I^2SDS
The Institute for Integrating Statistics in Decision Sciences

Technical Report TR-2010-13
May 27, 2010

A Krylov Subspace Approach to Large Portfolio Optimization

Isabelle Bajeux-Besnainou
Department of Finance
The George Washington University

Wachindra Bandara
Department of Finance
The George Washington University

Efstathia Bura
Department of Statistics
The George Washington University

A Krylov Subspace Approach to Large Portfolio Optimization

Isabelle Bajeux-Besnainou[1]

Department of Finance, The George Washington University

Wachindra Bandara

Department of Finance, The George Washington University

Efstathia Bura

Department of Statistics, The George Washington University

Abstract: Krylov subspaces are iterative method for solving large and sparse systems of linear equations. BFGMRES (Break Down Free Generalized Minimum RESidual) is a numerical algorithm that provides solutions to singular linear systems. With a large number of securities (N) and limited observations (T), deriving the global minimum variance portfolio requires the inversion of the singular sample covariance matrix of security returns. We use BFGMRES to estimate this portfolio without ex-ante assumptions on the covariance structure. Its cumulative returns outperform Ledoit and Wolf's while its Sharpe ratios have similar performances when applied to stock data between 1972 and 2008.

JEL classification: C02; G11

Keywords: Krylov subspaces; singular systems; algorithm; sample covariance matrix; global minimum portfolio

[1] Corresponding author: Finance Department, Funger Hall, Suite 501, 2201 G St, NW Washington DC 20052 Tel: 202 994 2559 Fax: 202 994 5014 E-mail: bajeux@gwu.edu

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. There have been significant advances in the field of applied mathematics during the last two decades following improvements in computer technology. Such are the Krylov subspace methods. They 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 (2006). Among these, GMRES (Generalized Minimum RESidual), introduced by Saad and Schultz (1986), has been most prominent for solving linear systems with a square non-singular matrix. However, the GMRES algorithm may breakdown when the matrix is singular, as in our application to portfolio management. Recent work by Reichel and Ye (2005) studies the various properties of GMRES algorithm at the point of breakdown and proposes an extension of the algorithm, BFGMRES (Breakdown Free Generalized Minimum RESdual), which can be applied to solving large linear systems with a singular matrix.

Assuming a mean-variance optimization objective, Markowitz (1952) showed that the 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 task (see Merton 1980), recent academic research (see 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:

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

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

Equation (1) shows that the portfolio weights of the global minimum variance portfolio depend on the inverse of the covariance matrix A. If A is non singular, the solution of these two equations is $x^* = A^{-1}\mathbf{1} / (\mathbf{1}^T A^{-1}\mathbf{1})$. 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 considered is larger than the number of available historical returns, the sample covariance matrix is singular and equations (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 (2003), Ledoit and Wolf (2003, 2004) . 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 always gives an invertible covariance matrix estimator. Their 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 use the break-down free version of the GMRES 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 estimator proposed by 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 controls the estimation errors reasonably well, we can eliminate the specification error introduced from using a specific shrinkage target.

The article is organized as follows. In section 2, we introduce the Krylov subspaces iterative methods. In section 3, we describe the GMRES and BFGMRES algorithms. We provide a short description of the Ledoit and Wolf estimates of the covariance matrix in section 4. Section 5 describes the data used for the analysis performed in section 5, where we compare the out-of-sample performance of our estimate

of the global minimum variance portfolio with the one derived from the shrinkage estimator. We conclude in section 6.

2. 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-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 *k*-th 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 $A = U\text{diag}(\theta_i)U^T$ is the spectral decomposition of the square matrix *A*, then $A^k = U\text{diag}(\theta_i^k)U^T$. As *k* gets large, the diagonal matrix of eigenvalues $\text{diag}(\theta_i^k)$ 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,

$$Ax = b,\ A \in R^{N \times N},\ x, b \in R^N \tag{3}$$

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

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 $A^0 = I$); that is,

$$K_k = \text{span}(b, Ab, ..., A^{k-1} b) \tag{4}$$

The best known Krylov subspace methods are the Arnoldi, Lanczos (Golub and Van Loan 1996), GMRES (generalized minimum residual) and BiCGSTAB (stabilized biconjugate gradient) 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.

3. GMRES and BFGMRES

The Generalized Minimal Residual Error algorithm, henceforth referred to as GMRES, for solving large linear systems as in (3) 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, even though with existing computer power 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, BFGMRES, that tackles breakdowns. The latter is used to obtain the solution of the linear system in (3) when *A* is a singular matrix.

3.1 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 $\|\mathrm{Ax_k} - b\|$, is reached at which point it terminates. Suppose this distance is reached at the k^{th} step and that the associated solution is x_k. Then, x_k is declared as the solution of the optimization problem given by

$$\min_{x_k \in K_k(A,b)} \|\mathrm{Ax_k} - b\| \tag{5}$$

where $K_k(A,b)$ is the k^{th} Krylov subspace generated by A and b defined in (3). An initial value for the solution, x_0, is needed to start the algorithm. Without loss of generality, we set $x_0 = 0$ [2].

Arnoldi Process

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 $\tilde{v}_{k+1} = Av_k - (h_{1,k}v_1 + ... + h_{k,k}v_k)$ where $h_{i,j} = v_i^T A v_j$. 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

$$V_k = \left[v_1, v_2, ..., v_k \right] \tag{6}$$

[2] 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 equation (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.

Additionally, the Arnoldi process, at each iteration, creates a $k \times k$ upper Hessenberg[3] matrix $V_k^T A V_k$.

From this we can write what is called the Arnoldi decomposition given by

$$AV_k = V_{k+1} H_k \tag{7}$$

where H_k is a matrix of size $(k+1) \times k$ given by $H_k = \begin{pmatrix} V_k^T A V_k \\ 0\ 0\ \ldots\ h_{k+1,k} \end{pmatrix}$. Here 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 k^{th} step and hence we have $x \in K_k(A,b)$. Then, there exists a y such that $x = V_k y$, 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 $Ax = AV_k y = V_{k+1} H_k y$ 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, we can write an equivalent form of the minimization problem as follows:

$$\min_{x \in k_k(A,b)} \|Ax - b\| = \min_{y \in \boldsymbol{R}^k} \|\beta e_1 - H_k y\| \tag{8}$$

This is a simpler and more familiar form of the minimization problem 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 k^{th} Krylov subspace. Hence, the k^{th} iterate of the

[3] 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).

solution in the k^{th} Krylov subspace $K_k(A,b)$ can be written as $x_k = V_k y$. The GMRES Algorithm is explicitly described in the Appendix.

3.2 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.

3.3 Breakdown-Free GMRES Algorithm

We first introduce notation used by Reichel and Ye (2005). The k^{th} 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 k^{th} axis vector. When $f_k \neq 0$ it is convenient to define the matrices

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

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 $U_1 = \left[v_p \right]$ with the old v_M that was replaced by $\hat{v}$. We will explain the reason below. From $AV_M = V_{M+1}\hat{H}_M$ we have that

$Av_{M-1} = h_{1,M-1}v_1 + h_{2,M-1}v_2 + ... + h_{M-1,M-1}v_{M-1} + h_{M,M-1}v_M$ and this can be written as

$Av_{M-1} = h_{1,M-1}v_1 + h_{2,M-1}v_2 + ... + h_{M-1,M-1}v_{M-1} + h_{M,M-1}u_1$ 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 $k = M, M+1, M+2, ...$ 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: $Av_k - V_k(V_k^T Av_k) - U_1(U_1^T Av_k) = f_k$ and then we can write

$$V_{k+1} = \left[V_k, \frac{f_k}{\| f_k \|} \right] \tag{9}$$

and

$$\hat{H}_k = \begin{bmatrix} H_k \\ \| f_k \| e_k^T \end{bmatrix} \tag{10}$$

We can now 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 $\hat{H}_k$ [4]. 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.

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 $U_2 = [U_1 \ \mathrm{v}_S]$. 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. Ledoit-Wolf Shrinkage Estimates of the Covariance Matrix

When the number of stocks *N* is larger than the number of historical returns per stock *T*, Ledoit and Wolf (2003, 2004) propose replacing the sample covariance matrix as an estimate of the true covariance matrix Σ with a weighted average of the sample covariance matrix and a low-variance target estimator, $\hat{\Sigma}_{target}$, an invertible positive definite symmetric $N \times N$ matrix. The noninvertible matrix $\hat{\Sigma}$ is replaced with the convex linear combination

$$\hat{\Sigma}_{LW} = (1-\lambda)\hat{\Sigma} + \lambda\hat{\Sigma}_{target} \tag{11}$$

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

where $\lambda \in [0,1]$ is the shrinkage intensity and controls how much weight is given to the target invertible matrix: for $\lambda = 1$ the shrinkage estimate equals the shrinkage target $\hat{\Sigma}_{target}$; for $\lambda = 0$, the unrestricted sample covariance matrix estimate $\hat{\Sigma}$ is recovered. This is a well-known technique in Statistics called shrinkage, originally developed by Stein (1956). The key advantage of this construction is that the regularized estimate $\hat{\Sigma}_{LW}$ is always invertible and outperforms the individual estimators $\hat{\Sigma}$ and $\hat{\Sigma}_{target}$ in terms of accuracy.

The target matrix Ledoit and Wolf use is Sharpe's (1963) single index model estimator, which imposes a lower dimensional structure on the estimator. In (12) the shrinkage intensity λ controls how much structure is imposed: the heavier the weight, the stronger the single-index model-based structure.

The shrinkage intensity is chosen by minimizing a risk function, $R(\lambda)$, such as the mean squared error (MSE), which in a matrix setting is the squared Frobenius norm:

$$R(\lambda) = E \,||\, \lambda\hat{\Sigma}_{target} + (1-\lambda)\hat{\Sigma} - \Sigma \,||^2 = E\left(\sum_{i=1}^{p}\sum_{j=1}^{p} \left(\lambda s_{ij}^{target} + (1-\lambda)s_{ij} - \sigma_{ij} \right)^2 \right) \qquad (12)$$

where $\Sigma = (\sigma_{ij})$, $\hat{\Sigma}_{target} = (s_{ij}^{target})$ and $\hat{\Sigma} = (s_{ij})$.

Ledoit and Wolf (2003) derived a methodology for choosing λ that guarantees minimizing $R(\lambda)$ without the need of having to specify any underlying distribution, and without requiring computationally expensive procedures such as MCMC (Markov Chain Monte Carlo), bootstrap, or cross-validation. In more detail, Sharpe's (1963) single-index model assumes that stock returns are generated by the model:

$$x_{it} = \alpha_i + \beta_i x_{0t} + \varepsilon_{it} \qquad (13)$$

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

$$\Phi = \sigma_{00}^2 \beta\beta^T + \Delta \tag{14}$$

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

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

$$F = s_{00}^2 bb^T + D \tag{15}$$

where s_{00}^2 is the sample variance of 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

$$\lambda^* = \max\left\{0, \min\left\{\frac{\hat{\kappa}}{T}, 1\right\}\right\} \tag{16}$$

where

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

with $\hat{\gamma} = \sum_{i=1}^{N}\sum_{j=1}^{N}(f_{ij} - s_{ij})^2$, $\hat{\pi} = \sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\pi}_{ij}$, $\hat{\pi}_{ij} = \sum_{t=1}^{T}\left\{(x_{it} - \overline{x}_{i\cdot})(y_{jt} - \overline{x}_{j\cdot}) - s_{ij}\right\}^2 / T$, and

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

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

The optimal shrinkage density is truncated at 0 and 1 in order to avoid overshrinkage or negative shrinkage (negative λ) as in finite samples $\hat{\kappa}/T$ may by higher than 1 or be negative. It is also worth noting that (17) is valid regardless of the sample size T. In particular, T can be substantially smaller than N, which is the case with the data considered in this article.

5. Data

We extracted equity returns from the Center for Research in Securities prices (CRSP) monthly database[5]. 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.

We use the monthly returns to generate the sample covariance matrix using data from January of year $t-10$ until December of year t. This covariance matrix is used to generate the weights for the global minimum variance portfolio. We build this portfolio in January of year $t+1$ and hold it until December of year $t+1$. In this case the in-sample period is from January 1st of year $t-10$ until December 31st of year t. The out-of-sample period is one year, from January 1st of year $t+1$ until December 31st of year $t+1$. To summarize, we build the minimum variance portfolio using the 10-year in-sample period

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

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 expected return, standard deviation of returns and Sharpe ratio on the out-of-sample period. To compute the Sharpe ratio, we use the monthly 3-month T-Bill rate (observed in the out-of-sample period) as a proxy for the risk-free rate. The T-Bill data are extracted from the Federal Reserve Economic Data (FRED).

For each 11- year in-sample-out-of-sample period we considered only the equities that were available for the entire 11-year period. Equities that entered the 11-year 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 within the window. The number of equities considered for all 27 periods ranges from 1226 to 2416. Since we have 120 observations in each period, the sample covariance matrix is always very singular.

6. Results

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

As this is an industry standard, we use the Sharpe ratio to compare the two methods. The Sharpe ratio is calculated as follows:

$$\text{Sharpe Ratio}=\frac{\text{Mean Monthly Out of Sample Return-Risk Free Rate}}{\text{Standard Deviation of Monthly Returns}}$$

In our calculations we use a one-year forecast window. Hence, we use the following formula to compute the Sharpe Ratio:

$$\text{Sharpe Ratio}=\frac{\text{Mean Out of Sample Returns Over 12 Months-Average 3-Month T-Bill Rate Over 12 Months}}{\text{Standard Deviation of Out of Sample Returns Over 12 Months}}$$

<Insert Table 1>

In Table 1, where the Sharpe Ratios for the two methods are reported, we see that in 17 out of 27 periods, or 63% of the time, the BFGMRES Sharpe ratios outperform the Ledoit and Wolf Sharpe ratios. To investigate whether the Sharpe ratios corresponding to the two methods over these 27 years are statistically significantly different, we first applied the Durbin-Watson Test for autocorrelation (see for example Gujarati 2003, p. 467-472) to the two sets of Sharpe ratios to test whether there is temporal dependence. In both cases, the 27 Sharpe Ratios were not found to be serially correlated (p-value= 0.56 for the BFGMRES Sharpe Ratios and p-value=0.66 for the Ledoit-Wolf Sharpe Ratios). Thus we can assume the BFGMRES and Ledoit-Wolf Sharpe Ratios are independently distributed over time. The correlation between the Sharpe Ratios of the two methods is 0.97 indicating they are highly correlated within a year and cannot be assumed to be independent samples. Moreover, with 27 observations we cannot assess whether the Sharpe Ratios of the two methods are normally distributed, which is required in order to use Jobson and Korkie's (1981) approach. We use the Wilcoxon signed-rank test (see, e.g. Hollander and Wolfe 1999, p. 36), the nonparametric analogue to the paired t-test for two related samples, to assess whether the median Sharpe Ratio for the GMRES method differs from the Ledoit-Wolf median Sharpe Ratio. The test has a p-value of 0.26 for

testing whether the two medians differ and a *p*-value of 0.13 for testing whether the median Sharpe Ratio for BFGMRES exceeds that of Ledoit-Wolf. Thus, we conclude that there is no statistically significant difference between the two and that both methods perform roughly the same with respect to Sharpe Ratios.

Ledoit and Wolf (2003) use standard deviations to compare the performance of the different shrinkage estimators of the covariance matrix. We also compute the standard deviations of the returns of the optimal portfolios over the 27 out-of-sample periods (1982 to 2008). Due to the closeness of the results, to make sure that we have a basis for comparison of the two methods, we apply Levene's test (Levene 1960; Gastwirth et al. 2009) to test whether the standard deviations corresponding to BFGMRES and Ledoit and Wolf generated returns are different. Levene's test makes no assumptions about the underlying distribution of the returns. We find that their standard deviations are not statistically significantly different (*p*-value=0.5298) and thus they cannot be used to compare the performance of the two methods.

As the standard deviations of the returns generated by BFGMRES and Ledoit-Wolf are not found to be statistically significantly different, we can compare the two methods with respect to the expected returns they generate. We use the expected annualized returns of the two methods for our comparison and followed the same analysis steps as for the Sharpe Ratios. The Durbin-Watson test for autocorrelation was not statistically significant in either case (*p*-value= 0.43 for the BFGMRES annualized returns and *p*-value=0.59 for the Ledoit-Wolf annualized returns) so that we can treat the two sets of annualized returns as samples of serially independent observations. The Wilcoxon signed rank test has a *p*-value of 0.055 for testing whether the BFGMRES

median annualized return exceeds that of Ledoit-Wolf and is marginally significant at level 5%. That is, we conclude the median BFGMRES return exceeds that of Ledoit-Wolf at any level greater than 5.5%.

<Insert Figure 1>

In Figure 1, we plot of the cumulative returns for the two methods. Over the period from 1982 to 2008, the cumulative returns generated using BFGMRES are consistently higher than the cumulative returns from Ledoit-Wolf's method. Even more interesting is the finding that the distance between the BFGMRES based cumulative returns and those of Ledoit and Wolf increases as time increases. That is, if one were using the BFGMRES method to generate the minimum variance portfolio rather than Ledoit-Wolf's, the improvement in cumulative returns would steadily increase. Over the 26 years simulation period, the cumulative returns of the BFGMRES method end up to outperform the Ledoit and Wolf cumulative returns by 31.7%.

7. 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. Yet, the performance of this purely numerical algorithm on average is better with respect to Sharpe ratios, although not statistically significantly so, and consistently better in terms of cumulative returns over the 27 years of our data analysis.

Although this article addresses only the issue of the global minimum variance portfolio, the entire efficient frontier can be derived using the same algorithm as long as the expected returns of the stocks can be estimated.

Appendix

GMRES Algorithm

Iteration k= 0:

Initialize: Set $x_0 = 0$, $v_1 = b / \| b \|$ and $V_1 = [v_1]$

Iteration k=1

Step 1: Create orthogonal vector $\tilde{v}_2 = Av_1 - h_{1,1}v_1$ with $h_{1,1} = v_1^T Av_1$

Step 2: Normalize orthogonal vector $v_2 = \tilde{v}_2 / \|\tilde{v}_2\|$

Step 3: Update the V matrix $V_2 = [v_1, v_2]$ and set $H_1 = \begin{pmatrix} h_1 \\ \|\tilde{v}_2\| \end{pmatrix}$

Step 4: Solve the simplified least squares optimization $\min_{y \in R} \|\beta e_1 - H_1 y\|$ and call the solution y_1

Step 5: Calculate $x_1 = V_1 y$

Iteration k >1 until convergence

Step 1: Create orthogonal vector $\tilde{v}_{k+1} = Av_k - (h_{1,k}v_1 + ... + h_{k,k}v_k)$ with $h_{i,j} = v_i^T Av_j$

Step 2: Normalize orthogonal vector $v_{k+1} = \tilde{v}_{k+1} / \|\tilde{v}_{k+1}\|$

Step 3: Update V matrix with $V_{k+1} = [v_1, v_2, ..., v_{k+1}]$ and H matrix with

$H_k = \begin{pmatrix} H_{k-1} & h_k \\ 0 & \|v_{k+1}\| \end{pmatrix}$ where $h_k = V_k^T Av_k$

Step 4: Solve the simplified least squares optimization $\min_{y \in R^k} \|\beta v_1 - H_k y\|$ and call the solution y_k

Step 5: Compute $x_k = V_k y_k$

BFGMRES Algorithm

Input *A, b*

Iteration *k* = 0:

Initialize: Set $x_0 = 0$, $v_1 = b / \| b \|, V_1 = [v_1]$, $U_0 = [\ ]$, $\hat{H}_0 = [\ ], G_0 = [\ ]$ and *tolerance*, a user defined minimum for the minimum condition number of $\hat{H}_k$. Set $p = 0$, where *p* counts break down points.

Iteration *k*=1, 2, 3,… until convergence

Step 1: Define $h_k = V_k^T A v_k$ and $g_k = U_p^T A v_k$. Create the orthogonal vector $\tilde{v}_{k+1} = A v_k - V_k h_k - U_p g_k$.

Step 2: Normalize the orthogonal vector $v_{k+1} = \tilde{v}_{k+1} / \left\| \tilde{v}_{k+1} \right\|$

Step 3: Update $\hat{H}_k = \begin{pmatrix} \hat{H}_{k-1} & h_k \\ 0 & \left\| \tilde{v}_{k+1} \right\| \end{pmatrix}$

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: Set $p = p + 1$ and $U_{p+1} = [U_p \ v_k]$

Step 6: Let $\hat{H}_{k-1}(k,:)$ denote the *k*th 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 of 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 $g_k = U_{p+1}^T A v_k$. Create orthogonal vector $\tilde{v}_{k+1} = A v_k - V_k h_k - U_{p+1} g_k$.

Step 9: Normalize orthogonal vector $v_{k+1} = \tilde{v}_{k+1} / \left\| \tilde{v}_{k+1} \right\|$

Step 10: Update $\hat{H}_k = \begin{pmatrix} \hat{H}_{k-1} & h_k \\ 0 & \left\| \tilde{v}_{k+1} \right\| \end{pmatrix}$

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 $G_k = \begin{bmatrix} G_{k-1} & g_k \end{bmatrix}$ else set $G_k = G_{k-1}$

Step 14: Solve $\min_{y_k \in R^k} \left\| \begin{pmatrix} \hat{H}_k \\ G_k \end{pmatrix} y_k - \left\| b \right\| e_1 \right\|$

Step 15: $x_k = V_k y_k$

References

1) Arnoldi, W. E. 1951. The principle of minimized iterations in the solution of the matrix eigenvalue problem, *Quarterly of Applied Mathematics*, 9:17–29.

2) Bengtsson, C. and Holst, J. 2002. On portfolio selection: Improved Covariance Matrix Estimation for Swedish Asset Returns. *Working paper*, Department of Economics, Lund University.

3) Chen, L. K. C., Karceski, J. and Lakonishok J., 1999. On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model. *Review of Financial Studies*, **12**: 937-974.

4) DeMiguel V., Garlappi L., Nogales F. and Uppal R., 2009. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms, Management Science, 1-15.

5) Gastwirth, J.L., Gel, Y.R. and Miao, W. 2009. The Impact of Levene's Test of Equality of Variances on Statistical Theory and Practice (to appear in *Statistical Science*).

6) Golub, G. H. and Van Loan, C. F. 1996 *Matrix Computations (3rd Ed.)*. Johns Hopkins University Press.

7) Gujarati, D. 2003 *Basic Econometrics (3rd Ed.)* McGraw-Hill

8) Hollander, M. and Wolfe, D. A. 1999. *Nonparametric Statistical Methods* (2nd ed.) Wiley and Sons.

9) Ipsen, I. and Meyer, C. 1998. The Idea behind Krylov *Methods The American Mathematical Monthly*, Vol. 105, No. 10, pp. 889-899

10) Jahannathan, R. and Ma, T. 2000. Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraint Helps. *Journal of Finance*, 54(4): 1651-1683.

11) Ledoit, O. and 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.

12) Ledoit, O. and Wolf, M. 2004. Honey, I shrunk the sample covariance matrix, *Journal of Portfolio Management*, **30**:110-117.

13) Levene, H. 1960. Robust tests for equality of variances, in Ingram Olkin, Harold Hotelling et al. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, Stanford University Press, pp. 278–292.

14) Markowitz, H. 1952. Portfolio Selection, *Journal of Finance* **7**:77-91.

15) Merton, R. C. 1980. On estimating the expected return on the market: An explanatory investigation. *Journal of Financial Economics* 8: 323-361.

16) Reichel, L. and Ye, Q. 2005. Breakdown-free GMRES for singular systems. SIAM J. Matrix Anal. Appl., 26(4): 1001-1021.

17) Saad, Y. 1992. *Numerical Methods for Large Eigenvalue Problems*, Manchester University Press.

18) Saad, Y. (2003). *Iterative methods for sparse linear systems* (2nd ed.). SIAM.

19) Saad, Y. and Schultz, M.H. (1986) GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems, *SIAM J. Sci. Stat. Comput.*, **7**:856-869.

20) Sharpe, W. F. (1963). A simplified model for portfolio analysis. *Management Science*, **9**(1):277-293.

21) Simoncini, V. and Szyld, D. B. (2007). Recent computational developments in Krylov subspace methods for linear systems. *Numerical Linear Algebra with Applications*, 14:1-59.

Figure 1

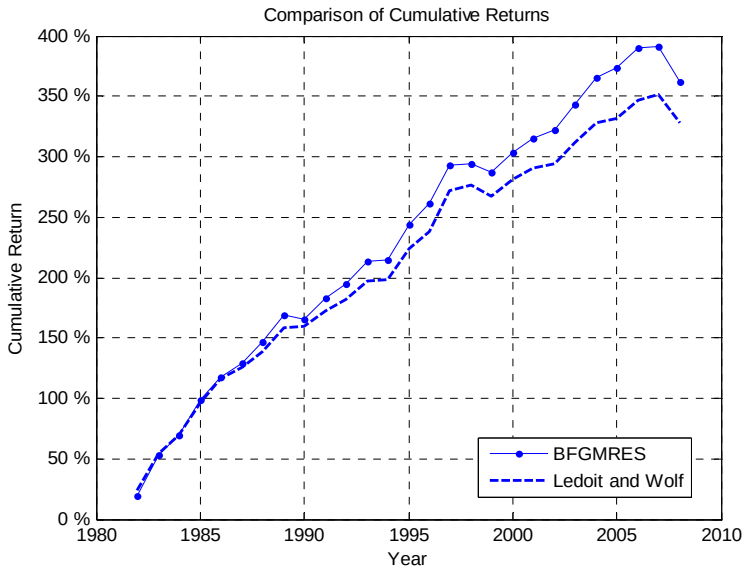

Figure 1: The cumulative returns generated by the two methods plotted against each year in the out-of-sample period, 1982-2008. The cumulative returns are computed as the sum of the annual returns for each out-of-sample year in the period. Over this period, BFGMRES yields higher cumulative returns compared to Ledoit and Wolf's method.

Table 1

	Sharpe Ratio	
Out-of-sample year	BFGMRES	Ledoit and Wolf
1982	0.70	1.18
1983*	5.29	4.76
1984	1.30	1.38
1985	4.00	4.33
1986	1.32	1.55
1987*	0.33	0.32
1988*	2.53	1.91
1989*	3.89	2.81
1990	-1.72	-1.28
1991*	2.37	1.46
1992	1.11	1.19
1993*	1.78	1.67
1994*	-0.38	-0.66
1995*	5.63	5.46
1996*	2.38	1.88
1997*	3.54	3.28
1998	-0.26	0.00
1999*	-1.18	-1.38
2000*	1.38	1.05

2001*	1.06	0.87
2002*	0.52	0.28
2003	3.00	3.49
2004*	2.83	2.38
2005*	0.71	0.00
2006	2.47	3.62
2007	-0.57	0.13
2008*	-1.91	-2.21

TABLE 1: The Sharpe ratios for each of the out-of-sample years between 1982 and 2008. The Sharpe ratios have been calculated on a monthly basis and then annualized i.e. the asset returns are multiplied by 12 and the standard deviations are multiplied by $\sqrt{12}$. The risk-free rate is computed as the sum of the monthly returns. The * indicates the years where BFGMRES outperforms the Ledoit and Wolf method: 17 out of the 27 years in the data set we have analyzed.