The University of Reading Business School
Department of Real Estate & Planning

Working Papers in Real Estate & Planning 12/06

The University of Reading

The Underlying Return Generating Factors for REIT Returns: An Application of Independent Component Analysis

Colin Lizieri (The University of Reading Business School)*

Stephen Satchell (Faculty of Economics, University of Cambridge)

Qi Zhang (Faculty of Economics, University of Cambridge).

Author for correspondence:

Professor Colin Lizieri
Department of Real Estate & Planning
University of Reading Business School
Whiteknights, Reading, RG66AW, UK
Tel +44 118 378 6339
Email: c.m.lizieri@reading.ac.uk

The Underlying Return Generating Factors for REIT Returns: An Application of Independent Component Analysis

Abstract:

Multi-factor approaches to analysis of real estate returns have, since the pioneering work of Chan, Hendershott and Sanders (1990), emphasised a macro-variables approach in preference to the latent factor approach that formed the original basis of the arbitrage pricing theory. With increasing use of high frequency data and trading strategies and with a growing emphasis on the risks of extreme events, the macro-variable procedure has some deficiencies. This paper explores a third way, with the use of an alternative to the standard principal components approach – independent components analysis (ICA). ICA seeks higher moment independence and maximises in relation to a chosen risk parameter. We apply an ICA based on kurtosis maximisation to weekly US REIT data using a kurtosis maximising algorithm. The results show that ICA is successful in capturing the kurtosis characteristics of REIT returns, offering possibilities for the development of risk management strategies that are sensitive to extreme events and tail distributions.

1. Introduction

The role of real estate in mixed asset portfolios depends on the return generating process when compared to other asset classes. Analysis in private real estate markets is badly hampered by data inadequacies, but the existence of traded real estate securities – REITs – enables factor models to be tested in real estate markets. As with common equities, single factor models and extended-CAPM models have proved limited in characterising risk-return characteristics and attention has switched to multi-factor models developed from the Arbitrage Pricing Theory framework. Early applications utilised the original direct principal components analysis (PCA) approach of Roll and Ross but increasingly have used macro-factor models. These are easier to interpret but leave questions concerning independence and missing variable specification problems. As a parallel development, greater attention has focussed on the distributional qualities of asset returns and, in particular, persistent evidence of non-normality in returns. This paper examines equity REIT data from the US to investigate whether non-normality results in the relatively weak performance of PCA relative to multi-factor models and introduces Independent Component Analysis – a modelling strategy that is more sensitive to the presence of kurtosis.

This paper is in the spirit of Chan, Hendershott and Sanders (1990) who provided an early application of multi factor analysis, examining EREIT returns using CAPM and APT approaches, with the latter based on a mimicking portfolio model, based around the Chen, Roll and Ross (1986) macro-factors. Chen *et al.* (1997) similarly compared macro-variable multi-factor models with derived, principal component models for REIT returns over the period 1974-1991. They suggest that the macro-variable model outperforms a factor loading model over the 1980-1995 period. Subsequent work has tended to utilise macro-factor approaches (see for example Ling, Naranjo and Ryngaert, 2000).

There are three principal reasons why it might be valuable to revisit the statistical approach. The first relates to the evolution of investment strategies that rely on active trading on a frequent basis. Macro-factor models require proxies for the key priced systematic risk factors, but these are rarely available at high frequency and are frequently subject to revision (notably macro-economic data). Given high frequency trading models, it is thus necessary to explore statistical procedures that may be used to inform investment policy and strategy. The second results from the interaction of the principal components analysis model and the observed distribution of real estate returns. Third, risk management strategies are increasingly focused on extreme events rather than general volatility.

Relative underperformance of the principal components approach may, in part, result from the nature of REIT return distributions. Non-normality is a feature that is evident in commercial real estate returns (noted a decade ago, for example, in Young & Graff (1995) for a review, see Lizieri & Ward, 2000) and is a characteristic of equity returns, with fat tails being commonplace. Given that principal components analysis focuses only on the first and second moments of the distribution, then standard PCA approaches may neglect information that is available in asset returns.

Principal components analysis attempts to capture as much variance as possible in the underlying data set. However, if variance is insufficient as a risk measure, one might prefer to maximise with respect to some other risk metric. Unfortunately, many risk measures, such as Value at Risk, Expected Loss or Semi-Variance, are not amenable to standard optimisation techniques, as they measure some aspects of joint and marginal tail distributions (for real estate examples see Bond & Patel, 2003;

Hamelink & Hoesli, 2004; Knight *et al.,* 2005). However, measures based on higher moments are somewhat more amenable for optimisation – hence our focus on kurtosis in this paper.

Recently, new procedures have emerged that attempt to handle these issues. One such procedure is Independent Component Analysis (ICA). Independent components analysis is a procedure analogous to principal components but based on higher moments than the second. It is one of a family of somewhat disparate techniques; we choose this particular approach – a kurtosis maximisation procedure - because of its applicability to portfolio construction, its ability to deal with large portfolios, and because of structural comparability with principal components analysis.

The algorithm we use iterates between orthogonalisation (enforcing zero correlation) followed by maximization of non-Gaussianity - in our case, maximization of kurtosis. We need to maximize kurtosis because the solution to zero correlation is not unique (it is 'rotationally invariant'). By maximizing kurtosis, we hope to find components which are as independent as possible. The model allows us to test whether kurtosis maximization indeed gives us independent components. By bootstrapping the variance of covariance, we test the null hypothesis that $cov\{g(S_i),h(S_j)\}=0$, where $g()$ and $h()$ are nonlinear transformations of S, and where the S refer to different components. We can reorder the independent components based on kurtosis and calculate what proportion of total kurtosis is explained by the first i independent components. This generalizes the variance decomposition used in PCA. For ICA, we tend to find that there is no value at which the plot of kurt(i) against i flattens out.

There are a number of practical implications of the model. For example, it is hoped that such calculations may find use in portable alpha applications by concentrating the kurtosis in a particular portfolio. This paper, however, will concentrate more on methodological and empirical issues rather than applications to fund management. We apply the ICA model to a sample of US REIT returns, comparing the results with the standard PCA approach. The next section briefly reviews some of the existing literature on multi-factor models in real estate. We then introduce the independent component analysis method. The fourth section provides details of the data set. We then present results and draw conclusions.

2. Multi-factor Models in Real Estate

The growth of the REIT market in the 1980s allowed researchers to explore the use of arbitrage pricing and multi-factor models in real estate. An early application of APT modelling was provided by Titman and Warga (1986) who use factor analytic procedures in an attempt to link NYSE and AMEX risk factors to REIT returns. In parallel, the principles of application were laid out in Blundell & Ward (1987). The paper by Chan, Hendershott and Sanders (1990) represents a significant step forward and lead to a focus on macro-economic or fundamental factors. The authors examine equity REIT returns using both CAPM and APT approaches, with the latter based on a mimicking portfolio model, based around the Chen, Roll and Ross (1986) macro-factors. They demonstrate that the excess returns to REITs observed using a simple CAPM framework "evaporate" when a multi-factor approach is used. They identify three key factors driving both equity market and REIT returns: changes in the risk and term structures and unexpected inflation. Other macro factors such as forward industrial production and expected inflation have lesser impact or switch between positive and negative influence. The impacts of macro factors on real estate returns are consistently lower than the impacts on general stock returns.

Chen *et al.* (1997), following the work of Chen & Jordan (1993), compare macro-variable multi-factor models with derived, principal component (factor loading) models for REIT returns over the period 1974-1991. Their macro-variable model utilises five factors – change in term structure, change in risk premium, the unanticipated inflation rate, changes in expected inflation and an orthogonalised market index residual. They suggest that the macro-variable model outperforms a factor loading model over the 1980-1995 period.

Subsequent work has tended to utilise macro-factor approaches. For example, Ling & Naranjo (1997) use a multi-factor model that includes the growth rate of real consumption, the term structure of interest rates, the real rate of interest, the term structure and inflation shocks to test the behaviour of real estate returns. They allow for time varying risk factors and point to the importance of the consumption variable as casting doubt on prior findings of abnormal real estate returns. Other applications include Ling & Naranjo (1999) and Ling, Naranjo and Ryngaert (2000). The emphasis on macro-variables has been reinforced by the use of long-run, cointegration methods

to examine the links between real estate and the equity market (for example, McCue & Kling, 1994; Okunev & Wilson, 1997, Brooks & Tsolacos, 1999).

While there are clear benefits from use of macro-variables, there are technical issues concerned with the derivation of innovations (for example, identification of inflation shocks requires a model of anticipated inflation) and with the inter-relationship between the identified risk factors. More significantly, the use of a macro factor approach may constrain the frequency of analysis which is dependent on the release of official statistics (inflation, industrial output, for example) and compromised by the frequent revision of official statistics. Given the move towards risk-management strategies based on higher frequency data and new investment vehicles and products that emphasise higher moments, there may be advantages in using an analytic approach that focuses on extreme events and implicit factors.

3. Independent Component Analysis

3.1 ICA Principles

Independent component analysis is a technique that aims to extract distinct signals from some generalised, commingled distribution. One way in which ICA principles are described is the so-called "cocktail-party" problem. Imagine there are two persons speaking simultaneously situated at different positions in a cocktail party. We want to know their original comments but can only hear combined noise. We recorded the two time signals We recorded the two time signals $x_1(t)$ and $x_2(t)$; they are the weighted sums of the original signal $s_1(t)$ and $s_2(t)$ with weight a_{ij}, where ij=1,2. To express it in linear equations: Expressed in linear equations:

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) \tag{1}$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) \tag{2}$$

It will be shown that we can solve for the original signals $s_1(t)$ and $s_2(t)$ solely by using the principles of statistical independence, with no further assumption needed to guarantee that the model can be estimated.

As will be demonstrated, we can apply this idea to the analysis of REITs. We observed the return series $x_n(t)$ for n REITs, and are interested in the underlying return generating factors $s_j(t)$. In mathematical notation:

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) + ... + a_{1j}s_j(t) = \sum_{i=1}^{j} a_{1i}s_i(t) \qquad (3)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) + ... + a_{2j}s_j(t) = \sum_{i=1}^{j} a_{2i}s_i(t) \qquad (4)$$

$$x_n(t) = a_{n1}s_1(t) + a_{n2}s_2(t) + ... + a_{nj}s_j(t) = \sum_{i=1}^{j} a_{ni}s_i(t) \qquad (5)$$

where $x_n(t)$ is the observed return for the nth REITs at time t; $s_1(t)$, $s_2(t)$ and $s_j(t)$ are the independent underlying factors at time t, and a_{nj} is the mixing matrix.

Even though we can use the assumption of independence to estimate the signals, there remain two ambiguities in the ICA model. First, we cannot determine the variances of the independent components because we can multiply any signal $s_j(t)$ by a scalar and divide the corresponding weight a_{ij} by the same scalar, while the observed signals remain unchanged. In this way, we can change the magnitude of the signal variances. To solve this, ICA assumes that E $\{s_j{}^2\}$=1, although this still leaves a problem of sign ambiguity. Second, we cannot determine the order of the independent components. We can change the lines of the weighting matrix and the order of independent components without changing the model.

It is more convenient to express the model in matrix notation:

$$X = AS \qquad (6)$$

Where $X^T = (x_1, x_2, ..., x_n)$, $A^T = (a_{1j}, a_{2j}, ..., a_{nj})$, $S^T = (s_1, s_2, ...s_j)$.

To estimate the independent components, we need to express them in terms of the observed data X. Denote this as $\hat{S} = \hat{W}X$, where $\hat{S}$ are the estimated independent

components. Obviously if $\hat{W}$ is the inverse of A, we can achieve $\hat{S}$ which is exactly the same as the true signal S. However, given the fact that we have no prior information about A, we cannot calculate $\hat{W}$ directly. We need to solve using the principle of statistical independence to find a close approximation. To estimate independence, we introduce a simple and intuitive principle: maximization of non-normality (non-Gaussianity). The central limit theorem says that the sum of N independent random variables tends toward a Gaussian distribution if N is large. In the case of ICA, it means that the sum of two independent components is more Gaussian than any of the two original signals.

By substituting X=AS, we can express $\hat{S}$ as a linear combination of S :

$$\hat{S} = \hat{W}X = \hat{W}AS = Q^T S \qquad (7)$$

where $Q^T = \hat{W}A$. Hyvarinen and Oja (2001) used the two signals case to explain how to apply the central limit theorem to achieve the independent components:

$$\hat{s}_1 = w_{11}x_1 + w_{12}x_2 = q_{11}s_1 + q_{12}s_2 = \sum_{i=1}^{2} q_{1i}s_i \qquad (8)$$

$$\hat{s}_2 = w_{21}x_1 + w_{22}x_2 = q_{21}s_1 + q_{22}s_2 = \sum_{i=1}^{2} q_{2i}s_i \qquad (9)$$

Since the sum of two independent components follows a "more Gaussian" distribution than the original components, $\hat{s}_1 = \sum_{i=1}^{2} q_{1i}s_i$ is "more Gaussian" than s_1 and is least Gaussian when $\hat{s}_1$ is actually equal to s_1; the same holds for $\hat{s}_2$[1]. Given the fact that we only have information for X, we can try different values for $\hat{W}$ and compare the distribution of $\hat{S} = \hat{W}X$. More precisely, we can solve ICA by maximizing the non-gaussianity of $\hat{S} = \hat{W}X$ by estimating the weights $\hat{W}$. This transforms the ICA problem to a numerical optimization problem.

---

[1] It can also be shown that minimizing mutual information is equivalent to maximizing nongaussianity. Hence, maximizing nongaussianity can achieve maximally independent components.

We can solve ICA on the basis of minimizing or maximizing certain contrast functions. There are several potential measures of non-gaussianity: for instance, negentropy, information theory, kurtosis. The negentropy of a random variable is the amount of information that can be captured by observing the variables: more unpredictable and unstructured data will have large entropy. Since the normal distribution is the least structured of all distributions, random variables with a normal distribution will have the largest negentropy; random variables with a non-gaussian distribution will have a lower negentropy. We can achieve independent components by minimizing negentropy. We can also maximize the independence by minimising the mutual information between the components. Theoretically in terms of statistical properties, they are good measures of non-normality; however, computationally they are rather difficult to use because we need to estimate the density function for negentropy estimating.

In econometrics, however, it is standard to measure normality with reference to kurtosis, which is a classic quantitative measure of non-normality, embedded in standard normality tests such as the Jarque-Bera statistic:

$$kurt(y) = E\{y^4\} - 3(E\{y^2\})^2 \tag{10}$$

The formula above has been simplified to account for zero-mean case. Furthermore, we approximate kurtosis by the fourth central moment. The excess kurtosis can be either positive or negative: if excess kurtosis is positive, it is supergaussian (or leptokurtic); if negative, it is subgaussian (or platykurtic); if it is zero, we get the gaussian distribution. In this paper, we will use the absolute value or squared value of kurtosis as a measure of non-gaussianity.

3.2 The ICA Algorithm

In what follows, we draw on insights from Hyvarinen *et al.* (2001). As discussed above, to find the independent components, we need to maximize the non-gaussianity of $\hat{S} = \hat{W}X$ by estimating the weights $\hat{W}$. In practice, we start from some weighting matrix $\hat{W}$, compute the direction in which the absolute value of kurtosis of X is

increasing most strongly, and use a gradient method to find a new $\hat{W}$. This is an iterative algorithm. To simplify and reduce the complexity initial pre-processing steps is conducted: for example, centring, whitening and dimensionality reduction.

The observed variables are centred by subtraction of their sample means. Thus, the signals are also zero mean by taking expectations on both sides of equation (7). The mixing matrix $\hat{W}$ is unaffected by this pre-processing. We can add W*E(X) to the centred estimates of S to obtain the original signals.

In order to whiten the data, we need to apply a linear transformation V on X such that the transformed data Z=VX are uncorrelated and have unit variance. This, again, reduces the complexity of the ICA.

$$Z=VX=VAS=\tilde{A}S \qquad\qquad (11)$$
$$E\{ZZ^T\}=I \qquad\qquad (12)$$

To whiten, we can use the eigenvalue decomposition of the covariance matrix:

$$E\{XX^T\}=EDE^T \qquad\qquad (13)$$

Where E is the orthogonal matrix of eigenvectors of E $\{XX^T\}$ and D is the diagonal matrix of its eigenvalues. Thus,

$$V=ED^{-1/2}E^T \qquad\qquad (14)$$

Where $D^{-1/2}=diag(d_1^{-1/2},d_2^{-1/2},\dots,d_n^{-1/2})$. It is easily to prove that[2]:

$$Var(Z)=E(ZZ^T)=E(VXX^TV^T)=I \qquad\qquad (15)$$

On the other hand, from equation (11), we obtain a new mixing matrix $\tilde{A}=VA$. As the variance of signals S and variable Z are constrained to unity, for whitened data, this is equivalent to constraining the norm of $\tilde{A}$ to be the unit matrix:

$$E(ZZ^T)=E(\tilde{A}SS^T\tilde{A}^T)=\tilde{A}\tilde{A}^T=I \qquad\qquad (16)$$

---

[2] We can also set $V=D^{-1/2}E^T$ and still get var(Z)=var(VX)=I. Whitening can only give up an orthogonal transformation. Any orthogonal transformation of Z can achieve unit variance.

9

As $\tilde{A}$ is an orthogonal matrix, we only need to estimate up to n(n-1)/2 degrees of freedom rather than n^2, considerably reducing estimation problems.

It is also useful to reduce the complexity further; a third common pre-processing step is to reduce the dimensionality of the data by retaining only larger eigenvalues in the covariance matrix. Whitening and dimension reduction can be achieved with principal components analysis.

However, this whitening process only determines that the components are uncorrelated. For non-gaussian data, independence is a much stronger assumption than lack of correlation. Thus, we can apply PCA to do most of the pre-processing and then perform ICA on the most important principal components to achieve independence.

As noted above, we use the principle of maximization of non-gaussianity to estimate independence. Here, we use kurtosis (as defined in equation 10, above) as a measure of non-gaussianity. We can solve ICA by maximizing the absolute value of kurtosis of $\hat{S} = \hat{W}X$ by estimating the weights $\hat{W}$. We test different values for $\hat{W}$ so as to maximize the absolute value of kurtosis for S. This transforms the ICA problem to a numerical optimization problem. The algorithm is based on the gradients of the cost function.

First, we show how to estimate one independent component:

$$\hat{s}_1 = \hat{w}_{11}x_1 + \hat{w}_{12}x_2 + ... + \hat{w}_{1n}x_n = \hat{w}_1^T X \tag{17}$$

where $\hat{w}_1^T = (\hat{w}_{11}, \hat{w}_{12}, ..., \hat{w}_{1n})$, $X^T = (x_1, x_2, ..., x_n)$.

Starting from some initial weighting vector $\hat{w}_1(0)$, we find the direction in which the kurtosis of $\hat{s}_1 = \hat{w}_1^T X$ is increasing most strongly if kurtosis is positive, or decreasing most strongly if kurtosis is negative, based on the observable data. A gradient method is then used to find a new vector $\hat{w}_1$. So we are trying to find the direction of $\hat{w}_1$ so that the projection $\hat{w}_1^T X$ maximizes the non-gaussianity, as measured by the absolute value of kurtosis.

The pre-processing steps, such as centering and whitening, guarantee that the variance of $\hat{w}_1^T X$ and the norm of $\hat{w}_1$ are constrained to unity. This transfers the ICA problem to an optimization problem with the constraint of optimizing on the unit circle:

$$Max\left|kurt(\hat{s}_1)\right| = Max\left|E\{\hat{w}_1^T X\}^4 - 3[E\{\hat{w}_1^T X\}^2]^2\right| \tag{18}$$
$$\text{subject to } \left\|\hat{w}_1\right\| = 1$$

$$\frac{\partial\left|kurt(\hat{w}_1^T X)\right|}{\partial\hat{w}_1} = 4sign(kurt(\hat{w}_1^T X))[E\{X(\hat{w}_1^T X)^3\} - 3\hat{w}_1\left\|\hat{w}_1\right\|^2] = 0 \tag{19}$$

We can form a Lagrangian function:

$$L(\hat{w}_1, \lambda_1) = \left|kurt(\hat{w}_1^T X)\right| + \lambda_1(\left\|\hat{w}_1\right\| - 1)$$
(20)

The gradients with respect to $\hat{w}_1$ and λ_1:

$$\frac{\partial L(\hat{w}_1^T, \lambda_1)}{\partial\hat{w}_1} = 4sign(kurt(\hat{w}_1^T X))[E\{X(\hat{w}_1^T X)^3\} - 3\hat{w}_1\left\|\hat{w}_1\right\|^2] + 2\lambda_1\hat{w}_1 \tag{21}$$

$$\frac{\partial L(\hat{w}_1^T, \lambda_1)}{\partial\lambda_1} = \left\|\hat{w}_1\right\| - 1 \tag{22}$$

We get the solution to the constrained optimization problem when the gradients are zero with respect to both $\hat{w}_1$ and λ_1. We need to start from some initial value of $\hat{w}_1$ and λ_1, then apply Newton iteration or other iteration methods to solve the set of equations.

If the constraint is simple, for example that some quadratic form of $\hat{w}_1$ is constant, we can apply another optimization technique: projections on the constraint set. In our case, we face the simple constraint that the norm of $\hat{w}_1$ is equal to 1. Thus we can use the projection method. This means that the maximization problem can be solved with an unconstrained learning rule; and after each iteration, we need to project $\hat{w}_1$ onto the unit sphere, which means dividing $\hat{w}_1$ by its norm. This is a two-step process:

$$\hat{w}_1(t) = \hat{w}_1(t-1) + \phi(t)\frac{\partial\left|kurt(\hat{w}_1(t-1)^T X)\right|}{\partial\hat{w}_1(t-1)} \tag{23}$$

$$\hat{w}_1(t) = \hat{w}_1(t)/\|\hat{w}_1(t)\| \tag{24}$$

There is a more convenient way to write the above rules, to conform with standard programming language:

$$\hat{w}_1 \leftarrow \hat{w}_1 + \phi\, \frac{\partial\big|kurt(\hat{w}_1^T X)\big|}{\partial \hat{w}_1} \tag{25}$$

$$\hat{w}_1 \leftarrow \hat{w}_1/\|\hat{w}_1\| \tag{26}$$

Iterations continue until convergence $\Delta \hat{w}_1 = 0$ is achieved. Here ϕ is a learning rate, which is an important factor for the speed of convergence. Too small a learning rate will result in low convergence; too large a learning rate can result in instability and local maxima.

A more efficient fixed-point iteration, which has neither a learning rate nor other adjustable parameters in the algorithm can be defined. As we are trying to maximize the kurtosis function under the constraint $\|\hat{w}_1\| = 1$, at the maximum point, the gradient must point in the same direction as $\hat{w}_1$, which means the gradient must be equal to w multiplied by some scalar constant. This is the condition for convergence: as only in this case, adding the gradient to $\hat{w}_1$ will not change the direction. Thus we can set $\hat{w}_1$ proportional to the gradient of kurtosis:

$$\hat{w}_1 \propto E\{X(\hat{w}_1^T X)^3\} - 3\hat{w}_1\|\hat{w}_1\|^2 \tag{27}$$

Where $\propto$ means "proportional to". We can obtain a fixed-point algorithm where

$$\hat{w}_1 \leftarrow E\{X(\hat{w}_1^T X)^3\} - 3\hat{w}_1 \tag{28}$$

We omit $\|\hat{w}_1\|^2$ from equation (24) as after each fixed-point iteration, we will project $\hat{w}_1$ onto the unit sphere. The final vector $\hat{w}_1$ will give us one of the independent components.

The algorithm presented above computes only one of the independent components, to get all the independent components, it is necessary to repeat the process. In principle, this can be done with different weighting vectors for $\hat{w}_1, \hat{w}_2, \ldots, \hat{w}_i$. In order to make

sure that different independent components do not converge to the same solution, the vectors $\hat{w}_i$ corresponding to different independent components must be orthogonal in the whitened space because $E\{(\hat{w}_i^T X)(\hat{w}_j^T X)\} = \hat{w}_i^T \hat{w}_j = 0$ for X with unit variance. As iteration algorithms do not guarantee orthogonality of weighting vectors $\hat{w}_i$ automatically, we need to orthogonalize the vectors after each iteration step. We can apply a deflationary orthogonaliztion where the vectors w_i are estimated one-by-one, or a symmetric orthogonalization where the vectors are estimated in parallel. Usually it is more desirable to use the symmetric approach to estimate the sources simultaneously, because the deflationary approach will cumulate the estimation error in the first vectors to the subsequent ones. In this paper, we employ a symmetric de-correlation to estimate all the $\hat{w}_i$ in parallel, until convergence. The symmetric orthogonalization of the matrix $W = (w_1,...,w_m)^T$ can be obtained by the classic method involving the matrix square root:

$$W \leftarrow (WW^T)^{-1/2}W \qquad\qquad (29)$$

where $(WW^T)^{-1/2} = E * diag(d_1^{-1/2},...,d_m^{-1/2}) * E^T$, which can be obtained from the eigenvalue decomposition.

4. Data

To test the ICA model against the standard PCA approach, a dataset of US equity REITs was assembled. Larger REITs (by market capitalisation) were selected from each of the sector categories[3] to reduce problems of the lack of trading activity for smaller stocks. Weekly data were collected from DataStream, starting in 1986, but it was decided to begin the analysis in 1992, to focus on what has been described as the "modern era" of the REIT market[4]. After removing any REITs where there were inexplicably large movements in share price, this left a usable sample of 46 REITs with continuous returns from 1992 to 2005 available for analysis. Inevitably there is some survivorship bias in the sample set, resulting both from REIT failure and from

---

[3] Information was drawn from the NAREIT REIT Watch pages which provides basic information including sector and market cap on each REIT analysed.
[4] That is the period after Kimko REIT IPO in 1991 that established the UPREIT structure and the 1992 IRS changes that allowed greater institutional holdings of REIT stocks (see, e.g., Ling & Ryngaert, 1997, Chan *et al.* 2003).

the process of consolidation observed in the market. The sample is felt to be sufficiently robust for conclusions to be drawn. Both price appreciation and total return (including dividends) series were collected. The results presented here relate to total returns (expressed as log differences).

Table 1 presents summary descriptive statistics for the REIT sample[5]. The results appear broadly consistent with the overall NAREIT index over the same period. The most striking feature of the dataset is its non-normality. For all 46 REITs, the Jarque-Bera statistic indicates strongly significant departure from normality – similar results were obtained using the Lilliefors test. Inspection indicates clearly that this results largely from kurtosis in the individual return series, rather than skewness. These results are consistent with prior analyses which show non-normality in real estate data (for a review, see Lizieri & Ward, 2000). The high significance of kurtosis in the data series provides strong support for our ICA approach.

Table 1: Descriptive Statistics: REIT Sample Total Returns

	------ *Statistics from Sample* ------			
Statistic:	*Mean*	*Median*	*Upper Quartile*	*Lower Quartile*
Mean	0.003	0.003	0.002	0.004
Median	0.001	0.001	0.000	0.002
Standard Deviation	0.039	0.030	0.027	0.040
Skewness	-0.222	0.037	-0.136	0.266
Kurtosis	15.031	6.885	4.889	9.467
Jarque-Bera	67063	477	115	1337

Note: Sample size = 46, Data are weekly total returns, 1992-2005, skewness and kurtosis normalised to zero.

5. Results

Following the steps discussed in section 3, initial processing involved subtraction of the mean of each REIT return series from the returns at time t, then normalization of the resulting values to give unit variance. As a preprocessing step of ICA, we employ the PCA technique to restrain our subspace. Based on the Kaiser criterion, which gives us a cut-off for choosing the number of components we keep i components at which λ_i still exceeds $\overline{\lambda} = \frac{1}{n}\sum_{i=1}^{n}\lambda_i$, where λ_i is the eigenvalue of the ith component

---

[5] Individual descriptive statistics are available from the authors by request but are not presented here for reasons of space.

(equivalent to retaining eigenvalues greater than one, that is that have greater explanatory power than an individual REIT series).

Using this criterion, twelve components, which together capture 70.8% of the total variance are retained. Figure 1 shows the cumulative proportion of variance explained by the principal components, Table 2 shows the twelve retained components, along with the proportion of variance explained, the cumulative variance and the kurtosis. Thus, the PCA procedure reduces the dimensionality significantly.

Figure 1 Principal Components: Variation Explained
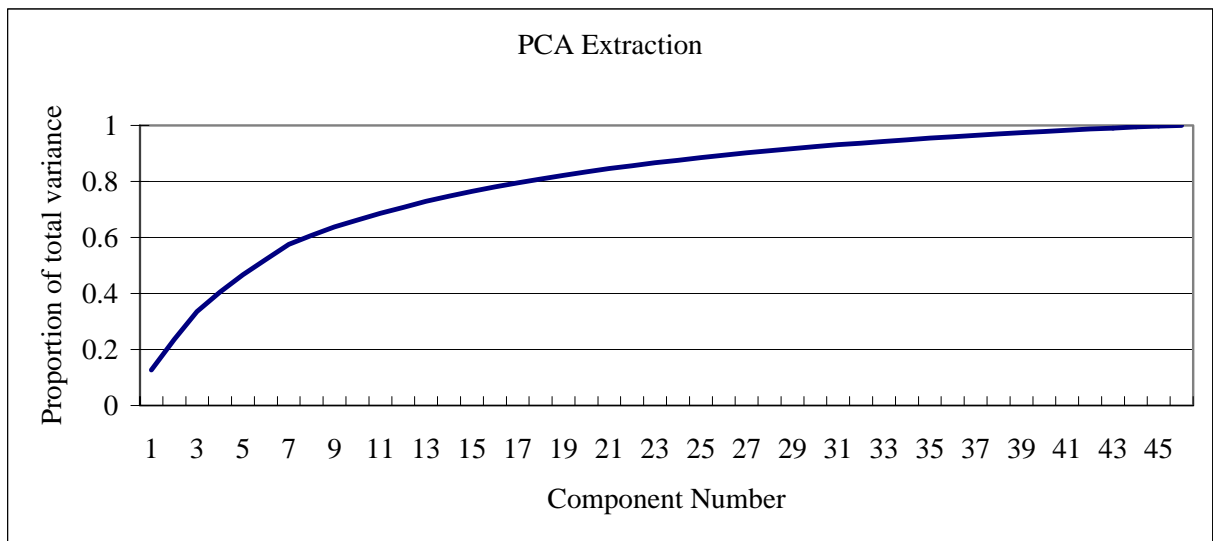
Table 2: PCA Retained Components

	Eigenvalue	%Var	Cu Var	Kurtosis
1	6.0514	12.7%	12.7%	6.166
2	5.2723	11.0%	23.7%	35.239
3	4.6978	9.8%	33.5%	17.404
4	3.3344	7.0%	40.5%	19.324
5	2.6990	6.1%	46.7%	4.334
6	2.5656	5.5%	52.2%	4.150
7	1.5614	5.4%	57.5%	7.670
8	1.4306	3.3%	60.8%	6.327
9	1.2126	2.9%	63.7%	6.957
10	1.1173	2.5%	66.2%	4.629
11	1.0633	2.3%	68.6%	5.382
12	1.0027	2.2%	70.8%	20.834

We now proceed to estimate Independent Components, based on our kurtosis criterion. By applying the iteration procedure for $\hat{W}$ to maximize the absolute value

of kurtosis, we obtain the estimated weighting matrix $\hat{A}$, $\hat{W}$ and 12 independent components S. Thus we successfully project the original 46-dimension vectors to the subspace spanned by these 12 independent components. This method is especially useful in a case such as this where we have multiple return series.

Table 3: ICA Kurtosis

	Kurtosis	Cum Kurtosis %
1	58.46	14.51%
2	44.76	25.62%
3	43.37	36.38%
4	42.96	47.04%
5	39.25	56.79%
6	36.96	65.96%
7	30.62	73.56%
8	29.50	80.88%
9	28.40	87.93%
10	16.52	92.03%
11	16.45	96.11%
12	15.68	100.00%

Figure 2: Kurtosis – ICA vs PCA
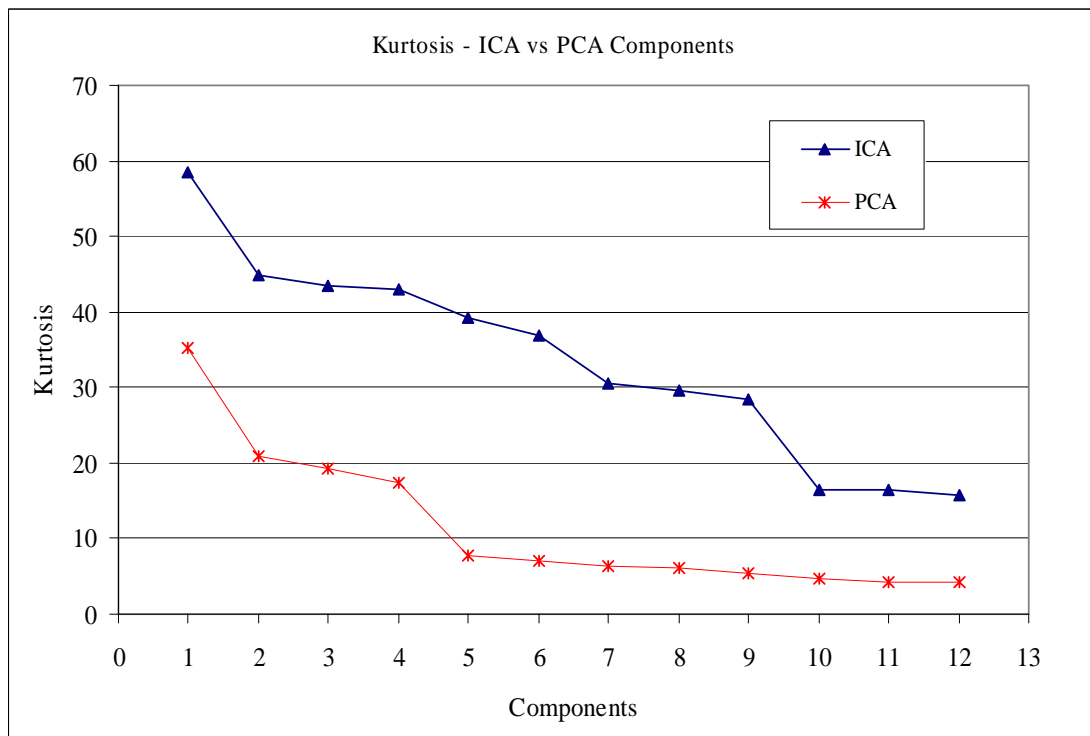
Figure 3 Cumulative Kurtosis, ICA vs PCA (ordered)
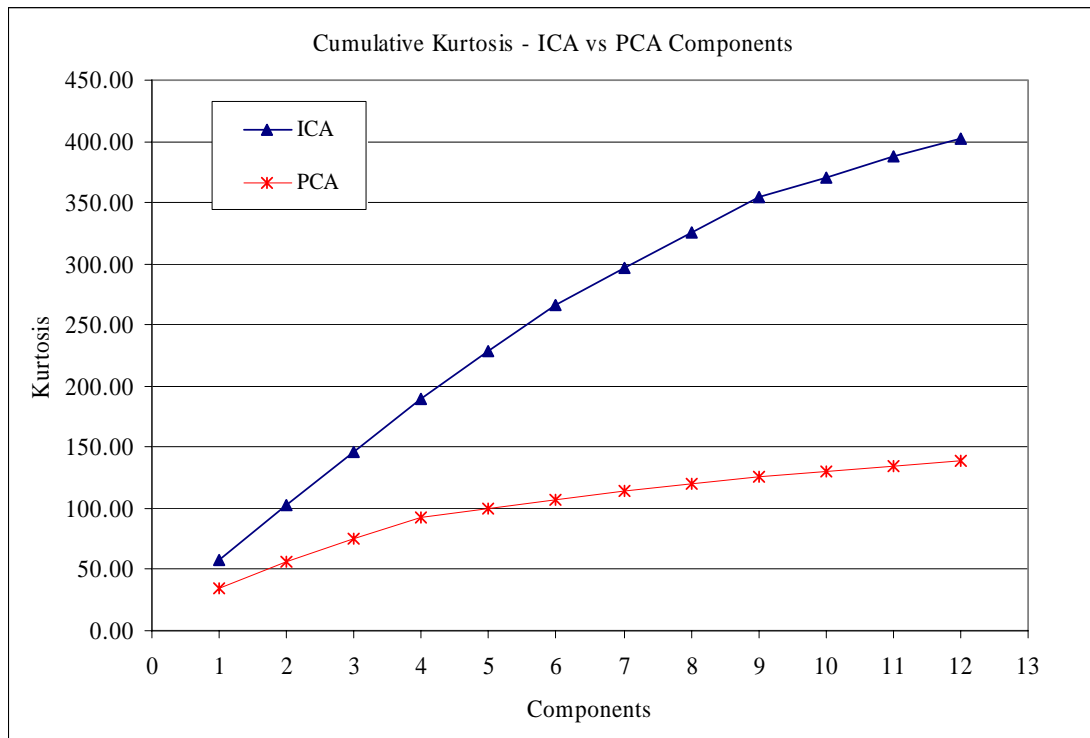
It is immediately apparent that the kurtosis of the independent components are significantly larger than those of the PCA components – Figures 2 and 3, which compare kurtosis measures from the twelve ICA components to those of the PCA components ordered by kurtosis, illustrates this graphically. The average kurtosis of the twelve ICA components is 33.6 compared to an average of 11.5 for the first twelve PCA components. The method has clearly been successful in creating highly kurtotic "portfolios" of REIT returns. This offers the opportunity to develop risk measures that are sensitive to exogenous factors that generate extreme events.

Figure 4 shows the factor scores from the independent components over time, scaled to unit variance. As can readily be seen, the components capture key return driver moments over time. It appears that many of the components capture periods of extreme volatility which suggests that they may well map onto particular shock events or factors in financial markets or the wider economy. That mapping exercise is outside the scope of this current paper but offers the possibility of characterising the components and identifying factors that drive extreme REIT returns.
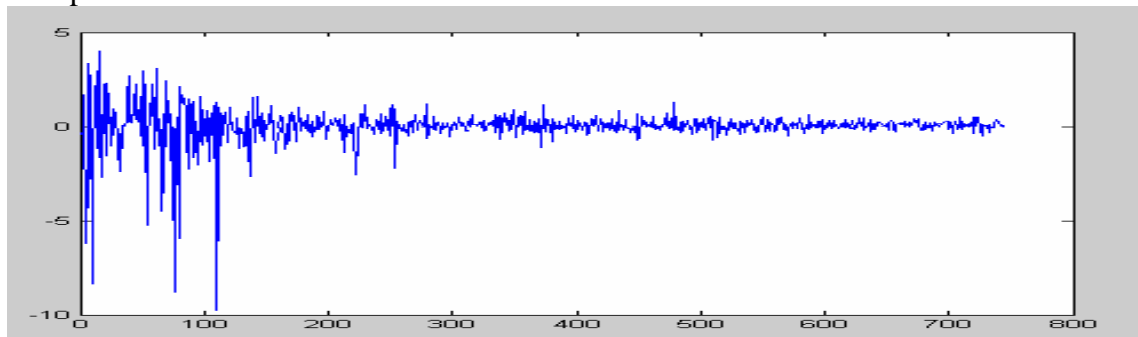
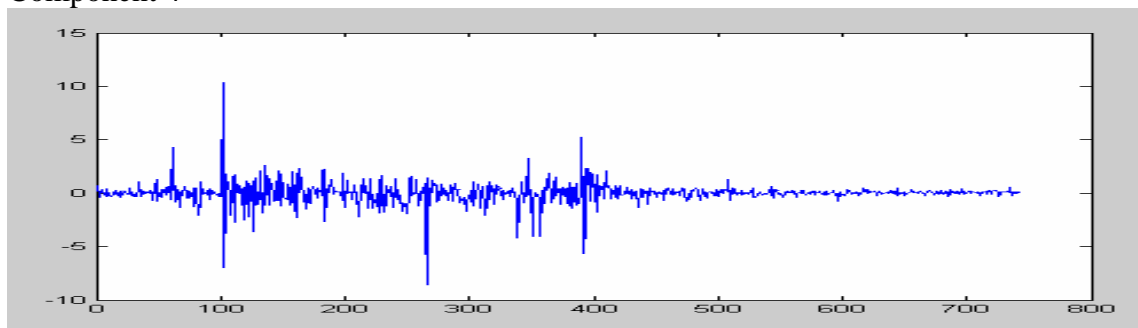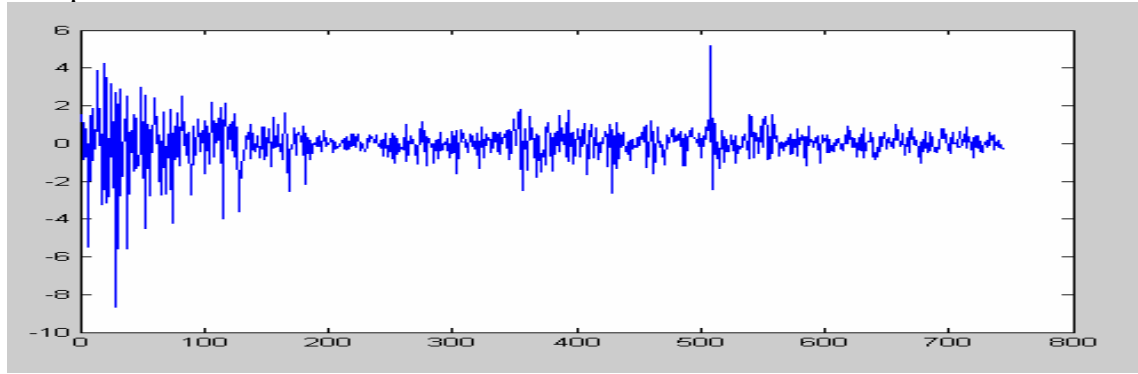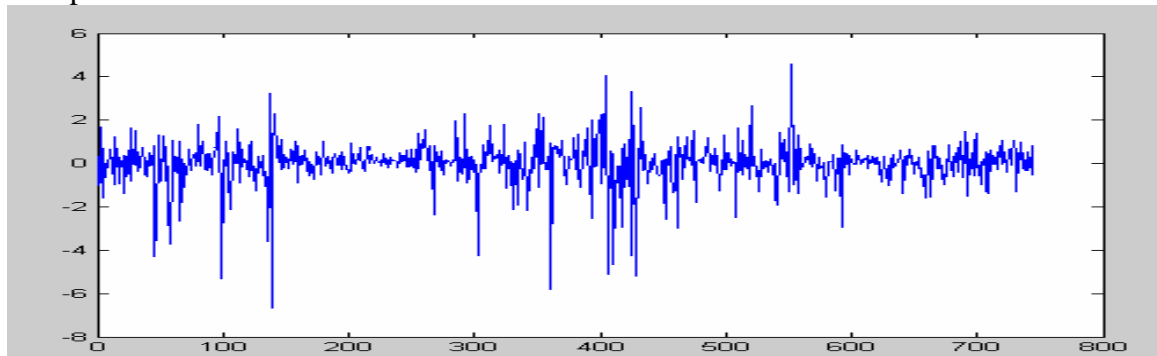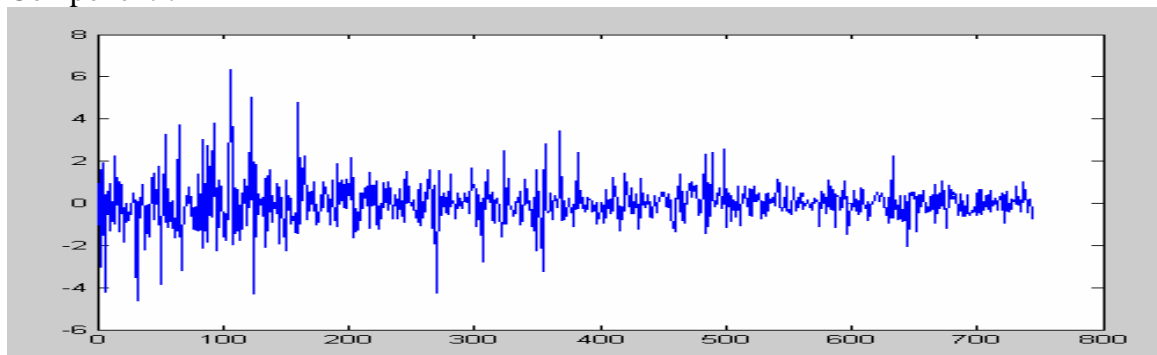Figure 4: ICA Factor Scores (Rescaled)

Component 1
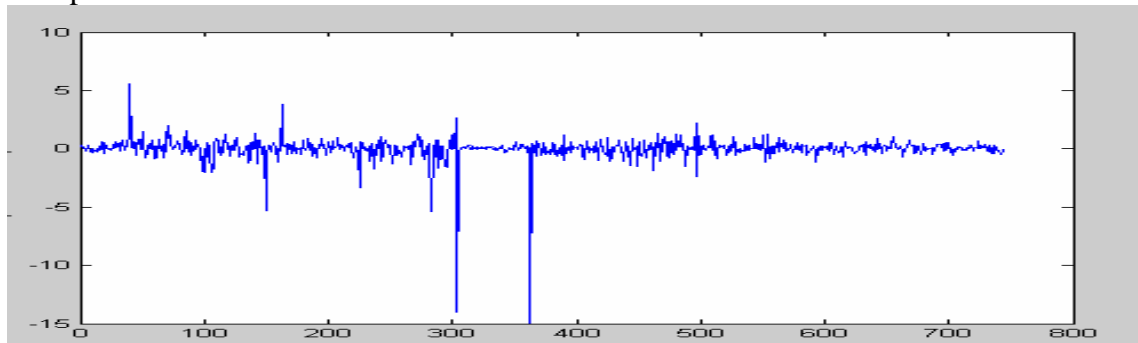
Component 2
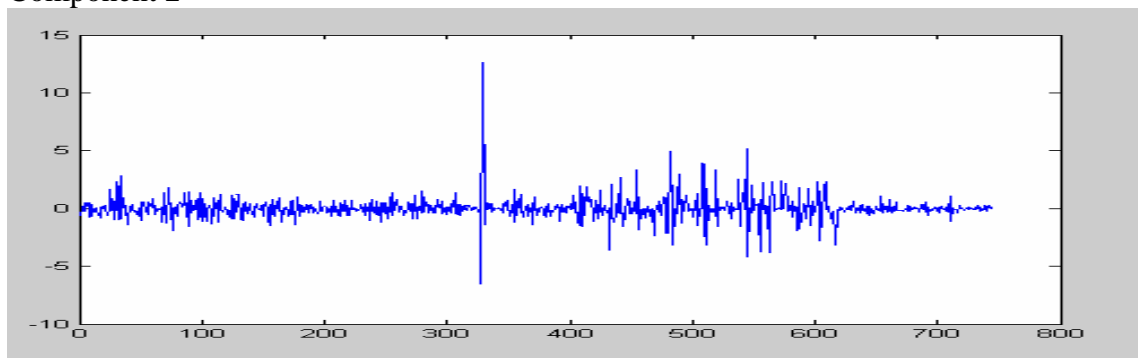
Component 3

Component 4

Component 5
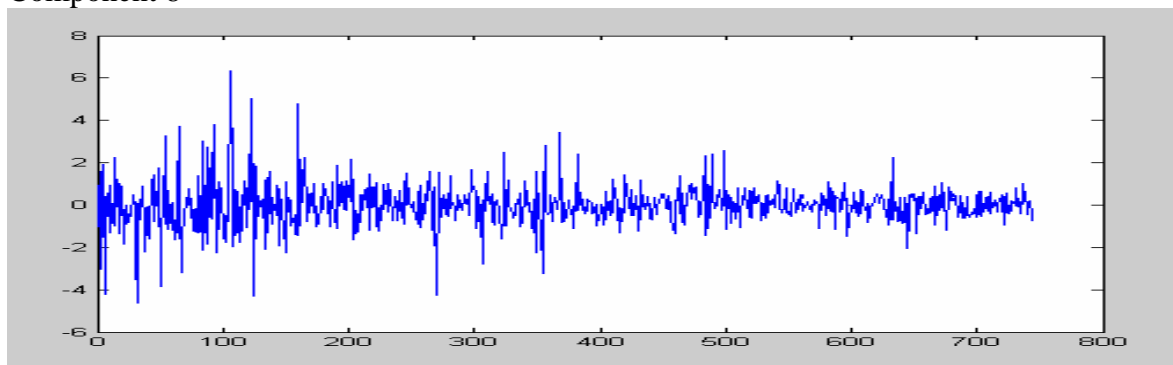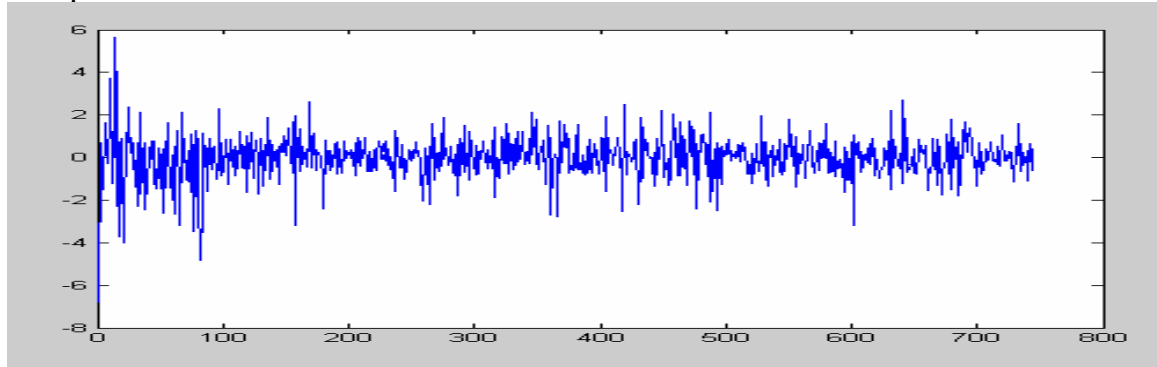
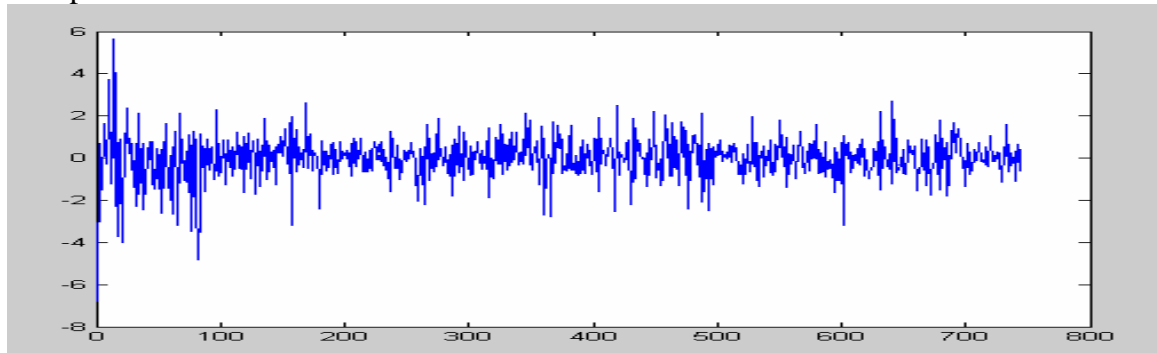

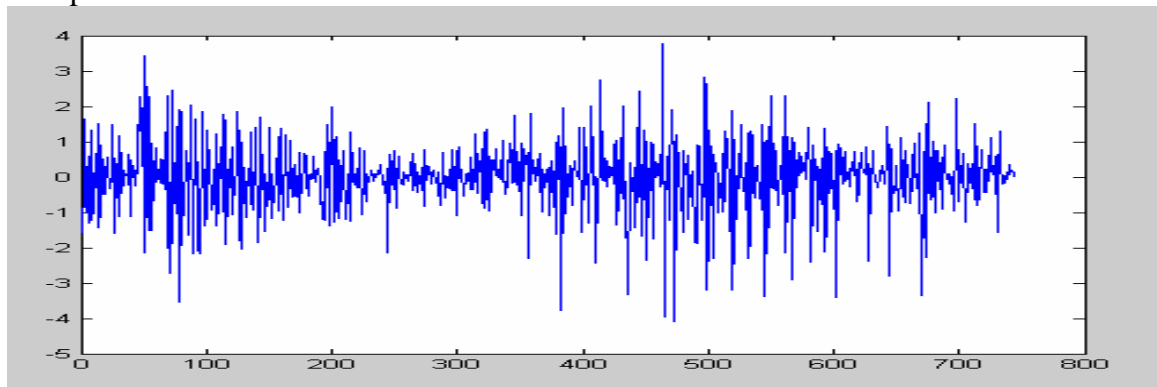Component 6

Component 7

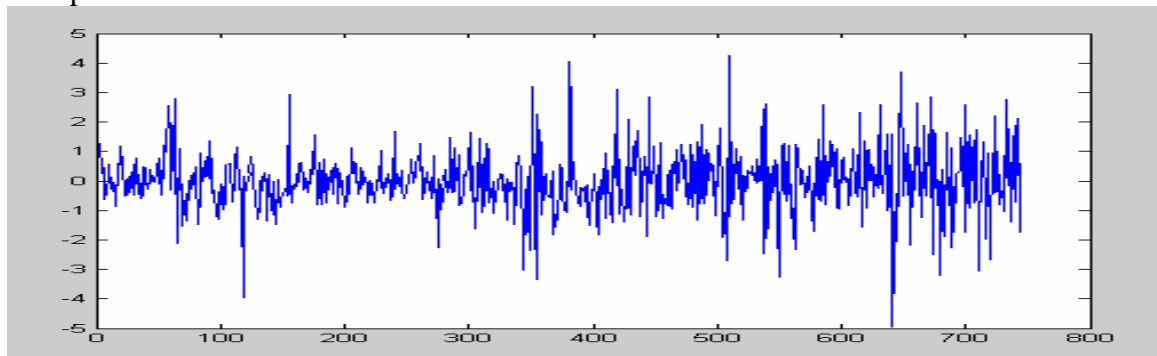Component 8

Component 9

Component 10

Component 11

Component 12

We should emphasise that our use of independence in this paper really only involves zero correlation for our ICA components. A more sophisticated procedure, currently under development, is to impose third order and fourth order zero cross moment constraints, to achieve fourth moment, rather than second moment, independence. Results, not reported here but available from the authors on request, show that the higher cross-moments are statistically insignificant.

6. Conclusions

The development of multi-factor models in real estate has seen an increasing focus on macro-factor approaches. While this has many advantages, the growing attention on extreme events, high frequency data and the data constraints of a macro variable approach suggest that there may be advantages in revisiting latent factor models as originally set out in the formulation of the arbitrage pricing model. However, the standard principal components model has certain deficiencies – notably that it emphasises only second moment independence. In this paper, we set out an alternative set of procedures – Independent Component Analysis – and provide a practical application using a kurtosis maximisation procedure. This is applied to individual weekly US equity REIT returns between 1992 and 2005.

The results demonstrate that ICA is successful in capturing much of the kurtosis in the REIT returns, the twelve ICA components explaining three times more kurtosis than the twelve retained PCA components (which, in turn, explain over 70% of the variance in REIT returns). Examination of the scaled components over time shows that they individually capture periods of extreme volatility in the market.

To generate a wider acceptance of the ICA approach, it will be necessary to develop a characterisation of the risk factors. This is a task outside the scope of the current paper. As a first step, one might examine the scaled factor score graphics, identify peaks and troughs and map these onto real world events and shocks to attempt to derive markers. These events are not necessarily equity market shocks – an extension would be to examine the extent to which general equities and REITs exhibit similar risk characteristics with respect to kurtosis-maximising Independent Components.

The ICA procedure set out here represents a first step in developing an analysis and techniques for management of extreme events. This is a critical task with the development of high time frequency trading models and the growing evidence that in market meltdowns, correlations tend to one – suggesting that risk management based on conventional mean-variance models are inadequate. Growing activity in real estate derivatives provides an added urgency for this task.

Bibliography

Back, A. & Weigend, A. (1997) A First Application of Independent Component Analysis to Extracting Structure from Stock Returns *International Journal of Neural Systems* 8, 473-484.

Blundell, G. and Ward, C. (1987) Property Portfolio Allocation: A Multi-Factor Model, *Land Development Studies,* 4, 2, 145-156.

Bond, S. and Patel, K. (2003) The Conditional Distribution of Real Estate Returns: Relating Time Variation in Higher Moments to Downside Risk Measurement, *Journal of Real Estate Finance & Economics,* 26, 319-339.

Brooks, C. and Tsolacos, S. (1999) The impact of economic and financial factors on UK property performance, *Journal of Property Research*, 16, 2, 139-152

Campbell, R., Koedijk, K. and Kofman, P. (2002) Increasing Correlation in Bear Markets: A Downside Risk Approach, *Financial Analysts Journal,* 58, 87-94.

Chan, K., Hendershott, P. & Sanders, A. (1990) Risk and Return on Real Estate: Evidence from Equity REITs, *AREUEA Journal,* 18, 431-452.

Chan SH, Erickson J, Wang K (2003) *Real estate investment trusts: structure, performance and investment opportunities*. Oxford University Press, ISBN 0-19-515534-3

Chen, N., Roll, R. & Ross, S. (1986) Economic Forces and the Stock Market: Testing the APT and Alternative Asset Pricing Theories, *Journal of Business,* 59, 383-403.

Chen, S., Hsieh, C. & Jordan, B., (1997) Real Estate and the Arbitrage Pricing Theory: Macrovariables vs. Derived Factors, *Real Estate Economics,* 25, 505-523.

Chen, S. and Jordan, B. (1993) Some Empirical Tests in the Arbitrage Pricing Theory *Journal of Banking and Finance,* 17, 1, 65-89.

Downs, D. (2000) Assessing the Real Estate Pricing Puzzle: A Diagnostic Application of the Stochastic Discounting Factor to the Distribution of REIT Returns, *Journal of Real Estate Finance and Economics*, 20:2, 155–75.

Hamelink, F. and Hoesli, M. (2004) Maximum Drawdown and the Allocation to Real Estate, *Journal of Property Research* 21(3), 5-29

Hyvarinen, A., Karunen, J. and Oja, E. (2001) *Independent Component Analysis* John Wiley & Sons

Knight, J., Lizieri, C. and Satchell, S. (2005) Diversification when It Hurts? The Joint Distributions of Real Estate and Equity Markets *Journal of Property Research,* 22, 4, 309-323

Ling, D. and Naranjo, A. (1997) Economic Risk Factors and Commercial Real Estate Returns, *Journal of Real Estate Finance and Economics* 14, 3, 283-307.

Ling, D. and Naranjo, A. (1999) The Integration of Commercial Real Estate Markets and Stock Markets, *Real Estate Economics*, 27, 483-515.

Ling, D., Naranjo, A. and Ryngaert, M. (2000) The Predictability of Equity REIT Returns: Time Variation and Economic Significance, *Journal of Real Estate Finance and Economics*, 20, 117-136.

Lizieri, C. and Ward, C. (2000), Commercial Real Estate Return Distributions: A Review of the Literature and Empirical Evidence, in Knight, J. and S. Satchell, (eds.) *Return Distributions in Finance*, London: Butterworth Heinemann.

McCue, T. and Kling, J. (1994) Real Estate Returns and the Macroeconomy: Some Empirical Evidence from Real Estate Investmetn Trust Data, 1972-1991. *Journal of Real Estate Research,* 9, 3, 277-287.

Titman, S. and Warga, A. (1986) Risk and Performance of Real Estate Investment Trusts: A Multiple Index Model, *AREUEA Journal,* 14, 414-431.

Young, M. and Graff, R. (1995) Real Estate Is Not Normal: A Fresh Look at Real Estate Return Distributions, *Journal of Real Estate Finance and Economics*, 10, 225-259.

Appendix One: PCA and ICA

Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are closed related. As a pre-processing step of ICA, PCA is mainly used to reduce the number of variables. The principal components (PC) are a set of variables that can capture the maximum amount of variation in a dataset and are orthogonal to all previous principal components. For example, principal component one (PC1) is defined as the eigenvector with the highest eigenvalue. A higher eigenvalue means that more variance has been captured. As there is a trade off between dimensionality and information, we choose the eigenvectors having the highest eigenvalues, so as to lose as little information as possible in the mean-square sense. The PCA algorithm focuses on the idea of zero correlation and uses only up to the second order statistical information to identify the components.

However, zero correlation does not mean independence for non-gaussian data, even though it is the case for Gaussian data. To deal with non-gaussian signals, as discussed earlier, we use ICA. In the model, we assume that the observed data variables are a linear mixture of some unknown non-gaussian latent variables with an unknown mixing system. We apply ICA to find the latent factors that are statistically independent - or as independent as possible. The latent variables are called independent components. Compared with PCA, ICA adds in the idea of independence and uses higher order statistical information to identify the underlying factors. It should thus be valuable in finance contexts characterised by non-normality.

Appendix 2: Estimation Procedures

1.   We identify the underlying return generating factors by applying Independent Component Analysis.

       The algorithm iterates between orthogonalisation (enforcing zero correlation) followed by maximization of non-Gaussianity: in our case, maximization of kurtosis. We need to maximize kurtosis because the solution to 'zero correlation' is not unique (it is 'rotationally invariant'). By maximizing kurtosis, we hope to find components which are as independent as possible.

2.   We test whether kurtosis maximization indeed gives us independent components. By bootstrapping the variance of covariance, we test the null hypothesis that cov{g(Si),h(Sj)}=0, where g() and h() are nonlinear transformations of S.

3.   We reorder the independent components based on kurtosis and calculate what proportion of total kurtosis is explained by the first i independent components.

      In PCA we can take the eigenvalues of the eigenvalue decomposition (EVD or SVD) and see them as the common variance, i.e. the variance that this component has in common over all measurements. The first component will thus have a high variance in measurements.