Portfolio Selection with Probabilistic Utility

Robert Marschinski¶, Pietro Rossi, Massimo Tavoni*, and Flavio Cocco

Prometeia S.r.l., Via Marconi 43, 40100 Bologna, Italy

E-mail: `robert.marschinski@pik-potsdam.de`, `pietro.rossi@prometeia.it`, `massimo.tavoni@feem.it`, `flavio.cocco@prometeia.it`

Abstract. We present a novel portfolio selection technique, which replaces the traditional maximization of the utility function with a probabilistic approach inspired by statistical physics. We no longer seek the single global extremum of some chosen utility function, but instead reinterpret the latter as a probability distribution of 'optimal' portfolios, and select the portfolio that is given by the mean value with respect to that distribution. This approach has several attractive features, when comparing it to the standard maximization of expected utility. First, it significantly reduces the over-pronounced sensitivity to external parameters that plague optimization procedures. Second, it mitigates the commonly observed concentration on too few assets; and third, it provides a natural and self consistent way to account for the incompleteness of information and the aversion to uncertainty. Empirical results supporting the proposed approach are presented by using artificial data to simulate finite-sample behavior and out-of-sample performance. With regard to the numerics, we carry out all integrals by using Markov Chain Monte Carlo, where the chains are generated by an adapted version of Hybrid Monte Carlo.

Keywords: Portfolio Selection, Estimation Error, Parameter Uncertainty, Probabilistic Utility, Asset Allocation.

1. Introduction

Classical portfolio selection (Markowitz, 1952) by <u>M</u>aximization of <u>E</u>xpected <u>U</u>tility (MEU) suffers from well-documented drawbacks (Michaud, 1989): it often leads to extreme and hardly plausible portfolio weights, which additionally are very sensitive to changes in the expected returns. Moreover, it does not take into account informational uncertainty, since the straightforward optimization procedure implies complete faith in the estimated parameters. Historical data provide some information on future returns, but it is well known that simple-minded use of this information often leads to nonsense because estimation uncertainty (noise) overwhelms the value of the information contained in the data. In fact, the position of the extrema of a function tends to be highly sensitive to irrelevant distribution details, and it is quite simple to build examples where a minimal parameter variation induces a very large shift in the location of the global extremum. In other words, over-fitting quickly becomes an issue when data - as in all cases - is limited.

¶ Now at Potsdam Institute for Climate Impact Research, 14473 Potsdam, Germany.
* Corresponding author. Current address and new affiliation: Fondazione Eni Enrico Mattei, C.so Magenta 63, 20100 Milano, Italy.

The issue of uncertainty in expected returns and its implications for portfolio selection has been extensively analyzed in the relevant literature: starting with the work of Barry (1974) and Bawa, Brown, and Klein (1979), many authors have since addressed the problem, often resorting to a Bayesian framework (Black and Litterman, 1992; Jobson and Korkie, 1980; Jorion, 1985; Frost and Savarino, 1986; Michaud, 1989; Chopra and Zembia, 1993). Stimulated by the recent debate on asset return predictability, the issue has again gained attention (Balduzzi and Liu, 2000; Barberis, 2000; Britten-Jones, 2002; ter Horst, de Roon, and Weker, 2002; Pastor, 2000; Polson and Tew, 2000), and is at the momement becoming one of the most debated topics in applied finance. A popular approach that - somewhat similar to ours - implicitly abandons the idea of maximizing expected utility has been proposed by Michaud (1998): in an attempt to maintain the flexibility of the efficient frontier optimization framework but still accommodate parameter uncertainty, he suggested to create a 'resampled frontier' by repeatedly maximizing the utility for a draw from a probability distribution and then averaging the optimal weights that result from each optimization. His approach has been criticized as "heuristic" (Scherer, 2002), but, nonetheless, offers interesting insights and has received supporting evidence (Markowitz and Usmen, 2003).

A second line of more theoretical research has started to investigate questions of ambiguity and uncertainty: assuming that investors - given the difficulty of estimating moments of asset returns - will likely form multiple priors rather than a single one, researchers have asked which implications ambiguity has for the models of portfolio choice. Supported by substantial evidence from economic experiments showing that agents are not neutral to the ambiguity arising from having multiple priors (Ellsberg, 1961), recent scholarship (Anderson, Hansen, and Sargent, 1999; Chen and Epstein, 2002; Uppal and Wang, 2003; Garlappi, Uppal and Wang, 2004) has developed models of decision making that allow for multiple priors and a decision maker that is not neutral to uncertainty. The most important difference between the Bayesian approach and the multi-prior approach is that in the latter the investor is averse to uncertainty.

Our work is related to both of the above strands of research. It retains the basic Bayesian idea of accepting that any knowledge about parameters is merely probabilistic and linked to the available data, though in a more aggregate form; and it is also related to the resampling technique inasmuch the maximization of expected utility is replaced by some average, but it does so in a more coherent framework. Finally, it addresses the issue of ambiguity aversion by introducing a parameter that can implicitly accommodate for it.

In the following, we propose a methodology which explicitly aims at the procedure for selecting portfolios - independent from the details of a given utility function or the functional form of the underlying distribution. We replace the traditional maximization problem by suggesting a different interpretation of the utility function: we consider it to be the logarithm of the probability density for the portfolio to assume a given composition, and we define as "optimal" the expected value of the portfolio's weights with respect to that distribution.

This approach was conceived through inspiration by statistical physics. The dynamics of the so-called canonical ensemble are strikingly similar to the problem of portfolio selection. In fact, such a system is subject to two competing influences: on one hand it tends to minimize its internal energy, but on the other it also tries to maximize its entropy. The relative weight between these two 'rivals' is determined by the system's (constant) temperature: at a high temperature entropy dominates, since strong internal fluctuations bar the particles from settling down in any particular and energetically convenient state; conversely, at low temperatures the system will be more likely to assume a state of low internal energy, even in a configuration that is quite particular, e.g. highly asymmetric. Now, if we view internal energy as utility with reversed sign, entropy as a measure of portfolio diversification, and temperature as a measure of missing information on parameters (i.e. estimation error), the exact formal analogy with portfolio selection becomes evident. Although noticed already by other authors, such as Bouchaud, Potters, and Aguilar (1997) and Piotrowski and Sladkowski (2001), this close resemblance has not yet been fully utilized and put into practice as suggested here†.

As will be shown by means of numerical examples, our probabilistic utility approach leads to an improved portfolio selection procedure, meaning that we are able to overcome the excessive sensitivity to external parameters on one hand, and take into account the *incomplete* information obtained from finite historical time series on the other. As a matter of fact, parameters (means, covariances, skewness, etc.) determined by observation of historical data are not the only source of trouble for portfolios based on optimization with respect to an objective function: all of the procedures based on expected utility maximization in addition suffer from the presence of a scalar parameter related to the investor's risk aversion. The value of this parameter cannot be set by theory, but the resulting portfolio composition still tends to be quite sensitive to its value. Actually, due to a complete lack of scale for this risk-aversion parameter, it is usually adjusted ex post by hand, i.e. by merely observing where "the dynamics happen", and defining an *ad hoc* scale according to the simple prescription "increase the parameter if you want a more aggressive - meaning riskier - portfolio". This turns out to produce highly instable and inconsistent results, i.e. portfolios that might change (and usually do change) significantly in response to an apparently small shift in risk-aversion, and might even become more "aggressive" than a neighboring portfolio with a lower risk-aversion. This will be discussed in more detail in Section 4.

The final contribution of our paper concerns the numerical routines employed to perform all of the relevant integrals. Most of these cannot be computed explicitly, and therefore we will resort to a dynamical Monte Carlo integration or "Markov Chain Monte Carlo". To enhance performance we have used a variation of the Metropolis-Hastings prescription known as "Hybrid Monte Carlo", that first appeared in the physics literature in 1987 (Duane *et al* , 1987).

The performance of our proposed method is evaluated by comparing our probabilistic utility estimator (PU from now on) with the traditional maximization of expected utility (MEU). To isolate the effects of our approach, we assume a common utility function and a stationary parameter distribution (no Bayesian modeling). We investigate in different ways the portfolio sensitivity to the estimation errors of market parameters (means, covariances). As is well known, asset log-returns can be very noisy, so that inferring the distribution with the required precision can turn out to be a difficult task. As will be shown in Section 4, our method outperforms the simplistic optimization: volatility is reduced and the average performance is either enhanced or remains the same. Finally, we are also able to report a significantly reduced instability of the selection

† Although in Bouchaud, Potters, and Aguilar (1997), a 'free-utility' $F = U - TS_q$ is defined, which is then minimized by slightly modifying the usual Markowitz formalism.

process with regard to the risk-aversion parameter.

The paper is organized as follows: in Section 2 we propose our method, in which MEU optimization is replaced by an expectation with respect to (a transformation of) the utility function. In Section 3 implementation issues are discussed, while results from numerical simulations are represented in Section 4. Conclusions and final remarks are presented in Section 5.

2. The Probabilistic Utility Approach

Let us denote with $\boldsymbol{\alpha}$ the set of parameters (weights) that identify a portfolio, and with U the set of parameters that characterize our utility function, like risk aversion, investment horizon, etc. Let us assume furthermore that the expected utility is computed with respect to a distribution characterized by parameters, for instance expected excess returns, which we will denote collectively by $\boldsymbol{\Phi}$. The estimated expected utility may then be written as a function

$$u = u(\boldsymbol{\alpha}, U, \boldsymbol{\Phi}).$$

In classical asset allocation theory, the prescription would be to select the portfolio $\boldsymbol{\alpha}$ that maximizes this expression, independent of the magnitude of the estimation error associated with $\boldsymbol{\Phi}$. In our framework, we consider the expected utility as proportional to the logarithm of a probability measure in portfolio space, fully conditional on U and $\boldsymbol{\Phi}$:

$$\boldsymbol{\alpha} \sim P(\boldsymbol{\alpha}|U, \boldsymbol{\Phi}) = Z^{-1}(\nu, U, \boldsymbol{\Phi}) \exp\Big(\nu u(\boldsymbol{\alpha}, U, \boldsymbol{\Phi})\Big). \quad (1)$$

The symbol $\sim$ has to be interpreted as "$\boldsymbol{\alpha}$ is distributed according to", and $Z(\nu, U, \boldsymbol{\Phi})$ is a normalization constant defined by

$$Z(\nu, U, \boldsymbol{\Phi}) = \int_{D(\boldsymbol{\alpha})} [d\boldsymbol{\alpha}] \exp\Big(\nu u(\boldsymbol{\alpha}, U, \boldsymbol{\Phi})\Big),$$

where $D(\boldsymbol{\alpha})$ stands for the integration domain of $\boldsymbol{\alpha}$. The recommended portfolio $\overline{\boldsymbol{\alpha}}$, given U and $\boldsymbol{\Phi}$, is defined as the expected or mean value of $\boldsymbol{\alpha}$:

$$\overline{\boldsymbol{\alpha}}(U, \boldsymbol{\Phi}) = Z^{-1}(\nu, U, \boldsymbol{\Phi}) \int_{D(\boldsymbol{\alpha})} [d\boldsymbol{\alpha}] \; \boldsymbol{\alpha} \; \exp\Big(\nu u(\boldsymbol{\alpha}, U, \boldsymbol{\Phi})\Big). \quad (2)$$

This definition is based on the following reasoning: If we consider portfolio selection as a decision making problem in which one tries to pick the optimal portfolio, no amount of data is sufficient to exclude any portfolio with certainty. Specifically, the data by which we estimate the parameters $\boldsymbol{\Phi}$ merely allows us to assign to *each* possible portfolio an individual but in any case non-zero probability $P(\boldsymbol{\alpha})$ of being the optimal one, i.e. the one that maximizes expected utility for the actual value of $\boldsymbol{\Phi}$.

Let us assume we had calculated (e.g. by Bayesian analysis) this probability. Which portfolio should then be selected? One possibility would be to choose according to the maximum probability criterion, but that could be quite risky, since it corresponds to an all or nothing bet, where 'a miss is as good as a mile'. More robust is the mean value of $\boldsymbol{\alpha}$, which - as is well known - is the preferred estimator whenever one wants to minimize the expected square error. This idea is formally expressed in Eq.(2).

Deriving the probability distribution $P(\boldsymbol{\alpha})$ analytically would clearly be very tedious if not impossible. However, it is actually sufficient to postulate two reasonable and simple properties (apart from continuity) in order to derive an approximation: (i), the probability distribution should faithfully represent the uncertainty about the parameters $\mathbf{\Phi}$, and (ii), within the limits of (i) it should concentrate as much of the probability measure as possible on portfolios with a high value of estimated expected utility (of course we expect $P(\boldsymbol{\alpha}) > P(\boldsymbol{\alpha}')$ if $u(\boldsymbol{\alpha}) > u(\boldsymbol{\alpha}')$).

Property (i) becomes a natural assumption when thinking of $P(\boldsymbol{\alpha})$ as representing the knowledge we have at disposition for solving the selection problem. Without any data and hence any knowledge, the distribution necessarily needs to be flat (all portfolios are equally likely), while on the other extreme, if we had an infinite amount of data, our knowledge would be complete and the only rational choice would be the portfolio that maximizes expected utility, hence implying a Delta distribution. In intermediate cases, the knowledge represented by $P(\boldsymbol{\alpha})$ must correspond to the knowledge we have acquired through, e.g., a recorded time series of length N. This request can be formalized through the introduction of the continuous form of the Shannon entropy H (Shannon and Weaver, 1949) of a given probability distribution

$$H[P(\boldsymbol{\alpha})] = -k \int_{D(\boldsymbol{\alpha})} [d\boldsymbol{\alpha}]\; P(\boldsymbol{\alpha}) \ln\left(P(\boldsymbol{\alpha})\right) , \tag{3}$$

defined up to a constant k, which we set to one. As can be verified easily, H attains its maximum value $\ln \Gamma$ only for the uniform distribution $P(\boldsymbol{\alpha}) = 1/\Gamma$, and approaches minus infinity for an increasingly sharply peaked probability density. If we denote by $I(D_N)$ the information about $\mathbf{\Phi}$ we can extract from a time series of length N, conservation of information can be expressed as $H[P(\boldsymbol{\alpha})] = \ln \Gamma - I(D_N)$, with, obviously, $dH/dN < 0$.

Property (ii) only means that the overall limited information we have should be used in the most efficient way to identify good candidate portfolios, i.e. those with a high expected utility. This concentration of the probability on high-utility portfolios can be measured by the expression $\int P(\boldsymbol{\alpha})u(\boldsymbol{\alpha})$, which, in order to increase as a whole, needs to concentrate more and more of the probability on those portfolios with a high value for $u(\boldsymbol{\alpha})$. Evidently, it merely represents the average utility $<u>$, and will thus be bounded from below and above by u_{min} and u_{max}, respectively.

Our prescription then becomes: find $P(\boldsymbol{\alpha})$ such that $<u>$ is maximized, subject to the entropy constraint $H[P(\boldsymbol{\alpha})] = S$. For the sake of simplicity, we illustrate the solution of this standard problem only for the discrete case, using the technique of Lagrange multipliers:

$$\frac{\partial}{\partial P(\boldsymbol{\alpha}_j)} \left(\sum_i P(\boldsymbol{\alpha}_i) u(\boldsymbol{\alpha}_i) + \lambda \left(S - \sum_i P(\boldsymbol{\alpha}_i) \ln\left(P(\boldsymbol{\alpha}_i)\right) \right) + \mu \left(1 - \sum_i P(\boldsymbol{\alpha}_i) \right) \right) \stackrel{!}{=} 0 \, ,$$

leads directly to

$$P(\boldsymbol{\alpha}_j) = \exp\left(\frac{u(\boldsymbol{\alpha}_j) - \mu}{\lambda} - 1 \right) \equiv Z(\nu)\; \exp\left(\nu u(\boldsymbol{\alpha}_j)\right) ,$$

where $Z(\nu)$ is determined by normalization, and ν, which incorporates the entropy constraint of Eq.(3), is fixed by the implicit equation

$$H\left[Z(\nu) \exp\left(\nu u(\boldsymbol{\alpha}_j)\right)\right] = \nu \frac{\partial}{\partial \nu} \ln\left(Z(\nu)\right) - \ln Z(\nu) \stackrel{!}{=} S \quad \Rightarrow \quad \nu = \nu(S) \, . \tag{4}$$

Although we cannot solve for $\nu(S)$ explicitly, from $\partial/\partial u\; P(u(\boldsymbol{\alpha})) > 0$ follows $\nu(S) \geq 0$, where equality holds only if S attains its maximum value, i.e. in the case of complete uncertainty. In fact, for $\nu = 0$, Eq.(4) correctly reads $H = -\ln Z(0) = \ln \Gamma = S_{max} \stackrel{!}{=} S$ (Γ denoting the total measure of the domain of $\boldsymbol{\alpha}$). How does $\nu(S)$ change if we acquire more information, and hence have a decreasing S? Implicit differentiation of Eq.(4) yields

$$\frac{d\nu}{dS} = -\frac{1}{\nu\, \sigma_u^2} < 0 \, ,$$

which is strictly negative, and hence ν must be an increasing function of the available information D_N, tending to infinity for $N \to \infty$. One possible interpretation of this parameter can be given by recalling that it represents the inverse of the Lagrange multiplier λ, and hence is related to the (reciprocal of a) shadow price of the reliability of the data, adjusted for the investor's aversion to ambiguity: in the extreme cases, i.e. whenever the available data has either complete or no informative value, the parameter ν attaches a weight of, respectively, zero or infinity to the entropy constraint. However, in intermediate cases, ν may be adjusted to reflect the investor's unease with the given uncertainty, i.e. how averse she is to ambiguity.

A second interpretation of ν can be given in terms of the aforementioned analogy to statistical physics, where the corresponding variable of the Boltzmann distribution, β, is inversely proportional to the system's temperature, and hence determines the relative weighting between entropy and energy. In our case, energy is represented by (minus) utility, and ν functions indeed as an adjustable weight-factor of the utility contribution $\exp(\nu u)$: for large ν, it will dominate and hence lead to portfolios with an expected utility close to the possible maximum, which will then coincide with traditional MEU in the limit of $\nu \to \infty$ (since all the measure is concentrated where $u(\boldsymbol{\alpha}, U, \boldsymbol{\Phi})$ has its maximum), while for small values of ν the utility function becomes increasingly smoothed out and hence tends to weigh the different portfolios more evenly. Utility considerations become irrelevant in the limit case of $\nu \to 0$, for which Eq.(2) leads to the uniform portfolio with an equal allocation for all assets, and hence - as expected from the picture of statistical physics - to the portfolio with the highest possible entropy.

For a rational investor, the uniform portfolio would in fact be preferred if no information on future asset performance was available. Conversely, any portfolio configuration that maximizes expected utility would be acceptable if we had perfect knowledge about the underlying return distributions. It follows that ν is in fact a measure of how much information we have or, said differently, how we assess the reliability of our estimation.

With a set of stationary historical data, we can bootstrap from the data, build scenarios, and compute unbiased estimates of the expected utility; or, if we have a data model, and we believe that historical data are drawn from some distribution, we can use the time series to estimate the distribution parameters. In both situations our confidence on the value of the expected utility will in some way be linked positively to the length of the available time series data set. As discussed earlier, we know for the asymptotic behavior of ν

$$\begin{aligned} \lim_{N \to 0} \nu(N) &= 0 \\ \lim_{N \to \infty} \nu(N) &= \infty \, , \end{aligned}$$

where N is the size of the data set. The simplest such form is

$$\nu = \rho N^{\gamma}, \tag{5}$$

with ρ and γ constant and strictly greater than zero. All of the simulations carried out in this paper will have $\rho = 1$ and $\gamma = 1$. The limit $\rho \to \infty$ will recover the standard maximization approach. It is obviously interesting to ask whether a more sophisticated relation between ν and N, e.g. $\sqrt{N}$ or $\nu = \frac{N}{J}$, with J representing the number of considered assets, could lead to a better algorithm, in particular to one that makes a more effective use of the available information. However, since already the simple relation $\nu = N$ leads to considerable improvements, these questions shall be addressed in future research.

Finally, since we choose to insist on a distribution to describe the portfolio, it is natural to identify the error associated with the estimate of $\overline{\boldsymbol{\alpha}}$ with its standard deviation in terms of the same distribution. An unbiased estimate of the standard deviation can be computed, at no extra cost, while computing the integral in Eq.(2).

3. Implementation

3.1. Markov Chain Monte Carlo Integration

The integral in Eq.(2) is easily carried out by Markov Chain Monte Carlo (MCMC) integration (Gilks, Richardson, and Spiegelhalter, 1996; Johannes and Polson, 2002). We do not know how to sample directly from this specific probability density function, and we are forced to devise a Markov chain that relaxes to the desired distribution. After several experiments with variations of the Metropolis-Hastings algorithm, we resorted to an implementation of the Hybrid Monte Carlo method. Details of the algorithm can be found in Appendix A. Once relaxation has been achieved, we can run the Markov chain for few more steps in order to perform measurements. Relaxation or thermalization is not a trivial issue, but a thorough discussion of the problems involved would bring us too far from the subject of this paper. We choose to defer this discussion to a forthcoming paper focusing on the implementation of the numerical integration scheme. As for the probability distribution of the asset returns, for the sake of simplicity we assume normality, though we want to point out that our approach works with all types of distributions.

3.2. The Utility Function

The selection of a 'good' utility function is not the subject of this paper, as the proposed method is meant to address problems that are present regardless of the specific utility chosen. Indeed, in order to demonstrate that our results do not depend on a particular utility function, we will show numerical results for two different types of such functions. The first is the mean-variance utility first introduced by Markowitz, a standard in financial economics and a long-lasting benchmark. In addition, we consider a utility function that belongs to the category of the so-called downside risk measures (Bouchaud and Potters, 2000), popularized by the application of concepts like Value at Risk in the risk management industry. In our investigation, we will employ a utility function of this type recently proposed by Consiglio, Cocco, and Zenios (2001), which we will refer to as CZC. Through this type of utility function, risk is measured in terms of the expected amount by which a specified target is not met: this might better describe how the investor perceives risk, as documented by results from behavioral finance (Hirshleifer, 2001), and is more in line with the recent ALM practice. The CZC function allows for the emergence of standard deviation terms that make the optimal portfolio dependent on the investment horizon - thus introducing a time scaling property. Details of the CZC function are presented in Appendix B.

4. Assessment of Performance: PU vs. MEU

In this section, the performance of our proposed PU method is analyzed under various aspects. First, we compare finite sample errors of the PU and the MEU approach by computing the mean realized utility as a function of the available sample length. The question is: to which extent do fluctuations in estimated average returns and covariances (whose variability stems from the inevitable finite sample estimation errors) affect the portfolio composition and, as a consequence, the achieved utility? Next, we compare out-of-sample performances in terms of realized portfolio returns. Finally, we assess the sensitivity towards the ubiquitous risk aversion parameter: what is the behavior of the portfolio composition in response to a variation of λ, and how is this behavior affected if we slightly modify our data sample? Historical data used to infer realistic distribution parameters consists of eight monthly indexes, covering the period from January 1988 to January 2002 (called *full sample* hereafter). A list of the employed titles is shown in Table 1.

Table 1. List of assets employed as basis of simulations. The data covers the period from Jan 1988 to Jan 2002. The used acronyms have the following meaning: MSCI = Morgan Stanley Capital Index, JPM = JP Morgan Index, ML = Merrill Lynch Index. Data source: Datastream. Data types: Price Index for equities, Total Return Index for bonds. All the samples are in local currency, unadjusted for inflation. The index titles refer to the Datastream mnemonics.

Assets	Description
MSNAMR	MSCI North America Equity
MSPACF	MSCI Pacific Equity
MSEROP	MSCI Europe Equity
JPMUSU	JPM US Government Bond
JPMJPU	JPM Japan Government Bond
JPMEIL	JPM Europe Government Bond
MLHMAU	ML US Corporate High Yield
JPEC3M	JPM Euro Cash

4.1. Robustness of Portfolio Selection under Finite Sample Effects

To compare the ability of PU and MEU to select 'correct' portfolios under realistic finite sample conditions, we designed the following test: taking distribution parameters as inferred from our empirical data sample, we generated a very long artificial time series with the exact same distribution parameters. To simulate the finite sample behavior associated with a time series of length L, we divided the long series in many non-overlapping pieces (around 1000) of length L. For each such piece, we re-estimated the distribution parameters, and let PU and MEU identify their recommended portfolio accordingly. With the portfolio obtained in this way, we computed the resulting 'true' value of the (expected) utility function of these portfolios by referring to the original distribution, and compared it to the maximum achievable value of utility, as obtained by seeking the extremum with respect to the original distribution. To measure the deviation from the optimal utility, we computed the quantity

$$\Delta U(L) = \frac{U_T - U(L)}{U_T},$$

where U_T stands for the true utility, and $U(L)$ for the finite sample utility obtained after re-estimating the distribution parameters using a time series of length L. We show mean and standard deviation of this quantity for both mean-variance utility and CZC utility in Fig.1 and Fig.2, respectively.

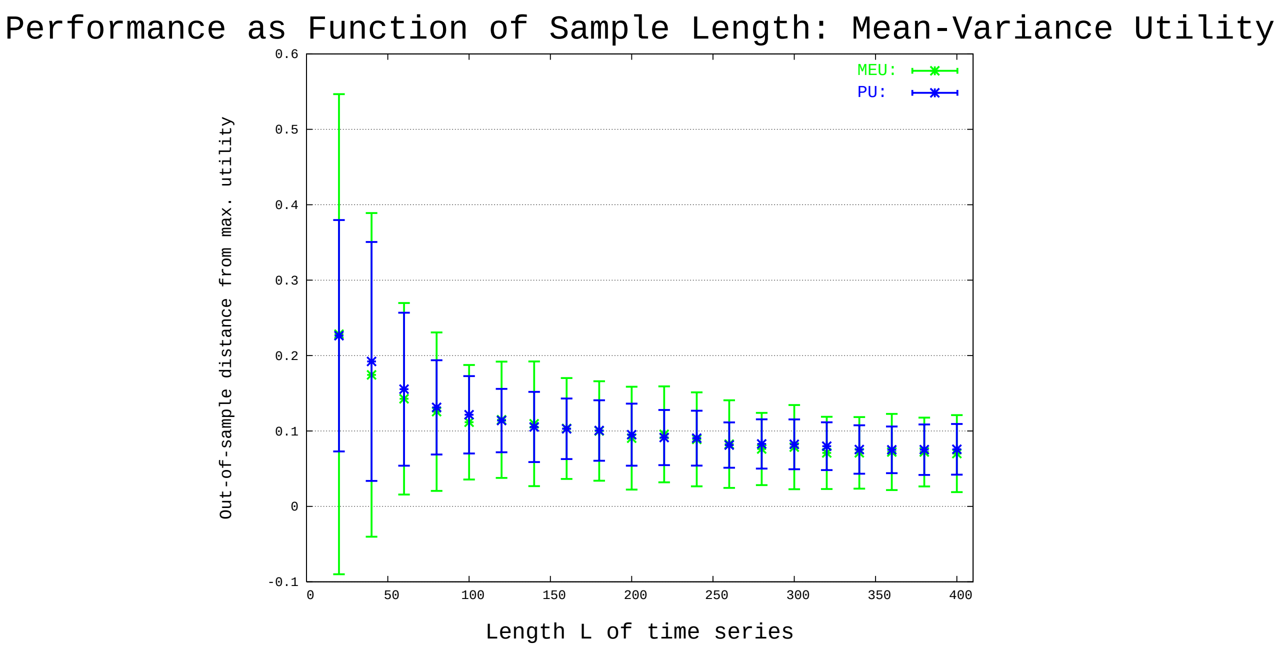

Figure 1. Average distance from the maximal achievable utility U_T in finite sample simulations for the mean-variance utility.

In either cases we would expect

$$\lim_{L\to\infty} \Delta U(L) = 0,$$

both for the mean value and the standard deviation. In fact, within the range of L shown in the two plots a convergent behavior can indeed be observed. With regard to the performance, the PU approach, while achieving the same average utility as MEU, clearly reduces the standard deviation for the mean-variance utility function, which means that it is less prone to finite sample effects. In case of the CZC utility function, the results are not sufficiently clear to discriminate between PU and MEU in terms of performance.

4.2. Out-Of-Sample Portfolio Return

We undertook a second simulation in order to compare the out-of-sample portfolio performance of PU and MEU: to this end, we took the full data set (Jan 1988-Dec 2002), and divided it into two parts. The first part comprised the 156 months from

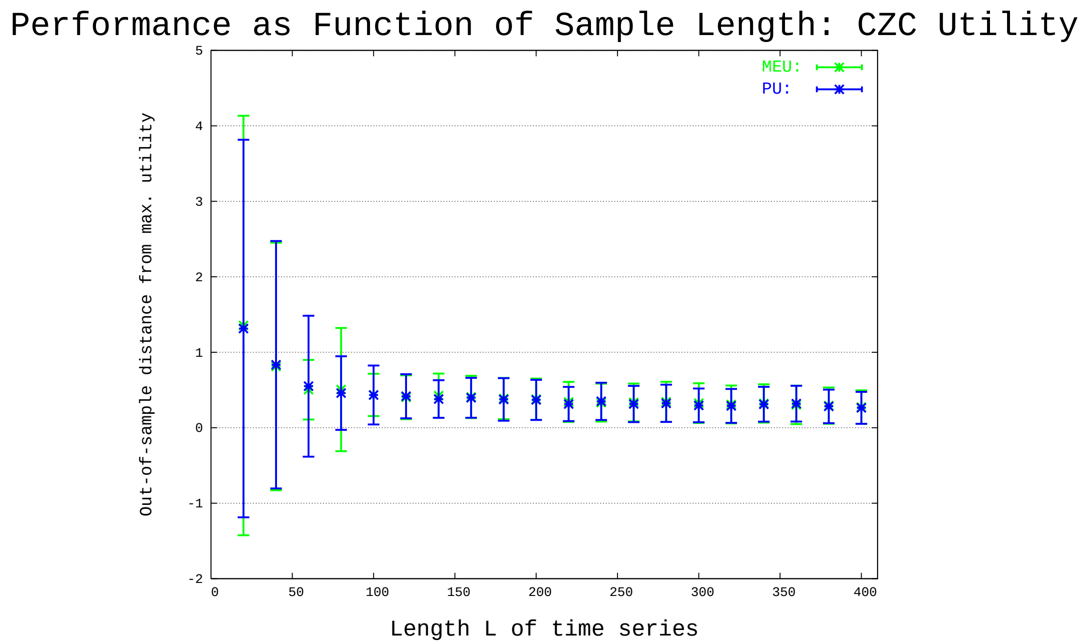

Figure 2. Average distance from the maximal achievable utility U_T in finite sample simulations for the CZC utility.

January 1988 through December 2000, and the second the remaining 24 months. For each data set we determined the mean returns, $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$, and the covariance matrices, $\boldsymbol{\Omega_1}$ and $\boldsymbol{\Omega_2}$. Based on the parameter set stemming from the longer series, $\boldsymbol{\mu_1}$ and $\boldsymbol{\Omega_1}$, we generated an artificial time series of length 156, and used it to re-compute distribution parameters and the associated optimal portfolio $\boldsymbol{\alpha}$. With the remaining parameter set, $\boldsymbol{\mu_2}$ and $\boldsymbol{\Omega_2}$, we generated a two year long scenario, and we tracked the performance of the portfolio over these 24 months. This process was repeated about 1000 times, and each time we measured the mean and the standard deviation of the portfolio return. This out-of-sample simulation was carried out for both the mean-variance and the CZC utility function, and yielded the results shown in Fig.3 and Fig.4. As can be seen, in case of the mean-variance utility, the PU approach consistently outperforms MEU, both in terms of mean return and variance. For the CZC utility, the average return remains the same, but the PU still shows a lower volatility, i.e. standard deviation.

Overall, the central values of both the finite sample and out-of-sample simulation show remarkable agreement between the MEU and the PU approach, but it has become quite clear that the PU approach, in terms of risk as measured by volatility, is always at least as good as the MEU approach, and mostly better. This result confirms the general idea behind our approach: a cautious attitude towards expected return distribution parameters, that in a consistent manner incorporates the inevitable parameter uncertainty induced by finite data samples, turns out to be a statistically winning approach when used to select portfolios in a realistic setting.

4.3. Sensitivity to Risk Aversion Parameter λ

As a third empirical investigation, we examine the algorithm's stability for one specific portfolio profile. As previously discussed, all expected-utility based procedures suffer

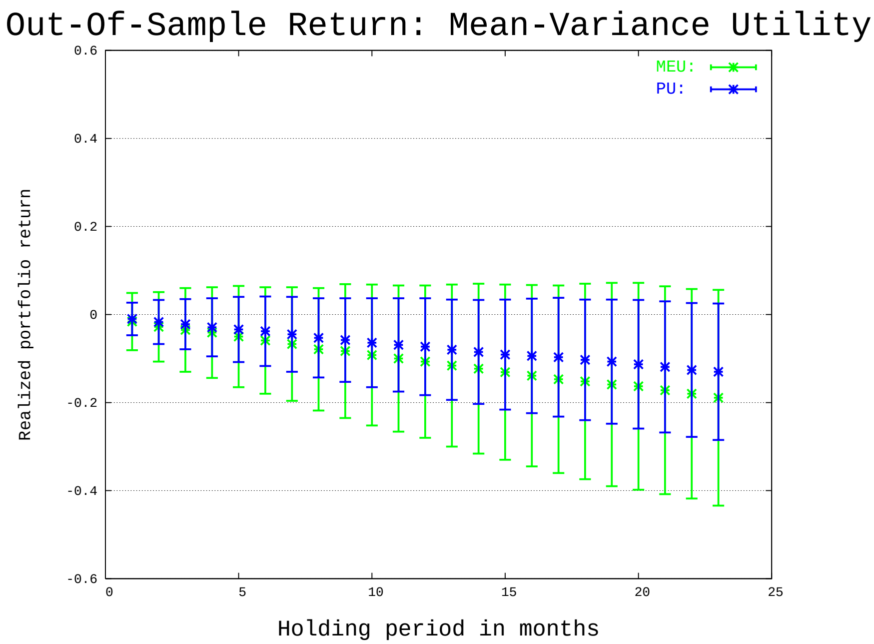

Figure 3. Average out-of-sample portfolio return and standard deviation for the mean-variance utility function.

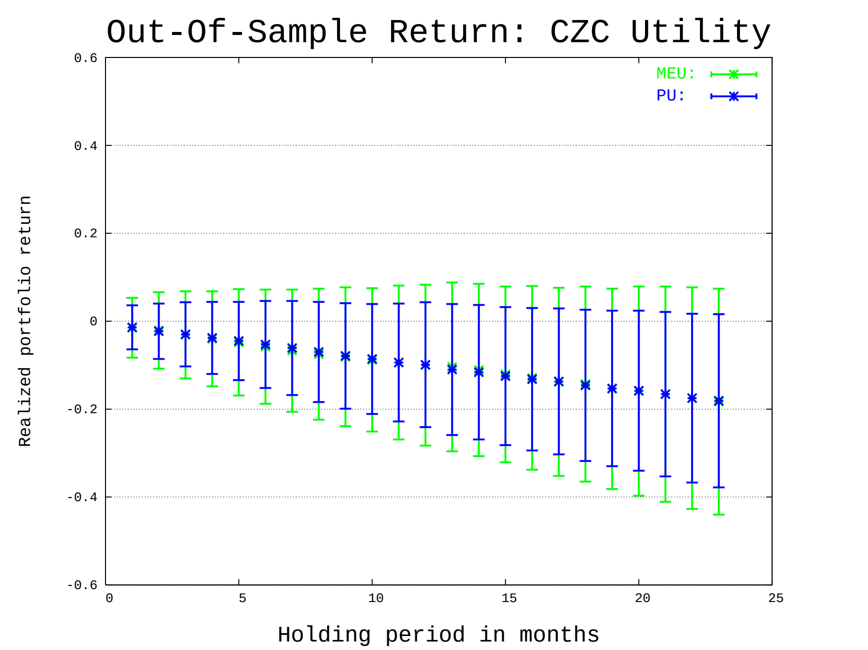

Figure 4. Average out-of-sample portfolio return and standard deviation for the CZC utility function.

from the presence of a risk aversion parameter, dimensionless and un-settable from theory. As a measure of instability, it then seems natural to compare the sensitivity to λ for both the MEU optimization prescription and the PU method. Specifically, we examine the behavior of a diversification indicator, i.e., an indicator that measures the degree of concentration within a portfolio, thereby allowing us to identify the range of the parameter λ that mostly affects the portfolio composition. The simplest of such a - as Bouchaud, Potters, and Aguilar (1997) put it - "entropy like measure" is the quantity

$$Y = \sum_{j=1}^{J} \alpha_j^2 \quad , \tag{6}$$

which ranges from $\frac{1}{J}$ (J denoting the number of assets, $J = 8$ here), when the portfolio is totally diversified (evenly spread), to $Y = 1$ in case of complete concentration on one asset.

In Fig.5 and Fig.6 we report the behavior of Y with respect to λ for two different data samples, the full sample and a slightly restricted 1988-2000 one. Looking at the MEU graph in Fig.5, the behavior appears very erratic and the significant (sensitive) range of λ is restricted to a relatively small interval, meaning that small changes in λ can produce large modifications in the portfolio composition. This is certainly not reassuring, given that λ is only loosely tied to the investor's risk aversion, and its setting can by no means be effected with certainty. What strikes even more is what happens when we look at the results for a different data sample, in this case shortened by the last two years of observations: the curve decidedly shifts to the right, and consequently the relevant range of λ does the same, leading to dangerous risk profile mis-identifications, and forcing one to re-calibrate (with all the associated uncertainties) the values of λ potentially each time new historical observations are added to the sample.

Coming now to the PU model, for which results are shown in Fig.6, the diversification indicator displays a very different pattern: it indicates a more conservative overall behavior, with values closer to the lower bound of $Y = \frac{1}{8}$. It never concentrates all the weights on a single asset, not even for the risk-neutral ($\lambda = 1$) case. Most importantly, Y exhibits a smooth progression. This reduces the danger of mis-setting λ, and its over-sensitivity to the chosen sample; for a different data sample, in fact, the curve remains rather similar, leaving the significant range of λ unaffected.

5. Conclusion And Final Remarks

The purpose of this work was to address and improve some of the well known weaknesses of portfolio selection by maximizing expected utility. In particular, we wanted to overcome the excessive sensitivity to external parameters and to account for informational uncertainty. We achieved this by employing a different interpretation of the utility function, namely as a measure of probability in 'portfolio space'. We have tested the proposed method against the benchmark of traditional expected utility maximization, using artificial data to simulate finite sample and out-of-sample behavior, and have been able to illustrate the superior performance of our method for two different types of utility functions. We also managed to significantly improve the intrinsic instability with respect to the risk-aversion parameter (lack of continuity) that plagues all maximization approaches.

As for future lines of research, it might be interesting to relax the normality assumption, for instance by modeling the data with a mixture of Gaussian distributions. Additionally, our theory included a parameter, namely ν from Eq.(5), that could easily be determined in its asymptotic behavior, but whose value within an intermediate range was not clear. Acting as a smoothing parameter within our probabilistic utility

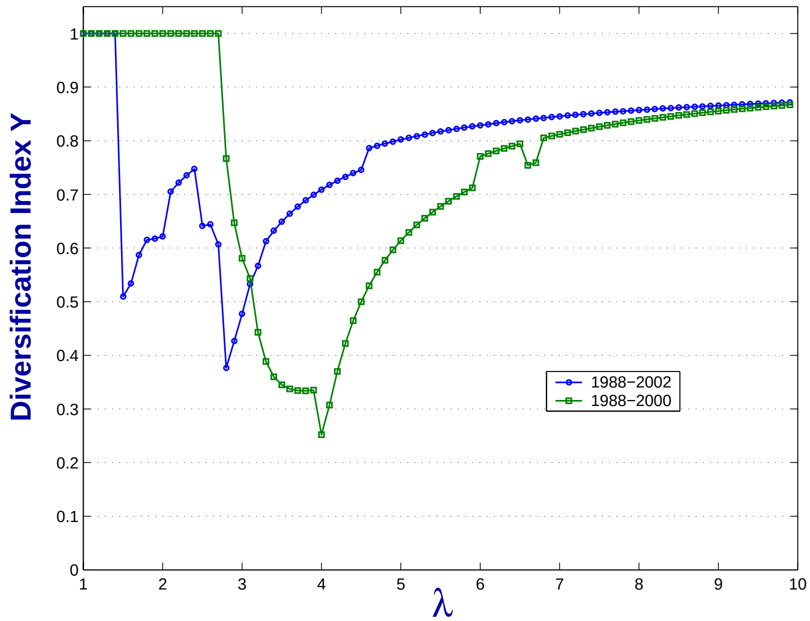

Figure 5. Portfolio diversification as a function of risk aversion λ and data sample: MEU selection procedure.

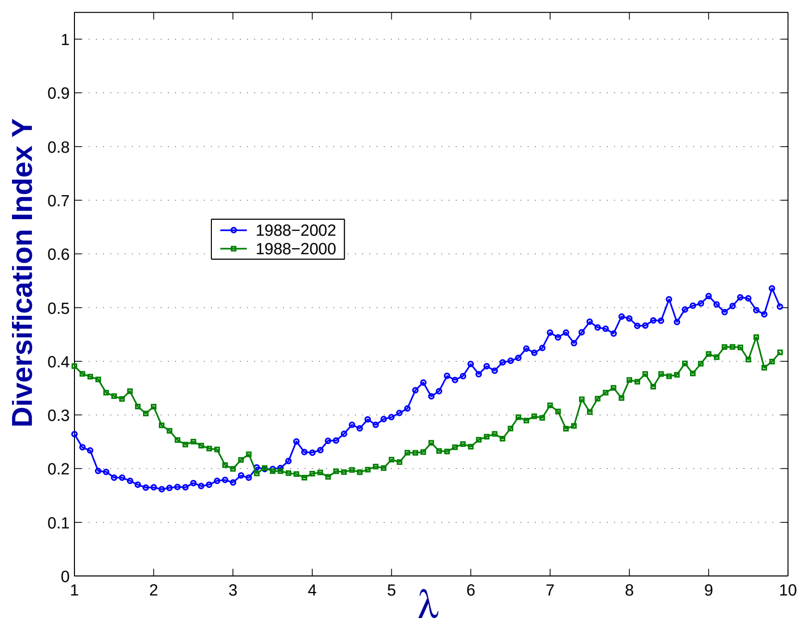

Figure 6. Portfolio diversification as a function of risk aversion λ and data sample: PU selection procedure.

approach and incorporating the investors aversion to ambiguity, it certainly can have an important influence on the overall performance of the proposed method and deserves closer investigation.

Acknowledgments

The authors thankfully acknowledge the exchange of opinions with T. Kennedy and A. Pellisetto for useful insight on the MCMC integration.

Appendix A. Hamiltonian MCMC for Portfolio Optimization

The algorithm we used in this paper is well known in the physics literature by the name of hybrid Monte-Carlo (Duane *et al* , 1987). In this appendix, we give a rough outline of the general concept employed. If we denote with $\boldsymbol{\alpha}$ the portfolio's weights, then expected values of functions of $\boldsymbol{\alpha}$ will not be affected if we replace $\exp\left[U(\boldsymbol{\alpha})\right]$ with the distribution

$$G(\boldsymbol{\alpha}, \boldsymbol{\pi}) \sim \exp\left(U(\mathbf{p}) - \frac{\boldsymbol{\pi}\cdot\boldsymbol{\pi}}{2}\right), \tag{A.1}$$

where $\boldsymbol{\pi}$ represents the conjugate momentum. Then, starting from a pair $(\boldsymbol{\alpha}_n, \boldsymbol{\pi}_n)$, the updating rule is defined as follows:

Step 1 Sample $\boldsymbol{\eta}$ as a normal variable with mean zero and unit variance.

Step 2 For a time interval T, integrate Hamilton's equations

$$\frac{d\pi_i}{dt} = -\frac{\partial U}{\partial \pi_i} \tag{A.2}$$

$$\frac{d\alpha_i}{dt} = \pi_i, \tag{A.3}$$

together with the boundary conditions

$$\boldsymbol{\pi}(0) = \boldsymbol{\eta}, \tag{A.4}$$

$$\mathbf{p}(0) = \mathbf{p}_n; \tag{A.5}$$

Step 3 With probability

$$\beta = \min\Big(1, \exp(G(\mathbf{p}(T), \pi(T)) - G(\mathbf{p}_n, \eta)\Big), \tag{A.6}$$

set $\mathbf{p}_{n+1} = \mathbf{p}(T)$, and with probability $1-\beta$ set $\mathbf{p}_{n+1} = \mathbf{p}_n$.

The clever idea behind this algorithm rests on the observation that, if step 2 is carried out exactly, Hamilton's equation enforce $G(\mathbf{p}(T), \pi(T)) = G(\mathbf{p}_n, \eta)$, and every proposed configuration is accepted. In general the acceptance rate will be controlled by the numerical accuracy of our integration scheme. A good scheme is the interleaved leap frog that, for finite integration step Δt and fixed trajectory length (that is, scaling the number of steps in the integration scheme with $1/\Delta t$), is guaranteed to have errors of the order $O(\Delta t^2)$.

Appendix B. The CZC Utility Function

The utility function is drawn from the article of Consiglio, Cocco, and Zenios (2001):

$$E\Big[u(\boldsymbol{\alpha}, L, \lambda, T)\Big] = \sum_{n=1}^{N_T} \Delta t[E(U(n\Delta t)) - \lambda E(D(n\Delta t))], \tag{B.1}$$

where $U(n\Delta t)$ and $D(n\Delta t)$ are the upside and downside, respectively, of the portfolio return at time $n\Delta t$ against a fixed target return L, and λ is a weight indicating the investor's risk aversion. The time horizon T is built out of N_T intermediate time intervals Δt, such that $T = N_T \Delta t$ is a sequence of N_T values for $\omega(n\Delta t),\ n = 1, \ldots, N_T$. The model takes a "target-all time" view, and the allocation is such that staying as close to the target return trajectory at all times is the primary concern.

Modeling the distribution for the single period log-return $\boldsymbol{\omega}$ with a normal $N(\mathbf{m}, \boldsymbol{\Omega})$,

$$\boldsymbol{\omega} \sim \exp\left(-\frac{(\boldsymbol{\omega}-\mathbf{m})^T\boldsymbol{\Omega}^{-1}(\boldsymbol{\omega}-\mathbf{m})}{2}\right)\frac{1}{\sqrt{(2\pi)^J|\boldsymbol{\Omega}|}}, \tag{B.2}$$

we obtain by straightforward (but tedious) Gaussian integration an explicit expression for the utility function

$$E\Big[u(\boldsymbol{\alpha}, L, \lambda, T)\Big] = \sum_{n=1}^{N_T} \Delta t\Big[n\Delta t M f_2(n\Delta t) - \sqrt{n\Delta t} S f_1(n\Delta t)\Big], \tag{B.3}$$

with

$$f_1(t) = (\lambda - 1)\frac{e^{-t\eta^2}}{\sqrt{2\pi}} \tag{B.4}$$

$$f_2(t) = \frac{1+\lambda}{2} - \frac{\lambda-1}{2}\mathrm{erfc}(\sqrt{t}\eta) \tag{B.5}$$

$$\eta = \frac{M}{2S} \tag{B.6}$$

$$M = \boldsymbol{\alpha}^T \cdot \mathbf{m} - L \tag{B.7}$$

$$S^2 = \boldsymbol{\alpha}^T \cdot \boldsymbol{\Omega} \cdot \boldsymbol{\alpha}. \tag{B.8}$$

As expected, the explicit solution for this specific form of utility is a function of the portfolio mean, variance and - most importantly - standard deviation. It incorporates a time-dependent competing effect between average return and standard deviation, since they have different time scaling properties: the standard deviation's contribution is proportional to $\lambda - 1$, and can be traced back to the imperfect cancellation of positive and negative deviations from the ideal line. We thus obtain the desired dependency of the optimal portfolio on the chosen time horizon: the longer the horizon (*ceteris paribus*), the more aggressive the optimal allocation.

References

Ait-Sahalia, Y. and M. Brandt. (2001). "Variable Selection for Portfolio Choice." *Journal of Finance* 56, 1297-1351.

Anderson, E., L. Hansen, and T. Sargent. (1999). "Robustness detection and the price of risk." Working paper, University of Chicago.

Balduzzi, P. and L. Liu. (2000). "Parameter Uncertainty and International Investment." Working Paper, Boston College.

Barberis, N. (2000). "Investing for the Long Run when Returns are Predictable." *Journal of Finance* 55, 225-264.

Barry, C.B. (1974). "Portfolio Analysis under Uncertain Means, Variances and Covariances." *Journal of Finance* 29, 515-522.

Bawa, V., S. Brown, and R. Klein. (1979). "Estimation Risk and Optimal Portfolio Choice." North Holland, Amsterdam.

Black, F. and R. Litterman. (1992). "Global Portfolio Optimization." *Financial Analyst Journal* 9, 28-43.

Bouchaud, J.P. and M. Potters. (2000). "Theory of Financial Risks." Cambridge University Press.

Bouchaud, J.P., M. Potters, and J.P. Aguilar. (1997). "Missing Information and Asset Allocation." Unpublished paper: available for download at www.science-finance.fr.

Britten-Jones, M. (2002). "Portfolio optimization and Bayesian regression." Unpublished working paper, London Business School.

Chen, Z. and L. Epstein. (2002). "Ambiguity, Risk, and Asset Returns in Continuous Time." *Econometrica* 70, 1403-43.

Chopra, V. and W.T. Zembia. (1993). "The Effect of Errors in Means, Variances and Covariances on Optimal Portfolio Choice." *Journal of Portfolio Management* Winter 1993, 6-11.

Consiglio, A., F. Cocco, and S.A. Zenios. (2001). "The value of integrative risk management for insurance products with guarantees." *Journal of Risk Finance* 2, 6-16.

Duane, S., A.D. Kennedy, B.J. Pendleton, and D. Roweth. (1987). "Hybrid Monte Carlo." *Phys.Lett.* B195, 216-222.

Ellsberg, D. (1961). "Risk, Ambiguity and the Savage Axioms." *Quarterly Journal of Economics* 75, 643-669.

Frost, P. and J. Savarino. (1986). "An Empirical Bayes Approach to Efficient Portfolio Selection." *Journal of Financial and Quantitative Analysis* 21, 293-305.

Garlappi, L., R. Uppal, and T. Wang. (2004). "Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach." Conference paper, 2004 UBC Finance Conference.

Gilks, W.R., S. Richardson, and D.J. Spiegelhalter. (1996). "Markov Chain Monte Carlo in Practice." Chapman & Hall.

Hirshleifer, D. (2001). "Investor Psychology and Asset Pricing." *Journal of Finance* 56, 1533-1597.

ter Horst, J.R., F.A. de Roon, and B.J.M. Weker. (2002). "Incorporating Estimation Risk in Portfolio Choice." Working paper, Tilburg University.

Jobson, J.D. and R. Korkie. (1980). "Estimation for Markowitz Efficient Portfolios." *Journal of the American Statistical Association* 75, 544-554.

Johannes, M. and N. Polson. (2002). "MCMC Methods in Financial Econometrics." In: Handbook of Financial Econometrics, edited by Y. Ait-Sahalia, forthcoming.

Jorion, P. (1985). "International Portfolio Diversification with Estimation Risk." *Journal of Business* 58, 259-278.

Jorion, P. (1992). "Portfolio Optimization in Practice." *Financial Analysts Journal* January/February, 68-74.

Kirkpatrick, S., C.D. Gelatt, and M.P. Vecchi. (1983). "Optimization by Simulated Annealing." *Science* 220 (4598), 671-680.

Markowitz, H. (1952). "Portfolio Selection." *Journal of Finance* 7, 77-91.

Markowitz, H. and Usmen, N. (2003). "Resampled Frontiers vs Diffuse Bayes: An Experiment." *Journal of Investment Management* 1 (4).

Michaud, R.O. (1989). "The Markowitz Optimization Enigma: Is Optimized Optimal?" *Financial Analyst Journal* 45, 31-42.

Michaud, R.O. (1998). "Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation." Oxford University Press.

Pastor, L. (2000). " Portfolio selection and asset pricing models." *Journal of Finance* 55(1), 179-223.

Piotrowski, E.W. and J. Sladkowski. (2001). "The Thermodynamics of Portfolios." *Acta Physica Polonica B* 32, 597-604.

Polson, N.G. and Tew, B.V. (2000). "Bayesian Portfolio Selection: An Empirical Analysis of the S&P 500 Index 1970-1996." *Journal of Business and Economic Statistics* 18 (2), 164-173.

Scherer, B. (2002). "Portfolio Resampling: Review and Critique." *Financial Analysts Journal* 58, 98-109.

Shannon, C.E. and W. Weaver. (1949). "The Mathematical Theory of Communication" University of Illinois Press, Urbana, IL.

Uppal, R. and T. Wang. (2003). "Model Misspecification and Under-diversification." *Journal of Finance* 58, 2465-2486.