Journal of Economic Dynamics & Control 36 (2012) 1688–1699

Contents lists available at SciVerse ScienceDirect

Journal of Economic Dynamics & Control

journal homepage: www.elsevier.com/locate/jedc

A Krylov subspace approach to large portfolio optimization

Isabelle Bajeux-Besnainou [a], Wachindra Bandara [a,*], Efstathia Bura [b]

[a] Department of Finance, The George Washington University, USA
[b] Department of Statistics, The George Washington University, USA

ARTICLE INFO

Article history:
Received 1 April 2010
Accepted 13 April 2012
Available online 8 June 2012

JEL classification:
C02
G11

Keywords:
Krylov subspaces
Singular systems
Algorithm
Sample covariance matrix
Global minimum portfolio

ABSTRACT

With a large number of securities (N) and fewer observations (T), deriving the global minimum variance portfolio requires the inversion of the singular sample covariance matrix of security returns. We introduce the Break-Down Free Generalized Minimum RESidual (BFGMRES), a Krylov subspaces method, as a *fully automated* approach for deriving the minimum variance portfolio. BFGMRES is a numerical algorithm that provides solutions to singular linear systems without requiring ex-ante assumptions on the covariance structure. Moreover, it is robust to illiquidity and potentially faulty data. US and international stock data are used to demonstrate the relative robustness of BFGMRES to illiquidity when compared to the "shrinkage to market" methodology developed by Ledoit and Wolf (2003). The two methods have similar performance as assessed by the Sharpe ratios and standard deviations for filtered data. In a simulation study, we show that BFGMRES is more robust than shrinkage to market in the presence of data irregularities. Indeed, when there is an illiquid stock shrinkage to market allocates almost 100% of the portfolio weights to this stock, whereas BFGMRES does not. In further simulations, we also show that when there is no illiquidity, BFGMRES exhibits superior performance than shrinkage to market when the number of stocks is high and the sample covariance matrix is highly singular.

1. Introduction

In this article, we introduce methods that are new in finance and have been developed in applied mathematics for solving large linear systems of equations. These systems are of the form $Ax = b$, where the matrix A is either invertible or non-invertible. Among the significant advances in the field of applied mathematics during the last two decades following improvements in computer technology are the Krylov subspace methods. These are iterative methods that generate a sequence of approximate solutions in the space spanned by powers of the matrix A as follows: starting with the vector b, we compute Ab, and then multiply that vector by A to find A^2b and so on. These algorithms are among the most efficient solution approximation methods currently available in numerical linear algebra (Saad, 2003). Krylov subspace methods are summarized in a review article by Simoncini and Szyld (2007). Among these, Generalized Minimum RESidual (GMRES), introduced by Saad and Schultz (1986), has been the most prominent for solving linear systems with a square non-singular matrix.. Their low computational complexity and storage requirement have been well documented (for example Gilli and Pauletto, 1998). However, the GMRES algorithm may breakdown when the matrix is singular, as in our application to

* Correspondence to: Finance Department, The George Washington University, Funger Hall, Suite 501, 2201 G Street, NW, Washington, DC 20052, USA. Fax: +1 202 994 5014.
E-mail address: wbandara@gwmail.gwu.edu (W. Bandara).

0165-1889/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jedc.2012.04.009

portfolio management. Recent work by Reichel and Ye (2005) studies the various properties of the GMRES algorithm at the point of breakdown and proposes an extension of the algorithm, Breakdown Free Generalized Minimum RESidual (BFGMRES), which can be applied to solve large linear systems with a singular matrix.

Assuming a mean-variance optimization objective, Markowitz (1952) showed that optimal portfolios can be represented by an efficient frontier in the expected return-standard deviation space. The derivation of this efficient frontier requires two types of inputs: the expected return of each stock and the variance–covariance matrix of stock returns. Since finding reliable estimates of the expected returns is the most difficult (see Merton, 1980) between these two tasks, recent academic research (see Tola et al., 2008; DeMiguel et al., 2009 or Ledoit and Wolf, 2003) has focused on the derivation of the global minimum variance portfolio that only requires estimates of the variance–covariance matrix of the returns. Specifically, if x represents the portfolio weights in the risky securities, x^T its transpose, A the variance–covariance matrix of the returns, and $\mathbf{1}$ the vector of ones, then the global minimum variance portfolio is the solution of two linear equations:

$$Ax = 1 \tag{1}$$

$$x^T\mathbf{1} = 1 \tag{2}$$

From Eq. (1), one can infer that the portfolio weights of the global minimum variance portfolio depend on the inverse of the covariance matrix A. In effect, if A is non-singular, the solution of these two equations is $x^{*} = A^{-1}1/(1^T A^{-1}1)$. In practice, the covariance matrix is often estimated from historical data available up to a given date, called the sample covariance matrix, and optimal portfolio weights are computed from this estimate. When the number of stocks to be included in the portfolio is larger than the number of available historical returns, the sample covariance matrix is singular and Eqs. (1) and (2) do not have a unique or well-defined solution.

There have been many attempts to find an invertible estimator of the covariance matrix. Currently, the most prominent estimators of the covariance matrix are the shrinkage estimators found in Chen et al. (1999), Bengtsson and Holst (2002), Jagannathan and Ma (2000), and Ledoit and Wolf (2003, 2004). A subsequent work by Golosnoy and Okhrin (2009) attempts to mitigate the inherent flaws in the traditional estimators by minimizing the estimation risk. The main idea is to substitute the singular matrix A by an invertible matrix, called the shrinkage estimator, derive the corresponding global minimum portfolio and measure its out-of-sample performance. A shrinkage estimator is a weighted average of the sample covariance matrix and an invertible covariance matrix estimator that imposes some type of structure. As such, a shrinkage estimator is always an invertible covariance matrix estimator. The main drawback is that one has to specify the structure of the invertible covariance matrix estimator, thus introducing specification error as it is difficult to "guess" what this structure is. Moreover, the derivation of shrinkage estimators requires solving an additional minimization problem to find the shrinkage intensity or optimal weights for the two matrices.

In this article, we introduce a competing new methodology to address the issue of the singularity of high dimensional sample covariance matrices. We use the break-down free version of the GMRES (BFGMRES) Krylov subspace algorithm to find an estimator of the global minimum variance portfolio. With this approach, we avoid imposing any a priori structure on the covariance matrix. We compute the out-of-sample performance of the global minimum variance portfolio and compare it with the performance of the shrinkage to market approach proposed in Ledoit and Wolf (2003). The intuition here is that if we can implement an algorithm that circumvents the singularity of the sample covariance matrix and control the estimation errors reasonably well, we can eliminate the specification error introduced from using a specific shrinkage target. We compare these two approaches for both domestic and international stocks. While analyzing international stock data, we identify unusual behavior of the annual returns when using the shrinkage to market approach. Specifically, we identify some long periods of zero returns for the annual returns that do not exist with the BFGMRES algorithm. We conjecture that this is due to the presence of illiquid stocks or/and faulty data and use a set of filters to remove illiquid stocks in the data. After filtering, the two methods have similar performance as assessed by Sharpe ratios and standard deviations.

In order to understand better the illiquidity effect, we proceed with simulations in a controlled environment to observe the optimal weights derived by these two methods when liquidity issues exist for one security as a test case. The shrinkage to market method allocates almost 100% to the illiquid stock, while BFGMRES keeps allocation to this particular stock low. While this feature of the shrinkage to market approach is understandable from a mathematical perspective, it is undesirable from a financial one.

We also conduct simulations in an environment with no illiquid stocks and find that the BFGMRES algorithm outperforms the shrinkage to market minimum variance portfolio with respect to risk as measured by the Sharpe ratio, especially when the number of stocks increases and the matrix becomes "more" singular.

The article is organized as follows. In Section 2, we discuss the importance of the inversion of high dimensional singular covariance matrices in methods used in finance and provide a quick survey of existing methods involving shrinkage that address this issue. We introduce the Krylov subspaces iterative methods focusing on the GMRES and BFGMRES algorithms in Section 3. Section 4 describes the data and provides the comparison between the shrinkage to market and the Krylov subspace approaches for the out-of-sample performance of the global minimum variance portfolios for both US and international stocks for both unfiltered and filtered data. Section 5 presents simulations in a controlled environment to study the liquidity issue. We conclude in Section 6.

2. Financial background and the shrinkage methodology

2.1. Financial background

Portfolio managers derive portfolios that optimally balance risk and returns. The most common measure of risk is the standard deviation. Markowitz (1952) showed that these optimal portfolios can be represented by an efficient frontier in the Expected return/Standard Deviation space and that this efficient frontier only depends on expected returns, standard deviations and correlation coefficients of individual securities' returns. Because estimating the expected returns is very problematic and involves approaches different from pure statistical analyses, in particular financial forecasting, we focus on the derivation of the global minimum variance portfolio which only requires estimates of the covariance matrix A of the securities' returns. The global minimum variance portfolio is the solution of the two linear Eqs. (1) and (2). Hence, when A is invertible, the portfolio weights x^* of the global minimum variance portfolio depend on A's inverse and are given by $x^*=A^{-1}1/(1^TA^{-1}1)$.

In practice, the covariance matrix is often estimated by the sample covariance matrix derived from historical data and optimal portfolio weights are computed from it. Moreover, portfolio managers try to include a large number of stocks in order to increase the diversification effect.[1]

One issue arises from this methodology: when the number of stocks included in the optimal portfolio is larger than the number of available historical returns, the sample covariance matrix is singular and Eqs. (1) and (2) do not have a unique or well-defined solution. As pointed out in Ledoit and Wolf (2003), "in typical applications, there can be over a thousand stocks to choose from, but rarely more than ten years of monthly data, i.e., $N=1000$ and $T=120$." Finding optimal approximations to the inverse of high dimensional singular sample covariance matrices is therefore crucial when deriving optimal portfolio allocations. The Ledoit and Wolf (2003, 2004) approach is to impose more structure on the estimator of the covariance matrix by using the shrinkage technique which is widely used in Statistics and dating back to Stein (1956).

2.2. The Ledoit–Wolf shrinkage to market methodology

Ledoit and Wolf (2003) propose replacing the sample covariance matrix as an estimate of the true covariance matrix Σ by a weighted average of the sample covariance matrix and a single-index regression model based covariance matrix estimator (Sharpe 1963). The latter is a low-variance target estimator, $\hat{\Sigma}_{\text{target}}$, which is a $N\times N$ invertible, positive-definite and symmetric matrix. The non-invertible matrix $\hat{\Sigma}$ is replaced with the convex linear combination:

$$\hat{\Sigma}_{LW}\sum=(1-\lambda)\hat{\Sigma}+\lambda\hat{\Sigma}_{\text{target}} \tag{3}$$

where $\lambda\in[0,1]$ is the shrinkage intensity that controls how much weight is given to the target invertible matrix: for $\lambda=1$ the shrinkage estimate equals the shrinkage target $\hat{\Sigma}_{\text{target}}$ for $\lambda=0$, the unrestricted sample covariance matrix estimate $\hat{\Sigma}$ is recovered. That is, the shrinkage intensity λ controls how much structure is imposed: the heavier the weight, the stronger the single-index model-based structure. The key advantage of this construction is that the regularized estimate $\hat{\Sigma}_{\text{shrink}}$ is always invertible and outperforms both individual estimators $\hat{\Sigma}$ and $\hat{\Sigma}_{\text{target}}$ in terms of accuracy.

The shrinkage intensity is chosen by minimizing a risk function, $R(\lambda)$, typically the mean squared error (MSE). Ledoit and Wolf (2003) use the squared Frobenius norm as a measure of distance between matrices:

$$R(\lambda)=E\|\lambda\hat{\Sigma}_{\text{target}}+(1-\lambda)\hat{\Sigma}-\Sigma\|^2=E\left(\sum_{i=1}^{p}\sum_{j=1}^{p}(\lambda s_{ij}^{\text{target}}+(1-\lambda)s_{ij}-\sigma_{ij})^2\right) \tag{4}$$

where $\Sigma=(\sigma_{ij})$, $\hat{\Sigma}_{\text{target}}=(s_{ij}^{\text{target}})$ and $\hat{\Sigma}=(s_{ij})$, and calculate the analytic solution to this minimization problem. Their method for choosing an optimal value of λ that minimizes $R(\lambda)$ is called Shrinkage to Market (SM). Among its attractive features are that no distributional assumptions are required for the data to satisfy nor computationally expensive procedures such as Markov Chain Monte Carlo (MCMC), bootstrap, or cross-validation are needed.

In more detail, Sharpe's (1963) single-index model assumes that stock returns are generated by the model:

$$x_{it}=\alpha_i+\beta_i x_{0t}+\epsilon_{it} \tag{5}$$

where the residuals $\in_{it}$ are uncorrelated to market returns x_{0t} and to one another. Also, within stocks the variance is constant, that is, $Var(\in_{it})=\delta_{it}$. The covariance matrix implied by this model is

$$\Phi=\sigma_{00}^2\beta\beta^T+\Delta \tag{6}$$

where σ_{00}^2 is the variance of market returns, β is the vector of slopes and Δ is the diagonal matrix with the residual variances δ_{ii} along its main diagonal. This model can be estimated by running a regression of stock i's returns on the market.

[1] Portfolio diversification means that given a certain level of expected returns you can reduce the standard deviation by increasing the number of securities in a portfolio and optimally allocating the portfolio among them.

If we let b_i denote the slope estimate and d_{it} the residual variance estimate, then the single-index model yields the following estimator for the covariance matrix of stock returns, $\Phi=(\phi_{ij})$,

$$F=s_{00}^2 bb^T+D \tag{7}$$

where s_{00}^2 is the sample variance of the market returns, b is the vector of slope estimates and D is the diagonal matrix with the residual variance estimates d_{ii} on the diagonal. Let f_{ij} denote the (i,j)th entry of F. The optimal shrinkage density is given by

$$\lambda^*=\max\left\{0,\min\left\{\frac{\hat{\kappa}}{T},1\right\}\right\} \tag{8}$$

where

$$\hat{\kappa}=\frac{\hat{\pi}-\hat{\rho}}{\hat{\gamma}}$$

$$\text{with}\quad \hat{\gamma}=\sum_{i=1}^{N}\sum_{j=1}^{N}(f_{ij}-s_{ij})^2,\quad \hat{\pi}=\sum_{i=1}^{N}\sum_{i=1}^{N}\hat{\pi}_{ij},$$

$$\hat{\pi}_{ij}=\frac{1}{T}\sum_{t=1}^{T}[(x_{it}-\overline{x}_{i.})(x_{jt}-\overline{x}_{j.})-s_{ij}]^2,$$

$$\hat{\rho}=\sum_{i=1}^{N}\hat{\pi}_{ii}+\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\frac{\overline{r}}{2}\left(\sqrt{\frac{s_{jj}}{s_{ii}}}\hat{\theta}_{ii,ij}+\sqrt{\frac{s_{ii}}{s_{jj}}}\hat{\theta}_{jj,ij}\right)\quad\text{and}$$

$$\hat{\theta}_{ii,ij}=\frac{1}{T}\sum_{t=1}^{T}\left[(x_{it}-\overline{x}_{i.})^2-s_{ii}\right]\left[(x_{it}-\overline{x}_{i.})(x_{jt}-\overline{x}_{j.})-s_{ij}\right]$$

The optimal shrinkage density is truncated at 0 and 1 in order to avoid over-shrinkage (λ higher than 1) or negative shrinkage (negative λ) as in finite samples $\hat{\kappa}/T$ may be higher than 1 or negative. It is also worth noting that the optimal shrinkage density λ^* in (8) is valid no matter what values the sample size T or the number of stocks N take. In particular, T can be substantially smaller than N, which is the case considered in this article.

3. Krylov subspaces, GMRES and BFGMRES

This section describes the general methodology proposed in this article to estimate the weights of the minimum variance portfolio in the case of a high dimensional singular covariance matrix of returns. This approach, contrary to shrinkage, does not require any particular ad-hoc structure on the estimator of the covariance matrix.

3.1. Krylov subspaces

Modern iterative methods for finding one (or a few) eigenvalues of large sparse matrices with only a few nonzero entries or for solving large systems of linear equations, focus on avoiding computationally expensive matrix operations, but rather multiply vectors by the matrix and work with the resulting vectors. Iterative methods generate a sequence of approximate solutions, where the main computational effort for constructing the kth approximant from the previous one consists of one or a few matrix–vector multiplications. This is why large and sparse systems are usually solved iteratively: Starting with a vector, b, one computes Ab, then multiplies that vector by A to find A^2b and so on.

The *power* method underlying Krylov subspace methods can find the largest eigenvalue of a matrix A. If $A=Udiag(\theta_i)U^T$ is the spectral decomposition of the square matrix A, then $A^k=Udiag(\theta_i^k)U^T$. As k gets large, the diagonal matrix of eigenvalues $diag(\theta_1^k)$ will be dominated by the largest eigenvalue θ_1^k. Also, $\|x_{k+1}\|/\|x_k\|$ will converge to the largest eigenvalue and $x_k/\|x_k\|$ to the associated eigenvector. If the largest eigenvalue has multiplicity greater than one, then $x_k/\|x_k\|$ will converge to a vector in the subspace spanned by the eigenvectors associated with the largest eigenvalue. Once the first eigenvalue and corresponding eigenvector have been obtained, one can successively restrict the algorithm to the null space of the known eigenvectors to get the other eigenvector/values.

In practice, if x_0 is a random vector and $x_{k+1}=Ax_k$, then $x_k/\|x_k\|$ approaches the eigenvector corresponding to the largest eigenvalue of A when k increases. This simple algorithm is applied iteratively but is typically not very accurate for computing many of the eigenvectors because any round-off error tends to introduce slight components of the more significant eigenvectors back into the computation. Pure power methods can also converge slowly, even for the first eigenvector. In the context of solving large linear systems,[2]

$$Ax=b,\quad A\in R^{N\times N},\quad x,b\in R^N \tag{9}$$

the power methodology described above starts at the seed vector b to form the so-called Krylov matrix: $[b,Ab,A^2b,\ldots,A^{k-1}b]$

[2] Here and henceforth R^N denotes the N-dimensional real number space.

The order-k Krylov subspace generated by A and the random vector b is the linear subspace spanned by the first k powers of A applied to b (starting from $A^0=I$):

$$K_\kappa = span(b, Ab, \ldots, A^{k-1}b) \tag{10}$$

The best known Krylov subspace methods are the Arnoldi (1951), Lanczos (Golub and Van Loan(1996)), generalized minimum residual error (GMRES) and stabilized biconjugate gradient (BiCGSTAB) methods. All these methods can break down when the matrix is singular. We use the break down-free variant of the GMRES algorithm, BFGMRES (Reichel and Ye, 2005) to address the problem of minimum variance portfolio.

The generalized minimal residual error algorithm henceforth referred to as GMRES, for solving large linear systems as in (9), is widely used among iterative methods. GMRES approximates the solution by the vector in a Krylov subspace with minimal residual. The Arnoldi iteration is used to find this vector. The method is well understood when A is non-singular. In this case, there are many competing numerical methods to solve such systems. We are interested in the application of the GMRES algorithm to covariance matrices that are large and rank deficient. In that case, the GMRES algorithm may break down and fail to provide a solution.

To set the stage for presenting of the breakdown-free version of the GMRES algorithm, we first present the basic GMRES algorithm and the intuition behind it. We discuss when and why this algorithm fails and present a modification, the breakdown-free GMRES (BFGMRES), which tackles breakdowns and is used to obtain the solution of the linear system in (9) when A is a singular matrix.

3.2. The basic GMRES algorithm

This is an iterative algorithm where each iteration brings the approximation, x_k, closer to the solution of the system. The algorithm is run until a predetermined threshold, i.e. a predetermined distance for $\|Ax_k-b\|$, is reached at which point it terminates. Suppose this distance is reached at the kth step and that the associated solution is x_k. Then, x_k is declared as the solution of the optimization problem given by

$$min_{x_k \in K_k(A,b)} \|Ax_k - b\| \tag{11}$$

where $K_k(A,b)$ is the kth Krylov subspace generated by A and b defined in (10). An initial value for the solution, x_0, is needed to start the algorithm. Without loss of generality, we set $x_0=0$[3].

At the heart of the GMRES algorithm is the Arnoldi process. The Arnoldi process can be thought of as the Gram–Schmidt orthogonalization process (see, for example, Golub and Van Loan, 1996, p. 230–231) tailored to Krylov subspaces. The algorithm begins by taking a vector $v_1=b/b$ as the first Krylov subspace K_1, i.e. $K_1=\text{span}(b)$, and it then iteratively generates the orthonormal basis for each subsequent Krylov subspace generated by A and b. In general, the orthonormal basis for $K_{k+1}(A,b)$ is generated from $K_k(A,b)$ by orthogonalizing the vector $Av_k(\in K_{k+1}(A,b))$ against the previous subspace $K_k(A,b)$. This can be done in a step analogous to that in the Gram–Schmidt process by taking $\tilde{v}_{k+1} = Av_k-(h_{1,k}v_1+\cdots+h_{k,k}v_k)$ where $h_{i,j}=v_i^TAv_j$. The orthonormal basis vector is given by $v_{k+1}=\tilde{v}_{k+1}/\|\tilde{v}_{k+1}\|$ and we define the matrix of orthonormal basis vectors of $K_k(A,b)$ as

$$V_k = [v_1, v_2, \ldots, v_k] \tag{12}$$

Additionally, the Arnoldi process, at each iteration, creates a $k \times k$ upper Hessenberg[4] matrix $V_k^TAV_k$. From this we can write what is called the Arnoldi decomposition given by

$$AV_k = V_{k+1}H_k \tag{13}$$

where H_k is a matrix of size $(k+1)\times k$ given by $H_k = \begin{pmatrix} & V_k^TAV_k & \\ 0 & 0\ldots & h_{k+1,k} \end{pmatrix}$.

We have established a relationship among $K_k(A,b)$, AV_k and $V_{k+1}H_k$. For example, suppose the algorithm terminates at the kth step and hence we have $x\in K_k(A,b)$. Then, there exists y such that $x=V_ky$, i.e. x is a linear combination of the columns of V_k.

Referring back to the minimization problem we began with, we can write $Ax=AV_ky=V_{k+1}H_ky$ by the Arnoldi decomposition. Similarly, since the first basis vector is $v_1=b/\|b\|$ we can write $b=\beta v_1=\beta V_{k+1}e_1$ where e_1 is the first column of the identity matrix and $\beta=\|b\|$.

With these relationships, the minimization problem can be written equivalently as follows:

$$min_{x\in K_k(A,b)} \|Ax_k - b\| = min_{y\in R^k} \|\beta e_1 - H_k y\| \tag{14}$$

[3] The algorithm was run with random vectors for x_0 and there was no change in the solution. Although Smoch (1999) attributes the breakdown of the GMRES algorithm partly to the selection of the initial value, the breakdown-free version of the algorithm seemed to not be sensitive to the selection of the initial value. Note that although this initial condition does not satisfy Eq. (2) and does not correspond to a vector defining a portfolio, the results of the simulation would be unchanged if we were starting with a portfolio with equal weights ($=1/N$) in all securities.

[4] An upper Hessenberg matrix is a matrix with zero entries below the first subdiagonal (the diagonal entries to the left and below the main diagonal).

This is a simpler and more familiar form of minimization that can be solved using a method such as ordinary least squares. This y by its definition is the vector of coefficients of the solution x_k in the kth Krylov subspace. Hence, the kth iterate of the solution in the kth Krylov subspace $K_k(A,b)$ can be written as $x_k = V_k y$.

3.3. Breakdown of the GMRES algorithm

If the matrix A is non-singular the solution of the system is guaranteed under GMRES. When a solution exists, GMRES's finding this solution is equivalent to the Krylov subspace not being augmented further, i.e. if the solution exists in $K_k(A,b)$ we have $v_{k+1} = 0$. This means that the last row of the matrix H_k is zero. Let H_k^- be H_k without its last row. Then, by the Arnoldi decomposition, $AV_k = V_k H_k^-$ since the matrix of basis vectors V is not augmented any further either.

We can also establish the following results: the column vectors in V_k span an invariant subspace of A. The eigenvalues of H_k^- are equal to the eigenvalues of A.

The non-singular A has no zero eigenvalues. Hence, H_k^- has no zero eigenvalues and it is non-singular. Using the Arnoldi decomposition, we have reduced the least squares problem to the solution of a non-singular linear system.

Here we see that the ability to solve the system depends on the fact that A is non-singular. However, when A is singular, which is the case we are interested in, we cannot obtain that H_k^- is non-singular. In this case the GMRES algorithm will not succeed and will break down. Reichel and Ye (2005) solve the problem of the Arnoldi process breaking down by creating a more general form of the Arnoldi process, which they use in their Breakdown-Free GMRES algorithm.

3.4. Breakdown-Free GMRES algorithm

We first introduce notation used by Reichel and Ye (2005). The kth iteration of the Arnoldi decomposition of the matrix A can be rewritten as $AV_k = V_k H_k + f_k e_k^T$ where $H_k \in R^{k \times k}$ is an upper Hessenberg matrix, $V_k \in R^{N \times k}$, $V_k e_1 = b$, $V_k^T V_k = I_k$, $V_k^T f_k = 0$, I_k denotes the identity matrix of order k, and e_k is the kth axis vector. When $f_k \neq 0$ it is convenient to define the matrices

$$V_{k+1} = \left[V_k, \frac{f_k}{\|f_k\|} \right] \in R^{N \times (k+1)} \quad \text{and} \quad \hat{H}_k = \begin{bmatrix} H_k \\ \|f_k\| e_k^T \end{bmatrix} \in R^{(k+1) \times k}.$$

We can then write the decomposition as $AV_k = V_{k+1} \hat{H}_k$ which is equivalent to the form of the Arnoldi decomposition presented in Section 3.2.

This procedure can be explained intuitively: suppose a breakdown occurs at step M of the algorithm. This means that the subspace $K_M(A,b)$ does not contain a solution of the system, i.e. the column vector $v_M \in V_M$ is not required to generate the solution.

However it can be shown that any solution of the system is in $K_{M-1}(A,b) + N(A^p)$ where $N(A^p)$ is the null space of A^p and p is the index of the matrix A, i.e. the largest Jordan block of A associated with the zero eigenvalue is of order p (Reichel and Ye, 2005, *Theorem 2.1*).

Thus, whenever there is a breakdown at step M of the process, the subspace $K_{M-1}(A,b)$ has to be extended to capture the component of the solution in $N(A^p)$ which is an eigenvector of A^p corresponding to the zero eigenvalue. This eigenvector can be approximated with a Krylov subspace generated by a new vector $\hat{v}$, which we can select randomly. We select a random vector $\hat{v}$ of the same size as v_M, orthogonalize this vector against the column vectors in V_M and replace v_M with this vector $\hat{v}$.

Next, we create a matrix $U_1 = [v_M]$ with the old v_M that is replaced by $\hat{v}$. We explain the reason below. From $AV_M = V_{M+1} \hat{H}_M$ we have: $Av_{M-1} = h_{1,M-1} v_1 + h_{2,M-1} v_2 + \ldots + h_{M-1,M-1} v_{M-1} + h_{M,M-1} v_M$ and this can be written as $Av_{M-1} = h_{1,M-1} v_1 + h_{2,M-1} v_2 + \ldots + h_{M-1,M-1} v_{M-1} + h_{M,M-1} u_1$ where the notation u_1 means that it is the first column of the matrix U_1. We can now proceed with the algorithm where for subsequent iterations $k = M, M+1, M+2, \ldots$ we require the iterates of the matrix V, V_k, satisfy the condition that the columns of V_k are orthogonal to the columns of V_{k-1} as well as orthonormal to the columns of U_1. This procedure can be written as follows: $Av_k - V_k(V_k^T A v_k) - U_1(U_1^T A v_k) = f_k$ and then we can write

$$V_{k+1} = \left[V_k, \frac{f_k}{\|f_k\|} \right] \tag{15}$$

and

$$\hat{H}_k = \begin{bmatrix} H_k \\ \|f_k\| e_k^T \end{bmatrix} \tag{16}$$

Next we test to see whether $\hat{H}_k$ is full rank. If $\hat{H}_k$ is not full rank, its condition number will be very large. To detect this, we set a predefined tolerance level for the condition number of $\hat{H}_k$.[5] If the condition number of $\hat{H}_k$ is greater than this tolerance level, we continue the generalized procedure. When the condition number of $\hat{H}_k$ falls below this tolerance level, we revert to the standard Arnoldi process. In this manner we have found a way to proceed with the algorithm even when a breakdown has occurred.

[5] The condition number of a matrix is the ratio of the largest and the smallest singular values of the matrix.

Suppose another breakdown occurs at step $S > M$. We proceed in a similar fashion. The vector v_s is appended to matrix U_1 giving a new matrix $U_2 = \begin{bmatrix} U_1 & v_S \end{bmatrix}$. A new random vector, $\tilde{v}$, is generated and orthogonalized against the columns of V_{S-1} and U_2. The last column of V_s is replaced by this $\tilde{v}$ and the steps are followed identically to what was done after the first breakdown. In this manner the BFGMRES algorithm averts the failure of the standard Arnoldi decomposition and finds a way to proceed. The algorithm is summarized in the Appendix.

4. Data and results

4.1. Data

We use monthly return data for the G-7 countries. We get equity returns for the US from the Center for Research in Securities prices (CRSP) monthly database[6] for the period between January 1926 and December 2009. We obtain the monthly stock returns for UK for the period January 1965–December 2009 and for Canada, France, Germany, Italy and Japan for the period from January 1973–December 2009. The international data are obtained from Datastream[7]. We focus on the period between January 1972 and December 2008. When generating covariance matrices, we consider the calendar year January 1st through December 31st.

To compare the shrinkage to market (Ledoit and Wolf, 2003) and BFGMRES methods, we use the monthly returns to generate the sample covariance matrix using data from January of year $t-5$ until December of year $t-1$.[8] This covariance matrix is used to generate the weights for the global minimum variance portfolio. We build this portfolio in January of year t and hold it until December of year t. In this case the in-sample period is from January 1st of year $t-5$ until December 31st of year $t-1$. The out-of-sample period is one year, from January 1st of year t until December 31st of the same year. To summarize, we build the minimum variance portfolio using the 5-year in-sample period and we compute its returns over the twelve months of the 1-year out-of-sample period. Using these twelve monthly returns, we compute the Sharpe ratio for the out-of-sample period.

For each 6 years in-sample-out-of-sample period we consider only the equities that were available for the entire 6 years period. Equities that entered the 6 years window after the beginning of the period or that dropped out before the end of the period are omitted from the calculations. Equities with any missing returns are also removed. The number of available equities ranges from 301 to 3608 for the US; 577 to 1757 for the UK; 106 to 2918 for Canada; 69 to 1293 for France; 119 to 1830 for Germany; 64 to 473 for Italy and 738 to 4069 for Japan. Due to the development of equity markets, the number of firms increases for each country over time. We eliminate the first four observations for Canada and the first observation for Italy because the number of firms is smaller than the number of observations giving a non-singular covariance matrix. All reported results are for cases where the covariance matrix is singular.

4.2. Empirical results

In Ledoit and Wolf (2003, 2004) different shrinkage methods are implemented and compared, and they find that "shrinkage to market" outperforms the other shrinkage methods in their data set. We refer to Ledoit and Wolf's method as the Shrinkage to Market (SM) method in the rest of the article and we only use the SM method as a basis of comparison.

We now turn to annual returns for the comparison of these two methods. In Fig. 1, we plot the annualized returns for the seven countries used in our study.

Looking at these annual returns, we identify an abnormality for all countries except the US. Indeed, for the Shrinkage to Market method, we always observe a period during which the portfolio has zero returns. These zero portfolio returns might come from heavy allocations in stocks with zero returns. In the data we use, we observe illiquidity for several stocks, which translates to zero observed returns. In order to understand the effect of illiquid stocks better and to investigate whether the BFGMRES method will outperform the SM method in the absence of illiquidity we conduct a simulation study in the next section.

4.3. Empirical results without illiquid stocks

As there is not a precise definition in the financial literature about what degree of trade inactivity constitutes stock illiquidity we impose the following criteria in selecting stocks included in the portfolio: (a) stocks with a standard deviation greater than 0.01 for the 60 month in-sample period and at least 0.02 for the out-of-sample period, (b) stocks that have returns of less than -0.9 in one month are excluded from the sample as these are stocks that have lost 90% of

[6] CRSP stores data on a security level rather than a company level. We only use a common stock for this analysis, which corresponds to equities with a share code of 10 or 11.

[7] Datastream stores data based on a return index. A stock is given the value 100 when it first appears on Datastream and stock returns are reflected in changes in the index.

[8] Note that Ledoit and Wolf (2003) use a 10 year in-sample period (from $t-10$ to t) while we are using 5 years. This choice was made to retain a sufficient amount of observations for countries where the sample period is shorter. In unreported results we applied the methods using a 10-year in-sample period for US stocks and found the results to be consistent.

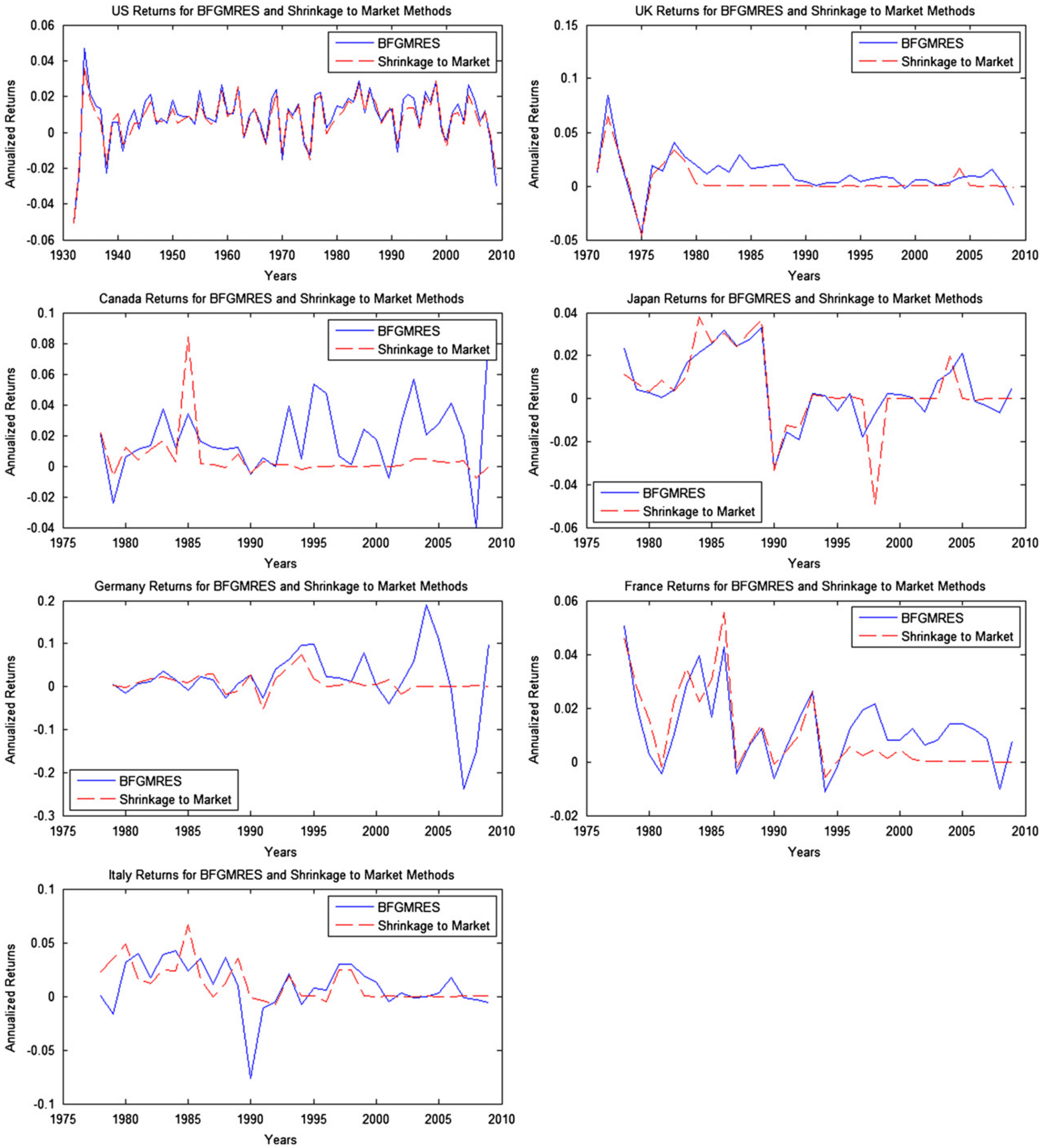

Fig. 1. Plots of the annualized returns for the G7 nations — the US, UK, Canada, Japan, Germany, France and Italy — using the BFGMRES and the Shrinkage to Market method of Ledoit and Wolf.

their market capitalization and are thus unlikely to have irregular behavior, (c) overall standard deviation for the entire 72-month period is greater than 0.05 and (d) there be at least 10 non-zero observations in the in-sample period and at least 4 non-zero observations in the out-of-sample period.

Now the number of available equities ranges from 300 to 3435 for the US; 557 to 1429 for the UK; 100 to 2412 for Canada; 67 to 845 for France; 110 to 1343 for Germany; 64 to 330 for Italy and 708 to 3215 for Japan.

The most standard measure of risk used in the finance literature is the standard deviation. However, the two global minimum variance portfolios generated by the two methods have very different return profiles. Therefore, we also use the

Table 1
Average return, standard deviation and Sharpe ratio for data from the G7 countries when filters (a)–(d) above are imposed.

	Average return		Std. deviation		Sharpe ratio	
	BFGMRES (%)	SM (%)	BFGMRES (%)	SM (%)	BFGMRES	SM
US	1.03	0.93	3.48	3.07	0.204	0.200
UK	1.47	1.48	3.47	4.22	0.287	0.238
Canada	1.87	1.49	4.93	6.21	0.292	0.171
Italy	1.42	1.48	8.88	4.65	0.092	0.186
Japan	0.57	0.68	3.79	3.98	0.077	0.100
Germany	1.17	1.46	4.07	3.70	0.210	0.311
France	1.50	1.50	4.39	4.65	0.242	0.229

Table 2
Average return, standard deviation and Sharpe ratio for data from the G7 countries when stocks that are not traded for at least two of the 60 in-sample months are excluded.

	Average return		Std. deviation		Sharpe ratio	
	BFGMRES (%)	SM (%)	BFGMRES (%)	SM (%)	BFGMRES	SM
US	0.84	0.76	3.46	2.99	0.150	0.148
UK	1.20	0.73	3.51	2.77	0.209	0.093
Canada	1.08	1.15	3.91	2.78	0.167	0.260
Italy	1.31	1.27	9.90	4.45	0.071	0.149
Japan	0.55	0.74	3.97	4.11	0.068	0.112
Germany	1.06	0.94	3.89	3.74	0.192	0.170
France	1.39	1.43	4.39	3.41	0.217	0.291

Sharpe ratio in addition to the standard deviation to give a risk-adjusted measure of the portfolio performance. Table 1 gives the results for the expected returns, standard deviation and the Sharpe ratio after filtering the data as described above.

Table 1 shows that BFGMRES has higher Sharpe ratios and smaller standard deviations than SM for four out of the seven countries.

To test whether the two strategies' Sharpe ratios and standard deviations are statistically significantly different we used the bootstrap based tests proposed by Ledoit and Wolf (2008, 2011). The tests were applied to the returns of the optimal portfolios produced by the two methods over the out-of-sample period. With the exception of Italy and Germany where the Sharpe ratios were found to be statistically significantly different at the 5% level (p-values < 0.05), for the other five countries there is no statistically significant difference and both methods perform roughly the same with respect to Sharpe ratios. With respect to standard deviation, the two methods have statistically significantly different performance only for Italy and the US (p-values < 0.05).

Next we consider the performance of these methods using a stricter filtering criterion. A referee pointed out that fund managers and finance practitioners are likely to exclude stocks from their portfolios that have been traded only in very few of the 60 months in the in-sample period. In Table 2, we report the results of our calculations when requiring that a stock is excluded if it has zero returns for two or more months in the in-sample period.

The number of available equities ranges from 203 to 3103 for the US; 266 to 1317 for the UK; 92 to 871 for Canada; 65 to 494 for France; 108 to 635 for Germany; 61 to 250 for Italy and 635 to 2357 for Japan. The stricter filtering criterion consistently reduces the number of available stocks in the in-sample period. Now BFGMRES has a higher Sharpe ratio for three out of the seven countries and higher expected returns for four out of seven, although SM outperforms BFGMRES in terms of standard deviation in six out of the seven countries.

In terms of assessing the statistical significance of the two measure comparison under this scenario, the bootstrap based tests (Ledoit and Wolf 2008, 2011) found all the Sharpe ratios to be statistically significantly different at the 5% level (p-values < 0.05) except for Germany, the US, and marginally the UK. With respect to standard deviation, the two methods have statistically significantly different performance for all countries except for Germany and Japan.

5. Simulations

These simulations are intended to examine the behavior of BFGMRES and SM in the presence of illiquidity. One illiquid stock is introduced in the simulation. We study how the BFGMRES and SM methods treat illiquidity by examining the weight allocation to this stock by the two methods.

We build a "controlled environment," where the number of periods T is 60 and the number of stocks N ranges from 75 to 500. The stock returns are generated using a Brownian motion with randomly selected means and volatilities for each

Table 3
Weights on the illiquid stock for the GMRES and Shrinkage to Market methods when the number of stocks varies from 75 to 500.

N	Weight on most illiquid stock	
	BFGMRES (%)	Shrinkage to Market (%)
75	11.47	99.99
100	5.14	99.98
125	3.33	99.98
150	2.49	99.97
175	1.93	99.97
200	1.59	99.96
225	1.36	99.97
250	1.17	99.96
275	1.07	99.97
300	0.95	99.96
325	0.86	99.97
350	0.80	99.98
375	0.73	99.97
400	0.67	99.96
425	0.62	99.96
450	0.58	99.96
475	0.56	99.96
500	0.52	99.96

stock[9]. We set one stock to be illiquid: all but one of its returns is set to 0. Table 3 reports the weight for this illiquid stock in the resulting MV portfolio.

Table 3 shows that in the presence of an illiquid stock the shrinkage to market method allocates almost 100% of the weight to the illiquid stock. Although this feature is justifiable from a mathematical point of view (this stock has near zero variance since its return is almost always zero), it is problematic in practice.

Indeed, actual data very often include stocks that are infrequently traded, especially at shorter trading frequencies, and appear to have zero returns because of their temporary illiquidity. Since precise criteria are difficult to impose when "cleaning" the data to eliminate illiquid stocks and because such criteria, even when they are imposed in the empirical finance literature tend to be ad hoc in nature, we view this feature as undesirable in this context. In contrast, the BFGMRES method does not allocate high weights to these illiquid stocks. Moreover, when the number of stocks included in the portfolio is high, the wealth allocation to illiquid stocks decreases.

We now analyze the performance of BFGMRES versus shrinkage to market with respect to the global minimum variance portfolio. For a fixed N, we assume that there is a rolling window of fifty periods each with 72 observations. The first 60 observations are assumed to be the in-sample period and the last 12 observations constitute the out-of-sample period. We assume fixed drift and volatility for the stocks within each 50 period block. We generate the global return for the global minimum variance portfolio for each of the 50 periods using the in-sample period to generate the weights and the out-of-sample period for the returns. We use the time series of returns generated in this manner to compute the Sharpe ratio for the global minimum variance portfolio as

$$\text{Sharpe Ratio} = \frac{\text{Mean Return over the 50 periods}}{\text{Standard Deviation of Returns over the 50 periods}}$$

for both BFGMRES and SM. In the simulation we assume that the risk-free rate is zero. Let the Sharpe ratio for BFGMRES be denoted by $SR_{BFGMRES}$ and the Sharpe ratio for SM by SR_{SM}. We repeat this process 100 times — each time with a different, randomly generated set of drifts and volatilities for the N stocks — and count the number of times $SR_{BFGMRES} > SR_{SM}$. For instance, Table 4 shows that when $N=75$ and there was no illiquidity issue, $SR_{BFGMRES} > SR_{SM}$ 11 times. Then, we increment N by 25 and repeat the process. Table 4 shows the results of this simulation for different values of N.

The No Illiquidity column shows the results for when all the stocks are liquid. The With Illiquidity column is for the case where one of the stocks is illiquid (by setting a majority of observations to 0 in the in-sample period)[10].

From Table 4 we see that the BFGMRES performance is superior to the shrinkage to market performance as the number of stocks N increases (the matrix becomes more singular) as long as there are no illiquidity issues. If there is an illiquid stock, the shrinkage to market method performs better, but this superior performance is the result of an undesirable property.

[9] This means that for each stock we randomly generate a mean-volatility pair (μ,σ) and simulate a path assuming that each stock follows a geometric Brownian motion.

[10] Here we experiment with different levels of illiquidity—from a single non-zero return for the illiquid stock to 30. In all cases, nearly 100% of the weight was allocated to this "most illiquid" stock and the result in Table 4 holds.

Table 4
Number of times (out of 100 simulations) the BFGMRES's Sharpe ratio is higher than the Shrinkage to Market's Sharpe ratio when all stocks are liquid and when one of the stocks is illiquid.

N	No illiquidity	With illiquidity
75	11	0
100	37	0
125	69	0
150	83	0
175	87	0
200	90	0
225	93	0
250	96	0
275	96	0
300	97	0
325	97	0
350	98	0
375	96	0
400	96	0
425	96	0
450	99	0
475	100	0
500	100	0

Liquidity is likely to be a concern in large portfolios of traded assets. If such illiquid stocks are present in the sample period used to calculate portfolio allocation weights, our simulation shows that the shrinkage to market method would consider this as the least risky stock and allocate almost 100% of the wealth to this illiquid stock. However, in this case, the return of the portfolio is not obtainable since allocating wealth to that stock would immediately render it liquid and its resulting return would most likely be different from zero. In practice, implementation of the shrinkage to market methodology to derive the minimum variance portfolio, in contrast to the BFGMRES methodology, requires up-front cleaning of the data and the removal of all illiquid stocks.

6. Concluding remarks

In this article we introduce a new methodology to tackle singular covariance matrices in mean-variance portfolio optimization. Following the approach of several authors (e.g. DeMiguel et al., 2009; Ledoit and Wolf, 2003, 2004) we consider the global minimum variance portfolio in order to avoid the estimation of expected returns. Contrary to Ledoit and Wolf's (2003, 2004) shrinkage based method, the BFGMRES method does not require any structural assumptions on the covariance matrix of historical returns. As acknowledged in Ledoit and Wolf (2003), choosing a priori the right shrinkage target is "an art." Our approach does not require any interventional choice on the part of the user.

We compare the minimum variance portfolios derived from the two methods using empirical US and international data. The shrinkage to market portfolios have long periods of zero returns in the case of international stocks, when raw, unfiltered data are used.

A simulation allows us to understand that this behavior comes from high wealth allocation to illiquid stocks, which is an undesirable feature. Further simulations show that when we eliminate illiquidity and use the Sharpe ratio as a risk measure, the BFGMRES method outperforms the shrinkage to market method as the number of stocks increases and the singularity of the covariance matrix becomes more pronounced.

In the past 20 years automated trading systems have become increasingly prevalent in financial markets. By some estimates, automated trading systems account for 70% of daily trading volume. With this gradual removal of human intervention in trading activity, it is important that we do not lose sight of financial sense in favor of mathematical accuracy. The advantage of the BFGMRES method is that it is robust to the presence of illiquid stocks and/or faulty data and does not require any choice on the part of the analyst while at the same time has similar and sometimes better performance compared to the competing shrinkage based approaches.

Acknowledgment

The authors thank three anonymous referees and the associate editor, Carl Chiarella, for their comments and suggestions that helped us improve the paper. We also thank Prof. Michael Wolf for sharing his R code for robust hypothesis testing with variances and Sharpe ratios with us.

Appendix

BFGMRES algorithm

Input A, b
Iteration $k=0$:
Initialize: Set $x_0=0$, $v_1=\frac{b}{\|b\|}$, $V_1=[v_1]$, $U_0=[\,]$, $\hat{H}_0=[]$, $G_0=[\,]$ and *tolerance*, a user defined minimum for the minimum condition number of $\hat{H}_0$.. Set $p=0$, where p counts break down points.
Iteration $k=1, 2, 3, \ldots$ until convergence
Step 1: Define $h_k=V_k^T A v_k$ and $g_k=U_p^T A v_k$. Create orthogonal vector $\tilde{v}_{k+1}=Av_k-V_k h_k-U_p g_k$.
Step 2: Normalize orthogonal vector $v_{k+1}=\frac{\tilde{v}_{k+1}}{\|\tilde{v}_{k+1}\|}$
Step 3: Update $\hat{H}_k=\begin{pmatrix}\hat{H}_{k-1} & h_k\\ 0 & \|\tilde{v}_{k+1}\|\end{pmatrix}$
Step 4: Compute condition number of $\hat{H}_k$. If the condition number of $\hat{H}_k$ is larger than predefined *tolerance* level, i.e. if the standard Arnoldi process breaks down, go to Step 5. Otherwise, go to Step 13
Step 5: $p=p+1$ and $U_{p+1}=\begin{bmatrix}U_p & v_k\end{bmatrix}$
Step 6: Let $\hat{H}_{k-1}(k,:)$ denote the kth row of $\hat{H}_{k-1}$. Define $G_{k-1}=\begin{pmatrix}G_{k-1}\\ \hat{H}_{k-1}(k,:)\end{pmatrix}$ and set $\hat{H}_{k-1}(k,:)=0$
Step 7: Generate random vector $\hat{v}$ such that $\hat{v}$ is a unit vector, $V_{k-1}^T\hat{v}=0$ and $U_{p+1}^T\hat{v}=0$ i.e. $\hat{v}$ is orthogonal to the columns in V_{k-1} and U_p. Replace the last column of V_k, v_k with $\hat{v}$.
Step 8: Define $h_k=V_k^T A v_k$ and $g_k=U_p^T A v_k$. Create orthogonal vector $\tilde{v}_{k+1}=Av_k-V_k h_k-U_p g_k$.
Step 9: Normalize orthogonal vector $v_{k+1}=\frac{\tilde{v}_{k+1}}{\|\tilde{v}_{k+1}\|}$
Step 10: Update as $\hat{H}_k=\begin{pmatrix}\hat{H}_{k-1} & h_k\\ 0 & \|\tilde{v}_{k+1}\|\end{pmatrix}$
Step 11: If the condition number of $\hat{H}_k$ is larger than predefined *tolerance level*, i.e. if the standard Arnoldi process breaks down go to Step 7. Otherwise, go to Step 12.
Step 12: Set $V_{k+1}=\begin{bmatrix}V_k & v_{k+1}\end{bmatrix}$
Step 13: If $p>0$ then $G_k=[G_{k-1}, g_k]$ else set $G_k=G_{k-1}$
Step 14: Solve $\min\limits_{y_k\in\Re^k}\left\|\begin{pmatrix}\hat{H}_k\\ G_k\end{pmatrix}y_k-\|b\|e_1\right\|$
Step 15: $x_k=V_k y_k$

References

Arnoldi, W.E., 1951. The principle of minimized iterations in the solution of the matrix eigenvalue problem. Quarterly of Applied Mathematics 9, 17–29.
Bengtsson, C., Holst, J., 2002. On Portfolio Selection: Improved Covariance Matrix Estimation for Swedish Asset Returns. Working paper, Department of Economics, Lund University.
Chen, L.K.C., Karceski, J., Lakonishok, J., 1999. On portfolio optimization: forecasting covariances and choosing the risk model. Review of Financial Studies 12, 937–974.
DeMiguel, V., Garlappi, L., Nogales, F., Uppal, R., 2009. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. Management Science, 1–15.
Gilli, M., Pauletto, G., 1998. Krylov methods for solving models with forward-looking variables. Journal of Economic Dynamics and Control 22 (8–9), 1275–1289.
Golosnoy, V., Okhrin, Y., 2009. Flexible shrinkage in portfolio selection. Journal of Economic Dynamics and Control 33 (2), 317–328.
Golub, G.H., Van Loan, C.F., 1996. Matrix Computations, 3rd ed. Johns Hopkins University Press.
Jagannathan, R., Ma, T., 2000. Risk reduction in large portfolios: why imposing the wrong constraint helps. Journal of Finance 54 (4), 1651–1683.
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, 603–621.
Ledoit, O., Wolf, M., 2004. Honey, I shrunk the sample covariance matrix. Journal of Portfolio Management 30, 110–117.
Ledoit, O., Wolf, M., 2008. Robust performance hypothesis testing with the Sharpe ratio. Journal of Empirical Finance 15, 850–859.
Ledoit, O., Wolf, M., 2011. Robust Performance Hypothesis Testing with the Variance. Wilmott Magazine, pp. 86–89.
Markowitz, H., 1952. Portfolio selection. Journal of Finance **7**, 77–91.
Merton, R.C., 1980. On estimating the expected return on the market: an explanatory investigation. Journal of Financial Economics 8, 323–361.
Reichel, L., Ye, Q., 2005. Breakdown-free GMRES for singular systems. SIAM Journal on Matrix Analysis and Applications 26 (4), 1001–1021.
Saad, Y., 2003. Iterative Methods for Sparse Linear Systems, 2nd ed. SIAM.
Saad, Y., Schultz, M.H., 1986. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. SIAM Journal on Scientific and Statistical Computing 7, 856–869.
Sharpe, W.F., 1963. A simplified model for portfolio analysis. Management Science 9 (1), 277–293.
Simoncini, V., Szyld, D.B., 2007. Recent computational developments in Krylov subspace methods for linear systems. Numerical Linear Algebra with Applications 14, 1–59.
Smoch, L., 1999. Some results about GMRES in the singular case, Numerical Algorithms 22, 193–212.
Stein, C., 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: Neyman, J., (Ed.), Proceedings of the Third Berkeley Symposium on Mathematical and Statistical Probability, Vol. I, University of California, Berkeley, pp. 197–206.
Tola, V., Lillo, F., Gallegati, M., Mantegna, R.N., 2008. Cluster analysis for portfolio optimization. Journal of Economic Dynamics and Control 32 (1), 235–258.