Revista Colombiana de Estadística
Diciembre 2011, volumen 34, no. 3, pp. 567 a 588

Comparison among High Dimensional Covariance Matrix Estimation Methods

Comparación entre métodos de estimación de matrices de covarianza de alta dimensionalidad

Karoll Gómez[1,3,a], Santiago Gallón[2,3,b]

[1]Departamento de Economía, Facultad de Ciencias Humanas y Económicas, Universidad Nacional de Colombia, Medellín, Colombia

[2]Departamento de Estadística y Matemáticas - Departamento de Economía, Facultad de Ciencias Económicas, Universidad de Antioquia, Medellín, Colombia

[3]Grupo de Econometría Aplicada, Facultad de Ciencias Económicas, Universidad de Antioquia, Medellín, Colombia

Abstract

Accurate measures of the volatility matrix and its inverse play a central role in risk and portfolio management problems. Due to the accumulation of errors in the estimation of expected returns and covariance matrix, the solution to these problems is very sensitive, particularly when the number of assets (p) exceeds the sample size (T). Recent research has focused on developing different methods to estimate high dimensional covariance matrixes under small sample size. The aim of this paper is to examine and compare the minimum variance optimal portfolio constructed using five different estimation methods for the covariance matrix: the sample covariance, RiskMetrics, factor model, shrinkage and mixed frequency factor model. Using the Monte Carlo simulation we provide evidence that the mixed frequency factor model and the factor model provide a high accuracy when there are portfolios with p closer or larger than T.

***Key words*:** Covariance matrix, High dimensional data, Penalized least squares, Portfolio optimization, Shrinkage.

Resumen

Medidas precisas para la matriz de volatilidad y su inversa son herramientas fundamentales en problemas de administración del riesgo y portafolio. Debido a la acumulación de errores en la estimación de los retornos esperados y la matriz de covarianza la solución de estos problemas son muy sensibles, en particular cuando el número de activos (p) excede el tamaño muestral (T).

[a]Assistant professor. E-mail: kgomezp@unal.edu.co
[b]Assistant professor. E-mail: santiagog@udea.edu.co

La investigación reciente se ha centrado en desarrollar diferentes métodos para estimar matrices de alta dimensión bajo tamaños muestrales pequeños. El objetivo de este artículo consiste en examinar y comparar el portafolio óptimo de mínima varianza construido usando cinco diferentes métodos de estimación para la matriz de covarianza: la covarianza muestral, el RiskMetrics, el modelo de factores, el shrinkage y el modelo de factores de frecuencia mixta. Usando simulación Monte Carlo hallamos evidencia de que el modelo de factores de frecuencia mixta y el modelo de factores tienen una alta precisión cuando existen portafolios con p cercano o mayor que T.

***Palabras clave*:** matrix de covarianza, datos de alta dimension, mínimos cuadrados penalizados, optimización de portafolio, shrinkage.

1. Introduction

It is well known that the volatility and correlation of financial asset returns are not directly observed and have to be calculated from return data. An accurate measure of the volatility matrix and its inverse is fundamental in empirical finance with important implications for risk and portfolio management. In fact, the optimal portfolio allocation requires solving the Markowitz's mean-variance quadratic optimization problem, which is based on two inputs: the expected (excess) return for each stock and the associated covariance matrix. In the case of portfolio risk assessment, the smallest and highest eigenvalues of the covariance matrix are referred to as the minimum and maximum risk of the portfolio, respectively. Additionally, the volatility itself has also become an underlying asset of the derivatives that are actively traded in the financial market of futures and options.

Consequently, many applied problems in finance require a covariance matrix estimator that is not only invertible, but also well-conditioned. A symmetric matrix is well-conditioned if the ratio of its maximum and minimum eigenvalues is not too large. Then it has full-rank and can be inverted. An ill-conditioned matrix has a very large ratio and is close to being numerically non-invertible. This can be an issue especially in the case of large-dimensional portfolios. The larger number of assets p with respect to the sample size T, the more spread out the eigenvalues obtained from a sample covariance matrix due to the imprecise estimation of this input (Bickel & Levina 2008).

Therefore, the optimal portfolio problem is very sensitive to errors in the estimates of inputs. This is especially true when the number of stocks under consideration is large compared to the return history in the sample. Traditionally the literature, the inversion matrix maximizes the effects of errors in the input assumptions and, as a result, practical implementation is problematic. In fact, those can produce the allocation vector that we get based on the empirical data can be very different from the allocation vector we want based on the theoretical inputs, due to the accumulation of estimation errors (Fan, Zhang & Yu 2009). Also, Chopra & Ziemba (1993) showed that small changes in the inputs can produce large changes in the optimal portfolio allocation. These simple arguments suggest that severe problems might arise in the high-dimensional Markowitz problem.

Covariance estimation for high dimensional vectors is a classical difficult problem, sometimes referred as the "curse of dimensionality". In recent years, different parametric and nonparametric methods have been proposed to estimate a high dimensional covariance matrix under small sample size. The most usual candidate is the empirical sample covariance matrix. Unfortunately, this matrix contains severe estimation errors. In particular, when solving the high-dimensional Markowitz problem, one can be underestimating the variance of certain portfolios, that is the optimal vectors of weights (Chopra & Ziemba 1993).

Other nonparametric methods such as 250-day moving average, RiskMetrics exponential smoother and exponentially weighted moving average with different weighting schemes have long been used and are widely adopted particularly for market practitioners. More recently, with the availability of high frequency databases, the technique of realized covariance proposed by Barndorff-Nielsen & Shephard (2004) has gained popularity, given that high frequency data provides opportunities for better inference of market behavior.

Parametric methods have been also proposed. Multivariate GARCH models –MGARCH– were introduced by Bollerslev, R. & Wooldridge (1988) with their early work on time-varying covariance in large dimensions, developing the diagonal vech model and later the constant correlation model (Bollerslev 1990). In general, this family model captures the temporal dependence in the second-order moments of asset returns. However, they are heavily parameterized and the problem becomes computationally unfeasible in a high dimension system, usually for $p \geq 100$ (Engle, Shephard & Sheppard 2008).

A useful approach to simplifying the dynamic structure of the multivariate volatility process is to use a factor model. Fan, Fan & Lv (2008) showed that the factor model is one of the most frequently used effective ways to achieve dimension-reduction. Given that financial volatilities move together over time across assets and markets is reasonable to impose a factor structure (Anderson, Issler & Vahid 2006). The three factor model of Fama & French (1992) is the most widely used in financial literature. Another approach that has been used to reduce the noise inherent in covariance matrix estimators is the shrinkage technique by Stein (1956). Ledoit & Wolf (2003) used this approach to decrease the sensitivity of the high-dimensional Markowitz-optimal portfolios to input uncertainty.

In this paper we examine and compare the minimum variance optimal portfolios constructed using five methods of estimating high dimensional covariance matrix: the sample covariance, RiskMetrics, shrinkage estimator, factor model and mixed frequency factor model. These approaches are widely used both by practitioners and academics. We use the global portfolio variance minimization problem with the gross exposure constraint proposed by Fan et al. (2009) for two reasons: *i*) to avoid the effect of estimation error in the mean on portfolio weights and *ii*) the error accumulation effect from estimation of vast covariance matrices.

The goal of this study is to evaluate the performance of the different methods in terms of their precision to estimate a covariance matrix in the high dimensional

minimum variance optimal portfolios allocation context.[1] The simulated Fama-French three factor model was used to generate the returns of $p = 200$ and $p = 500$ stocks over a period of 1 and 3 years of daily and intraday data. Using the Monte Carlo simulation we provide evidence than the mixed frequency factor model and the factor model using daily data show a high accuracy when there are portfolios with p closer or larger than T.

The paper is organized as follows. In Section 2, we present a general review of different methods to estimate high dimensional covariance matrices. In Section 3, we describe the global portfolio variance minimization problem with the gross exposure constraint proposed by Fan et al. (2009), and the optimization methodology used to solve it. In Section 4, we compare the minimum variance optimal portfolio obtained using simulated stocks returns and five different estimation methods for the covariance matrix. Also in this section we include an empirical study using the data of 100 industrial portfolios by Kenneth French web site. Finally, in Section 5 we conclude.

2. General Review of High Dimensional Covariance Matrix Estimators

In this Section, we introduce different methods to estimate the high dimensional covariance matrix which is the input for the portfolio variance minimization problem. Let us first introduce some notation used throughout the paper. Consider a p-dimensional vector of returns, $\boldsymbol{r}_t = (r_{1t}, \ldots, r_{pt})'$, on a set of p stocks with the associated $p \times p$ covariance matrix, $\boldsymbol{\Sigma}_t$, $t = 1, \ldots, T$.

2.1. Sample Covariance Matrix

The most usual candidate for estimating $\boldsymbol{\Sigma}$ is the empirical sample covariance matrix. Let $\boldsymbol{R}$ be a $p \times T$ matrix of p returns on T observations. The sample covariance matrix is defined by

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{T-1}\boldsymbol{R}\left(\boldsymbol{I} - \frac{1}{T}\boldsymbol{\imath}\boldsymbol{\imath}'\right)\boldsymbol{R}' \tag{1}$$

where $\boldsymbol{\imath}$ denotes a $T \times 1$ vector of ones and $\boldsymbol{I}$ is the identity matrix of order T.[2] The (i,j)th element of $\boldsymbol{\Sigma}$ is $\boldsymbol{\Sigma}^{ij} = (T-1)^{-1}\sum_{t=1}^{T}\left(r_t^i - \bar{r}^i\right)\left(r_t^j - \bar{r}^i\right)$ where r_t^i and r_t^j are the ith and jth returns of the assets i and j on $t = 1, \ldots, T$, respectively; and $\bar{r}^i$ is the mean of the ith return.

[1] Other authors have compared a set of models which are suitable to handle large dimensional covariance matrices. Voev (2008) compares the forecasting performance and also proposes a new methodology which improves the sample covariance matrix. Lam, Fung & Yu (2009) also compare the predictive power of different methods.

[2] When $p \geq T$ the rank of $\hat{\boldsymbol{\Sigma}}$ is $T-1$ which is the rank of the matrix $\boldsymbol{I} - \frac{1}{T}\boldsymbol{\imath}\boldsymbol{\imath}'$, thus it is not invertible. Then, when p exceeds $T-1$ the sample covariance matrix is rank deficient, (Ledoit & Wolf (2003)).

Although the sample covariance matrix is always unbiased estimator is well known that the sample covariance matrix is an extremely noisy estimator of the population covariance matrix when p is large (Dempster 1979).[3] Indeed, estimation of covariance matrix for samples of size T from a p-variate Gaussian distribution, $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}_p)$, has unexpected features if both p and T are large such as extreme eigenvalues of $\boldsymbol{\Sigma}_p$ and associated eigenvectors (Bickel & Levina 2008).[4]

2.2. Exponentially Weighted Moving Average Methods

Morgan's RiskMetrics covariance matrix, which is very popular among market practitioners, is just a modification of the sample covariance matrix which is based on an exponentially weighted moving average method. This method attaches greater importance on the more recent observations while further observations on the past have smaller exponential weights. Let us denote $\boldsymbol{\Sigma}_{RM}$ the RiskMetrics covariance matrix, the (i, j)th element is given by

$$\boldsymbol{\Sigma}_{RM}^{ij} = (1-\omega)\sum_{t=1}^{T} \omega^{t-1} \left(r_t^i - \bar{r}^i\right)\left(r_t^j - \bar{r}^j\right) \tag{2}$$

where $0 < \omega < 1$ is the decay factor. Morgan (1996) suggest to use a value of 0.94 for this factor. It can be write also as follows:

$$\boldsymbol{\Sigma}_{RM,t} = \omega \boldsymbol{r}_{t-1}\boldsymbol{r}_{t-1}' + (1-\omega)\boldsymbol{\Sigma}_{RM,t-1}$$

which correspond a BEKK scalar integrated model by Engle & Kroner (1995).

Other straightforward methods such as rolling averages and exponentially weighted moving average using different weighting schemes have long been used and are widely adopted specially among practitioners.

2.3. Shrinkage Method

Regularizing large covariance matrices using the Stein (1956) shrinkage method have been used to reduce the noise inherent in covariance estimators. In his seminal paper Stein found that the optimal trade-off between bias and estimation error can be handled simply taking properly a weighted average of the biased and unbiased estimators. This is called shrinking the unbiased estimator full of estimation error towards a fixed target represented by the biased estimator.

This procedure improved covariance estimation in terms of efficiency and accuracy. The shrinkage pulls the most extreme coefficients towards more central values, systematically reducing estimation error where it matters most. In summary, such method produces a result to exhibit the following characteristics: i) the

[3]There is a fair amount of theoretical work on eigenvalues of sample covariance matrices of Gaussian data. See Johnstone (2001) for a review.

[4]For example, the larger p/T the more spread out the eigenvalues of the sample covariance matrix, even asymptotically.

estimate should always be positive definite, that is, all eigenvalues should be distinct from zero and ii) the estimated covariance matrix should be well-conditioned.

Ledoit & Wolf (2003) used this approach to decrease the sensitivity of the high-dimensional Markowitz-optimal portfolios to input uncertainty. Let us denote $\mathbf{\Sigma}_S$ the shrinkage estimators of the covariance matrix, which generally have the form

$$\mathbf{\Sigma}_S = \alpha \boldsymbol{F} + (1-\alpha)\widehat{\mathbf{\Sigma}} \tag{3}$$

where $\alpha \in [0,1]$ is the shrinkage intensity optimally chosen, $\boldsymbol{F}$ corresponds to a positive definite matrix which is the target matrix and $\widehat{\mathbf{\Sigma}}$ represents the sample covariance matrix.

The shrinkage intensity is chosen as the optimal α with respect to a loss function (risk), $L(\alpha)$, defined as a quadratic measure of distance between the true and the estimated covariance matrices based on the Frobenius norm. That is

$$\alpha^* = \arg\min \mathbb{E}\left[\left\|\alpha \boldsymbol{F} + (1-\alpha)\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|^2\right]$$

Given that α^* is non observable, Ledoit & Wolf (2004) proposed a consistent estimator of α for the case when the shrinkage target is a matrix in which all pairwise correlations are equal to the same constant. This constant is the average value of all pairwise correlations from the sample covariance matrix. The covariance matrix resulting from combining this correlation matrix with the sample variances, known as equicorrelated matrix, is the shrinkage target.

Ledoit & Wolf (2003) also proposed to estimate the covariance matrix of stock returns by an optimally weighted average of two existing estimators: the sample covariance matrix with the single-index covariance matrix or the identity matrix.[5]

An alternative method frequently used proposes banding the sample covariance matrix or estimating a banded version of the inverse population covariance matrix. A relevant assumption, in particular for time series data, is that the covariance matrix is banded, meaning that the entries decay based on their distance from the diagonal. Thus, Furrer & Bengtsson (2006) proposed to shrink the covariance entries based on this distance from the diagonal. In other words, this method keeps only the elements in a band along its diagonal and gradually shrinking the off-diagonal elements toward zero.[6] Wu & Pourahmadi (2003) and Huang, Liu, Pourahmadi & Liu (2006) estimate the banded inverse covariance matrix by using thresholding and L_1penalty, respectively.[7]

2.4. Factor Models

The factor model is one of the most frequently used effective ways for dimension reduction, and a is widely accepted statistical tool for modeling multivariate

[5]The single-index covariance matrix corresponds to a estimation using one factor model given the strong consensus about the use of the market index as a natural factor.

[6]This method is also known how "tapering" the sample covariance matrix.

[7]Thresholding a matrix is to retain only the elements whose absolute values exceed a given value and replace others by zero.

volatility in finance. If few factors can completely capture the cross sectional variability of data then the number of parameters in the covariance matrix estimation can be significatively reduced (Fan et al. 2008). Let us consider the $p \times 1$ vector $\boldsymbol{r}_t$. Then the K-factor model is written as

$$\boldsymbol{r}_t = \boldsymbol{\Lambda}\boldsymbol{f}_t + \boldsymbol{\nu}_t = \sum_{k=1}^{K} \boldsymbol{\lambda}_k \cdot f_{kt} + \boldsymbol{\nu}_t \tag{4}$$

where $\boldsymbol{f}_t = (f_{1t}, \ldots, f_{Kt})'$ is the K-dimensional factor vector, $\boldsymbol{\Lambda}$ is a $p \times K$ unknown constant loading matrix which indicates the impact of the kth factor over the ith variable, and $\boldsymbol{\nu}_t$ is a vector of idiosyncratic errors. $\boldsymbol{f}_t$ and $\boldsymbol{\nu}_t$ are assumed to satisfy

$$\begin{aligned}
&\mathbb{E}(\boldsymbol{f}_t \mid \Im_{t-1}) = \mathbf{0}, \quad \mathbb{E}(\boldsymbol{f}_t \boldsymbol{f}_t' \mid \Im_{t-1}) = \boldsymbol{\Phi}_t = \operatorname{diag}\{\phi_{1t}, \ldots, \phi_{Kt}\}, \\
&\mathbb{E}(\boldsymbol{\nu}_t \mid \Im_{t-1}) = \mathbf{0}, \quad \mathbb{E}(\boldsymbol{\nu}_t \boldsymbol{\nu}_t' \mid \Im_{t-1}) = \boldsymbol{\Psi} = \operatorname{diag}\{\psi_1, \ldots, \psi_p\}, \\
&\mathbb{E}(\boldsymbol{f}_t \boldsymbol{\nu}_t' \mid \Im_{t-1}) = \mathbf{0}.
\end{aligned}$$

where $\Im_{t-1}$ denotes the information set available at time $t-1$.

The covariance matrix of $\boldsymbol{r}_t$ is given by

$$\boldsymbol{\Sigma}_{F,t} = \mathbb{E}(\boldsymbol{r}_t \boldsymbol{r}_t' \mid \Im_{t-1}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}_t\boldsymbol{\Lambda}' + \boldsymbol{\Psi} = \sum_{k=1}^{K} \boldsymbol{\lambda}_k \boldsymbol{\lambda}_k' \phi_{kt} + \boldsymbol{\Psi} \tag{5}$$

where all the variance and covariance functions depend on the common movements of f_{kt}.

The multi-factor model which utilizes observed market returns as factors has been widely used both theoretically and empirically in economics and finance. It states that the excessive return of any asset r_{it} over the risk-free interest rate satisfies the equation above. Fama & French (1992) identified three key factors that capture the cross-sectional risk in the US equity market, which have been widely used. For instance, the Capital Asset Pricing Model −CAPM− uses a single factor to compare the excess returns of a portfolio with the excess returns of the market as a whole. But it oversimplifies the complex market. Fama and French added two more factors to CAPM to have a better description of market behavior. They proposed the "small market capitalization minus big" and "high book-to-price ratio minus low" as possible factors. These measure the historic excess returns of small caps over big caps and of value stocks over growth stocks, respectively. Another choice is macroeconomic factors such as: inflation, output and interest rates; and the third possibility are statistical factors which work under a purely dimension-reduction point of view.

The main advantage of statistical factors is that it is very easy to build the model. Fan et al. (2008) find that the major advantage of factor models is in the estimation of the inverse of the covariance matrix and demonstrate that the factor model provides a better conditioned alternative to the fully estimated covariance matrix. The main disadvantage is that there is no clear meaning for the factors. However, a lack of interpretability is not much of a handicap for portfolio optimization. Peña & Box (1987), Chan, Karceski & Lakonishok (1999), Peña & Poncela (2006), Pan & Yao (2008) and Lam & Yao (2010) among others have studied the covariance matrix estimate based on the factor model context.

2.5. Realized Covariance

More recently, with the availability of high frequency databases, the technique of realized volatility introduced by Andersen, Bollerslev, Diebold & Labys (2003) in a univariate setting has gain popularity. In a multivariate setting, Barndorff-Nielsen & Shephard (2004) proposed the realized covariance $-RCV-$, which is computed by adding the cross products of the intra-day returns of two assets. Dividing day t into M non-overlapping intervals of length $\Delta = 1/M$, the realized covariance between assets i and j can be obtained by

$$\mathbf{\Sigma}^{\Delta}_{RCV,t} = \sum_{m=1}^{M} r^{i}_{t,m} r^{j}_{t,m} \tag{6}$$

where $r^{i}_{t,m}$ is the continuously compounded return on asset i during the mth interval on day t.

The RCV based on the synchronized discrete observations of the latent process is a good proxy or representative of the integrated covariance matrix. Barndorff-Nielsen & Shephard (2004) showed that this is true in the low dimensional case. However, in the high dimensional case, i.e. when the dimension p is not small compared with T, it is in general not a good proxy (Zheng & Li 2010). This is a consequence of several issues related with non-synchronous trading, market microstructure noise and spurious intra-day dependence.

Indeed, estimating high dimensional integrated covariance matrix has been drawing more attention. Several solutions have been proposed that are robust to these frictions. Bannouh, Martens, Oomen & van Dijk (2010) propose a Mixed-Frequency Factor Model $-$MFFM$-$ for estimating the daily covariance matrix for a vast number of assets, which aims to exploit the benefits of high-frequency data and a factor structure. They proposed to obtain the factor loadings in the conventional way by linear regression using daily stock information, and calculated the factor covariance matrix and residual variances with high precision from intra-day data. Using this approach they can avoid non-synchronicity problems inherent in the use of high frequency data for individual stocks.

Considering the same linear factor structure specified in (4), the covariance matrix can be defined as before:

$$\mathbf{\Sigma}_{MFFM} = \mathbf{\Lambda \Pi \Lambda}' + \mathbf{\Theta} \tag{7}$$

where $\mathbf{\Pi} = \mathbb{E}(\mathcal{F}\mathcal{F}')$ is the realized covariance matrix obtained using $\mathcal{F}$ high-frequency factor return observations. $\mathbf{\Lambda}$ denotes the factor loadings, and $\mathbf{\Theta}$ the idiosyncratic residuals, which are obtained using $\boldsymbol{\nu} = \mathcal{R} - \mathbf{\Lambda}\mathcal{F}$ where $\mathcal{R}$ denotes the high-frequency matrix return observations.

This methodology has several advantages over the realized covariance matrix. First, the advantages of dimension reduction in the context of the factor model based purely on daily data continue to hold in the MFFM. Second, the MFFM makes efficient use of high-frequency factor data while bypassing potentially severe biases induced by microstructure noise for the individual assets. Third, we can

easily expand the number of assets in the MFFM approach while this is more difficult with the RC matrix for which the inverse does not exist when the number of assets exceeds the number of return observations per asset. For additional details see Bannouh et al. (2010).

Wang & Zou (2009) also develop a methodology for estimating large volatility matrices based on high frequency data. The estimator proposed is constructed in two stages: first, they propose to calculate the average of the realized volatility matrices constructed using tick method and pre sampling frequency, which is called ARVM estimator. Then, regularize ARVM estimator to yield good consistent estimators of the large integrated volatility matrix. Other proposal have been introduced by Barndorff-Nielsen, Hansen, Lunde & Shephard (2010), Zheng & Li (2010), among others.

3. Portfolio Variance Minimization Problem with the Gross Exposure Constraint

In this section, we start recalling the portfolio variance minimization problem proposed by Fan et al. (2009). The noteworthy innovation in their proposal is to relax the gross exposure constraint in order to enlarge the pools of admissible portfolios generating more diversified portfolios.[8] Moreover, they showed that there is no accumulation of estimation errors thanks to the gross exposure constraint. We also present, in a different subsection, the LARS algorithm developed by Efron, Hastie, Johnstone & Tibshirani (2004), which permits to find efficiently the solution paths to the constrained variance minimization problem.

3.1. The Variance Minimization Problem with Gross Exposure Constraint

Following the proposal of Fan et al. (2009), we suppose a portfolio with p assets and corresponding returns $\boldsymbol{r} = (r_1, \ldots, r_p)'$ to be managed. Let $\boldsymbol{\Sigma}$ be its associated covariance matrix, and $\boldsymbol{w}$ be its portfolio allocation vector. as a consequence, the variance of the portfolio return $\boldsymbol{w}'\boldsymbol{r}$ is given by $\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}$. Considering the variance minimization problem with gross-exposure constraint as follows:

$$\begin{aligned} &\min_{\boldsymbol{w}} \Gamma(\boldsymbol{w}, \boldsymbol{\Sigma}) = \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}, \\ &\text{subject to: } \boldsymbol{w}'\boldsymbol{\imath} = 1 \qquad \text{(Budget constraint)} \\ &\qquad\qquad \|\boldsymbol{w}\|_1 \leq c \qquad \text{(Gross exposure constraint)} \end{aligned} \tag{8}$$

where $\|\boldsymbol{w}\|_1$ is the L_1 norm. The constraint $\|\boldsymbol{w}\|_1 \leq c$ prevents extreme positions in the portfolio. Notice that when $\|\boldsymbol{w}\|_1 = 1$, ie $c = 1$, no short sales are allowed as studied by Jagannathan & Ma (2003); when $c = \infty$, there is no constraint on

[8]The portfolio optimization with the gross-exposure constraint bridges the gap between the optimal no-short-sale portfolio studied by Jagannathan & Ma (2003) and no constraint on short-sale in the Markowitz's framework.

short sales as in Markowitz (1952). Thus, the proposal of Fan et al. (2009) is a generalization to the work of them.[9]

The solution to the optimization problem $\boldsymbol{w}^*$ depends sensitively on the input vectors $\boldsymbol{\Sigma}$ and its accumulated estimation errors, but under the gross-exposure constraint, with a moderate value of c, the sensitive of the problem is bounded and these two problems disappear. The upper bounds on the approximation errors is given by

$$\left|\Gamma\left(\boldsymbol{w}, \boldsymbol{\Sigma}\right) - \Gamma\left(\boldsymbol{w}, \widehat{\boldsymbol{\Sigma}}\right)\right| \leq 2a_n c^2 \tag{9}$$

where $\Gamma\left(\boldsymbol{w}, \boldsymbol{\Sigma}\right)$ and $\Gamma(\boldsymbol{w}, \widehat{\boldsymbol{\Sigma}})$ correspond to the theoretical and empirical portfolio risks, $a_n = ||\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}||_\infty$ and $\widehat{\boldsymbol{\Sigma}}$ is an estimated covariance matrix based on the data with sample size T.

They point out that this holds for any estimation of covariance matrix. However as long as each element is estimated precisely, the theoretical minimum risk and the empirical risk calculated from the data should be very close, thanks to the constraint on the gross exposure.

3.2. The Optimization Methodology

The risk minimization problem described in the equation (8) takes the form of the Lasso problem developed by Tibshirani (1996). For a complete study of Lasso (Least Absolute Shrinkage and Selection Operator) method see Buhlmann & van de Geer (2011). The connection between Markowitz problem and Lasso is conceptually and computationally useful. The Lasso is a constrained version of ordinary least squares −OLS−, which minimize a penalized residual sum of squares. Markowitz problem also can be viewed as a penalized least square problem given by

$$\begin{aligned} \boldsymbol{w}^*_{\text{Lasso}} = \arg\min \sum_{t=1}^{T} \left(y_t - b - \sum_{j=1}^{p-1} x_{tj} w_j \right)^2 \\ \text{subject to } \sum_{j=1}^{p-1} |w_j| \leq d \qquad (L_1 \text{ penalty}) \end{aligned} \tag{10}$$

where $y_t = r_{tp}$, $x_{tj} = r_{tp} - r_{tj}$ with $j = 1, \ldots, p-1$ and $d = c - \left|1 - \sum_{j=1}^{p-1} w_j^*\right|$. Thus, finding the optimal weight $\boldsymbol{w}$ is equivalent to finding the regression coeffcient $\boldsymbol{w}^* = (w_1, \ldots, w_{p-1})'$ along with the intercept b to best predict y.

Quadratic programming techniques can be used to solve (8) and (10). However, Efron et al. (2004) proposed to compute the Lasso solution using the LARS algorithm which uses a simple mathematical formula that greatly reduced the

[9]Let w^+ and w^- be the total percent of long and short positions, respectively. Then, under $w^+ - w^- = 1$ and $w^+ - w^- \leq c$, we have $w^+ = (c+1)/2$ and $w^- = (c-1)/2$. These correspond to percentage of long and short positions allowed. The constraint on $\|\boldsymbol{w}\|_1 \leq c$ is equivalent to the constraint on w^-, which is binding when the portfolio is optimized.

computational burden. Fan et al. (2009) showed that this algorithm provides an accurate solution approximation of problem (8).

The LARS procedure works roughly as follows. Given a collection of possible predictors, we select the one having largest absolute correlation with the response y, say x_{j_1} ,and perform simple linear regression of y on x_{j1}. This leaves a residual vector orthogonal to x_{j_1}, which now is considered to be the response. We project the other predictors orthogonally to x_{j1} and repeat the selection process. Doing the same procedure after s steps this produce a set of predictors $x_{j_1}, x_{j_2}, \ldots, x_{j_s}$ that are then used in the usual way to construct a s-parameter linear model (Efron et al. (2004)). For more details, the LARS algorithm steps are summarized in the Appendix A

The LARS algorithm applied to the problem (10) produces the whole solution path $\boldsymbol{w}^*(d)$, for all $d \geq 0$. The number of non-vanishing weights varies as d ranges from 0 to 1. It recruits successively one stock, two stocks, and gradually all the stocks of the portfolio. When all stocks are recruited, the problem is the same as the Markowitz risk minimization problem, since no gross-exposure constraint is imposed when d is large enough (Fan et al. 2009).

4. Comparison of Minimum Variance Optimal Portfolios

In this section, we compare the minimum variance optimal portfolio constructed using five different estimation methods for the covariance matrix: the sample covariance, RiskMetrics, factor model, mixed frequency factor model and shrinkage method.

4.1. Dataset

We use a simulated return of p stocks considering 1 and 3 years of daily data, this is $T = 252, 756$. The simulated Fama-French three factor model is used to generate the returns of $p = 200$ and $p = 500$ stocks, using the specification in (4) and following the procedure employed by Fan et al. (2008). We carry out the following steps:

1. Generate p factor loading vectors $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_p$ as a random sample of size p from the trivariate normal distribution $N(\boldsymbol{\mu}_{\boldsymbol{\lambda}}, \mathrm{cov}_{\boldsymbol{\lambda}})$. This is kept fixed during the simulation.

2. Generate a random sample of factors $\boldsymbol{f}_1$, $\boldsymbol{f}_2$ and $\boldsymbol{f}_3$ of size T from the trivariate normal distribution $N(\boldsymbol{\mu}_{\boldsymbol{f}}, \mathrm{cov}_{\boldsymbol{f}})$.

3. Generate p standard deviations of the errors $\psi_1, \ldots, \psi_p$ as a random sample of size p from a gamma distribution with parameters $\alpha = 3.3586$ and $\beta = 0.1876$. This is also kept fixed during the simulation.

4. Generate a random sample of p idiosyncratic noises $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_p$ with size T from the p-variate normal distribution $N(0, \boldsymbol{\Psi})$, and also from Student's t distribution t-Stud$(6, \boldsymbol{\Psi})$.

5. Calculate a random sample of returns $\boldsymbol{r}_t$, $t = 1, \ldots, T$ using the model (4) and the information generated in steps 1, 2 and 4.

6. By means of this simulated returns we calculated the following covariance matrix using: the sample covariance, RiskMetrics, factor model and shrinkage method, as was discussed in Section 2.

The parameters used in steps 1, 2 and 3, were taken from Fan et al. (2008) who fit three-factor model using the three-year daily data of 30 Industry Portfolios from May 1, 2002 to August 29, 2005, available at the Kenneth French website. They calculated the sample means and sample covariance matrices of $\boldsymbol{f}$ and $\boldsymbol{\lambda}$ denoted by $(\boldsymbol{\mu}_{\boldsymbol{f}}, \text{cov}_{\boldsymbol{f}})$ and $(\boldsymbol{\mu}_{\boldsymbol{\lambda}}, \text{cov}_{\boldsymbol{\lambda}})$. These values are reported in Appendix B, Table 4.

Additionally, to implement the Mixed-Frequency Factor Model we simulated, as proposed by Bannouh et al. (2010), five minutes high frequency factor data $\mathcal{F}$ from a trivariate Gaussian distribution, $N(0, \text{cov}_f)$ and high frequency idiosyncratic noises from a p-variate normal distribution $N(0, \boldsymbol{\Psi})$. In practice high-frequency financial asset prices bring problems such as non-synchronous trading and are contaminated by market microstructure noise.

We implement non-synchronous trading by assuming trades arrive following a Poisson process with an intensity parameter equal to the average number of daily trades for the S&P500.[10] Also, we include a microstructure noise component in the model, $\boldsymbol{\eta} \sim N(0, \boldsymbol{\Delta})$ where $\boldsymbol{\Delta} = (1/4\tau)(\boldsymbol{\Lambda\Pi\Lambda}' + \boldsymbol{\Theta})$ with τ the high frequency sample size returns. Using this we also calculate the random sample of high frequency returns $\mathcal{R} = \boldsymbol{\Lambda}\mathcal{F} + \boldsymbol{\nu} + \boldsymbol{\eta}$ and by means of these returns we calculate (7).

Finally, from the estimated covariance matrices obtained using the different methods, we find an approximately optimal solution to problem (8) using the LARS algorithm. For this calculation, we take the no short sale constraint optimal portfolio as dependent variable in (10). Thus, having the optimal portfolio weights and the estimated covariance matrix we calculate the theoretical and empirical minimum variance optimal risk. In this paper, the risk of each optimal portfolio is referring to the standard deviation of the quantities $\Gamma(\boldsymbol{w}, \boldsymbol{\Sigma})$ and $\Gamma\left(\widehat{\boldsymbol{w}}, \widehat{\boldsymbol{\Sigma}}\right)$, calculated as the square-root thereof.

4.2. Simulation results

Fan et al. (2009) showed that the unknown theoretical minimum risk, $\Gamma(\boldsymbol{w}, \boldsymbol{\Sigma})$, and the empirical minimum risk, $\Gamma\left(\widehat{\boldsymbol{w}}, \widehat{\boldsymbol{\Sigma}}\right)$, of the invested portfolio are approximately the same as long as: i) the c is not too large and ii) the accuracy of

[10] This value corresponds to 19.385 which is the average number of daily trades over the period November 2006 through May 2008 (Bannouh et al. (2010)Z).

estimated covariance matrix is not too low. Based on this result, we are going to compare the theoretical solution path of the minimum variance optimal portfolios with the solution path obtained using five different estimation methods for high dimensional covariance matrix: the sample covariance, RiskMetrics, factor model, shrinkage and mixed frequency factor model.

We first examine the results in case of $p = 200$ with 100 replications. In Table 1, we present the mean value of the minimum variance optimal portfolio in three cases: i) when no short sales are allowed, that is $c = 1$, as studied by Jagannathan & Ma (2003), ii) under a gross exposure constraint equal to $c = 1.6$ as proposed by Fan et al. (2009), which correspond to a typical choice and iii) when $c = \infty$, that is, no constraint on short sales as in Markowitz (1952).[11]

The results show that the empirical minimum portfolio risk obtained using the covariance matrix estimated from mixed frequency factor model method has the smaller difference with respect to the theoretical risk. Thus, the MFFM method produces the better relative estimation accuracy among the competing estimators. The gains come from the fact that this model exploits the advantages of both high frequency data and the factor model approach. The factor model also permits a precise estimation of the covariance matrix, which is closer to the MFFM. The accuracy of the covariance matrix estimated from the shrinkage method is also fairly similar to the factor models and slightly superior to the sample covariance matrix.[12] Finally, all estimation methods overcome the RiskMetrics, especially when no short sales are allowed. We have the same results when we used three years of daily returns, presented at the bottom of Table 1.

Table 1: Theoretical and empirical risk of the minimum variance optimal portfolio High dimensional case ($p = 200$).

True covariance matrix		Competing estimators				
c	$\boldsymbol{\Sigma}$	$\widehat{\boldsymbol{\Sigma}}$	$\boldsymbol{\Sigma}_{RM}$	$\boldsymbol{\Sigma}_{S}$	$\boldsymbol{\Sigma}_{F}$	$\boldsymbol{\Sigma}_{MFFM}$
$T = 252$						
1	21.23	19.16	16.65	19.72	19.88	20.12
1.6	7.64	6.76	5.92	7.29	7.35	7.42
∞	1.32	0.88	0.85	0.93	1.03	1.01
$T = 756$						
1	19.84	18.53	15.53	19.01	19.15	19.45
1.6	5.85	3.82	3.05	4.95	5.05	5.53
∞	1.25	0.69	0.61	0.87	0.94	0.98

As we can see in Table 1, in all cases the theoretical risk is greater than the empirical risk, although in some cases the difference is slim. The intuition of

[11] The corresponding values for parameter d in each case is: 0, 0.7, and 12.8.

[12] We used as target matrix the identity which works well as was shown by Ledoit & Wolf (2003) and also the shrinkage target actually proposed by them. The practical problem in applying the shrinkage method is to determine the shrinkage intensity. Ledoit & Wolf (2003) showed that it behaves like a constant over the sample size and provide a way to consistently estimate it. Following the Ledoit & Wolf (2003) proposal we found $\alpha^* = 0.7895$. However, we check the stability of the results using different values for α chosen ad hoc. The results show that the as long as the shrinkage intensity is lower than α^* the methods tends to underestimate a little bit more the risk. However, this method maintains his superiority with respect to sample and RiskMetrics estimated covariance matrices. Detailed results are available upon request.

these results is that having to estimate the high dimensional covariance matrix at stake here leads to risk underestimation. In other words, in general the covariance matrix estimation leads to overoptimistic conclusions about the risk. The most dramatic case occurs with the RiskMetrics portfolio, which shows the lower risk.

Additionally, the results show that constrained short sale portfolios are not diversified enough, as also was found by Fan et al. (2009). For instance, the risks can be improved by relaxing the gross exposure constraint, which implies allowing some short positions. However, allowing the possibility of extreme short or long positions in the portfolio we can get a lower optimal risk; extremely negative weights are difficult to implement in practice. Actually, practical portfolio choices always involve constraints on individual assets such as the allocations are no larger than certain percentages of the median daily trading volume of an asset. This result is true no matter what method is used to estimate the covariance matrix and which sample size is used.

Figure 1, shows the whole path solution of the risk for a selected portfolio as a function of LARS steps. The path solution was calculated for each of the five competing methods and the true covariance matrix, using the LARS algorithm. This figure illustrates the decrease in optimal risk when we move from a portfolio with no short sale to allowed short sale portfolio, which is more diversified and therefore less risky. In other words, the graph suggests that the optimal risk decreases as soon as in each step the parameter d is growing. This occurs as long as the LARS algorithm progresses.[13] This implies that the higher value of optimal risk is reached in the case of no short sale.

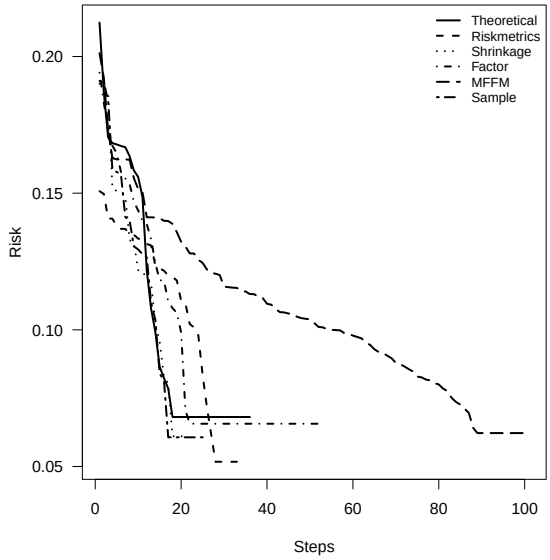

FIGURE 1: LARS solution path of the optimal risk for each minimum variance portfolio

[13]The number of steps required to complete the algorithm and have the entire solution path can be different in each case

In consequence, once the gross exposure constraint is relaxed the number of selected stocks increases and the portfolio becomes more diversified. In fact, at the first step when d is relaxed the LARS algorithm identifies the stock that permit reduction of the minimum optimal risk under no short sale restriction, permitting this stock to enter into the optimal portfolio allocation with a weight that can be positive or negative. This process is continued until the entire set of stocks are examined and as result in each step you will have a decreasing optimal risk but increasing short percentage. This process is illustrated in Figure 2. Each graph in the panel corresponds to a profile of optimal portfolio weights obtained solving the problem (10) using the true covariance matrix and each estimated covariance matrix.

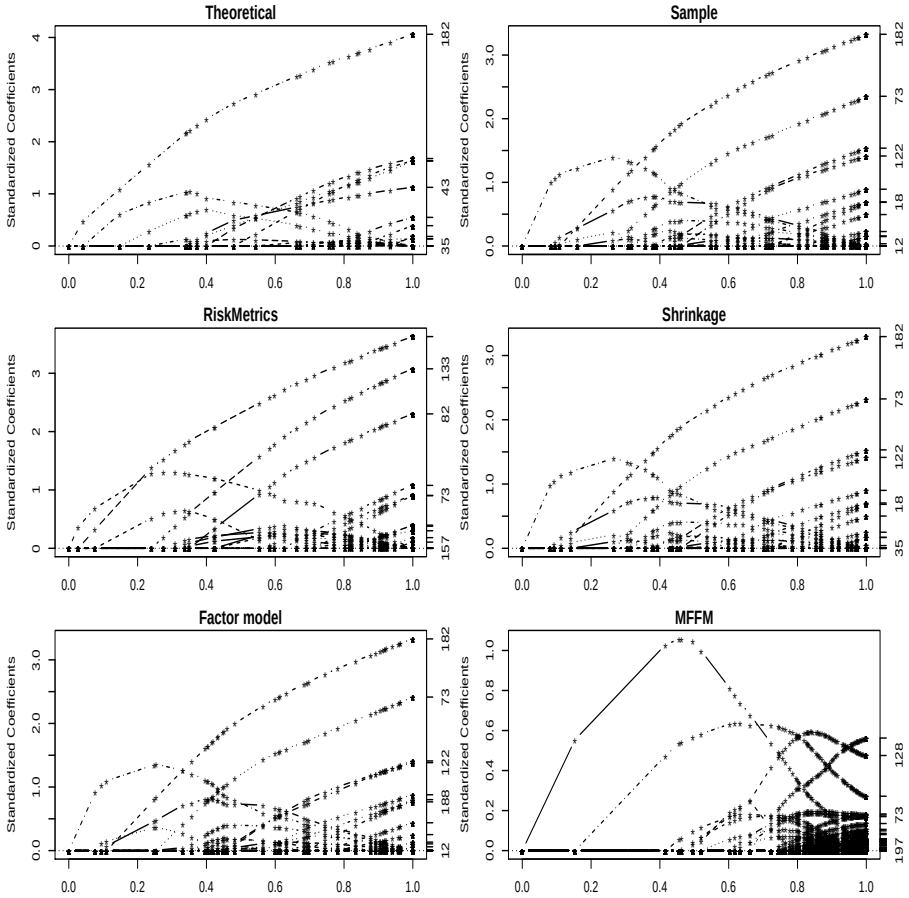

FIGURE 2: Estimated optimal portfolio weights via the Lasso. The abscissae correspond to the standardized Lasso parameter, $s = d/\sum_{j=1}^{p-1} |w_j|$.

The figure shows the optimal portfolio weights as a function of the standardized Lasso parameter $s = d/\sum_{j=1}^{p-1} |w_j|$. Each curve represents the optimal weight of a particular stock in the portfolio as s is varied. We start with no short sale portfolio

at $s = 0$. The stocks begin to enter in the active set sequentially as d increases, allowing us to have a more diversified portfolio. Finally, at $s = 1$, the graph shows the stocks that are included in the active stock set where short sales are allowed with no restriction. The number of some of them are labeled on the right side in each graph.[14]

We now examine the results in case of $p = 500$, again with 100 replications. The results, considering this very high dimensional case, are presented in Table 2. Similarly, this table contains the mean value of the minimum variance optimal portfolio risk using different estimation methods for covariance matrix. First of all, as we can see, sampling variability for the case with 500 stocks is smaller than the case involuing 200 stocks. These are due to the fact that with more stocks, the selected portfolio is generally more diversified and hence the risks are generally smaller. This result is according with the founded results by Fan et al. (2009).

Table 2: Theoretical and empirical risk of the minimum variance optimal portfolio Very high dimensional case ($p = 500$).

True covariance matrix		Competing estimators				
c	$\boldsymbol{\Sigma}$	$\widehat{\boldsymbol{\Sigma}}$	$\boldsymbol{\Sigma}_{RM}$	$\boldsymbol{\Sigma}_{S}$	$\boldsymbol{\Sigma}_{F}$	$\boldsymbol{\Sigma}_{MFFM}$
$T = 252$						
1	15.49	13.89	12.85	14.28	14.07	14.16
1.6	4.91	1.89	1.22	4.17	4.04	4.14
∞	1.21	0.40	0.38	1.11	0.98	1.09
$T = 756$						
1	14.04	13.03	12.23	13.71	13.11	13.55
1.6	3.58	1.32	1.00	3.05	1.55	3.68
∞	1.01	0.17	0.01	0.89	0.64	0.78

Additionally, simulation results show that the shrinkage method offers an estimated covariance matrix with superior estimation accuracy. This is reflected in the fact that the minimum optimal portfolio risk using this method is just a little different with respect to the theoretical risk. The mixed frequency factor model and the factor model using daily data also have a high accuracy. However, as can be seen, the factor model, the MFFM and shrinkage method offer a quite close estimation accuracy of the covariance matrix. Finally, all estimation methods overcome the sample covariance matrix, however, its performance is quite similar to the RiskMetrics.

4.3. Empirical Results

In the same way than Fan et al. (2009), data from Kenneth French was obtained is website from January 2, 1997 to December 31, 2010. We use the daily returns of 100 industrial portfolios formed on size and book to market ratio, to estimate according to four estimators, the sample covariance, RiskMetrics, factor model and

[14]The active stock set refers to the stocks with weight different from zero. This set changes as the LARS algorithm progresses. Actually, it can increase or decrease in each step depending if a particular stock is added or dropped from the active set. This is the reason why in Figure 2, some curves at the last step are at zero.

the Shrinkage, the covariance matrix of the 100 assets using the past 12 months' daily returns data.[15] These covariance matrices, calculated at the end of each month from 1997 to 2010, are then used to construct optimal portfolios under three different gross exposure constraints. The portfolios are then held for one month and rebalanced at the beginning of the next month. Different characteristics of these portfolios are presented in Table 3.

TABLE 3: Returns and Risks based on Fama French Industrial Portfolios, $p = 100$.

c	Mean	Standard deviation	Sharpe ratio	Min weight	Max weight
Sample covariance					
1	20.89	12.03	1.80	0.00	0.30
1.6	22.36	8.06	2.22	−0.05	0.28
∞	15.64	7.13	1.86	−0.11	0.25
Factor model					
1	21.49	12.09	1.82	0.00	0.29
1.6	22.56	8.26	2.24	−0.04	0.24
∞	16.73	7.40	1.90	−0.11	0.22
Shrinkage					
1	21.34	11.90	1.79	0.00	0.29
1.6	22.46	8.06	2.23	−0.05	0.23
∞	15.94	7.16	1.88	−0.11	0.22
RiskMetrics					
1	17.07	9.23	1.43	0.00	0.46
1.6	18.89	7.83	1.56	−0.07	0.44
∞	15.80	6.87	1.48	−0.13	0.42

We found that the optimal no short sale portfolio is not diversified enough. It is still a conservative portfolio that can be improved by allowing some short positions. In fact, when $c = 1$, the risk is greater than when we allowed short positions. These results hold using all covariance matrices measures. Also, we found that the portfolios selected by using the RiskMetrics have lower risk which coincides with Fan et al. (2009) results. Thus, according our simulation and empirical results, RiskMetrics give us the most overoptimistic conclusions about the risk.

Finally, the Sharpe ratio is a more interesting characterization of a security than the mean return alone. It is a measure of risk premium per unit of risk in an investment. Thus the higher the Sharpe Ratio the better. Because of the low returns showed by Riskmetrics, it has also a lower Sharpe ratio. Although differences between the other three methods are not important, the factor model has the higher Sharpe ratio. This result indicates that the return of the portfolio better compensates the investor for the risk taken.

5. Conclusions

When p is small, an estimate of the covariance matrix and its inverse can easily obtained. However, when p is closer or larger than T, the presence of

[15] We do not include the mixed frequency factor model because of the impossibility to have access to high frequency data.

many small or null eigenvalues makes the covariance matrix not positive definite any more and it can not be inverted as it becomes singular. That suggests that serious problems may arise if one naively solves the high-dimensional Markowitz problem. This paper evaluates the performance of the different methods in terms of their precision to estimate a covariance matrix in the high dimensional minimum variance optimal portfolios allocation context. Five methods were employed for the comparison: the sample covariance, RiskMetrics, factor model, shrinkage and realized covariance.

The simulated Fama-French three factor model was used to generate the returns of $p = 200$ and $p = 500$ stocks over a period of 1 and 3 years of daily and intraday data. Thus using the Monte Carlo simulation we provide evidence than the mixed frequency factor model and the factor model using daily data show a high accuracy when we have portfolios with p closer or larger than T. This is reflected in the fact that the minimum optimal portfolio risk using these methods is just a little different with respect to the theoretical risk. The superiority of the MFFM, comes from the fact that this model offers a more efficient estimation of the covariance matrix being able to deal with a very large number of stocks (Bannouh et al. 2010).

Simulation results also show that the accuracy of the covariance matrix estimated from shrinkage method is also fairly similar to the factor models with slightly superior estimation accuracy in a very high dimensional situation. Finally, as have been found in the literature all these estimation methods overcome the sample covariance matrix. However, RiskMetrics shows a low accuracy and in both studies (simulation and empirical) leads to the most overoptimistic conclusions about the risk.

Finally, we discuss the construction of portfolios that take advantage of short selling to expand investment opportunities and enhance performance beyond that available from long-only portfolios. In fact, when long only constraint is present we have an optimal portfolio with some associated risk exposure. When shorting is allowed, by contrast, a less risky optimal portfolio can be achieved.

Acknowledgements

We are grateful to the anonymous referees and the editor of the *Colombian Journal of Statistics* for their valuable comments and constructive suggestions.

[Recibido: septiembre de 2010 — Aceptado: marzo de 2011]

References

Andersen, T., Bollerslev, T., Diebold, F. & Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**(2), 579–625.

Anderson, H., Issler, J. & Vahid, F. (2006), 'Common features', *Journal of Econometrics* **132**(1), 1–5.

Bannouh, K., Martens, M., Oomen, R. & van Dijk, D. (2010), Realized mixed frequency factor models for vast dimensional covariance estimation, Discussion paper, Econometric Institute, Erasmus Rotterdam University.

Barndorff-Nielsen, O., Hansen, P., Lunde, A. & Shephard, N. (2010), 'Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading', *Journal of Econometrics* **162**(2), 149–169.

Barndorff-Nielsen, O. & Shephard, N. (2004), 'Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics', *Econometrica* **72**(3), 885–925.

Bickel, P. & Levina, E. (2008), 'Regularized estimation of large covariance matrices', *The Annals of Statistics* **36**(1), 199–227.

Bollerslev, T. (1990), 'Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH approach', *Journal of Portfolio Management* **72**(3), 498–505.

Bollerslev, T., R., E. & Wooldridge, J. (1988), 'A capital asset pricing model with time varying covariances', *Journal of Political Economy* **96**(1), 116–131.

Buhlmann, P. & van de Geer, S. (2011), *Statistics for High-Dimensional Data: Methods, Theory and Applications*, Springer Series in Statistics, Springer. Berlin.

Chan, L., Karceski, J. & Lakonishok, J. (1999), 'On portfolio optimization: Forecasting covariances and choosing the risk model', *Review of Financial Studies* **12**(5), 937–974.

Chopra, V. & Ziemba, W. (1993), 'The effect of errors in means, variance and covariances on optimal portfolio choice', *Journal of Portfolio Management* **19**(2), 6–11.

Dempster, A. (1979), 'Covariance selection', *Biometrics* **28**(1), 157–175.

Efron, B., Hastie, T., Johnstone, I. & Tibshirani, R. (2004), 'Least angle regression', *The Annals of Statistics* **32**(2), 407–499.

Engle, R. & Kroner, K. (1995), 'Multivariate simultaneous generalized ARCH', *Econometric Theory* **11**(1), 122–150.

Engle, R., Shephard, N. & Sheppard, K. (2008), Fitting vast dimensional time-varying covariance models, Discussion Paper Series 403, Department of Economics, University of Oxford.

Fama, E. & French, K. (1992), 'The cross-section of expected stock returns', *Journal of Financial Economics* **47**(2), 427–465.

Fan, J., Fan, Y. & Lv, J. (2008), 'High dimensional covariance matrix estimation using a factor model', *Journal of Econometrics* **147**(1), 186–197.

Fan, J., Zhang, J. & Yu, K. (2009), Asset allocation and risk assessment with gross exposure constraints for vast portfolios, Technical report, Department of Operations Research and Financial Engineering, Princeton University. Manuscrit.

Furrer, R. & Bengtsson, T. (2006), 'Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants', *Journal of Multivariate Analysis* **98**(2), 227–255.

Hastie, T., Tibshirani, R. & Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer. New York.

Huang, J., Liu, N., Pourahmadi, M. & Liu, L. (2006), 'Covariance matrix selection and estimation via penalized normal likelihood', *Biometrika* **93**(1), 85–98.

Jagannathan, R. & Ma, T. (2003), 'Risk reduction in large portfolios: Why imposing the wrong constraints helps', *Journal of Finance* **58**(4), 1651–1683.

Johnstone, I. (2001), 'On the distribution of the largest eigenvalue in principal components analysis', *The Annals of Statistics* **29**(2), 295–327.

Lam, C. & Yao, Q. (2010), Estimation for latent factor models for high-dimensional time series, Discussion paper, Department of Statistics, London School of Economics and Political Science.

Lam, L., Fung, L. & Yu, I. (2009), Forecasting a large dimensional covariance matrix of a portfolio of different asset classes, Discussion Paper 1, Research Department, Hong Kong Monetary Autority.

Ledoit, O. & Wolf, M. (2003), 'Improved estimation of the covariance matrix of stock returns with an application to portfolio selection', *Journal of Empirical Finance* **10**(5), 603–621.

Ledoit, O. & Wolf, M. (2004), 'Honey, I shrunk the sample covariance matrix', *Journal of Portfolio Management* **30**(4), 110–119.

Markowitz, H. (1952), 'Portfolio selecction', *Journal of Finance* **7**(1), 77–91.

Morgan, J. P. (1996), Riskmetrics, Technical report, J. P. Morgan/Reuters. New York.

Pan, J. & Yao, Q. (2008), 'Modelling multiple time series via common factors', *Boimetrika* **95**(2), 365–379.

Peña, D. & Box, G. (1987), 'Identifying a simplifying structure in time series', *Journal of American Statistical Association* **82**(399), 836–843.

Peña, D. & Poncela, P. (2006), 'Nonstationary dynamic factor analysis', *Journal of Statistics Planing and Inference* **136**, 1237–1257.

Stein, C. (1956), Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, *in* J. Neyman, ed., 'Proceedings of the Third Berkeley Symposium on Mathematical and Statistical Probability', Vol. 1, University of California, pp. 197–206. Berkeley.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the Lasso', *The Journal of Royal Statistical Society, Series B* **58**(1), 267–288.

Voev, V. (2008), *Dynamic Modelling of Large Dimensional Covariance Matrices*, High Frequency Financial Econometrics, Springer-Verlag. Berlin.

Wang, Y. & Zou, J. (2009), 'Vast volatility matrix estimation for high-frequency financial data', *The Annals of Statistics* **38**(2), 943–978.

Wu, W. & Pourahmadi, M. (2003), 'Nonparametric estimation of large covariance matrices of longitudinal data', *Biometrika* **90**(4), 831–844.

Zheng, X. & Li, Y. (2010), On the estimation of integrated covariance matrices of high dimensional diffusion processes, Discusion paper, Business Statistics and Operations Management, Hong Kong University of Science and Technology.

Appendix A.

In this appendix we present the LAR algorithm with the Lasso modification proposed by Efron et al. (2004), which is an efficient way of computing the solution to any Lasso problem, especially when $T \ll p$.

Algorithm. LARS: Least Angle Regression algorithm to calculate the entire Lasso path

1. Standardize the predictors to have mean zero and unit norm. Start with the residual $r = y - \bar{y}$, and $w_j = 0$ for $j = 1, \ldots, p-1$.
2. Find the predictor x_j most correlated with r.
3. Move w_j from 0 towards its least-squares coefficient $\langle x_j, r \rangle$, until some other competitor x_k has as much correlation with the current residual as does x_j.
4. Move w_j and w_k in the direction defined by their joint least squares coefficient of the current residual on (x_j, x_k), until some other competitor x_l has as much correlation with the current residual. If a non-zero coefficient hits zero, drop its variable from the active set of variables and recompute the current joint least squares direction.
5. Continue in this way until all p predictors have been entered. After a number of steps no more than $\min(T-1, p)$, we arrive at the full least-squares solution.

Source: Hastie, Tibshirani & Friedman (2009)

Appendix B.

Table 4: Parameters used in the simulation.

Parameters for factor loadings				Parameters for factor returns			
$\boldsymbol{\mu_\lambda}$		$\mathrm{cov}_{\boldsymbol{\lambda}}$		$\boldsymbol{\mu_f}$		$\mathrm{cov}_{\boldsymbol{f}}$	
0.7828	0.029145			0.023558	1.2507		
0.5180	0.023873	0.053951		0.012989	−0.0349	0.31564	
0.4100	0.010184	−0.006967	0.086856	0.020714	−0.2041	−0.0022526	0.19303

Source: Fan et al. (2008).