \right)$$

The first equation determines the optimal consumption policy, while the second equation determines the optimal portfolio policy. The first term of the second equation is the myopic portfolio rule: a function of the relative risk aversion and the price of risk. The second term is the inter-temporal hedging demand. Basically, this term is not zero if investment opportunities are time varying $\sigma_S > 0$.

Now, the policy still depends on the value function J which is determined by substituting the optimal policy rules back in the H-J-B equation. This leads to a second-order partial differential equation for the value function. Once we have solved for the value function, we have the policy rule.

15

10 The dynamic portfolio policy with generalized preferences

Now, we show how to derive the optimal policy with the preferences described in section 4. The investment opportunity set is the same as the one described in the last section. As argued by Prelec a compact representation of preferences is given by x^δ for the utility function in certainty and the following PWF $v(p) = \exp(-(-\log p)^\alpha)$. We have argued in section 7 that the PWF is treated in the estimation process. The idea is the following. Instead of trying to estimate the process in each detail (Jump diffusion, stochastic volatility) and solve very complex models, our decision maker opts for a rule of thumb. He is aware that his rule of thumb is "wrong" and adjusts the parameters according to his aversion to uncertainty. The parameters are thus estimated with the subjectively transformed likelihood function.

The H-J-B equation for our problem is :

$$0 = \max_{C,\alpha}\{C^\delta + J_W[(\alpha\,(\mu_p - r) + r)W - C] + J_S\mu_S + \frac{\partial J}{\partial t}$$

$$+ \frac{1}{2}J_{WW}\alpha^2 W^2\sigma_P^2 + J_{WS}\alpha W\sigma_p\sigma_s\rho_{ps} + \frac{1}{2}J_{ss}\sigma_s^2\}$$

If we derive the optimal policies, we get the same policy for α as in the section above. The consumption policy is the following: $\delta C^{\delta-1} = J_W$ hence $C = \frac{1}{\delta}J_W^{\frac{1}{\delta-1}}$. In order to determine $J(W,S,t)$ we replace C with $\frac{1}{\gamma}J_W^{\frac{1}{\delta-1}}$ in the above equation. $J(.)$ can now be determined with numerical techniques. Finally, by deriving J with respect to W and S we can calculate α, the weight of the risky asset.

11 Conclusion

We developed a tractable way to choose portfolios consistent with non-linear transformation of observed probability distributions. Our interpretation is that the estimated density gives an indication on the set of possible probability distribution, but is not enough to pin down one unique distribution. In that sense, our DM faces "imprecise risk" and picks the most uncertain distribution, given his preferences and given the data. We invoke instrumentalism and argue that observed behavior, notably Prelec's weighting function, can be explained by maximizing behavior. In fact, it is shown that Prelec's probability weighting function can be recovered by maximizing entropy under the constraint of observed moments (mean, variance...).

Since the estimated likelihood function also retraces one of different possible realizations, the above mentioned reasoning can be applied to it. We argue that violations of the independence axiom or sure thing principle imply violations of the weak conditionality principle, which is central for the likelihood principle and parametric inference. Our DM is careful and doesn't trust the likelihood function and hence maximizes a weighted likelihood function. We argue that this is a way to take robust decisions and consistent with maximum entropy principle. Portfolio weights can then be chosen in a simple way by simply putting the robust estimated mean variance parameters in the standard mean variance portfolio choice model.

This is a first attempt to solve complex portfolio problems with non-linear probability weighting behavior. Our method is based on the standard information theory entropy concept. However, developing dynamic portfolio strategies with generalized entropy, deduced from axioms of individual behavior, may be an interesting avenue for future research. Moreover, developing tractable techniques to determine customer profiles from individual data is important to make the method useable for private banking institutions.

11.1 A1:

The appendix collects axioms from Prelec (1998). They consist in an axiomatization of the sign- and rank-dependent representation. Let $\succsim$ represent a preference relation on the set Π of probability distributions $P, Q, ...$ on $X = [x^-, x^+]$, with $x^- < 0 < x^+$. The following five axioms are assumed to hold on Π without restriction:

A1. Weak Order: $\succsim$ is complete and transitive.

A2. Strict Stochastic Dominance: $P > Q$ if $Q \neq P$ and P stochastically dominates Q.

A3. Certainty Equivalent Condition: $\forall P, \exists x \text{ s.t. } (x) \sim P$.

A4. Continuity in Probabilities: If $(y, p) > (x)$, $0 < p < 1$, then there exists q, r s.t. $q < p < r$, $(y, q) > (x)$, and $(y, r) > (x)$. If $(y, p) < (x)$, $0 < p < 1$, then there exists q, r s.t. $q < p < r$, $(y, q) < (x)$ and $(y, r) < (x)$.

Let's define $S(k, n)$, $0 \leq k \leq n$, as the set of all k non-positive and $(n-k)$ nonnegative rank-ordered n-tuples from X, $S(k, n) = \{(x_1, ..., x_n) \in X : x_1 \leq ... \leq x_k \leq 0$

A5. Simple-Continuity: For any probability vector $(p_1, ..., p_n)$ the preference relation induced on each S(k,n) is continuous.

Let $(x, p_i; x_{-i}, p_{-i})$ denote a prospect with outcome x of rank "i" singled out, and let $\Pi(k, n, p)$ denote the set of all sign- and rank-order compatible prospects that have a p-chance of yielding a negative outcome: $\Pi(k, n, p) = \{(x_1, p_1; ...; x_n, p_n) : (x_1, ..., x_n) \in S(k, n) and : p_1 + ... + p_k = p\}$

A6. Tradeoff Consistency: There do not exist eight prospects, $(x, p_i; a_{-i}, p_{-i})$, $(x, p_i; b_{-i}, p_{-i})$, $(y, q_j; d_{-j}, q_{-j})$, such that the first four and the second four belong to the same sign- and rank-order compatible set, and that:

$$(x, p_i; a_{-i}, p_{-i}) \succsim (y, p_i; a_{-i}, p_{-i}),$$

$$(x', p_i; a_{-i}, p_{-i}) \precsim (y', p_i; b_{-i}, p_{-i})$$

$$(x', q_j; c_{-j}, q_{-j}) \succsim (y', q_j; d_{-j}, q_{-j})$$

$$(x, q_j; c_{-j}, q_{-j}) \precsim (y, q_j; d_{-j}, q_{-j})$$

If A1 to A6 are satisfied, then the preference relationship can be represented by the following functional:

$$V(P) = \sum_{i=1}^{k} \left(v^- (\sum_{j=1}^{i} p_j) - v^- (\sum_{j=1}^{i-1} p_j) \right) u(x_j) + \sum_{i=k+1}^{n} \left(v^+ (\sum_{j=1}^{n} p_j) - v^- (\sum_{j=i+1}^{n} p_j) \right) u(x_j)$$

$$(38)$$

11.2 A2:

Let's consider the following weighting function $v = \exp\left(-(-\ln p)^\alpha\right)$. If we denote $cdfn(l(\theta))$ the cumulative distribution function v takes the following form $v = \exp\left(-(-\ln(cdfn(l(\theta))))^\alpha\right)$. We assume that v as well as $cdfn(l(\theta))$ are C^1. Differentiating v leads to

$$dv = \alpha \frac{l(\theta)}{cdfn(l(\theta))} \left(-\ln(cdfn(l(\theta)))\right)^{\alpha-1} \exp\left(-(-\ln(cdfn(l(\theta))))^\alpha\right) \, d\theta$$

$$(39)$$

where $l(\theta)$ denotes the likelihood.

11.3 A3:

We review the developments of the different Jaynes papers collected in Jaynes Papers on Probability, Statistics and Statistical Physics. First we develop the conceptually simpler discrete case. We then discuss the continuous case. Consider a quantity x that can take n different values $(x_1, x_2, ..., x_n)$ and functions $f_1(x), ...f_m(x)$ where $m < n$. The idea of maximum entropy is to find a probability assignment $p(x_i) = p_i$ that maximizes uncertainty given a few functional contraints from the data.

As we will see later another interpretation (in terms of frequency) is that the maximum entropy distribution, is that frequency distribution which can be realized in the greatest number of ways and hence is the most likely to occur.

More formally the maximum entropy formalism is the following:

$$Max \quad -\sum_{i=1}^{n} p_i \log p_i \quad s.t.$$

$$p_i \geq 0$$

$$\sum_{i=1}^{n} p_i f_k(x_i) = \mu_k$$

$$\sum_{i=1}^{n} p_i = 1$$

The solution is

$$p_i \equiv \frac{1}{Z(\lambda_1, ..., \lambda_m)} \exp\left[-\lambda_1 f_1(x_i) - ... - \lambda_m f_m(x_i)\right]$$

with the partition function :

$$Z(\lambda_1, ..., \lambda_m) \equiv \sum_{i=1}^{n} \exp\left\{-\lambda_1 f_1(x_i) - ... - \lambda_m f_m(x_i)\right\}$$

The Lagrange multipliers are determined from

$$\mu_k = -\frac{\partial \log Z(.)}{\partial \lambda_k} \qquad k = 1, .., m$$

The above density is the "maximum non-commital" distribution given the sample information. Jaynes suggests to use this method to form prior probabilities in a least ad hoc way. Notice that the above proabilities are not defined as frequencies, but are subjective "maximum non-commital" probabilities. Now, Jaynes also discusses the frequency case. Consider the above random experiment that is repeated N times, the result x_i will be obtained m_i times, $i = 1, 2, ..., n$. The probability (frequency) p_i is now given by $\frac{m_i}{N}$ and m_i is not uniquely determined. Now we can ask what is the "best" estimate, to be discussed afterwards. Now, in N repetitions of the random experiment, there are a priori n^N conceivable

results. For observed $m_i's$, out of the original n^N , how many would lead to a given set of observed $m_i's$? The answer is given by the multinomial coefficient :

$$W = \frac{N!}{m_1! \cdot \bullet \cdot m_n!} = \frac{N!}{(Np_1)! \cdots (Np_n)!}$$

Hence the set of $p_i's$ that can be realized in the greatest number of ways is the one which maximizes W subject to the above mentioed constraints. Now, in the limit for $N \to \infty$, we have by the Stirling formula,

$$\lim_{N \to \infty} \frac{1}{N} \log W = \lim_{N \to \infty} \frac{1}{N} \log \left[\frac{N!}{(Ng_1)! \cdots (Ng_n)!} \right] = - \sum_{i=1}^{n} p_i \log p_i$$

.

12   Bibliographic References

Anderson, E., Hansen, L.P. and Sargent, Th. (2003) "A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection", Working Paper.

Anscombe, F. J. and Aumann, R. J. (1963) "A Definition of Subjective Probability", *Ann. Math. Stat.* Vol. 34, p. 199-205.

Berger, J. (1985) *Statistical Decision Theory and Bayesian Analysis* , Springer Series in Statistics, Second Edition.

Billingsley, P. (1986) *Probability and Measure,* New York: Wiley

Blackwell, D, and Girsnick, M. (1965) *Theory of Games and Statistical Decisions*, Wiley

Camerer, C. and Ho, T. (1994) "Violations of the Betweenness Axiom and Nonlinearity in Probability", *Journal of Risk and Uncertainty*, Vol. 8, p. 167-196.

Cavadini, F., Sbuelz, A. and Trojana, F. (2001) "A Simplified Way of Incorporating Model Risk, Estimation Risk and Robustness in Mean Variance Portfolio Management", Working Paper.....

Chateauneuf, A. and Wakker, P. (1999) "An Axiomatization of Cumulative Prospect Theory for Decision under Risk", *Journal of Risk and Uncertainty*, Vol. 18.

Choi, P. (2001) "A Failure in Mean-Variance Approximation to the Utility Function of Prospect Theory", Working Paper, Texas A&M University.

Chow, G. (1995) "Portfolio Selection based on Return, Risk, and Relative Performance", *Financial Analysts Journal*, p. 54-60.

Diecidue, E. and Wakker, P. (2000) "Comonotonic Book-making with nonadditive probabilities", Working Paper, *CentER, Tilburg University.*

Epstein, L.G. (1999) "A Definition of Uncertainty Aversion", *Review of Economic Studies*, 66, 579-608.

Epstein, L. and Zhang, J. (2001) "Subjective Probabilities on Subjectively Unambiguous Events", *Econometrica,* Vol. 69, p. 265-306.

Gilboa, I. (1987) "Expected Utility with Purely Subjective Non-additive Probabilities", *Journal of Mathematical Economics*, Vol. 16, p. 65-88.

Gilboa, I. and Schmeidler, D. (1989) "Maxmin Expected Utility with Non-unique Prior", *Journal of Mathematical Economics*, Vol. 18, p. 141-153.

Goldfarb, D. and Iyengard, G. (2003) "Robust portfolio selection problems", CORC Report 2002-2003.

Grant, S. and Kajii, A. (1998) "AUSI expected utility: An anticipated utility theory of relative disappointment aversion", *Journal of Economic Behavior and Organization,* Vol. 37.

Gul, F. (1991) "A Theory of Disappointment Aversion", *Econometrica,* Vol. 59, p. 667-686

Hansen, L.P., Sargent, Th., Turmuhambetova, G. and Williams N. "Robust Control and Model Misspecification", Working Paper.

Jaynes, E. T. (1957) "Information Theory and Satistical Mechanics", Reprinted from The Physical Revie, Vol. 106; No. 4, 620-630

Jaynes, E. T. (1982) "On the Rationale of Maximum-Entropy Methods", *Proceedings of the IEEE*, Vol. 70, No. 9

Jaynes, E. T. (1984) "The intuitive inadequacy of Classical Statistics", Epistemologia VII, in *Probability, Statistics and Inductive Logic*, pp. 43-74.

Jaynes, E. T. (1988) "How does the Brain do Plausible Reasoning", *Maximum-Entropy and Bayesian Methods in Science and Engineering*, Vol. 1, p. 1-24, Kluwer Academic Publishers.

Kahneman, D. and Tversky, A. (1979) "Prospect Theory: An Analysis of Decision under Risk", *Econometrica*, Vol. 47, p. 263-291.

Kahneman, D. and Tversky, A. (1992) "Advances in Prospect Theory: Cumulative Representation of Uncertainty", *Journal of Risk and Uncertainty,* p. 297-323.

Karni, E. (1993) "A Definition of Subjective Probabilities with State-dependent Preferences", *Econometrica*, Vol. 61, p. 187-198.

Karni, E. and Schmeidler, D. (1991) "Utility Theory with Uncertainty", *Handbook of Mathematical Economics,* Vol. 4, Elsevier Science Publishers Ed.

Kitamura, Y. and Stutzer, M. (1997) "An Information-Theoretic Alternative to Generalized Method of Moments Estimation", *Econometrica*, Vol. 56, No. 4, p. 861-874.

Kojadinovic, I. , Marichal, J.-L. and Roubens, M. (2001) "An axiomatic approach to the definition of the entropy of a discrete Choquet capacity", Working Paper, Institute of Mathématics, Université de Liège.

Kroll, Y., Levy, H. and Markowitz, H. (1984) "Mean-Variance Versus Direct Utility Maximization", Journal of Finance 39 (1), 47-61.

Levy, H. and Markowitz, H. (1979) "Approximating Expected Utility by a Function of Mean and Variance", American Economic Revie 69(3), p. 308-317.

Machina, M. (1982) "Expected Utility Analysis without the Independence Axiom", *Econometrica*, Vol. 50, p. 277-323.

Machina, M. (1989) "Dynamic Consistency and Non-Expected Utility Models of Choice Under Uncertainty", *Journal of Economic Literature*, Vol. 27, p. 1622-1668.

Machina, M. and Schmeidler, D. (1992) "A more robust definition of Subjective Probability", *Econometrica*, Vol. 60, p. 745-780.

Marichal, J.-L. (2001) "Entropy of Discrete Choquet Capacities", Working Paper, Institute of Mathematics, Université de Liège.

Marichal, J.-L. and Roubens, M. (?) "Entropy of Discrete Fuzzy Measures", Institute of Mathematics, Université de Liège.

McFadden, D (1999) "Rationality for Economists ?", *Journal of Risk and Uncertainty*, Vol 19, p. 73-105.

Perret-Gentil, C. and Victoria-Feser, M.-P. (2003) "Robust Mean Variance Portfolio Selection", Working Paper, HEC University of Geneva.

Prelec, D. (1998) "The Probability Weighting Function", *Econometrica*, Vol. 66, p. 497-527.

Savage, L. (1972) *The Foundations of Statistics*, Dover Publications , 2nd edition.

Schmeidler, D. (1989) "Subjective Probability and Expected Utility without Additivity", *Econometrica*, Vol. 57, p. 571-587.

Simon, H.A. (1989) "Rationality as process and as product of thought", in *Decision making: Desriptive, normative, and prescriptive interactions,* p. 58-77.

Skiadas, C. (1997) "Conditioning and Aggregation of Preferences", *Econometrica*, Vol. 65, No. 2, p. 347-367.

Sugden, R. (1993) "An Axiomatic Foundation of Regret Theory", Journal of Economic Theory, Vol. 60, p. 159-180.

Verlaine (2003) "Transformations of Distributions, Robustness and Coherent Risk Measures", Working Paper GRID.

Victoria-Feser, M.-P. (2000) "Robust Portfolio Selection", Working Paper, University of Geneva.

Zhang, (2002) "Subjective ambiguity, expected utility and Choquet expected utility", *Economic Theory* 20, p. 159-181.

Wu, G. and Gonzales, R. (1996) "Curvature of the Probability Weighting Function", *Management Science*, Vol. 42, p. 1676-1690.