On Portfolio Selection: Improved Covariance Matrix Estimation for Swedish Asset Returns

Christoffer Bengtsson*
Department of Economics
Lund University

Jan Holst
Mathematical Statistics
Lund Institute of Technology

4th October 2002

Abstract

Mean-Variance (MV) theory for portfolio selection is based on assumptions involving parameters that have to be estimated using historical data. Depending on the method of estimation, the estimates will suffer from estimation error and/or specification error, both of which will effect the portfolio optimization in such a way that the resulting optimal portfolio is not the true optimal portfolio. It is therefore of interest to make the estimates as good as possible, in order to avoid as much as possible the effects of this uncertainty.

In this paper we focus on the estimation of the covariance matrix for stock returns on the Swedish market. This is one of the two input parameters of MV optimization, the other being the expected return vector. We do this using Bayesian shrinkage and principal component analysis in combination with random matrix theory. Our empirical results implies that such an approach is better than all those previously proposed.

1 Introduction

The Mean-Variance (MV) theory of Markowitz (1952,1959) for portfolio selection is by many considered as one of the most important contributions to modern financial economics. The method is however sensitive to small changes in the two input parameters, the expected return vector and the covariance matrix for the returns. In this paper we will concentrate on improving the performance of MV optimization for portfolio selection on the Swedish stock market by improving the estimation of the covariance matrix. For this purpose we construct a contest between different estimators where the metric is the out-of-sample volatility of the global minimum variance portfolio implied by the different estimators.

It is often argued that estimating the covariance matrix is less important than estimating the expected returns. The argument is that the portfolio weights are more sensitive to

*E-mail christoffer.bengtsson@nek.lu.se, phone +46 (0)46 - 222 79 09.

changes in the expected return vector than changes in the covariance matrix. But, at the same time, the difficulty of estimating expected returns compared to estimating the covariance matrix implies that the most improvement that can actually be made on MV optimization lies in covariance matrix estimation. Secondly, the estimation of the expected returns should preferably use restrictions from economic theory since it implies a return to equilibrium models such as the Capital Asset Pricing Model (CAPM) or factor models like the Arbitrage Pricing theory (APT), while the estimation of the covariance matrix is mainly based on financial econometrics.

The usual sample covariance matrix is rarely used since the portfolios it produces often are practically worthless, at least when no portfolio weight restrictions are imposed to control for extreme long and short positions caused by estimation error. The primary problem is that the sample covariance matrix requires estimating too many parameters because it imposes too little structure and that it overfits the sample data. The cure is naturally to impose some structural assumptions, but at the same time it is usually hard to decide on how much and what factor structure to impose. In this paper we will impose structure by combining Bayesian Stein shrinkage with principal component analysis where we rely on random matrix theory to determine how many principal components it is meaningful to include.

The proposed estimators are combinations of the work on random matrix theory for stock correlations by Plerou et al. (2001) and the estimators proposed by Ledoit & Wolf (2001) and Jagannathan & Ma (2000). Ledoit & Wolf (2001) develop a shrinkage estimator that is an optimally weighted average of the sample covariance matrix and the covariance matrix implied by the market model of Sharpe (1963). Jagannathan & Ma (2000) extend this notion by adding a third matrix that is the diagonal part of the sample covariance matrix. In both cases we replace the covariance matrix implied by the market model with a K factor principal component covariance matrix where the choice of the number of factors K is inspired by Plerou et al. (2001).

The paper is structured in the following way. In section two we give a short introduction of MV theory, which is the basis and motivation for the entire paper. In section three we discuss a few techniques on how to estimate the covariance matrix. This includes a discussion on principal component analysis and random matrix theory, which are the tools needed for improving the estimators of Ledoit & Wolf (2001) and Jagannathan & Ma (2000). In section four we discuss the shrinkage estimator of Ledoit & Wolf (2001) and make the modifications needed to replace the covariance matrix implied by the market model of Sharpe (1963) with a K factor principal component covariance matrix. Which estimator is the best is an empirical issue. We will devote section five to this by letting the developed covariance matrix estimators compete against those proposed in literature. We conclude the paper with a brief summation in section six.

2 Mean-Variance Portfolio Selection

The MV portfolio selection theory orders different strategies by only taking into account the first two moments of the probability distributions of the asset returns, mean and variance, the variance being our risk measure. Consequently, a rational investor will only choose between those portfolios that for a certain expected return offers the least possible amount of risk, or alternatively, maximizes the level of expected return given a certain willingness to take on risk.

We define $\boldsymbol{\mu} = [E[r_1], E[r_2], ..., E[r_N]]^T$ as the vector of expected returns and $\boldsymbol{\Sigma}$ as the covariance matrix for the returns where $\sigma_{ii} = \sigma_i^2 = \mathsf{Var}[r_i]$ and $\sigma_{kl} = \mathsf{Cov}[r_k, r_l]$. A portfolio consisting of N assets is then described by vector of weights, $\mathbf{w} = [w_1, w_2, ..., w_N]^T$, where the ith weight w_i is the fraction of the total amount of invested capital that is placed on asset i. By definition the weights must sum up to one, i.e. $\mathbf{w}^T\mathbf{1} = 1$, which is the investor's budget constraint. The expected return and variance of the portfolio is then simply

$$\begin{aligned} \mathsf{E}[r_p] &= \mathbf{w}^T\boldsymbol{\mu} && (1) \\ \mathsf{Var}[r_p] &= \mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}. && (2) \end{aligned}$$

Not including any risk free asset and allowing for unlimited short selling, we can write the simplest MV optimization problem as

$$\max_{\mathbf{w}} \left\{ t\mathbf{w}^T\boldsymbol{\mu} - \frac{1}{2}\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} \quad | \quad \mathbf{w}^T\mathbf{1} = 1 \right\}, \tag{3}$$

where $\mathbf{1}$ denotes a $N \times 1$ vector of ones and the parameter t is the MV investor's risk tolerance parameter which relates expected return to risk. Solving equation 3 yields the well known formula for the optimal portfolio as

$$\mathbf{w}^* = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{c} + t\left(\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - \frac{\mathbf{1}a}{c}\right)\right), \tag{4}$$

where

$$\begin{aligned} a &= \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\ c &= \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}. \end{aligned}$$

For future references we also note that the only portfolio that will not require an estimate of $\boldsymbol{\mu}$, i.e. that only depends on the covariance matrix, is the minimum variance portfolio (MVP) which is given by

$$\mathbf{w}_{MVP} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{c}. \tag{5}$$

3 Some Covariance Matrix Estimators

3.1 The Sample Covariance Matrix

The sample covariance matrix has the appealing quality of being maximum likelihood (ML) under the assumption of normality, meaning that it is the best unbiased estimator. However, it also comes with some major drawbacks. First of all, the fact that the sample

covariance matrix is ML may not only be positive. ML means that we put all our trust in the data, which is a sound principle if have enough data. In small samples, however, we may therefore run the risk of overfitting the data. This means that although the sample covariance matrix performs the best in-sample, it may perform very poorly out-of-sample. Secondly, we have to estimate quite a lot of parameters, namely $N(N+1)/2$. As an example, this means that if we have say 100 assets to choose from, we have to estimate 5050 parameters. Estimating so many parameters is clearly a problem in the sense that we again need a lot of data. This can be illustrated in the following way.

Define $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_T]$ as an $N \times T$ matrix where the rows are the time series of historical returns for the different assets and

$$\bar{\mathbf{r}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{r}_t = \frac{1}{T}\mathbf{R}\mathbf{1}.$$

We can then write the sample covariance matrix as

$$\mathbf{S} = \frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{r}_t - \bar{\mathbf{r}})(\mathbf{r}_t - \bar{\mathbf{r}})^T = \frac{1}{T-1}\mathbf{R}\left(\mathbf{I} - \frac{1}{T}\mathbf{1}\mathbf{1}^T\right)\mathbf{R}^T, \tag{6}$$

where $\mathbf{I}$ denotes the identity matrix and T equals the number of historical observations on each of the assets[1]. From equation 6 we make the important observation that the sample covariance matrix will never be invertible if $N \geq T$, even if true the covariance matrix is. This follows since the maximum rank of $\mathbf{S}$ will be the rank of $\mathbf{I} - \mathbf{1}\mathbf{1}^T/T$ which is $T-1$. Thus if the the number of assets exceeds the number of available observations, the sample covariance matrix will be non-invertible. This is particularly alarming since we saw in equation 4 that the investor's optimal portfolio choice depends upon the inverse of the covariance matrix.

3.2 The Market Model

In this context, factor models assume that the returns are generated by specific exogenous factors, usually observables with some economical interpretation. This means that we impose some form of structural assumptions and the fewer the factors, the stronger the structure. Not surprisingly, the most important factor usually proves to be the market. The well known capital asset pricing model (CAPM) can for example can be thought of as a one factor model depending on the market. This is Sharpe's (1963) single-index market model.

We formulate the single-index model for an individual asset as follows

$$r_{it} = \alpha_i + \beta_i r_{Mt} + \varepsilon_{it} \tag{7}$$

or in matrix form for all N assets as

$$\mathbf{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\boldsymbol{r}_{Mt} + \boldsymbol{\varepsilon}_t, \tag{8}$$

[1]Not be confused with the superscript T, which always denotes the transpose of a matrix/vector.

where $\boldsymbol{\varepsilon}_t$ is a $N \times 1$ vector containing the zero mean uncorrelated residuals ε_{it}. The covariance matrix for the asset returns, implied by the market model, then becomes

$$\boldsymbol{\Phi} = \sigma_M^2 \boldsymbol{\beta}\boldsymbol{\beta}^T + \boldsymbol{\Sigma}_\varepsilon, \tag{9}$$

where σ_M^2 is the variance of the market portfolio. We can estimate this model by e.g. performing a multivariate regression. We denote the estimated model by

$$\mathbf{F} = s_M^2 \mathbf{b}\mathbf{b}^T + \hat{\boldsymbol{\Sigma}}_\varepsilon. \tag{10}$$

We note that in choosing a single-index covariance matrix we must only estimate $2N + 1$ parameters, which is a major reduction compared to the sample estimator. Since we have more data per estimated parameter, we can therefore expect a reduction in estimation error. This is of course at the expense of introducing specification error instead, due to the rather restrictive assumption that the asset returns are generated by a simple linear function of the market. On a more technical note, when estimating covariances it is sometimes the case that an equally weighted index is better at explaining covariances than the actual value weighted index (see Ledoit & Wolf (2001)). This is a slight departure from CAPM theory where the composition of r_M is of great importance. Therefore, when we use the single-index covariance matrix in the remainder of the paper, we will examine both the case when r_M is a value weighted index and the case when it is a simple average over all assets.

3.3 Principal Component and Factor Analysis

3.3.1 Sample Principal Components

The goal of principal component analysis (PCA) is to explain covariance structures using only a few linear combinations of the original stochastic variables. What PCA accomplishes is (1) data reduction and (2) interpretability. For an $N \times N$ covariance matrix, N principal components are needed to reproduce all variability in the system, but often, however, most variability can be explained by a lesser number K, $K < N$, principal components without losing much information. As a bonus, PCA often prove to reveal relationships that otherwise would have been hard to see or even suspect.

Definition 1. *The (sample) principal components are those linear combinations*

$$f_i = \boldsymbol{\ell}_i^T \mathbf{r} = \sum_{n=1}^{N} \ell_{ni} r_n \quad i = 1, 2, ..., N \tag{11}$$

for which the variances and covariances are given by

$$\begin{aligned} \mathsf{Var}[f_i] &= \boldsymbol{\ell}_i^T \mathbf{S} \boldsymbol{\ell}_i \quad i = 1, 2, ..N \\ \mathsf{Cov}[f_i, f_j] &= \boldsymbol{\ell}_i^T \mathbf{S} \boldsymbol{\ell}_j \quad i, j = 1, 2, ..N \end{aligned} \tag{12}$$

of the N stochastic variables (in our case asset returns) that gives the direction of maximum variance in the sample in such a way that the principal components are uncorrelated to each other and $\boldsymbol{\ell}_i^T \boldsymbol{\ell}_i = 1$, $i = 1, 2, ..., N$.

Let the stochastic vector $\mathbf{r} = [r_i, r_2, ..., r_N]^T$ be any observation on $\mathbf{r}_t = [r_{it}, r_{2t}, ..., r_{Nt}]^T$ for $t = 1, 2, ..., T$ and let $\mathbf{r}$ have the (sample) covariance matrix $\mathbf{S}$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N \geq 0$. We then find the ith principal component f_i from the solution of

$$\max_{\ell_i} \left\{ \frac{\boldsymbol{\ell}_i^T \mathbf{S} \boldsymbol{\ell}_i}{\boldsymbol{\ell}_i^T \boldsymbol{\ell}_i} \quad | \quad \mathsf{Cov}[f_i, f_j] = 0, \forall 0 < j < i \right\}. \tag{13}$$

The optimal weight functions $\boldsymbol{\ell}_i$ are thus functions of the covariance matrix alone. As it turns out, we have the following result that enables us to extract the optimal weight functions from the covariance matrix without solving a number of optimizations.

Proposition 1. *Let* $\mathbf{S}$ *be the (sample) covariance matrix for the stochastic vector* $\mathbf{r} = [r_i, r_2, ..., r_N]^T$ *and let* $\mathbf{S}$ *have eigenvalue-eigenvector pairs* $(\mathbf{e}_1, \lambda_1)$, $(\mathbf{e}_2, \lambda_2)$, ..., $(\mathbf{e}_N, \lambda_N)$ *where* $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N \geq 0$ *and* $\mathbf{e}_j = [e_{1j}, e_{2j}, ..., e_{Nj}]^T$. *Then the* i*th (sample) principal component is given by*

$$f_i = \mathbf{e}_i^T \mathbf{r} = \sum_{n=1}^{N} e_{ni} r_n \quad i = 1, 2, ..., N \tag{14}$$

and

$$\begin{aligned} \mathsf{Var}[f_i] &= \mathbf{e}_i^T \mathbf{S} \mathbf{e}_i = \lambda_i \quad i = 1, 2, ..N \\ \mathsf{Cov}[f_i, f_j] &= \mathbf{e}_i^T \mathbf{S} \mathbf{e}_j = 0 \quad i \neq j. \end{aligned} \tag{15}$$

Proof. See Johnson & Wichern (1992). □

Thus, the principal components are determined by the eigenvectors of $\mathbf{S}$ and their variances equal the corresponding eigenvalues. For future discussions the following proposition will also be useful.

Proposition 2. *Let* $\mathbf{S}$ *be the (sample) covariance matrix for the stochastic vector* $\mathbf{r} = [r_i, r_2, ..., r_N]^T$ *and let* $\mathbf{S}$ *have eigenvalue-eigenvector pairs* $(\mathbf{e}_1, \lambda_1)$, $(\mathbf{e}_2, \lambda_2)$, ..., $(\mathbf{e}_N, \lambda_N)$ *where* $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N \geq 0$ *and* $\mathbf{e}_j = [e_{1j}, e_{2j}, ..., e_{Nj}]^T$. *Let* $f_1 = \mathbf{e}_1^T \mathbf{r}$, $f_2 = \mathbf{e}_2^T \mathbf{r}$, ..., $f_N = \mathbf{e}_N^T \mathbf{r}$ *be the (sample) principal components. Then the (sample) correlation between the* i*th asset and the* j*th principal component is given by*

$$\mathsf{Corr}[r_i, f_j] = \frac{e_{ij} \sqrt{\lambda_j}}{s_{ii}}. \tag{16}$$

Proof. See Johnson & Wichern (1992). □

3.3.2 Standardized Principal Components

Sometimes variables measured on different scales with differing ranges, e.g. assets with widely different magnitudes of volatility, are standardized to avoid an unreasonably large impact on the principal components by only a few variables. Their interpretational properties may also become clearer. Especially for financial assets it is intuitively more appealing to investigate the correlations rather than the covariances. In this way the first principal component for asset returns usually proves to represent a kind of market factor. Other principal components often mirror industry specific effects.

The standardization of the variables is performed in the following manner

$$\mathbf{z}_t = \mathbf{V}^{-1/2}(\mathbf{r}_t - \bar{\mathbf{r}}) \quad t = 1, 2, ..., T, \tag{17}$$

where $\mathbf{V}$ is a diagonal matrix containing the sample variances. For the matrix of standardized observations $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_T]$, the standardized sample covariance matrix is given by

$$\mathbf{S}_z = \mathbf{C} = \frac{1}{T-1}\mathbf{Z}\mathbf{Z}^T. \tag{18}$$

The principal components of $\mathbf{Z}$ are obtained in the same way as before by simply replacing $\mathbf{S}$ with $\mathbf{C}$ in proposition 1. However, the principal components will be greatly affected by the standardization, as they should be, and the quantities derived from $\mathbf{C}$ are generally not the same as those derived from $\mathbf{S}$. In fact, the two sets of principal components will in no way be simple functions of each other.

3.3.3 The Principal Component Method for Orthogonal Factor Models

As before, let the N dimensional stochastic vector $\mathbf{r}$ be any of the stochastic variables $\mathbf{r}_t$ for $t = 1, 2, ..., T$ with (sample) mean $\mathbf{m}$ and (sample) covariance matrix $\mathbf{S}$. The factor model is based on the assumption that the variables all depend on a number of underlying and unobservable stochastic factors, denoted $F_1, F_2, ..., F_K$, as well as the variable specific errors/variations $\varepsilon_1, \varepsilon_2, ..., \varepsilon_N$. Written in matrix form the factor model looks as

$$\mathbf{r} - \mathbf{m} = \mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon}. \tag{19}$$

We call the coefficients ℓ_{ij} of $\mathbf{L}$ the loading on variable i by factor j. Here $\mathbf{F}$ is an unobservable vector of the factors, not to be confused with the single-index covariance matrix also denoted $\mathbf{F}$. Furthermore we assume that

$$\begin{aligned} \mathsf{E}[\mathbf{F}] = 0 \quad \mathsf{Var}[\mathbf{F}] = \mathsf{E}[\mathbf{F}\mathbf{F}'] = \mathbf{I} \\ \mathsf{E}[\boldsymbol{\varepsilon}] = 0 \quad \mathsf{Var}[\boldsymbol{\varepsilon}] = \mathsf{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \boldsymbol{\Psi}, \end{aligned} \tag{20}$$

where $\boldsymbol{\Psi}$ is a diagonal matrix containing the residual variances and $\mathsf{Cov}[\mathbf{F}, \boldsymbol{\varepsilon}] = 0$. The covariance structure for $\mathbf{r}$ implied by our assumptions then becomes

$$\begin{aligned} \mathsf{Var}[\mathbf{r}] &= \mathsf{E}[(\mathbf{r} - \mathbf{m})(\mathbf{r} - \mathbf{m})^T] = \mathsf{E}[(\mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon})(\mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon})^T] \\ &= \mathbf{L}\mathsf{E}[\mathbf{F}\mathbf{F}^T]\mathbf{L}^T + \mathsf{E}[\boldsymbol{\varepsilon}\mathbf{F}^T]\mathbf{L}^T + \mathbf{L}\mathsf{E}[\mathbf{F}\boldsymbol{\varepsilon}^T] + \mathsf{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \mathbf{L}\mathbf{L}^T + \boldsymbol{\Psi} \end{aligned} \tag{21}$$

and we have

$$\begin{aligned} \mathsf{Var}[r_i] &= \ell_{i1}^2 + \ell_{i2}^2 + ... + \ell_{iK}^2 + \psi_i = h_i^2 + \psi_i \\ \mathsf{Cov}[r_i, r_j] &= \ell_{i1}\ell_{j1} + \ell_{i2}\ell_{j2} + ... + \ell_{iK}\ell_{jK} \\ \mathsf{Cov}[r_i, F_j] &= \ell_{ij} = \sqrt{\lambda_j}e_{ij}. \end{aligned} \tag{22}$$

One method for estimating factor loadings and specific factors is the principal component method. The method is based on a spectral decomposition of the sample covariance matrix.

Proposition 3. *Let* $\mathbf{S}$ *have the eigenvalue-eigenvector pairs* $(\mathbf{e}_1, \lambda_1)$, $(\mathbf{e}_2, \lambda_2)$, ..., $(\mathbf{e}_N, \lambda_N)$ *where* $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N \geq 0$. *Then*

$$\begin{aligned}\mathbf{S} &= \lambda_1 \mathbf{e}_1 \mathbf{e}_1^T + \lambda_2 \mathbf{e}_2 \mathbf{e}_2^T + ... + \lambda_N \mathbf{e}_N \mathbf{e}_N^T \\ &= \begin{bmatrix} \sqrt{\lambda_1}\mathbf{e}_1 & \sqrt{\lambda_2}\mathbf{e}_1 & \ldots & \sqrt{\lambda_N}\mathbf{e}_N \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1}\mathbf{e}_1^T \\ \sqrt{\lambda_2}\mathbf{e}_2^T \\ \vdots \\ \sqrt{\lambda_N}\mathbf{e}_N^T \end{bmatrix}. \end{aligned} \tag{23}$$

Proof. See Johnson & Wichern (1992). □

We see that the decomposition in equation 23 fits exactly into equation 21 with the specific variances equal to zero, i.e. $\mathbf{\Psi} = \mathbf{0}$. We can therefore write equation 23 as

$$\mathbf{S} = \bar{\mathbf{L}}\bar{\mathbf{L}}^T, \tag{24}$$

where $\bar{\mathbf{L}}$ is an $N \times N$ matrix. We notice that the loadings on the jth factor are the coefficients in the jth principal component times a scale factor $\sqrt{\lambda_j}$. This is an exact representation, but we are however interested in describing the covariance structure using only a few common factors. We know from the previous section that this can be done by using only the first K principal components, so neglecting the contribution of the smallest eigenvalues we can write an approximation of equation 23 as

$$\begin{aligned}\mathbf{S} &\approx \lambda_1 \mathbf{e}_1 \mathbf{e}_1^T + \lambda_2 \mathbf{e}_2 \mathbf{e}_2^T + ... + \lambda_K \mathbf{e}_K \mathbf{e}_K^T \\ &= \begin{bmatrix} \sqrt{\lambda_1}\mathbf{e}_1 & \sqrt{\lambda_2}\mathbf{e}_1 & \ldots & \sqrt{\lambda_K}\mathbf{e}_K \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1}\mathbf{e}_1^T \\ \sqrt{\lambda_2}\mathbf{e}_2^T \\ \vdots \\ \sqrt{\lambda_K}\mathbf{e}_K^T \end{bmatrix} = \mathbf{L}\mathbf{L}^T, \end{aligned} \tag{25}$$

where $\mathbf{L}$ is an $N \times K$ matrix. If we also assume that the specific errors are significant, we can take them as the diagonal of $\mathbf{S} - \mathbf{L}\mathbf{L}^T$ and the approximation becomes

$$\mathbf{S} \approx \mathbf{L}\mathbf{L}^T + \mathbf{\Psi}, \tag{26}$$

where $\psi_{ii} = s_{ii} - h_i^2$. This is of extra relevance when analyzing the sample correlation matrix since it makes sure that diagonal elements in the correlation matrix still will be equal to one.

3.3.4 Choosing How Many Factors to Include

A commonly accepted rule of thumb, but with weak theoretical support, is that one should only use eigenvalues larger than one, i.e. the ones that explain at least $1/N$ percent of the total standardized variance. For our purpose it is however of interest to use a rule that is a bit more theoretically founded. We find such a rule in the relatively new field of econophysics and the work of Plerou et al. (2001), Plerou et al. (1999), Laloux et al. (1998) and Guhr (2001). The idea is to try to separate the real correlation from the estimation error by comparing the properties of the sample correlation matrix with known results for

a completely random correlation matrix.

Let $\tilde{\mathbf{C}}$ be defined as

$$\tilde{\mathbf{C}} = \frac{1}{T-1}\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T \tag{27}$$

where $\tilde{\mathbf{Z}}$ contains mutually uncorrelated time series with zero mean and unit variance from an empirical asset distribution. It can been shown that[2] when N and T tend to infinity such that $Q = T/N$ is fixed, the eigenvalue distribution $f_{RM}(\lambda)$ of a matrix such as $\tilde{\mathbf{C}}$ is given by

$$f_{RM}(\lambda) = \frac{Q}{2\pi}\frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda}, \tag{28}$$

where

$$\lambda_{min} \leq \lambda \leq \lambda_{max}$$

and the minimum and maximum eigenvalue bounds in turn are given by

$$\lambda_{min}^{max} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}. \tag{29}$$

On US stock data, comparing the predicted distribution $f_{RM}(\lambda)$ with the estimated probability density function (EDF) of the eigenvalues of $\tilde{\mathbf{C}}$, Plerou et al. (2001) find no significant deviations, while $f_{RM}(\lambda)$ compared with the EDF of $\mathbf{C}$ shows that about 2% of the eigenvalues differ significantly. These deviations from the remaining bulk of eigenvalues consequently imply real correlations. The by far largest deviation is (naturally) for the largest eigenvalue and Plerou et al. (2001) interpret it as the market factor. The other deviating eigenvalues are closely linked to specific regions and industries. Generally, however, simply looking at the deviation of the eigenvalue spectrum of $\mathbf{C}$ from $f_{RM}(\lambda)$ is not enough to conclude that $\mathbf{C}$ is dominated by noise and that most of the genuine correlations are to be found in the deviation from equation 29. However, Plerou et al. (2001) provide a very extensive examination of the statistical properties of $\mathbf{C}$ and find no evidence against such a conclusion. In fact, the deviating eigenvalues with corresponding eigenvectors seem to be quite stable through time, implying that true correlations are also stable through time. Due to the close relation between the principal components and the eigenvalue-eigenvector pairs, our new rule of thumb will thus be to only choose those principal components that are determined by the eigenvectors belonging to the eigenvalues that deviates significantly from the maximum eigenvalue prediction of equation 29.

3.4 Constrained Portfolio Optimization

Imposing constraints on the portfolio weights such as no short selling restrictions and upper bounds, may sometimes prove to reduce out-of-sample volatility of the investor's optimal portfolio. The no short selling restrictions can also be seen as a way of imposing a little more realism into the portfolio choice problem, since it is often very hard, or even impossible, for an ordinary investor to short sell an asset. However, there is a more mathematical way of understanding the no short selling constraints, as is done by Jagannathan & Ma (2002). No short selling constraints can be interpreted as a means of

[2] See e.g. Plerou et al. (2001) for references.

shrinking the largest variances and covariances, which cause the negative weights, towards more standard values, since it can be argued that these extreme estimates are those most likely caused by estimation error.

Denote the solution of

$$\min_{\mathbf{w}} \left\{ \frac{1}{2} \mathbf{w}^T \hat{\mathbf{\Sigma}} \mathbf{w} \quad | \quad \mathbf{w}^T \mathbf{1} = 1, \mathbf{w} \geq \mathbf{0} \right\} \tag{30}$$

as $\mathbf{w}^*(\hat{\mathbf{\Sigma}})$ where $\hat{\mathbf{\Sigma}}$ should be taken as an estimated covariance matrix, e.g. the sample covariance matrix or the single-index covariance matrix. Also, denote the lagrangian multiplier associated with non-negativity restriction i for this solution as η_i. It is now possible to construct a new covariance matrix $\tilde{\mathbf{\Sigma}}$ so that its *unconstrained* MVP is $\mathbf{w}^*(\hat{\mathbf{\Sigma}})$. Jagannathan & Ma (2002) derive an expression for this matrix as

$$\tilde{\mathbf{\Sigma}} = \hat{\mathbf{\Sigma}} - \begin{bmatrix} 2\eta_1 & \eta_1 + \eta_2 & \dots & \eta_1 + \eta_N \\ \eta_2 + \eta_1 & 2\eta_2 & \dots & \eta_2 + \eta_N \\ \vdots & \vdots & \ddots & \vdots \\ \eta_N + \eta_1 & \eta_N + \eta_2 & \dots & 2\eta_N \end{bmatrix}, \tag{31}$$

which additionally is also non-singular. The proof is left out here. We see that for every asset i for which the non-negativity restriction is binding, the covariances are reduced by $\eta_i + \eta_j$ (a positive positive number) and the variance is reduced by $2\eta_i$. Obviously, as the elements in row i of $\hat{\mathbf{\Sigma}}$ increases in magnitude, the more one wants to break the restrictions and the larger the lagrangian multipliers will be and the more the variance and covariances are reduced. Hence, the new covariance matrix $\tilde{\mathbf{\Sigma}}$ is constructed by shrinking the largest elements in the initial covariance matrix estimate towards more standard values. There are of course many covariance matrices that also have the MVP $\mathbf{w}^*(\hat{\mathbf{\Sigma}})$. Jagannathan & Ma (2002) however provide a proof for the case when the initial estimate is the sample covariance matrix, that $\tilde{\mathbf{\Sigma}}$ is the covariance matrix that lies closest to the sample covariance matrix in the sense that it is the *constrained* ML estimator. This proof is also left out here.

A similar discussion can be carried out in the case when also imposing upper bounds on the portfolio weights, but instead of shrinking large elements, it corresponds to enlarging small elements. Empirically however, upper bound constraints does not seem to have much effect on reducing out-of-sample volatility when lower bound constraints are already in place, see e.g. Jagannathan & Ma (2002).

4 Shrinkage

Shrinkage is a typical Stein estimator, which in turn is a Bayesian statistical procedure. Characteristic for Bayesian statistics is to assume a prior which should reduce the dependency on purely estimated parameters by representing some form of structure. The shrinkage estimator is an optimally weighted average between the sample covariance matrix and the prior. The idea is that the prior should become increasingly more important as the uncertainty of the sample estimate grows larger and that there exists an optimal

balance between estimation error and specification error. The prior can e.g. be a reasonable guess or the result of an assumption. We denote the weight assigned to the prior covariance matrix in the shrinkage estimator by α, $0 \leq \alpha \leq 1$, so that the larger α is, the more structure we impose. In statistical terms, α is referred to as shrinkage intensity and the prior can also be called shrinkage target. For a fairly general choice of prior, the shrinkage estimator will always be invertible. This section is based on the shrinkage estimator originally proposed by Ledoit & Wolf (2001).

4.1 The Shrinkage Estimator of Ledoit & Wolf (2001)

In this section the prior is the single-index covariance matrix of Sharpe (1963), as suggested by Ledoit & Wolf (2001). This can be interpreted as choosing between two extremes, an N-factor model and a one-factor model. We denote the estimated covariance matrix implied by the single-index market model of Sharpe (1963) as $\mathbf{F}$ and as $T \to \infty$, it is assumed to converge to $\mathbf{\Phi}$, for which we have $\mathbf{\Phi} \neq \mathbf{\Sigma}$. We write the resulting shrinkage estimator as

$$\alpha\mathbf{F} + (1-\alpha)\mathbf{S}. \tag{32}$$

It is also assumed that the asset returns are independent and identically distributed (iid) and that they have finite fourth moments. Furthermore, N is kept fixed while $T \to \infty$. The difficulty lies in finding the optimal shrinkage intensity α. For a fixed N, $\mathbf{S}$ will be consistent as $T \to \infty$, while $\mathbf{F}$ will not. Thus the shrinkage intensity should vanish asymptotically. The most important contribution of Ledoit & Wolf (2001) is the formulation of an objective function that does not depend upon the inverse of the covariance matrix, as opposed to all previous work. Consequently, this is the first shrinkage estimator that does not break down when $N \geq T$. The objective is quite natural. It is a quadratic measure of the distance between the true covariance matrix and the shrinkage estimator.

Definition 2. *The squared Frobenius norm of a $N \times N$ symmetric matrix $\mathbf{A}$ with entries a_{ij}, $i, j = 1, 2, ..., N$, and eigenvalues λ_i, $i = 1, 2, ...N$, is defined by*

$$\|\mathbf{A}\|_F^2 = \text{tr}(\mathbf{A}^2) = \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^2 = \sum_{i=1}^{N} \lambda_i^2. \tag{33}$$

Using the Frobenious norm, Ledoit & Wolf (2001) defines the quadratic loss function between the shrinkage estimator and the true covariance matrix as

$$L(\alpha) = \|\alpha\mathbf{F} + (1-\alpha)\mathbf{S} - \mathbf{\Sigma}\|_F^2. \tag{34}$$

Then, simply setting the first derivative of the expected value of $L(\alpha)$ equal to zero yields the optimal shrinkage intensity as

$$\alpha^* = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} \mathsf{Var}[s_{ij}] - \mathsf{Cov}[f_{ij}, s_{ij}]}{\sum_{i=1}^{N}\sum_{j=1}^{N} (\mathsf{Var}[f_{ij} - s_{ij}] + (\phi_{ij} - \sigma_{ij})^2)}. \tag{35}$$

Proposition 4. *Let π denote the sum of asymptotic variances of the entries in the sample covariance matrix scaled by $\sqrt{T}$, i.e. $\pi = \sum_{i=1}^{N}\sum_{j=1}^{N} \mathsf{AsyVar}[\sqrt{T}s_{ij}]$. Similarly let ρ denote the sum of asymptotic covariances of the entries in the single-index*

covariance matrix with entries in the sample covariance matrix, also scaled by $\sqrt{T}$*, i.e.* $\rho = \sum_{i=1}^{N}\sum_{j=1}^{N} \mathsf{AsyCov}[\sqrt{T}f_{ij}, \sqrt{T}s_{ij}]$*. Finally let* γ *be a measure of the misspecification of the single-index covariance matrix, i.e.* $\gamma = \sum_{i=1}^{N}\sum_{j=1}^{N}(\phi_{ij} - \sigma_{ij})^2$*. Then the optimal shrinkage intensity satisfies*

$$\alpha^* = \frac{1}{T}\frac{\pi - \rho}{\gamma} + O\left(\frac{1}{T^2}\right). \tag{36}$$

Proof. See Ledoit & Wolf (2001). □

Determined by the constant

$$\kappa = \frac{\pi - \rho}{\gamma} \tag{37}$$

we can thus write the asymptotically optimal shrinkage estimator as

$$\frac{\kappa}{T}\mathbf{F} + \left(1 - \frac{\kappa}{T}\right)\mathbf{S}. \tag{38}$$

We note that, just as Ledoit & Wolf (2001) points out, through the calculations in this section we have never actually used the fact that $\mathbf{F}$ is the single-index covariance matrix, only that it is an asymptotically biased estimator. As a matter of fact, equations 38 & 37 extend to (almost) any biased covariance matrix estimate of choice.

4.2 Estimating the Optimal Shrinkage Intensity

Now, it is important to realize that the parameters π, ρ and γ are not directly observable, so consequently they have to be estimated. To start, the parameters are decomposed so that $\pi = \sum_{i=1}^{N}\sum_{j=1}^{N}\pi_{ij}$ where $\pi_{ij} = \mathsf{AsyVar}[\sqrt{T}s_{ij}]$, $\rho = \sum_{i=1}^{N}\sum_{j=1}^{N}\rho_{ij}$ where $\rho_{ij} = \mathsf{AsyCov}[\sqrt{T}f_{ij}, \sqrt{T}s_{ij}]$ and $\gamma = \sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij}$ where $\gamma_{ij} = (\phi_{ij} - \sigma_{ij})^2$. Also, let s_{iM} be the sample covariance between the return of asset i and the market. The consistent estimators of π, ρ and γ are then proposed by Ledoit & Wolf (2001) in the following lemmas.

Lemma 1. *A consistent estimator for* π_{ij} *is given by*

$$\hat{\pi}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\left((r_{it} - \bar{r}_i)(r_{jt} - \bar{r}_j) - s_{ij}\right)^2. \tag{39}$$

Proof. See Ledoit & Wolf (2001). □

Lemma 2. *A consistent estimator for* ρ_{ii} *is given by* $\hat{\rho}_{ii} = \hat{\pi}_{ii}$*. For the off-diagonal elements a consistent estimator is given by* $\hat{\rho}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\hat{\rho}_{ijt}$*, where*

$$\hat{\rho}_{ijt} = \frac{s_{jM}s_{MM}x_{it} + s_{iM}s_{MM}x_{jt} - s_{iM}s_{jM}x_{Mt}}{s_{MM}^2}x_{Mt}x_{it}x_{jt} - f_{ij}s_{ij} \tag{40}$$

and

$$\begin{aligned} x_{Mt} &= r_{Mt} - \bar{r}_M \\ x_{it} &= r_{it} - \bar{r}_i \\ x_{Mt} &= r_{jt} - \bar{r}_j. \end{aligned}$$

Proof. See Ledoit & Wolf (2001). □

Lemma 3. *A consistent estimator for* $\gamma_{ij} = (\phi_{ij} - \sigma_{ij})^2$ *is given by its sample equivalent* $\hat{\gamma}_{ij} = (f_{ij} - s_{ij})^2$.

Proof. See Ledoit & Wolf (2001). □

Thus, the consistent estimator for the optimal shrinkage constant can be written as

$$\hat{\kappa} = \frac{\hat{\pi} - \hat{\rho}}{\hat{\gamma}} \tag{41}$$

and consequently the asymptotically optimal shrinkage estimator towards the single-index covariance matrix is

$$\frac{\hat{\kappa}}{T}\mathbf{F} + \left(1 - \frac{\hat{\kappa}}{T}\right)\mathbf{S}. \tag{42}$$

4.3 Shrinkage Towards a Principal Component Model

We will now extend the estimator of Ledoit & Wolf (2001) to the case when the shrinkage target is a K factor principal component covariance matrix. Inspired mostly by the results of Plerou et al. (2001), we extract the principal components from the sample correlation matrix rather than directly from the sample covariance matrix itself. The intuition in this is that since the largest eigenvalues with belonging eigenvectors may sometimes be attributed to either market or industry phenomena, we want to examine whether or not we can improve the estimator proposed by Ledoit & Wolf (2001) by in a (admittedly very mathematical) way adding industry effects. Another more statistical line of logic can be seen by separating the sample correlation matrix into two parts. One that we are pretty sure describes true correlations such as the market and industry effects and one that is mostly influenced by estimation noise, but still carries some degree of information. Plerou et al. (2001) suggest that we should throw away the second part completely and disregard any information it may or may not contain. What we would hope to accomplish in this section is that rather than removing it completely and losing potential information, we merely reduce its effect in an (asymptotically) optimal way. In this way the part of the sample correlation matrix that we are more confident in becomes more prominent. This is easily seen using equations 23 & 25. Since the K largest eigenvalues and their corresponding eigenvalues are contained in both the sample estimate and the shrinkage target, they will receive the weight $(1 - \alpha) + \alpha = 1$. The influence of the last $N - K$ eigenvalues on the other hand, will be reduced since the weight they receive is only $1 - \alpha \leq 1$.

If we take a look at the parameters π, ρ and γ we note that, just as Ledoit & Wolf (2001) again points out, the formulas for $\hat{\pi}$ and $\hat{\gamma}$ does not change with the choice of shrinkage target. The formula for $\hat{\rho}$, however, needs some modification.

We denote the final principal component covariance matrix by $\mathbf{P}_K$ and we have

$$\mathbf{P}_K = \mathbf{V}^{1/2}\left(\mathbf{L}\mathbf{L}^T + \mathbf{\Psi}\right)\mathbf{V}^{1/2}, \tag{43}$$

where K is the number of eigenvalues of the sample correlation matrix that deviates significantly from the maximum eigenvalue prediction of equation 29. It follows from the

definition of $\mathbf{\Psi}$ in equation 26 that the diagonal of $\mathbf{P}_K$ is equal to the diagonal of the sample covariance matrix, i.e. $p_{K,ii} = s_{ii}$. For the off-diagonal elements we use the fact that the coefficients ℓ_{ij} are equal to $\sqrt{\lambda_j}e_{ij}$. Combining this with proposition 2 we get the off-diagonal elements of $\mathbf{P}_K$ as

$$\begin{aligned}
p_{K,ij} &= s_i s_j \sum_{k=1}^{K} \ell_{ik}\ell_{jk} = s_i s_j \sum_{k=1}^{K} \sqrt{\lambda_k}e_{ik}\sqrt{\lambda_k}e_{jk} \\
&= s_i s_j \sum_{k=1}^{K} \mathsf{Corr}[z_i, f_k]\mathsf{Corr}[z_j, f_k] \\
&= s_i s_j \sum_{k=1}^{K} \frac{\mathsf{Std}[z_i]\mathsf{Corr}[z_i, f_k]\mathsf{Std}[f_k]}{\mathsf{Std}[z_i]\mathsf{Std}[f_k]} \frac{\mathsf{Std}[z_j]\mathsf{Corr}[z_j, f_k]\mathsf{Std}[f_k]}{\mathsf{Std}[z_j]\mathsf{Std}[f_k]} \\
&= s_i s_j \sum_{k=1}^{K} \frac{\mathsf{Cov}[z_i, f_k]\mathsf{Cov}[z_j, f_k]}{\mathsf{Var}[f_k]} = \sum_{k=1}^{K} \frac{s_i\mathsf{Cov}[z_i, f_k]s_j\mathsf{Cov}[z_j, f_k]}{\mathsf{Var}[f_k]} \\
&= \sum_{k=1}^{K} \frac{\mathsf{Cov}[r_i, f_k]\mathsf{Cov}[r_j, f_k]}{\mathsf{Var}[f_k]},
\end{aligned} \tag{44}$$

where, for the sake of readability, $\mathsf{Var}[\cdot]$, $\mathsf{Std}[\cdot]$, $\mathsf{Cov}[\cdot,\cdot]$ and $\mathsf{Corr}[\cdot,\cdot]$ should be taken as the sample estimates. Slightly modifying lemma 2, we now propose the lemma for consistently estimating ρ when shrinking towards a K factor principal component covariance matrix.

Let s_{if_k} denote the sample covariance between the return of asset i and the kth principal component of the sample correlation matrix.

Lemma 4. *A consistent estimator for ρ_{ii} is given by $\hat{\rho}_{ii} = \hat{\pi}_{ii}$. For the off-diagonal elements a consistent estimator is given by $\hat{\rho}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\hat{\rho}_{ijt}$, where*

$$\hat{\rho}_{ijt} = \sum_{k=1}^{K}\left[\frac{s_{jf_k}s_{f_kf_k}x_{it} + s_{if_k}s_{f_kf_k}x_{jt} - s_{if_k}s_{jf_k}f_{kt}}{s^2_{f_kf_k}} f_{kt}x_{it}x_{jt}\right] - p_{K,ij}s_{ij} \tag{45}$$

and we use

$$\begin{aligned}
x_{jt} &= (r_{jt} - \bar{r}_j) \\
x_{it} &= (r_{it} - \bar{r}_i) \\
\bar{f_k} &= 0, k = 1, 2, ..., K.
\end{aligned}$$

Proof. On the diagonal $p_{K,ii} = s_{ii}$ and hence $\hat{\rho}_{ii} = \hat{\pi}_{ii}$. For the off-diagonal elements we have $p_{K,ij} = \sum_{k=1}^{K} s_{if_k}s_{jf_k}/s_{f_kf_k}$. A first order Taylor series expansion then yields

$$\begin{aligned}
\rho_{ij} &= \mathsf{AsyCov}[\sqrt{T}\sum_{k=1}^{K}\frac{s_{if_k}s_{jf_k}}{s_{f_kf_k}}, \sqrt{T}s_{ij}] = \sum_{k=1}^{K}\mathsf{AsyCov}[\sqrt{T}\frac{s_{if_k}s_{jf_k}}{s_{f_kf_k}}, \sqrt{T}s_{ij}] \\
&\approx \sum_{k=1}^{K}\left[\frac{\sigma_{jf_k}}{\sigma_{f_kf_k}}\mathsf{AsyCov}[\sqrt{T}s_{if_k}, \sqrt{T}s_{ij}] + \frac{\sigma_{if_k}}{\sigma_{f_kf_k}}\mathsf{AsyCov}[\sqrt{T}s_{jf_k}, \sqrt{T}s_{ij}]\right. \\
&\left. - \frac{\sigma_{if_k}\sigma_{jf_k}}{\sigma^2_{f_kf_k}}\mathsf{AsyCov}[\sqrt{T}s_{f_kf_k}, \sqrt{T}s_{ij}]\right].
\end{aligned} \tag{46}$$

Consistent estimators for σ_{if_k}, σ_{jf_k} and $\sigma_{f_kf_k}$, $k = 1, 2, .., K$, are the sample estimates s_{if_k}, s_{jf_k} and $s_{f_kf_k}$, $k = 1, 2, .., K$. The *usual* estimator for $\mathsf{AsyCov}[\sqrt{T}s_{qf_k}, \sqrt{T}s_{ij}]$ where $q = i, j, f_k$, $k = 1, 2, .., K$, is

$$\frac{1}{T}\sum_{t=1}^{T}\left((r_{qt} - \bar{r}_q)(f_{kt} - \bar{f}_k) - s_{qf_k}\right)\left((r_{it} - \bar{r}_i)(r_{jt} - \bar{r}_j) - s_{ij}\right). \tag{47}$$

Putting this into equation 46 together with some manipulation yields equation 45. □

Finally we can write the asymptotically optimal shrinkage estimator towards a K factor principal component covariance matrix as

$$\frac{\hat{\kappa}}{T}\mathbf{P}_K + \left(1 - \frac{\hat{\kappa}}{T}\right)\mathbf{S}, \tag{48}$$

where $\hat{\kappa}$ is given by

$$\hat{\kappa} = \frac{\hat{\pi} - \hat{\rho}}{\hat{\gamma}} \tag{49}$$

and $\hat{\pi}$, $\hat{\rho}$ and $\hat{\gamma}$ are estimated using lemmas 1, 4 & 3 respectively.

A nice property of this estimator compared to the one of Ledoit & Wolf (2001), is that it does not depend on an exogenously given covariance matrix, such as the one from the single-index model. We simply re-use some of the building blocks of the sample covariance matrix. Our estimator may therefore be useful in applications beyond the economical context, where a clear and natural factor such as the market is not present.

4.4 An Alternative Shrinkage Estimator

Jagannathan & Ma (2000) argue that different plausible covariance matrix estimators based on different assumptions make errors in different directions. One should therefore avoid putting all ones belief in one single estimator and try to diversify ones beliefs by using a portfolio of estimators where the errors hopefully cancel out. The intuition behind this is very similar to the one behind the shrinkage estimator. Jagannathan & Ma (2000) recommend a portfolio consisting of the sample covariance matrix, the single-index covariance matrix and a matrix consisting of the diagonal part of the sample covariance matrix. In other words an N factor model, a *one* factor model and a *zero* factor model. Jagannathan & Ma (2000) argue that since little is known about the covariance structure of the estimation errors of the different estimators, equal portfolio weights is the safest approach. Nothing is said to further motivate the choice of estimators.

5 Empirical Results

5.1 Data

The data consists of monthly returns on Swedish stocks from the year 1977 to 1997[3] corrected for dividends and capital changes such as splits etc. The stocks are those from

[3]Courtesy of Hossein Asgharian, Department of Economics at the School of Economics and Management, University of Lund.

the Swedish *A1-listan* (excluding banks and financial firms) and together they make up more than 95% of the total value of the market.

5.2 Evaluation Methodology

To test the performances of the different estimators we follow Ledoit & Wolf (2001) and Chan et al. (1999). Starting 1st January 1985 we use historical data from eight years back in order to estimate the covariance matrices. This is the in-sample period. We then form the respective MVP's and hold them for six months. This is the out-of-sample period. At the end of the six months the covariance matrices are re-estimated, again using data eight years back in time and the portfolios are re-balanced accordingly and held for another six months. This process is repeated until 1st July 1997, which is the last re-balancing point. The interesting metric is the out-of-sample standard deviation of this portfolio strategy. It is important to add that we allow ourselves to disregard stocks whose returns do not cover the entire out-of-sample period. Obviously this is not an option for an actual investor, but we motivate it by the fact that it further increases the stress on the covariance matrix. A valid question is why we only use the MVP as opposed to all other MV efficient portfolios. The reason is quite clearly that the impact of the covariance matrix on the portfolio is maximized when the portfolio depends solely on the covariance matrix, as the MVP does. All other portfolios also need an estimate of the expected returns, which is not a trivial matter.

Our situation differ in two ways from the one of Ledoit & Wolf (2001). Firstly, we only have access to 20 years of stock returns whereas Ledoit & Wolf (2001) uses 33 years altogether. The implications of this is that we have to shorten the in-sample and the out-of-sample periods. Ledoit & Wolf (2001) use 120 months and twelve months respectively. We use 96 months and six months. In this way we get 26 re-balancing points. We will discuss this choice in more detail in section 5.7. The second difference is that although our sample represents a large portion of the Swedish market, it is small in comparison to the sample of stocks used by Ledoit & Wolf (2001) which is in the range of 909 to 1,314. The number of stocks in our portfolios vary between 44 and 61. However, since this paper is concerned with covariance matrix estimation for portfolio selection on the Swedish market, this difference is somewhat irrelevant. Worth mentioning is also that Jagannathan & Ma (2000) consider 30-56 stocks in their studies and that Plerou et al. (2001) use 422-1000 stocks, but on very short time scales (down to 30-min returns).

5.3 Covariance Matrix Estimators to be Considered

The sample covariance matrix An important note is that given our sample size, we will not run into the problem of it not being invertible. Thus we do not have to use the pseudo-inverse to invert it and we can make a fair comparison to the other covariance matrix estimators. We also include a diagonal model which is a matrix consisting of only the diagonal elements of the sample covariance matrix.

The single-index market model This is the covariance matrix implied by the market model of Sharpe (1963), defined in equation 10. For the market index we consider two

alternatives. (1) a simple average over all stocks, as in Ledoit & Wolf (2001) and (2) a value weighted market index, as in the work of Jagannathan & Ma (2000).

The K factor principal component model This is the covariance matrix estimator defined in equation 43. In the work of Ledoit & Wolf (2001), a five factor estimator is used. They are however extracted from the actual sample covariance matrix and not from the sample correlation matrix, as we do. Plerou et al. (2001) use thirteen principal components, although they do not call it a principal component model.

Shrinkage towards the single-index market model This is the covariance matrix estimator defined in equation 42, as proposed by Ledoit & Wolf (2001). Again, we will try both alternatives for the market index when computing the single-index covariance matrix.

Shrinkage towards a K factor principal component model This is the covariance matrix estimator proposed in equation 48.

A portfolio of estimators This is the covariance matrix estimator covered in section 4.4. The estimator, as proposed by Jagannathan & Ma (2000), is an equally weighted portfolio of the sample covariance matrix, a matrix containing only the diagonal elements of the sample covariance matrix and the single-index covariance matrix computed from a value weighted index. We will try to modify it in two ways. (1) by replacing the value weighted market index with a simple average over all stocks and (2) by replacing the single-index covariance matrix with a K factor principal component covariance matrix.

Constrained portfolio optimization We will also see if there is any reduction in out-of-sample standard deviation as a result of imposing no short selling restrictions.

As a benchmark, a naive investment strategy of investing an equal amount in all stocks is also included.

5.4 Analysis of the Out-of-Sample Standard Deviations

Reported in table 1 are the out-of-sample standard deviations for the various minimum variance portfolios calculated from the above mentioned covariance matrix estimators. The standard deviations are annualized through multiplication by $\sqrt{12}$ and expressed in percents. Unconstrained standard deviation means that there are no restrictions beside the budget restriction. Constrained standard deviation means that we have imposed the no short selling restrictions. The so called standard errors on the estimates are obtained by delaying the start of the evaluation procedure by one month at a time and recording the results until we reach July 1985 and then taking the standard deviation of the outcomes for each estimator. All single-index covariance matrices are calculated from a value weighted market index. This because the single-index covariance matrix itself showed improvement when we calculated it from the value weighted market index[4] and it is first and foremost

[4] An unconstrained out-of-sample standard deviation of 18.35 vs. 18.64.

	Unconstrained Standard Deviation	Constrained Standard Deviation
Sample Estimate	26.89 (3.46)	18.90 (0.13)
Equally Weighted Portfolio	24.50 (0.17)	24.50 (0.17)
Diagonal Model	22.41 (0.14)	22.41 (0.14)
Single-Index Market Model	18.35 (0.18)	18.30 (0.08)
Single Factor PC Model	18.06 (0.20)	18.01 (0.08)
Shrinkage To Market	17.80 (0.23)	18.32 (0.10)
Average of Sample, Diagonal and Market	17.58 (0.13)	18.32 (0.06)
Shrinkage To Single PC	17.55 (0.24)	18.20 (0.09)
Average of Sample, Diagonal and Single PC	17.40 (0.13)	18.22 (0.06)

Table 1: Unconstrained an constrained annualized out-of-sample standard deviations expressed in percents. The standard errors on the estimates are given in parenthesis.

a good market model we are after. The shrinkage estimator and the portfolio of estimators actually performed slightly better when the market index was taken as the simple average[5], but from a strictly economical point of view we still feel it is safer to choose the value weighted market index. We also found that using only one principal component in the shrinkage estimator and the portfolio of estimators performs by far the best, why we write *single* principal component covariance matrix as opposed to a K factor principal component covariance matrix. This will be discussed in the next section.

For the unconstrained estimates we see that the sample covariance matrix performs the worst, followed by the naive investment strategy. Apparently the sample covariance matrix contains so much estimation error that we even are better off by simply investing in an equally weighted portfolio without concerning ourselves with neither variances nor covariances. The large influence of estimation error in the sample covariance matrix also shows itself in the sense that the standard error on the performance is by far the largest, implying that it is also very sensitive to small changes in data. In fact, the only standard error that is large is the one of the sample covariance matrix, all other standard errors are quite small. The estimator where we have replaced the single-index covariance matrix in the portfolio of estimators with the single principal component covariance matrix performs the best. It is also clear that imposing some form of structure is very beneficial. Together with the sample covariance matrix, the single-index covariance matrix and the single principal component covariance matrix are the only estimators that performs better when we impose the no short selling restrictions. The sample covariance estimator is the one that benefits the most. This comes as no surprise since this estimator is the one that is influenced the most by estimation error and thus should gain the most from the shrinking effect that the constraints has on large variances and covariances. The improvement for the single-index covariance matrix and the single principal component covariance matrix is minor in comparison. For all other estimators the performance is reduced by

[5]Unconstrained out-of-sample standard deviations of 17.80 vs. 17.66 and 17.58 vs. 17.42.

the constraints. This is reassuring since if we have a good covariance matrix estimate to begin with, the performance should not be improved when imposing further constraints, it should at best be unaffected. Another observation we can make about the constrained estimators is that the performance of the last six estimators in table 1 is about the same, perhaps excluding the single principal component covariance matrix. What is interesting is that this similarity does not seem to reflect itself in the individual portfolio weights. Although the weights show similar overall movements, they are still somewhat spread out. Thus, even though the portfolio weights are sensitive to differences in the covariance matrix, this does not have a lot of impact on the actual out-of-sample performance. This can be interpreted as if the multi-dimensional function surface that we optimize over is quite flat.

It is interesting that both the shrinkage estimator and the portfolio estimator is improved when we replace the single-index covariance matrix with the single principal component covariance matrix. On their own, the single principal component covariance matrix is also better than the single-index covariance matrix. We can interpret this in the following way. The first principal component is given by the first eigenvector of the sample correlation matrix and this eigenvector can in turn be re-scaled and interpreted as a portfolio. If we take a look at its elements we find that they are all positive and about equally large (see figure 1). Furthermore, there are no particular groupings of the stocks. It is therefore a factor common to all stocks. This is what we expected and we interpret it as the market factor. We conclude therefore that, compared to the simple average used by Ledoit & Wolf (2001), the first eigenvector might be a better representative of the prevailing market portfolio within the available universe of stocks. However, although the single factor principal component covariance matrix is constructed through the linear combination of the *original* returns that explain the market structure in terms of their correlations, we must realize that the actual principal component cannot really be interpreted as a market index. Because we extracted it from the sample correlation matrix, the principal component is a linear combination of the standardized returns and not the real ones. The single principal component covariance matrix can therefore not be interpreted as the result of a regression on a market index. Thus, to test if it can really serve as a market portfolio, we simply take the eigenvector, re-scale it and use it to calculate a market index for the original variables. Using it to estimate the single-index covariance matrix yields an unconstrained out-of-sample standard deviation of 18.38, which is a clear improvement. We also plug it into the shrinkage estimator and the portfolio of estimators and the results are 17.59 and 17.51 respectively, which both are improvements. If we allow ourselves to be a little bit bold, there is also another way of looking at the situation. If we accept the fact that the first eigenvector may be a good estimate of the prevalent market portfolio, the principal factor solution may simply be a better way of estimating the market model covariance matrix than running a simple linear regression. In this way it may also be better than the actual value weighted index.

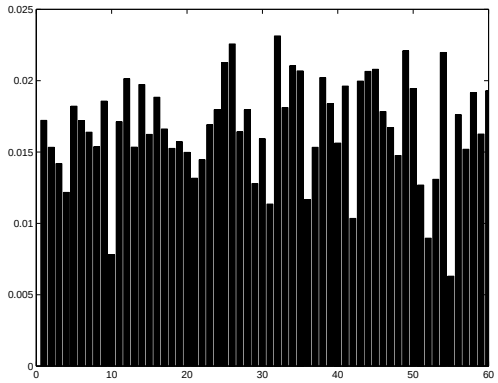

Figure 1: The elements in the re-scaled eigenvector belonging to the largest eigenvalue. The eigenvector is from the sample correlation matrix in the last in-sample period of our test.

5.5 Analysis of Our Shrinkage Estimator

It is a bit surprising that it does not seem to be optimal to use K's larger than one in the shrinkage estimator towards a principal component covariance matrix and the portfolio of estimators. This implies that there are no significant industry factors explaining covariances on the Swedish market, at least not in the sense of principal components. What we have for $K = 1$ is thus still a kind of market model. This absence of industry effects is nicely illustrated by looking at the random matrix prediction for the maximum eigenvalue bound, as defined in equation 29. For $N = 60$ and $T = 96$ we calculate $Q = T/N = 1.6$ and thus the prediction for the maximum eigenvalue is $\lambda_{max} = 3.21$. Plerou et al. (2001) found that several eigenvalues deviated significantly from this prediction[6], we do not. In figure 2 we present the EDF of the eigenvalues. The first eigenvalue of the last sample correlation matrix in our evaluation equals 23.47. That is a strong deviation, but the second eigenvalue, however, equals 3.50. That hardly differs from the prediction at all. In figure 3 we plot the first and the second eigenvalues for all 26 re-balancing points together with the maximum eigenvalue prediction of equation 29. What we see is that the largest eigenvalue always deviates significantly from the prediction, while the second largest eigenvalue actually lies under the prediction most of the time. It is therefore close at hand to suspect that all eigenvalues except the largest one are more or less dominated by estimation noise. Nothing can therefore be gained by ever including them in the prior.

If we also compare our study with the study of Plerou et al. (2001) by computing the percentage of eigenvalues used to construct the principal component covariance matrix, it turns out that we use almost the same percentage. We use $100 \cdot (1/60) = 1.67\%$ and Plerou et al. (2001) use $100 \cdot (13/1000) = 1.30\%$. One main difference between our two studies is the number of stocks considered and this is obviously a result of this. Hence, one possible reason as to why we do not detect any industry factors may simply be that

[6] Their prediction was 1.94.

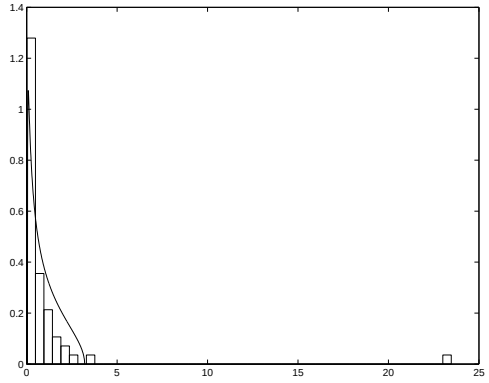

Figure 2: The EDF for the eigenvalues of the sample correlation matrix in the last in-sample window together with the theoretical distribution (smooth curve).

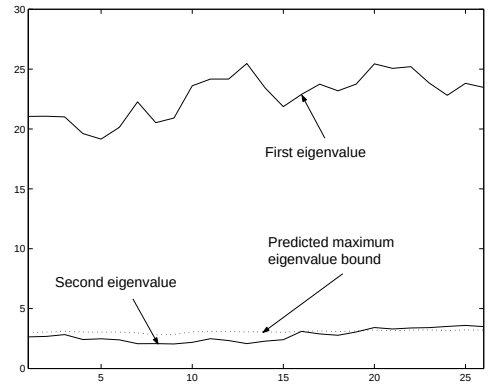

Figure 3: The first and second eigenvalue throughout the 26 re-balancing points plotted against the maximum eigenvalue prediction for a random correlation matrix.

we do not have a large enough sample for the effects to appear, at least not in a clear and stable way.

5.6 Shrinkage Intensities

In figure 4 we show how the optimal shrinkage intensities for the single-index covariance matrix and the single principal component covariance matrix have behaved over the 26 re-balancing points. We can see that they indeed behave rather like constants over time and that they tend to vary around 0.7[7]. This can be interpreted as if there is about twice as much estimation error in the sample covariance matrix as there is bias in the two shrinkage targets. It is interesting that the shrinkage intensity for the single principal component covariance matrix is generally larger than the one for the single-index covariance matrix. An explanation might be that there is a lesser amount of bias in single principal compo-

[7]The mean values are 0.65 and 0.69 respectively.

nent covariance matrix than in the single-index covariance matrix. That they are similar in their behavior again hints at that the single principal components covariance matrix also is a kind of market model.

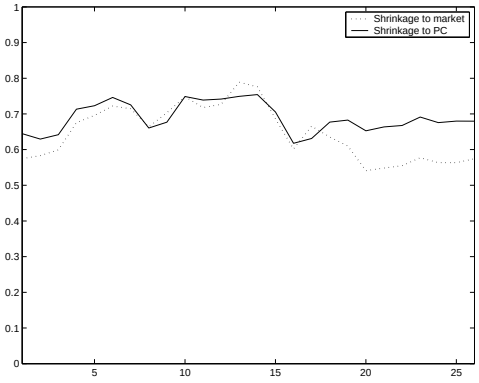

Figure 4: Optimal shrinkage intensity estimates, $\hat{\kappa}/T$, throughout the 26 re-balancing points for shrinkage towards the single-index market covariance matrix and shrinkage towards the single principal component covariance matrix.

But are the intensities really optimal? Jagannathan & Ma (2002) argue that the answer is *no*. To test some of the validity of this statement, we examine how the two shrinkage estimators perform when the intensities are either (1) completely random[8] or (2) equal to 0.5. For case (1) we find that the annualized out-of-sample standard deviations are in the range 18-20. For case (2) that the annualized out-of-sample standard deviation is 17.97 for shrinkage towards the single-index covariance matrix and 17.87 for shrinkage towards the single principal component covariance matrix. Thus, unlike Jagannathan & Ma (2002), we find that we have little reason to doubt that the shrinkage intensities we have estimated are in fact the best guesses we can produce without looking out-of-sample.

5.7 The Estimation Window

Amongst all talk about optimal estimators, there are still two parameters that usually are chosen very much arbitrarily, namely the length of the in-sample and out-of-sample periods. We have not yet really motivated our choices other than that we had to shorten the out-of-sample length in order to extract a sufficient number of covariance matrices to evaluate from our relatively small sample. It is obvious that we form twice as many portfolios when we use six month out-of-sample windows compared to using twelve months (as in the case of Ledoit & Wolf (2001) and Jagannathan & Ma (2000)). Besides from being more or less a necessity, this choice seems rather innocent. But to shorten the in-sample, or estimation window, (relative Ledoit & Wolf (2001) and Jagannathan & Ma (2000)) from 120 months to 96 months is a bit more questionable. Of course by doing so,

[8] Uniformly distributed between zero and one.

we increase the number of out-of-sample periods from 22 to 26[9], but this is not that large a difference. Looking at the situation from another angle, there should intuitively be an optimal length of the estimation window. If we use a too large window we run the risk of using information that is too old and hence the quality of our estimates will reduced. On the other hand, if we use too little data, our estimates will become unstable. This discussion is completely left out in the literature we have come across. Without further motivation, either 60 months or 120 months is normally chosen. We would however like to argue that this choice is also important, since all improvements that can be made on the covariance matrix estimates are important. Especially so in our sample, since the results in table 1 seem to be rather sensitive to changes in the length of the estimation window. This comes as somewhat of a surprise since no such behavior is prevalent in the work of Jagannathan & Ma (2000). Nor is it mentioned in Ledoit & Wolf (2001). Whether it is a result of our small sample size, differences between Swedish and US stock returns or choosing sub-optimal estimation windows is hard to say without further studies. Nevertheless, to test the effect of the estimation window size, we compare the results from using 120, 96, 72 and 60 month estimation windows[10]. As expected, the sample covariance matrix favors a lot of data, mainly since it becomes quite ill-conditioned for 72 and 60 months, while the diagonal model is virtually unaffected. The results for the single-index covariance matrix and the single principal component covariance matrix are also quite stable and besides that the single principal component model shows an out-of-sample standard deviation of 18.06 for 120 months vs. 18.13 for 96 months, they both show the best out-of-sample performance for the 96 month estimation window. The remaining four estimators all show significantly better out-of-sample performance for 96 months than they do for 120, 72 and 60 months. Thus, since roughly all estimators perform the best when we use 96 month estimation windows, this seems to be (at least empirically) the optimal length. Our motivation for choosing 96 months is then that it is reasonable to make the comparison between the estimators when they all perform at their best. Furthermore, since we do not work with the same data set as Jagannathan & Ma (2000) and Ledoit & Wolf (2001), the numerical values are not directly comparable anyway and we can take this liberty.

5.8 Portfolio Weights

In table 2 we present the maximum and minimum portfolio weights averaged over the 26 out-of-sample periods. Firstly, we note that besides being good estimators, the two portfolio estimators are also those that on average provide us with the most attractive portfolios in the sense that they require the least amount of short selling. Short selling is, as we mentioned earlier, sometimes hard to carry out in practice for most investors. Secondly, a pairwise similarity between the single-index covariance matrix and the single principal component covariance matrix, between the two shrinkage estimators and between the two portfolio estimators is evident. In fact, the six last estimators all show very similar values. Could it be that the difference between the portfolios we form is not that large?

[9] Also, there is on average eight more assets in the portfolios when we use the 96 month estimation window as there is for the 120 month estimation window.

[10] For all estimation window lengths, we begin estimation 1st January 1987 so that we always form 22 portfolios.

	Max Weight	Min Weight	Short Interest
Sample Estimate	34.94	-41.09	229.33
Diagonal Model	3.63	0.35	0
Single-Index Market Model	16.14	-15.05	67.98
Single Factor PC Model	16.18	-14.64	70.37
Shrinkage To Market	16.95	-16.67	71.78
Average of Sample, Diagonal and Market	15.56	-7.43	42.20
Shrinkage To Single PC	17.15	-16.11	72.86
Average of Sample, Diagonal and Single PC	15.62	-7.35	43.85

Table 2: The average maximum and minimum portfolio weights together with the average amount of short interest, which essentially is the sum of all negative portfolio weights. All values are given in percents.

And if so, when do they separate? As a measure of the difference between the portfolios in a certain month, we use the variance of the six out-of-sample portfolio return observations of that month. We present the result in figure 5. What is interesting is that the largest differences occurs around 1992. This is a well known turbulent period in the Swedish (and international) economy[11]. The conclusion is therefore that the shrinkage estimators and portfolio estimators are better equipped for handling such extreme periods of time[12]. This is also nicely shown in figure 6 where we have plotted the cumulative sum of the de-meaned and squared out-of-sample realizations for the last six estimators in table 2. As expected, the slopes show a clear divergence between 1991 and 1995, i.e. it is this period that causes most of the difference in out-of-sample performance between the estimators. There is also some divergence between 1985 to 1988. As a matter of fact, the market seems to have two rather clear states in our sample. The slope between 1985 to 1988 and 1991 to 1995 is about the same. This can be interpreted as a risky state. Meanwhile, the slope between 1988 to 1991 and 1995 to 1998 is about the same. This can be interpreted as a normal state. To illustrate this further, we re-estimate table 1 for the filtered sample where we have removed the years 1991, 1992, 1993 and 1994. The result is presented in table 3. Focusing on the last six estimators, we see that the unconstrained performance of the two portfolio estimators (especially) and the two shrinkage estimators is relatively unchanged, while the performance of the single-index covariance matrix and the single principal component covariance matrix shows clear improvement. What has also happened is that the spread of the standard errors on the estimates have decreased compared to those in table 1. Hence, under normal market conditions the performance of the last six estimators is resonably similar, but as we deviate from normal conditions the only estimators showing a stable level of performance are the portfolio estimators and the shrinkage estimators. If we consider the constrained estimators, we see that the difference from the unconstrained estimators

[11] It was e.g. in this period (September 1992) that the Swedish central bank raised the marginal lending rate to 500%.

[12] We here mean events that span over longer periods of time and not events such as those in October 1987 (Black Monday) which is a singular event that effected all estimators exactly the same.

is not as large as it was when we used the entire sample. The importance of being able to short sell stocks thus seems to be of greater importance when the market is in risky states than when it is in a normal state. We also see that when the market is in this normal state, the constrained sample covariance matrix performs rather well.

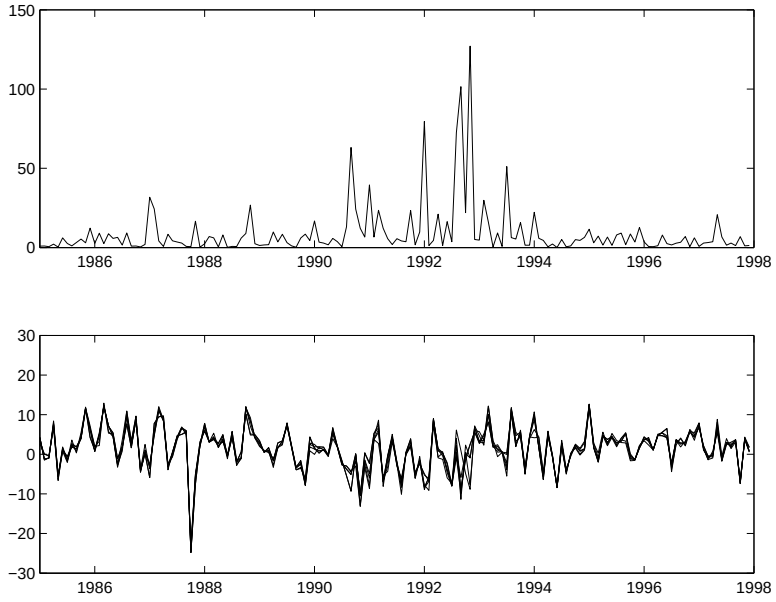

Figure 5: Top: Measure of the difference between the unconstrained portfolios of the last six estimators in table 2 every month between January 1985 and December 1997. Bottom: The out-of-sample portfolio realizations between January 1985 and December 1997 for the last six estimators in table 2.

6 Conclusion

In this paper we have focused on the improvement of covariance matrix estimation for asset returns on the Swedish market. This is one of the two input parameters in the Mean-Variance theory of Markowitz (1952,1959) for portfolio selection, the other being the expected return vector. We have looked at the use of Bayesian shrinkage and principal component analysis together with random matrix theory.

We have proposed two estimators. (1) A shrinkage estimator where we shrink the sample covariance matrix towards a single principal component covariance matrix, based on the work of Ledoit & Wolf (2001). We simply replaced the single-index market model covariance matrix used by Ledoit & Wolf (2001), with the single principal component covariance matrix. Doing this, we had to slightly modify the formula for the optimal shrinkage intensity. The intention behind this estimator was to decrease the contribution from the lower (noisy) part of the spectrum of the sample correlation matrix while we

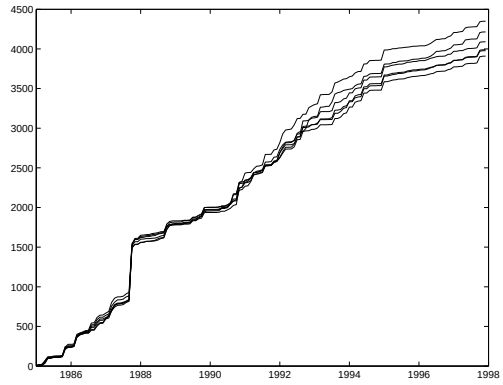

Figure 6: The cumulative sum of the de-meaned and squared out-of-sample portfolio realizations for the last six estimators in table 2.

	Unconstrained Standard Deviation	Constrained Standard Deviation
Sample Estimate	23.78 (1.00)	17.95 (0.12)
Equally Weighted Portfolio	20.61 (0.14)	20.61 (0.14)
Diagonal Model	20.44 (0.13)	20.44 (0.13)
Single-Index Market Model	17.96 (0.12)	17.89 (0.10)
Single Factor PC Model	17.65 (0.12)	17.81 (0.13)
Shrinkage To Market	17.64 (0.13)	17.68 (0.12)
Shrinkage To Single PC	17.57 (0.11)	17.73 (0.13)
Average of Sample, Diagonal and Market	17.47 (0.12)	17.89 (0.10)
Average of Sample, Diagonal and Single PC	17.45 (0.11)	17.88 (0.11)

Table 3: Table 1 re-estimated when the years 1991 to 1995 have been removed from the sample.

kept the (stable) part that stems from the largest eigenvalue and its eigenvector intact. We motivated this by the work of Plerou et al. (2001) and the prediction for the largest eigenvalue of a completely random correlation matrix. An attractive property of our modification to the shrinkage estimator of Ledoit & Wolf (2001) is that it does not depend on an exogenously given covariance matrix such as the single-index covariance matrix. We simply re-use some of the building blocks from the sample covariance matrix and as it turns out, we still manage to keep some economical interpretation in that the largest eigenvalue and its eigenvector can be interpreted as a market factor. Since the form of our shrinkage estimator is rather general in that it is not based on any explicit economical assumptions, it may be useful in applications outside an economic context. (2) An equally weighted portfolio consisting of the sample covariance matrix, the diagonal part of the sample covariance matrix and the single principal component covariance matrix, based on the work of Jagannathan & Ma (2000). This can either be seen as diversifying amongst

estimators or like extending the shrinkage estimator of Ledoit & Wolf (2001) to the case of several estimators. Again, we simply replaced the single-index covariance matrix with the single principal component covariance matrix. The downside of this estimator might be that it is not sufficiently motivated mathematically, at least not in contexts other than economics. The upside is of course that it is easily calculated.

The reason as to why we write *single* principal component covariance matrix as opposed to a K factor principal component covariance matrix is that our analysis showed that it is only fruitful to use one principal component on the Swedish market. Only the first eigenvalue of the sample correlation matrix ever deviates significantly from the maximum eigenvalue prediction of a random correlation matrix.

Empirically, both our estimators show improvement relative to their predecessors and our portfolio of estimators performs the best out of all competing estimators. The shrinkage estimator comes in second. It would be interesting to try out the two estimators on US data in order to investigate if any improvement can be made for $K > 1$. The use of US data would also be interesting as a comparison since the results from the Swedish market seem to be quite sensitive to the length of the estimation window.

References

Asgharian, H. & Hansson, B., 2002, *Factors Explaning Covariances of Equity Returns for a Small Open Economy: A Risk Management Perspective*, Department of Economics, School of Economics and Management, Lund University.

Best, M. & Grauer, R., 1990, *The Efficient Set Mathematics When Mean-Variance Problems Are Subject to General Linear Constraints*, Journal of Economics and Business, 42, 105-120.

Best, M. & Grauer, R., 1991, *On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results*, The Review of Financial Studies, 2, 315-342.

Chan, L., Karceski, J., Lakonishok, J. & Potters, M., 1999, *On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model*, Working Paper 7039 (http://papers.nber.org/papers/w7039), National Bureau of Economic Research.

Chopra, V. & Ziemba, W., 1993, *The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice*, Journal of Portfolio Management, Winter.

Guhr, T., 2001, *Econophysics*, Lecture Notes, Department of Mathematical Physics, Lund Institute of Technology.

Halldin, R., 2002, *Scenario Trees for Inflow Modelling in Stochastic Optimisation for Energy Planning*, Licentiate Theses, Mathematical Statistics, Center for Mathematical Sciences, Lund Institute of Technology.

Jagannathan, R. & Ma, T., 2000, *Three Methods for Improving the Precision in Covariance Matrix Estimation*, Working Paper.

Jagannathan, R. & Ma, T., 2002, *Risk Reduction in Large Portfolios: A Role for Portfolio Weight Constraints*, Working Paper.

Johnson, R. & Wichern, D., 1992, *Applied Multivariate Statistical Analysis, Third Edition*, Prentice Hall, Englewood Cliffs, New Jersey.

Laloux, L., Cizeau, P., Bouchaud, J-P & Potters, M., 1998, *Noise Dressing of Financial Correlation Matrices*, arXiv:cond-mat/9810255 v1.

Ledoit, O. & Wolf, M., 2001, *Improved Estimation of the Covariance Matrix of Stock Returns With an Application to Portfolio Selection*, Working Paper, Department of Economics and Business, Universitat Pompeu Fabra.

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

Markowitz, H., 1959, *Portfolio Selection: Efficient Diversification of Investment*, Yale University Press, New Haven, CT, USA.

Michaud, R., 1998, *Efficient Asset Management*, Harvard Buisiness School Press, Boston, Massachusetts, USA.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L.A.N., Guhr, T. & Stanley, H.E., 2001, *A Random Matrix Approach to Cross-Correlations in Financial Data*, arXiv:cond-mat/0108023.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L.A.N. & Stanley, H.E., 1999, *Universal and Nonuniversal Properties of Cross Correlations in Financial Time Series*, Physical Review Letters 83, American Physical Soc.

Sharpe, W., 1963, *A Simplified Model for Portfolio Analysis*, Management Science, 79, 277-231.

Varian, H., 1999, *Intermediate Microeconomics: A Modern Approach*, Fifth Edition, W.W. Norton & Company.