Journal Of Investment Management, Vol. 6, No. 1, (2008), pp. 1–17

www.joim.com

BAYES VS. RESAMPLING: A REMATCH*

Campbell R. Harvey [a], *John C. Liechty*[b]
and Merrill W. Liechty[c]

We replay an investment game that compares the performance of a player using Bayesian methods for determining portfolio weights with a player that uses the Monte Carlo based resampling approach advocated in Michaud (Efficient Asset Management. Boston: Harvard Business School, 1998). Markowitz and Usmen (Journal of Investment Management **1***(4), 9–25, 2003), showed that the Michaud player always won. However, in the original experiment, the Bayes player was handicapped because the algorithm that was used to evaluate the predictive distribution of the portfolio provided only a rough approximation. We level the playing field by allowing the Bayes player to use a more standard algorithm. Our results sharply contrast with those of the original game. The final part of our paper proposes a new investment game that is much more relevant for the average investor—a one-period ahead asset allocation. For this game, the Bayes player always wins.*

1 Introduction

In the classic mean–variance portfolio selection problem, the investor is presumed to have complete knowledge of the inputs, i.e., exact knowledge of expected returns, variances, and covariances. Most often this assumption is considered innocuous, ignored, or perhaps not fully understood by asset managers. There have been many advances in dealing with parameter uncertainty.[1] In an important recent article, Markowitz and Usmen (2003) report the results of an experiment which compares the performance of two competing methods for

*The authors appreciate the detailed comments of the referees.

[a]Campbell R. Harvey is J. Paul Sticht Professor of International Business, Fuqua School of Business, Duke University, Durham NC 27708 and Research Associate, National Bureau of Economic Research, Cambridge, MA 02138, e-mail: *cam.harvey@duke.edu*, Tel.: (919) 660–7768.

[b]John C. Liechty is Associate Professor, Department of Marketing, Smeal College of Business, and Department of Statistics, Pennsylvania State University, University Park, PA 16803, e-mail: *jcl12@psu.edu*, Tel.: (814) 865-0621.

[c]Merrill W. Liechty is Assistant Professor, Department of Decision Sciences, LeBow College of Business, Drexel University, Philadelphia, PA 19104, e-mail: *merrill@drexel.edu*, Tel.: (215) 895–2459.

determining optimal portfolio weights, where each method explicitly accommodates the uncertainty in the parameter estimates. We revisit this comparison.

In the first approach, portfolio weights are found by integrating out these uncertainties using Bayesian methods. In the second approach, a competing set of weights are obtained using the Resampled Efficient Frontiers™ method found in Michaud (1998).[2] Markowitz and Usmen (2003) conduct an experiment using synthetic data and find that the resampled weights perform better than the weights implied by a Bayesian method.

We replay the same investment simulation game, with two main differences from the way the game was played by Markowitz and Usmen (2003). First, while they use uniform prior distributions for the mean and covariance, we use a hierarchical Bayesian model with diffuse, conjugate prior distributions that mimic uniform prior distributions. This facilitates the second and more important difference, i.e., the use of a Markov Chain Monte Carlo (MCMC) algorithm, as opposed to an Importance Sampling algorithm. While both approaches are used in the literature, the MCMC algorithm is almost always preferred in part because of well documented problems that can arise with regards to the variance of Importance Sampling approximations, see Robert and Casella (1998) and Bernardo and Smith (1994). In addition, Markowitz and Usmen (2003) probably used too few samples to approximate these high dimensional integrals.

Under the MCMC inference method, we find that the results from the investment game sharply differ from the original experiment. In our rematch, there are many cases where weights from the Bayesian method perform better than weights from the resampling method, using the same performance criteria as the initial experiment. In this rematch, we found that the Bayesian method does better at low levels of risk aversion and the resampling method does better at high levels of risk aversion. We provide the economic intuition for the role of risk aversion.

We also consider a second competition, a one-step ahead version of the investment problem, which is more relevant from the investor's perspective. In this competition, additional returns are generated and one-step ahead portfolio returns are calculated for all of the different historical data sets. We find that the Bayes approach dominates the resampled efficient frontier approach when the data are drawn from a distribution that is consistent with the data in each history, i.e., drawn from the predictive distribution conditional on each history. Our results lead us to conjecture that the resampled frontier approach has practical merit when the future returns are not consistent with the historical returns (e.g., when the underlying statistical model has been misspecified or the data is drawn from a distribution other than the predictive distribution) or when the investor has a very long investment horizon, as implied by the criteria used in the initial competition, and is not very risk averse. Later we explore why risk aversion impacts the success of these approaches for both competitions.

The remainder of the paper is organized as follows. In Section 2, we review the simulation competition and the set of utility functions that are considered. We briefly review the equivalent Resampled Efficient Frontier™ approach that we use in Section 3, and we discuss our modification of the specification of the Bayesian investor in Section 4. In Section 5, we explore the one-step ahead investment problem and conclude with a discussion of the results and potential reasons for the differences between the original experiment and the new experiment. We also discuss settings where the resampled frontier approach may offer a more robust solution to the portfolio allocation problem. Some concluding remarks are offered in Section 6.

2 The investment game

Following Markowitz and Usmen (2003), we conduct a simulated investment "game" with two players and a referee. The referee generates 10 "true" parameter sets for a multivariate normal density. Each "true" parameter set summarizes the behavior for a group of eight asset returns in the sense that for each "truth" the monthly percent returns for these eight assets are assumed to be *i.i.d.* normal with means, variances, and covariance given by the corresponding "true" set of parameters. As in the original experiment, we mimic the asset allocation task discussed in Michaud (1998), where the assets that are being considered are a collection of six equity indices (Canada, France, Germany, Japan, the United Kingdom, and the United States) and two bond indices (United States Treasury bond and a Eurodollar bond).

The referee starts with an original set of parameters, which are the Maximum Likelihood Estimates (MLE) of the mean and covariance for these eight assets based on their monthly percent returns over the 216 months from January 1978 to December 1995; see Chapter 2 of Michaud (1998) for the exact values. The referee then generates 10 sets of perturbed parameters, or "truths," by generating 216 draws from a multivariate normal density using the original parameters and a new random seed; the perturbed or "true" parameters are the MLE estimates from each corresponding sets of draws. Using each of the 10 truths, the referee then generates 100 histories (each with 216 simulated observations), which form the basis of the games, see Figure 1 for a summary. To be more explicit, let μ_{OP} and Σ_{OP} be the mean and covariance matrix representing the original parameters. The referee creates the ith set of "true" parameters (μ_{Ti}, Σ_{Ti}), by generating $r_{in} \sim N(\mu_{OP}, \Sigma_{OP})$, for $n = 1, \ldots, 216$ and letting

$$\mu_{Ti} = \frac{1}{216}\sum_{n=1}^{216} r_{in} \quad \text{and}$$

$$\Sigma_{Ti} = \frac{1}{216}\sum_{n=1}^{216}(r_{in} - \mu_{Ti})(r_{in} - \mu_{Ti})'. \quad (1)$$

For each (μ_{Ti}, Σ_{Ti}), the referee generates 100 histories, where the kth history for the ith set of "true"

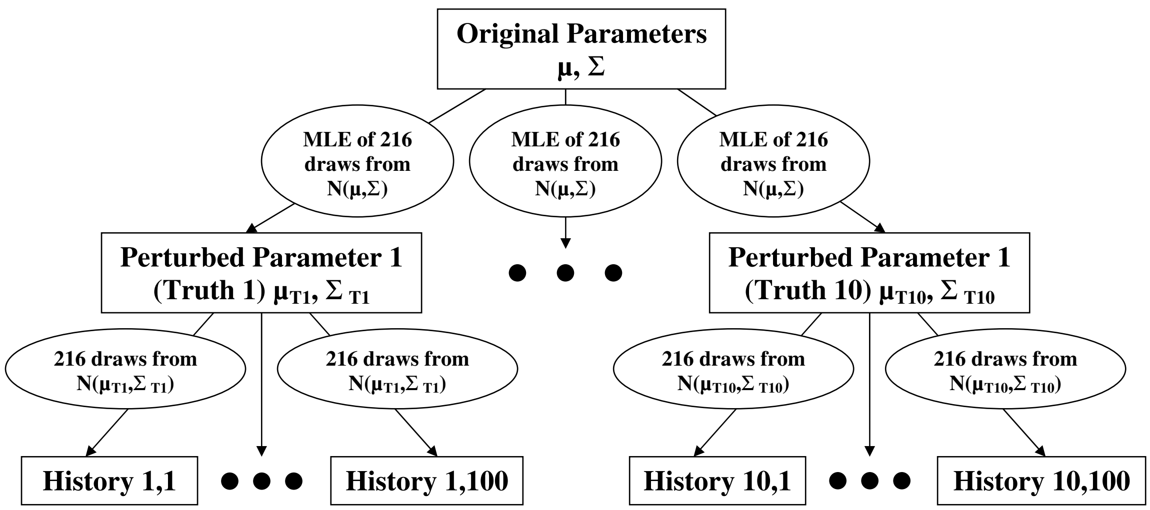

Figure 1 Graphical representation of the histories and truths used both in this paper and in Markowitz and Usmen (2003).

parameters is as follows:

$$H_{ik} = \{r_{ikn} : r_{ikn} \sim N(\mu_{Ti}, \Sigma_{Ti}),\ n = 1, \ldots, 216\}. \quad (2)$$

The investment game is played as follows. The referee gives each player a simulated history and the players tell the referee the portfolio weights that they believe will maximize the expected utility under three different utility functions; the utility functions are given by

$$u_\lambda(\omega, r_{n+1}) = \omega' r_{n+1} - \lambda(\omega'(r_{n+1} - E[r_{n+1}|H]))^2,\quad \lambda = \{0.5, 1.0, 2.0\}, \quad (3)$$

where $E[r_{n+1}|H]$ is the predictive mean given history H, ω are the portfolio weights, r_{n+1} are the predictive returns (e.g., the distribution of returns for the next month, month 217, conditional on the observed returns, months 1–216) and λ reflects risk aversion and takes three different values. In addition to returning the optimal portfolio weights, the players also tell the referee their own estimate of the expected utility using their optimal weights. The referee compares each players' weights by calculating the players' expected utility using the true parameter values in place of the predictive mean and covariance. For each of the 100 histories, the player with the weights that result in a higher expected utility, using the true parameter values, is determined to have won.

As shown in Markowitz and Usmen (2003) and Harvey *et al.* (2006), the expected utility of (3), given a specific history H is a function of predictive moments (mean and covariance). The predictive mean is equal to the posterior mean, which will be very close to the MLE estimate of μ (the historical average returns). The predictive covariance matrix, however, is composed of *two* different summaries of uncertainty: (1) the posterior mean of Σ, which will be very close to the MLE estimate of Σ (the historical covariance matrix of the returns) and (2) the posterior mean of the *covariance* of μ, which reflects our uncertainty with respect to the mean returns μ given the data that has been observed. So the posterior mean of Σ captures both the covariance of the return as well as summarizing the inherent uncertainty in estimating the average return or the uncertainty with respect to μ. (See Appendix A.1 for more details).

The Bayes player finds the weights, ω_B, which maximize the expected utility with respect to the predictive moments for each history, while the Michaud player finds the weights, ω_M, using the resampling scheme. The referee compares both sets of weights assuming the true parameters used to generate the history, μ_T, Σ_T, are the predictive mean and covariance, or

$$Eu_{1\lambda}(\omega|H) = \omega'\mu_T - \lambda\omega'\Sigma_T\omega. \quad (4)$$

Before describing the details of how the Michaud and Bayes player obtain their portfolio weights, it is worth observing that the referee and two players are not using consistent frameworks. The Bayes player uses a utility function based on predictive returns and the Michaud player uses a utility function based on parameter estimates. The referee evaluates performance based on the "true" parameters, which ignores the contribution to Σ that comes from the inherent uncertainty regarding the "true" average return. If this extra variance is missing, the estimate for the portfolio variance will be lower than they should be, which will lead to suboptimal portfolio allocation.

3 The resampling player

As in the original experiment, we consider the basic version of the resampled frontier approach. Markowitz and Usmen (2003) form the resampled frontier by calculating the resampled weights for an appropriate grid of portfolio standard deviations.

In our experiment, we implement the alternative, but equivalent approach, of constructing a resampled frontier by calculating portfolio weights for a range of linear utility functions, see Michaud (1998, p. 66) for a discussion. The advantage of using this version of the resampling approach is that the resampled frontier only needs to be calculated for values of λ that are of interest to the referee and there is no need to calculate the frontier for a grid of portfolio standard deviations.

For each history H_{ik}, the Michaud player uses the corresponding standard parameter estimates μ_{ik} and Σ_{ik} and generates 500 additional histories, which we will denote as resampled histories H^R_{ikm}, by drawing 216 *i.i.d.* normal draws using μ_{ik} and Σ_{ik}. For each resampled history, a discrete approximation of the efficient frontier is calculated, or a set of 101 weights describing the efficient frontier based on the standard estimates from each resampled history are calculated. Next, the set of weights which gives the highest utility value is selected for each resampled history.

The Michaud weights are equal to the average of the best weights for each resampled history.[3] Stated more explicitly, in the original experiment, the Michaud player calculates an efficient frontier for each resampled history,[4] H^R_{ikm}, and for a discrete grid of 101 equally spaced portfolio standard deviations $(\sigma_{ikm,\min}, \sigma_{ikm,1}, \ldots, \sigma_{ikm,99}, \sigma_{ikm,\max})$, they calculate a set of weights $W_{ikm} = (\omega_{ikm,\min}, \omega_{ikm,1}, \ldots, \omega_{ikm,99}, \omega_{ikm,\max})$ that maximize the portfolio expected return for the corresponding standard deviation; these weights form a discrete estimate of the efficient frontier for the corresponding resampled history, H^R_{ikm},—one draw from the resampled frontier. In the original experiment, for each value of λ, the Michaud player selects the weights as follows:

$$\omega_{\lambda ikm} = \arg\max\{\omega'\mu^R_{ikm} - \lambda\omega'\Sigma^R_{ikm}\omega : \omega \in W_{ikm}\}. \qquad (5)$$

Then the "resampled" weights, $\overline{\omega}_{\lambda ik}$, reported by the Michaud player for the kth history associated with the ith "truth," are the average optimal weights over the corresponding resampled histories, or

$$\overline{\omega}_{\lambda ik} = \frac{1}{500}\sum_m \omega_{\lambda ikm}. \qquad (6)$$

Alternatively, the maximized weights for each μ^R_{ikm} and Σ^R_{ik} and for each λ can be obtained directly by solving the standard quadratic programming problem of

$$\omega_{\lambda ikm} = \arg\max\left\{\omega'\mu^R_{ik} - \lambda\omega'\Sigma^R_{ik}\omega : 0 \le \omega, \sum_p \omega_p = 1\right\}. \qquad (7)$$

Finding the optimal weights for each λ in this fashion has two advantages, first it requires fewer optimizations (3 compared to 101) and it obtains a set of weights for each resampled history which is at least as good as the weights using the original experiment.

4 The Bayes player

In our rematch, the Bayes player will use a different approach for calculating the expected utility. In both the original and current game, the Bayes player assumes that asset returns are *i.i.d.* and follow a normal distribution with mean μ and covariance matrix Σ; see Appendix A.2 for an exact specification of the model.

We modify the Bayes player in two ways: we alter the prior distribution and we use the MCMC algorithm. In the original game, the Bayes player assumes a uniform prior distribution on μ and Σ, where the distributions are truncated to include all "reasonable" parameter values. This allows equal probability, *a priori*, over the range of possible parameters, reflecting a diffuse prior distribution. In our current experiment, we assume diffuse conjugate prior distributions for μ and Σ or

$$\mu \sim N(\overline{\mu}, \tau^2 I), \tag{8}$$

and

$$\Sigma^{-1} \sim \text{Wishart}(\nu, SS), \tag{9}$$

where $\bar{\mu} = 0$, $\tau^2 = 100$, $SS = I$, $\nu = 5$, and I is an identity matrix. The intuition is as follows. The prior distribution for a model parameter, such as μ, is considered to be conjugate, if the resulting distribution, conditional on the data and the remaining parameters is the same type of distribution as the prior distribution, (e.g., if the prior for μ is a Normal distribution, then the distribution for μ, conditional on Σ and the data is also a Normal distribution). By picking appropriate values for τ^2, ν, and SS, these distributions can be such that they are diffuse, and have no impact on the final parameter estimates. Both the uniform prior and the diffuse conjugate prior are equivalent with regards to the information they bring to the analysis. However, the conjugate prior makes it easier to do the MCMC calculations. While the calculations could still be done with a uniform prior, they would be more cumbersome. Hence the reason for choosing the conjugate prior is purely computational. See Bernardo and Smith (1994) for a more complete discussion of prior distributions. See Appendix A.2 for a discussion of how both model specifications are similarly diffuse.

The most important difference between the original experiment and our experiment is the use of the MCMC algorithm to estimate the expected utility; see Gilks *et al.* (1998) for a discussion of the MCMC algorithm. In the original experiment, the Bayes player used an Importance Sampling scheme, based on 500 draws from a proposal distribution to approximate the expected value of (3) (see Appendix A.1 for more details); while the Importance Sampler has attractive computational properties, it can result in integral estimates with unbounded or extremely large variances, which is problematic because the weights for points with high posterior probability can be large, leading to infrequent selection from the proposal distribution; see Robert and Casella (1998) and Bernardo and Smith (1994).

To contrast the two inference approaches, the MCMC algorithm generates samples from the predictive density and uses these draws to approximate the expected utility integral, where the Importance Sampling scheme generates draws from an alternative density and reweights these draws in order to approximate the integral with respect to the predictive density. In other words this MCMC algorithm samples directly from the predictive density, whereas the Importance Sampler obtains samples from the predictive density in a round about way. An important difference between our implementation of the MCMC algorithm and Markowitz and Usmen's (2003) implementation of the Importance Sampler has to do with the number of samples that were used. In the original experiment, they used only 500 samples, whereas we use 25,000 draws from the predictive density. The relatively small number of draws, with respect to the dimension of the space being integrated over (44 dimensions), is one potential reason for the differences in the two experiments.[5]

5 Results of the rematch

The results using the MCMC algorithm for inference and using the original performance criteria (i.e., evaluating each weight using the proposed "true" parameter values as the predictive mean and covariance as detailed in (4)), are markedly different from the results reported from the original experiment. In the original experiment, the Michaud player won for every "truth" and for every value of λ in that the portfolio weights reported by the

Michaud player gave a larger average utility over the 100 histories as evaluated by the referee. In the new experiment, the Bayes player wins in 7 out of the 10 histories when $\lambda = 0.5$, and the Michaud player wins in 8 out of 10 histories and in 6 out of 10 histories when $\lambda = 1$ and 2, respectively; see Table 1 for a summary of the results.[6]

The main difference between the original experiment and the current experiment comes from the choice of inference used by the Bayes player (i.e., the difference between using the Importance Sampling and the MCMC algorithms to approximate the expected utility). As a result, investors should use caution when determining which approach to use for selecting an optimal portfolio in practice.

In the original game, the referee chooses a criteria that handicaps the Bayes player and that reflects an investment strategy that is much different from the investment strategies pursued in practice. Specifically, the players select an optimal set of weights based on a history and then the referee uses a criteria that is not consistent with that history (he/she evaluates the weights using the "true" mean and covariance, which are different from the predictive mean and covariance associated with the history). This would be reasonable, if the investor does not expect future returns to match historical returns. Since this is not the case in the original game, the Bayes player is handicapped as he/she is operating under the assumption that the future returns distribution will match the past returns distribution, and it is interesting that even with this handicap the Bayes player performs at a comparable level to the Michaud player.

From an investment perspective, the referee's criteria implicitly assumes that each player is going to take their derived weights and hold a portfolio based on these weights until all uncertainty from the parameter estimates is gone. Stated differently, the referee is determining the performance of a set of portfolio weights by assuming that each player will hold their respective portfolio forever (or at a minimum for the rest of the player's life). It is inconceivable that a real world investor will never adjust their portfolio.

5.1 One period ahead asset allocation

In order to explore the performance of these two approaches in a setting that is more relevant to an investor with a shorter investment horizon and where the Bayes player is not handicapped, we conducted a new experiment. In this out of sample asset allocation game, the referee assumes that the investor will only hold the portfolio for one period and where the referee draws returns that are consistent with the history that has been presented to the player (i.e., the return is drawn from the predictive density, given the history).

To be more precise, for each history H_{ik}, both players calculate weights as described in Sections 3 and 4. The referee draws 100 asset returns for the next period ($t = 217$) from the predictive distribution

$$r_{ik217} \sim N\left(\mu_{|H_{ik}}, \Sigma_{|H_{ik}}\right),$$

and using the Michaud players weights, $\omega_{MH_{ik}}$, and Bayes players weights, $\omega_{BH_{ik}}$, the referee calculates the portfolio return for each draw

$$R_{Mik} = \omega'_{MH_{ik}} r_{ik217} \quad \text{and} \quad R_{Bik} = \omega'_{BH_{ik}} r_{ik217}. \tag{10}$$

The referee calculates the players utility for each "truth" (for each i), by calculating the mean and variance of the one-step ahead portfolio returns and putting that into the quadratic utility function, or given a λ and estimates of the portfolio mean and variance calculated in the usual way

$$\mu_{\text{port}_i} = \frac{1}{10{,}000} \sum_{kt} R_{ikt} \quad \text{and}$$

$$\Sigma_{\text{port}_i} = \frac{1}{10{,}000} \sum_{kt} (R_{ikt} - \mu_{\text{port}_i})^2. \tag{11}$$

Table 1 Player's choice of portfolio.

λ:	0.5	0.5	0.5	0.5	1	1	1	1	2	2	2	2
Player:	Bayes	Bayes	Michaud	Michaud	Bayes	Bayes	Michaud	Michaud	Bayes	Bayes	Michaud	Michaud
Eval. by:	player	referee	player	referee	player	referee	player	referee	player	referee	player	referee
Panel A: EU averaged over 100 histories, for each of 10 truths												
Truth 1:	0.02036	0.01916	0.02019	0.01902	0.01800	0.01684	0.01840	0.01714	0.01403	0.01293	0.01508	0.01366
Truth 2:	0.01214	0.01046	0.01186	0.01032	0.01006	0.00854	0.01024	0.00866	0.00700	0.00585	0.00755	0.00599
Truth 3:	0.00739	0.00487	0.00710	0.00486	0.00617	0.00418	0.00635	0.00421	0.00464	0.00333	0.00515	0.00326
Truth 4:	0.01623	0.01362	0.01599	0.01360	0.01457	0.01251	0.01480	0.01246	0.01194	0.01036	0.01279	0.01072
Truth 5:	0.01167	0.01035	0.01143	0.01014	0.00926	0.00790	0.00966	0.00827	0.00598	0.00483	0.00694	0.00551
Truth 6:	0.00830	0.00632	0.00804	0.00622	0.00663	0.00494	0.00683	0.00500	0.00431	0.00315	0.00486	0.00312
Truth 7:	0.00763	0.00481	0.00736	0.00491	0.00597	0.00374	0.00613	0.00372	0.00400	0.00252	0.00438	0.00222
Truth 8:	0.00966	0.00692	0.00935	0.00693	0.00812	0.00593	0.00827	0.00594	0.00578	0.00419	0.00638	0.00427
Truth 9:	0.01021	0.00699	0.00993	0.00699	0.00841	0.00578	0.00865	0.00580	0.00613	0.00430	0.00665	0.00407
Truth 10:	0.00751	0.00512	0.00724	0.00504	0.00606	0.00402	0.00646	0.00450	0.00413	0.00271	0.00513	0.00349
Mean	0.01111	0.00886	0.01085	0.00880	0.00932	0.00744	0.00958	0.00757	0.00679	0.00542	0.00749	0.00563
Std. Dev.	0.00425	0.00464	0.00429	0.00459	0.00401	0.00425	0.00405	0.00428	0.00344	0.00349	0.00358	0.00370
No. times better		7		3		2		8		4		6
Panel B: Number of "wins" out of 100 histories, for each of 10 truths												
Truth 1:		73		27		20		80		6		94
Truth 2:		70		30		41		59		24		76
Truth 3:		54		46		43		57		46		54
Truth 4:		61		39		61		39		26		74
Truth 5:		75		25		14		86		8		92
Truth 6:		66		34		43		57		42		58
Truth 7:		40		60		51		49		79		21
Truth 8:		52		48		53		47		32		68
Truth 9:		53		47		40		60		71		29
Truth 10:		63		37		15		85		9		91
Avg. No. wins		60.7		39.3		38.1		61.9		34.3		65.7
No. times better		9		1		3		7		2		8

Table 1 (*Continued*)

λ:	0.5	0.5	0.5	0.5	1	1	1	1	2	2	2	2
Player:	Bayes	Bayes	Michaud	Michaud	Bayes	Bayes	Michaud	Michaud	Bayes	Bayes	Michaud	Michaud
Eval. by:	player	referee	player	referee	player	referee	player	referee	player	referee	player	referee
Panel C: Standard deviation of EU over 100 histories, for each of 10 truths												
Truth 1:	0.00418	0.00124	0.00419	0.00119	0.00399	0.00108	0.00405	0.00108	0.00362	0.00081	0.00373	0.00083
Truth 2:	0.00338	0.00126	0.00339	0.00108	0.00315	0.00108	0.00320	0.00102	0.00247	0.00059	0.00267	0.00074
Truth 3:	0.00281	0.00081	0.00283	0.00077	0.00246	0.00057	0.00257	0.00080	0.00193	0.00042	0.00212	0.00076
Truth 4:	0.00322	0.00106	0.00323	0.00089	0.00280	0.00080	0.00293	0.00074	0.00258	0.00064	0.00256	0.00076
Truth 5:	0.00439	0.00182	0.00438	0.00171	0.00391	0.00126	0.00401	0.00131	0.00307	0.00074	0.00333	0.00088
Truth 6:	0.00332	0.00097	0.00332	0.00081	0.00302	0.00075	0.00310	0.00070	0.00240	0.00039	0.00254	0.00053
Truth 7:	0.00319	0.00082	0.00320	0.00066	0.00286	0.00056	0.00295	0.00057	0.00219	0.00038	0.00235	0.00061
Truth 8:	0.00308	0.00051	0.00306	0.00045	0.00287	0.00049	0.00290	0.00059	0.00251	0.00029	0.00259	0.00071
Truth 9:	0.00375	0.00087	0.00379	0.00074	0.00340	0.00057	0.00351	0.00068	0.00274	0.00039	0.00293	0.00069
Truth 10:	0.00262	0.00138	0.00259	0.00124	0.00241	0.00092	0.00246	0.00120	0.00183	0.00063	0.00222	0.00090
Avg. Std. Dev.		0.00107		0.00095		0.00081		0.00087		0.00053		0.00074
No. times better		0		10		6		4		10		0

Each player's utility is given by

$$E[u_\lambda] = \mu_{\text{port}} - \lambda \Sigma_{\text{port}}. \tag{12}$$

In the one-step ahead asset allocation game, using the draws from the "predictive" density, the Bayes player wins for all of the "truths"; the Bayes player has a higher expected utility for 10 out of 10 "truths" for all of the utility functions, see Table 2 for a summary.

The results of the experiment show that the Bayesian approach will outperform and potentially dominate the resampling approach, depending on the perspective that the investor wants to adopt. If the investor assumes that the distribution of future returns will match the distribution of past returns and the investor has a short investment time horizon, then they should avoid the resampling approach; alternatively, if there is some ambiguity about the distribution of past returns and the investor has a very long time horizon, the resampling approach has some advantages.

5.2 *Interpreting the relative performances: Bayes vs. Resampling*

In replaying the original game, it appears that there may be a pattern in the performance of the two approaches. The difference in the average expected utility between the two approaches across all of the histories is influenced by the investor's risk aversion (or λ). In the original game, as the investor's risk aversion increases (λ gets bigger), the resampling approach performs better on average, see Figure 2.

In contrast, although the Bayes approach dominates in the new game, the level of dominance increases as the investor becomes more risk averse.

Table 2 EU calculated using one-step ahead draws from predictive distributions.

λ: Player: Eval. by:	0.5 Bayes referee	0.5 Michaud referee	1 Bayes referee	1 Michaud referee	2 Bayes referee	2 Michaud referee
Truth 1:	0.01929	0.01926	0.01612	0.01597	0.01150	0.01054
Truth 2:	0.00968	0.00959	0.00742	0.00703	0.00439	0.00348
Truth 3:	0.00595	0.00578	0.00457	0.00425	0.00255	0.00168
Truth 4:	0.01491	0.01472	0.01278	0.01248	0.00962	0.00869
Truth 5:	0.01078	0.01067	0.00778	0.00764	0.00394	0.00326
Truth 6:	0.00667	0.00660	0.00472	0.00455	0.00212	0.00153
Truth 7:	0.00628	0.00615	0.00430	0.00416	0.00176	0.00129
Truth 8:	0.00826	0.00790	0.00639	0.00610	0.00372	0.00290
Truth 9:	0.00852	0.00821	0.00632	0.00608	0.00351	0.00284
Truth 10:	0.00639	0.00610	0.00445	0.00408	0.00201	0.00050
Grand mean	0.00967	0.00950	0.00749	0.00723	0.00451	0.00367
Std. Dev.	0.00434	0.00438	0.00395	0.00397	0.00333	0.00330
No. times better	10	0	10	0	10	0

Note: This table shows averages of expected utility calculated from one-step ahead draws from predictive distributions for each player. Specifically, for risk aversion $\lambda = 0.5, 1.0$, and 2.0, as indicated by the row labeled "Lambda," and for each player, and as indicated by the row labeled "Player."

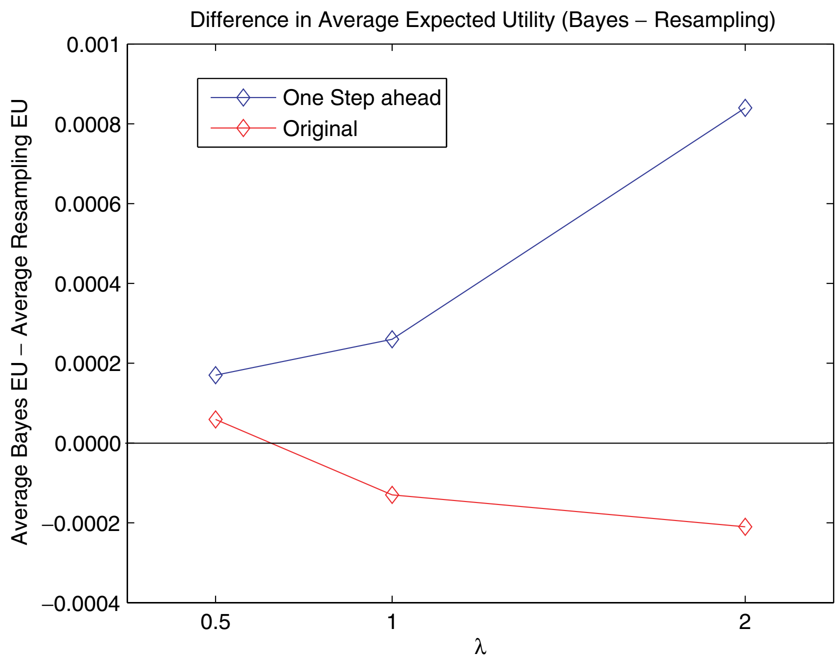

Figure 2 Difference in the average expected utility (Bayes–Resampling) as a function or risk aversion (or λ). Results are for the original game and for the new game (or one-step ahead game).

The influence of risk aversion on the difference in performances is much larger for the new game than for the original game and it is in the opposite direction. The economic reason for these differences can be understood by investigating how the average portfolio mean and the average portfolio variance (the two components of the quadratic utility function) change as a function of λ. As the investor becomes more risk averse, the average portfolio mean and variance, for both approaches across both games, decreases as we would expect. However, the decrease in the average variance and the average mean for the Bayes approach is larger (particularly the decrease in the average variance) when compared with the resampling approach, see Table 3. This gives us the key insight that while the resampling approach tends to result in a larger average portfolio mean, this comes at the expense of a larger average portfolio variance, and this difference in the average variance increases dramatically as an investor's risk aversion increases.

The two games can be framed in terms of the investment time-frame: a long-term investor in the original game and a short-term investor in the new or one-step ahead game. While investors in both games use the same amount of information (216 data points) to find their weights, the referee uses very different criteria for each game. In the original game the referee evaluates weights using the "true" parameters, which implies that the investor is holding the portfolio for a very long time. By using the "true" parameters, the referee is ignoring the extra variance that comes from the uncertainty about the estimates of the average return. As a result the average portfolio variance for the original or long-term game are smaller than the average portfolio variance from the second or one-step ahead game, again see Table 3. The most striking difference between the two games is in terms of the average portfolio variance. For both the Bayes and resampling approach, the average portfolio variance is roughly twice as large for the new game when compared with the

Table 3 Summary of Average Portfolio Mean and Variance, by Game and approach.

		$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$
Original game	Bayes: Average portfolio mean	0.0102	0.0094	0.0081
	Resampling Ave. portfolio mean	0.0100	0.0098	0.0091
	Bayes: Average portfolio variance	0.0026	0.0020	0.0013
	Resampling Ave. portfolio variance	0.0025	0.0022	0.0017
One-Step ahead game	Bayes: Average portfolio mean	0.0124	0.0113	0.0095
	Resampling Ave. portfolio mean	0.0120	0.0117	0.0109
	Bayes: Average portfolio variance	0.0054	0.0038	0.0025
	Resampling Ave. portfolio variance	0.0049	0.0045	0.0036

Note: The striking difference between the two games is the difference in the average portfolio variance. For both Bayes and resampling approach, the average portfolio variance is roughly twice as large for the new game (one-step ahead) when compared with the original game, again see Table 1. In the original game, the smaller variances from the Bayes strategy does not compensate for the relative change in mean, which results in the resampling strategy performing better. However, for the new game, the average portfolio variances are roughly doubled while the average portfolio means are only marginally better (on the order of 1.2 times larger). As a result the naturally smaller portfolio variance of the Bayes strategy becomes increasingly important. We feel that the new game is the proper way to assess the performance of both of these methods as both strategies are calibrated conditional on the historical data and they have to account for both uncertainty due to unexplained randomness and uncertainty due to our inability to predict the mean.

original game. In the original game, the smaller variances from the Bayes strategy does not compensate for the relative change in mean, which results in the resampling strategy performing slightly better as λ increases. However, for the new game the average portfolio variances are roughly doubled while the average portfolio means are only marginally better (on the order of 1.2 times larger). As a result the naturally smaller portfolio variance of the Bayes strategy becomes increasingly important and leads to the dominate performance of the Bayes approach.

All investors will have to deal with making asset allocation decisions in the face of both the unexplained uncertainty and uncertainty about the mean. In addition, the dramatic difference in the average portfolio variance obtained by using the Bayes approach demonstrates the value of the Bayes approach as the uncertainty facing the investor increases and/or as the investor becomes more averse to risk.

6 Conclusion

Our paper replays the investment simulation game that pits a Bayesian investor against an investor that uses the resampling approach advocated by Michaud (1998). In the original game, Markowitz and Usmen (2003) find that the resampling player always wins. We level the playing field by allowing the Bayes player to use a more standard technique to approximate the moments of the predictive distribution. With this minor change, the game ends up essentially in a tie.

We also offer an investment game that more closely approximates the practical situation that investors face—a one-step ahead portfolio allocation. Here our results depend on the distributional assumptions. If the future distribution is just like the past, the Bayes player always wins. However, if there is a change in the distribution (i.e., the predictive distribution is different from the historical distribution), the resampling player shows advantages.

The dominate performance of the Bayes player, for the one-step ahead game, comes about because the investor faces more uncertainty (they have uncertainty about both the variability of the returns and about their ability to predict the mean) and because the Bayes approach results in a smaller average portfolio variance as the investor's risk aversion increases.

The Bayesian and resampling literature consider a broader interpretation of risk by focusing on parameter uncertainty. The Bayesian handles parameter uncertainty by averaging over parameter values in a way that is consistent with the data, the assumed distribution, and the prior beliefs, whereas the resampler resorts to a Monte Carlo simulation to deal with the uncertainty.

There is a third level of risk sometimes referred to as ambiguity. One can think of this as uncertainty about the distribution or uncertainty about the basic model. That is, while we might have a prior for a particular distribution, there are many possible distributions. Our results show that the resampling approach shows some robustness to distributional uncertainty. Our future research will focus on a Bayesian implementation to handle this third type of certainty.

Appendix: Details for Bayesian analysis

A.1 Posterior moments

Conditional on diffuse priors and the data gives a posterior density, $f(\mu, \Sigma|H)$, for each history. The predictive distribution, for the next observation in a history, is obtained by integrating out the model parameters with respect to the posterior density,

$$f(r_{n+1}|H) = \int_{\mu,\Sigma} f(r_{n+1}|\mu, \Sigma \times f(\mu, \Sigma|H) d\mu d\Sigma. \quad (A.1)$$

As shown in Markowitz and Usmen (2003) and Harvey *et al.* (2006), the expected value of the utility given in (3) and a specific history H becomes,

$$E[u_\lambda(\omega, r_{n+1})|H] = \omega'\hat{\mu} - \lambda\omega'\hat{\Sigma}\omega - \lambda\omega' Cov(\mu - \hat{\mu})\omega, \quad (A.2)$$

where $\hat{\mu}$ is the predictive mean, which is equal to the posterior mean,

$$\hat{\mu} = E[r_{n+1}|H] = E[\mu|H], \quad (A.3)$$

and where the predictive covariance matrix can be rewritten as the sum of the posterior mean of Σ and the posterior mean of the covariance of $\mu - \hat{\mu}$, or

$$\hat{\Sigma} = E[\Sigma|H] \quad \text{and}$$
$$\text{Cov}(\mu - \hat{\mu}) = E[(\mu - \hat{\mu})(\mu - \hat{\mu})'|H]. \quad (A.4)$$

Parameter uncertainty is taken into account by including this extra term $\text{Cov}(\mu - \hat{\mu})$ in the predictive covariance.

A.2 Model specification

The Bayes player assumes that all returns follow a normal probability model, or

$$f(r|\mu, \Sigma) = |\Sigma|^{-1}\left(\frac{1}{2\pi}\right)^{\frac{p}{2}} \times \exp\left[-\frac{1}{2}(r-\mu)'\Sigma^{-1}(r-\mu)\right], \quad (A.5)$$

where p is the number of assets, and assumes a set of diffuse conjugate priors for μ and Σ, or

$$\mu \sim N(\overline{\mu}, \tau^2 I), \quad (A.6)$$

and

$$\Sigma^{-1} \sim \text{Wishart}(\nu, SS). \quad (A.7)$$

By choosing diffuse hyper-parameters, the conjugate prior specification can be made to mimic the uniform prior specification used in the original experiment. (For example by letting $\overline{\mu} = 0$ and letting τ^2 be large, the prior for μ becomes essentially constant over the range of "reasonable" parameter values. The same can be obtained for Σ^{-1}, by letting $\nu = p + \delta$, letting $SS = \delta I$ and letting δ be small.) To illustrate how both modeling approaches can result in equally "objective" diffuse priors over the range of "reasonable" parameters values, consider a prior on μ. When there is only one asset, μ is a scalar. If we assume that the range of "reasonable" values for μ is between -100 and 100, then the uniform prior is given by

$$f_{\text{UniformPrior}}(\mu) = \frac{1}{200} I\{-100 < \mu < 100\}, \tag{A.8}$$

where $I\{\}$ is the indicator function, see Figure 3 for a graphical representation. If we assume a conjugate prior for μ, which is the Normal distribution, and set the hyper-parameters (or parameters of this prior distribution) to be equal to 0 for the mean and τ^2 for the variance, or

$$f_{\text{ConjugatePrior}}(\mu) = \frac{1}{\sqrt{2\pi\tau}} \exp\left\{-\frac{\mu^2}{2\tau^2}\right\}, \tag{A.9}$$

then the difference between these two prior specifications, for the "reasonable" values for μ disappears as τ^2 increases, see Figure 3 for an illustration. Similar prior specifications can be chosen for the covariance matrix Σ.

A.3 Approximating expected utility

In order to approximate the expected utility, with respect to the predictive distribution, the Bayes player generates samples from the posterior distribution

$$\mu^m, \Sigma^m \sim f(\mu, \Sigma | H, \overline{\mu}, \tau^2, n, SS), \tag{A.10}$$

and in turn generates samples from the predictive distribution for each draw from the posterior

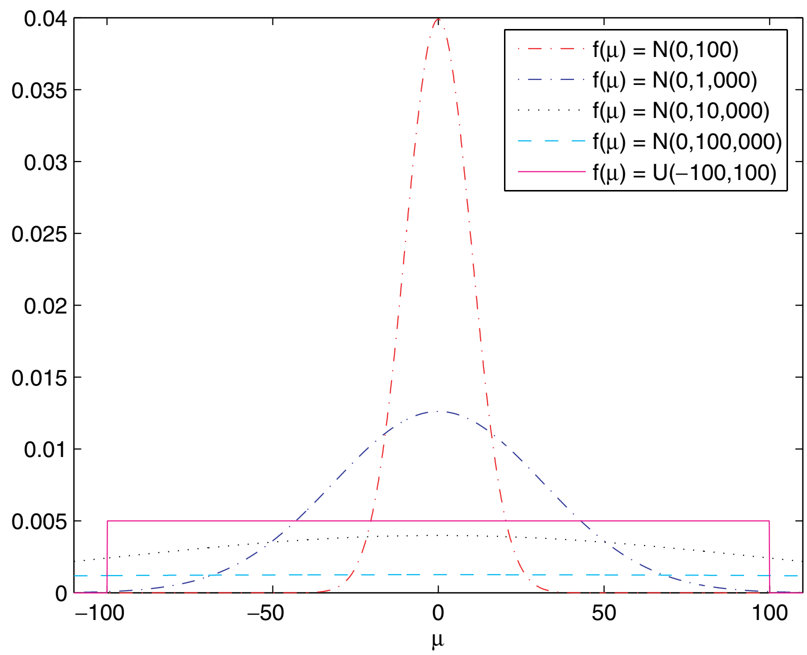

Figure 3 This figure shows several normal densities, and one uniform density that could be used as priors for μ. A normal density can be a non-informative prior by setting the standard deviation to be large. Markowitz and Usmen (2003) use a uniform density as their non-informative prior.

distribution,

$$r_{n+1}^{m,\varrho} \sim f(r|\mu^m, \Sigma^m). \tag{A.11}$$

In the implementation for the new experiment, the Bayes player ran the MCMC algorithm for a burn-in of 10,000 iterations (to allow the MCMC algorithm converge in distribution) and then generated 25,000 draws from the posterior and predictive densities (i.e., one sample from the predictive density for each posterior draw). The approximation of the expected utility for the Bayes player is calculated as follows:

$$E[u_\lambda(\omega, r_{n+1})|H] \cong \frac{1}{25,000}\sum\nolimits_{m,\varrho} \omega' r_{n+1}^{m,\varrho} - \lambda(\omega'(r_{n+1}^{m,\varrho} - \hat{\mu}))^2, \tag{A.12}$$

where

$$\hat{\mu} \cong \frac{1}{25,000}\sum\nolimits_{m,\varrho} r_{n+1.}^{m,\varrho} \tag{A.13}$$

For each history, the Bayes player finds and reports the weights that maximize (A.12).

Notes

[1] Estimation error has been examined by Bawa *et al.* (1979), Britten-Jones (1999), Chen and Brown (1983), Frost and Savarino (1986), Jobson and Korkie (1980 and 1981), Jorion (1985 and 1986), Klein and Bawa (1976), and Michaud (1989).

[2] Several authors have considered resampling including Bey *et al.* (1990), Broadie (1993), Christie (2005), diBartolomeo (1991 and 1993), Herold and Maurer (2002), Jorion (1992), Harvey *et al.* (2006), Michaud (2001), Mostovoy and Satchell (2006), and Scherer (2002, 2006).

[3] This approach is guaranteed to produce weights that result in an expected utility that is less than the maximum expected utility because the resampled weights will be different than the Bayes weights (see Harvey *et al.* (2006) for a discussion).

[4] Each resampled efficient frontier is based on μ_{ikm}^R and Σ_{ikm}^R, which are the standard estimates based on H_{ikm}^R.

[5] In order to explore the robustness of the results from the original experiment, we opted to use the MCMC algorithm and have the Bayes player generate samples from the posterior distribution and in turn generate samples from the predictive distribution for each draw from the posterior distribution. (Even though we are using conjugate priors, the joint, posterior density of μ and Σ is non-standard and cannot be integrated out analytically; hence the need to take a sampling based approach (MCMC) to integrate out the parameters with respect to the predictive density.) The approximation of the expected utility for the Bayes player is calculated by taking the average utility based on the draws from the predictive density. For each history, the Bayes player finds and reports the weights that maximize this average utility; see Appendix A.3 for the exact formulas.

[6] Table 1 follows the same format as Table 3 in Markowitz and Usmen (2003).

References

Bawa, V.S., Brown, S.J., and Klein, R.W. (1979). *Estimation Risk and Optimal Portfolio Choice*. New York: North-Holland Publishing Company.

Bernardo, J.M. and Smith, A.F.M. (1994). *Bayesian Theory*. New York: Wiley.

Bey, R., Burgess, R., and Cook, P. (1990). "Measurement of Estimation Risk in Markowitz Portfolios." University of Tulsa Unpublished paper.

Britten-Jones, M. (1999). "The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights." *Journal of Finance* **LIV**, 655–671.

Broadie, M. (1999). "Computing Efficient Frontiers Using Estimated Parameters." *Annals of Operations Research: Financial Engineering* **45**, 21–58.

Chen, S.-N. and Brown, S.J. (1983). "Estimation Risk and Simple Rules for Optimal Portfolio Selection." *The Journal of Finance* **38**, 1087–1093.

Christie, S. (2005). "Strategic and Tactical Asset Allocation in the Presence of Sampling Error." Unpublished paper. (Macquarie University).

diBartolomeo, D. (1991). "Estimation Error in Asset Allocation." Unpublished paper.

diBartolomeo, D. (1993). Portfolio Optimization: The Robust Solution Prudential Securities Quantitative Conference.

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

Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (1998). *Markov Chain Monte Carlo In Practice*. New York: Chapman and Hall.

Harvey, C.R., Liechty, J.C., Liechty, M.W., and Müller, P. "Portfolio Selection with Higher Moments." Working Paper. Duke University.

Herold, U. and Maurer, R. (2002). "Portfolio choice and estimation risk: A comparison of Bayesian approaches to resampled efficiency." Working Paper. Johann Wolfgang Goethe University.

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

Jobson, J.D. and Korkie, B. (1981). "Putting Markowitz Theory to Work." *Journal of Portfolio Management* 7(4), 70–74.

Jorion, P. (1992). "Portfolio Optimization in Practice." *Financial Analysts Journal* **48**, 68–74.

Jorion, P. (1986). "Bayes-Stein Estimation for Portfolio Analysis." *Journal of Financial and Quantitative Analysis* **21**, 279–292.

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

Klein, R.W. and Bawa, V.S. (1976). "The effect of estimation risk on optimal portfolio choice." *Journal of Financial Economics* **3**, 215–232.

Markowitz, H.M. and Usmen, N. (2003). "Resampled Frontiers Versus Diffuse Bayes: An Experiment." *Journal of Investment Management* **1**(4), 9–25

Michaud, R.O. (2001). *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. New York: Oxford University Press.

Michaud, R.O. (1998). *Efficient Asset Management*. Boston: Harvard Business School Press.

Michaud, R.O, (1989). "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?" *Financial Analysts Journal* 31–42.

Mostovoy, D. and Satchell, S.E. (2006). "Robust inference in quantitative finance—especially robust optimization." Alpha Strategies Investment Seminar Presentation, Duke University.

Robert, C.P. and Casella, G. (1998). *Monte Carlo Statistical Methods*. New York: Springer Verlag.

Scherer, B. (2006). "A Note on the Out-of-Sample Performance of Resampled Efficiency." *Journal of Asset Management* 7, 170–178.

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

Keywords: Bayesian decision problem; parameter uncertainty; optimal portfolios; utility function maximization; resampling.

Glossary

Conjugate prior—a prior distribution for a parameter, where the resulting full-conditional distribution (the distribution conditional on the remaining parameters and the data) is from the same family of distributions as the prior distribution. For example, for the models considered in this paper, if we assume μ follows a Normal distribution, before observing any data, then the distribution of μ conditional on Σ and the data is a Normal distribution.

Diffuse Bayesian analysis—Summary of parameter distributions, assuming a Bayesian model, where the prior distributions are chosen to be vague or non-informative.

Diffuse prior—a prior distribution that is vague or non-informative, where the information provided by the data dominates the information provided in the prior.

Hierarchical Bayesian model—a statistical model that is specified in a hierarchical fashion; typically the distribution of the observed data is given conditional on a set of parameters (random variables) and the (prior) distribution of these parameters is given conditional on another set (or hierarchy) of parameters.

Importance Sampling—A Monte Carlo technique for sampling, where samples are drawn from a proposed distribution and then are re-weighted according to a target distribution in order to obtain a sample from the target distribution.

Inverse Wishart distribution—a family of distributions for covariance matrices. To contrast with the Normal distribution, if excess returns r are Normally distributed, this describes the distribution of returns; in contrast an Inverse Wishart distribution describes the distribution of Covariance matrices.

Markov Chain Monte Carlo (MCMC)—Monte Carlo integration using Markov Chains. Samples from a distribution of interest (for example a posterior distribution) are obtained by repeatedly sampling from the distribution of each parameter, conditional on the most recently sampled values of the remaining parameters and the data. This forms a Markov Chain, that results in samples from the joint distribution of interest.

Predictive distribution (density)—The distribution of the data in the future, conditional on all of the observed data and the prior distributions. For example, the distribution of tomorrow's excess returns, conditional on a set of historical excess returns and prior beliefs.

Prior distribution—a distribution placed on a parameter before any data is observed. This can represent an expert's prior opinion or be vague and non-informative.

Posterior distribution—The distribution of the model parameters, conditional on all of the observed data and the prior distributions. For example, the distribution of the average excess returns μ and the covariance matrix Σ conditional on a set of historical excess returns and prior beliefs.