Portfolio Investment Modeling Using High Frequency Data

By

David E. Allen\textsuperscript{1} and Seyed-Ali Hosseini-Yekani\textsuperscript{2}

\textsuperscript{1}School of Accounting, Finance and Economics, Edith Cowan University
\textsuperscript{2}Department of Agricultural Economics, Shiraz University

Correspondence author:

David E. Allen
School of Accounting, Finance and Economics
Faculty of Business and Law
Edith Cowan University
Joondalup, WA 6027
Australia
Phone: +618 6304 5471
Fax: +618 6304 5271
Email: d.allen@ecu.edu.au
ABSTRACT

In this paper the contribution and advantages of utilizing high frequency data for optimal portfolio selection purposes are investigated via the application of three different risk measures: Variance, VaR and CVaR in a portfolio selection risk-return framework. A new computational method for the calculation of VaR with smaller dimensions and the additional capability of simultaneously calculating CVaR, which is necessarily more efficient in large scale optimisation, is proposed. The results show that the effectiveness of VaR as a risk measure in optimal portfolio selection is demonstrated to be more apparent than Variance and CVaR, especially in the case of this particular high frequency data set. The results confirm the advantages of utilising the information inherent in high frequency data for short term investment horizons. The actual losses of the portfolios obtained by using historical data are bigger than in the parallel situations of those chosen using high frequency data in all cases.

Keywords: Portfolio selection; VaR; CVaR; High frequency data
1. Introduction

The availability of intraday databases of financial trading information has stimulated an enormous amount of interest and research into the analysis and modeling of high frequency data. Engle (2000) refers to 'ultra high frequency' data in the limiting case when all events are recorded. This availability has opened up a number of interesting fields in financial economics and financial econometrics. The fact that trades are unequally intervalled in time and stochastic has lead to some interesting statistical challenges, see for example, Ait-Sahalia and Mykland (2003). New ways of modeling realized volatility have been developed, see Barndorff-Neilsen and Shephard (2002). Several methods have recently been proposed in the ultra high frequency financial literature to remove the effects of microstructure noise and to obtain consistent estimates of the integrated volatility as a true measure of daily volatility; see for example, Bandi and Russell (2006) and Asai, McAleer and Medeiros (2007). The fact that sequences of prices and trades are auto correlated led Engle and Russell (1994, 1998) to develop the ACD model which facilitates the measurement and forecasting of durations between price changes and trades in a manner analogous to the forecasting of volatility using GARCH models.

Given the above interest in obtaining better and more immediate estimates of volatility from ultra high frequency data, a natural corollary is to address estimates of the covariance structure and portfolio estimation using such data. Engle (2002) and Cappiello, Engle and Shephard (2006) have approached the former using the DCC model and in the latter paper, introduce the AG-DCC process which extends previous specifications along two dimensions: it allows for series-specific news impact and smoothing parameters and permits conditional asymmetries in correlation dynamics.

Much less has been accomplished on the analysis of portfolios using high frequency data. Engle, Ferstenberg and Russell (2006) have analysed execution costs in a manner analogous to portfolio analysis using a notion of liquidation value at risk (LVAR). Furthermore, emphasis on the analysis of portfolios using ‘down-side risk’ has become more apparent in the last couple of decades. See for example Sortino and Satchell (2001). This approach has been popularized under the acronym ‘post-modern portfolio theory’. A theoretical culmination of this approach can be seen in the work of Rockafellar, Uryasev and Zabarankin (2006) who derive CAPM type general deviation measures using these metrics in a ‘master fund’ context. This leads on to a natural intersection of the use of these types of portfolio selection metrics in the context of high frequency data sets. This is the
principle focus of the current paper; which is to engage in portfolio analysis using high frequency data. This involves an emphasis on relatively short-term investment horizons.

Investors and speculators might invest for either long term or short term gains. Investors typically use a variety of decision aids including fundamental analysis to assess intrinsic values which may give an indication of longer-term equilibrium values. In the short-term, speculators with an information advantage, real or assumed, may be investing for gains over much shorter time horizons. They contribute to market liquidity and use a variety of tools including various forms of technical analysis, various charting techniques, and different kinds of Moving Average analysis techniques. These approaches focus on individual stocks rather than portfolios. Investors also have to make decisions about their portfolio selection strategies in the short or long-term.

Optimal portfolio selection metrics using a variety of risk-return approaches are the customary approach. This might involve maximizing return at a given level of risk or minimizing risk at a target level of return. The degree of risk aversion of the investors concerned will also impact on the optimum choice and also condition the choice of risk measure. The first and most important step in calculation of these risk measures is specifying the sample probability density function of loss in the considered investment period. There are various simulation, forecasting and approximation methods available for specifying the sample PDF of losses. These are also dependent on our knowledge and assumptions about the distribution of the prices data being utilised. Sample PDFs can be determined by making different probable scenarios of prices conditions/outcomes in the future. The accuracy of optimal portfolio selection outcomes is obviously dependent on the accuracy of the specification of the sample PDF. There is not much gain to the adoption of a complex and precise risk-return portfolio selection metric if there is a misspecification in the base-line sample PDF adopted.

It is customary to use historical price data for optimal portfolio selection because price data often doesn’t follow a specific distribution pattern. However, in the short term, historical data is less suitable for optimal portfolio selection given short term investment horizons. Using 6 months or one year’s historical data might be suitable for an investment period of 1-3 months in terms of recognizing the primary trends (the tides), to use a nautical analogy, but this type of data may not illustrate the secondary trends (the waves) needed for effective short term investing. For example if the investment period is the next 3-5 days, utilizing historical data older that last two weeks might not yield much accurate
information. The most important function of risk-return approaches in optimal portfolio selection is the means they afford for controlling the unsystematic risk of portfolio by minimization of various risk measures. Short term investors are affected by short term fluctuations which may largely be driven by the unsystematic risk of shocks to stocks supply and demand. If long term historical data is used for portfolio selection purposes then the systematic risk of the stock market will be added to unsystematic risk of the portfolio. These strong systematic risks could affect the amount of risk measures by decreasing the contribution of unsystematic risk on them. This contribution could be hidden under the strength of systematic risks and the selected portfolio using such risk measures may not be really optimal.

On the other hand there isn’t enough new historical data in order to facilitate scenario making required for short term investing. For example, in optimal portfolio selection for an investment horizon of the next 3-5 days, there are not more than 10 price observations for each stock in the last two weeks of historical data and this is insufficient for constructing a sample PDF for this period. The insufficiency of new observation price data is partly because of the customary summarizing of all transaction data in one daily index such as the closing price. Summarizing all available high frequency data in terms of a closing price also omits plenty of useful intra-day information for specifying the appropriate sample PDF relevant to short term investing. There is a lot of information in high frequency data about intra day price changes stemming from intra day supply and demand changes which is potentially relevant to short term investing but which is not available in the closing price data. Intra day conditions relating to a stock’s supply and demand changes are very important for forecasts of the behavior of sellers and buyers over the next few days.

In this paper, we attempt to test the contribution of the use of high frequency data sets for choosing optimal portfolio for short term investors. We use this data and select optimal portfolios using three different risk measures: the Variance, Value-at-Risk (VaR) and Conditional VaR (CVaR) of the portfolios as three alternative measures of risk in a risk-return framework.

Many studies have focused on application of these risk measures to optimal portfolio selection problems. For example see Allen (2005), Alexander et al. (2006), Campbell et al. (2001), Consigly (2002), Duffie and Pan (2001), Fusai and Luciano (2001), Gaivoronski and Pflug (2005), Kluppelberg and Korn (1998), Rockafeller and Uryasev (2000),

We undertake optimal portfolio selection analyses for short investment time horizon periods using both daily and high frequency data using each of these three measures and the results are then compared. We also propose a new computational method for estimating VaR in an optimal portfolio selection context which is more efficient than previous methods used; especially in the case of large scale optimizations. This is the major contribution of this paper. In addition, this method also provides a means of estimating the CVaR levels of optimal portfolio which can be calculated simultaneously.

2. Methodology

Consider an investor who wants to select an optimal portfolio between \( m \) \((i = 1,2,\ldots,m)\) stocks for investing in time horizon \( \bar{T} \). This investor could select different positions for decision vector \( X \in R^t \) such as:

\[
X = (X_1, X_2,\ldots,X_m)
\]

(1)

where \( X_i \geq 0 \) shows that all positions taken are long. If the initial prices of these stocks are \( p = (p_1, p_2,\ldots,p_m) \) the initial value of selected portfolio is determined by the investor’s budget limit:

\[
X^T p = \nu
\]

(2)

The subsequent prices of the selected stocks over the next few days are unknown quantity for investors. These prices at the end of investing time horizon \( \bar{T} \) might be:

\[
Y = (Y_1, Y_2,\ldots,Y_m)
\]

Then the investor is confronted by a random price vector of \( Y \in R^n \) in his/her optimization. Assuming the rationality of this investor, he/she looks for a portfolio with a low probability level of loss. The amount of loss is a function of both decision vector and the market price vector:

\[
\Lambda = f(X,Y)
\]

This function could be shown to be something like:

\[
\Lambda = \nu^{-1}(\nu - X^TY) = \frac{\nu - \sum_{i=1}^{m} X_i Y_i}{\nu}
\]

(3)
Then, for each specific portfolio $X \in R^r$ the random vector of loss has a distribution function of $F(u)$:

$$F(u) = P\{\Lambda \in R : \Lambda \leq u\} = P\{y \in R^n : X^T Y \geq \nu - u\nu\} \quad (4)$$

The expected level of prices at the end of time horizon $\bar{T}$ determines the expected loss of the selected portfolio:

$$E(\Lambda) = E[v^{-1}(\nu - X^T Y)] = v^{-1}(\nu - X^T E(Y)) \quad (5)$$

Selecting the portfolio with the minimum level of expected loss is inefficient in terms of returns foregone. The investor could choose a level of expected loss such as $\rho$, higher than that minimal level, but one that reduces his/her risk in terms of replacement value, or terminal value of the selected portfolio. If the level of risk in the selected portfolio is $\mathcal{R}(\Lambda)$ the optimization problem of this investor could be shown to be equivalent to:

$$\begin{align*}
\text{Min}_X \quad & \mathcal{R}(\Lambda) \\
E(\Lambda) & \leq \rho \\
X^T \nu & = \nu \\
X & \geq 0
\end{align*} \quad (6)$$

In this paper, we want to choose an optimal portfolio drawn from the top 30 stocks in the US market for an investor who enters the market with a view to achieving portfolio gains over a short investment horizon of the next few days. Both end of day and high frequency data for these top 30 stocks was taken from the TAQTIC database of the Securities Industry Research Centre of Asia-Pacific (SIRCA). As at the time of data query, the last data available from this database was up to the end of September 2007, the high frequency data for all these top 30 stocks was downloaded for the second and third weeks of September 2007. The data pertaining to the last week of September 2007 was not downloaded, because we decided to use the closing prices taken from this week for evaluation of the results in terms of a hold out sample and consequently do not use them in the optimization procedure. Also a complete one year of end of day prices data for the same stocks was downloaded from October 2006 up to the end of September 2007. Again last the last week’s data was not used for optimization but as a hold out sample.

The Variance, VaR and CVaR are the three risk measures that are used in this paper for selecting the optimal portfolio in a risk-return framework. The Variance of losses ($\nu(\Lambda)$) according to definition is:
The VaR of losses ($\zeta_\beta(\Lambda)$) is defined variously in literature, but this does not affect the outcomes.

$$\mathcal{R}(\Lambda) = \text{VaR}(\Lambda) = \zeta_\beta(\Lambda)$$

VaR could be defined as “a loss that will not be exceeded at some specified confidence level”[18]. In the other word, “the 100a% h-day VaR is that number x such that the probability of losing x, or more, over the next h days equals 100a%”[3]. But formally $\zeta_\beta(\Lambda)$ is defined as $\beta$ percentile of loss distribution function [17], then $\zeta_\beta(\Lambda)$ is a smallest value such that probability that loss does not exceed to this value is bigger or equals to $\beta$[20].

$$\zeta_\beta(\Lambda) = \text{Min}\{\zeta \in R : P\{\Lambda \in R : \Lambda \leq \zeta\} \geq \beta\}$$

CVaR of losses $\omega_\rho(\Lambda)$ is the expectation of losses conditioned on exceeding or being equal the $\zeta_\beta(\Lambda)$ [17].

$$\mathcal{R}(\Lambda) = \text{CVaR}(\Lambda) = \omega_\rho(\Lambda)$$

$$\omega_\rho(\Lambda) = \frac{P\{\Lambda \in R : \Lambda \geq \zeta_\beta(\Lambda)\}}{1 - \beta} E(\Lambda|\Lambda \geq \zeta_\beta(\Lambda)) + \left(1 - \frac{P\{\Lambda \in R : \Lambda \geq \zeta_\beta(\Lambda)\}}{1 - \beta}\right) \zeta_\beta(\Lambda)$$

where, $\omega_\rho(\Lambda) = E(\Lambda|\Lambda \geq \zeta_\beta(\Lambda))$ if $P\{\Lambda \in R : \Lambda \geq \zeta_\beta(\Lambda)\} = 1 - \beta$.

In order to calculation and optimization of these above-mentioned risk measures, in this paper the sample PDF of losses is simulated using the actual historical end of day prices and high frequency data without assuming any specific distribution of loss function. As the target of our supposed short term investor are gains obtained during the next $\Delta t$ days of investing days or short-term investment horizon, a necessary requirement is the specification of a sample PDF of losses up to that date via simulations of the probable loss conditions. If the historical end of day price of $m$ stocks in time $t$ is $h^t = (h^t_1, h^t_2, ..., h^t_m)$, $r_j = (r_{j1}, r_{j2}, ..., r_{jm})$ is the $j^{th}$ scenario of probable rates of price changes over the next $\Delta t$ days.

$$r_j = \frac{h^t_j}{h^{t-\Delta t}_j}$$

$$t = 1 + \Delta t, 2 + \Delta t, ..., T$$
Then, there are $N = T - \Delta t$ scenarios when $T$ end of day historical data are used for simulating the sample PDF of losses. If high frequency data are utilized for scenario making, there are $K$ price data and then $K$ scenarios in each day. In such a situation $r_j$ is:

$$r_j = \frac{h^k_j}{h^{T-\Delta t}},$$

$$h^k = (h^k_1, h^k_2, ..., h^k_m),$$

where $h^k_m$ is $k^{th}$ transaction price data of $m^{th}$ stock in the $t^{th}$ historical date. Then, there are $N = K(T - \Delta t)$ scenarios when $T - \Delta t$ days high frequency data are used for simulation of the sample PDF of losses.

Utilizing $r_j$ for $j = 1, 2, ..., N$ in:

$$p^T r_j = y_j$$

We have a sample of $Y \in R^n$ with $N$ members that the $j^{th}$ member is $y_j = (y_{j1}, y_{j2}, ..., y_{jm})$. Also $\bar{y} = (\bar{y}_1, \bar{y}_2, ..., \bar{y}_m)$ is the vector of mean value of each stock price in all scenarios:

$$\bar{y}_j = \frac{1}{N} \sum_{j=1}^{N} y_{ij}$$

Now, there are $N$ members in the sample of $\Lambda \in R$ corresponding $y_j$, with the $j^{th}$ member of:

$$\Lambda_j = \frac{\nu - \sum_{i=1}^{m} X_i y_{ij}}{\sigma}$$

Having obtained the sample PDF of losses, it is possible to optimize the portfolio selected using the previously mentioned risk measures and thereby select the optimal portfolio for a short term investor by calculating the efficient risk-return frontier.

In order to choosing the optimal portfolio with a minimal level of Variance, this nonlinear model is used in this paper:
\[
\text{Min}_x \quad \nu(X) = v^{-2} X^T \left[ \frac{1}{N} \sum_{j=1}^{N} (y_j - \bar{y})(y_j - \bar{y}) \right] X \\
v^{-1}(v - X^T \bar{y}) \leq \rho \\
X^T p = v \\
X \geq 0
\] (15)

It was illustrated in the literature [20] that the minimum level of CVaR is achieved by minimizing the following:
\[
\text{Min} \omega_\beta(\Lambda) = \text{Min}\left\{\zeta_\beta(\Lambda) + (1 - \beta)^{-1} E \max[\Lambda - \zeta_\beta(\Lambda), 0]\right\} (16)
\]

Then this linear model could be utilized in order to selecting the optimal portfolio with minimum level of CVaR [17, 20, 21]:
\[
\text{Min}_{x, \zeta, z} \quad \omega_\beta(\Lambda) = \zeta + \frac{1}{(1 - \beta)N} \sum_{j=1}^{N} Z_j \\
v^{-1}(v - X^T y_j) - \zeta - Z_j \leq 0 \\
v^{-1}(v - X^T \bar{y}) \leq \rho \\
X^T p = v \\
Z_j \geq 0 \\
X \geq 0
\] (17)

\(Z_j\) is a auxiliary variable that is used to selecting the \(\text{Max}[\Lambda - \zeta_\beta(\Lambda), 0]\) in the above model. This is because when we proceed according to definition, \(\omega_\beta(\Lambda)\) is the expectation of losses conditioned on exceeding or being equal to the \(\zeta_\beta(\Lambda)\).

Concerning the optimization of VaR, according to the definition, the VaR of losses is the \([\beta N]^{th}\) minimum of all loss scenarios in the sample PDF [17]:
\[
\zeta_\beta(\Lambda) = \text{Min}\left\{v^{-1}(v - X^T y_1), v^{-1}(v - X^T y_2), ..., v^{-1}(v - X^T y_N)\right\} (18)
\]

Where \([\beta N]\) is the smallest integer non-smaller than \(\beta N\).

Then, in order to find the optimal portfolio with the minimum level of losses VaR, we should minimize the \([\beta N]^{th}\) minimum of all scenarios in the sample PDF of loss:
\[
\text{Min}_x \quad \zeta_\beta(\Lambda) = \text{Min}\left\{v^{-1}(v - X^T y_1), v^{-1}(v - X^T y_2), ..., v^{-1}(v - X^T y_N)\right\} (19)
\]

A major contribution of this paper is the following simplification of the portfolio estimation procedure in which we propose a new optimization and more efficient model of VaR as follows:
\[
\min_{X, \Omega, \Delta} \xi_{\rho}(\Lambda) = \left[ \sum_{j=1}^{N} d_j \varphi(\Omega_j) \right]^{-1} \sum_{j=1}^{N} d_j \varphi(\Omega_j) \left( v^{-1}(v - X^T y_j) \right)
\]

\[
\sum_{j=1}^{N} \varphi(\Delta_{gj}) - \Omega_j = \beta N
\]

\[
\Delta_{gj} - \left( v^{-1}(X^T y_g - X^T y_j) \right) = 0
\]

\[
\sum_{j=1}^{N} d_j \varphi(\Omega_j) \geq 1
\]

\[
v^{-1}(v - X^T y) \leq \rho
\]

\[
X^T p = v
\]

\[
X \geq 0
\]

where \( d_j \) is a binary \((0,1)\) variable, \( \Omega_j \) and \( \Delta_{gj} \) are two auxiliary variables (which \( g = 1, 2, \ldots, N \) is an index that just shows the number of scenarios as previously given as \( j \)) and \( \varphi(z) \) is a conditional function such that:

\[
\varphi(z) = \begin{cases} 
1 & \text{if } z \geq 0 \\
0 & \text{otherwise}
\end{cases}
\]

This new optimization model is much more efficient than previous methods because it has smaller dimensions. Smaller dimensions are a major requirement in the optimization of large scale models such as those encountered using high frequency data models. In some of cases, VaR optimization with the previous methods using high frequency price data prove to be impossible, given the curse of dimensionality. The model proposed here is a mixed integer nonlinear model that can be solved with the related solvers such as ALPHAECP, BARON, DICOPT, LINDOGLOBAL, OQNLP and SBB.

Notice that in the cases that we have a locally optimum solution, the \( \varphi(z) \) function could be produced for the other definable values for different scenarios of negative \( z \) values such as \( \varepsilon \leq z < 0 \) -where \( \varepsilon \) has a little absolute value- in order to achieve smoothing in this model. This method of smoothing sometimes was used in the literature [4, 17]. However, this method could stray away from the solution for the exact amount of VaR in the case of semi-large \( \varepsilon \) values. Although selecting small amounts of \( \varepsilon \) can’t solve the non-smoothing problem. Then the optimal selection of \( \varepsilon \) requires trial and error methods.

The other way round this difficulty is via the relaxation of the \( \varphi(z) \) function in a relaxed mixed integer nonlinear programming framework using the related solvers. We used the COINPOPT solver in the GAMS environment for the optimization of this model.
Another advantage of this model is that it is possible to calculate the CVaR value of the optimal VaR portfolio at the same time. This will be done adding the following constraint to the model:

$$\omega_{\beta}(\Lambda) - \left[ \left( (1 - \beta) N \right)^{-1} \sum_{j=1}^{N} \psi(\Omega_j) \left( \nu^{-1}(\nu - X^T y_j) \right) \right] = 0 \quad (21)$$

where $\psi(z)$ is another conditional function as shown below:

$$\psi(z) = \begin{cases} 
0 & \text{if } z \leq 0 \\
1 & \text{otherwise} 
\end{cases}$$

3. Results and discussions

As we discussed in the introduction to the paper, our central goal in the paper is to assess the effectiveness of using intra-day high frequency data to select optimal portfolios over short investment time horizons. In order to investigate this hypothesis we need to have a benchmark or base for making comparisons. The typical base case is to use the results from choosing optimal portfolios using historical data as is customary in the literature. We used the three described risk measures utilizing a total of a full 1 year’s end of day historical data from October 2006 up to third week of September 2007. Optimal portfolios are then determined in four scenarios of investment horizons using each method. We have supposed that the last day of third week of September in 2007 is the investment decision day. Then the four different scenarios of investment horizons are the 2\(^{\text{nd}}, 3\(^{\text{th}}, 4\(^{\text{th}}\) and 5\(^{\text{th}}\) day of the last week of September in 2007 which is the first out of sample week. The amount of maximum level of expected loss- $\rho$ parameter- is taken to be equal to -0.3\%, -0.45\%, -0.6\% and -0.7\% for the next 2-5 days investment horizons respectively. The results are illustrated in Table 1 below.

The row of actual losses shows the amount of loss faced on the assumption that we select the corresponding portfolio in each of the scenarios using the actual prices of that day. The first part of Table 1 shows the results of optimal portfolio selection achieved by minimizing the variance of losses. Actual loss is smaller than the maximum level of expected loss in only the 4 days investment horizon scenario. It means that we have a smaller return than we had expected in the other scenarios. There is a positive actual loss in the 2 day investment horizon scenario. This condition continues for the portfolios which we selected using minimizing the CVaR and VaR of losses in the case of the first scenario. The
optimal portfolios determined by the parameters obtained from historical data were not efficient in this scenario of a 2 day investment horizon. The second and third parts of Table 1 are results that are related to minimizing the CVaR of losses using the usual linear method and the minimizing of the VaR utilizing the new proposed approach respectively. The results for the simultaneously calculated CVaR of optimal portfolios with the minimum level of VaR are also shown in Table 1.

Table 1. Results of optimal portfolio selection using historical data (%)
The portfolio weights are shown in the column after the company abbreviation.

<table>
<thead>
<tr>
<th>Investing horizon</th>
<th>2 days</th>
<th>3 days</th>
<th>4 days</th>
<th>5 days</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variance Minimization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal portfolio</td>
<td>ETR 12.3</td>
<td>IBM 66.7</td>
<td>XOM 21</td>
<td>ETR 16.8</td>
</tr>
<tr>
<td>Actual loss</td>
<td>0.21</td>
<td>-0.37</td>
<td>-0.78</td>
<td>-0.5</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.13</td>
<td>0.17</td>
<td>0.2</td>
<td>0.19</td>
</tr>
<tr>
<td><strong>CVaR 99% Minimization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal portfolio</td>
<td>ETR 5.2</td>
<td>HPQ 5</td>
<td>IBM 87.5</td>
<td>ETR 9.5</td>
</tr>
<tr>
<td>Actual loss</td>
<td>0.13</td>
<td>-0.45</td>
<td>-0.79</td>
<td>-0.5</td>
</tr>
<tr>
<td>VaR</td>
<td>3.72</td>
<td>4.44</td>
<td>5.28</td>
<td>4.77</td>
</tr>
<tr>
<td>CVaR</td>
<td>3.8</td>
<td>4.61</td>
<td>5.93</td>
<td>5.03</td>
</tr>
<tr>
<td><strong>VaR 99% Minimization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal portfolio</td>
<td>ETR 13.6</td>
<td>IBM 68</td>
<td>XOM 18.4</td>
<td>ETR 18</td>
</tr>
<tr>
<td>Actual loss</td>
<td>0.2</td>
<td>-0.49</td>
<td>-0.8</td>
<td>-0.5</td>
</tr>
<tr>
<td>VaR</td>
<td>3.7</td>
<td>4.4</td>
<td>5.1</td>
<td>4.7</td>
</tr>
<tr>
<td>CVaR</td>
<td>3.94</td>
<td>4.76</td>
<td>5.93</td>
<td>5.54</td>
</tr>
</tbody>
</table>

Optimal portfolios obtained by minimizing VaR seem more successful than those which were obtained by minimizing the CVaR of losses as well as the Variance of losses according to the amounts of the actual losses. But the rates attached to the calculated VaR and CVaR using both methods are not really small enough for short term investors. For example, there is not much reassurance in the statement that "We are 99% certain that you will not lose more than 5.28% of your capital in the next four days" to an investor who arrives in the market with an expected rate of return of 0.6%. These large rates are the result...
of entering many different conditions into the sample PDF of losses by virtue of the use of historical data when the investment horizon is short. Although utilizing historical data in optimal portfolio selection means that we have an exact sample for calculating the PDF, it seems that this apparent precision could be illusory and not particularly helpful for a short term investor with a short-term investment horizon.

This opens the door to the use of high frequency data. In the case of choosing a portfolio using intra-day high frequency data, we have just the conditions which would be likely to continue to occur in the normal trend of the market over the next few days. This kind of data does not have complete information about the future but does have sufficient and valuable information if the focus is on the next few days which is the requirement for short term investors. The results of utilizing this type of high frequency data in the different scenarios is shown in Table 2.

These results are achieved using high frequency data for the two middle weeks of September 2007. As is shown in the table, the actual losses are smaller than in the similar situations based on historical data in all of the cases. These amounts are also smaller than maximum level of expected losses in most of the cases. Also there is not a positive amount of loss in any case. In terms of the relative effectiveness of using the different risk measures, the results in Table 2 show that optimal portfolios determined by minimizing the VaR of losses have smaller rates of actual losses. Table 2 also shows that the calculated rates of VaR and CVaR are very small and in most of the cases are negative. The negative rate of VaR shows that our return will not decrease from the absolute amount of that rate in the next few days given a confidence level of 99%. It is clear that the systematic risk of market, or longer term trends are not a concern in this measurement as we are focused on market conditions of stocks over a short period of time.

Although comparisons of Table 1 and Table 2 clearly illustrate the relative advantages of high frequency data versus historical data in optimal portfolio selection for short investment horizons, the relative rankings of stocks chosen depicted in Figure 1 show this advantage from another viewpoint.
Table 2. Results of optimal portfolio selection using high frequency data (%)

The portfolio weights are shown in the column after the company abbreviation. The results in the table follow from the optimization routines set out in expressions (20) and (21). It has to be born in mind that the optimisation involves minimising a given loss level set at \( \rho \). The loss levels are displayed as positive numbers in the table. “Gains” the reverse of losses, are indicated as negative numbers.

This counter-intuitive representation is consistent with the programming algorithm adopted.

<table>
<thead>
<tr>
<th>Investing horizon</th>
<th>2 days</th>
<th>3 days</th>
<th>4 days</th>
<th>5 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal portfolio</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JNJ</td>
<td>34.1</td>
<td>JNJ</td>
<td>16.8</td>
<td>HNZ</td>
</tr>
<tr>
<td>KO</td>
<td>6.4</td>
<td>KO</td>
<td>12.7</td>
<td>HNZ</td>
</tr>
<tr>
<td>PEP</td>
<td>2.8</td>
<td>PEP</td>
<td>8.1</td>
<td>MDT</td>
</tr>
<tr>
<td>PG</td>
<td>63.1</td>
<td>PG</td>
<td>63</td>
<td>PG</td>
</tr>
<tr>
<td>WMT</td>
<td></td>
<td>WMT</td>
<td>5.7</td>
<td></td>
</tr>
</tbody>
</table>

| Actual loss       | -0.33      | -1.15      | -1.07      | -0.56      |
| Standard Deviation| 0.05       | 0.06       | 0.07       | 0.04       |

| Optimal portfolio |            |            |            |            |
| JNJ               | 31.3       | JNJ        | 30.2       | HNZ        | 7.3        |
| KO                | 1.5        | KO         | 8.2        | GE         | 9.4        |
| PEP               | 30.5       | PEP        | 37.6       | MDT        | 42         |
| PG                | 38.2       | PG         | 21         | HNZ        | 23         |
| WMT               |            | WMT        | 9.7        |            |            |

| Actual loss       | -0.09      | -0.7       | -0.91      | -0.66      |
| VaR               | 0.22       | -0.01      | -0.3       | -0.45      |
| CVaR              | 0.23       | 0          | -0.29      | -0.42      |

| Optimal portfolio |            |            |            |            |
| HNZ               | 1.3        | HNZ        | 2.9        | HNZ        | 43         |
| JNJ               | 29.1       | JNJ        | 12.7       | HNZ        | 43         |
| KO                | 3.4        | KO         | 8.2        | HPQ        | 1.3        |
| PEP               | 66.2       | PEP        | 7.3        | MDT        | 19.6       |
| PG                |            | PG         | 64.2       | MDT        | 57.4       |
| WMT               |            | WMT        | 4.7        | PG         | 10.7       |

| Actual loss       | -0.37      | -1.22      | -1.08      | -0.55      |
| VaR               | 0.19       | -0.1       | -0.31      | -0.47      |
| CVaR              | 0.46       | 0          | -0.11      | -0.41      |

![Figure 1. Average composition of selected portfolios using historical and high frequency data](image)

13
Figure 1 shows the average composition of portfolios utilizing historical and high frequency data in all scenarios when using all of the methods. The numbers beside the indices for each stock in this figure introduce the average rank of that stock during the last week of September 2007. The first thing that is clearly illustrated in this figure is that the composition of portfolios is completely different when based on these two kinds of data sets. Another point apparent in this figure is the relative rank of the stocks in these two sets of portfolios. In the case of the high frequency data, about 60% of all the investment portfolios always consist of the top 5\textsuperscript{th}, 6\textsuperscript{th} and 7\textsuperscript{th} stocks in terms of market performance in the investment period. Whilst using historical data leads to chosen portfolios that are compositions of the 13\textsuperscript{th} and 17\textsuperscript{th} stocks in terms of market performance in about 90% of all the cases.

In this paper we have used just two weeks (ten working days) intra-day high frequency data to set up portfolios, but we are not asserting that the optimal period of sample data in order to undertake portfolio selection is necessarily two weeks of data to set up short-term investment portfolios. Determining the optimal size of this sample period needs more investigation and clearly differs from time to time and from market to market. We have argued that using historical long term data could be illusory and uninformative when making short-horizon investment decisions, whilst utilizing the end of day short term data sets is insufficient, in terms of not providing sufficient information in order to make meaningful selections of short term optimal portfolios. A viable alternative is to use intra-day high frequency data sets and simulations to obtain the required pdfs, and the results presented in the paper illustrate the viability and benefits of utilizing intra-day high frequency data for portfolio choices by investors with short term investment horizons.

4. Conclusion

In this paper the contribution and advantages of utilizing high frequency data for optimal portfolio selection purposes are investigated via the application of three different risk measures: Variance, VaR and CVaR in a portfolio selection risk-return framework. A new computational method for the calculation of VaR with smaller dimensions and the additional capability of simultaneously calculating CVaR which is necessarily more efficient in large scale optimization is proposed. The results show that the effectiveness of VaR as a risk measure in optimal portfolio selection is demonstrated to be more apparent than Variance and CVaR, especially in the case of this particular high frequency data set.
The results confirm the advantages of utilising high frequency data for short term investment horizons. The actual losses of the portfolios obtaining by using historical data are bigger than in the similar situation of choices made by using high frequency data in all cases. Also we do not have any positive amounts for actual losses (in effect we are making gains) in the cases where we use high frequency data. Whilst in the cases of historical data these amounts are positive (i.e. Losses) in 25% of all cases.
References:


