

Statistical methods for large portfolio risk management

JIAN ZOU* AND YU WANG

Portfolio allocation is one of the most important problems in financial risk management. It involves dividing an investment portfolio among different assets based on the volatilities of the asset returns. In the recent decades, it gains popularity to estimate volatilities of asset returns based on high-frequency data in financial economics. In this article, we focus on the portfolio allocation problem using high-frequency financial data. The paper starts with a review on portfolio allocation and high-frequency financial time series. Then we introduce a new methodology to carry out efficient asset allocations using regularization on estimated integrated volatility via intra-day high-frequency data. We illustrate the methodology by comparing the results of both low-frequency and high-frequency price data on stocks traded in the New York Stock Exchange in 2011. The numerical results show that portfolios constructed using high-frequency approach generally perform well by pooling together the strengths of regularization and estimation from a risk management perspective.

1. INTRODUCTION

In financial risk management, the question of how to allocate assets in a large portfolio is of utmost interest to most investors. It is necessary to understand the returns to expect from the portfolio, in addition to the volatility the portfolio can experience about that expectation. The question was first explored in a seminal paper Markowitz [18], which was the original milestone for modern portfolio theory on the mean-variance analysis by solving an unconstrained quadratic optimization problem. This approach has had a profound impact on financial economics. But this model, while appealing in its simplicity, fails to capture some important characteristics of the capital markets. In particular, it assumes that 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. For example, asset volatilities and correlations both increase during periods of market stress, and global asset correlations have increased in the recent decades as global capital markets have become increasingly integrated.

The drawback of requiring constant known parameters for this model has led to great interest in novel methods of modeling these parameters and analysis of financial time series. More recently, motivated by wide availability and the extra amount of information contained in the high-frequency asset price data, researchers are able to develop more accurate estimators for the volatility matrix. The volatility matrix of daily asset returns is a key input in portfolio allocation, option pricing and risk management. However, the main challenge is that when the number of assets is large, the volatility matrix cannot be estimated accurately. To address this problem, several innovative approaches for volatility matrix estimation were proposed in the past decade. Estimation methods for the univariate case include realized volatility (RV)(Andersen et al. [1]), bi-power realized variation (BPRV)(Barndorff-Nielsen and Shephard [3]), two-time scale realized volatility (TSRV)(Zhang et al. [24]), wavelet realized volatility (WRV)(Fan and Wang [9]), kernel realized volatility (KRV)(Barndorff-Nielsen et al. [2]), and Fourier realized volatility (FRV)(Mancino and Sanfelici [17]). For multiple assets, we face a non-synchronization issue, which is referred to as high-frequency price data that are not aligned properly for different assets, hence are recorded at various mismatched time points. For a pair of assets, Hayashi and Yoshida [13] and Zhang et al. [24] have developed methods based on overlap intervals and previous ticks to estimate co-integrated volatility of the two assets. The development of these new estimators has, in turn, led to investigations into their practical benefits in investment decisions. Fleming et al. [11] 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 [5]). Liu [16] finds that the benefits of high-frequency data in portfolio allocation exist only when the investor rebalances daily or the volatility estimation window is fewer than six months.

From a financial risk management perspective, Jagannathan and Ma [14] analyzed the impact of weight constraints in a large portfolio allocation. They show that solving the global minimum variance portfolio problem with some constraints on weights is equivalent to use a shrink-

*Corresponding author.

age estimate of the covariance matrix. Fan et al. [7] studied portfolio allocation with gross-exposure constraint combining vast volatility matrix estimation using different sampling schemes. However, there exists an interesting question as of when and how an investor will benefit from using high-frequency financial data in his/her portfolio allocation decisions. This paper aims to perform a comparative study on the performance of portfolios constructed using both low-frequency and high-frequency financial data. Therefore, we hope to shed some light on the benefits of high-frequency data in risk management.

Pertinent to the high dimensionality issues, penalized likelihood methods such as LASSO have been extensively studied for high dimensional variable selection and a considerable amount of research was dedicated recently to development of methods of risk management based on regularization (See e.g., Fan and Lv [8], Tibshirani [20]). Although the LASSO estimator can discover the correct sparse representation of the model (Donoho and Huo [4]), it is in general a biased estimator especially when the true coefficients are relatively large. Several remedies, including the smoothly clipped absolute deviation (SCAD) (Fan and Li [6]) and the adaptive LASSO (ALASSO) (Zou [25]) have been proposed to discover the sparsity of the true models, while producing consistent estimates for nonzero regression coefficients. Yuan and Lin [22] and Levina et al. [15] proposed estimators of the precision matrix via graphical models. Guo et al. [12] extended this methodology to multiple graphical models. In this paper, we adopt the idea of LASSO/LARS to come up with a regularized version of the estimated ARVM estimator (Wang and Zou [21]) for the daily volatility matrix. Zou and Wu [26] developed a similar approach using a different regularization method.

The data that motivated this research comprise the transaction-by-transaction stock prices from the 30 DJIA composite constituents traded in the New York Stock Exchange (NYSE). The data set is huge with ultra high-frequency observations since these stocks are highly liquid with a vast trading volume. These high-frequency financial data also possess unique features such as price discreteness, unequally spaced time intervals, and nonsynchronized trading (see e.g., Wang and Zou [21] for some illustrations of these issues). The normal trading hours of the NYSE are from 09:30 until 16:00. Thus, for simplicity, we discarded any transactions beyond these hours from our analysis.

The rest of this paper is structured as follows. Section 2 elaborates the framework of our methodology of portfolio allocation for high-frequency data. Section 3 presents numerical evidence on the performance comparison of our method under both low- and high-frequency scenarios. Section 4 concludes this paper.

2. METHODS

Suppose that a portfolio consists of p assets and their log price process $X_t = (X_{1t}, \dots, X_{pt})^T$ obeys an Itô process governed by

$$(1) \quad dX_t = \boldsymbol{\mu}_t dt + \boldsymbol{\sigma}_t dW_t, \quad t \in [0, L],$$

where W_t is a p -dimensional standard Brownian motion, $\boldsymbol{\mu}_t$ is a drift taking values in R^p , and $\boldsymbol{\sigma}_t$ is a diffusion variance of $p \times p$ matrix. Both $\boldsymbol{\mu}_t$ and $\boldsymbol{\sigma}_t$ are assumed to be continuous in t .

For the portfolio with allocation vector \boldsymbol{w} (i.e., percentage of each asset in the portfolio) and a holding period T , the variance (risk) of the portfolio return is given by $R(\boldsymbol{w}, \Sigma) = \boldsymbol{w}^T \Sigma \boldsymbol{w}$. However, it is well known that the estimation error in the mean vector $\boldsymbol{\mu}_t$ could severely affect the portfolio weights and produce suboptimal portfolios. This motivates us to adopt another popular portfolio strategy: the *global minimum variance portfolio*, which is the minimum risk portfolio with weights that sum to one. These weights are usually estimated proportional to the inverse covariance matrix, i.e., $\boldsymbol{w} \propto \Sigma^{-1}$. Following Jagannathan and Ma [14] and [10], we consider the following risk optimization with two different constraints:

$$(2) \quad \min \boldsymbol{w}^T \Sigma \boldsymbol{w}, \quad s.t. \|\boldsymbol{w}\|_1 \leq c \text{ and } \boldsymbol{w}^T \mathbf{1} = 1$$

where c is the gross exposure parameter which specifies the total exposure allowed in the portfolio, and $\|\cdot\|_1$ is the standard vector l_1 norm. The summation to one constraint ensures weight percentages sum to 100%, inducing a full investment. An additional common requirement is no short selling (e.g., 401k, IRAs, and mutual funds), which adds the nonnegative optimization constraint. This corresponds to the case $c = 1$ as in Jagannathan and Ma [14]. The second optimization case is the global minimum risk portfolio where the gross exposure constraint $c = \infty$. Note that we only consider these two cases for simplification of the problem. Other cases with varying c can be easily generalized in our methodology.

2.1 Average realized volatility matrix for high-frequency financial data

The average realized volatility matrix (ARVM) estimator computed from the high-frequency data on the prices of the p assets is used to estimate the integrated volatility over a given period of time. Let's take a day as a unit time. The integrated volatility for the ℓ -th day is a matrix defined to be

$$(3) \quad \Sigma_x(\ell) = \int_{\ell-1}^{\ell} \sigma_s \sigma_s^\dagger ds, \quad \ell = 1, \dots, L.$$

We first consider estimation of $\Sigma_x(1)$ on day one. For $i = 1, \dots, p$, suppose that the high-frequency data for the i -th asset on day one are observed at times $t_{ij}, j = 1, \dots, n_i$, and we denote by $Y_i(t_{ij})$ the observed log price of the i -th asset at time t_{ij} . Because transactions may be made at distinct times for different assets, the high-frequency price data for p -assets are often recorded at various mismatched time points.

This is referred to as a non-synchronized problem in high-frequency financial data.

Another complication in high-frequency financial data is that due to microstructure noise, the observed log price $Y_i(t_{ij})$ of the i -th asset is a noisy version of its true log price $X_i(t_{ij})$ and is usually assumed to obey an additive noise model

$$(4) \quad Y_i(t_{ij}) = X_i(t_{ij}) + \varepsilon_i(t_{ij}),$$

where $\varepsilon_i(t_{ij})$ are assumed to be i.i.d. noise with mean zero and finite fourth moments and independent of $X_i(t)$.

We consider the average realized volatility matrix (ARVM) estimator based on previous-tick times. Let $n = n_1 + \dots + n_p$ and assume that n_i/n are bounded away from zero. Take a predetermined sampling frequency $\tau_k, k = 1, \dots, m$. One such example is to select τ_i be a regular grid. For each $k = 1, \dots, m$, we choose the corresponding observation time for the i -th asset by

$$\tau_{ik} = \max\{t_{ij} \leq \tau_k, j = 1, \dots, n_i\}, \quad i = 1, \dots, p.$$

The realized co-volatility is defined to be a $p \times p$ matrix $\hat{\Sigma}_y(1)$ whose (i_1, i_2) element is given by

$$\begin{aligned} & \hat{\Sigma}_y(1)[i_1, i_2] \\ &= \sum_{k=1}^m [Y_{i_1}(\tau_{i_1,k}) - Y_{i_1}(\tau_{i_1,k-1})] [Y_{i_2}(\tau_{i_2,k}) - Y_{i_2}(\tau_{i_2,k-1})]. \end{aligned}$$

We estimate $\Sigma_x(1)$ by $\hat{\Sigma}_y(1)$. Zhang [23] proved that with i.i.d. microstructure noise and $m \sim n^{2/3}$,

$$\hat{\Sigma}_y(1) - \Sigma_x(1) = O_P(n^{-1/4}).$$

We apply the method in Wang and Zou [21] to the price observations on the ℓ -th day and obtain the average realized volatility matrix estimator $\tilde{\Sigma}_y(\ell)$ of integrated volatility $\Sigma_x(\ell)$ on the ℓ -th day. For the purpose of estimation of daily volatility, we have

$$(5) \quad \tilde{\Sigma}_y(\ell) - \Sigma_x(\ell) = O_P(n^{-\eta}), \quad \ell = 1, \dots, L,$$

where rates of convergence can be as slow as (optimal) n exponent η is equal to $1/4$ for i.i.d. noise and $1/2$ for no noise in the price data. However, since there are other sources of randomness in the data, such as price discreteness and non-synchronization errors, Wang and Zou [21] showed that the rate η is on the order of $1/6$. Subsequently, [19] established the optimal minimax risk for estimating a large volatility matrix under the subgaussian tail assumption.

3. NUMERICAL STUDIES

3.1 Portfolio optimization with low frequency data

First, we applied our methodology to the daily closing prices of Dow 30 stocks in the year 2011. The data are pub-

licly available in many online search engines such as Yahoo Finance.

Since the number of stocks we considered here is 30, we calculated the minimal variance portfolio consisting from 2 stocks to 30 stocks. The main challenge here is the possible portfolio combination increases exponentially when the number of stocks increases. If considering all the possibilities, the total case would be on the order of 2^{30} . This would be prohibitive to do in real applications. Hence we resort to a sampling method to get around this issue. For a fixed number of stocks, we randomly generate 10,000 samples from the 30 stocks and then performed the calculation of yield and variance based on each subsample. We recorded the global minimal variance portfolio and the corresponding sharp ratio, which is defined as

$$(6) \quad S = \frac{E[R - R_f]}{\sigma} = \frac{E[R - R_f]}{\sqrt{\text{var}[R_a - R_b]}}$$

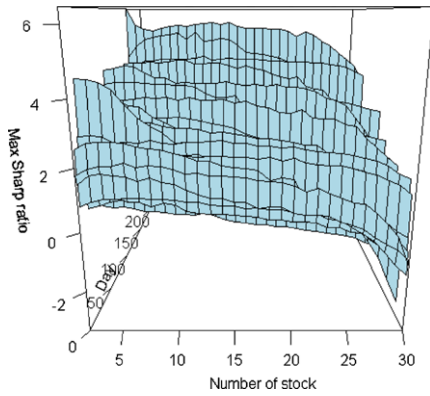
with R_f being the risk free rate. It is a measure of the excess return (or risk premium) per unit of deviation in an investment asset, typically referred to as risk (and is a deviation risk measure).

In most cases, the maximum sharp ratio first increases when we include more stock but decreases when the number of stock exceeds a certain threshold (Figure 1a). For each day, the cusp point is different, ranging from two to seven visually. If we take the transaction fees into consideration, it may even support the preference of choosing those portfolios consisting of a small number of stocks. The maximum sharp ratio across trading days can vary a lot (Figure 1b) and its volatility is highly correlated with the market performance (Figure 1c).

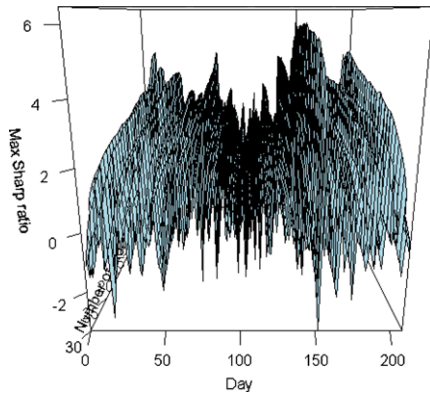
To check how the number of sampling could affect the possibility of capturing the maximum sharp ratio, we run another experiment with 100,000 sampling and the comparison is shown in Figure 2. With the increased number of sampling, the distribution of maximum sharp ratio does shift right a little. Overall the difference is not significant. Considering the big increase of computational cost when increasing the sampling from 10,000 to 100,000, we kept the sampling to 10,000 in the following experiments.

3.2 Portfolio optimization with high-frequency data

In this section, we focus on the portfolio allocation problem using high-frequency financial data. High-frequency financial data usually refer to intra-day observations. The economic value of analyzing high-frequency financial data is now obvious, both in the academic and financial world. It is the basis of intra-day and daily risk monitoring and forecasting, an input to the portfolio allocation process, and also for the increasingly popular high-frequency trading among investment banks and hedge funds. The high-frequency trading data for this study are downloaded from the Whar-



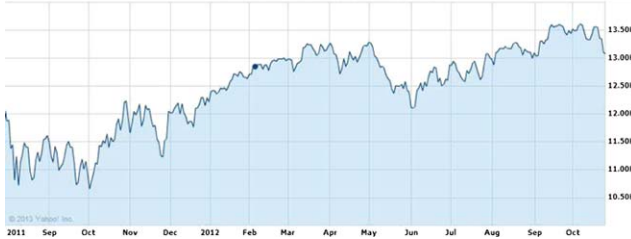
(a) Relationship between maximum sharp ratio and number of stocks



(b) Dynamics of maximum sharp ratio across the whole year

Prices						
Date	Open	High	Low	Close	Volume	Adj Close*
Aug 30, 2011	11,532.13	11,630.07	11,429.39	11,559.95	1,820,900	11,559.95
Aug 29, 2011	11,286.85	11,541.78	11,286.58	11,539.25	1,775,400	11,539.25
Aug 26, 2011	11,149.82	11,326.43	10,929.20	11,284.54	2,444,100	11,284.54
Aug 25, 2011	11,321.02	11,406.39	11,106.76	11,149.82	2,550,500	11,149.82
Aug 24, 2011	11,175.78	11,331.57	11,113.04	11,320.71	2,273,800	11,320.71
Aug 23, 2011	10,854.58	11,176.84	10,854.43	11,176.76	2,441,300	11,176.76
Aug 22, 2011	10,820.37	11,020.55	10,820.37	10,854.65	2,267,200	10,854.65
Aug 19, 2011	10,989.75	11,086.40	10,801.41	10,817.65	3,363,700	10,817.65
Aug 18, 2011	11,406.27	11,406.50	10,881.60	10,990.58	3,085,200	10,990.58
Aug 17, 2011	11,392.01	11,529.67	11,322.30	11,410.21	1,712,800	11,410.21
Aug 16, 2011	11,480.48	11,488.01	11,292.63	11,405.93	1,878,000	11,405.93
Aug 15, 2011	11,269.85	11,484.60	11,269.85	11,482.90	1,881,200	11,482.90
Aug 12, 2011	11,143.46	11,346.67	11,142.18	11,269.02	2,280,300	11,269.02
Aug 11, 2011	10,729.85	11,278.90	10,729.85	11,143.31	3,931,900	11,143.31
Aug 10, 2011	11,228.00	11,228.00	10,686.49	10,719.94	3,963,000	10,719.94
Aug 9, 2011	10,810.91	11,244.01	10,604.07	11,239.77	4,314,100	11,239.77
Aug 8, 2011	11,433.93	11,434.09	10,809.85	10,809.85	4,799,800	10,809.85
Aug 5, 2011	11,383.98	11,555.41	11,139.00	11,444.61	4,063,100	11,444.61
Aug 4, 2011	11,893.86	11,893.94	11,372.14	11,383.68	3,007,600	11,383.68
Aug 3, 2011	11,863.74	11,904.91	11,700.34	11,896.44	1,982,200	11,896.44
Aug 2, 2011	12,129.77	12,130.30	11,865.56	11,866.62	2,070,600	11,866.62
Aug 1, 2011	12,144.22	12,282.42	11,998.08	12,132.49	1,828,200	12,132.49

* Close price adjusted for dividends and splits.



(c) The high volatility region in Figure 1b corresponding to the high volatility of the whole market

Figure 1.

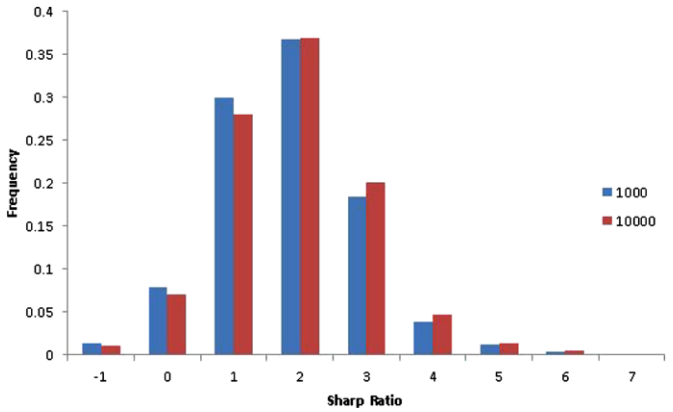


Figure 2. The distribution of maximum sharp ratio if sampling 10,000 compared with that with 100,000 sampling.

ton Research Data Service (WRDS) Database, which contains transaction-by-transaction stock prices from the 30 DJIA composite constituents traded in New York Stock Exchange (NYSE) in 2011. For this high-frequency data set, we estimated the volatility matrix between the Dow 30 stocks for each day by adopting the methodology in Section 2.1.

We found that the high-frequency data possess some unique features such as price discreteness, unequally spaced time intervals, non-synchronized trading, and leverage effect. To demonstrate the features of the realized volatility, we select the most traded stock, Bank of America (BAC), as an example.

The synchronous order of the price is dealt in this way: for every 5 min start 9:30 to 16:00, the last price at each 5-min interval is recorded as the price for the 5-min interval. Let y_t represent the 5-min interval prices and the return is expressed as $\log(y_t) - \log(y_{t-1})$. Then the daily integrated volatility is defined as

$$IV_i = \int_{T_{i-1}}^{T_i} \sigma_t^2 dt, \text{ where } T_i - T_{i-1} = 1 \text{ day}$$

and the realized volatility estimate for a univariate case is

$$RV_i = \sum_{j=1}^N r_{t-j, \Delta}^2, \text{ where } \Delta = 1 \text{ day} / N$$

where $N = 78$ for a five-minute return interval. For illustrative purposes, we found that BAC has a mean price of \$ 9.9 while the lowest daily trading volume is 6.3×10^6 and the highest trading volume is 8.6×10^8 . The realized volatilities (RV) are plotted in Figure 3 and its ACF and PACF plots are shown in Figure 4. We can see that there are some highly volatile days between the period of Day 150 to Day 200, which corresponds to the high volatility period during the first half of August in 2011 as shown in Figure 1c. The high-frequency data also exhibit some leverage effect as mentioned in the following section.

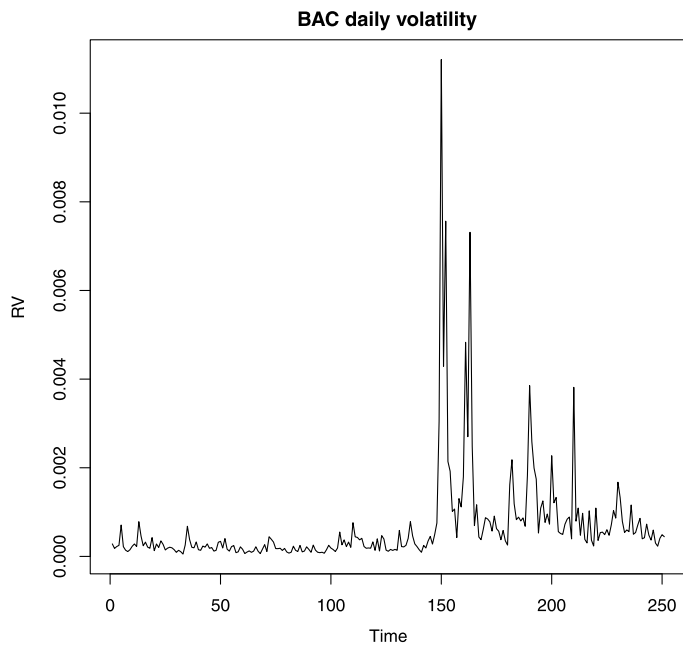


Figure 3. Daily Volatility of BAC.

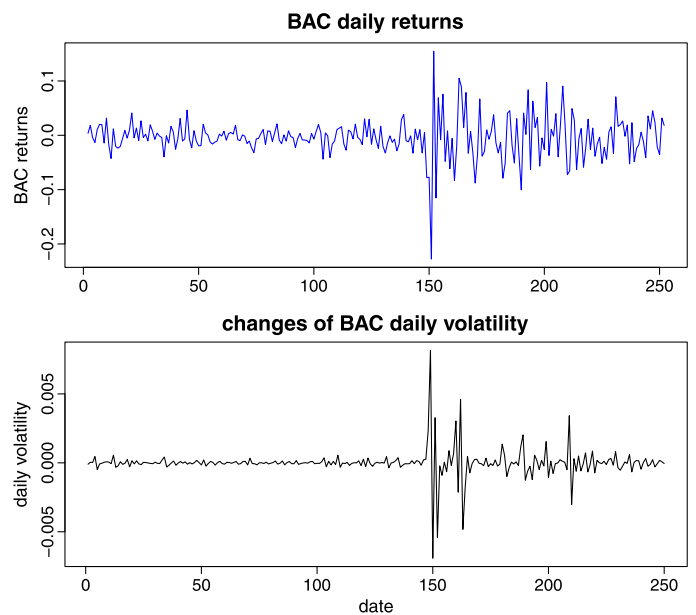


Figure 5. Leverage Effect of BAC year 2011.

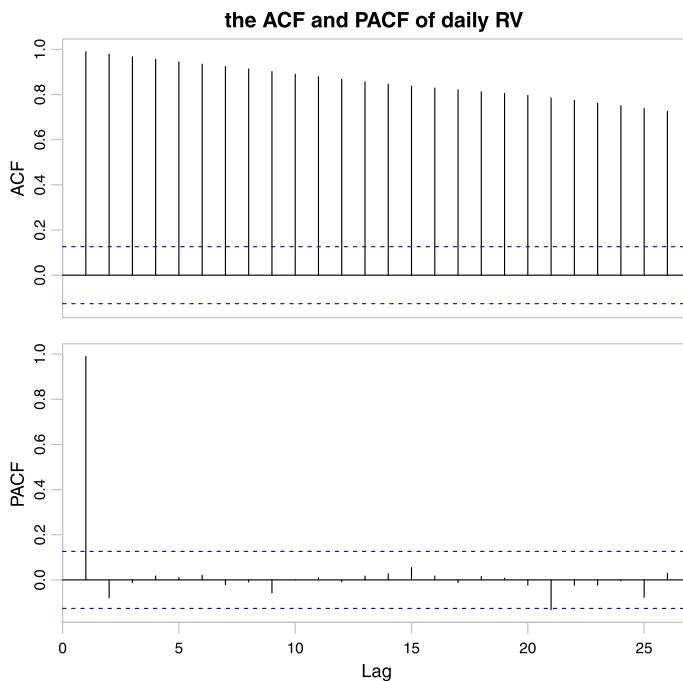


Figure 4. The ACF Plot for the integrated volatility of BAC.

Leverage effect

The leverage effect refers to the negative correlation between the daily returns and changes of the integrated daily volatility. Figure 5 shows that the daily returns of BAC and the changes of daily volatility are closely related to each other. In particular, the plot illustrates that the return and

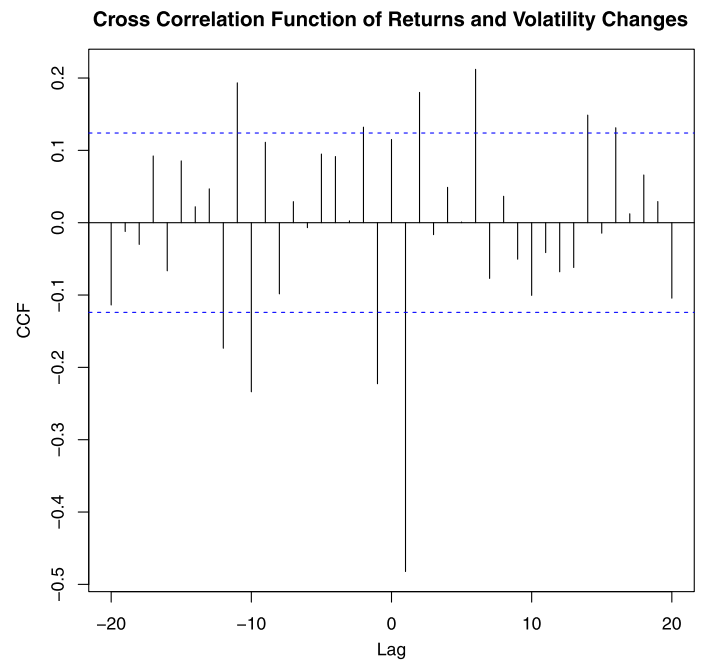
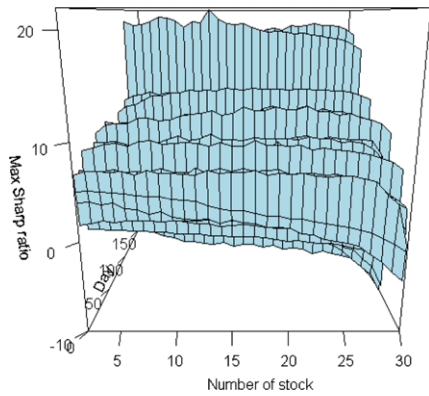
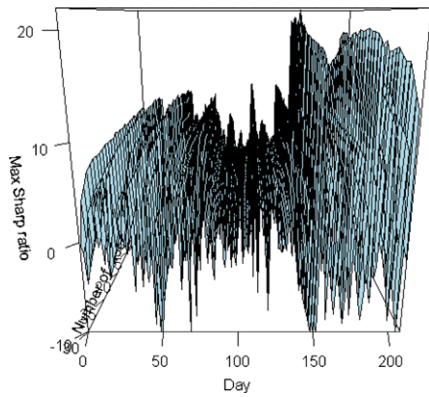


Figure 6. CCF for returns and changes of volatility of BAC.

change of volatility have a negative correlation. For example, at the highest peak where the date is around 150, there is a large return drop while the volatilities increase dramatically. Following the leverage effect plot, we also computed the cross correlation functions (CCF) between the returns and changes of volatilities, as shown in the Figure 6. A negative lag trend is also implied and confirming the existence of the leverage effect in the high-frequency data setting.



(a) Relationship between maximum sharp ratio and number of stocks



(b) Dynamics of maximum sharp ratio across the whole year

Figure 7.

Then we performed a similar permutation approach for portfolio selection as before. The relationship between the maximum sharp ratio and the number of stocks shown in the low frequency data still holds (Figure 7a), although the maximum sharp ratio patterns vary depending on the number of stocks in the portfolio.

The maximum sharp ratio across trading days (Figure 7b) also varies a lot since the co-volatilities between stocks are more dynamic. Figure 7b also shows volatility clustering, leverage effects and the non-stationary process of the stock price movement. The portfolio based on high-frequency data can achieve a much higher sharp ratio during the high volatility range of the whole market when compared with Figure 1b. This relatively strong performance is universal for the global minimum variance portfolio consisting from 2 stocks to 30 stocks (Figure 7a). From the comparison of the distribution of those larger than zero maximum sharp ratios between high-frequency data and low-frequency data (Figure 8), we could see that sharp ratios from high frequency data have a more disperse range and right shift towards higher values than those from low frequency data. It attributes to the fact that high frequency data could capture the true co-volatilities between stocks.

We also checked the composite of the maximum sharp ratio portfolio. Figure 9 shows an example of the maximum

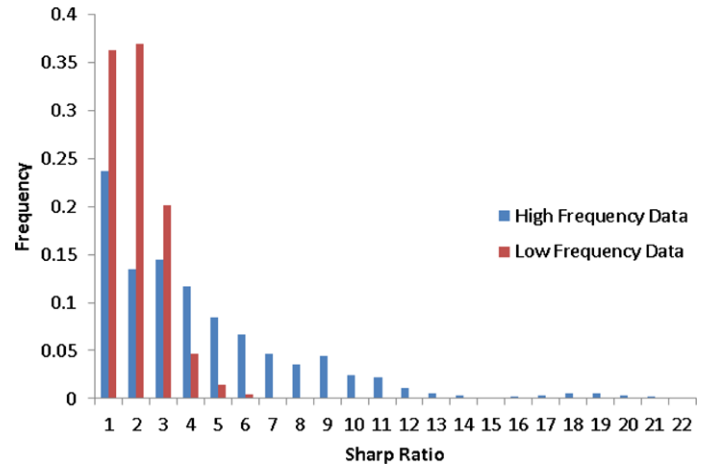


Figure 8. The distribution of maximal sharp ratios from high and low frequency data.

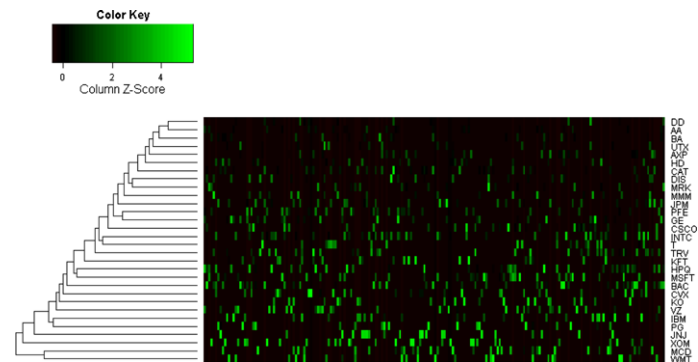


Figure 9. The composite of the maximum sharp ratio portfolio of five stocks: from left to right is the trading day in 2011 and from top to bottom lists the Dow30 stocks. The value of the heat map is scaled value based on the weight of each stock in the portfolio.

sharp ratio portfolio of five stocks. The composite changes across trading days, indicating rebalance is needed to keep maximum sharp ratio. Nevertheless, some interesting patterns conform to our existing knowledge. For example for the technical stocks, GE, CSCO, INTC, and T closely clustered together. Also PG and JNJ, BA and UTX form a cluster implying their similarity in the underlying sector.

3.3 LASSO approximation for risk minimization problem

As we pointed out before, when the number of assets are large, the permutation approach applied in the previous section is not feasible as the complexity of the optimization problem grows exponentially. Recently, Fan et al. [10] proposed that the risk minimization problem (2) for portfolio selection can be approximately transformed into a regression model with certain constraints.

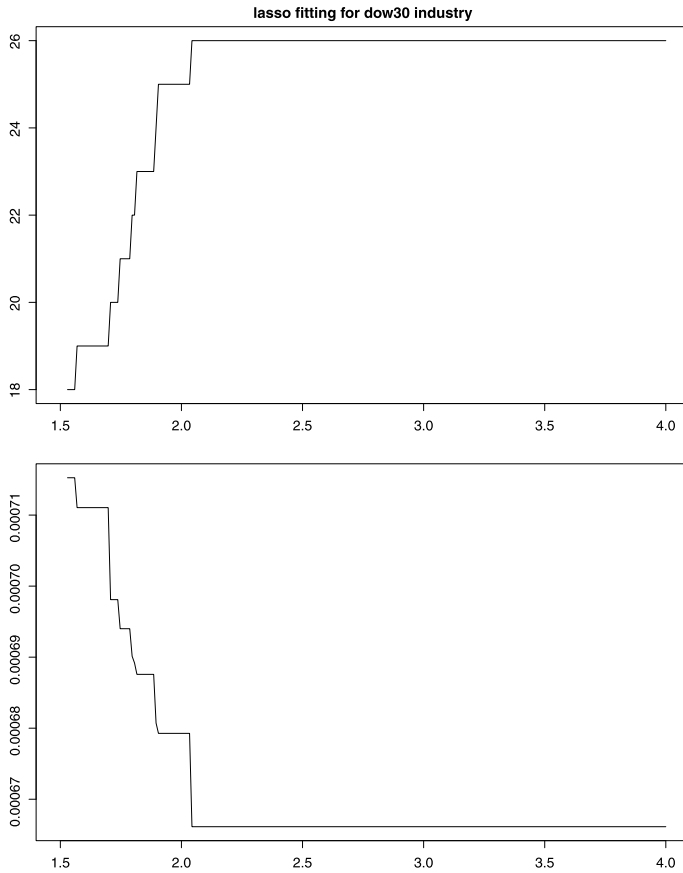


Figure 10. Lasso Fitting Results for Dow30 Industry Stocks.

(7)

$$\text{var}(\mathbf{w}^T \mathbf{R}) = \min_b \mathbb{E}(\mathbf{w}^T \mathbf{R} - b)^2$$

(8)

$$= \min_b \mathbb{E}(Y - w_1 X_1 - \dots - w_{p-1} X_{p-1} - b)^2$$

(9)

$$\approx \min_{b, \|\mathbf{w}\|_1 \leq d} \mathbb{E}(Y - \mathbf{w}^T \mathbf{X} - b)^2$$

where \mathbf{R} is the $n * p$ return matrix, $Y = R_p$, $X_j = R_p - R_j$, $j = 1, \dots, p - 1$, and \mathbf{w} is the weight vector.

We adopt the methodology to the Dow 30 portfolio allocation problem. The solution is given in Figure 10 with l_1 norm constraint d increasing from 1.5 to 4.0. As the constraint becomes larger, more stocks are selected into the model and when it comes to 2.2, almost all stocks are covered. As the constraint increases and more stocks are allowed to enter the portfolio, the risk decreases and eventually falls close to 0. Note that the LASSO optimized portfolio covariance matrix can be calculated with the following product:

$$\text{cov}(\mathbf{w}^T \mathbf{R}) = \text{diag}(\mathbf{w})^T \boldsymbol{\Sigma} \text{diag}(\mathbf{w})$$

where \mathbf{R} is the $n * p$ return matrix and \mathbf{w} is the weight vector.

The LASSO solution path plot in Figure 11 gives information about the order of the variable which enters the solution

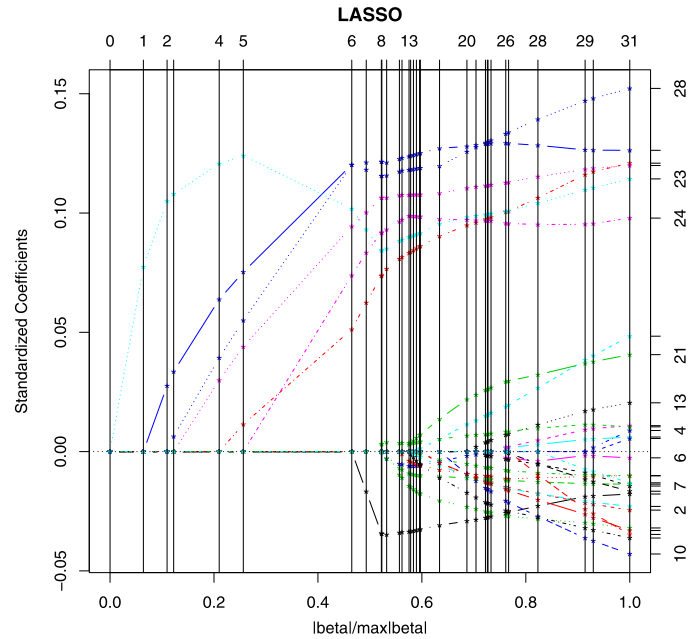


Figure 11. Lasso Solution Path.

and its values. The vertical lines convey the order of entering, and the right panel marks the index of the variables. The x-axis marks the fraction of the sum of the entered weights to the total weight. The left y-axis is the value of the coefficients. Here we applied LASSO risk minimization method to the Dow 30 data with BAC returns as the response. For example, the pink curve at the vertical line marked as 1 represents the index 23, which is PFE, enters the solution at the first place. Then following stock with index 16, which is JPM, enters the solution path at the second place.

4. DISCUSSION AND CONCLUSIONS

The portfolio optimization under the Markowitz's framework has been applied widely over the years among practitioners. While appealing in its simplicity, Markowitz's portfolio fails to capture some important characteristics of the capital markets, especially concerning wide availability of high-frequency financial data nowadays. This paper proposed an innovative methodology to perform asset allocation using high-frequency data. The empirical comparison between portfolios using low-frequency and high-frequency illustrated the efficiency of this method. The penalization based portfolio selection strategy that we proposed offers several advantages. It can significantly reduce the accumulation of small component-wise estimation errors which in turn leads to a larger overall portfolio risk. This helps answer questions about performance and risk as part of a broader investment decision-making process. Moreover, our framework can easily accommodate different positions such as short sale constraints. We illustrate the methodology with the high-frequency price data on stocks traded in New York

Table 1. Dow 30 Company Index

Company	Index
1	AA
2	AXP
3	BA
4	BAC
5	CAT
6	CSCO
7	CVX
8	DD
9	DIS
10	GE
11	HD
12	HPQ
13	IBM
14	INTC
15	JNJ
16	JPM
17	MDLZ
18	KO
19	MCD
20	MMM
21	MRK
22	MSFT
23	PFE
24	PG
25	T
26	TRV
27	UTX
28	VZ
29	WMT
30	XOM

Stock Exchange in 2011. The numerical results show that our approach performs well in portfolio allocation while pooling together the strengths of regularization and estimation from a high-frequency finance perspective. The benefits of applying the regularization approach is that we avoided looking at all possible combinations as in the permutation approach, which is computationally more attractive in practice. Furthermore, the whole solution path of the optimal portfolio gives insights on the order of dependence of different assets on the whole portfolio with respect to return and risk. This information may be valuable to practitioners when evaluating different hedging strategies. Further work could be done involving out-of-sample comparisons of the performances of the different allocations derived from different statistical methods of asset allocation strategies.

ACKNOWLEDGEMENT

We thank the editor and an anonymous referee for helpful comments that improved the quality of this paper.

Received 14 June 2013

REFERENCES

- [1] ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X., and LABYS, P. (2001). The distribution of realized exchange rate volatility. *J. Amer. Statist. Assoc.* **96**(453) 42–55. [MR1952727](#)
- [2] BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A., and SHEPHARD, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* **76**(6) 1481–1536. [MR2468558](#)
- [3] BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**(2) 253–280. [MR1904704](#)
- [4] DONOHO, D. L. and HUO, X. (2002). Beamlets and multiscale image analysis. In *Multiscale and Multiresolution Methods, Lect. Notes Comput. Sci. Eng.*, vol. 20, Springer, Berlin, pp. 149–196. [MR1928566](#)
- [5] ENGLE, R. F. and KRONER, K. F. (1995). Multivariate simultaneous generalized arch. *Econometric Theory* **11**(1) 122–150. [MR1325104](#)
- [6] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96**(456) 1348–1360. [MR1946581](#)
- [7] FAN, J., LI, Y., and YU, K. (2012a). Vast volatility matrix estimation using high frequency data for portfolio selection. *J. Am. Stat. Assoc.* To appear. [MR2949370](#)
- [8] FAN, J. and LV, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statist. Sinica* **20**(1) 101–148. [MR2640659](#)
- [9] FAN, J. and WANG, Y. (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *J. Amer. Statist. Assoc.* **102**(480) 1349–1362. [MR2372538](#)
- [10] FAN, J., ZHANG, J., and YU, K. (2012b). Asset allocation and risk assessment with gross exposure constraints for vast portfolios. *J. Am. Stat. Assoc.* To Appear.
- [11] FLEMING, J., KIRBY, C., and OSTDIEK, B. (2003). The economic value of volatility timing using “realized” volatility. *Journal of Financial Economics* **67**(3) 473–509.
- [12] GUO, J., LEVINA, E., MICHAILIDIS, G., and ZHU, J. (2011). Joint estimation of multiple graphical models. *Biometrika* **98**(1) 1–15. [MR2804206](#)
- [13] HAYASHI, T. and YOSHIDA, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* **11**(2) 359–379. [MR2132731](#)
- [14] JAGANNATHAN, R. and MA, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance* **58** 1651–1684.
- [15] LEVINA, E., ROTHMAN, A., and ZHU, J. (2008). Sparse estimation of large covariance matrices via a nested Lasso penalty. *Ann. Appl. Stat.* **2**(1) 245–263. [MR2415602](#)
- [16] LIU, Q. (2009). On portfolio optimization: How and when do we benefit from high-frequency data? *Journal of Applied Econometrics* **24**(4) 560–582. [MR2668786](#)
- [17] MANCINO, M. E. and SANFELICI, S. (2008). Robustness of Fourier estimator of integrated volatility in the presence of microstructure noise. *Comput. Statist. Data Anal.* **52**(6) 2966–2989. [MR2424771](#)
- [18] MARKOWITZ, H. M. (1952). Portfolio selection. *Journal of Finance* **7** 77–91.
- [19] TAO, M., WANG, Y., YAO, Q., and ZOU, J. (2011). Large volatility matrix inference via combining low-frequency and high-frequency approaches. *J. Amer. Statist. Assoc.* **106**(495) 1025–1040. [MR2894761](#)
- [20] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58**(1) 267–288. [MR1379242](#)
- [21] WANG, Y. and ZOU, J. (2010). Vast volatility matrix estimation for high-frequency financial data. *Ann. Statist.* **38**(2) 943–978. [MR2604708](#)

- [22] YUAN, M. and LIN, Y. (2007). Model election and estimation in the gaussian graphical model. *Biometrika* **94** 19–35. [MR2367824](#)
- [23] ZHANG, L. (2006). Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. *Bernoulli* **12**(6) 1019–1043. [MR2274854](#)
- [24] ZHANG, L., MYKLAND, P. A., and AÏT-SAHALIA, Y. (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.* **100**(472) 1394–1411. [MR2236450](#)
- [25] ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101**(476) 1418–1429. [MR2279469](#)
- [26] ZOU, J. and WU, Y. (2012). Large portfolio allocation using high-frequency financial data. Manuscript.

Jian Zou
Department of Mathematical Sciences
Indiana University-Purdue University Indianapolis
Indianapolis, IN 46202
USA
E-mail address: zoujian@iupui.edu

Yu Wang
Department of Mathematical Sciences
Indiana University-Purdue University Indianapolis
Indianapolis, IN 46202
USA