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

Technical Report TR-2008-5
June 6, 2008

Sufficient Dimension Reduction and Optimal Portfolio Selection

Efstathia Bura
Department of Statistics
The George Washington University, Washington, DC

Sufficient Dimension Reduction and Optimal Portfolio Selection

Efstathia Bura*

June 6, 2008

Abstract

An optimal portfolio is a weighted sum of p stocks, or variables, that maximizes the mean return and has minimum variance. Finding such a portfolio is the classical mean-variance optimal portfolio selection problem. Markowitz (1952) computed the solution to this problem, which involves the inversion of the covariance matrix of the returns of the p stocks. Two main approaches seeking to find a well-conditioned estimator of the covariance matrix have been proposed. Ledoit and Wolf (2004) used the idea of shrinkage toward a target covariance matrix, typically the identity, to regularize the sample covariance. Another approach is to impose a low-dimensional factor structure on the covariance estimator via a K-factor model with uncorrelated residuals so that it is non-singular. Yet, there is no consensus either on the identity of the factors or on the number of factors K. In this paper we present an approach where K is the rank of a random matrix. We estimate K using a sequential asymptotic chi-square test for the rank of a random matrix. The resulting non-singular estimate of the covariance matrix is used to solve the portfolio optimization problem.

*Department of Statistics, George Washington University, Washington, DC 20052.

Key Words: Dimension Reduction; Covariance matrix; Markowitz Mean-Variance Efficiency; SAVE; SIR.

1 Portfolio Selection

In Portfolio Selection capital is allocated over a number of available assets in order to maximize the return on the investment while minimizing the risk. Markowitz's mean-variance efficiency (1952, 1959) gives the classic paradigm of modern finance for efficiently allocating capital among risky assets: diversification always reduces investment risk since real assets almost never have returns perfectly correlated with others.

Markowitz viewed investment choice as a result of utility maximizing behavior in a world of uncertainty. The utility assigned to a given wealth level can be expressed as a quadratic utility function $U(w) = aw - bw^2$, where a and b are constants. Since a portfolio's return can be defined as end-of-period wealth divided by initial wealth, the quadratic utility function implies that investors will select portfolios based solely on their expected return and standard deviation of return.

Consider a universe of p stocks whose returns are distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. By assuming that investors' preference is fully defined by the mean and the variance of portfolio returns over a single holding period, Markowitz showed that, given either an upper bound on the risk that the investor is willing to take or a lower bound on the return the investor is willing to accept, the optimal portfolio can be obtained by solving a convex quadratic programming problem:

$$\min \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} = \min \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij} = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j=1, i \neq j}^{n} w_i w_j \sigma_{ij} \qquad (1)$$

subject to

$$\mathbf{w}^T \mathbf{1}_p = 1$$
$$\mathbf{w}^T \boldsymbol{\mu} = \mathrm{E}(R_p)$$

where $\mathbf{1}_p$ denotes a p-vector of ones, $\mathrm{E}(R_p)$ is the expected rate of return that is required on the portfolio, w_i is the investment portion of the weights of the ith stock, σ_i^2 is the variance of the ith stock, and σ_{ij} is the covariance between the return of the ith and jth stocks. The solution is:

$$\mathbf{w} = \frac{C - qB}{AC - B^2} \boldsymbol{\Sigma}^{-1} \mathbf{1}_p + \frac{qA - B}{AC - B^2} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \tag{2}$$

where $A = \mathbf{1}_p^T \boldsymbol{\Sigma}^{-1} \mathbf{1}_p$, $B = \mathbf{1}_p^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$, and $C = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$. Equation (2) shows that optimal portfolio weights depend on the inverse of the covariance matrix. This may cause difficulty if the covariance matrix estimator is not invertible or if it is numerically ill-conditioned, which means that inverting it amplifies estimation error tremendously (Michaud, 1989). This mean-variance model has had a profound impact on the economic modeling of financial markets and the pricing of assets.

In practice, the covariance matrix is estimated from historical data available up to a given date, optimal portfolio weights are computed from this estimate, then the portfolio is formed on that date and held until the next rebalancing occurs. The performance of a covariance matrix estimator is measured by the variance of this optimal portfolio after it is formed. It is a measure of out-of-sample performance, or of predictive ability. An estimator that overfits in-sample data can turn out to work very poorly for portfolio selection, which is why imposing some structure is beneficial. In the specific context of risk management and portfolio allocation, the number p of stocks can be in the order of hundreds, which is typically of the same or higher order as the sample size n. For example, when $p = 200$ there are more than 20,000 free parameters in the covariance matrix. Yet, the available sample size is usually in the

hundreds or a few thousands because longer time series (larger n) increase modeling bias.

1.1 Factor Models

- *Capital Asset Pricing Model (CAPM):* The computation of risk reduction as proposed by Markowitz is tedious. For a more efficient computation, the Capital Asset Pricing Model (CAPM), independently developed by Sharpe (1964), Lintner (1965) and Mossin (1966), becomes an immediate logical extension of Markowitz's portfolio theory.

  The CAPM assumes the financial market is efficient and in equilibrium and states that that the expected return of any risky asset is a linear function of its tendency to co-vary with the market portfolio:

$$\mathrm{E}(R_j) = R_f + \beta_j[\mathrm{E}(R_m) - R_f] \tag{3}$$

  where $\mathrm{E}(R_j)$ and $\mathrm{E}(R_m)$ are the expected returns to asset j and the market portfolio, respectively, R_f is the risk free rate, and β_j measures the tendency of asset j to co-vary with the market portfolio.

  The market portfolio, usually described as an "index," is the single most important factor that influences the variability of each individual asset and variability from all other influences rapidly disappears when a portfolio is formed. In other words, the risk associated with a risky asset now can be decomposed into systematic risk and an idiosyncratic risk and only the idiosyncratic risk can be diversified away.

  Since the market portfolio is not observable in practice, some publicly-traded security indices are commonly used as its proxy. However, the real market also consists of other non-publicly-traded investments that make up so much of the

economy's net wealth. Therefore, the universe of the publicly-traded securities does not exactly represent all of the capital risks investors are exposed to. As a result, this miss-specification of the market portfolio has led to the failure of the CAPM: many empirical studies of stock returns have shown that the market beta alone can not explain the cross-sectional variation in returns.

- *Arbitrage Pricing Theory (APT):* Unlike the CAPM, which is a model of financial market equilibrium, the APT starts with the premise that arbitrage opportunities should not be present in efficient financial markets. This assumption is much less restrictive than those required to derive the CAPM. The APT starts by assuming that there are K factors which cause asset returns to systematically deviate from their expected values. However, the theory does not specify how large the number K is, nor does it identify the factors. It simply assumes that these factors cause returns to vary together. There may be other, firm-specific reasons for returns to differ from their expected values, but these firm-specific deviations are not related across stocks. Since the firm-specific deviations are not related to one another, all return variation not related to the K common factors can be diversified away. Based on these assumptions, Ross (1976) shows that, in order to prevent arbitrage, an asset's expected return must be a linear function of its sensitivity to the K common factors:

$$\mathrm{E}(R_j) = R_f + \sum_{i=1}^{n} \beta_{ji} f_i \tag{4}$$

where $\mathrm{E}(R_j)$ and R_f are defined as before. Each β_{ji} coefficient represents the sensitivity of asset j to risk factor f, and f_i represents the risk premium for factor i.

As with the CAPM, we have an expression for expected return that is a linear function of the asset's sensitivity to systematic risk. Under the assumptions

of APT, there are K sources of systematic risk, whereas there is only one in a CAPM world.

- *Some extensions of Capital Asset Pricing Model:* Both the CAPM and the APT are static, or single-period models. As such, they ignore the multi-period nature of participation in the capital markets. Merton's (1973) intertemporal capital asset pricing model (ICAPM) was developed to capture this multi-period aspect of financial market equilibrium. The ICAPM framework recognizes that the investment opportunity set might shift over time, and investors would like to hedge themselves against unfavorable shifts in the set of available investments. If a particular security tends to have high returns when bad things happen to the investment opportunity set, investors would want to hold this security as a hedge. This increased demand would result in a higher equilibrium price for the security (all else held constant). One of the main insights of the ICAPM is the need to reflect this hedging demand in the asset pricing equation. The form of the ICAPM is very similar to that of the APT. There are subtle differences, however. The first factor of the ICAPM is explicitly identified as being related to the market portfolio. Further, while the APT gives little guidance as to the number and nature of factors, the factors that appear in the ICAPM are those that satisfy conditions such that they describe the evolution of the investment opportunity set over time and investors care enough about them to hedge their effects. We still don't know exactly how many factors there are, but the ICAPM at least gives us some guidance. It is clear from well-established stylized facts that the unconditional security return distribution is not normal and the mean and variance of returns alone are insufficient to characterize the return distribution completely. This has led researchers to pay attention to the third moment - skewness- and the fourth moment - kurtosis. Investors are generally compensated for taking high risk as measured by high systematic variance

and systematic kurtosis. Investors also forego the expected returns for taking the benefit of a positively skewed market. It also has been documented that skewness and kurtosis cannot be diversified away by increasing the size of portfolios (Arditti, 1971). Many researchers extensively investigated the validity of the CAPM in the presence of higher-order co-moments and their effect on asset prices and extend the CAPM to incorporate skewness in asset valuation models and provided mixed results.

1.2 K-factor Covariance Estimation

An optimal portfolio is a weighted sum of the p stocks, or variables, that maximizes the mean return and has minimum variance. Finding such a portfolio is the classical mean-variance optimal portfolio selection problem. Markowitz (1952) computed the solution to this problem, which involves the inversion of the covariance matrix, $\mathbf{\Sigma}_x$, of the returns of the p stocks. Two main approaches seeking to find a "well-conditioned" estimator of $\mathbf{\Sigma}_x$ have been proposed. Ledoit and Wolf (2004) used the idea of shrinkage toward a target covariance matrix, typically the identity, to regularize the sample covariance estimator of $\mathbf{\Sigma}_x$. Another approach is to impose a certain structure on $\mathbf{\Sigma}_x$. Fan *et al.* (2006), motivated by the Arbitrage Pricing Model, proposed a model-based covariance matrix estimator that is non-singular and asymtotically normally distributed.

If we let X_i, $i = 1, \ldots, p$, denote the excessive returns over the risk-free interest rate of the p assets, $\eta_1, \ldots, \eta_K$ be the factors that can be used to model an efficient market portfolio (see Campbell, Lo and MacKinlay, 1997) and $\alpha_{ij}, i = 1, \ldots, p$, $j = 1, \ldots, K$, be the unknown factor loadings, the K-factor model is

$$X_i = \alpha_{i1}\eta_1 + \alpha_{i2}\eta_2 + \ldots + \alpha_{iK}\eta_K + \epsilon_i \tag{5}$$

where the errors e_i are uncorrelated given the factors $\eta_1, \ldots, \eta_K$. The factors η_i in (5)

are observable and of pre-set dimension K.

If a few factors can completely capture the cross-sectional risks, the number of parameters in covariance matrix estimation can be significantly reduced. For example, using the Fama-French three-factor model (Fama and French (1992, 1993)), there are $4p$ instead of $p(p+1)/2$ parameters to be estimated. Despite the popularity of factor models in the literature, the impact of dimensionality on the estimation errors of covariance matrices and its application to portfolio allocation and risk management are poorly understood. The factor model has been extensively studied in the literature [see, e.g. Scott (1966) and (1969), Browne (1987), Browne and Shapiro (1987), and Yuan and Bentler (1997)], but traditional work assumes the sample size n tends to infinity while the stocks in the portfolio p and the number of factors K are fixed.

The factor model (5) is expressed in matrix form as

$$\mathbf{x} = \mathbf{A}\boldsymbol{\eta} + \boldsymbol{\epsilon} \tag{6}$$

where $\mathbf{x} = (X_1, \ldots, X_p)$, $\mathbf{A} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p)^T$ with $\boldsymbol{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{ip})^T$ $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_K)^T$ and $\boldsymbol{\epsilon} = (e_1, \ldots, e_p)^T$. Throughout we assume that $\mathrm{E}(\boldsymbol{\epsilon}|\boldsymbol{\eta}) = 0$ and $\mathrm{Cov}(\boldsymbol{\epsilon}|\boldsymbol{\eta}) = \boldsymbol{\Sigma}_0$ is diagonal.

Let $(\boldsymbol{\eta}_1, \mathbf{x}_1), \ldots, (\boldsymbol{\eta}_n, \mathbf{x}_n)$ be n independent and identically distributed (i.i.d.) samples of $(\boldsymbol{\eta}, \mathbf{x})$. Also, let

$$\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{x}) \quad \mathbf{F} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n) \quad \mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \quad \mathbf{E} = (\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n)$$

Then, model (6) yields

$$\boldsymbol{\Sigma} = \mathbf{A}\mathrm{Cov}(\boldsymbol{\eta})\mathbf{A}^T + \boldsymbol{\Sigma}_0 \tag{7}$$

Since model (6) is a regression model, one can estimate $\boldsymbol{\Sigma}$ by least squares. That is,

$$\widehat{\boldsymbol{\Sigma}} = \hat{\mathbf{A}}\widehat{\mathrm{Cov}}(\boldsymbol{\eta})\hat{\mathbf{A}}^T + \widehat{\boldsymbol{\Sigma}}_0 \tag{8}$$

where $\hat{\mathbf{A}} = \mathbf{X}\mathbf{F}^T(\mathbf{F}\mathbf{F}^T)^{-1}$ is the matrix of estimated regression coefficients, $\widehat{\mathrm{Cov}}(\boldsymbol{\eta}) = (n-1)^{-1}\mathbf{F}\mathbf{F}^T - \{n(n-1)\}^{-1}\mathbf{F}\mathbf{1}\mathbf{1}^T\mathbf{F}^T$ is the sample covariance matrix of the factors $\boldsymbol{\eta}$, and $\widehat{\boldsymbol{\Sigma}}_0 = \mathrm{diag}(n^{-1}\widehat{\mathbf{E}}\widehat{\mathbf{E}}^T)$, where $\widehat{\mathbf{E}} = \mathbf{X} - \hat{\mathbf{A}}\mathbf{F}$ is the residual matrix.

The sample covariance matrix estimator is written as

$$\widehat{\boldsymbol{\Sigma}}_x = (n-1)^{-1}\mathbf{X}\mathbf{X}^T - \{n(n-1)\}^{-1}\mathbf{X}\mathbf{1}\mathbf{1}^T\mathbf{X}^T \tag{9}$$

Fan *et al.* (2006) showed that $\widehat{\boldsymbol{\Sigma}}$ is always invertible, even when $p > n$. The main advantage of the factor model lies in the estimation of the inverse of the covariance matrix, not the estimation of the covariance matrix itself. Optimal portfolio allocation involves the inverse of the covariance matrix and the factor model based estimates yield substantial gain. They also showed that the random matrix $\widehat{\boldsymbol{\Sigma}}$ is asymptotically normal.

In this paper, we present an estimation method for the number K of factors that are truly needed to model stock returns. This approach is one of the sufficient dimension reduction methods that have been developed over the last seventeen years. It is especially suited to the formulation of the K factor model as its formulation is that of reduced-rank regression.

Reduced rank regression assumes that the coefficient matrix in a multivariate regression model is not of full rank. The unknown rank is traditionally estimated under the assumption of normal responses. Bura and Cook (2003) derived an asymptotic test for the rank that only requires the response vector have finite second moments. The test was extended to the non-constant covariance case. Linear combinations of the components of the predictor vector that are estimated to be significant for modelling the responses were obtained. In our context, these serve as the factor projections sufficient to capture the common features of the portfolio's asset returns.

In the next section we present an overview of sufficient dimension reduction methodology (SDR) and then we proceed by focusing on estimating the rank of the regression in (6).

2 Introduction to Sufficient Dimension Reduction (SDR)

Let $\mathbf{X} = (X_1, \ldots, X_p)^T$ denote a predictor vector and Y a response variable. Throughout vectors and matrices will be boldfaced in contrast to univariate quantities. Sufficient dimension reduction via inverse regression is based on the idea that $\mathbf{X}$ can be replaced with a lower-dimensional projection $P_S\mathbf{X}$ without loss of information about the conditional distribution of $Y|\mathbf{X}$, where P_S is the orthogonal projection onto the vector space S in the usual inner product. No pre-specified model for $Y|\mathbf{X}$ is required. The intersection of all subspaces $S \subset R^p$ with $F(Y|\mathbf{X}) = F(Y|P_S\mathbf{X})$, where $F(\cdot|\cdot)$ is the conditional distribution function of the response Y given the second argument, is the *central* subspace, $S_{Y|X}$ (Cook, 1996, 1998). The dimension $d = \dim(S_{Y|X})$ is called the structural dimension of the regression of Y on $\mathbf{X}$ and can take on any value in the set $\{0, 1, \ldots, p\}$. When $d < p$, the *structural* dimension of the regression is smaller than the numbers of predictors. If $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_d)$ is a basis for $S_{Y|\mathbf{X}}$, $P_{\boldsymbol{\eta}}\mathbf{X}$, or equivalently, the d linear combinations $\boldsymbol{\eta}^T\mathbf{X} = (\boldsymbol{\eta}_1^T\mathbf{X}, \ldots, \boldsymbol{\eta}_d^T\mathbf{X})$ contain all the information in $\mathbf{X}$ about Y.

For numerical stability and computational simplicity and efficiency, the predictor is standardized to have zero mean and identity variance-covariance structure; that is, if $\boldsymbol{\Sigma}_x$ denotes the covariance matrix of $\mathbf{X}$, $\mathbf{Z} = \boldsymbol{\Sigma}_x^{-1/2}(\mathbf{X} - \mathrm{E}(\mathbf{X}))$ is its standardized version. There is no loss of generality in working in the $\mathbf{Z}$-scale as $S_{Y|X} = \boldsymbol{\Sigma}_x^{-1/2} S_{Y|Z}$. The estimation of the central subspace in almost all sufficient dimension reduction (SDR) techniques is based on finding a kernel matrix $\mathbf{M}$ so that

$$S(\mathbf{M}) \subset S_{Y|\mathbf{Z}} \tag{10}$$

Kernel matrices based on first moments are used in SIR (Li, 1991) with $\mathbf{M} = \mathrm{Cov}(\mathrm{E}(\mathbf{Z}|Y))$, and polynomial inverse regression (Bura and Cook, 2001b), with $\mathbf{M} = \mathrm{E}(\mathbf{Z}|Y)$. Kernel matrices based on both first and second moments are used in pHd (Li,

1991), with $\mathbf{M} = \mathrm{E}((Y - \mathrm{E}(Y))\mathbf{Z}\mathbf{Z}^T)$, SAVE (Cook and Weisberg, 1991), with $\mathbf{M} = \mathrm{E}(\mathbf{I}_p - \mathrm{Cov}(\mathbf{Z}|Y))^2$, and SIRII (Li, 1991) with $\mathbf{M} = \mathrm{E}(\mathrm{Cov}(\mathbf{Z}|Y) - \mathrm{E}(\mathrm{Cov}(\mathbf{Z}|Y)))^2$.

For the span of a kernel matrix to be a subspace of or equal to $S_{Y|Z}$ at least one of two conditions on the marginal moments of the predictors must hold. For first moment based kernel methods the following linearity condition is needed:

$$\mathrm{E}(\mathbf{Z}|P_{S_{Y|Z}}\mathbf{Z}) = P_{S_{Y|\mathbf{Z}}}\mathbf{Z} \tag{11}$$

For second moment based kernel methods condition (11) and also the constant variance condition

$$\mathrm{Var}(\mathbf{Z}|P_{S_{Y|Z}}\mathbf{Z}) = Q_{S_{Y|Z}} \tag{12}$$

are required to hold.

To assess $d = \dim(S_{Y|Z})$, typically a test statistic for dimension that is a function of the singular (or eigen) values of an estimate of the kernel matrix $\widehat{\mathbf{M}}$ is used. The test statistic is generally of the form $L_k = n\sum_i f(\hat{\lambda}_i)$, where $\hat{\lambda}_i$ are the singular or eigenvalues of $\widehat{\mathbf{M}}$ in decreasing order and f is a smooth non-negative valued function. The dimension is usually estimated via sequential hypothesis testing of $H_0 : d = k$ against $H_a : d > k$, starting at $k = 0$, which corresponds to independence of Y and $\mathbf{Z}$. Assessment of the accuracy of the estimation requires knowledge of the asymptotic distribution of the test statistic, whose computation comprises an important aspect of all SDR techniques.

The Unification of SDR Methods: At the heart of most SDR methods is the estimation of a kernel matrix $\mathbf{M}$ so that (10) is satisfied. In general, $\mathbf{M}$ is of order $p \times q$. If we let $r = \mathrm{rank}(\mathbf{M})$, then r is a lower bound of the dimension d of the central subspace $S_{Y|X}$. If the estimation of the central subspace is exhaustive, that is $S(\mathbf{M}) = S_{Y|X}$, then $d = r$. Let $\lambda_1 \geq \ldots \geq \lambda_r$ be the singular values of $\mathbf{M}$, and $\mathbf{u}_1, \ldots, \mathbf{u}_r$ be the corresponding singular vectors. Under (10), $\mathrm{span}(\mathbf{u}_1, \ldots, \mathbf{u}_r) \subset S_{Y|X}$

so that SDR estimation methods can be formulated as eigen-decomposition problems where estimating the dimension of $S(\mathbf{M})$ amounts to estimating the rank of the kernel matrix $\mathbf{M}$, r, and estimation of the subspace itself amounts to estimating the r left singular vectors of $\mathbf{M}$, $\mathbf{u}_1, \ldots, \mathbf{u}_r$. In a separate project (Bura, 2007, work in progress), the rank r is estimated via sequential hypotheses testing. Special cases of this test are the tests for the most widely used SDR methods of SIR (Li, 1991; Bura and Cook, 2001(a,b)) and SAVE (Yin, 2005).

Determining the rank of a matrix is a difficult task. The approach in Bura and Pfeiffer (2007) is based on the fact that the rank of the matrix equals the number of its non zero eigen- or singular values. A formal test for the rank of a matrix is equivalent to testing how many of the eigen- or singular values of the matrix are equal to zero. The two rank tests proposed in Bura and Pfeiffer (2007), based on the smallest eigen- or singular values of the estimated matrix, generalize and unify previous results in SDR methodology. The weighted chi-squared test statistic is based on an important result for the asymptotic distribution of the singular values of a random matrix by Eaton and Tyler (1994). The perturbation theory based test is asymptotically chi-squared and is expected to be more accurate since it does not involve estimation of weights.

Furthermore, a Wald type test for assessing the contribution of any given subset of the individual predictors $X_1, \ldots, X_p$ to the SDR projections is developed in Bura and Pfeiffer (2007). This test allows one to carry out variable selection prior to fitting any model to the data.

2.1 Reduced-rank regression and SDR

Let $\mathcal{S} = \text{span}(\mathbf{A}\boldsymbol{\eta})$ denote the subspace spanned by the columns of $\mathbf{A}\boldsymbol{\eta}$. This subspace represents the fewest linear combinations of $\boldsymbol{\eta}$ that are needed for the regression. If $\mathbf{b}$ is a known matrix whose columns form a basis for $\mathcal{S}$ then we can replace $\boldsymbol{\eta}$ with $\mathbf{b}$

without loss of information on the regression. The dimension of $\mathcal{S}$, the minimum number of reduced predictors, is equal to the rank of $\mathbf{A}$: Since $\text{rank}(\mathbf{A}\boldsymbol{\eta}) = \text{rank}(\boldsymbol{\eta}^T\mathbf{A}^T)$,

$$\dim(\mathcal{S}) = \text{rank}(\boldsymbol{\eta}^T\mathbf{A}^T) = \text{rank}(\mathbf{A}\boldsymbol{\eta}\boldsymbol{\eta}^T\mathbf{A}^T) = \text{rank}(\mathbf{A}) \tag{13}$$

because $\boldsymbol{\eta}\boldsymbol{\eta}^T$ is a $K \times K$ positive definite matrix (see *A4.4*, Seber, 1977) by our assumption on the rank of $\boldsymbol{\eta}$. In consequence, inference on the dimension of $\mathcal{S}$ can be based solely on $\mathbf{A}$ in the sense that an estimate of the rank of $\mathbf{A}$ constitutes an estimate of the dimension of $\mathcal{S}$.

When the dimension of $\mathcal{S}$ is less than $\min(K, p)$, model (6) corresponds to the basic reduced-rank regression model (see Reinsel and Velu, 1998, ch. 2). Reduced-rank regression models were introduced by Anderson (1951) and are used mostly when there is a need to reduce the number of parameters in (6). They have a wide spectrum of applications in fields such as chemometrics (Frank and Friedman, 1993), psychometrics (Anderson and Rubin, 1956), econometrics (Velu, Reichern and Wichern, 1986), and financial economics (Zhou, 1995). The typical analysis of a reduced rank regression model is based on the assumption that the coefficient matrix $\mathbf{A}$ is not of full rank. The elements of $\mathbf{A}$ are subsequently estimated for a given value of the rank of $\mathbf{A}$.

Recall that $\hat{\mathbf{A}}$ is the OLSE for $\mathbf{A}$, and assume that $\mathbf{G}_n = (\mathbf{F}\mathbf{F}^T/n)^{-1}$ has a $K \times K$ positive definite limit matrix $\mathbf{G}$. Let $\mathbf{H}_n = \boldsymbol{\Sigma} \otimes (\mathbf{F}\mathbf{F}^T/n)^{-1}$ denote the covariance matrix of $\sqrt{n}\text{vec}(\hat{\mathbf{A}} - \mathbf{A})$.

Bura and Cook (2003) showed that under model (6), if $\mathbf{G}_n$ has a positive definite limit matrix, and that $\widehat{\boldsymbol{\Sigma}}$ is any consistent estimate of $\boldsymbol{\Sigma}$, then

$$\Lambda_d = n \sum_{j=d+1}^{\min(K,p)} \phi_j^2 \tag{14}$$

is asymptotically distributed as a $\chi^2_{(K-d)(p-d)}$ random variable, where $\phi_1 \geq \phi_2 \geq \ldots \geq \phi_{\min(K,p)}$ are the ordered singular values of $\hat{\mathbf{A}}_{\text{std}}$, $\hat{\mathbf{A}}_{\text{std}} = \mathbf{G}_n\hat{\mathbf{A}}\widehat{\boldsymbol{\Sigma}}^{-1/2}$.

We use Λ_k as a test statistic for the rank of $\mathbf{A}$, or equivalently, for the rank of the model. We start by testing the hypothesis that $d = 0$, that is, by comparing Λ_0 to the percentage points of a chi-squared distribution with Kp degrees of freedom. If the test is significant, the rank is estimated to be 0. Otherwise, we sequentially test $d = 1, 2, \ldots, \min(K, p)$ till we encounter a significant value which serves as the estimate of the rank of the model. This is a fairly standard general procedure to estimate a rank (see, for example, Rao, 1965, p. 474). As an aside, it is easy to see that the chi-square asymptotic test coincides with the usual F-test for testing $d = 0$; that is, that all the coefficients are zero, when $p = 1$.

The d left singular vectors $\mathbf{u}_j$ of $\hat{\mathbf{A}}$ that correspond to its d largest singular values provide an estimated basis for $\mathcal{S}$. The linear combinations of $\boldsymbol{\eta}$ needed for the regression are then constructed as $\mathbf{u}_j^T \mathbf{F}$, $j = 1, \ldots, d$.

We will use this result to estimate how many factors are necessary to model the three-year daily data of 30 Industry Portfolios from May 1, 2002 to Aug. 29, 2005, which are available at the website

http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html using several factors. Fama and French (1993) proposed a three factor model. The first factor is the excess return of the proxy of the market portfolio, which is the value-weighted return on all NYSE, AMEX and NASDAQ stocks (from CRSP) minus the one-month Treasury bill rate (from Ibbotson Associates). The other two factors are constructed using six value-weighted portfolios formed on size and book-to-market. These three factors will be included along with a few others. The structural dimension of the (6) model will be assessed with the chi-square test and the resulting projected factors will be used in the estimation of $\boldsymbol{\Sigma}$ using (8). Our results will be compared to those of Fan *et al.* (2006).

References

Anderson, T. W., Estimating linear restrictions on regression coefficients for multivariate normal distributions, *Ann. Math. Statist.* **22** (1951), 327–351.

Anderson, T. W. and Rubin, H., Statistical inference in factor analysis, *in* "Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability" (J. Neyman, Ed.), pp 111–150, University of California Press, Berkeley, 1956.

Browne, M. W. (1987). Robustness of statistical inference in factor analysis and related models. *Biometrika* **74**, 375384.

Browne, M. W. and Shapiro, A. (1987). Adjustments for kurtosis in factor analysis with elliptically distributed errors. *J. Roy. Statist. Soc. Ser. B*, **49**, 346352.

Bura, E. and Cook, R. D. (2001a). Estimating the structural dimension of regressions via parametric inverse regression. *J. R. Statist. Soc.* B **63**, pp. 393-410.

Bura, E. and Cook, R. D. (2001b). Extending SIR: The Weighted Chi-Square Test, *Journal of the American Statistical Association* **96**, 996–1003.

Bura, E. and Cook, R. D. (2003). Rank Estimation in Reduced-Rank Regression. *Journal of Multivariate Analysis*, **87**, 159-176.

Bura, E. and Pfeiffer, R. (2007). On the distribution of the left singular vectors of a random matrix and its applications. To appear in *Statistics & Probability Letters*.

Campbell, J. Y., Lo, A.W. andMacKinlay, A. G. (1997). *The Econometrics of Financial Markets*. Princeton University Press, New Jersey.

Cook, R. D. (1996). Graphics for regressions with a binary response. *Journal of the American Statistical Association* **91**, 983–992.

Cook, R. D. (1998). *Regression Graphics: Ideas for studying regressions through graphics*. New York: Wiley.

Cook, R. D. (2004). Testing predictor contributions in sufficient dimension reduction. *Annals of Statistics* **32**, 1062–1092.

Cook, R. D. and Lee, H. (1999). Dimension reduction in regressions with a binary response. *Journal of the American Statistical Association.* **94**, 1187–1200.

Cook, R. D. and Li, B. (2002). Dimension reduction for the conditional mean in regression. *Annals of Statistics* **30**, 455–474.

Cook, R. D. and Weisberg, S. (1991). Discussion of Li (1991). *Journal of the American Statistical Association* **86**, 328–332.

Diaconis, P. and Freedman, D. (1984). Asymptotics of graphical projection pursuit. *Annals of Statistics* **12**, 793–815.

Donoho, D. L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. Aide-Memoire of a Lecture at AMS Conference on Math Challenges of the 21st Century.

Eaton, M. L. and Tyler, D. E., The asymptotic distribution of singular values with applications to canonical correlations and correspondence analysis, *J. Mult. Anal.* **34** (1994), 439–446.

Eckhardt, M. J., Ryback, R.S., Rawlings, R.R., Graubard, B.I. (1981). Biochemical diagnosis of alcoholism - a test of the discriminating capabilities of γ-Glutamyl Transpeptidase and mean corpuscular volume. *Journal of the American Medical Association* **246**, 2707–2710.

Fama, E. and French, K. (1992). The cross-section of expected stock returns. *Jour. Fin.* **47**, 427465.

Fama, E. and French, K. (1993). Common risk factors in the returns on stocks and bonds. *Jour. Fin. Econ.* **33**, 356.

Fan, J., Fan, Y. and Lv, J. (2006). High Dimensional Covariance Matrix Estimation Using a Factor Model. http://arxiv.org/abs/math.ST/0701124v1

Frank, I. E. and Friedman, J. H., A statistical view of some chemometrics regression tools (with discussion), *Technometrics* **35** (1993), 109–148.

Hall, P. and Li, K. C. (1993). On almost linearity of low dimensional projections from high dimensional data. *Annals of Statistics* **21**, 867–889.

Harville, D. A., "Matrix Algebra From a Statistician's Perspective," Springer-Verlag, New York, 1997.

Izenman, A. J., Reduced-rank regression for the multivariate linear model, *J. Mult. Anal.* **5** (1975), 248–264.

Kato, T. (1982) *A Short Introduction to Perturbation Theory for Linear Operators.* Springer-Verlag, New York.

Ledoit O, Wolf M (2004). A well-conditioned estimator for large-dimensional covariance matrices *Journal of Multivariate Analysis*, **88**, 365-411.

Li, K. C. (1991). Sliced inverse regression for dimension reduction (with discussion). *Journal of the American Statistical Association* **86**, 316–342.

Li, L., Cook R. D. and Nachtsheim C. J. (2005). Model-free variable selection. *Journal of the Royal Statistical Society, Series B***67**, 285–299.

Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance* **7**, 7791.

Markowitz, H. M. (1959). *Portfolio Selection: Efficient Diversification of Investments*. John Wiley and Sons, New Jersey.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* New York: Wiley.

Rao, C. R., "Linear Statistical Inference and Its Applications," Wiley, New York, 1965.

Rao, C.R. (1973). *Linear Statistical Inference and Its Applications, 2nd Edition.* New York: Wiley.

Ratsimalahelo, Z. (2003). Rank test based on matrix perturbation theory. from *Economics Working Paper Archive at WUSTL*, http://econpapers.repec.org/paper/wpawuwpem/0306008.htm.

Reinsel, G. C. and Velu, R. P., "Multivariate Reduced-Rank Regression: Theory and Applications, Lecture Notes in Statistics," Springer-Verlag, New York, 1998.

Ross, S. A. (1976). The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory* **13**, 341360.

Scott, J. T. (1966). Factor analysis and regression. *Econometrica* **34**, 552562.

Seber, G.A.F. (1977). *Linear Regression Analysis.* New York: Wiley.

Tyler, D. E. (1981). Asymptoric inference for eigenvectors. *Annals of Statistics*, 9, 725-736.

Velu, R. P., Reinsel, G. C. and Wichern, D. W., Reduced rank models for multiple time series, *Biometrika* **73** (1986), 105–118.

Yuan, K.-H. and Bentler, P. M. (1997). Mean and covariance structure analysis: theoretical and practical improvements. *J. Amer. Statist. Assoc.* **438**, 767774.

Zhou, G., Small sample rank tests with applications to asset pricing, *Journal of Empirical Finance* **2** (1995), 71–93.