Portfolio Selection With High-Frequency Data

Xiaoyang Zhuang
Economics 201FS/Spring 2011
May 4, 2011

The Duke Community Standard was upheld in the completion of this project.
I. Introduction

For an investor, the question of how to allocate his wealth amongst a basket of assets is an important one. It is necessary to understanding the returns he can expect from his portfolio, in addition to the volatility the portfolio can experience about that expectation. The question was first explored in a seminal paper by Harry Markowitz (1952). Given a set of assets amongst which an investor could allocate his wealth, Markowitz derives what has since become known as the “Markowitz bullet”: For a desired level of expected return, the point on the Markowitz bullet describes the portfolio of assets that would achieve that return with the least volatility. The portfolios on the Markowitz bullet, which have a closed-form solutions, are known as the minimum-variance portfolios.

But this model, while appealing in its simplicity, fails to capture some important characteristics of the capital markets. In particular, it assumes that the investor optimizes in one-period; under this framework, asset expected returns, variances, and covariances remain constant and known. Empirically, this assumption runs counter to the evidence of time-varying expectations, variances, and covariances. In particular, asset volatilities and correlations both increase during periods of market stress, and global asset correlations have increased in the 20th and 21st centuries as global capital markets have become increasingly integrated (see Buraschi, Porchia, and Trojani (2010) for a guide to the empirical literature).

The necessity of knowing expected returns, variances, and covariances for this model (as well as the other widely-used investment model, the Black-Scholes-Merton option pricing model) has led to wild interest in the estimation of these parameters and analysis of financial time-series. While not developed for this purpose initially, GARCH estimators have been the choice tool used in the financial services industry, using the second moments of daily returns to
estimate asset and market volatility. More recently, motivated by the availability of high-frequency asset price data, the work of Andersen, Bollerslev, Diebold, and Labys (2003) has led to the use “realized” models. The realized volatility for a given day $t$ is computed by taking the sum of squares of each of the five-minute (or one-minute, etc.) returns. The theoretical benefit of these high-frequency estimators is facilitated by the Law of Large Numbers: as the number of samples (i.e. returns) goes to infinity, the realized volatility and realized covariance measures converge to the theoretical “integrated volatility” and “integrated covariance,” respectively.

The development of these new estimators has, in turn, led to investigations into their practical benefits in investment decisions. Fleming, Kirby, and Ostdiek (2003) use a “volatility timing” approach to test the benefit of high-frequency data on portfolio allocation decisions. Specifically, they yield a 50- to 200- basis point benefit relative to a covariance matrix estimation using a special case of the BEKK multivariate GARCH model (Engle and Kroner, 1995). Liu (2009), then, finds that the benefits of high-frequency data exist only when the investor rebalances daily or his estimation window is fewer than six months.

In this paper, we begin with an equi-weighted portfolio in five stocks: Alcoa (AA), DuPont (DD), Ford (F), JPMorgan Chase (JPM), and Walmart (WMT). Then, step-by-step, we examine the performance gains when increasingly sophisticated estimation procedures are added. First, I add a covariance matrix estimator that is estimated using daily returns, constant (10%) expected returns, and daily rebalancing using the Markowitz mean-variance framework with a target annual return of 12%. Then, I add conditional expected returns. After that, I use a high-frequency estimator for the covariance matrix. The portfolios’ performance will be tracked between April 9, 1997 and December 23, 2010.

The rest of this paper will proceed as follows: Section 2 will describe the tools that will
be employed over the course of this study. Section 3 examines the results with a particular view toward portfolio performance gains. Section 4 will conclude.

II. Finance and Econometric Tools

i) The Markowitz Mean-Variance Framework

In the Markowitz model, investors consider two exogenous factors in making their investment decisions: expected returns and the variance of these expected returns. So, for a desired portfolio return \( \mu_p \), they solve the following optimization problem:

\[
\min (\alpha_t) \quad \sigma^2 = \alpha_t \Sigma_t \alpha_t \quad \text{subject to} \quad \alpha_t^T q = 1, \quad \alpha_t^T \mu_t = \mu_p
\]

where \( \sigma^2 \) is the portfolio variance, \( \Sigma_t \) is the covariance matrix of \( n \) assets, \( \alpha \) is an \( n \times 1 \) vector of portfolio weights, \( q \) is an \( n \times 1 \) vector of ones, and \( \mu \) is an \( n \times 1 \) vector of expected returns. In particular, I constrain the portfolio weights to sum to 1, so that we do not consider a risk-free asset.

ii) Covariance Matrix Estimation Using Daily Returns

In estimating our covariance matrix that uses daily returns, we will follow Fleming, Kirby, and Ostdiek (2003). The formula is as follows:

\[
\hat{\Sigma}_t = \sum_{k=1}^{T} \Omega_{t-k} \otimes e_{t-k} e'_{t-k}
\]

Define \( n \) as the number of assets that the investor is allocating between. \( \hat{\Sigma}_t \), then, is the \( n \times n \) matrix of asset variances and covariances at time \( t \). \( e_{t-k} \) is an \( n \times 1 \) vector of daily return innovations (i.e. the difference between the actual return and expected return for each asset), so \( e_{t-k} e'_{t-k} \) is an \( n \times n \) matrix of variances and covariances computed using the standard definitions.
\( \Omega_{t,k} \) is an \( n \times n \) matrix of weights (to represent how much the asset covariances during previous days factor into the covariance matrix estimation at day \( t \)), \( T \) is the number of lags (in days) used in computing the covariance matrix estimator, and \( \odot \) denotes element-by-element multiplication. Each element of the weighting matrix has the following form:

\[
a \times \exp(-k \times \alpha)
\]

where we set \( \alpha \) equal to 0.06, a reasonable value based on the RiskMetrics methodology ("RiskMetrics Technical Document").

**iii) Conditional Expected Returns**

Expected returns for a period are highly dependent on the state of the markets in the previous period. When the economy is strong – marked by low unemployment, high output, high asset prices, and the like – stock market returns are high and expectations for future returns are high as well. To incorporate this phenomenon into our model, we use a rolling estimator of expected returns:

\[
\mu_t = \text{mean}(R_{t-1}, R_{t-2}, \ldots, R_{t-T})
\]

where \( \mu_t \) is the \( n \times 1 \) vector of expected asset returns on day \( t \), and \( T \) is the number of lags (in days) used in calculating our rolling estimator.

**iv) High Frequency Estimator for the Covariance Matrix**

It is natural for our covariance matrix, estimated using high-frequency data, to share the same general shape as the one estimated using daily returns data. So, as before, this estimator will have the following structure:
In this case, $V_{t-k}$, the contribution on day $t-k$, is estimated using high-frequency returns data, as opposed to daily returns data. The $(i,j)^{th}$ element of $V_{t-k}$ is as follows:

$$
\{V_{t-k}\}_{i,j} = \sum_{j=1}^{M} (r_{i,t+j\Delta})(r_{j,t+j\Delta})
$$

where $r_{i,t+j\Delta}$ represents the logarithmic return of asset $i$ between times $t+j\Delta$ and $t+(j-1)\Delta$. Finally, $\Delta$ is the (high-frequency) sampling interval, and $M$ represents the total number of such sampling intervals over the course of a day.

In the analysis, we take $\Delta = 5$ minutes, and then $M = 76$, as per our data set. Again, each day’s contribution $V_{t,k}$ is then weighted by $\alpha*\exp(-k*\alpha)$, and $\alpha = 0.06$ is the parameter of choice.

III. Results

i) Portfolio Returns: The Equiweighted Portfolio

We begin with a benchmark case. We consider the performance of a static portfolio with 20% of its funds invested in each of the following stocks over its lifetime (April 9, 1997 to December 23, 2010): Alcoa (AA), DuPont (DD), Ford (F), JPMorgan Chase (JPM), and Walmart (WMT). No rebalancing is ever conducted.

Without consideration for dividends, the return on this portfolio at time $t$ is simply the (arithmetic) average return of each of the individual stocks over $t$. This method gives us a portfolio return of 36.43% between 1997 and 2010, driven in particular by Walmart’s 132% return, JPMorgan Chase’s 42.94% return, and Ford’s 22.57% return. The losers during this period were Alcoa (which lost -9.64% of its value over the period) and DuPont, which lost
6.19%. Figure 1 shows the daily returns of the equiweighted portfolio during this time period.

![Figure 1: Daily Returns of the Equiweighted Portfolio](image)

<table>
<thead>
<tr>
<th>Mean daily return (%)</th>
<th>Median daily return (%)</th>
<th>Standard Error (%)</th>
<th>Sharpe Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0107</td>
<td>0.0235</td>
<td>0.0311</td>
<td>0.343766</td>
</tr>
</tbody>
</table>

Table 1: Sample Statistics for the Daily Returns of the Equiweighted Portfolio

The equiweighted portfolio’s mean daily return is 0.0107%, and its median daily return is 0.0235%, so the mean is biased down by large negative returns, which is supported by the literature. The standard error of these returns is 0.0311%, so the Sharpe ratio is 0.3438 (there are 3404 data points) if we assume that returns are i.i.d. We now use this portfolio as a performance benchmark for portfolios derived using more advanced methods. In the next section, the first tools I will add are daily portfolio rebalancing to meet the optimal Markowitz allocations, an estimator for the covariance matrix using daily returns, constant expected returns (10%), and a target annual return of 12%.
ii) Portfolio Returns: Adding Mean-Variance Optimization (with Overnight Returns in Covariance Matrix Computation)

In the analysis for this section, we move away from the equiweighted portfolio. Instead, we optimize using the Markowitz one-period framework, i.e.

\[ \min(\alpha_t) \sigma^2 = \alpha_t \Sigma_t \alpha_t \quad \text{subject to} \quad \alpha_t^T q = 1, \alpha_t^T \mu_t = \mu_p \]  

Doing so requires us to estimate the covariance matrix \( \Sigma_t \), determine the vector of expected returns \( \mu_t \), and set a target return \( \mu_p \). So, we add the following tools:

- Covariance matrix estimation using daily returns, as described in Section 2ii), with lag \( k = 5 \)
- A vector of expected (daily) returns \( \mu_t \) with \( 0.1^{1/252} \) in every slot (i.e. the expected return of every asset is 10% per year)
- Set the target (daily) portfolio return \( \mu_p \) to \( .12^{1/252} \) (i.e. the target return of the portfolio is 12% per year)
- Note that overnight returns are used in making these portfolio allocation decisions

Using the Lagrange Multiplier method to solve (1) for the optimal portfolio \( \alpha_t \), we have the following:

Define \( A = q^T \Sigma_t^{-1} q \)

Define \( B = \mu_t^T \Sigma_t^{-1} q \)

Define \( C = \mu_t^T \Sigma_t^{-1} \mu_t \)

\[ \alpha_t = \frac{C - \mu_p B}{AC - B^2 \Sigma_t^{-1} q} + \frac{\mu_p A - B}{AC - B^2 \Sigma_t^{-1} \mu_t} \]

This portfolio actually performed worse than the equiweighted portfolio. The (arithmetic) mean daily return was \(-0.0658\%\), with a standard error of \(0.0551\%\), leading to a disappointing Sharpe ratio of \(-1.1942\) (see Table and Figure 2 for the results). A likely reason for the high standard
error relative to the mean is the inclusion of large and often distortionary overnight returns in the computation of the covariance matrix. So, our next step is to exclude the overnight returns, and our results are presented in Table and Figure 2.

**Figure 2: Daily returns of the Markowitz portfolio that uses overnight returns in computing the covariance matrix**

<table>
<thead>
<tr>
<th>Mean daily return (%)</th>
<th>Median daily return (%)</th>
<th>Standard Error (%)</th>
<th>Sharpe Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.0658 (target: 0.0450)</td>
<td>-0.0053</td>
<td>0.0551</td>
<td>-1.1942</td>
</tr>
</tbody>
</table>

**Table 2: Sample statistics corresponding to Figure 2**

**iii) Portfolio Returns: Adding Mean-Variance Optimization (without Overnight Returns in Covariance Matrix Computation)**

This section uses the same portfolio selection framework as described in the previous section, except overnight returns are excluded in the computation of the covariance matrix (and thus the optimal portfolio). The mean daily return is improved in magnitude from the previous section, at
-0.0133%. The standard error is also slightly improved, at 0.0515%, leading to a Sharpe Ratio of -0.2078. These results do not indicate strong distortionary effects by the inclusion of overnight data because of the small change in the portfolio standard error. Regardless of whether or not we include the overnight data, the daily returns are still insignificantly nonzero at any standard significance level. These results are still disappointing relative to the much simpler equiweighted portfolio. The results are presented in Figure 3 and Table 3.

Figure 3: Daily returns of the Markowitz portfolio that excludes overnight returns in computing the covariance matrix

<table>
<thead>
<tr>
<th>Mean daily return (%)</th>
<th>Median daily return (%)</th>
<th>Standard Error (%)</th>
<th>Sharpe Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.0133 (target: 0.0450)</td>
<td>-0.0136</td>
<td>0.0515</td>
<td>-0.2583</td>
</tr>
</tbody>
</table>

Table 3: Sample statistics corresponding to Figure 3
iv) Portfolio Returns: Adding Time-Dependent Expected Returns

Again, we use a framework that is identical to the framework described in subsection i), with a couple of key modifications.

- We add a less simplistic expected returns vector. In particular, the expected return of asset $i$ on day $t$ is the arithmetic mean of the actual returns of asset $i$ in the five days preceding $t$.
- Again, we exclude overnight returns in estimating the covariance matrix and expected returns vector.

The mean daily return is quite small at 0.0051%, as is the Sharpe Ratio of 0.06702. However, it is evident in comparing these results with those from subsections ii) and iii) that a conditional expected return does yield something of an improvement in the portfolio selection decision.

![Daily Returns: The Markowitz Portfolio, Rebalanced Daily](image)

**Figure 4:** Daily returns of the Markowitz portfolio when a conditional expected return vector is applied
<table>
<thead>
<tr>
<th>Mean daily return (%)</th>
<th>Median daily return (%)</th>
<th>Standard Error (%)</th>
<th>Sharpe Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0051 (target: 0.0450)</td>
<td>0.0089</td>
<td>0.0761</td>
<td>0.06702</td>
</tr>
</tbody>
</table>

Table 4: Sample statistics corresponding to Figure 4

v) Portfolio Returns: Adding High-Frequency Estimates for the Covariance Matrix

This section again uses the Markowitz framework outlined in subsection i), but with several changes:

- The estimator for the expected return vector is a rolling estimator of the realized returns in the previous five days
- No overnight data is used in estimating the expected return vector or the covariance matrix
- The covariance matrix estimator uses high-frequency (5-minute) returns to compute each day’s realized asset variances and realized covariances

The results here are a little more promising. The mean daily returned increased to 0.01132, slightly bettering the static portfolio and yielding a Sharpe Ratio that is quite close to the one for the static portfolio. By no means compelling, it remains an important question what necessary steps must be taken in order to condition the covariance matrix estimators to yield stronger results.
Figure 5: Daily returns of the Markowitz portfolio when high-frequency returns are used to estimate the covariance matrix

<table>
<thead>
<tr>
<th>Mean daily return (%)</th>
<th>Median daily return (%)</th>
<th>Standard Error (%)</th>
<th>Sharpe Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01132 (target: 0.0450)</td>
<td>0.3501</td>
<td>0.0361</td>
<td>0.3136</td>
</tr>
</tbody>
</table>

Table 5: Sample statistics corresponding to Figure 5

IV. Conclusions

The relative benefits to using high frequency data in the Markowitz portfolio setting are obvious – the Sharpe ratio increased nearly four times when high-frequency measures were added.

However, in our perhaps overly simple use of historical data, the optimal portfolio performed only slightly better than the static equiweighted portfolio on the mean daily return side (0.01132% vs. 0.0107%) and actually fared worse than the equiweighted portfolio in the Sharpe ratio.
This somewhat bizarre result may be attributable to several issues that will be explored after submission of this draft, including the impact of having a small number of stocks in the portfolio (five stocks, to be exact), choosing stocks out of pure luck that performed unusually strongly over the past decade (specifically, Walmart, JPMorgan, and Ford), and lack of conditioning on the covariance matrix estimators.
References

Andersen, Torben, Tim Bollerslev, Francis Diebold, and Paul Labys.  


