The Incompleteness Problem of the APT Model

Peter S. Karlsson
Jönköping University
SE-551 11 Jönköping
Sweden

Abstract: The Arbitrage Pricing Theory provides a theory to quantify risk and the reward for taking it. While the theory itself is sound from most perspectives, its empirical version is connected with several shortcomings. One extremely delicate problem arises because the set of observable asset returns rarely has a history of complete observations. Traditionally, this problem has been solved by simply excluding assets without a complete set of observations from the analysis. Unfortunately, such a methodology may be shown to (i) lead for any fixed time period to selection bias in that only the largest companies will remain and (ii) lead to an asymptotically empty set containing no observations at all. This paper discusses some possible solutions to this problem and also provides a case study containing Swedish OMX data for demonstration.

Keywords: APT, principal components, asymptotics, bias, high-dimensional data.

1 Introduction

One fundamental question in financial economics, and particularly in portfolio theory, concerns the tradeoff between risk and expected return of assets. The theory is based on the principle that investors want to maximize the expected return and to minimize the risk (the variance of the return on the asset) (Markowitz 1952). When considering a portfolio of *n* assets, the variance can be divided into the sum of each asset variance (unsystematic risk) and the sum of the covariances between assets (systematic risk). It turns out that if the number of assets in the portfolio approaches infinity, then by the law of large numbers the unsystematic risk of the portfolio's variance limits zero, whereas the systematic risk limits an average covariance term. Hence, the only risk an investor is concerned about is the covariation of the asset returns. However, we also need to determine the price of risk, i.e., how much extra expected return is required for an investor to be willing to invest in a risky asset. To solve this problem, Sharpe (1964), Lintner (1965), Mossin (1966) developed the Capital Asset Pricing Model (CAPM), in which the central point is that in equilibrium the expected rates of return on all risky assets are linear functions of their covariance with the market portfolio. Researchers such as Hansen and Jagannathan (1997) have shown that the use of assets covariance with the market portfolio as risk measure does not accurately capture all the risk that the investors face. To account for all risk the investors face, Ross (1976) developed the Arbitrage Pricing Theory model (APT), which takes several risk components or risk factors into account. While the APT provides an important understanding of asset pricing from a theoretical point of view, it is less straightforward to apply in empirical works. More specifically, APT requires that principal components or factors are extracted from the covariance matrix of asset returns (Roll and Ross 1980; Faff 1988, 1992; Shukla and Trzcinka 1990; Chen 1983). In other words, an empirical analysis of the APT model is strongly dependent on the estimation of a covariance matrix. In an ideal perspective the dataset is complete and thus the covariance matrix is easily estimated, but for most practical cases some of the assets in the dataset lack a complete history of observations. For example, a company might have been acquired by another company that may be listed or not listed on the stock exchange, resulting in a sudden end to the series. The same pattern can prevail if a company goes bankrupt due to some financial distress. Also,

new companies may enter the stock exchange, producing stocks which cannot initially be observed in the return matrix. Consequently, the matrix of returns is incomplete. One frequent solution to this problem is to simply drop assets with an incomplete history (relative to other assets) and conduct analysis on the complete dataset (McElroy and Burmeister 1988). However, this method has two serious shortcomings. First, it is usually the assets of very large companies that have a full history of observations and consequently dropping some assets will lead to a bias in the analysis. Second, such a method will asymptotically (as the number of time observations increase) lead to an empty set containing no asset returns at all, an obvious weakness of a statistical analysis. In this paper we will consider different solutions to this problem. Basically, there are two ways out. First, one may impose restrictions on the distributional form of the true unknown covariance matrix so that the number of parameters to estimate is fixed and independent of the dimension of the covariance matrix. By doing so, one can achieve consistent estimates even if some elements in the covariance matrix are based on incomplete stocks. Second, one may extract principal components from an incomplete dataset without estimating the covariance matrix (Cristoffersson 1970; Ruhe 1974; Wiberg 1976). When the principal components have been extracted, they may be used as the risk factors in the APT model.

The paper is arranged as follows. In the next section we present the traditional APT model and discuss the incompleteness problem of the APT model. In Section 3 a model-based approach of the covariance matrix for the incompleteness problem is discussed. In Section 4 we present a solution to the incompleteness problem using a model-independent approach for extraction of principal components using the Singular Value Decomposition (SVD). In the final section an empirical application of the model-independent approach from Section 4 is applied to a dataset from the Stockholm stock exchange.

2 The APT model and the incompleteness problem

Ross (1976) states the APT model as a factor model in the sense that a vector of n assets returns may be formulated as the following model:

$$\mathbf{r} = E[\mathbf{r}] + \mathbf{BF} + \boldsymbol{\varepsilon}, \tag{1.1}$$

where $\mathbf{r}$ is a vector of *n* asset returns, $\mathbf{F}$ is an unobservable vector of the *k* factors, $\mathbf{B}$ is an $(n \times k)$ matrix of the *n* assets sensitivities to the factors and $\boldsymbol{\varepsilon}$ is a noise vector. The assumptions of the factor model are: $E[\boldsymbol{\varepsilon}] = \mathbf{0}$, $Cov(\varepsilon_i, \varepsilon_j) = 0 \; \forall i \neq j$, $E[\mathbf{F}] = \mathbf{0}$, $V(\mathbf{F}) = \mathbf{I}$ and $Cov(\boldsymbol{\varepsilon}, \mathbf{F}) = \mathbf{0}$. These assumptions enable the following decomposition of the covariance matrix of the returns of the assets:

$$\boldsymbol{\Sigma}_{\mathbf{r}} = E\left[\left(\mathbf{r} - E[\mathbf{r}] \right) \left(\mathbf{r} - E[\mathbf{r}] \right)' \right] = \mathbf{B} E[\mathbf{F}\mathbf{F}'] \mathbf{B}' + E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \mathbf{B}\boldsymbol{\Sigma}_{\mathbf{F}}\mathbf{B}' + \mathbf{V}. \qquad (1.2)$$

Given that the factors are orthonormal, (1.2) reduces to:

$$\boldsymbol{\Sigma}_{\mathbf{r}} = \mathbf{B}\mathbf{B}' + \mathbf{V} \qquad (1.3)$$

Moreover, given that $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \mathbf{V}$ is a diagonal matrix in combination with Ross's arbitrage condition, (Ross 1976) states the pricing relationship:

$$E[r_i] = \lambda_0 + \lambda_1 \hat{\beta}_{i1} + \cdots + \lambda_k \hat{\beta}_{ik}, \quad \forall i = 1, \ldots, n, \qquad (1.4)$$

where λ_0 is the return on a risk-free asset and $\lambda_1, \cdots, \lambda_k$ are risk premiums on each risk factor. Chamberlain and Rothschild (1983) show that for a large economy the pricing relationship (1.4) only needs to hold approximate in a large economy, hence:

$$E[r_i] \approx \lambda_0 + \lambda_1 \hat{\beta}_{i1} + \cdots + \lambda_k \hat{\beta}_{ik} \quad \forall i = 1, \ldots, n. \qquad (1.5)$$

The main implication of the APT is that expected returns on assets are linear in the sensitivities to the factors. In an ideal world, the procedure of estimating the APT model consists of performing a factor analysis on the covariance matrix (1.3), which in turn gives the estimated loadings. The loadings are then used as independent variables in cross-sectional regressions for each time point (1.4), where the risk

premiums are estimated. But in reality the task at hand is not trivial since, in general, a typical return matrix will have the pattern shown in the figure below:

$$\mathbf{r}_{4\times5} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} & \cdot \\ r_{21} & r_{22} & r_{23} & r_{24} & r_{25} \\ r_{31} & \cdot & r_{33} & r_{34} & r_{35} \\ r_{41} & \cdot & r_{42} & r_{44} & r_{45} \end{bmatrix}.$$

The above matrix shows an incomplete return matrix consisting of 5 assets measured over a period of 4 time units, where the second asset exited at $t = 3$ and the fifth asset entered at $t = 2$. At this point it should be stressed that the empty cells are not to be considered as missing values in the traditional sense, i.e., values that should have been there but they were lost for some reasons. The empty cells are rather generated by the same economic mechanism that generated the values that are actually there. Obviously, as T grows, the number of stocks with a complete set of observations decreases, while the number of stocks with an incomplete set of observations increases. This is what we refer to as *the incompleteness problem* of the APT model. Due to the fact that the incompleteness problem of the APT model does not involve missing values in the traditional meaning, it does not make sense to impute values to the empty cells. The important fact is that as soon as a stock has been delisted from the stock exchange there will not be any further observations, hence these missing observations do not occur at random. Methods based on "filling the holes" in the data matrix such as the EM algorithm (Depster et al, 1977) assuming that these are occurring randomly (MAR) are hence not fully applicable. Moreover, the EM algorithm is based on distributional assumptions and the returns on assets are known to have non-normal distributions (Rydberg 2000), therefore, there is an additional complication of determining an appropriate distribution when using this method. As a result, other techniques should be considered instead.

So far I have only discussed the use of the $(n \times n)$ covariance matrix for the factor analysis, but there is also the possibility of using the $(T \times T)$ covariance matrix. To avoid the restriction that the number of time periods has to be larger than the number of assets in a factor analysis performed on the $(n \times n)$ covariance matrix, Connor and Korajczyk (1986, 1988) develop a new procedure called Asymptotic Principal

Component (APC). In APC, the first *k* eigenvectors of the $(T \times T)$ covariance matrix are viewed as asymptotically non-singular linear transformations of the true factors from the $(n \times n)$ covariance matrix. This method, however, does not solve the incompleteness problem in cases where the $(T \times T)$ covariance matrix has empty entries. Hence, regardless of in what dimension (or mode) we analyze the data, the incompleteness problem and its implications must be considered. In particular we notice the following:

– In the case of a $(T \times T)$ covariance matrix, the empty cells will show up first in the off-diagonal furthest away from the diagonal. This is due to the fact that as long as there is at least one asset with a full set of observations, the $(T \times T)$ covariance matrix will not have any empty entry.

– For an $(n \times n)$ covariance matrix, the empty cells can show up in any entries except in the diagonal. In order to have a complete $(n \times n)$ covariance matrix, it is required that all assets should have observations in at least two joint successive time points.

One conceivable way to solve the incompleteness problem is to first restrict the analysis to the $(T \times T)$ mode and then assume ex-ante that there is a specific underlying covariance structure, such as a band matrix or a Toeplitz matrix, so that the number of parameters is either fixed or at the most of order *T* (to be compared to the $T(T+1)/2$ unique parameters of a general covariance matrix). By doing so, one can achieve consistent estimates even if some elements in the covariance matrix are based on incomplete data, therefore the number of observations for the element at hand will not increase as we increase the size of the data matrix. This will be discussed further in Section 3.

3 Model-based covariance matrix estimation

When considering the increasing size problem together with the incompleteness problem discussed in the previous section, it may seem a hopeless task to consistently estimate the covariance matrix. However, there are some special types of covariance matrices that turn out to be estimable in certain senses. The large sample properties of

a statistic are usually evaluated with respect to some specific norm. It should be evident that in a setting with a parameter of increasing dimension calculated from an incomplete dataset, only weak modes of convergence, if any, can be established. Hence we will here consider a simple scalar-valued measure and demonstrate that a specific type of matrices converge with respect to this measure. We will define an element-wise partial loss function as follows:

$$\textbf{Def:}\quad \ell_r^0\left(\boldsymbol{\Sigma},\hat{\boldsymbol{\Sigma}}\right) := \left(1/(T-r)\right)\sum_{i=1}^{T-r}\left(\hat{\sigma}_{i,(i+r)} - \sigma_{i,(i+r)}\right)^2. \tag{2.1}$$

Note that the ℓ_r^0 measure is simply the arithmetic mean of the squares of the r first off-diagonal lines of $\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$, where the dimension of $\boldsymbol{\Sigma}$ is $(T \times T)$. Also, since ℓ_r^0 is essentially a scalar-valued measure, it does not express the convergence of $\hat{\boldsymbol{\Sigma}}$ in a genuinely multivariate sense.

To demonstrate Increasing Dimension Asymptotics (IDA) convergence of band matrices we define:

$$\boldsymbol{\Sigma}_{T\times T} = \begin{bmatrix} \sigma^2 & \sigma_1 & \sigma_2 & \\ \sigma_1 & \sigma^2 & \ddots & \sigma_2 \\ \sigma_2 & \ddots & \ddots & \sigma_1 \\ & \sigma_2 & \sigma_1 & \sigma^2 \end{bmatrix},\ \hat{\boldsymbol{\Sigma}}_{T\times T} = \begin{bmatrix} S^2 & S_1 & S_2 & \\ S_1 & S^2 & \ddots & S_2 \\ S_2 & \ddots & \ddots & S_1 \\ & S_2 & S_1 & S^2 \end{bmatrix}, \tag{2.2}$$

where $\sigma_r \le c_r \le \infty$, uniformly in T, and $S_r = \sum_{t=1}^{T-r} w_{r,t} S_{r,t}^2$, $S_{r,t} = \left(1/n_{t,t'}\right)\sum_{i=1}^{n_{t,t'}}\left(X_{i,t} - \bar{X}_t\right)\left(X_{i,(t+r)} - \bar{X}_{t+r}\right)$ where $n_{t,\,t+r}$ is the number of common (i.e., non-empty cross products) observations in $\left\{X_{i,t}, X_{i,(t+r)}\right\}$ and $w_{t,t+r} = n_{t,t+r} \Big/ \sum_{i=1}^{p} n_{t,t+r}$.

Theorem 1: $\underset{T\to\infty}{p\lim}\, \ell_r^0\left(\hat{\boldsymbol{\Sigma}},\boldsymbol{\Sigma}\right) = 0$ for any fixed (predetermined) r.

Proof: See Appendix A.

It is also possible to impose even stronger restrictions of the functional form of $\mathbf{\Sigma}$, such as for example exponentially decaying bands, e.g., $\sigma_{i,i+h} = \rho^h$, so that the matrix only contains a single parameter. For example, if $\mathbf{\Sigma} = \mathbf{\Sigma}(\theta)$, then an estimator may be obtained by minimizing $tr\left(\mathbf{\Sigma}(\theta) - \mathbf{S}\right)^2$ over a range of values of θ. Hence any subsequent analysis involving the covariance matrix could be applied to $\hat{\mathbf{\Sigma}}$, regardless of incompleteness of the data. Thus one may assume an underlying covariance structure and then estimate the covariance matrix based on the set of available stocks and then extract the related eigenvectors and eigenvalues from the estimated covariance matrix, which enables us to extract principal components to be used in the APT model. However, it should also be stressed that the issue of consistency need not be important in all applications. For example, suppose we are conducting an analysis of a set of variables observed over a finite time span, and this specific time span is the only time period of interest for the analysis. Then any sample statistic calculated from the dataset must be considered to be the true parameter. This is more or less what the original idea of Ross's APT model is about. The original theory concerns an analysis of a relatively short time period, so that the parameters may be considered fairly constant over time and all relevant assets are included in the dataset. Hence the risk of imposing unrealistic parametric restrictions on the covariance matrix may not be necessary. Instead, one may want to estimate the covariance matrix without assumptions of the functional form of the covariance structure, (say $\mathbf{S}$) and extract factors/principal components from $\mathbf{S}$. Unfortunately, it will not be possible to do so by traditional methods. For example, if the analysis is to be conducted in the $(n \times n)$ mode, the $\mathbf{S}_{(n \times n)}$ matrix will (unless the dataset is complete) contain empty cells and as a consequence factors or principal components cannot be obtained. There are however, approximation methods to extract the principal components directly without first estimating the covariance matrix. Surprisingly, these methods do not seem to have been appreciated in econometric contexts previously. In the next section we will investigate such a method that is of specific interest for the APT model application.

4 Model-independent extraction of principal components

Singular Value Decomposition (SVD) decomposes the structure of a data matrix $\mathbf{X}_{(T\times n)}$ of any size into (i) a set of orthonormal eigenvectors corresponding to $\mathbf{X}'\mathbf{X}$, (ii) a set of orthonormal eigenvectors corresponding to $\mathbf{X}\mathbf{X}'$, and (iii) a set of singular values (the square root of the related positive eigenvalues) of $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}\mathbf{X}'$, (Schott 2005). Thus, by using SVD it is possible to extract principal components and thereby summarize the data by a few linear combinations. However, when the data matrix is incomplete a number of problems arise, such as how to perform SVD on an incomplete dataset. The problem of extracting principal components when data are missing has been studied both in statistics and computer graphics (Gabriel and Zamir 2008; Chen 2002). Cristoffersson (1970) suggests a solution for estimating a one-component model when the data matrix is incomplete. He bases the extraction of the first principal component on iterative minimization of the sum of squared difference between the observed value and the predicted value of the one-component model over all complete observations. To solve the iterative minimization, a nonlinear iterative partial least squares algorithm (NIPALS) developed by Wold and Lyttkens (1969) is used. Later on, Ruhe (1974) proposed a more efficient minimization method for the case of the one-component PCA model, which was later extended by Wiberg (1976) to the case of extracting an arbitrary number of principal components when the data matrix is incomplete. It should be stressed that the extraction of the principal components is done without estimating the covariance matrix and due to this, the method requests that the number of principal components has to be determined in advance. The method is described as follows:

Case of complete data matrix: Suppose that we have a complete data matrix. Then, for an *r*-component model (where $r < Min\{n,T\}$) we may write:

$$\mathbf{X}_{(T\times n)} - \mathbf{1}\boldsymbol{\mu}' \approx \mathbf{P}_{(T\times T)}\mathbf{D}_{(T\times n)}\mathbf{Q}'_{(n\times n)} \tag{3.1}$$

where $\mathbf{PD}$ gives the principal components of interest and the singular values in $\mathbf{D}$ are given by the submatrix $\boldsymbol{\Lambda} = diag\left(\sqrt{\lambda_1},\cdots,\sqrt{\lambda_r}\right)$, where λ_i is the *i*-th eigenvalue. The approximation is due to the fact that all the singular values are not included. Hence,

the problem lies in determining the best rank *r* approximation of $\mathbf{X}$ by minimizing the Frobenius norm:

$$\underset{\mathbf{P},\mathbf{D},\mathbf{Q}'}{Min} \quad \left\|\left(\mathbf{X}-\mathbf{1}\boldsymbol{\mu}'\right)-\mathbf{PDQ}'\right\|_F^2 . \tag{3.2}$$

By rewriting (3.2) and substituting $\mathbf{Y}=\mathbf{X}-\mathbf{1}\hat{\boldsymbol{\mu}}'$, $\mathbf{U}=\mathbf{PD}$ and $\mathbf{V}'=\mathbf{Q}'$, we get:

$$\underset{\mathbf{U},\mathbf{V}}{Min} \quad \left\|\mathbf{Y}-\mathbf{U}\mathbf{V}'\right\|_F^2=\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{T}\left|\left(Y_{i,j}-\mathbf{u}_i'\mathbf{v}_j\right)\right|^2} \text{ where } \mathbf{U}=\begin{bmatrix}\mathbf{u}_1'\\ \vdots \\ \mathbf{u}_T'\end{bmatrix} \text{ and } \mathbf{V}=\begin{bmatrix}\mathbf{v}_1'\\ \vdots \\ \mathbf{v}_n'\end{bmatrix}. \tag{3.3}$$

Case of incomplete data matrix: If the data matrix is incomplete, we can rewrite (3.3) to:

$\underset{\mathbf{U},\mathbf{V}}{Min} \quad \left\|\boldsymbol{\Delta}\odot\left(\mathbf{Y}-\mathbf{U}\mathbf{V}'\right)\right\|_F^2$, where $\odot$ is the direct product and Δ is a matrix such that:

$$\boldsymbol{\Delta}_{i,j}=\begin{cases}1 & \text{if } X_{i,j} \text{ is observed}\\ 0 & \text{else}\end{cases} . \tag{3.4}$$

Hence, in case of an incomplete data matrix we need to minimize the following function:

$$\varphi\left(\mathbf{U},\mathbf{V}\right)=\frac{1}{2}\sum_{I}\left(X_{i,j}-\bar{x}_j-\mathbf{u}_i'\mathbf{v}_j\right)^2, \text{ where } I=\left\{\left(i,j\right):X_{i,j} \text{ is observed}\right\}. \tag{3.5}$$

In order to obtain a solution of (3.5) it is necessary that $m\geq r\left(n+T-r\right)+n$, where *m* is the number of observable elements in $\mathbf{X}$ and *r* is the number of extracted principal components. To get a full solution, the criteria $n>\sup\left\{\boldsymbol{\omega}_i\right\}_{i=1}^{n}$ also need to be met, where $\boldsymbol{\omega}_i$:= sequence of incomplete observations for the *i*-th stock. To minimize (3.5), it is necessary to either normalize the left matrix $\mathbf{U}$ or the right matrix $\mathbf{V}$, which enforces more constraints. By writing $\mathbf{X}$ as an *m*-dimensional vector $\mathbf{x}=vec(\mathbf{X}')$, we

may write the expression $X_{i,j} - \bar{x}_j - \mathbf{u}_i'\mathbf{v}_j$ as $f_i = x_i - \bar{x}_{p(i)} - \mathbf{u}_i'\mathbf{v}_j$ so that (3.5) becomes:

$$\arg\min_{\mathbf{v},\mathbf{u}} \varphi = \frac{1}{2}\mathbf{f}'\mathbf{f}, \tag{3.6}$$

where $\mathbf{f}$ is an $(m\times 1)$ vector and x_i is the $\left(f(i), p(i)\right)$-th component of $\mathbf{X}$. The vector $\mathbf{f}$ in (3.6) can be calculated as follows:

$$\mathbf{f} = \mathbf{y} - \mathbf{F}\mathbf{u} = \mathbf{y} - \mathbf{G}\mathbf{v}, \tag{3.7}$$

where the structures of matrixes $\mathbf{F}_{(m\times rT)}, \mathbf{G}_{(m\times rn)}$ are $\begin{bmatrix} \mathbf{v}_1' & & \\ \vdots & & \\ \mathbf{v}_n' & & \\ & \ddots & \\ & & \mathbf{v}_1' \\ & & \vdots \\ & & \mathbf{v}_n' \end{bmatrix}$ and $\begin{bmatrix} \mathbf{u}_1' & & \\ & \ddots & \\ & & \mathbf{u}_1' \\ & \vdots & \\ \mathbf{u}_T' & & \\ & \ddots & \\ & & \mathbf{u}_T' \end{bmatrix}$.

To solve the minimization problem (3.6) the partial derivatives with respect to $\mathbf{u}$ and $\mathbf{v}$ of the objective function should be zero and hence:

$$\min_{\mathbf{u},\mathbf{v}} \varphi(\mathbf{u},\mathbf{v}) = \begin{bmatrix} \mathbf{F}'\mathbf{F}\mathbf{u} - \mathbf{F}'\mathbf{y} \\ \mathbf{G}'\mathbf{G}\mathbf{v} - \mathbf{G}'\mathbf{y} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{3.8}$$

If the Λ matrix is fixed, then $\mathbf{F}$ is a function of $\mathbf{v}$ and $\mathbf{G}$ is a function of $\mathbf{u}$, thus (3.8) is a nonlinear equation. The nonlinear optimization problem can fortunately be linearized as follows: By treating either one of $\mathbf{u}$ or $\mathbf{v}$ as a fixed variable, let us say $\mathbf{v}$ is fixed, then $\mathbf{u}$ is treated as a function of the former, i.e., $\mathbf{u}(\mathbf{v})$, and the minimization problem (3.8) becomes:

$$\min_{\mathbf{v}} \varphi(\mathbf{u}(\mathbf{v}),\mathbf{v}) = \begin{bmatrix} \mathbf{F}'\mathbf{F}\mathbf{u}(\mathbf{v}) - \mathbf{F}'\mathbf{y} \\ \mathbf{G}'\mathbf{G}\mathbf{v} - \mathbf{G}'\mathbf{y} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{3.9}$$

Thus, the optimization problem is now linear and can easily be solved. One way to find the minimum of φ is to use Alternated Least Square (ALS), (Okatani and Deguchi 2007). Since our goal is to find the minimum of φ, we need to search for solutions to the two partial derivatives of φ with respect to $\mathbf{v}$ and $\mathbf{u}$, i.e., the solutions to:

$$\partial\varphi/\partial\mathbf{v} = \mathbf{F}'\mathbf{F}\mathbf{u} - \mathbf{F}'\mathbf{y} = 0 \text{ and } \partial\varphi/\partial\mathbf{u} = \mathbf{G}'\mathbf{G}\mathbf{v} - \mathbf{G}'\mathbf{y} = 0. \quad (3.10)$$

As the equations in (3.10) can be considered independent of each other, we can temporarily treat a subset of the parameters as fixed and solve for the other subset of parameters. This gives us the ordinary least square solution as follows:

$$\hat{\mathbf{u}} = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{y} \text{ and } \hat{\mathbf{v}} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{y}. \quad (3.11)$$

Thus, the ALS algorithm starts by choosing arbitrary initial values on either $\mathbf{u}$ or $\mathbf{v}$ and then solves for $\mathbf{u}$ or $\mathbf{v}$ in an alternating way. The iteration procedure stops when some convergence criteria is met. In the case where $\mathbf{F}'\mathbf{F}$ and/or $\mathbf{G}'\mathbf{G}$ is singular, then the inverses are calculated by using the Moore-Penrose inverse. I present below a flowchart for extracting principal components in the case of an incomplete data set:

ALS algorithm for estimating PC in an APT application:

1. Start by centering each asset in the dataset by its sample mean, i.e., let $\underset{(T\times n)}{\mathbf{Y}} = \underset{(T\times n)}{\mathbf{X}} - \underset{(T\times n)}{\mathbf{\Delta}} \odot \underset{(T\times 1)}{\mathbf{1}} \underset{(1\times n)}{\bar{\mathbf{x}}'}$.

2. Form the vector $\underset{(m\times 1)}{\mathbf{y}} = vec(\mathbf{Y})$. Note that this vector only consists of nonempty entries.

3. Decide the number of principal components (r) that should be extracted and the convergence criteria (for example, $\alpha > \left\|\mathbf{\Delta} \odot (\mathbf{Y} - \mathbf{U}\mathbf{V}')\right\|_F^2$).

4. Specify a start value $\underset{(rT\times 1)}{\hat{\mathbf{u}}}$ to generate $\underset{(rT\times 1)}{\mathbf{u}}^{(0)}$ (either a vector of a constant value or, preferably, a randomly generated value (ten Berge, 1993)).

5. Calculate $\underset{(rn\times 1)}{\hat{\mathbf{v}}} = \left(\underset{(rn\times m)}{\mathbf{G}'}\underset{(m\times rn)}{\mathbf{G}}\right)^{-1}\underset{(rn\times m)}{\mathbf{G}'}\underset{(m\times 1)}{\mathbf{y}}$; this gives $\underset{(rn\times 1)}{\hat{\mathbf{v}}}^{(i)}$.

6. Calculate $\underset{(rT\times 1)}{\hat{\mathbf{u}}} = \left(\underset{(rT\times m)}{\mathbf{F}'}\underset{(m\times rT)}{\mathbf{F}}\right)^{-1}\underset{(rT\times m)}{\mathbf{F}'}\underset{(m\times 1)}{\mathbf{y}}$; this gives $\underset{(rT\times 1)}{\hat{\mathbf{u}}}^{(i)}$.

7. If the convergence criteria are met then stop, otherwise go to step 5.

8. Finally, the estimated principal components (each with length *T*) are stacked in the vector $\underset{(rT\times 1)}{\hat{\mathbf{u}}}$.

In other words, the vector $\hat{\mathbf{u}}$ contains the principal components to be used as approximations of the unobservable factors of the APT model (2.1).

It should be stressed that this method is an model-independent approach which, unlike the traditional methods of dealing with missing observations such as the EM-algorithm, does not make any assumptions about distribution. The application for this model-independent approach stretches beyond the case of incomplete datasets since it

is also a method for imputing values in case of randomly missing observations. In order to demonstrate the potential of this method, we will apply the method on a real dataset in the following section.

5 An empirical application – The Stockholm exchange

The empirical analysis will consist of an APT modeling of the Stockholm stock exchange. The analysis involves the suggested method for model-independent extraction of principal components and a model dependent approach based on the EM algorithm for comparison. The data used is a sampled set from a full dataset consisting of monthly returns of 250 stocks listed at the Stockholm stock exchange from the database Compustat Global. It consists of 250 columns (stocks) and 228 rows (monthly returns) with the date ending in December 2006, giving a data matrix with a total of 57,000 entries. Of these 57,000 entries there are only 26,852 entries (about 47 %) containing observations and the rest of the entries are empty. Since the data base employs "the hot dec method" (i.e. if a stock was not traded during any sampling time point but was still listed on the stock market then this missing value is replaced by the most recent daily non-missing value), all the empty entries are because a company stock either entered the stock market at a later time point than $t=1$, ceased to be listed on the stock market at an earlier time point than $t=228$ or both. Further, only 14 of the 250 stocks have a complete set of observations and out of the 62,500 entries in the $(n\times n)$ covariance matrix, 1,784 have stock pairs without simultaneous observations, hence the $(n\times n)$ covariance matrix is not estimable in a traditional way. When empirically investigating the dataset, it was found that the elements in the $(T\times T)$ covariance matrix are not homogeneous along the diagonal or along any of the off-diagonals; further it did not have a diminishing absolute value of any of the rows, i.e., in any row *t* letting τ increase did not result in a diminishing value of $|\sigma_{t,\tau}|$. Hence, this disqualifies any solution based on assuming a covariance model, i.e., a specified structure of the covariance matrix.

To investigate whether the ratio between ordered eigenvalues changes as a function of the number of incomplete observations, the original dataset $\mathbf{X}_{(n\times T)}$ is rearranged into overlapping subsets of 20 stocks/row in each subset, i.e.,

$\mathbf{X}^{(1)} = \{\mathbf{x}_1\, \mathbf{x}_2 \cdots \mathbf{x}_{20}\}$, $\mathbf{X}^{(2)} = \{\mathbf{x}_{11}\, \mathbf{x}_{12} \cdots \mathbf{x}_{30}\}$, and so on, where the dataset is sorted by the number of incomplete observations for each stock, so that the first row of the dataset has the least number of incomplete observations and as one goes down along the row dimension, the number of incomplete observations increases. Each subset $\mathbf{X}_{(20\times T)}$ hence contains an overlap of ten stocks from the earlier subdataset and ten additional stocks, working like a moving window across the dataset. To investigate if the relations between eigenvalues are approximately the same for the different subsets, we extract the possible twenty eigenvalues (since $\text{Rank}(\mathbf{x}_{(20\times T)}) \leq 20$) from each subdataset and calculate the eigenvalue ratios λ_1/λ_9, λ_1/λ_{15}. This makes it possible to see if the relations between eigenvalues are the same for subsets with a few incomplete observations as for subsets with many.

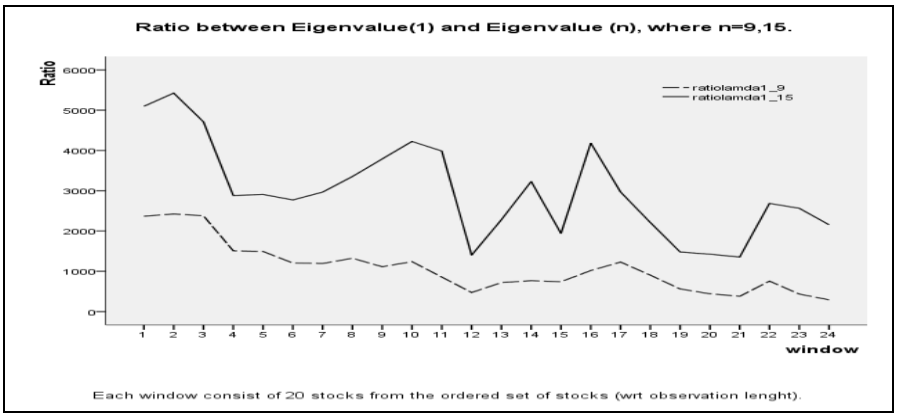

Figure 1: Eigenvalue ratios for the different subpopulations of stocks.

As seen in Figure 1, the eigenvalue ratios tend to decrease as one moves to the right along the x-axis, hence letting the subsets have more and more incomplete observations. This suggests that the information caught by the first principal component is slowly decreasing, hence becoming less informative. In other words, this suggests that the different subsets contain different amounts of information and that all available data should be included in the analysis, since the information is not the same across the different subsets of the data. If the analysis excludes variables with incomplete history of entries the result will be biased.

Another issue is how the number of observations increases if the dimension n increases, the dimension T increases, or both n and T simultaneously increase for the Swedish OMX data. Hence the question is how the stocks with a incomplete

history contribute to the overall number of observations. As we can see in Figure 2, the ratio $r = (\#\text{of non-empty off-diagonals})/(\#\text{diagonals})$ grows as we increase T so that $n = n(T)$. Note though that the rate is not constant, indicating that the proportion of incomplete variables is not equal during different time periods.

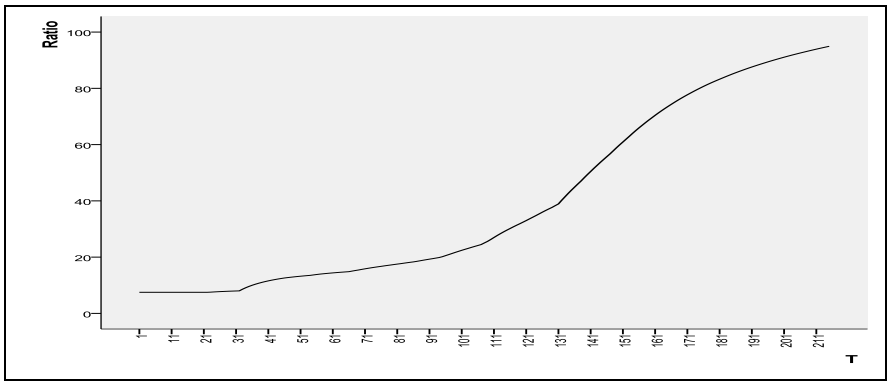

Figure 2: *T* growing and *n* is a function of *T*.

Estimation procedure. Due to computer limitations, the dataset is restricted to the dimension (190×185), where the assets included are chosen by taking a simple random sample out of the 250 existing assets and limiting the number of time observations to 190 with the same end point as before. This should be contrasted to the traditional APT analysis where the population is restricted to include only those stocks that have a complete history, giving a biased result. The sampled data set contains 46 % incomplete observations and out of the 34,225 entries in the $(n \times n)$ covariance matrix, 368 have stock pairs without simultaneous observations, so the $(n \times n)$ covariance matrix is not estimable in a traditional way. Next, the number of PCs that is to be extracted has to be determined. By extracting the eigenvalues of the $(T \times T)$ covariance matrix we are able to get a hint of how many PCs/factors should be included in the APT model. Further, the empirical specification of the APT model used here is based on the technique that was initially proposed by (Roll and Ross 1980) where they estimate factors and loadings in a first step and then use these loadings as explanatory variables in a second step and estimate the risk premium of each risk factor by performing generalized least squares cross-sectional regressions in each time point. Hence, the estimation procedure is as follows:

Step 1: For each asset a time series regression is performed, where the asset's return sensitivities to each factor realization (PC) is estimated. The time series regressions are of the form:

$$r_{i,t} - \overline{r}_i = \sum_{j=1}^{4} \beta_{i,j} PC_{j,t} + \varepsilon_{i,t}. \tag{3.12}$$

Step 2: For each time point the following cross-sectional regression (3.13) is estimated, were the estimated sensitivities are used as independent variables. The cross-sectional regressions are of the form:

$$r_{i,t} = \lambda_{0,t} + \sum_{j=1}^{4} \lambda_{j,t} \hat{\beta}_{i,j} + u_{i,t}. \tag{3.13}$$

Empirical result. The scree plot of the eigenvalues related to the cross-product matrix is studied with the purpose of getting a proxy of how many principal components are needed to be able to describe the data matrix sufficiently well. To solve the problem of negative eigenvalues, these are set to the value zero. This is due to the fact that each eigenvalue corresponds to each component variance and a negative eigenvalue would then imply that the component has a negative variance, which is not possible since variance is a squared quantity. The scree plot (Figure 3) shows a knee appearing after the fourth eigenvalue indicating that the dimension of the data matrix can be reduced to four principal components.

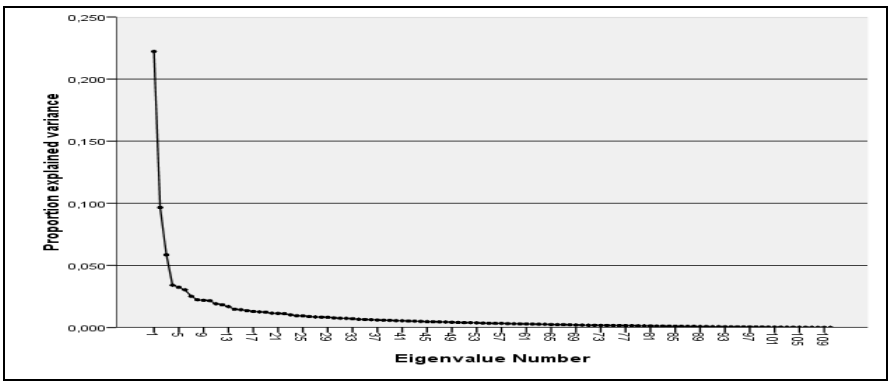

Figure 3: Scree plot of the cross-product matrix.

Based on the above analysis, the number of principal components is set to four in the further analysis and we can expect any model based on these principal components to have an upper bound of the explained variation close to 40 % of the total variation.

Estimation of factor sensitivities. In the first step each asset's returns are regressed on the four principal components extracted by the ALS algorithm as specified on earlier, where the estimated parameters are each asset's sensitivity to each factor/principal component. To see if there is any relation at all, an *F*-test is performed where the regression coefficients are simultaneously tested. Hence for each time series regression, i.e., for each asset ($i = 1, \ldots, 185$) we test:

$$H_0 : \beta_{i,1} = \beta_{i,2} = \beta_{i,3} = \beta_{i,4} = 0 \text{ against } H_A : \text{at least one } \beta_{i,j} \neq 0.$$

The *p*-value for each test is presented in the histogram below and the result is quite satisfactory, since most time series regressions have at least one coefficient significantly different from zero. Out of the 185 regressions, 80 % has a *p*-value less than or equal to 0.01, and about 92 % of them have a *p*-value less than or equal to 0.10. This implies that the model (3.12) has quite good explanatory power, which is also supported by the histogram of the coefficients of multiple determination in Figure 6. As we can see in the histogram below (Figure 6), the R^2 values are moderate and the arithmetic mean is equal to $\bar{R}^2 = 0.34$, suggesting that the models based on four factors explain on average 34 % of the variation in the returns. Since the number of factors is limited to four here, there is a possibility of getting higher R^2 values by including additional factors.

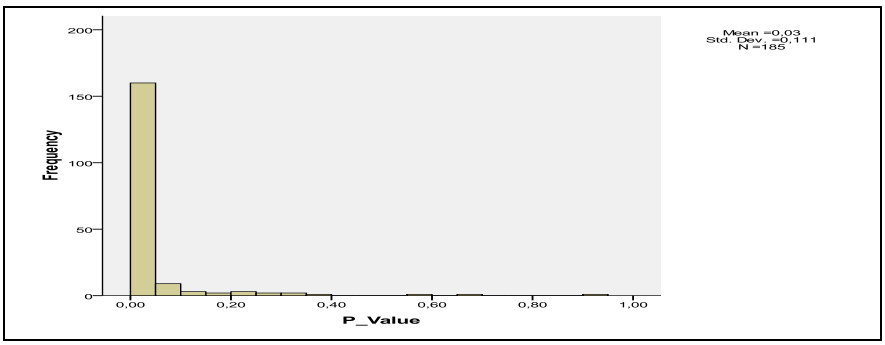

Figure 4: *P*-values for estimated sensitivities, ALS case.

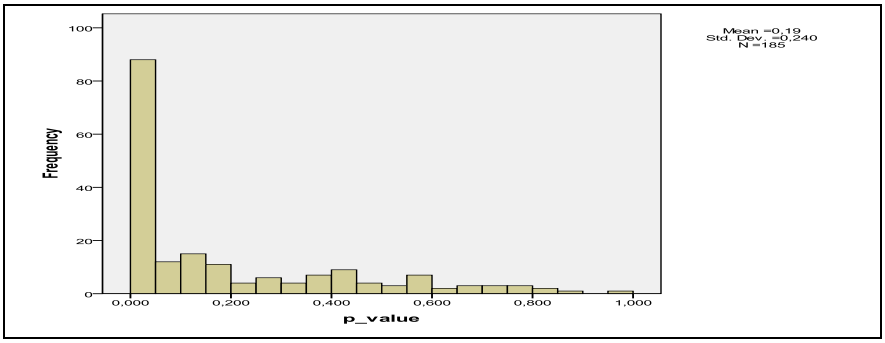

Figure 5: *P*-values for estimated sensitivities, EM case.

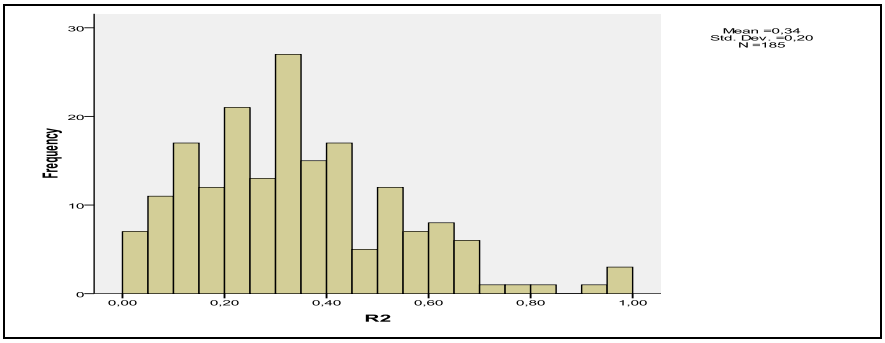

Figure 6: Coefficients of multiple determination for estimated sensitivities, ALS case.

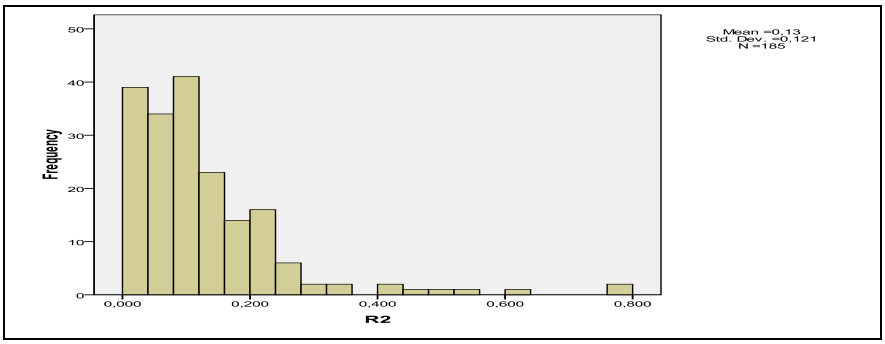

Figure 7: Coefficients of multiple determination for estimated sensitivities, EM case.

For the purpose of comparison, the analogous analysis is also performed using the EM algorithm based on the $t_{(4)}$ distribution. The Normal distribution was also considered but it is omitted from the presentation due to its poor performance. Since the EM algorithm failed to converge, it was stopped when it reached 50 iterations (allowing the maximum number of iterations to be much larger, e.g. 1000, 5000, resulted in divergence and therefore it was set to the typical maximum number of iterations of 50 which is used in many standard softwares). In this first step, each asset's returns is also regressed on the four extracted principal components from the imputed data matrix. Comparing Figure 4 and 5 above, the average *p*-value, increases from 0.03 to 0.19. Hence the number of significant time series regressions depreciates as we shift method from model-independent extraction of principal components to the model

dependent EM algorithm. Comparing Figure 6 and 7 also shows that the EM algorithm produced an average $\bar{R}^2$ of 0.13, as compared with 0.34 of the ALS algorithm. This indicates that the models based extraction explain on average less of the variation compared to the model free approach.

Estimation of factor risk premiums. In the second step, (3.13) is estimated as a cross-sectional regression in each of the 190 time points. The dependent variable is now defined by the excess return. For each cross-sectional regression (i.e., for each time point $t = 1,\ldots,190$), the following hypotheses are tested to see if there is any significant risk premiums:

$$H_0 : \lambda_{1,t} = \lambda_{2,t} = \lambda_{3,t} = \lambda_{4,t} = 0 \text{ against } H_A : \text{at least one } \lambda_{j,t} \neq 0 .$$

About 52 % of the cross sectional regression hypotheses were rejected at $\alpha = 0.01$, and if the size is increased to $\alpha = 0.10$, then approximately 69 % of the hypotheses were rejected.

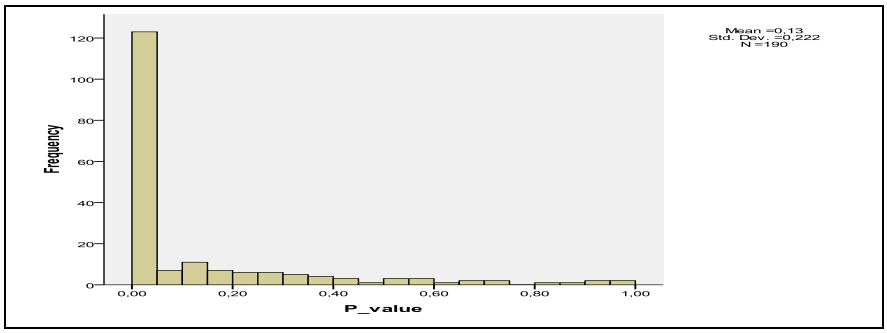

Figure 8: *P*-values for factor risk premiums, ALS case.

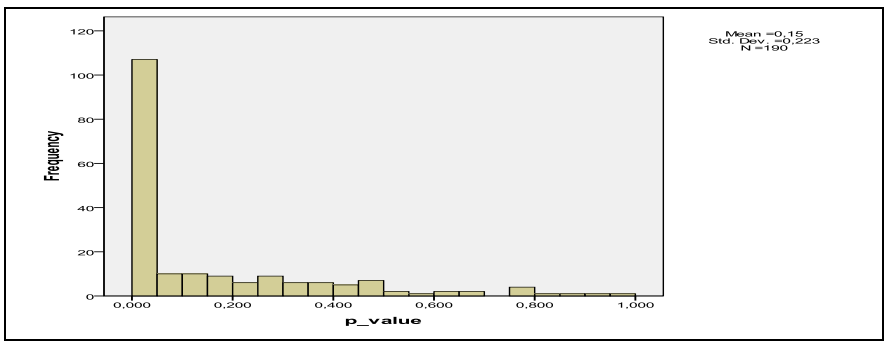

Figure 9: *P*-values for factor risk premiums, EM case.

The next question involves the extent to which the model (3.13) is able to explain the cross sectional variation of the excess returns. From the histogram of the coefficients of multiple determination (Figure 10), it is seen that the R^2 coefficients have a wide spread. About 50 % of the regressions has $R^2 < 0.22$. This implies that the APT model for the data at hand has a somewhat low explanatory power when the estimation of the APT model is performed as a cross-sectional regression.

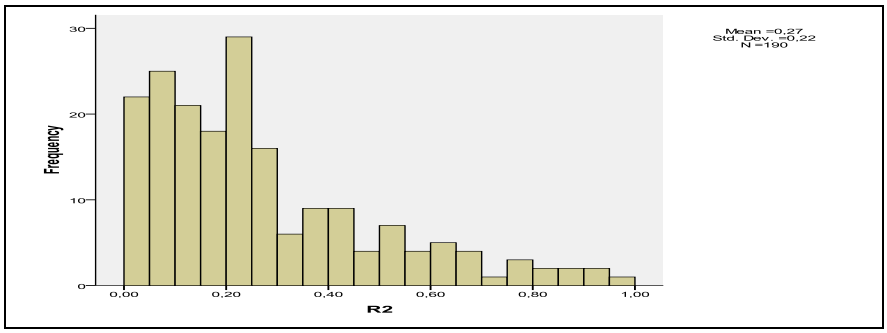

Figure 10: Coefficients of multiple determination for factor risk premiums, ALS case.

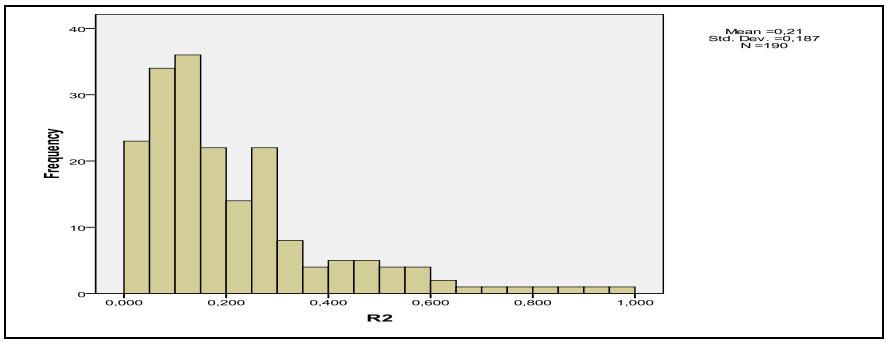

Figure 11: Coefficients of multiple determination for factor risk premiums, EM case.

Comparing the result above with that based on the EM algorithm, we find that the average *p*-values for the cross-sectional regressions are 0.13 for the ALS algorithm and 0.15 for the EM algorithm (Figures 8 and 9, respectively). The inferiority of the EM algorithm to the ALS algorithm is again reflected in Figures 10 and 11, where it is seen that the average $\bar{R}^2$ are 0.27 and 0.21, respectively. Hence, in this APT application with incomplete data, the model-independent extraction of principal components is in most senses superior to the EM algorithm approach.

6 Summary

In this paper I have considered the Incompleteness problem of the APT model. It is argued that since datasets of assets rarely have a complete history of observations in practice, the school-book approach to empirical APT analysis is not valid. In particular, the frequently-used approach to this problem by simply excluding assets without a complete history of observations will lead to an asymptotically empty set of zero observations and will in any fixed finite time period lead to a selection bias. The first fact follows almost trivially, while the bias part is demonstrated empirically. Also, as soon as a stock has been delisted from the stock exchange, there will not be any further observations. Hence these missing observations do not occur at random, which is an underlying assumption in the traditional imputation solutions such as the EM algorithm, and so other methods need to be considered instead. More specifically, we propose that in most practical situations, one should either estimate the covariance matrix (the main ingredient of the APT model) by a parametric model containing only a few parameters, or use some optimization algorithm for the data at hand. In particular, we demonstrate that band matrices may be consistently estimated with respect to a certain norm, regardless of whether the data matrix is incomplete and of increasing dimension. Moreover, it is also argued that non-parametric methods may be less risky, when compared to parametric approaches and that appropriate techniques for this purpose have already been developed, such as Wiberg's method, though they do not seem to have been used in this context previously. We also conduct a case study on the Stockholm OMX data with respect to the APT model. It is empirically shown that the method of only including assets with a long history will lead to a selection bias. I also demonstrated that the non-parametric optimization technique for extracting principal components works surprisingly well and should be implemented in all similar pricing models. Moreover, it is also empirically shown that the model-independent approach for extracting principal components outperforms the model approach based on the EM algorithm. The empirical specification of the APT model in this paper is based on the two-step technique that was initially proposed by Roll and Ross. We extract four factors and for the data at hand we find that a high proportion of the cross-sectional regressions are significant and hence the extracted risk factors are priced.

References

ten Berge, J. M. F. (1993). Least Squares Optimization in Multivariate Analysis. Leiden, The Nederlands: DSWO Press.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets. Econometrica, 51(5), 1281-1304.

Chen, H. (2002). Principal Component Analysis with Missing Data and Outliers. Technical Report. Electrical and Computer Engineering Department, Rutgers University, Piscataway, NJ.

Chen, N. (1983). Some Empirical Tests of the Theory of Arbitrage Pricing. The Journal of Finance, 38(5), 1393-1414.

Christofferson, A. (1970). The one component model with incomplete data. Ph.D. Thesis, Institute of Statistics. Uppsala University, Uppsala.

Connor, G. and Korajczyk, R. A. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. Journal of Financial Economics, 15(3), 373-394.

Connor, G. and Korajczyk, R. A. (1988). Risk and return in an equilibrium APT: Application of a new test methodology. Journal of Financial Economics, 21(2), 255-289.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society, Series B, 39(1), 1-38.

Faff, R. W. (1988). An Empirical Test of the Arbitrage Pricing Theory on Australian Stock Returns. 1974-85. Accounting and Finance, 28(2), 23-43.

Faff, R. W. (1992). A Multivariate Test of an Equilibrium APT with Time Varying Risk Premia in the Australian Equity Market. Australian Journal of Management, 17(2), 233-259.

Gabriel, K. R. and Zamir, S. (2008). Lower Rank Approximation of Matrices by Least Squares with any Choice of Weights. Technometrics, 21(4), 489-498.

Hansen, L. P. and Jagannathan, R. K. (1997). Assessing Specification Errors in Stochastic Discount Factor Models. Journal of Finance, 52(2), 557-590.

Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. Review of Economics and Statistics, 47(1), 13-37.

Markowitz, H. (1952). Portfolio Selection. Journal of Finance, 7(1), 77-91.

McElroy, M. B. and Burmeister, E. (1988). Arbitrage Pricing Theory as Restricted Nonlinear Multivariate Regression Model: Seemingly Unrelated Regression Estimates. Journal of Business & Economic Statistics, 6(1), 29-42.

Mossin, J. (1966). Equilibrium in a Capital Asset Market. Econometrica, 34(4), 768-783.

Okatani, T. and Deguchi, K. (2007). On the Wiberg Algorithm for Matrix Factorization in the Presence of Missing Components. International Journal of Computer Vision, 72(3), 329–337.

Roll, R. and Ross, S. A. (1980). An Empirical Investigation of the Arbitrage Pricing Theory. Journal of Finance, 35(5), 1073-1103.

Ross, S. A. (1976). The Arbitrage Theory of Capital Asset Pricing. Journal of Economic Theory, 13(3), 341-360.

Ruhe, A. (1974). Numerical computation of principal components when several observations are missing. Technical report. UMINF-48-74, Dept. Information Processing, Umeå University, Umeå.

Rydberg, T. H., (2000). Realistic statistical modelling of financial data. International Statistical Review, 68(3), 233-258.

Schott, J. R. (2005). Matrix Analysis for Statistics, second edition. New York: Wiley-Interscience.

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. Journal of Finance, 19(3), 425-442.

Shukla, R. and Trzcinka, C. (1990). Sequential Tests of the Arbitrage Pricing Theory: A Comparison of Principal Components and Maximum Likelihood Factors. Journal of Finance, 45(5), 1541-1564.

Wiberg, T. (1976). Computation of principal components when data are missing. Proc. Second Symp. Computational Statistics, 229–236.

Wold, H. and Lyttkens, E. (1969). Nonlinear iterative partial least squares (NIPALS) estimation procedures. Bulletin of the International Statistical Institute, 43, 29–51.

Appendix A

Proof of Theorem 1:

Restricting ourselves to the case $r=1$, expansion of (2.1) yields:

$$\ell^{0}_{r=1}\left(\hat{\boldsymbol{\Sigma}},\boldsymbol{\Sigma}\right)=\left(1/(2T-1)\right)\left\{\sum_{t=1}^{T}\left(S^{2}-\sigma^{2}\right)^{2}+\sum_{t=1}^{T-1}\left(S_{1}-\sigma_{1}\right)^{2}\right\}=$$

$$\frac{T\left(S^{2}-\sigma^{2}\right)^{2}+(T-1)\left(S_{1}-\sigma_{1}\right)^{2}}{2T-1}=\frac{\left(S^{2}-\sigma^{2}\right)^{2}}{2-(1/T)}+\frac{1-(1/T)}{2-(1/T)}\left(S_{1}-\sigma_{1}\right)^{2}. \qquad (4.1)$$

To prove that (4.1) converges in probability to zero we first calculate the expected value:

$$E\left[S^{2}-\sigma^{2}\right]=E\left[\sum_{t=1}^{T}w_{t}S_{0,t}^{2}\right]-\sigma^{2}=\sum_{t=1}^{T}w_{t}E\left[S_{0,t}^{2}\right]-\sigma^{2}=\sigma^{2}\sum_{t=1}^{T}w_{t}-\sigma^{2}=0.$$

$$E\left[S_{1}-\sigma_{1}\right]=E\left[\sum_{t=1}^{T-1}w_{t,t+1}S_{1,t}^{2}\right]-\sigma_{1}=\sum_{t=1}^{T-1}w_{t,t+1}E\left[S_{1,t}^{2}\right]-\sigma_{1}=\sigma^{2}\sum_{t=1}^{T-1}w_{t,t+1}-\sigma_{1}=0.$$

To investigate the convergence of the variance of (4.1) we use an upper bound of the variance, which can be easily proved. The variance of (4.1) is bounded above by:

$$Var\left[\hat{\sigma}^{2}\right]\leq\frac{k-\sigma^{4}}{T}+\frac{3\sigma^{4}-k}{T\,\mathrm{Min}\left(\left\{n_{t}\right\}_{t=1}^{T}\right)}+4\sigma_{1}^{2}\frac{T-1}{T^{2}\,\mathrm{Min}\left(\left\{n_{t}\right\}_{t=1}^{T}\right)}\to 0 \text{ as } \begin{matrix} n\to\infty \\ T\to\infty \end{matrix}, \text{ and}$$

$Tn^{-1}\to c,\ 0\leq c<k<\infty,$ and for the first off-diagonal:

$$V\left[S_{1}\right]\leq\frac{\left(k'-\sigma_{1}^{2}\right)}{(T-1)}+\frac{\left(\sigma^{4}-k'+2\sigma_{1}^{2}\right)}{(T-1)\,\mathrm{Min}\left[\left\{n_{t,t+1}\right\}_{t=1}^{T-1}\right]}+\frac{2(T-2)\left(k''-\sigma_{1}^{2}\right)}{(T-1)^{2}}+$$

$$\frac{2(T-2)\left(2\sigma_{1}^{2}-k''\right)}{(T-1)^{2}\,\mathrm{Min}\left[\left\{n_{t,t+1}\right\}_{t=1}^{T-1}\right]}\to 0 \text{ as } \begin{matrix} n\to\infty \\ T\to\infty \end{matrix} \text{ and } Tn^{-1}\to c, \quad \begin{matrix} 0\leq c<k'<\infty \\ 0\leq c<k''<\infty \end{matrix}.$$

For the case of complete data, the inequality above reduces to:

$$V\left[S_1-\sigma_1\right]=V\left[S_1\right]\le\frac{\left(k'-\sigma_1^2\right)}{\left(T-1\right)}+\frac{\left(\sigma^4-k'+2\sigma_1^2\right)}{\left(T-1\right)n}+\frac{2\left(T-2\right)\left(k''-\sigma_1^2\right)}{\left(T-1\right)^2}+$$

$$\frac{2\left(T-2\right)\left(2\sigma_1^2-k''\right)}{\left(T-1\right)^2 n}\to 0 \text{ as } \begin{matrix} n\to\infty \\ T\to\infty \end{matrix} \text{ and } Tn^{-1}\to c, \begin{matrix} 0\le c<k'<\infty \\ 0\le c<k''<\infty \end{matrix}.$$

Hence, both bias and variance of S^2 and S_1 converge to zero so that $\operatorname{plim}_{T\to\infty}\ell^0_{r=1}\left(\hat{\mathbf{\Sigma}},\mathbf{\Sigma}\right)=0$, uniformly in T. A similar proof may be given for an arbitrary fixed r and this completes the proof of Theorem 1. □