Consistency regions and frontiers: using density forecasting to find consistent portfolios

N. Meade[1] and J.E. Beasley[2,3]

[1]Imperial College Business School, London SW7 2AZ, UK, n.meade@ic.ac.uk

[2]Mathematical Sciences, Brunel University, Uxbridge UB8 3PH, UK, john.beasley@brunel.ac.uk

[3]JB Consultants, Morden SM4 4HS, UK, john.beasley@jbconsultants.biz

November 2014

Abstract

Our objective is to develop a methodology to detect regions in risk-return space where the out-of-sample performance of portfolios is consistent with their in-sample performance. We use the Berkowitz statistic to evaluate the accuracy of the density forecast, derived from in-sample portfolio returns, of out-of-sample portfolio returns. Defined by its co-ordinates in risk-return space, a portfolio is ‘consistent’ if this statistic indicates that the in-sample return density is a good predictor of out-of-sample portfolio returns.

Given a track record of portfolio selections by an algorithm, we create a framework that defines a portfolio's position in risk-return space and describes a ‘consistency region’ bounded by a ‘consistency frontier’. We argue that, for a given level of expected return, a consistent portfolio is preferable to an inconsistent one. This methodology for the creation of a consistency region is independent of the portfolio selection algorithm.

We use mean-variance portfolio selection, to firstly validate the approach by using well-behaved simulated multivariate normal returns; variation of the time interval for estimation demonstrates that the consistency region behaves intuitively reasonably. Secondly, we show that with real data, the behaviour of the consistency region is driven by asset return dynamics and consequently less stable.

Two investment strategies using consistency are created and compared with a minimum variance portfolio benchmark strategy. In nearly all cases, consistent portfolios lead to out-of-sample outperformance of the benchmark. We also investigate the consistency of the naive diversified (*1/N*) portfolio.

Keywords: Berkowitz statistic; Density forecasting; Portfolio selection; Risk

JEL Classification: G11, G17

1. Introduction

A key stage in asset allocation is the construction of an efficient frontier, a set of dominant portfolios that offer the lowest available risk for a given level of return. The best known approach is that of Markowitz [1], where return is measured by the mean return of the portfolio and risk is measured by standard deviation in portfolio return. Since the density of returns is assumed to be normal, then in effect the efficient frontier identifies the density function of the returns of a portfolio at any point on the frontier (since the normal distribution is defined by just two parameters, the mean and the standard deviation). A typical example of an efficient frontier (using the universe of the Dow Jones Industrial 30 Stock index, henceforth the Dow Jones) is shown in the upper plot in Figure 1. In the lower plot, the predicted density of returns is super-imposed for three portfolios on the frontier; note the effect of the inclusion in this representation of the uncertainty in portfolio returns on the scale of the return axis. Simply stated the return distribution covers a much wider range than might be implied by a naive interpretation of the efficient frontier. This is important since an investor, in choosing a portfolio from the efficient frontier, is implicitly exposing themselves to the implied return distribution. To make matters worse, it is entirely conceivable that the return distribution implied by the efficient frontier (produced via optimisation using in-sample data) does not accurately represent the achievable out-of-sample return distribution. In these circumstances the investor is not making an investment decision under uncertainty, but in total ignorance; the investor has no reliable information about the selected portfolio's out-of-sample performance.

The construction of an efficient frontier in risk-return space is an attractively straightforward way of representing a set of portfolios of assets with the appropriate risk-return characteristics for an investor. The steps in this construction involve the collection of historical data; processing this data into the necessary inputs (such as a vector of estimates of expected asset returns and an estimate of their covariance matrix); computation of asset weightings within the set of selected portfolios. If the input data is wholly representative of the asset data generating process which will generate the out-of-sample returns, then the efficient frontier reflects valid investment information. Conversely, if this condition is not met, then the investor will be misled. However, inside the efficient frontier, there may be dominated portfolios which do yield out-of-sample returns consistent with their co-ordinates in risk-return space. An alternative characterisation is to say that the parameters of the density function of returns for a portfolio on the efficient frontier may be estimated less accurately than those for a dominated portfolio situated some way inside the frontier. Here we examine the accuracy of the return density functions resulting from portfolios both on and within the efficient frontier.

We introduce the term 'consistent' to describe a portfolio whose position in risk-return space indicates that the in-sample information allows its out-of-sample return density to be accurately predicted. With a consistent portfolio, the investor is making a decision under well-calibrated uncertainty. In order to assess consistency, a track record of comparable portfolio investments is necessary. This record is composed of a series of these actions: portfolio selection; distillation of return density from the in-sample portfolio returns; observation of the out-of-sample return. To compile this track record, we set up a grid based on the efficient frontier, with this grid we give co-ordinates to those portfolios on the frontier and to the dominated portfolios. The scale on the expected return axis of the grid runs from a minimum of the expected return of the minimum variance portfolio through to the maximum expected return. The scale on the risk axis is conditional on the level of expected return and is defined by the range the variance of a portfolio can take. For each point on this grid, the sequence of in-sample densities and the corresponding sequence of out-of-sample returns allows the construction of a time series of realised probabilities, i.e. the probability of the portfolio achieving the observed out-of sample return, given its in-sample density function. These probabilities allow the accuracy of the in-sample densities to be measured. We use the Berkowitz statistic, a recently developed and powerful tool for the evaluation of the accuracy of density forecasts. We call a portfolio '***consistent***' when the Berkowitz statistic indicates that the in-sample return density is a good predictor of the out-of-sample portfolio return, i.e. the out-of-sample return density is consistent with the in-sample parameters. Collectively consistent portfolios define a '***consistency region***', and we call the boundary between the consistency region and inconsistent portfolios the '***consistency frontier***'.

In our analysis we first demonstrate that the consistency regions behave intuitively reasonably under the 'laboratory' conditions represented by simulated multivariate normal data. We then use data from two markets (US: Dow Jones 30 constituents and Japan: Nikkei 225 constituents) to show how the consistency regions react to the changing dynamics of the market. To evaluate the economic implications of consistency, we examine two types of investment strategies based on the use of consistent portfolios and compare them with the benchmark strategy of investing in the minimum variance portfolio. We also investigate the consistency of the naive diversified (*1/N*) portfolio.

In summary, we propose an innovative methodology, based on the measurement of density forecasting accuracy, to identify the consistency region, that region of risk-return space where portfolios will exhibit the same performance out-of-sample as they exhibited in-sample. We will demonstrate that investment strategies that make use of consistency information add value, that is they outperform strategies that do not use such information. This paper has six sections; in the next section we review the literature concerned with estimation errors in the efficient frontier and give the background to accuracy measurement in density forecasting. In Section 3, we describe our data and our methodology for using density forecasting. In Section 4, we use simulated and real data to investigate the behaviour of the consistency region. We explore the investment implications of our analysis in Section 5. In Section 6 we draw our conclusions.

2. Literature review

2.1 Estimation for portfolio selection

The portfolio selection problem considers a universe of N assets whose performance is summarised by a return vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$. Markowitz [1] showed that quadratic programming can be used to establish an efficient frontier of portfolios of assets by finding the optimal set of weights on the assets, $\boldsymbol{\omega}$, such that the portfolios on the frontier have minimum risk for a given level of expected return. In practice, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown and have to be estimated from historical data. However, when the true (but unknown) parameters are replaced by sample estimates, the search for a minimum variance portfolio exploits errors in in-sample variances and covariances. Michaud [2] suggested that unless mean-variance optimisation was used with '*sophisticated adjustment of the inputs*', its use '*may often do more harm than good*'. Broadie [3] expresses the trade-off in estimation: using too few historical observations leads to estimation errors; using too many historical observations means that the parameter estimates may be outdated. This second point implies that the parameters are time dependent and would be better represented as $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$. Broadie concludes that '*points on the estimated frontier are optimistically biased predictors of actual portfolio performance*'. He suggests that improving estimation of returns is the most effective way to achieve bias reduction. Exploration of the problem of estimation bias has followed two paths. One path looks at ways of making parameter estimation more robust. The other path looks at modifying the constraints and the objective function applied when optimising to lessen the impact of estimation errors.

Robust estimation using Stein or shrinkage estimators (see James and Stein [4]) was introduced by Jobson and Korkie [5]. Jorion [6], Chopra, Hensel and Turner [7], among others, explored this topic. In a survey of work on robust portfolios, Fabozzi, Huang and Zhou [8] point out that the Bayesian approach is appropriate when the true data generating process is unknown. Jorion [9] compares Bayes-Stein estimation with CAPM based estimation of mean asset returns, he finds that the latter method outperforms Bayes-Stein estimation which itself outperformed classical estimation. Ledoit and Wolf [10] applied shrinkage estimation to the covariance matrix; they produce an optimally weighted average estimate using the sample covariance matrix and the single-index (CAPM) covariance matrix. Fabozzi, Huang and Zhou [8] also point out that variance does not capture all components of what many investors consider as risk, but they further suggest that no single measure exists as a universal solution. In this context, Tu and Zhou [11] devised an approach for the inclusion of uncertainty in the data generation process to account for fat tails in return density. Although this led to changes in portfolio weights, they concluded that the certainty-equivalent penalty for ignoring fat tails was low and that for a mean-variance investor the normality assumption worked well.

Using a constraint-based-approach, Jagannathan and Ma [12] motivate their analysis by drawing attention to the practical problem of selecting a portfolio from a universe of 500 assets. With a maximum of around 900 monthly observations of returns available there are fewer than four degrees of freedom for each parameter estimate. They focus on why the imposition of constraints helps reduce risk. If an asset has large estimated covariances with other assets, then it is likely to be drawn into the minimum variance portfolio (with the opposite signed weight to those on the other assets, assuming short-selling is allowed). Imposing a non-negativity constraint on the asset weight (i.e. not allowing short-selling) thus has the effect of reducing the covariances of assets bound by the constraint. Similarly, the imposition of upper bounds on weights tends to affect assets with low covariance with other assets, in this case effectively increasing some covariances. They point out that the effect of these constraints is similar to using shrinkage in estimation.

Several authors have pointed out the practical difficulties in selecting an efficient portfolio. A summary of their findings is given in Table 1, which gives their conclusion, the data underlying their analysis and the criterion used for any comparisons.

Looking at the conclusion column in Table 1, we find support for schemes that constrain the vector of asset weights. There is support for shrinkage estimation of mean returns, but the emphasis has moved to the minimum variance portfolio, i.e. ignoring mean return estimates, (see De Miguel, Garlappi and Uppal [21]). Shrinkage estimation of the covariance matrix receives some support, but simpler ways of averaging estimates are reported to work as well.

Table 1. A summary of papers critical of portfolio selection using efficient frontiers based on mean and covariance estimates (ordered by year of publication)

Paper	Conclusion	Data	Criterion
Frankfurter, Phillips and Seagle [13]	Mean-variance portfolios are no more likely to be efficient than randomly selected portfolios	Simulated multivariate normal data	Frequency of domination
Bloomfield, Leftwich and Long [14]	Average return on tangency portfolio not significantly better than equal weighted portfolio	US equities	Comparison of means
Jorion [15]	Out-of-sample, shrinkage estimators significantly outperform sample mean	7 National indices, monthly, Jan 1972 – Dec 1983	Sharpe ratio
Broadie [3]	'*points on the estimated frontier are optimistically biased predictors of actual portfolio performance*'	Simulated data	Distances between 3 frontiers: using estimated parameters, true parameters and 'actual' frontier based on estimated weights with true parameters
Chan, Karceski and Lakonishok [16]	Out-of-sample performance of minimum variance portfolio dominates equal weighted portfolio	NYSE and AMEX stocks (1968 – 1998)	Sharpe ratio
Polson and Tew [17]	Found significant out-performance of index using daily data for volatility estimation and upper and lower bounds on asset weights	S&P 500 Daily (1970 - 1996)	Distance to the ex-post efficient frontier of optimised portfolio relative to the S&P index
Wang [18]	Aversion to model uncertainty leads to portfolios that are not mean-variance efficient	NYSE, AMEX and NASDAQ stocks (1963 - 1998)	Single period mean-variance utility function
Disatnik and Benninga [19]	Found no advantage in using shrinkage estimation of covariance matrix. Averages of sample covariance matrix with single index matrix and others worked just as well.	NYSE (1964 - 2003)	Ex-post standard deviations of returns of minimum variance portfolio
DeMiguel, Garlappi and Uppal [20]	Constraining the norm of the weight vector often leads to higher Sharpe ratios than shrinkage or other approaches	4 sets of US portfolios (1963 - 2004) plus a CRSP set (1968 - 2005)	Out-of-sample portfolio variance, Sharpe ratio, and portfolio turnover

2.2. Accuracy measurement in density forecasting

Density forecasting is the prediction of the density function of a random variable from a time series of observations of the variable. In our case, the random variable is the out-of-sample return on a portfolio selected at time *t*, say R_t. This return is forecast as having a cumulative density function (*cdf*), say $F_t(\ )$. Since $F_t(\ )$ may change over time, the approach commonly followed in measuring the accuracy of the density forecast is to use the empirical *cdf* (or probability integral transform) as the datum. The measurement of the accuracy of the density forecast is then based on a sequence of realised observations of out-of-sample returns and corresponding predicted *cdf*s. One approach to the comparison of density functions develops the use of Kolmogorov-Smirnov methodology for density comparison; recent developments and the background to this approach are given by Corradi and Swanson [22]. Another approach centres on the uniform distribution. Diebold, Gunther and Tay [23] note that if the *cdf* is predicted accurately, then its empirical evaluation will be a uniform random variable between zero and one; i.e. $F_t(r_t) \sim U(0,1)$ where r_t is the realised out-of-sample return. Hong and Li [24] and Hong, Li and Zhao [25] develop this approach.

Mitchell and Hall [26] and Bao, Lee and Saltoglu [27] give full discussions of the measurement of density forecasting accuracy. In line with their findings, we use an approach due to Berkowitz [28]. Under the null hypothesis, that the *cdf* of the out-of-sample return has been correctly identified, the sequence of *cdf*s (evaluated at the observed return) should be uniform (0,1) random variables. We use the Berkowitz test, where the *cdf*s are transformed to normal random variables; the reasoning is that tests are well developed to detect departures from normality. The test statistic, for *K* observed returns r_k $k=1,\dots,K$, is given below.

$$\text{Berkowitz statistic} = 2\sum_{k=1}^{K}\left(\ln\left(\frac{1}{\hat{\sigma}}\phi\left(\frac{y_k-\hat{\mu}}{\hat{\sigma}}\right)\right)-\ln\left(\phi\left(y_k\right)\right)\right) \qquad (1)$$

where $y_k = \Phi^{-1}\left(\hat{F}_k\left(r_k\right)\right)$ and $\hat{\mu}, \hat{\sigma}$ are the sample mean and standard deviation of y_k; $\phi(\ )$ and $\Phi(\ )$ are the density and cumulative density functions of the standard normal random variable. Under the null hypothesis that *F(.)* is correctly identified, the Berkowitz statistic is χ_2^2, see Berkowitz [28].

3. Experimental procedure

3.1 Data

Since we demonstrate our analysis throughout with examples, we now define the two data sets used from the US and Japan. Weekly data was collected for the members of the Dow Jones Industrials 30 share index from January 1988 to November 2009, the 28 companies who were members of the index throughout this period form the universe of assets (i.e. N = 28); this is the primary data set and is used to demonstrate issues throughout this paper. In our analysis, prior to using real-world data, we also use simulated data. We generate a set of asset returns constructed to be multivariate normally distributed random variables using the means and the covariance matrix of the Dow Jones estimated from the data described. The time series of vectors of asset returns are generated following a Cholesky factorization of the covariance matrix. To broaden our results, we also use a Japanese data set for the analyses in Section 5. Weekly data was collected for members of the Nikkei 225 index over the same time period as for the Dow Jones, January 1988 to November 2009. The 191 companies who were members throughout this period form the universe in this case. For both our data sets, the asset universes consist of those members of the index that survived throughout the period considered. Since our analysis is a comparison of strategies, each with access to the same information, we feel that survivorship bias is not an issue.

3.2 Methodology

The initial objective of our investigation is to discover where in risk-return space a portfolio selection strategy produces out-of-sample returns that are consistent with the in-sample portfolio mean and variance. In order to explore the space dominated by the efficient frontier, we need to identify portfolios within this space. The methodology for the identification of consistency regions is separate from the portfolio selection methodology. Here, for simplicity and convenience, we use straightforward mean-variance analysis, the procedure is as follows.

At a particular time period, t_k, we construct an efficient frontier of portfolios from a universe of N assets using an estimation interval of M historical observations of returns. We consider long only portfolios for convenience to avoid the issues associated with short selling such as excessive transaction costs and its possible prohibition in some countries. We start with the minimum variance portfolio; we use this term throughout to describe the minimum variance portfolio with long only holdings. The weights on the assets in the portfolio are denoted by the column vector $\boldsymbol{\omega}$, its transpose by $\boldsymbol{\omega}'$, the weights $\boldsymbol{\omega}_0$ in the minimum variance portfolio are the solution to:

$$\textit{Minimise}\ \boldsymbol{\omega}'\hat{\boldsymbol{\Sigma}}_{\mathbf{t}_k}\boldsymbol{\omega} \quad \textit{subject to}\ \boldsymbol{\omega} \geq 0\ \textit{and}\ \sum_{i=1}^{N} \omega_i = 1 \qquad (2)$$

The expected return on the minimum variance portfolio is $R_0(t_k) = \sum_{i=1}^{N} \omega_{0,i} \hat{\mu}_{i,t_k}$ and its standard deviation is $SD_0(t_k) = \sqrt{\boldsymbol{\omega}_0' \hat{\boldsymbol{\Sigma}}_{\mathbf{t}_k} \boldsymbol{\omega}_0}$. The ^ symbol denotes that the covariance matrix and expected return vector are estimated from sample data. The efficient frontier is then evaluated at *B* equally spaced points (on the return axis) between the expected return $R_0(t_k)$ associated with the minimum variance portfolio and the asset with the highest expected return, $R_B(t_k) = \max(\hat{\mu}_{1,t_k}, ..., \hat{\mu}_{N,t_k})$, i.e. we examine expected return levels $(R_0(t_k), R_1(t_k), \ldots, R_{B-1}(t_k))$. We use the term 'efficient frontier' as shorthand for the section of the whole efficient frontier, subject to the long only constraint.

The next step is to identify the dominated portfolios within the efficient frontier. For each of these expected return values, $R_b(t_k), b = 0, \ldots, B-1$, 5000 random portfolios are generated with non-negative weights summing to one giving an expected portfolio return $R_b(t_k)$, i.e. satisfying both the constraints in (2) plus $R_b(t_k) = \sum_{i=1}^{N} \omega_i \hat{\mu}_{i,t_k}$. The range of standard deviations of these portfolios is found from the maximum portfolio standard deviation, denoted as $SD_{b,C-1}(t_k)$ to the minimum, denoted as $SD_{b,0}(t_k)$. By definition the portfolio with minimum standard deviation (and expected return $R_b(t_k)$) lies on the efficient frontier. We define *C* equally spaced points (on the standard deviation axis), $(SD_{b,0}(t_k), SD_{b,1}(t_k), \ldots, SD_{b,c}(t_k), \ldots, SD_{b,C-1}(t_k))$.

By means of the above we have divided the region inside the efficient frontier into a grid structure which has within it *B*×*C* points, where a point (*b,c*) in the grid has associated expected return $R_b(t_k)$ and standard deviation $SD_{b,c}(t_k)$.

The specification of the mean and standard deviation for a portfolio at grid point (*b,c*) does not define the portfolio weights unambiguously. To achieve as smooth a transition (in weights and in subsequent forecasting behaviour) as possible between portfolios in neighbouring grid positions, we define the portfolio weights uniquely (in nearly all cases). We choose the weights by minimising their variance. The weights at grid point (*b,c*) are hence the solution to:

$$Min(Variance(\omega_1, \ldots, \omega_N)) \text{ subject to: } \omega_i \geq 0 \ \forall i \ and \sum_{i=1}^{N} \omega_i = 1,$$

$$\sum_{i=1}^{N} \omega_i \hat{\mu}_{i,t_k} = R_b(t_k) \text{ and } \boldsymbol{\omega}' \hat{\boldsymbol{\Sigma}}_{\mathbf{t}_k} \boldsymbol{\omega} = (SD_{b,c}(t_k))^2 . \quad (3)$$

Minimising the variance of the weights tends to make the weights all small and similar. This minimisation with a non-linear constraint uses an algorithm due to Schittkowski [29], the starting point for the minimisation is the nearest random portfolio.

We use the data set constructed to be multivariate normally distributed returns to demonstrate the output from the exercise so far. The efficient frontier and the associated dominated random portfolios are shown in Figure 2 where there are $N = 28$ assets and $M = 260$ historical observations of returns (an estimation interval equivalent to five years); we generate $C = 50$ portfolios for each of $B = 11$ return values. It can be seen that the distribution of portfolio standard deviations is sometimes slightly irregular. This arises for computational reasons as, if a solution to (3) is not found (which may happen as (3) is a quadratically constrained quadratic program that is hard to solve numerically), the nearest random portfolio is used instead.

For each of the $B \times C$ portfolios considered, we note the realised value of the random variable, $r^{b,c}_{t_k,H}$, the out-of-sample portfolio return over an investment horizon of H time periods from $t_k + 1$ to $t_k + H$. We hypothesise that the out-of-sample portfolio return is drawn from a distribution defined by its historical (in-sample) performance. To provide the material to test this hypothesis, we compute the empirical *cdf* of the out-of-sample return using the in-sample returns. There are $M–H+1$ in-sample (overlapping) returns: $\left(r^{b,c}_{t,H} \mid \Theta_{t_k} \text{ for } t = t_k - M + 1 \text{ to } t_k - H + 1 \right)$ where Θ_{t_k} represents the information set available at time t_k when the portfolio weights were estimated. The derivation of the empirical *cdf* of the out-of-sample return, $\hat{F}(R^{b,c}_{t_k,H} \mid \Theta_{t_k})$, is explained in Appendix A. Our null hypothesis is:

$$\text{the } cdf \text{ of the random variable } r^{b,c}_{t_k,H} \text{ is } \hat{F}(R^{b,c}_{t_k,H} \mid \Theta_{t_k}) \qquad (4)$$

In other words, under the null hypothesis we believe that the observed distribution of out-of-sample returns is consistent with the observed in-sample return distribution.

In order to measure the accuracy of the density forecast provided by the empirical *cdf*, we need repeated observations for each grid point (*b,c*). We consider portfolio selection at K different origin times, $\left(t_1, t_2, \ldots, t_k, \ldots, t_K\right)$. At each origin t_k, for $k = 1, \ldots, K$, for each portfolio with expected return $R_b\left(t_k\right)$ and standard deviation, $SD_{b,c}\left(t_k\right)$ the within-sample historical portfolio returns and the out-of-sample return are calculated and the empirical *cdf* of the out-of-sample return computed. Once this has been done for each of the K origins, the procedure concludes with the computation of the Berkowitz statistic for each grid point (*b,c*).

We are interested in those portfolios (whose positions in the set of possible portfolios are defined by their grid coordinates (*b,c*)) where the out-of-sample returns are consistent with the null hypothesis given in (4). This means we are concerned with '***accepting***' this null hypothesis, rather than with rejecting it (as is more common in hypothesis testing). In order to make acceptance of the null hypothesis a demanding requirement, we need to set the probability of rejecting the null hypothesis when it is true (a type I error) at a higher value, say 10%, than we would if our concern was to detect slight departures from the null hypothesis. As mentioned earlier in Section 2.2, under the null hypothesis the Berkowitz statistic is a χ_2^2 random variable.

Since this use of a χ_2^2 random variable is an asymptotic result, we investigated whether some adjustment was necessary for small values of *M* or *K*. The experiment and the correction factor developed are described in Appendix B. In order to test the sensitivity of the Berkowitz statistic to departures from the null hypothesis, the following experiment was conducted. We generate *M* = 260 observations of in-sample data by draws from a normal distribution, $N\left(0.005,\left(0.01\theta\right)^2\right)$, with $\theta = 1$. The out-of-sample observation is a draw with θ, the multiplier of the standard deviation, taking on a range of values from 0.5 to 1.5 (the investment horizon, *H,* is one period). The power of the Berkowitz statistic (for K = 78 origins, a value used in our later analysis) to reject the null hypothesis at levels of significance of 10%, 5% and 1%, are plotted in Figure 3, calculated using 10,000 replications. For $\theta = 1$, the probabilities of rejection of a correct hypothesis are at their expected values of 10%, 5% and 1%. The probability of detection of a departure from the identified distribution increases with $\left|\theta - 1\right|$. The probability of rejection reaches 50% for the 10% test for $\theta < 0.85$ and $\theta > 1.15$. In other words, if the standard deviation of the out-of-sample return differs by 15% from its in-sample value, there is a probability of 0.5 of detecting this (and hence rejecting the null hypothesis) using a test with a 10% significance level.

4. Empirical analysis

As explained in Section 3.2, the use of density forecasting for the discovery of consistency regions is independent of the portfolio selection procedure, thus in our empirical analysis we use standard approaches to the estimation of the inputs to the mean-variance optimisation algorithm. We estimate the vector of expected asset returns using straightforward averages. We estimate the covariance matrix using either the shrinkage method of Ledoit and Wolf [10] (when using actual data and a relevant market index is available) or, straightforward sample estimates (when using simulated data and an index is not available).

4.1. Validation using simulated data

It has long been known that asset returns are not well described by the normal distribution; see for example, Blattberg and Gonedes [30], Kon [31], Barndorff-Nielson [32] and a fuller survey in Meade [33]. Departures from normality and other phenomena, such as time dependency in parameters, may lead to the selection of inconsistent portfolios. Therefore it is important to validate our proposed procedure in a clearly defined environment. Thus we examine the accuracy of return density forecasts in the idealised situation of a data set known to be multivariate normally distributed, that is we use the simulated data described in Section 3.1. This data set was used to produce the example efficient frontier and dominated portfolios shown in Figure 2 and has $N = 28$ assets and $M = 260$ historical observations of returns. The efficient frontiers were estimated using the sample covariance matrix. For each value of expected portfolio return, there are $C = 50$ portfolios designed to be equally spaced along the standard deviation axis. The investment period, H, was chosen to be 4 weeks, the empirical *cdf* is found from the M-H+1 overlapping observations of 4 week returns.

The presentation of the measure of density forecasting accuracy makes use of the grid points (b,c) and entails reporting the number of bands ($B = 11$) by the number of risk levels ($C = 50$), in this case 550 grid points and values of the Berkowitz statistic. We decided to present this information in graphical form. In Figure 4, we give examples in the form of a version of Figure 2, summarising $K = 78$ origins (since the investment horizon is four weeks, this equates to considering density forecasts over six years). The means and standard deviations are averaged for each combination of b and c. The six plots shown correspond to the six different values of M used (M = 52, 104, 156, 208, 260, 312 equivalent to estimation intervals of one to six years). The (green) crosses signal where the null hypothesis (that the out-of-sample returns are generated by the observed in-sample distribution) cannot be rejected at 10%; if the hypothesis can be rejected then a (red) dash is shown. For convenience, we refer to the (green) area of the graph where the null hypothesis is accepted as the '***consistency region***'. ***Portfolios in the consistency region have out-of-sample return characteristics that are consistent with those displayed in-sample***.

Due to the stochastic nature of hypothesis testing where the values of the Berkowitz statistic are close to the critical values used, the boundary of the reject/accept region (between inconsistent and consistent portfolios) is fuzzy. To decrease this fuzziness and make the transition from rejection of the hypothesis to acceptance appear more smoothly for each band (of constant expected return) in the plots, we use the lower quartile of the Berkowitz statistics starting from the efficient frontier up to the value of *c* as the criterion for acceptance and thus the consistency frontier. This boundary between the consistency region and the other portions of risk-return space (containing inconsistent portfolios) we call the '***consistency frontier***'. Using the consistency region as our guide to the accuracy of density forecasting, we find the following.

As the estimation interval, *M*, increases, the consistency region expands towards the efficient frontier, equivalently the consistency frontier moves towards the efficient frontier. Examining the bottom line of the plots (the return associated with the minimum variance portfolio) we see that this level of return is only, for *M* = 52, 104 and 156, associated with consistent portfolios with a higher variance. For these values of *M* therefore the expected return and standard deviation of any portfolio on the efficient frontier cannot be used to predict out-of-sample behaviour. For *M* = 208, the consistency region does include the efficient frontier for the lower values of expected returns $\left(b \le 2\right)$. With an increase in the estimation interval to *M* = 260, the consistency region includes nearly all portfolios, except those close to the frontier for the higher expected returns on the efficient frontier $\left(b \ge 6\right)$. For *M* = 312, the consistency region includes all but the top of the efficient frontier and the dominated portfolios. For this idealised case of multivariate normal returns with means and covariances unchanging with time, an estimation period of $M \ge 312$ (weekly) observations allows the efficient frontier to accurately predict out-of-sample returns.

4.2. Validation using market data

In this section the data for the *N* = 28 members of the Dow Jones was used directly. To summarise market behaviour during the period considered, the Dow Jones index is plotted in Figure 5. The density forecasts for an investment horizon of *H* = 4 weeks, for *K* = 78 (equivalent to six years) consecutive investments are collected for two evaluation dates, 27 November 2001 and 22 September 2009. These dates are identified in Figure 5. In the period before the first date, we can see that the index value is mostly trending upwards. The index experiences some wide fluctuations in the period between the first and second date.

In Figure 6, we plot the averaged efficient frontiers, the averaged dominated portfolios and the consistency regions for three sizes of estimation intervals (*M* = 260, 312 and 416) for both evaluation dates. Since with real data, we have access to the index, in order to calculate the efficient frontier, we use the Ledoit and Wolf method to estimate the covariance matrix of returns.

For the first date, the efficient frontier and most of the dominated portfolios are not consistent. In other words, during the periods before this date, an investor choosing any efficient (or inefficient) portfolio would not have experienced out-of-sample returns that in-sample observations would have led them to expect. By contrast, for the second date, the lower part of the efficient frontier and most of the dominated portfolios are consistent. In other words, in the periods before this date an investor choosing any efficient portfolio at a lower return level, or an inefficient portfolio at a higher return level, would have experienced out-of-sample returns that in-sample observations would have led them to expect. For the second evaluation date, (M = 312 and 416) there is a stronger resemblance to Figure 4 (which showed comparable results for the time-invariant multivariate normal data).

An explanation for the inconsistency of the portfolios shown in Figure 6 for the first evaluation date in November 2001 lies in Figure 5. In the period before this evaluation date we can see that the Dow Jones index followed a strong upward trend, this is evidenced in the scale of the expected return axis in the left hand plots in Figure 6. After 1996, the volatility tends to be higher and more variable. As the estimation interval, M, increases more of the pre-1996 data is used, causing the predicted mean returns to be overestimates, and the predicted volatility of asset returns to be underestimates.

The greater consistency of the portfolios for the second evaluation date (September 2009) may be explained as follows. The period before the second evaluation date has no long term price trend and periods of high volatility, thus in Figure 6 the expected returns are lower and the average standard deviation is higher. The absence of a long term price trend means there is less over or under-estimation of returns.

Although the corresponding consistency regions for the two evaluation dates are very dissimilar, these periods are eight years apart. The consistency region for a given value of M evolves over time. As an example, in Figure 7, we show six intermediate plots of the efficient frontier and dominated portfolios inside and outside the consistency region, for M=312 and K=78. These plots form an equally spaced series when taken together with the two plots for M=312 in Figure 6 at the beginning and end of the series. In the first plot in Figure 7 (21/01/2003) the consistency region is very small, well inside the efficient frontier; it then grows slightly in the second (17/02/2004). In the third (12/04/2005) and fourth plots (06/06/2006), the consistency region stretches out behind a consistent minimum variance portfolio on the efficient frontier. The consistency region shrinks in the fifth plot (03/07/2007). By the sixth (29/07/2008) plot, the consistency region has disappeared. The consistency region then re-establishes itself as shown in the plot in Figure 6.

We see that with real data, where the return vector and covariance matrix are varying over time, the consistency region tends to be smaller, sometimes non-existent, and its position is far more variable than in the idealised circumstances examined in Section 4.1 and Figure 4.

5. Investment strategy implications of the consistency region

In terms of practically useful information, it is clear that the portfolios on the efficient frontier belong to the consistency region intermittently. In this section we examine two possible investment strategies that follow the evolution of the consistency region over time.

To judge the worth of investing in portfolios that belong to the consistency region (or not) we use the Sharpe ratio of portfolio returns as a summary investment performance measure. The use of the Sharpe ratio for demonstrating differences in performance is a common practice, even when there is parameter or model uncertainty, see for example, Garlappi, Uppal and Wang [34]. Lo [35] formulates the standard error of the ratio under a range of different conditions and Ledoit and Wolf [36] suggest a robust hypothesis test using the ratio. Here we use average Sharpe ratios to summarise differences in performance but do not attempt any formal hypothesis testing.

As an example, we continue to use the case where the Berkowitz statistic is calculated over the previous six years (K = 78) using a rolling six years of data for estimation (M = 312). The Sharpe ratio is calculated for the in-sample and the out-of-sample returns. For each case, two average Sharpe ratios are calculated, one is averaged over portfolios in the consistency region and one is averaged over those portfolios outside the region (inconsistent). The in-sample Sharpe ratios, plotted in Figure 8, trend downwards due to market conditions.

The average in-sample Sharpe ratio for inconsistent portfolios is always greater than that for consistent portfolios. Thus an investor using the in-sample Sharpe ratio as an investment criterion is likely to choose an inconsistent portfolio. However if we look at the average out-of-sample Sharpe ratios the situation reverses. For most of the time, the average out-of-sample Sharpe ratio for the consistent portfolios exceeds that for the inconsistent portfolios. We can also see that the difference between the average in-sample and out-of-sample Sharpe ratios for consistent portfolios is much less than that for inconsistent portfolios. This greater accuracy in forecasting the Sharpe ratio for consistent portfolios is in line with our expectation that consistency implies better prediction of out-of-sample behaviour. The line plotted at the foot of Figure 8 shows the proportion of consistent portfolios and gives a continuous summary of the behaviour of the consistency region, discussed in Section 4.2 and shown in the plots 56 weeks apart in Figure 7.

In summary here then our results indicate that choosing to invest in consistent portfolios (if they exist) will lead to better out-of-sample Sharpe ratios.

We investigate two types of investment strategy making use of the methodology we have developed above. The first type of strategy exploits the consistency region. The second type of strategy uses the Berkowitz statistic to choose between portfolios on efficient frontiers generated under different conditions.

5.1 Consistency region based strategies

Here we choose a portfolio from the 'more efficient' consistency frontier boundary of the consistency region. As can be seen from Figure 7, the consistency frontier is not always a convenient parabolic shape like the efficient frontier. Using information summarised as in Figure 7, define the minimum mean, minimum variance, consistent (MMMVC) portfolio to be the portfolio lying at the bottom of the consistency region on the left-hand side. MMMVC can be categorised by the values of b and c associated with it. For each period, the latest return data is used to produce an updated version of Figure 7, from this we identify the b and c co-ordinates of MMMVC; the latest return data is then used to update the efficient frontier and the dominated portfolios; we invest in the (dominated) portfolio corresponding to the identified values of b and c. We chose this criterion as it relates to investing in the consistent portfolio most similar to the minimum variance portfolio. Indeed these investment strategies may coincide when the lower part of the efficient frontier belongs to the consistency region. Note that if this consistent portfolio is very close to maximum risk, we do not invest (for minimum risk $c=0$; for maximum risk $c=C$; and we do not invest if $c \geq 0.8C$). Investing in the minimum variance (MV) portfolio (which must by definition lie on the efficient frontier) at each period is the benchmark strategy.

In Figure 9, (continuing to use the M = 312 weeks and K = 78 as an example) we show the cumulative return accruing to investors following these strategies. If there are no consistent portfolios, as in early 2002, (see the lower plot in Figure 8) we do nothing. The cumulative return of the consistent strategy increases faster than the minimum variance strategy from March 2003 to January 2004. From May 2006, the strategies converge as the minimum variance portfolio becomes consistent (the line plotted at the bottom of Figure 9 is unity when the minimum variance portfolio is consistent, zero otherwise), they separate again briefly in July 2008.

The above analysis was repeated for different values of the estimation interval, M, and the number of out-of-sample returns, K, considered in the Berkowitz statistic, to discover their effect on the relative profitability of these investment strategies (MMMVC and MV). To gain an idea of the robustness of a strategy choice over time, cumulative returns over the first and second halves of the period are shown separately. The results are shown in Table 2.

In Table 2, the cumulative returns for strategies that match or exceed the MV benchmark in the investment interval are shown in bold, followed by a star (*). If the strategy outperforms the MV benchmark in the first half and then continues to outperform in the second interval (52 to 103), the cumulative returns are shown in bold italics followed by a star (*). For three out of the eight cases considered, investment in the MMMVC portfolio delivers excess return over MV over the first half of the evaluation period. For all three cases, the strategy continued to generate positive excess returns in the second half of the evaluation period.

Table 2. Using the Dow Jones data set - the cumulative return of the minimum mean, minimum variance consistent (MMMVC) portfolio as compared with that of the minimum variance (MV) portfolio. Covariance matrix estimated using method of Ledoit and Wolf. The investment horizon is four weeks (*H* = 4).

Estimation Interval (M)	**Period**	**Cumulative return (MV)**	**Cumulative return (MMMVC) Values of K**		
			65	**78**	**104**
260	2 to 51	0.065	0.040	-0.359	-0.170
	52 to 103	0.004	-0.220	0.013	0.035
312	2 to 51	0.049	0.036	**0.362***	**0.152***
	52 to 103	0.045	0.045	***0.050****	***0.078****
416	2 to 51	0.056	**0.166***	0.033	
	52 to 103	0.090	***0.168****	0.189	

Note: there is insufficient data in our time series to evaluate the case where M = 416 and K = 104.

5.2 Strategies using minimum variance portfolios ranked by the Berkowitz statistic

Here the strategy focuses only on the lower end of the efficient frontier, dominated portfolios are ignored. A set of minimum variance portfolios are considered, where the set is defined by the range of values of M and K considered, here we use the same experimental output as described in Table 2. The actual minimum variance portfolio depends only on M; however for a given M, there is one Berkowitz statistic for each value of K. The investment strategy is to invest in the case with the lowest Berkowitz statistic; the strategy can be varied by always investing, or investing only when the value of the Berkowitz statistic places the portfolio in the consistency region.

The results for a number of possible definitions of K are given in Table 3; the covariance matrix estimation was carried out using the method of Ledoit and Wolf. As a benchmark strategy, we chose investing in a minimum variance portfolio using $M = 260$, this value of M is optimal during the first investment interval (2 to 51), (see Table 2). As in Table 2, the cumulative returns for strategies that match or exceed the benchmark in the first investment interval are shown in bold followed by a star (*). If the strategy continues to outperform the benchmark in the second interval (52 to 103), the cumulative returns are shown in bold italics followed by a star (*).

As an example, we consider $M = 312$ and invest only if the lowest Berkowitz statistic (corresponding to the values of K) is consistent (row 3 in the table), the cumulative return is 0.338 in the first investment interval. This compares to 0.049 from the strategy ignoring consistency (row 18 in table). In this case, the portfolio is consistent throughout the second investment interval, so the return is the same (0.045) for both cases. If we consider the strategy where we take consistency into account and use $M = 416$ and $K = 65$, the cumulative return over the first investment interval is 0.065 and this is less than the benchmark.

Table 3. Using the Dow Jones data set - the cumulative return of the choice of minimum variance portfolio for different values of *M* and *K* is given for intervals 2 to 51 and 52 to 103. Covariance matrix estimated using method of Ledoit and Wolf. The investment horizon is four weeks (*H* = 4).

Definition of Set				Cumulative Return	
K	M	Consistent	Number in set	2 to 51	52 to 103
any	any	yes	8	**0.354***	***0.076****
any	260	yes	3	**0.361***	***0.004****
any	312	yes	3	**0.338***	***0.045****
any	416	yes	2	**0.142***	***0.091****
65	any	yes	3	**0.296***	-0.006
78	any	yes	3	**0.146***	***0.058****
104	any	yes	2	**0.078***	-0.003
65	260	yes	1	**0.296***	-0.070
65	312	yes	1	**0.254***	-0.034
65	416	yes	1	0.065	0.043
78	260	yes	1	**0.089***	***0.004****
78	312	yes	1	**0.087***	***0.045****
78	416	yes	1	**0.142***	***0.091****
104	260	yes	1	**0.078***	***0.004****
104	312	yes	1	**0.101***	***0.045****
any	any	no	8	**0.069***	***0.076****
any	260	no	3	**0.065***	***0.004****
any	312	no	3	0.049	0.045
any	416	no	2	0.056	0.091
65	any	no	3	0.055	0.069
78	any	no	3	**0.070***	***0.058****
104	any	no	2	**0.065***	-0.003

Note: any implies *K* is either 65, 78 or 104; or *M* is either 260, 312 or 416.

We see that in most cases, it is preferable to invest in the MV portfolio only when it is consistent. For the strategies based on consistent portfolios, 14 out of 15 outperform the benchmark in the first investment interval; of these 14, 10 continue to outperform in the second investment interval. For the strategies ignoring consistency, 4 out of 7 outperform the benchmark in the first investment interval; three of these four continue to outperform in the second investment interval.

In summary here then we have seen above that in nearly all cases use of the methodology that we have developed to identify whether (or not) portfolios are consistent, and subsequently choosing to invest in consistent portfolios, enables us to make investment choices that outperform the MV benchmark.

5.3 Using a larger universe of assets – the Japanese Nikkei 225

The empirical analysis so far has been based on the relatively small universe of the Dow Jones; in order to explore the scalability of the results seen so far a larger universe is now used. The time period used for this universe is the same as that for the Dow Jones data, January 1988 to November 2009. As before, we use the Ledoit and Wolf method for variance estimation; the analyses leading to Tables 2 and 3 are repeated and the results shown in Tables 4 and 5 respectively.

Table 4. Using the Nikkei data set - the cumulative return of the minimum mean, minimum variance consistent (MMMVC) portfolio as compared with that of the minimum variance (MV) portfolio. Covariance matrix estimated using the method of Ledoit and Wolf. The investment horizon is four weeks (*H* = 4).

Estimation Interval (M)	Period	Cumulative return (MV)	Cumulative return (MMMVC) Values of K		
			65	78	104
260	2 to 51	0.188	**0.284***	**0.284***	**0.239***
	52 to 103	-0.265	***-0.204****	***-0.199****	-0.287
312	2 to 51	0.190	**0.264***	**0.278***	**0.216***
	52 to 103	-0.296	***-0.197****	***-0.212****	***-0.099****
416	2 to 51	0.171	**0.189***	**0.171***	
	52 to 103	-0.277	***-0.244****	***-0.223****	

In Table 4, we see that, during the first investment period, for all 8 combinations of M and K, the MMMVC portfolio outperforms the corresponding strategy of investing in the minimum variance (MV) portfolio. In 7 of these 8 cases, outperformance is carried through to the second investment period. In Table 5, there are fewer cases (7 out of 15 where consistency is required, 1 out of 7 where it is not) where the cumulative return during the first investment interval outperforms the benchmark. In all of the cases where consistency is required, outperformance is carried through to the second investment interval. For the single case ignoring consistency, outperformance is not achieved in the second investment interval.

In summary here then we have shown that, in this case with the larger universe, knowledge of the consistency (or not) of portfolios allows outperformance of the MV benchmark.

Table 5. Using the Nikkei data set - the cumulative return of the choice of minimum variance portfolio for different values of *M* and *K* is given for intervals 2 to 51 and 52 to 103. Covariance matrix estimation for the minimum variance portfolio uses the method of Ledoit and Wolf. The investment horizon is four weeks (*H* = 4).

Definition of Set				Cumulative Return	
K	M	Consistent	Number in set	2 to 51	52 to 103
any	any	yes	8	0.167	-0.156
any	260	yes	3	**0.191***	***0.307****
any	312	yes	3	**0.209***	***-0.215****
any	416	yes	2	0.171	-0.148
65	any	yes	3	**0.222***	***-0.143****
78	any	yes	3	0.171	-0.156
104	any	yes	2	**0.222***	***-0.059****
65	260	yes	1	0.168	0.307
65	312	yes	1	0.147	-0.159
65	416	yes	1	**0.225***	***-0.154****
78	260	yes	1	0.129	0.293
78	312	yes	1	0.162	-0.230
78	416	yes	1	0.171	-0.148
104	260	yes	1	**0.230***	***0.000****
104	312	yes	1	**0.205***	***-0.059****
any	any	no	8	0.167	-0.285
any	260	no	3	0.188	-0.265
any	312	no	3	0.190	-0.296
any	416	no	2	0.171	-0.277
65	any	no	3	0.168	-0.266
78	any	no	3	0.171	-0.285
104	any	no	2	**0.207***	-0.296

5.4 Investigating the consistency of naive diversified portfolios; the 1/N portfolio

It is an obvious point that optimised portfolios, including the minimum variance portfolio, tend to weight assets with desirable characteristics more heavily than others. For example, for the minimum variance portfolios computed for the Dow Jones data, the maximum weight on an asset had an inter-quartile range of 0.23 to 0.28. The comparable range for the Nikkei data was 0.13 to 0.21. If the portfolios had equally weighted assets, then for the Dow Jones with N=28 the weight on each asset would be 1/28=0.036 and for the Nikkei with N=191 the weight on each asset would be 1/191=0.005. The implication here then is that the returns on a minimum variance portfolio will tend to be influenced by a few assets and are thus likely to be more idiosyncratic than the returns on an equally weighted portfolio.

In an extensive comparison of models over seven data sets, DeMiguel, Garlappi and Uppal [21] found that no optimisation based model regularly outperformed the simple strategy of investing equally in each asset in the universe considered. The motivation for using this equally weighted *1/N* strategy is to avoid reliance on past data and to spread risk as widely as possible. It is important to note that the equally weighted portfolio can be very closely correlated with the market index. For our two data sets the correlation between the weekly returns for the index and those for the *1/N* portfolio was 0.97 for the Dow Jones and 0.95 for the Nikkei (in other words for the data we examined the *1/N* portfolio is effectively a closet index tracker).

Here we use probability density forecasting in the same way as in the previous sub-sections to investigate if the consistency of the portfolio offers useful information. Although the calculation of the returns of the *1/N* portfolio is trivial, the determination of consistency still depends on the estimation interval, *M*, and the number of periods, *K*, considered in the Berkowitz statistic. The value of *M* does not affect the portfolio returns, but affects the estimation of the *cdf*s. The cumulative returns are calculated for the straightforward *1/N* strategy, a comparison is made between the cumulative returns achieved if investment is always made or made only if the most recent Berkowitz statistic indicates consistent behaviour. The results are given in Table 6.

In Table 6, during the first investment period, we see that, for 7 of the 8 combinations of *M* and *K*, the consistent *1/N* portfolio outperforms the corresponding strategy of always investing in the *1/N* portfolio. In 5 of these 7 cases, outperformance is carried through to the second investment period. We also found that if we substituted the Dow Jones index for the *1/N* portfolio (i.e. examined a portfolio that exactly replicated the index), we achieved similar results.

In the case of the *1/N* strategy applied to the Nikkei data (where *N* is 191), the Berkowitz statistic remained low in value for all combinations of *M* and *K* so that the portfolio never became inconsistent. The cumulative return in the first investment period was 0.534 and -0.382 during the second investment period. We found that the Nikkei 225 index (more precisely a portfolio that exactly replicated the index) exhibited the same consistent behaviour as the *1/N* portfolio, i.e. the Berkowitz statistic remained insignificant throughout the time period considered.

Table 6. Using the Dow Jones data set - the cumulative return of the consistent *1/N* portfolio as compared with that of the *1/N* portfolio. The investment horizon is four weeks (*H* = 4).

Estimation Interval (M)	Period	Cumulative return *1/N*	Cumulative return Consistent *1/N* Values of K: 65	78	104
260	2 to 51	-0.003	**0.122***	**0.007***	-0.009
	52 to 103	0.002	-0.265	-0.299	-0.299
312	2 to 51	-0.003	**0.079***	**0.007***	**0.000***
	52 to 103	0.002	***0.119****	***0.318****	***0.002****
416	2 to 51	-0.003	**0.068***	**0.007***	
	52 to 103	0.002	***0.383****	***0.318****	

6. Conclusions

We have introduced the concept of a consistency region in risk-return space, where a portfolio's in-sample performance is a reliable predictor of its out-of-sample return. We have set up a methodology where the measurement of density forecasting accuracy can be carried out for a series of portfolio selections in the same relative position to the efficient frontier. The methodology does not depend on a specific portfolio selection algorithm; it merely requires an identification of an efficient frontier in risk-return space.

We validated the concept using well-behaved data generated to be time-invariant multivariate normal asset returns and showed that the consistency region behaves in an intuitively reasonable manner. For short estimation periods, the consistency region is composed of dominated portfolios with higher standard deviations than those on the efficient frontier with the same mean returns. As the estimation period is increased, the consistency region increases and gradually moves towards, and then moves up, the efficient frontier. When used with real data, the behaviour of the consistency region reflects the changing dynamics of asset returns. This means that for the same size estimation period, the consistency region may fluctuate between non-existence and comprising most of the efficient frontier.

As evidence that the return density function of consistent portfolios is estimated more accurately than those of inconsistent portfolios, we showed that the out-of-sample Sharpe ratio was, on average, more accurately predicted by the in-sample ratio for consistent portfolios than for inconsistent portfolios. In other words, the consistency of a portfolio implies more accurate density forecasts, which in turn implies more accurate prediction of functions of density parameters, such as the Sharpe ratio.

Two strategies for exploiting the consistency of portfolios were developed: one investing in a consistent portfolio within the efficient frontier; the other investing in the minimum variance portfolio, but only when it is consistent. We showed that, in nearly all the cases where outperformance of the benchmark occurred in the first half of the period considered, outperformance persisted in the second half of the investment period. We also investigated the consistency of the naive diversified (*1/N*) portfolio.

Appendix A

Consider a set of returns, $R_1,\ldots,R_n$, which form the in-sample data; the out-of-sample return is R_{n+1}. We define the following from the in-sample data: $R'_1=\min\left(R_1,\ldots,R_n\right)$; $R'_n=\max\left(R_1,\ldots,R_n\right)$ and R'_L is the L^{th} observed return in ascending order of magnitude. We also define $\hat{\mu}_R$ and $\hat{\sigma}_R$ as the sample mean and standard deviation of $R_1,\ldots,R_n$. For the calculation of the empirical *cdf*, we consider two cases. The first case is when the out-of-sample return falls within the in-sample range, $R'_1<R_{n+1}<R'_n$:

$$\hat{F}(R_{n+1}\mid\Theta)=\frac{L+\frac{1}{2}+\left(\frac{R_{n+1}-R'_L}{R'_{L+1}-R'_L}\right)}{n+1}$$

where there are L in-sample observations less than R_{n+1} and R'_L is the greatest in-sample observation less than R_{n+1}. The second case is when R_{n+1} falls outside of the in-sample range, $R_{n+1}\le R'_1$ or $R_{n+1}\ge R'_n$.

$$\hat{F}(R_{n+1}\mid\Theta)=\begin{cases}\dfrac{1}{2\left(n+1\right)}\left(\dfrac{\Phi\left(\dfrac{R_{n+1}-\hat{\mu}_R}{\hat{\sigma}_R}\right)}{\Phi\left(\dfrac{R'_1-\hat{\mu}_R}{\hat{\sigma}_R}\right)}\right) & \text{if } R_{n+1}\le R'_1\\[2ex] 1-\dfrac{1}{2\left(n+1\right)}\left(\dfrac{1-\Phi\left(\dfrac{R_{n+1}-\hat{\mu}_R}{\hat{\sigma}_R}\right)}{1-\Phi\left(\dfrac{R'_n-\hat{\mu}_R}{\hat{\sigma}_R}\right)}\right) & \text{if } R_{n+1}\ge R'_n\end{cases}$$

In this case, we use the normal distribution to approximate the behaviour of the out-of-sample return.

Appendix B

For different values of M and K, the values of the Berkowitz statistic under the null hypothesis were simulated 10000 times and the 90, 95 and 99 percentiles of the statistic were found (let Q_α be the $100(1-\alpha)$ percentile for $\alpha = 0.10, 0.05, 0.01$). Following linear regression of $\ln\left(Q_\alpha / \chi^2_{2,\alpha} \right)$ on different powers of M and K, the following determination of the critical values was made; in the range considered the value of K was not significant.

$$\textit{Critical value}(\alpha) = \chi^2_{2,\alpha} \exp\left(-0.0276 + 1.289 / \sqrt{M} \right).$$

References

[1] Markowitz, H., 1952. Portfolio selection. *Journal of Finance*, 7(1), 77–91.
[2] Michaud, R.O., 1989, The Markowitz optimization enigma: is 'optimized' optimal? *Financial Analysts Journal*, 45(1), 31-42.
[3] Broadie, M., 1993, Computing efficient frontiers using estimated parameters. *Annals of Operations Research*, 45(1), 21-58.
[4] James, W. and C. Stein, 1961, Estimation with quadratic loss. *Proceedings of the 4th Berkeley Symposium on Probability and Statistics 1*. Berkeley: University of California Press, 361-379.
[5] Jobson, J.D. and B. Korkie, 1980, Estimation for Markowitz efficient portfolios, *Journal of the American Statistical Association*, 75(371), 544-554.
[6] Jorion, P., 1986, Bayes-Stein estimation for portfolio analysis. *The Journal of Financial and Quantitative Analysis*, 21(3), 279-292.
[7] Chopra, V.K., C.R. Hensel and A.L. Turner, 1993, Massaging mean-variance inputs: returns from alternative global investment strategies in the 1980s. *Management Science*, 39(7), 845-855.
[8] Fabozzi, F.J., D.S. Huang, and G.F. Zhou, 2010, Robust portfolios: contributions from operations research and finance, *Annals of Operations Research*, 176(1), 191-220.
[9] Jorion, P., 1991, Bayesian and CAPM estimators of the means: Implications for portfolio selection, *Journal of Banking and Finance*, 15(3), 717-727.
[10] Ledoit, O. and M. Wolf, 2003, Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance,* 10(5), 603-621.
[11] Tu, J. and G.F. Zhou, 2004, Data-generating process uncertainty: What difference does it make in portfolio decisions? *Journal of Financial Economics*, 72(2), 385-421.
[12] Jagannathan, R. and T.S. Ma, 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 58(4), 1651-1683.
[13] Frankfurter, G.M., H.E. Phillips and J.P. Seagle, 1971, Portfolio selection, the effects of uncertain means, variances, and covariances. *Journal of Financial and Quantitative Analysis*, 6(5), 1251-1262.
[14] Bloomfield T., R. Leftwich and J.B. Long, 1977, Portfolio strategies and performance. *Journal of Financial Economics*, 5(2), 201-218.
[15] Jorion, P., 1985, International portfolio diversification with estimation risk. *Journal of Business*, 58(3), 259-278.
[16] Chan, L.K.C., J. Karceski and J. Lakonishok, 1999, On portfolio optimization: forecasting covariances and choosing the risk model. *The Review of Financial Studies*, 12(5), 937-974.
[17] Polson, N.G. and B.V. Tew, 2000, Bayesian portfolio selection: An empirical analysis of the S&P 500 index 1970-1996, *Journal of Business & Economic Statistics*, 18(2), 164-173.
[18] Wang, Z.Y., 2005, A shrinkage approach to model uncertainty and asset allocation, *The Review of Financial Studies*, 18(2), 673-705.
[19] Disatnik, D.J. and S. Benninga, 2007, Shrinking the covariance matrix, *Journal of Portfolio Management*, 33(4), 55-63.
[20] DeMiguel, V., L. Garlappi, F.J. Nogales and R. Uppal, 2009, A generalized approach to portfolio optimization: improving performance by constraining portfolio norms, *Management Science*, 55(5), 798-812.
[21] DeMiguel, V., L. Garlappi, and R. Uppal, 2009, Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?, *The Review of Financial Studies,* 22(5), 1915-1953.
[22] Corradi V. and N.R. Swanson, 2006, Predictive density and conditional confidence interval accuracy tests, *Journal of Econometrics,* 135(1-2), 187-228.
[23] Diebold F.X., T.A. Gunther and A.S. Tay, 1998, Evaluating density forecasts with applications to financial risk management, *International Economic Review*, 39(4), 863-883.
[24] Hong, Y.M. and H.T. Li, 2005, Nonparametric specification testing for continuous-time models with applications to term structure of interest rates, *The Review of Financial Studies*, 18(1), 37-84.
[25] Hong Y.M., H.T. Li and F. Zhao, 2007, Can the random walk model be beaten in out-of-sample density forecasts? Evidence from intraday foreign exchange rates, *Journal of Econometrics*, 141(2), 736-776.

[26] Mitchell, J. and S.G. Hall, 2005, Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR 'fan' charts of inflation. *Oxford Bulletin of Economics and Statistics*, 67, Supplement s1, 995-1033.
[27] Bao, Y., T-H. Lee and B. Saltoglu, 2007, Comparing density forecast models. *Journal of Forecasting*, 26(3), 203-225.
[28] Berkowitz, J., 2001, Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19(4), 465-474.
[29] Schittkowski, K. 1986, NLPQL: A FORTRAN subroutine solving constrained nonlinear programming problems, *Annals of Operations Research*, 5(1-4), 485-500.
[30] Blattberg, R.C. and N.J. Gonedes, 1974, A comparison of the stable and Student distributions as statistical models for stock prices, *Journal of Business*, 47(2), 244-280.
[31] Kon, S.J., 1984, Models of stock returns - a comparison, *Journal of Finance*, 39(1), 147-165.
[32] Barndorff-Nielsen, O.E., 1997, Normal inverse Gaussian distributions and stochastic volatility modelling, *Scandinavian Journal of Statistics*, 24(1), 1-13.
[33] Meade, N., 2010, Oil prices - Brownian motion or mean reversion? A study using a one year ahead density forecast criterion, *Energy Economics*, 32(6), 1485-1498.
[34] Garlappi, L., R. Uppal and T. Wang, 2007, Portfolio selection with parameter and model uncertainty: A multi-prior approach, *The Review of Financial Studies*, 20(1), 41-81.
[35] Lo, A.W., 2002, The statistics of Sharpe ratios, *Financial Analysts Journal*, 58(4), 36-52.
[36] Ledoit, O. and M. Wolf, 2008, Robust performance hypothesis testing with the Sharpe ratio, *Journal of Empirical Finance*, 15(5), 850-859.

Figure 1. An efficient frontier constructed using a universe of Dow Jones stocks. The upper plot shows the frontier alone, the lower plot shows the return density functions (assuming normality) superimposed.

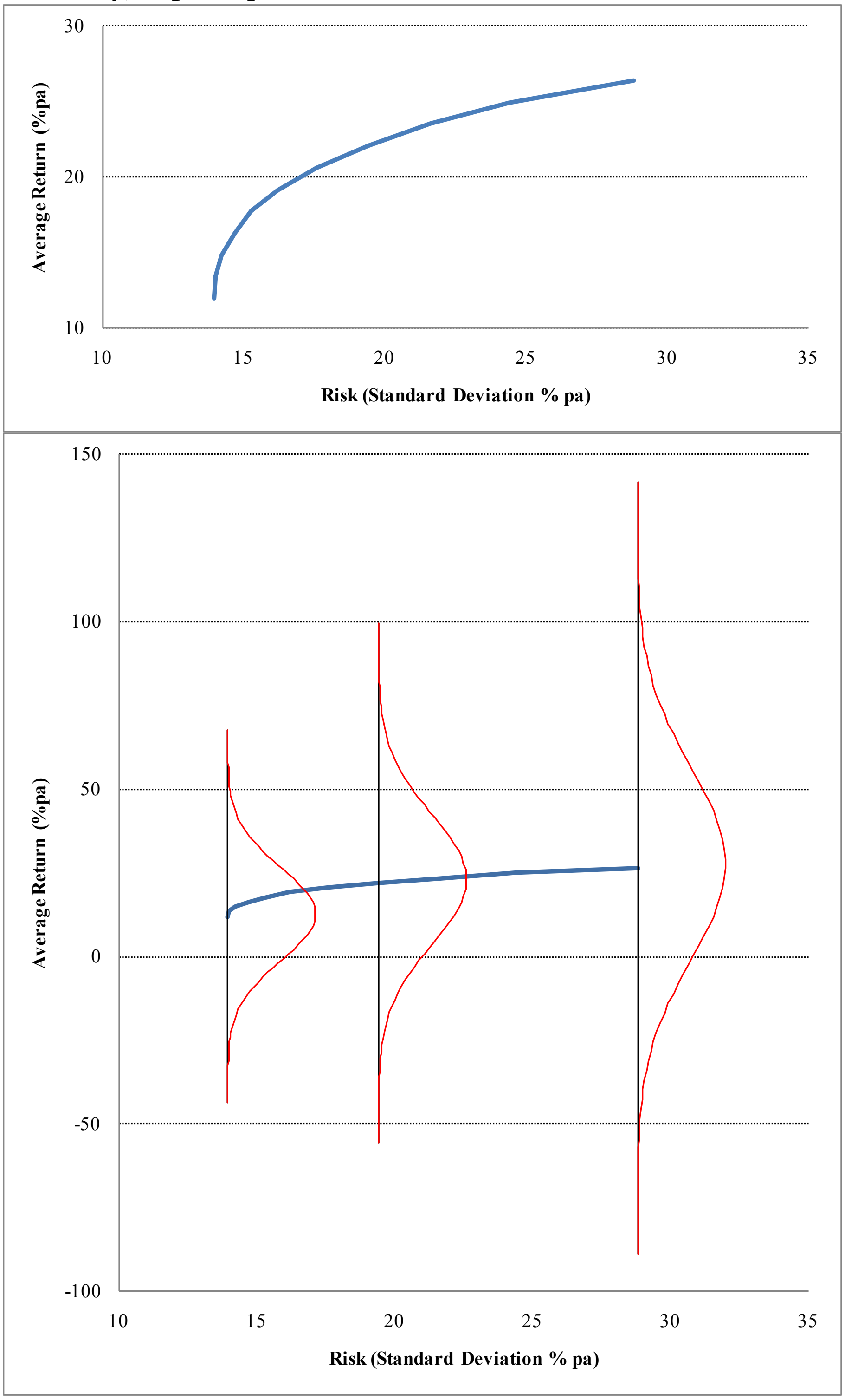

Figure 2. The efficient frontier and the dominated portfolios in risk-return space for a universe of $N = 28$ assets, and estimation regions of $M = 260$ (5 years).

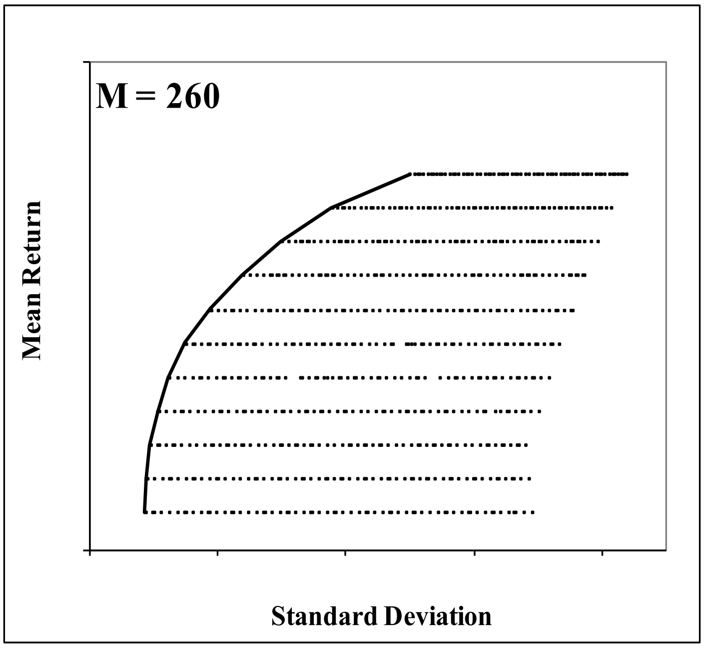

Figure 3. The power of the Berkowitz statistic to detect departures from the null hypothesis (for $K = 78$ and $M = 260$).

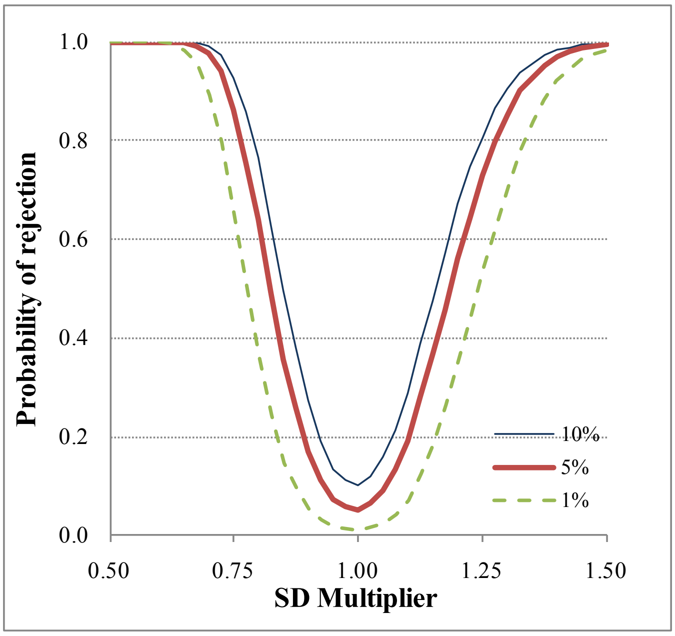

Figure 4. Summary efficient frontiers over K=78 origins for a range of M estimation regions using a multivariate normal data set. *The symbols indicate the significance of the Berkowitz statistic, a (red) dash indicates that the consistency hypothesis can be rejected at 10% significance, a (green) cross indicates acceptance.*

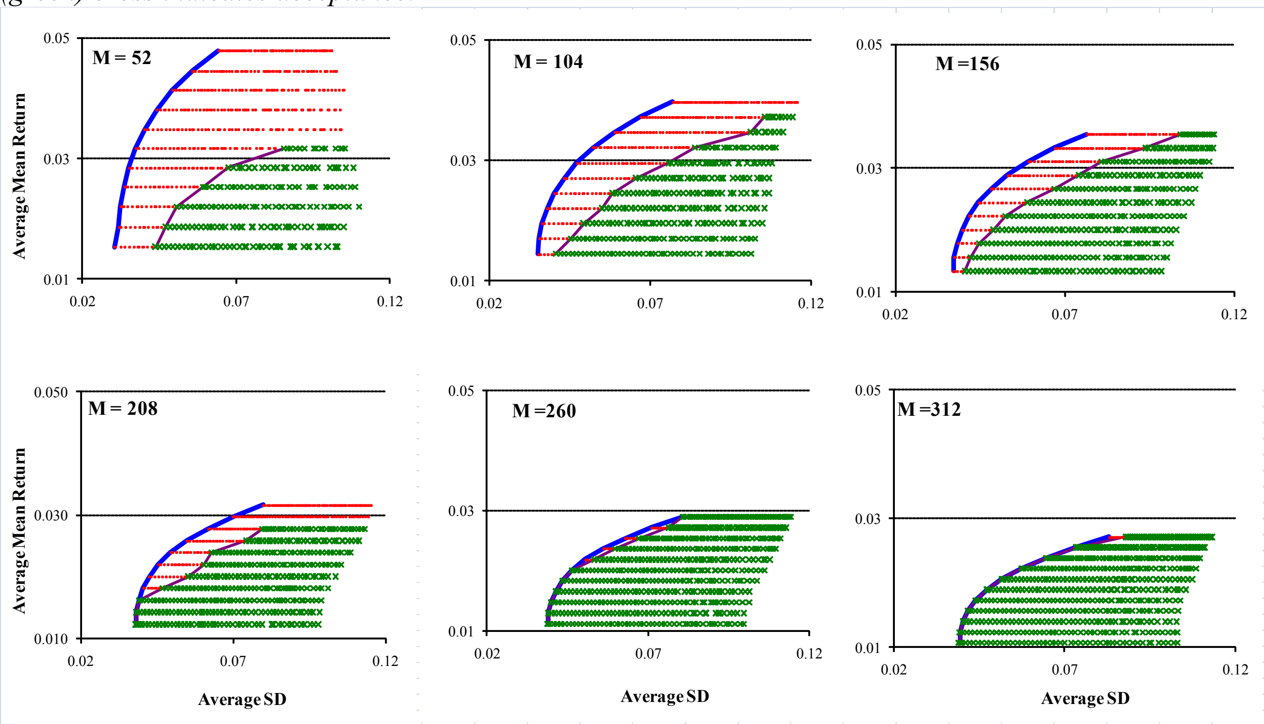

Figure 5. The Dow Jones index from January 1988 to November 2009. The index value and a rolling 30 day volatility are shown. The evaluation dates used are indicated.

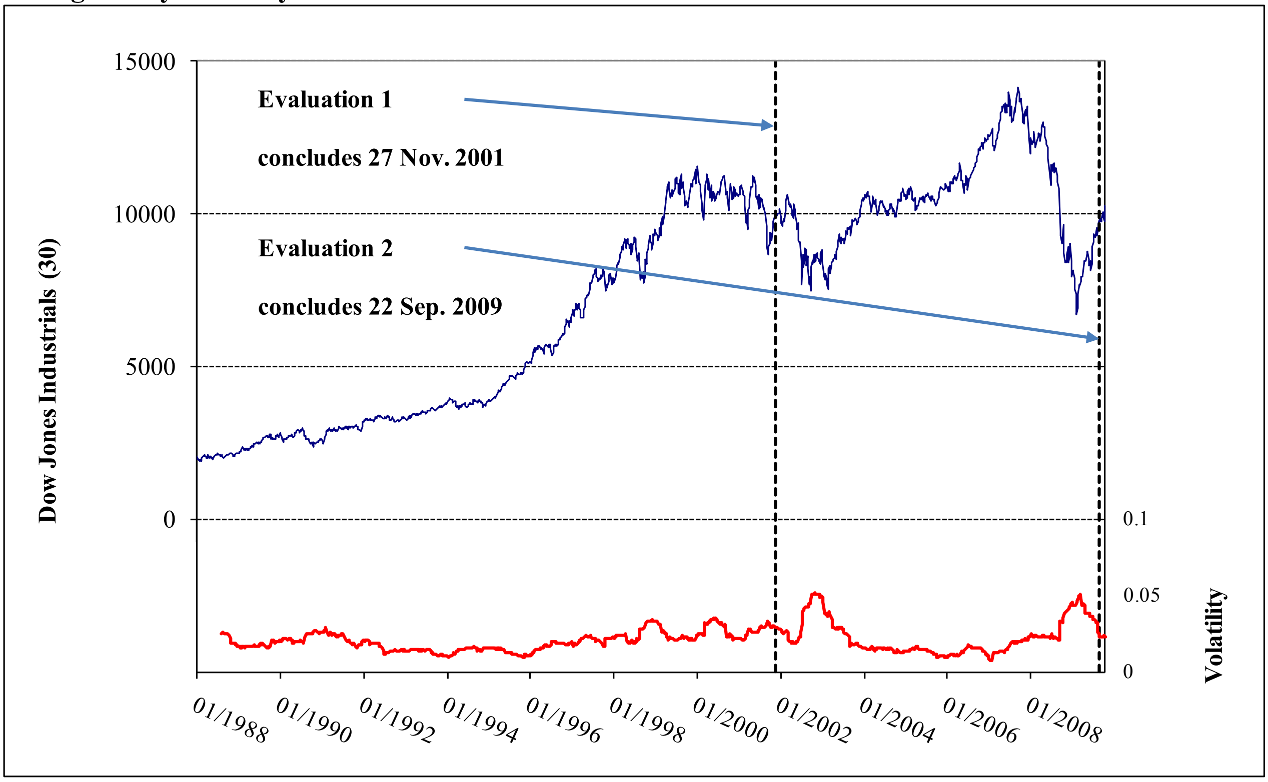

Figure 6. Summary efficient frontiers for a universe of 28 members of the Dow Jones over *K*=78 origins for a range of *M* (=260, 312, 416) estimation regions for the two evaluation dates defined in Figure 5. *(The symbols indicate the significance of the Berkowitz statistic, a (red) dash indicates that the consistency hypothesis can be rejected at 10% significance, a (green) cross indicates acceptance.)*

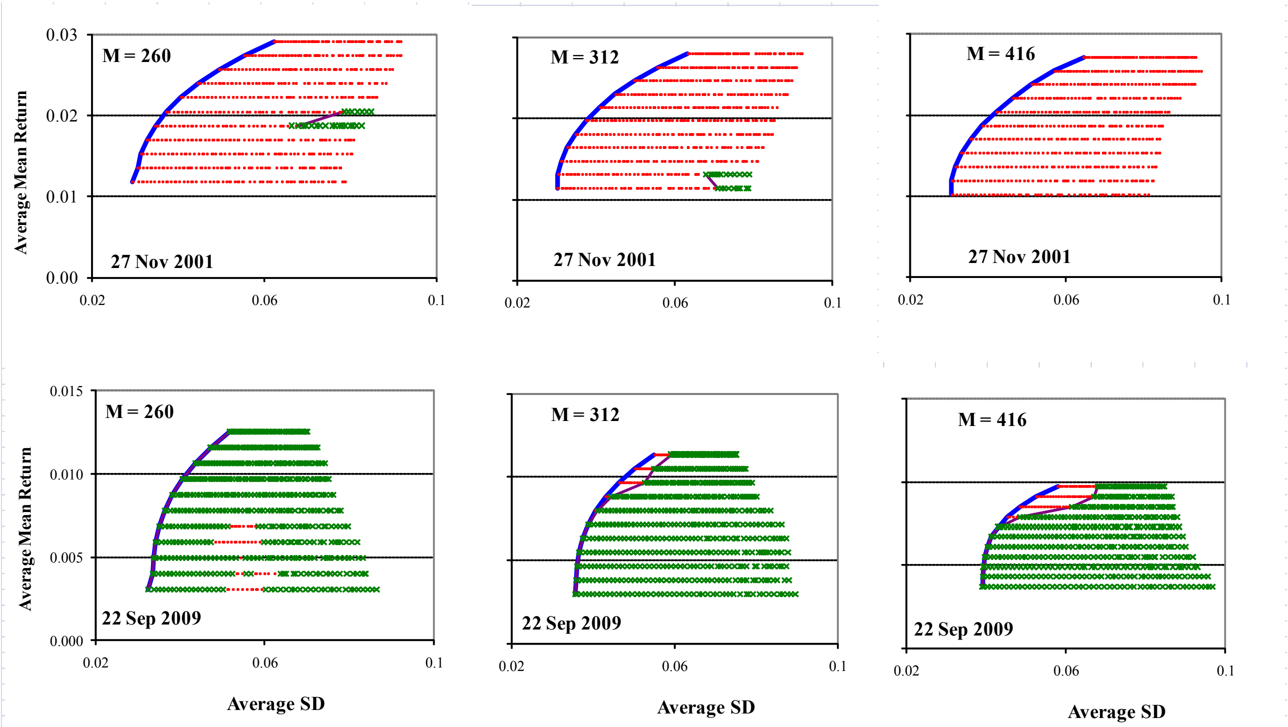

Figure 7. Summary efficient frontiers over K=78 origins for M=312 evaluated for evenly spread origins between 27/11/2001 and 22/09/2009 using Dow Jones data. *(The symbols indicate the significance of the Berkowitz statistic, a (red) dash indicates that the consistency hypothesis can be rejected at 10% significance, a (green) cross indicates acceptance.)*

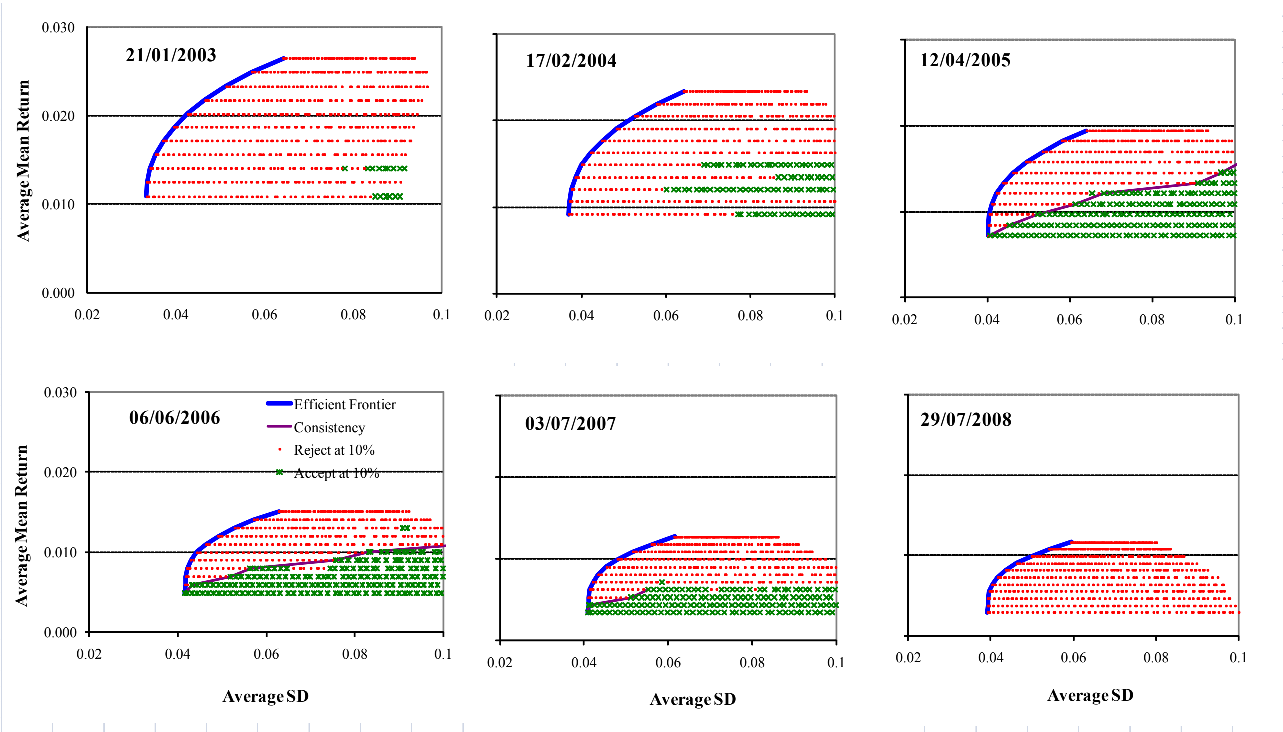

Figure 8. Average Sharpe ratios for inconsistent and consistent portfolios evaluated for in-sample (IS) and out-of-sample (OS) returns. The values associated with each date refer to the previous 78 periods of 4 weeks (K = 78, M = 312). The lower plot shows the percentage of all portfolios considered that were consistent.

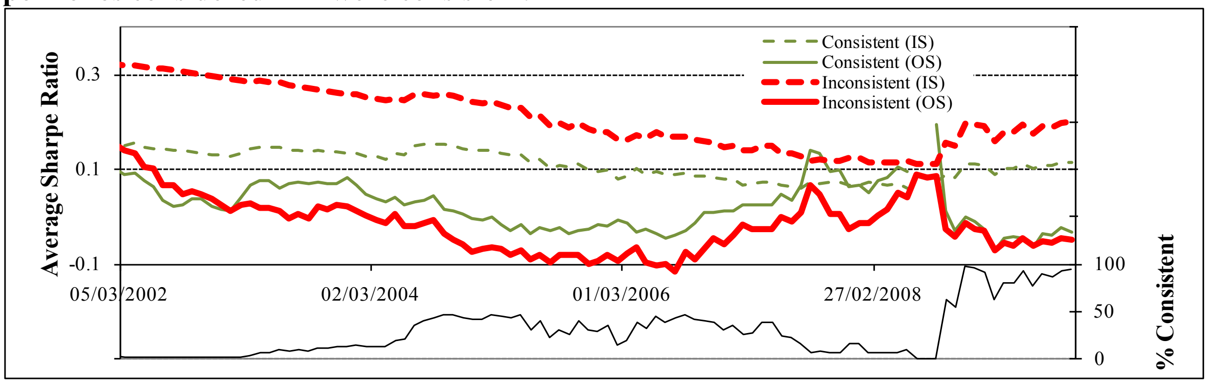

Figure 9. The cumulative returns of the minimum mean, minimum variance consistent portfolio and the minimum variance portfolio (K = 78, M = 312).