A Computational Methodology For Modelling the Mispricing DYNAMICS

Georgios Pipis

A thesis submitted to the Heriot-Watt University for the degree of Msc in Financial Mathematics

UNIVERSITY OF EDINBURGH
AND
HERIOT-WATT UNIVERSITY

SCHOOL OF MATHEMATICS
AND
COMPUTER SCIENCE

SEPTEMBER 2007

Abstract

Statistical arbitrage trading is a quantitative and computational approach to equity trading which is widely applied by hedge funds to produce market-neutral returns. The simplest and most popular version of the strategy is known as pairs trading and involves the identification of pairs of assets that are believed to have some long-run equilibrium relationship. By taking an appropriate long-short position on this pair when the spread has diverged sufficiently from the equilibrium value, a profit will be made if the spread converges back to equilibrium by unwinding the position. Similar ideas govern more complicated strategies which consider a larger basket of assets.

The scope of this thesis is to examine the statistical models that underpin statistical arbitrage implementing tools from the fields of econometrics, statistics and time-series forecasting. Initially we focus on pairs trading strategy endeavouring to specify precisely the concept of the long-run equilibrium relationship between two assets and then we try to describe and apply a computational methodology for modelling the mispricing dynamics. Thereafter we attempt to apply the *statistical arbitrage trading strategy* to larger basket of assets. The final step is the implementation of an empirical study the objective of which is the evaluation of this strategy at the Greek financial market enabling us to answer the question if we could have made money consistently.

ACKNOWLEDGMENTS

I would like to express my special gratitude to my supervisor Professor Alexander J. McNeil for suggesting to me the problem which led to this thesis, and for helping me at various stages through the course of its solution. Finally I would like to thank him for his encouragement during this work.

I would like also to thank my uncle Ioannis Kontogiannis for affording this education opportunity.

I would like to thank my parents, Prokopis and Renata for their love and support.

My love and thanks to my sisters Elena and Stella.

Thanks to Nikolaos Vasilakis for his support.

Thanks to my classmate Charikleia Liakopoulou.

Thanks to my grandmother Eleni for her love.

Table of Contents

Chapter 7: Rolling Analysis

Chapter 8: Trading Design

INTRODUCTION

The scope of this thesis is the attempt to exploit regularities or patterns in asset price dynamics. It could be assumed that the asset prices follow a *'random walk'*[1] which simply implies that none investor could predict the future movement of the stock prices. This is consistent with the *'Efficient Market Hypothesis'*[2] (EMH) which states that is not possible to outperform the market by using any information that the market already knows. As a result investment strategies such as 'Technical Analysis' and 'Fundamental analysis' which are based on historical share prices or other financial data will not be able to consistently produce excess returns. Thus given the above argument one could claim that is futile trying to predict stock returns, but as we shall show throughout this thesis, it may be possible to construct portfolios whose dynamics are stationary time series and indeed predictable. On this argument is based the *'Statistical Arbitrage Strategy'* where in a wide sense is any strategy that is approximately *'beta neutral'* and uses tools from the area of statistics and econometrics in order to provide signal for execution which are usually generated through contrarian *mean-reversion* property[3]. Thus our aim is to construct statistical fair-price relationships between stock prices such that the deviations (mispricings) have a potentially predictability and to apply an appropriate trading strategy buying stocks which are forecasted to outperform and selling stocks which are forecasted to underperform. The mispricing component can be considered as a *'synthetic asset'*[4] which is a stationary process and has a strong mean-reverting property which implies a certain degree of potentially predictable behaviour. The concept of 'Statistical Arbitrage'[5] is that statistical regularities in combinations of stock prices can be exploited as the basis of profitable trading strategy, regardless the presence or absence of a theoretical fair-price relationship between the stocks which are involved in the '*synthetic asset*' .

Before starting the analysis it is important to clarify that statistical arbitrage trading is not a riskless strategy and thus an investor who follows this strategy should be alert. All these issues will be discussed through this dissertation. The effectiveness of the methodology is demonstrated by extensive empirical analysis of real data sets, in particular daily closing prices of Athens Stock Exchange (ASE) from 10/7/2002 until 12/2/2007.

[1] Fama, E., 1965, *'The Behavior of Stock Market Prices'*
[2] Fama, E., 1970, *'Efficient Market Theory'*
[3] By this we mean the property which have some stochastic process to oscillate around its mean value.
[4] Burgess, A.N., 1999, '*A Computational Methodology for Modelling the Dynamics of Statistical Arbitrage*'
[5] Wong, M.A., 1993, '*Fixed-Income Arbitrage: Analytical Techniques and Strategies'*

1.1 Selection of Pairs

For the beginning we restrict ourselves to the bivariate case known as '*Pairs Trading Strategy*' and later on we will extend the theory considering a larger basket of stocks. Pairs trading is a well-known *statistical arbitrage strategy* and is one of Wall Street's quantitative methods of speculation which dates back to the mid-1980s[6]. The aim of this strategy is to construct a portfolio of two stocks which are in long-run equilibrium. Then we take an appropriate position when the spread has diverged significantly of its equilibrium. A profit may be made by unwinding the position upon convergence of the spread, or the measure of relative mispricing. From the above discussion it is clear that we are seeking for stocks whose price movements are strongly correlated in order to have chances to implement the pairs trading strategy. Since in this empirical study we have chosen around 90 stocks from Athens Stock Exchange (ASE) it will be a time-consuming process to examine $\frac{90!}{2!(90-2)!}$ pairs separately in order to verify if they are indeed co-integrated[7]. Hence we are looking for some simple and straightforward approaches which enable us to exclude some pairs and find some potentially co-integrated pairs. The simplest approaches that we have used are described briefly in the next section.

1.2 Simple Pairs Selection Methods

The simplest method to define potentially co-integrated pairs is the computation of the correlation of stock prices considering around 1000 daily closing prices. A widely-used method is the '*distance method*'[8] where the co-movement in a pair is measured by what is know as the distance, or the sum of squared differences between the two normalized price series. Finally a rational method is to consider the logarithm of the stock prices and then to compute the correlation of them.

In this study we applied all of these methods where we found approximately the same results. In order to compute these 'measures' we considered about 1000 observations in order to verify the long-run equilibrium. Later on, the same calculations are made taking the 350 more recent observations in order to avoid misleading results, since there is a chance that two stocks are in equilibrium for several years, and then due to some reasons, such as changes in the fundamentals of a company, a take over bid, an announcement of split etc, will not be in equilibrium anymore. This was the first step of the analysis and the results were to exclude around 70 stocks leaving only 23 stocks. The most important criterion of stocks to qualify to further analysis was the correlation coefficient of the logarithm of the stock prices to be grater than 0.89.

In sequel we will explain the reason for why we have transformed the stock prices to a logarithmic scale and the notion of co-integration.

A Correlation Matrix of the logarithm of the 23 stock prices and some plots are illustrated in *Appendix 1.1*. Below we represent a time-series plot which illustrates

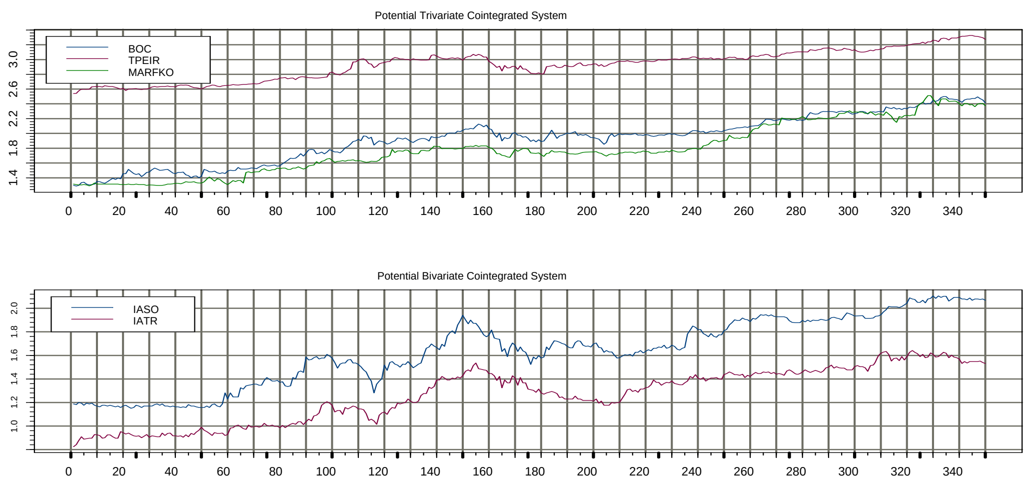
[6] Vidyamurthy G., 2004, '*Pairs Trading: Quantitative Methods and Analysis'*

[7] Two or more Time-Series are called co-integrated if they are non-stationary and integrated of the same order and have the same trend. An appropriate combination of them leads us to a stationary process.

[8] Gatev et al 1999, '*Pairs Trading: Performance of a Relative Value Arbitrage Rule*'

some stocks which seem to be in long-run equilibrium and intuitively we can argue that they are indeed co-integrated. Of course we need much further analysis but this was the first step.

Figure 1.1



Source: www.euro2day.gr. The plot above represents the logarithm of the daily closing prices of three stocks. These three stocks are Greek banks and it seems to be in equilibrium. The plot below represents the logarithm of the daily closing prices of two stocks where both of them are private hospitals. These two stocks is a potential trading pair. The prices are obtained from 21/9/2005 to 12/2/2007.

In the next section there is a brief description of more advanced and complicated methods for selecting potentially co-integrated pairs.

1.3 Advanced Pairs Selection Methods

In case of pairs trading, the scope is to construct a portfolio of the following form:

$$\log(p_t^A) - \gamma \log(p_t^B) = \mu + \varepsilon_t \qquad (1.3.1)$$

where p_t^A and p_t^B are the daily closing prices of stocks A and B respectively. The coefficient γ is the co-integration coefficient and the stochastic term ε_t is the mispricing, which we would like to be stationary and to have the mean-reverting property in order to exploit its predictability. Finally the constant μ is the mean value of the long-run equilibrium. Hereafter we define the term $\log(p_t^A) - \gamma \log(p_t^B)$ as '*spread*'. In order to have chances to achieve a good fitting the model above (1.3.1) a basic requirement is the two stocks to be highly correlated.

Several studies and empirical results have shown that the logarithm of the stock prices is a non-stationary process and could be expressed as follows:

$$\log(p_t) = n_t + \varepsilon_t \qquad (1.3.2)$$

Where n_t represents the trend or the random walk component which usually is stochastic but it could be also deterministic. Clearly this is a non-stationary process, and the term ε_t is the specific component of the time series which is a stationary process. Notice that some times in model (1.3.2) some authors use to add another component known as 'seasonality'. Since we have no evidence for seasonality and for simplicity we excluded this term of the model. A basic requirement for two time-series to be co-integrated, is that their common trends must be identical up to a scalar. Considering the (1.3.1) equations and taking the first differences we get:

$$\log(p_t) - \log(p_{t-1}) = n_t - n_{t-1} + \varepsilon_t - \varepsilon_{t-1} \tag{1.3.3}$$

or

$$r_t = r_t^c + r_t^s \tag{1.3.4}$$

where r_t^c is the return of the non-stationary trend component, and r_t^s is the return of the stationary component[9]. The equation above implies that if two stocks are co-integrated, then the returns from their common trends must be identical up to a scalar. This argument implies that an appropriate measure to define potential co-integrated pairs is to measure the correlation of the log-returns. But this argument is not absolutely correct because we made the assumption that all stocks are exposed to the same risk factors which is not true. For this reason we introduce some more appropriate methods known in wide sense as '*Factor Modelling*'.

1.3.1 Sharpe's Single Index Model

Sharpe's single factor model is the simplest and most famous macroeconomic model which could be expressed in the following form:

$$R_{it} = \alpha_i + \beta_i R_{Mt} + \varepsilon_{it},\ i = 1,\ldots,N \text{ and } t = 1,\ldots,T \tag{1.3.5}$$

where R_{Mt} is the returns of the market index in period t and R_{it} denotes the returns of the asset i in period t. The covariance matrix of assets from the single factor model can be expressed using matrix notation as follows:

$$\mathbf{V} = \sigma_M^2 \beta\beta' + \mathbf{D} \tag{1.3.6}$$

where σ_M^2 is the variance of the market index returns, $\beta = (\beta_1, \beta_2, \ldots, \beta_N)'$ and $\mathbf{D}$ is a diagonal matrix with $\sigma_i^2 = \text{var}(\varepsilon_{it})$ along the diagonal.

We can estimate the model using multivariate regression. The results that we obtained are illustrated below.

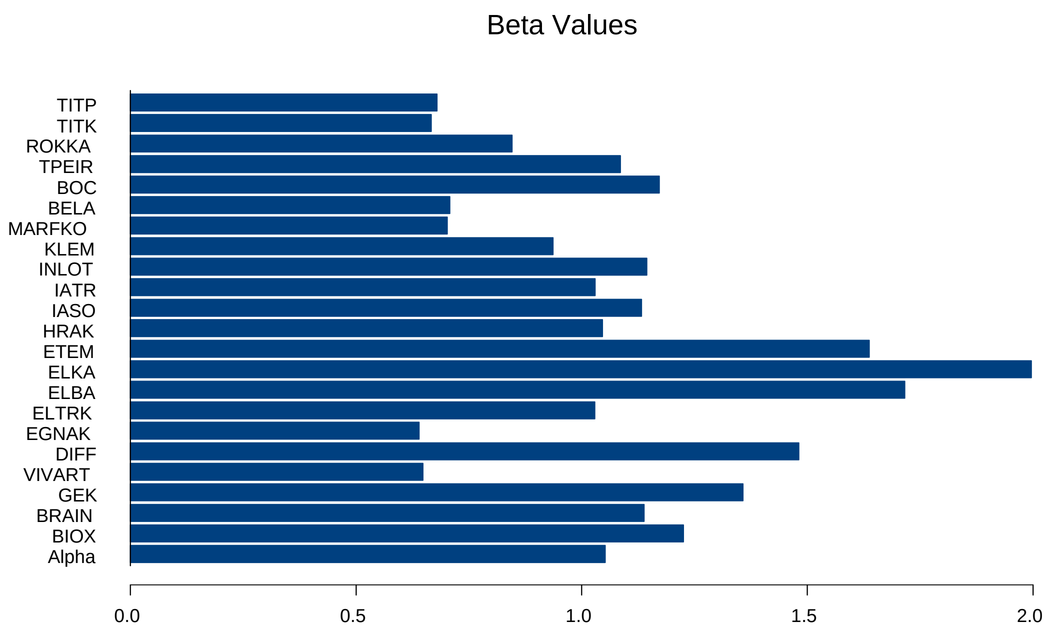
[9] Vidyamurthy G., 2004, '*Pairs Trading: Quantitative Methods and Analysis*'

Table 1.3.1

	beta.hat	s.error	r.square
Alpha	1.05	0.011	0.500
BIOX	1.23	0.018	0.361
BRAIN	1.14	0.034	0.115
GEK	1.36	0.019	0.372
VIVART	0.65	0.015	0.181
DIFF	1.48	0.031	0.211
EGNAK	0.64	0.022	0.088
ELTRK	1.03	0.022	0.209
ELBA	1.72	0.023	0.389
ELKA	2.00	0.030	0.344
ETEM	1.64	0.028	0.292
HRAK	1.05	0.015	0.353
IASO	1.13	0.025	0.198
IATR	1.03	0.020	0.247
INLOT	1.14	0.020	0.274
KLEM	0.94	0.019	0.225
MARFKO	0.70	0.021	0.115
BELA	0.71	0.017	0.167
BOC	1.17	0.020	0.280
TPEIR	1.09	0.012	0.478
ROKKA	0.85	0.018	0.211
TITK	0.67	0.014	0.206
TITP	0.68	0.014	0.224

The table above illustrates the '*beta value*' of each stock as well as the specific risk and the R^2 measure. Below we represent a 'bar plot' of the 'beta values' for each stock.

Figure 1.2



As we expected all the stocks have positive beta values. A suggestion for potential co-integrated pairs is the same values of beta. Indeed the higher correlated pairs have approximately the same values of beta. Since all estimated β_i values are positive, all the values in the single factor covariance will be positive. In 'Appendix 1.2' there is a presentation of some of the single factor correlations.

1.3.2 Fundamental Factor Model

Fundamental factor models use observable asset specific characteristics known as 'fundamentals' such as market capitalization, industry classification etc in order to determine the common risk factors. In practice, these models could be estimated in two ways. The first approach in known as the '*BARRA*'[10] approach, and the second one approach is known as the '*Famma-French*'[11] approach. In this study there is an implementation of the '*BARRA*' approach, where the observable asset specific fundamentals are treated as the factors betas, and the factor realizations at time t are unobservable, which are estimated by running T cross-section regressions, given the factor betas[12]. The '*Fundamental Factor Modelling*' is used in order to model the asset returns. In aspect of '*Statistical Arbitrage*' is a sensible way to determine the correlation of the asset returns. In this empirical study we implemented two models. The first one was based on industry classification where we considered four different classifications, and the other on was bas on market capitalization where we separated the assets to 'High Capitalization' and 'Middle and Small Capitalization'.

The general form of the model is as follows:

$$R_{it} = \alpha_i + \beta_{1i} f_{1t} + \beta_{2i} f_{2t} + \ldots + \beta_{Ki} f_{Kt} + \varepsilon_{it} \tag{1.3.7}$$

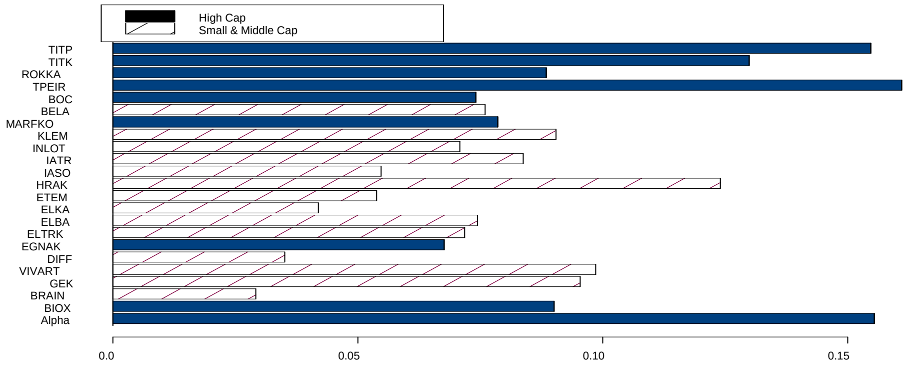
where R_{it} is the real return on asset i ($i = 1,\ldots,N$) in time period t ($t = 1,\ldots,T$), α_i is the intercept, f_{kt} is the k^{th} common factor ($k = 1,\ldots,K$), β_{ki} is the factor loading or factor beta for asset i on the k^{th} factor, and ε_{it} is the asset specific factor. In our case we obtained the factor realizations f_{kt} as well as the factor correlation matrices applying the *weighted least square estimation* (GLS). Notice that the GLS estimates of the factor realizations are just linear combinations of the observed returns in each common risk factor. These linear combinations sum to unity so that they can be interpreted as *factor mimicking portfolios*[13] and they are weighted averages of the returns in each industry where the weights on each asst are based on the size of the residual variance. Below we represent the results that we obtained using this approach.

1st Case: *Fundamental Factor Model based on Market Capitalization*

The plot below illustrates the factor loadings for each stock. We have considered two common factors based on market capitalization (i.e. the size of the company).

[10] Sheikh, A., 1995, '*BARRA's Risk Models*'
[11] Fama, E. , French, K.R., 1992, *'The Cross-Section of Expected Stock Returns'*
[12] See Zivot, E., Wang, J.2006, '*Modeling Financial Time Series with S-PLUS*' pages 580-589
[13] Zivot, E., Wang, J.2006, '*Modeling Financial Time Series with S-PLUS*'

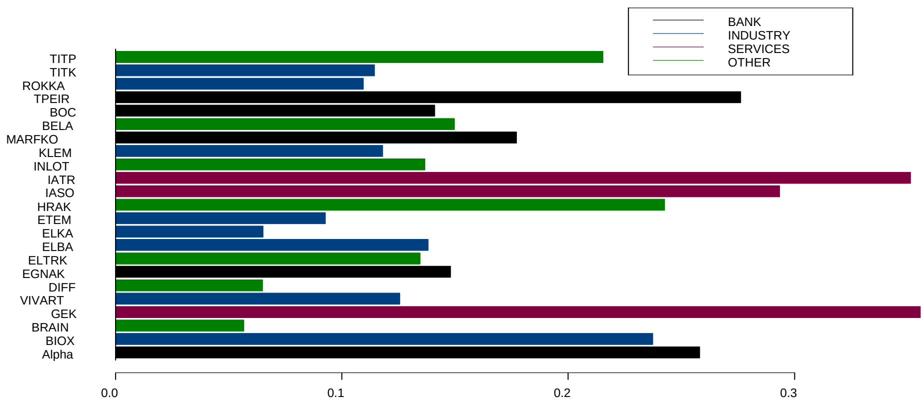
Figure 1.3



2nd Case: *Fundamental Factor Model based on Industry Classification.*

In this case we have separated the stocks into four types of industry which are defined to be '*Banks*', '*Industry*' for companies which produce industrial products, '*Services*' for tourist agencies, hospitals etc, and '*Other*'. The factor loadings for each stock are illustrated below.

Figure 1.4



These loadings can be used to find candidate co-integrated pairs. In *'Appendix 1.3'* there is a presentation of industry factor model correlation matrices for each case.

1.3.3 Statistical Factor Model

Statistical Factor Models could be divided into two main methods called '*Traditional Factor Analysis'* and *'Principal Component Analysis'* which are based on the $N \times N$ sample covariance matrix[14]

[14] An excellent overview of factor analysis is given in Zivot, E., Wang, J.2006, '*Modeling Financial Time Series with S-PLUS*' pages 590-614

$$\widehat{\mathbf{\Omega}}_N = \frac{1}{T}\mathbf{R}\mathbf{R}' \tag{1.3.8}$$

where $\mathbf{R}$ is the $N \times T$ matrix of observed returns. In this section we focus on the traditional factor analysis where we assume a time invariant *orthogonal factor structure*[15]

$$\underset{(N\times1)}{\mathbf{R}_t} = \underset{(N\times1)}{\mu} + \underset{(N\times K)}{\mathrm{B}}\underset{(K\times1)}{\mathbf{f}_t} + \underset{(N\times1)}{\varepsilon_t} \tag{1.3.9}$$
$$\operatorname{cov}(\mathbf{f}_t, \varepsilon_s) = 0, \text{ for all } t, s$$
$$E[\mathbf{f}_t] = E[\varepsilon_t] = 0$$
$$Var[\mathbf{f}_t] = \mathbf{I}_K$$
$$Var[\varepsilon_t] = \mathbf{D}$$

where $\mathbf{D}$ is a diagonal matrix with σ_i^2 along the diagonal and $\mathbf{B}$ is an $(N \times K)$ matrix of factor loadings. Then, the return covariance matrix $\mathbf{\Omega}$ could be expressed as follows:

$$\mathbf{\Omega} = \mathbf{BB'} + \mathbf{D}$$

For a given asset i, the return variance could be expressed as:

$$Var(R_{it}) = \sum_{j=1}^{K} \beta_{ij}^2 + \sigma_i^2 .$$

Now under the assumption that the returns are *iid* and normally distributed, we can obtain maximum likelihood estimates of $\mathbf{B}$ and $\mathbf{D}$. Given these two estimates we can construct an empirical version of the factor model as follows:

$$\mathbf{R}_t - \hat{\boldsymbol{\mu}} = \hat{\mathbf{B}}\mathbf{f}_t + \hat{\boldsymbol{\varepsilon}}_t \tag{1.3.10}$$

where $\hat{\boldsymbol{\mu}}$ is the sample mean vector of $\mathbf{R}_t$. Using the above equation we may estimate the factor realizations in a given time period t, $\mathbf{f}_t$ implementing the cross-sectional generalized least squares regression(GLS).

Since the number of factors is not specified a priori, the adequacy of the K factors may be tested applying the likelihood ratio test which is of the form:

$$LR(K) = -(T - 1 - \frac{1}{6}(2N+5) - \frac{2}{3}K) \cdot \left(\log\left|\hat{\mathbf{\Omega}}\right| - \log\left|\hat{\mathbf{B}}\hat{\mathbf{B}}' + \mathbf{D}\right| \right)$$

where $LR(K)$ follows a chi-square distribution (χ_ν^2) with $\nu = \frac{1}{2}((N-K)^2 - N - K)$ degrees of freedom[16].

[15] See Johnson and Wichern (1998), '*Factor Analysis using S-PLUS*'
[16] Zivot, E., Wang, J., 2006, '*Modeling Financial Time Series with S-PLUS*' pages 591-592

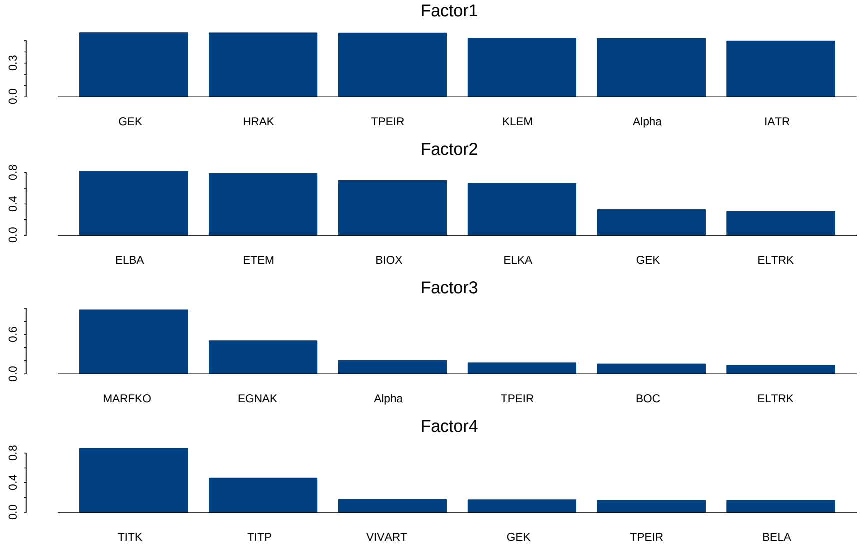
At this point we implement the traditional factor analysis. Applying the likelihood ratio test, we found that four factors are sufficient to determine the model. Below there is an illustration of the factor loadings.

Table 1.3.2

	Factor1	Factor2	Factor3	Factor4
Alpha	0.519	0.200	0.204	0.141
BIOX	0.407	0.697		0.146
BRAIN	0.357	0.183	0.131	
GEK	0.570	0.327		0.169
VIVART	0.388	0.169	0.113	0.176
DIFF	0.411	0.267		
EGNAK	0.216		0.504	
ELTRK	0.445	0.304	0.132	
ELBA	0.371	0.817		0.160
ELKA	0.417	0.663		0.120
ETEM	0.318	0.788		0.108
HRAK	0.568	0.269	0.124	0.137
IASO	0.488	0.162		0.117
IATR	0.496	0.257	0.112	
INLOT	0.475	0.185	0.130	0.143
KLEM	0.522	0.247		
MARFKO	0.120	0.104	0.978	
BELA	0.373	0.116		0.163
BOC	0.492	0.231	0.152	0.141
TPEIR	0.568	0.240	0.167	0.164
ROKKA	0.477	0.138		0.143
TITK	0.169		0.109	0.866
TITP	0.301	0.260	0.109	0.462

The next figure represents a '*bar plot*' of factor loadings.

Figure 1.5



1.3.4 Principal Component Analysis

At this point we discuss briefly the notion of *principal component analysis* (PCA) which is a dimension reduction technique used to explain the information in the sample covariance matrix returns. There are as many principal components as the number of assets, which are linear combinations of asset returns and are constructed and ordered in such a way that the first principal component explains the largest portion of the sample covariance matrix, the second one explains the next largest portion and so on. By construction, principal components are orthogonal to each other and they have unit length. The K most important principal components are the factor realizations and the factor loadings on these observed factors can be estimated using regression techniques[17]. The way of how we get each principal component is described below:

Let $\hat{\mathbf{V}}$ be the sample covariance matrix of returns and $\mathbf{R}_t$ denote the asset returns at time t. The first sample component is $\mathbf{x}_1^{*'}\mathbf{R}_t$ where the $(N \times 1)$ vector $\mathbf{x}_1^{*'}$ solves

$$\max_{x_1} \mathbf{x}_1'\hat{\mathbf{V}}\mathbf{x}_1 \quad \text{s.t. } \mathbf{x}_1'\mathbf{x}_1 = 1 .$$

The solution $\mathbf{x}_1^{*'}$ is the eigenvector associated with the largest eigenvalue of $\hat{\mathbf{V}}$. Similarly the second principal component is $\mathbf{x}_2^{*'}\mathbf{R}_t$ where $\mathbf{x}_2^{*}$ solves

$$\max_{x_2} \mathbf{x}_2'\hat{\mathbf{V}}\mathbf{x}_2 \quad \text{s.t. } \mathbf{x}_2'\mathbf{x}_2 = 1 \text{ and } \mathbf{x}_1^{*'}\mathbf{x}_2 = 0 .$$

The same process is repeated until the kth principal component. The factor loadings β_i for each asset and the residual variances σ_i^2 can be estimated applying the *Ordinary Least Square* (OLS) approach from the following equation[18]:

$$R_{it} = a_i + \boldsymbol{\beta}_i'\hat{\mathbf{f}}_t + \varepsilon_{it}, \quad t = 1,\ldots,T \tag{1.3.11}$$

The factor model covariance matrix of returns is:

$$\hat{\mathbf{V}} = \hat{\mathbf{B}}\hat{\mathbf{V}}^F\hat{\mathbf{B}}' + \hat{\mathbf{D}} \tag{1.3.12}$$

where

$$\hat{\mathbf{B}} = \begin{pmatrix} \hat{\beta}_1' \\ \vdots \\ \hat{\beta}_2' \end{pmatrix}, \qquad \hat{\mathbf{D}} = \begin{pmatrix} \hat{\sigma}_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{\sigma}_N^2 \end{pmatrix}$$

and

[17] Zivot, E., Wang, J., 2006, '*Modeling Financial Time Series with S-PLUS*' page 597

[18] Zivot, E., Wang, J., 2006, '*Modeling Financial Time Series with S-PLUS*' page 598

$$\hat{\mathbf{V}}^F = \frac{1}{T-1}\sum_{t=1}^{T}(\hat{\mathbf{f}}_t - \bar{\mathbf{f}})(\hat{\mathbf{f}}_t - \bar{\mathbf{f}})'.$$

Implementing the PCA to the data we obtained the *factor loadings* known as *betas* which are illustrated below. Four factors is a sufficient number to explain most of the variability of the assets.

Figure 1.6

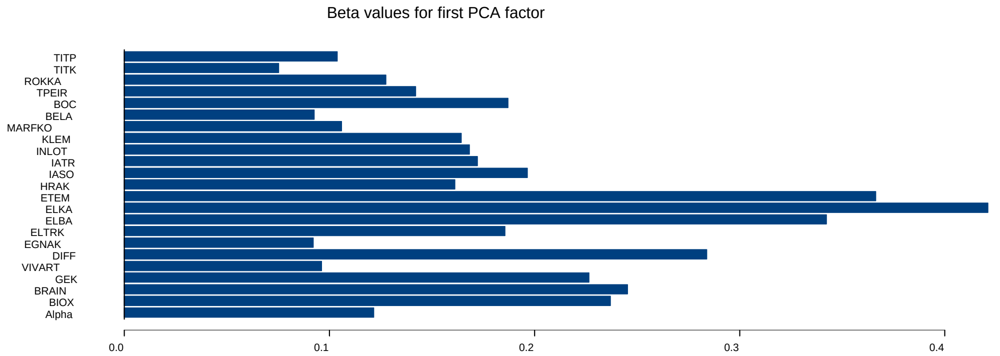
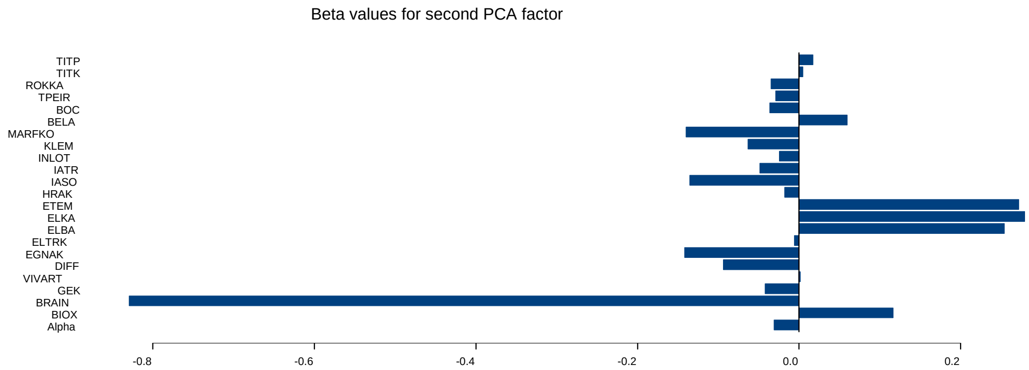


Figure 1.7



A sample of the correlation matrix of returns as well as the *beta values* of the third and fourth principal component factor are representing in the '*Appendix 1.4*'.

We may observe a lot of similarities between *figure 1.6* and *figure 1.2*. From the factor analysis is clear that there not exist assets which are exposed to the same factors. But as we have mentioned several times, the factor modelling is a useful tool to determine possible co-integrated stocks, since the stocks which are exposed to the same common factors, are expected to have similar returns and as a result a common trend. The above results could give us some hints about which stocks to test for the co-integrated modelling. An important observation is that different approaches sometimes suggest different assets to be co-integrated. To be more specific, for instance *PCA, Statistical Factor Model* and *Single Factor Model* approaches seems to suggest that the stocks '*ELBA*' and '*ETEM*' are co-integrated whereas *Fundamental Factor Model* suggests that possibly are not co-integrated. The above argument does not imply that '*Factor Modelling*' is not an efficient approach. Contrary is a tool that an analyst should take it into account in order to continue his/her analysis. Now the next step is to examine which stocks have a common trend, using Granger's and Johansen's approach determining the co-integration vectors or the co-integration coefficient for the bivariate case.

2.1 Co-Integration

At this point it would be appropriate to give a brief description about the notion of co-integration. It is clear that non-stationary time-series are not suitable for financial modeling and for that reason analysts used to difference them in order to make them stationary. In 1987 *Engle* and *Granger*[19] showed that it is possible for a linear combination of integrated time-series to be stationary. If this is the case then the time-series are said to be co-integrated.

To give an example, let $\mathbf{X}_t = (x_{1t},\dots,x_{nt})'$ denote an $(n\times 1)$ vector of $I(1)$ [20] time-series. Then $\mathbf{X}_t$ is said to be co-integrated if there exists an $(n\times 1)$ vector $\boldsymbol{\gamma} = (\gamma_1,\dots,\gamma_n)'$ such that:

$$\boldsymbol{\gamma}'\mathbf{X}_t = \gamma_1 x_{1t} + \dots + \gamma_n x_{nt} = \varepsilon_t \sim I(0) \qquad (2.1)$$

where ε_t is the *equilibrium error*, or in our case the *spread*. The linear combination $\boldsymbol{\gamma}'\mathbf{X}_t$ is referred to as a *long-run equilibrium* relationship. The idea is that $I(1)$ time-series with such a relationship cannot diverge too far apart from the equilibrium because economic forces, which sometimes cannot be explained, will act to re-establish the *equilibrium* relationship.

Generally the components of the $\mathbf{X}_t$ vector are said to be co-integrated of order *d, b* denoted by $\mathbf{X}_t \sim CI(d,b)$ if:

1. All the components of $\mathbf{X}_t$ are integrated of order *d*

2. There exists a vector $\boldsymbol{\gamma}'$ such that the linear combination $\boldsymbol{\gamma}'\mathbf{X}_t$ of order $(d-b)$ where $b>0$.

In our case, since the logarithm of the stock prices are $I(1)$ processes, we consider the case $\mathbf{X}_t \sim CI(1,1)$ [21]. Finally If $\mathbf{X}_t$ has n components there may be as many as $n-1$ linearly independent co-integrating vectors.

To be more specific, let $\mathbf{X}_t$ be a $(n\times 1)$ co-integrated vector, then there may be $0<r<n$ linearly independent co-integrated vectors. For instance, let $n=3$ and assume that there are $r=2$ co-integrated vectors $\boldsymbol{\gamma}_1 = (\gamma_{11},\gamma_{12},\gamma_{13})'$ and $\boldsymbol{\gamma}_2 = (\gamma_{21},\gamma_{22},\gamma_{23})'$. Then

$$\boldsymbol{\gamma}_1'\mathbf{X}_t = \gamma_{11}x_{1t} + \gamma_{12}x_{2t} + \gamma_{13}x_{3t} \sim I(0)$$

---

[19] Engle, R.F., Granger, C.W.J., *1987, 'Co-Integration and Error Correction: Representation, Estimation and Testing,'*, Econometrica, 55, 251-276

[20] $I(1)$ is read as "Integrated of order one" and we refer to time-series which are not stationary and we should take the first difference in order to obtain a stationary process. On contrary $I(0)$ referred to stationary time-series.

[21] Notice: All variables must be integrated of the same order.

and

$$\boldsymbol{\gamma}_2^{'}\mathbf{X}_t = \gamma_{21}x_{1t} + \gamma_{22}x_{2t} + \gamma_{23}x_{3t} \sim I(0)$$

The following matrix:

$$\boldsymbol{\Gamma}^{'} = \begin{pmatrix} \boldsymbol{\gamma}_1^{'} \\ \boldsymbol{\gamma}_2^{'} \end{pmatrix} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \end{pmatrix}$$

forms a *basis* for the space of co-integrating vectors.

2.2 Common Trends

Recall the equation (1.3.2). As we have mentioned in *section 1.3* the logarithm of the stock price may be considered as the sum of two components, which are a stationary process and a non-stationary process, and the basic requirement for two or more time-series to be co-integrated is to share a common trend. Formally if the $(n \times 1)$ vector $\mathbf{X}_t$ is co-integrated with $0 < r < n$ co-integrating vectors then there exist $n - r$ *common $I(1)$ stochastic trends*. To illustrate the duality between co-integration and common trends, let

$$\mathbf{X}_t = (x_{1t}, x_{2t})^{'} \sim I(1)$$

where

$$x_{1t} = \log(p_{1t}) = \gamma_2 \sum_{i=1}^{t} \varepsilon_{1i} + \varepsilon_{3t} \quad \text{and} \quad x_{2t} = \log(p_{2t}) = \sum_{i=1}^{t} \varepsilon_{1i} + \varepsilon_{2t}$$ [22]

and

$$\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})^{'} \sim I(0).$$

Clearly in this case, the common stochastic trend is $\sum_{i=1}^{t} \varepsilon_{1i}$ and $\mathbf{X}_t$ is co-integrated with co-integrating vector $\boldsymbol{\gamma} = (1, -\gamma_2)^{'}$. Notice that the co-integrating relationship eliminates the common stochastic trend, obtaining a stationary process:

$$\boldsymbol{\gamma}^{'}\mathbf{X}_t = \gamma_2 \sum_{i=1}^{t} \varepsilon_{1i} + \varepsilon_{3t} - \gamma_2 \left(\sum_{i=1}^{t} \varepsilon_{1i} + \varepsilon_{2t} \right) = \varepsilon_{3t} - \gamma_2 \varepsilon_{2t} \sim I(0).$$

2.3 Residual-Based Test for Co-integration

The question which arises now, is how can we define if the $\mathbf{X}_t$ is indeed co-integrated. Recall that the $(n \times 1)$ vector $\mathbf{X}_t \sim I(1)$ is co-integrated with $0 < r < n$ co-integrating vectors if there exist an $(r \times n)$ matrix $\boldsymbol{\Gamma}^{'}$ such that

---

[22] Let p_{1t} be the price of the stock 1, as in equation (1.3.2), and $n_{p_t} = \sum_{i=1}^{t} \varepsilon_{1i}$ respectively.

$$\mathbf{\Gamma}'\mathbf{X}_t = \begin{pmatrix} \boldsymbol{\gamma}_1'\mathbf{X}_t \\ \vdots \\ \boldsymbol{\gamma}_r'\mathbf{X}_t \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{rt} \end{pmatrix} \sim I(0)$$

At this point we will explain the co-integration test which was originally considered by Engle and Granger[23] where we test if *there is at most one co-integrating vector* and is a simple two-step residual-based testing procedure based on regression techniques. The two-step procedure for determining if the $(n \times 1)$ vector $\boldsymbol{\gamma}$ is a co-integrating vector is as follows:

- Form the co-integrating residual $\boldsymbol{\gamma}'\mathbf{X}_t = \varepsilon_t$ (or $\hat{\boldsymbol{\gamma}}'\mathbf{X}_t = e_t$) estimating the co-integrating vector from the data, usually using regression techniques.
- Perform a unit root test on ε_t (or on e_t) to determine if it is $I(0)$.

The null hypothesis in the *Engle-Granger procedure* is that there is no co-integrating vector, and the alternative is that there is at most one. Hence the hypothesis to be tested is:

$$H_0: \ \varepsilon_t = \boldsymbol{\gamma}'\mathbf{X}_t \sim I(1)$$
$$H_1: \ \varepsilon_t = \boldsymbol{\gamma}'\mathbf{X}_t \sim I(0)$$

The above hypotheses may be tested implementing any *unit root tes*t statistic, where the most widely-used are the *ADF* and *PP* which we will discuss in the next section. Co-integration is found if the unit root test rejects the no-cointegration null hypothesis.

2.4 Unit Root Tests and Stationarity

Economic and finance theory often suggests the existence of long-run equilibrium relationships among non-stationary time-series variables. If these variables are $I(1)$ and they share a common trend, then co-integration techniques can be used to model these long-run relations. Hence in our case we have to determine if the times-series variables are $I(1)$ and then applying the *Engle-Granger procedure* to determine if the residuals are $I(0)$ (*see section 2.3*). Thus, pre-testing for unit roots is often a first step in the co-integration modelling.

For the beginning, we start with the autoregressive unit root tests. Considering a simple $\mathrm{AR}(1)$ model of the form:

$$y_t = \phi y_{t-1} + \varepsilon_t, \text{ where } \varepsilon_t \sim WN(0, \sigma^2)$$

[23] Engle, R.F., Granger, C.W.J., *1987, 'Co-Integration and Error Correction: Representation, Estimation and Testing,'*, Econometrica, 55, 251-276

the hypotheses to be tested are:

$$H_0: \ \phi = 1 \Rightarrow y_t \sim I(1)$$
$$H_1: \ |\phi| < 1 \Rightarrow y_t \sim I(0)$$

The test statistic is[24]:

$$t_{\phi=1} = \frac{\hat{\phi} - 1}{SE(\hat{\phi})}$$

where $\hat{\phi}$ is the least squares estimates and $SE(\hat{\phi})$ is the usual standard error estimate[25].

The limiting distribution of $t_{\phi=1}$ is called the *Dickey-Fuller* (DF) distribution and does not have a closed representation. As a result, quantiles of the distribution must be computed by numerical approximation or by simulation[26].

When testing for unit roots, it is very important to specify the null and the alternative hypotheses in an appropriate way in order to characterize the trend properties of the data. For this reason we outline the two most popular trend cases[27].

Case I: Constant only

In this case the regression form is as follows:

$$y_t = c + \phi y_{t-1} + \varepsilon_t$$

The hypotheses to be tested are:

$$H_0: \ \phi = 1 \Rightarrow y_t \sim I(1) \text{ without drift}$$
$$H_1: \ |\phi| < 1 \Rightarrow y_t \sim I(0) \text{ without non-zero mean}$$

Case II: Constant and Time Trend

In this case the regression form is as follows:

$$y_t = c + \delta t + \phi y_{t-1} + \varepsilon_t$$

The hypotheses to be tested are:

$$H_0: \ \phi = 1 \Rightarrow y_t \sim I(1) \text{ with drift}$$
$$H_1: \ |\phi| < 1 \Rightarrow y_t \sim I(0) \text{ with deterministic time trend}$$

---

[24] See Phillips, P.C.B., 1987, '*Time Series Regression with a Unit Root*', Econometrica, 55, 227-301
[25] Hamilton, J., 1994, '*Time Series Analysis*' Princeton University Press
[26] Dickey, D. Fuller, W., 1979, '*Distribution of the Estimators for Autoregressive Time Series with a Unit Root*', Econometrica, 49, 1057-1072
[27] Zivot, E., Wang, J., 2006, '*Modeling Financial Time Series with S-PLUS*' pages 119-121

2.4.1 Dickey-Fuller Unit Root Test

The unit root test described above are valid of the time series y_t which are well characterized by an $\mathrm{AR}(1)$ model with white noise errors. Since many financial time series have more complicated dynamics, *Said and Dickey (1984)* augment the basic autoregressive unit root test in order to accommodate general $ARMA(p,q)$ models. This test is referred to as the *augmented Dickey-Fuller* test (ADF)[28] and tests the null hypothesis that a time-series y_t is $I(1)$ against the alternative that it is $I(0)$ assuming that the dynamics in the data have an $ARMA$ structure. The ADF test is based on estimating the test regression:

$$y_t = \beta'\mathbf{D}_t + \phi y_{t-1} + \sum_{j=1}^{p} \psi_j \Delta y_{t-j} + \varepsilon_t \tag{2.4.1}$$

where $\mathbf{D}_t$ is a vector of deterministic terms (constant, time trend etc). The p lagged difference terms, Δy_{t-j} are used in order to approximate the $ARMA(p,q)$ structure of the errors, and the ε_t are serially uncorrelated and assuming that are not heteroskedastic. An important practical issue for the application of the *augmented Dickey-Fuller* test is the appropriate specification of the lag length p.

A widely-used lag length selection is as follows[29]:

- Set an upper bound $p_{\max}$[30] for p.
- Estimate the ADF test regression with $p = p_{\max}$. If the absolute value of the t-statistic for testing the significance of the last lagged difference is grater than 1.6 then set $p = p_{\max}$ and perform the unit root test. Otherwise, reduce the lag length by one and repeat the process.

2.4.2 Philips-Perron Unit Root Tests

Another popular unit root test is the Philips-Perron (PP) test[31] which differs from the ADF test in how it deal with serial correlation and heteroskedasticity in the error terms. Particularly, PP test ignores any serial correlation in the test regression which has the form:

$$\Delta y_t = \beta'\mathbf{D}_t + \pi y_{t-1} + u_t$$

---

[28] Said, S.E., Dickey, D., 1984, 'Testing for Unit Roots in Autoregressive Moving-Average Models with Unknown Order', Biometrika, 71, 599-607

[29] Ng, S., Perron, P., 1995, '*Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag*', Journal of the American Statistical Association, 90, 268-281

[30] Schwert (1989) suggests that $p_{\max} = \left[12 \cdot \left(\frac{T}{100}\right)^{1/4}\right]$.

[31] Phillips, P.C.B., Perron, P., 1988, '*Testing for Unit Roots in Time Series Regression*', Biometrika, 75, 335-346

The great advantages of the PP test over the ADF test is that the former is robust to general forms of heteroskedasticity in the error terms, and that the analyst does not have to specify a lag length for the test regression.

2.4.3 Stationarity Tests

As we have already mentioned, the ADF and PP unit root test are for the null hypothesis that a times-series variable is $I(1)$, whereas the stationarity tests are for the null hypothesis that the time-series variable is $I(0)$. The most popular stationarity test is the KPSS[32] test which is of the form:

$$y_t = \beta'\mathbf{D}_t + \mu_t + u_t$$

where

$$\mu_t = \mu_{t-1} + \varepsilon_t,\ \varepsilon_t \sim WN(0, \sigma_\varepsilon^2).$$

The null hypothesis is that y_t is $I(0)$ and is of the form:

$$H_0:\ \sigma_\varepsilon^2 = 0$$
$$H_1:\ \sigma_\varepsilon^2 > 0$$

The KPSS test statistic is the Lagrange multiplier (LM) and is given by:

$$KPSS = \left(T^{-2} \sum_{t=1}^{T} \hat{S}_t^2 \right) / \hat{\lambda}^2$$

where

$\hat{S}_t = \sum_{j=1}^{t} \hat{u}_j$ and $\hat{\lambda}^2$ is a consistent estimate of the long-run variance of u_t using $\hat{u}_t$.

In '*Appendix section 2.1*' there is an extensive example of how can we test for co-integration in practice.

[32] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992, '*Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root*' , Journal of Econometrics, 54, 159-178

3.1 Error Correction Models

The fact that the time-series variables are co-integrated implies that there is some adjustment process which prevents the errors in the long-run relationship becoming larger and larger. Engle and Granger[33] have shown that any co-integrated series have an error correction representation. The converse is also true, in that co-integration is a necessary condition for error correction models to hold. For simplicity we consider a bivariate $I(1)$ vector $\mathbf{X}_t = (x_{1t}, x_{2t})'$ and assume that $\mathbf{X}_t$ is co-integrated with a normalized co-integrating vector $\boldsymbol{\gamma} = (1, -\gamma_2)'$ so that $\boldsymbol{\gamma}'\mathbf{X}_t = x_{1t} - \gamma_2 x_{2t}$ is $I(0)$. Then the co-integration implies the existence of an *error correction model* (ECM) of the form:

$$\Delta x_{1t} = c_1 + a_1(x_{1t-1} - \gamma_2 x_{2t-1}) + \sum_j \psi_{11}^j \Delta x_{1t-j} + \sum_j \psi_{12}^j \Delta x_{2t-j} + \varepsilon_{1t} \tag{3.1.1}$$

$$\Delta x_{2t} = c_2 + a_2(x_{1t-1} - \gamma_2 x_{2t-1}) + \sum_j \psi_{21}^j \Delta x_{1t-j} + \sum_j \psi_{22}^j \Delta x_{2t-j} + \varepsilon_{2t} \tag{3.1.2}$$

These two equations describe the dynamic behavior of x_{1t} and x_{2t} respectively. The ECM links the long-run equilibrium relationship implied by co-integration with the short-run dynamic adjustment mechanism that describes how the time-series variables react when they move out of the long-run equilibrium. Let us now interpret the above equations. The left-hand side is the increment to the time-series variables at each time step. The term $x_{1t-1} - \gamma_2 x_{2t-1}$ is representative of the deviation from the long-run equilibrium and a_1 is the error correction rate, indicative of the speed with which the time series corrects itself to maintain equilibrium. Hence as the two time-series evolve with time, the deviations from the long-run equilibrium are caused by a white noise ε_{it} and these deviations are corrected in the future time steps. As we may observe from the equations above, in order to estimate the parameters of the equations, it is necessary to know the value of the co-integrating vector $\boldsymbol{\gamma} = (1, -\gamma_2)'$. As mentioned in *Chapter 2* the Engle-Granger procedure consists of two steps. First we estimate the co-integrating vector $\boldsymbol{\gamma}$ by OLS and then test for stationarity of the residuals. One doubt about the Engle-Granger procedure concerns the estimate of the co-integrating vector. As we know the variables x_{1t} and x_{2t} are non-stationary. Hence, what is the sense of estimating $\boldsymbol{\gamma}$ by OLS? Further investigation has shown that the properties of the OLS estimators are not all bad. Indeed, under some general assumptions, they are consistent. But the problem is that the OLS estimate of the standard error of $\hat{\gamma}_2$ is not consistent in this case. The above arguments indicate that the OLS estimator of the co-integrating vector could be improved upon, A simple improvement is suggested by Stock and Watson[34] and will be discussed briefly in the next section.

[33] Engle, R.F., Granger, C.W.J., *1987, 'Co-Integration and Error Correction: Representation, Estimation and Testing,'*, Econometrica, 55, 251-276
[34] Stock, J.H., Watson, M.W., 1993, '*A Simple Estimator of Co-integrating Vectors in Higher Order Integrated Systems*', Econometrica, 61, 783-820.

3.2 Stock and Watson's Efficient Lead/lag Estimator

At this point we represent an improved method of estimating the co-integrating vector $\boldsymbol{\gamma}$ which proposed by Stock and Watson (1993) who show a relative simple method for estimating an asymptotically efficient estimator for the normalized co-integrating vector $\boldsymbol{\gamma}_2$ and a formula for computing its asymptotic variance[35]. Below we represent the *Stock and Watson's* procedure which could also be used in the bivariate case in order to obtain the co-integration vector $\boldsymbol{\gamma} = (1, -\gamma_2)'$ and then to apply this estimator to equations (3.1.1) and (3.1.2).

Let $\mathbf{X}_t = (x_{1t}, \mathbf{X}_{2t}^{'})'$ and let the co-integrating vector be normalized as $\boldsymbol{\gamma} = (1, -\boldsymbol{\gamma}_2^{'})'$. Stock and Watson's efficient estimator procedure is:

- Augment the co-integrating regression of x_{1t} on $\mathbf{X}_{2t}^{'}$ as follows:

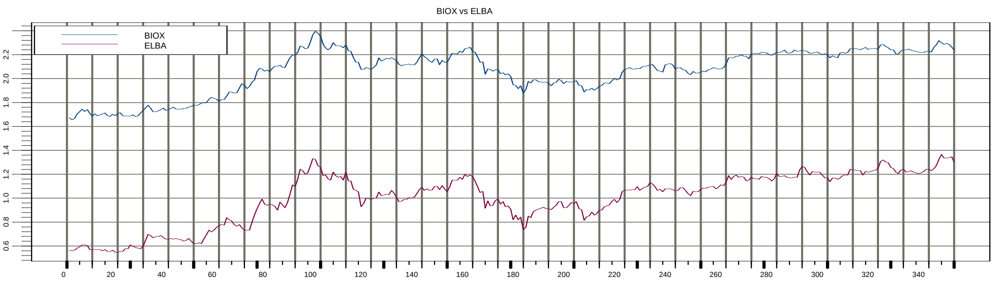
$$x_{1t} = \boldsymbol{\beta}'\mathbf{D}_t + \boldsymbol{\gamma}_2^{'}\mathbf{X}_{2t} + \sum_{j=-p}^{p} \psi_j^{'} \Delta\mathbf{X}_{2t+j} + u_t \qquad (3.2.1)$$

where $\mathbf{D}_t$ is a vector of deterministic terms (constant, time trend etc).

- Estimate the (3.2.1) equation by OLS. The $\boldsymbol{\gamma}_2$ estimator is called the dynamic ordinary least square estimator (DOLS).

Below we represent an example of the DOLS estimation of co-integrating vector considering again the logarithm of two stocks ('BIOX' and 'ELBA').

Figure 3.1



Source: www.euro2day.gr. The prices are obtained from 21/9/2005 to 12/2/2007.

In our case the model that we would like to estimate is as follows:

$$BIOX = \hat{\mu} + \hat{\gamma}_2 ELBA$$

where 'BIOX' and 'ELBA' are transformed into a logarithmic scale. Appling the OLS the results that we obtain are representing below.

[35] See Hamilton (1994) chapter 19, and Hayashi (200) chapter 10

Table 3.1

```
Coefficients:
               Value Std. Error  t value Pr(>|t|)
(Intercept)   1.2316   0.0108  114.1260   0.0000
       ELBA   0.8299   0.0106   78.6056   0.0000

Regression Diagnostics:

         R-Squared 0.9467
Adjusted R-Squared 0.9465
Durbin-Watson Stat 0.1460
```

Hence the model is of the form: $BIOX = 1.2316 + 0.8299 \times ELBA$.

At this point we estimate the co-integrating vector using least Stock and Watson's dynamic OLS estimator computed from (3.2.1) with x_{1t}=BIOX, $\mathbf{D}_t$=1, $\mathbf{X}_{2t}$=ELBA and $p = 3$. The results are representing below.

Table 3.2

```
Coefficients:
                  Value Std. Error  t value Pr(>|t|)
  (Intercept)    1.2236   0.0176    69.4832   0.0000
         ELBA    0.8392   0.0174    48.3669   0.0000
 D.ELBA.lead3    0.0637   0.0917     0.6954   0.4873
 D.ELBA.lead2    0.1122   0.0915     1.2264   0.2209
 D.ELBA.lead1    0.1220   0.0930     1.3126   0.1902
  D.ELBA.lag0   -0.1538   0.0856    -1.7972   0.0732
  D.ELBA.lag1   -0.1203   0.0756    -1.5904   0.1127
  D.ELBA.lag2   -0.1084   0.0845    -1.2827   0.2005
  D.ELBA.lag3   -0.1143   0.0736    -1.5522   0.1216

 Regression Diagnostics:

          R-Squared 0.9475
 Adjusted R-Squared 0.9463
 Durbin-Watson Stat 0.1268
```

In this case the estimated model is $BIOX = 1.2236 + 0.8392 \times ELBA$ which is slightly different than the previous one, but in this case the estimator is asymptotically efficient and should be preferred.

3.3 Error Correction Models in Practice

In order to estimate Error Correction Models of the form (3.1.1) and (3.1.2) the time-series variables must be co-integrated and vice-versa. Let us continue our example. First of all both variables are $I(1)$ and the co-integrating vector is obtained by OLS or DOLS. Now the next step is to examine if the residuals are indeed $I(0)$ and stationary. The results of the ADF and PP tests are representing below:

Table 3.3

Test for Unit Root: Augmented DF Test	Test for Unit Root: Phillips-Perron Test
Null Hypothesis: there is a unit root Type of Test: t-test Test Statistic: -2.983 **P-value: 0.002896**	**Null Hypothesis: there is a unit root** Type of Test: t-test Test Statistic: -3.32 **P-value: 9.447e-4**

As we can see both tests reject the null hypothesis. The next step now is to estimate ECM by least squares. In this example we consider the ECM of the form:

$$\Delta_{Bt} = c_B + a_B(B_{t-1} - \hat{\gamma}E_{t-1} - \hat{\mu}) + \psi_{11}\Delta_{Bt-1} + \psi_{12}\Delta_{\mathrm{E}t-1} + \varepsilon_{Bt}$$
$$\Delta_{Et} = c_E + a_E(B_{t-1} - \hat{\gamma}E_{t-1} - \hat{\mu}) + \psi_{21}\Delta_{Bt-1} + \psi_{22}\Delta_{\mathrm{E}t-1} + \varepsilon_{Et}$$

where the reactions of 'BIOX' and 'ELBA' to the disequilibrium error are captured by the adjustment coefficients a_B and a_E. Notice that in order to estimate the ECM, the number of lags need to be determined. This could be done using test statistics for the significance of the lagged terms or model selection criteria like AIC and BIC. Using the estimator obtained by DOLS, we are in position to estimate the ECM. The results are representing below:

$$\Delta_{Bt} = 0.0015 - 0.0253(B_{t-1} - \hat{\gamma}E_{t-1} - \hat{\mu}) + 0.155\Delta_{Bt-1} - 0.0568\Delta_{\mathrm{E}t-1}$$
$$\Delta_{Et} = 0.0022 + 0.0569(B_{t-1} - \hat{\gamma}E_{t-1} - \hat{\mu}) + 0.0960\Delta_{Bt-1} - 0.0540\Delta_{\mathrm{E}t-1}$$

Since the lagged terms are not found statistically significant, the ECM could be re-estimated and the results are representing below:

$$\Delta_{Bt} = 0.0015 - 0.0185(B_{t-1} - \hat{\gamma}E_{t-1} - \hat{\mu})$$
$$\Delta_{Et} = 0.0022 + 0.0625(B_{t-1} - \hat{\gamma}E_{t-1} - \hat{\mu})$$

In '*Appendix 3.1*' there is a presentation of the calculations with more details. Note that the adjustment coefficients a_B and a_E are very low which implies that the error correction mechanism does not work very fast, and may take long time for the time-series variable to return back to their equilibrium relationship. An important observation is that the adjustment coefficients are set to opposite signs, which is very important for error correcting behaviour. In '*Appendix 3.1*' one may observe that in our example, the *p-value* of the adjustment coefficients is very high, which implies that these values can be considered equal to zero, and as a result the two time series not to be co-integrated. Of course we must be alert that is very difficult to find stocks which share exactly a common trend. The most examples in finance regarding the co-integrated time series are referred to 'income-consumption', exchange interest rate etc, where economic forces act to retain this equilibrium relationships. In case of stocks, the economic forces are unobservable, expect that some stocks are exposed to the same risk factors as we have explained in '*Chapter 1*'. Since this study is an empirical one, it is very important to see how these co-integrated series work in practice. For this reason we represent again the same strategy as that one which is described in '*Chapter 2*' for the stocks '*BIOX*' and '*ELBA*'.

In order to test how a pairs trading strategy works in practice we have fitted the model applying the dynamic OLS taking 350 observations (21/9/2005-12/2/2007), and then we applied the strategy to the next 99 trading days (13/2/2007-6/7/2007). As a threshold of a trading signal we considered again a standard error of the estimated residuals. The next plot illustrates the future movements of the logarithm of these two stocks.

Figure 3.2

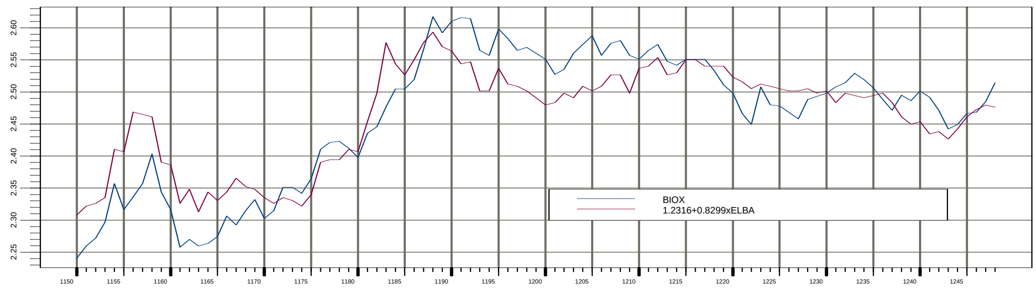
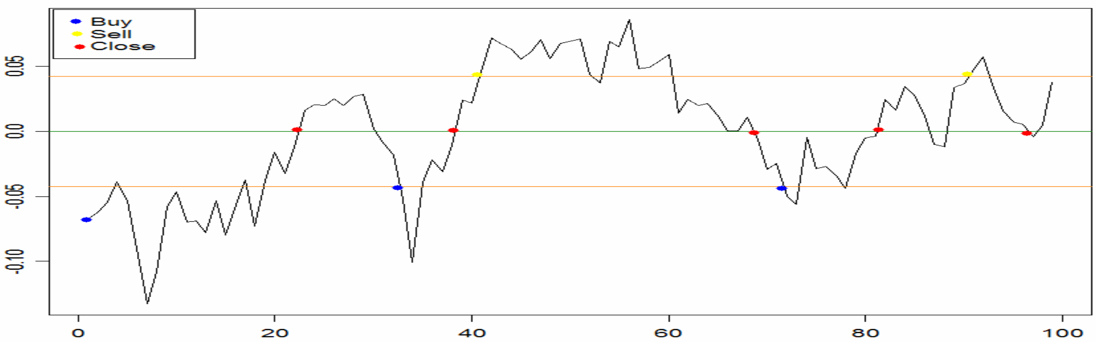


Figure 3.3



The results that we obtained are as follows:

13/2/2007: Buy the spread at 6.36€
16/3/2007: Sell the spread back at 7.36€
30/3/2007: Buy the spread at 7.72€
11/4/2007: Sell the spread back at 9.41€
13/4/2007: Short-Sell the spread at 9.47€
24/5/2007: Close the position paying back 8.58€
30/5/2007: Buy the spread at 7.88€
13/6/2007: Sell the spread back at 8.53€
26/6/2007: Short-Sell the spread at 8.58€
4/7/2007 : Close the position paying back 8.09€

The net profit of the strategy is equal to **4.71€** and as we can see the model shows a relative strong mean reverting strategy since it oscillates around zero 10 times in 100 days time period.

4.1 VAR Models and Co-Integration

As we have seen earlier, the Granger-Engle procedure links co-integration to error correction models. In 1988, the mathematician S. Johansen[36] managed to root co-integration and error correction models in a vector autoregression framework, known as vector error correction models (VECM). In this chapter we try to give a brief description of Johansen's approach to co-integration modeling and then to implement it. First of all let us start with the *vector autoregressive* (VAR) *model* which is an extension of the univariate autoregressive model to dynamic multivariate time-series variables, and has proven to be useful for describing the dynamics of financial time-series and forecasting.

4.2 Stationary VAR model

The basic p-lag Vector Autoregressive model has the following form:

$$\mathbf{Y}_t = \mathbf{c} + \sum_{i=1}^{p} \mathbf{\Pi}_i \mathbf{Y}_{t-i} + \varepsilon_t \tag{4.2.1}$$

where $\mathbf{\Pi}_i$ is an $n \times n$ coefficient matrix and ε_t is an $n \times 1$ white noise vector process, with zero mean and serially uncorrelated or independent with time invariant covariance matrix $\mathbf{\Sigma}$. For simplicity let us consider a bivariate $VAR(2)$ model of the form:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \pi_{11}^1 & \pi_{12}^1 \\ \pi_{21}^1 & \pi_{22}^1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \pi_{11}^2 & \pi_{12}^2 \\ \pi_{21}^2 & \pi_{22}^2 \end{pmatrix} \begin{pmatrix} y_{1t-2} \\ y_{2t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Notice that in lag operator notation, the $VAR(p)$ could be written as follows:

$$\mathbf{\Pi}(L)\mathbf{Y}_t = \mathbf{c} + \varepsilon_t$$

where $\mathbf{\Pi}(L) = \mathbf{I}_n - \sum_{i=1}^{p} \mathbf{\Pi}_i L^i$ and the $VAR(p)$ is stable if the roots of

$$\det\left(\mathbf{\Pi}(L)\right) = 0 \tag{4.2.2}$$

lie outside the unit circle or in other words, if the eigenvalues of the matrix

[36] Johansen, S., 1988, *'Statistical Analysis of Cointegrating Vectors'*, Journal of Economic Dynamics and Control, 12, 231-254

$$\mathbf{F} = \begin{pmatrix} \mathbf{\Pi}_1 & \mathbf{\Pi}_2 & \cdots & \mathbf{\Pi}_n \\ \mathbf{I}_n & 0 & \cdots & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & \mathbf{I}_n & 0 \end{pmatrix}$$

have modulus less than one.

The basic $VAR(p)$ which we have discussed above, sometimes seems to be not sufficient to describe the dynamics of the time-series variables and for that reason some other deterministic terms, such as a time trend, or seasonality may be required to represent the data properly. Hence the general form of the $VAR(p)$ model with deterministic terms is of the form:

$$\mathbf{Y}_t = \mathbf{\Phi}\mathbf{D}_t + \sum_{i=1}^{p} \mathbf{\Pi}_i \mathbf{Y}_{t-i} + \varepsilon_t$$

where $\mathbf{D}_t$ is an $l \times 1$ matrix of deterministic terms (constant, trend, seasonal dummy variables etc.) and $\mathbf{\Phi}$ is the parameter matrix.

4.3 Vector Error Correction Model

In the previous section we mentioned that a VAR model is stationary if in equation (4.2.2) all roots lie outside the complex unit circle. Since our scope is to model the mispricing of stock dynamics where they are $I(1)$ processes the implementation of a VAR model is not the appropriate one because the co-integrating relationships are not explicitly apparent. Notice that if (4.2.2) has a root on the unit circle then this implies that some or all of the variables in $\mathbf{Y}_t$ are $I(1)$ and they may be co-integrated i.e. there exists some linear combination of the variables in $\mathbf{Y}_t$ that is $I(0)$. Now if we re-parameterize the (4.2.1) model to a *vector error correction model* (VECM) then the co-integrating relationships become apparent. The VECM is of the form:

$$\Delta\mathbf{Y}_t = \mathbf{\Phi}\mathbf{D}_t + \mathbf{\Pi}\mathbf{Y}_{t-1} + \sum_{i=1}^{p-1} \mathbf{\Gamma}_i \Delta\mathbf{Y}_{t-i} + \varepsilon_t \tag{4.3.1}$$

where the *long-run impact matrix* $\mathbf{\Pi} = \sum_{i=1}^{p} \mathbf{\Pi}_i - \mathbf{I}_n$ and the *short-run impact matrix* $\mathbf{\Gamma}_k = -\sum_{j=k+1}^{p} \mathbf{\Pi}_j$, k=1,…,$p-1$. Clearly since in our case the $\mathbf{Y}_t$ are $I(1)$, in VECM (4.3.1) , $\Delta\mathbf{Y}_t$ and its lags are $I(0)$, whereas the term $\mathbf{\Pi}\mathbf{Y}_{t-1}$ is the only one which includes potential $I(1)$ variables. But since the left hand side of the VECM is $I(0)$ then it must be the case that $\mathbf{\Pi}\mathbf{Y}_{t-1}$ is also $I(0)$. In other words, $\mathbf{\Pi}\mathbf{Y}_{t-1}$ must contain the co-integrating relationships if they exist.

Hence the question of co-integration may be formulated in terms of the rank of the $\mathbf{\Pi}$ matrix. If the rank of the $(n \times n)$ matrix $\mathbf{\Pi}$ is n, then it will be non-singular and one may obtain a long-run equilibrium solution for the system of stochastic difference equations in (4.2.1). This also implies that the elements of $\mathbf{Y}_t$ are covariance stationary. Now if the $VAR(p)$ model has unit roots then from (4.2.2) we can see that $\mathbf{\Pi}$ is a singular matrix, which implies that it has *reduced rank*, which is $rank(\mathbf{\Pi}) = r < n$. Thus the cases that we have to examine are the following:

I. If $r = 0$ (i.e. $rank(\mathbf{\Pi}) = 0$), implies that $\mathbf{\Pi} = 0$ and the VECM (4.3.1) reduces to a $VAR(p-1)$ model in the first differences of the time-series, implying that all the elements of $\mathbf{Y}_t$ are $I(1)$ and that they are <u>not</u> co-integrated.
II. If $0 < rank(\mathbf{\Pi}) = r < n$ then this implies that $\mathbf{Y}_t$ is $I(1)$ with r linearly independent co-integrating vectors and $n-r$ common stochastic trends. In this case the matrix $\mathbf{\Pi}$ may be expressed as $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $(r \times n)$, each with rank r. The columns of $\boldsymbol{\beta}$ are the co-integrating vectors and $\boldsymbol{\beta}'\mathbf{Y}_{t-1} \sim I(0)$ yields the r co-integrating relations between the elements of $\mathbf{Y}_t$, also called *error correction terms.* Hence, they represent the stationary discrepancies of the observed values of the elements of $\mathbf{Y}_t$ at time $t-1$ from their long-run equilibrium values. On the other hand the elements of the $\boldsymbol{\alpha}$ distribute the impact of the co-integrating vectors to the evolution of $\Delta\mathbf{Y}_t$ and represent the short-run adjustments in the elements of $\mathbf{Y}_t$ to the deviations from the long-run relations[37].

It is important to understand that the factorization $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ is not unique and it only identifies the space spanned by the co-integrating relations. In order to obtain unique values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ requires further restrictions on the model[38].

4.4 Johansen's Methodology for Modelling Co-Integration

At this point we outline the four basic steps in Johansen's methodology.

I. Identify and estimate a $VAR(p)$ model for $\mathbf{Y}_t$.
II. Construct likelihood ration tests for the rank of matrix $\mathbf{\Pi}$ in order to determine the number of co-integrating vectors.
III. Impose normalization and identifying restrictions on the co-integrating vectors.
IV. From the normalized co-integrating vectors estimate the co-integrated vector error correction model (VECM) by maximum likelihood.

[37] See Johansen, S. (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

[38] See Zivot, E., Wang, J., 2006, '*Modeling Financial Time Series with S-PLUS*' Chapter 12.

To understand how deterministic terms enter the VEC model, we need to realize that it contains both the first-differenced form (the short-run part) and the level form (the long-run part). Thus, how they enter the level will determine how they will appear in the first-difference part. Hence in order to identify the $VAR(p)$ model one needs to specify the deterministic terms in (4.3.1) which are restricted to the form:

$$\mathbf{\Phi D}_t = \boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$$

In case where the deterministic terms are unrestricted the time-series variables $\mathbf{Y}_t$ may show quadratic trends there may be a linear trend term in the co-integrating relationships. The trend behavior of $\mathbf{Y}_t$ may be classified into five cases:

I. Model $H_2(r):\boldsymbol{\mu}_t = 0$ (*no constant*). In this case the restricted VECM is:

$$\Delta\mathbf{Y}_t = \boldsymbol{\alpha\beta}'\mathbf{Y}_{t-1} + \sum_{i=1}^{p-1}\mathbf{\Gamma}_i\Delta\mathbf{Y}_{t-i} + \varepsilon_t$$

which implies all the time-series variables in $\mathbf{Y}_t$ are $\mathrm{I}(1)$ without drift and the co-integrating relationships $\boldsymbol{\beta}'\mathbf{Y}_t$ have zero mean.

II. Model $H_1^*(r):\boldsymbol{\mu}_t = \boldsymbol{\alpha\rho}_0$ (*restricted constant*). The restricted VECM is:

$$\Delta\mathbf{Y}_t = \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\rho}_0) + \sum_{i=1}^{p-1}\mathbf{\Gamma}_i\Delta\mathbf{Y}_{t-i} + \varepsilon_t,$$

which implies that $\mathbf{Y}_t$ are $\mathrm{I}(1)$ without drift and the co-integrating relationships $\boldsymbol{\beta}'\mathbf{Y}_t$ have non-zero means equal to $\boldsymbol{\rho}_0$.

III. Model $H_1(r):\boldsymbol{\mu}_t = \boldsymbol{\mu}_0$ *(unrestricted constant).* The restricted VECM is:

$$\Delta\mathbf{Y}_t = \boldsymbol{\mu}_0 + \boldsymbol{\alpha\beta}'\mathbf{Y}_{t-1} + \sum_{i=1}^{p-1}\mathbf{\Gamma}_i\Delta\mathbf{Y}_{t-i} + \varepsilon_t$$

and the time-series variables $\mathbf{Y}_t$ are $\mathrm{I}(1)$ with drift vector $\boldsymbol{\mu}_0$ and the co-integrating relationships $\boldsymbol{\beta}'\mathbf{Y}_t$ may have a non-zero mean.

IV. Model $H^*(r):\boldsymbol{\mu}_t = \boldsymbol{\alpha\rho}_1 t + \boldsymbol{\mu}_0$ (*restricted trend*). The restricted VECM is:

$$\Delta\mathbf{Y}_t = \boldsymbol{\mu}_0 + \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\rho}_1 t) + \sum_{i=1}^{p-1}\mathbf{\Gamma}_i\Delta\mathbf{Y}_{t-i} + \varepsilon_t$$

In this case the series in $\mathbf{Y}_t$ are $\mathrm{I}(1)$ with drift vector $\boldsymbol{\mu}_0$ and the co-integrating relationships $\boldsymbol{\beta}'\mathbf{Y}_t$ have a linear trend term $\boldsymbol{\rho}_1 t$.

V. Model $H(r):\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$ *(unrestricted constant and trend).* The unrestricted VECM is:

$$\Delta\mathbf{Y}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t + \boldsymbol{\alpha\beta}'\mathbf{Y}_{t-1} + \sum_{i=1}^{p-1}\mathbf{\Gamma}_i\Delta\mathbf{Y}_{t-i} + \varepsilon_t.$$

In this case the time-series variables in $\mathbf{Y}_t$ are $\mathrm{I}(1)$ with linear trend and the co-integrating relations $\boldsymbol{\beta}'\mathbf{Y}_t$ have a linear trend.

4.5 Testing for the Presence of Co-Integration

Johansen's Trace Statistic

Recall that the presence of co-integration was to be tested in terms of the rank of the long-run matrix $\mathbf{\Pi}$. There are two ways of setting up this hypothesis. The *first* one may be formulated as the condition that the rank of $\mathbf{\Pi}$ is less than or equal to r. Let us $H(r)$ denote the unrestricted co-integrated VECM (4.3.1) and consider a nested set of models:

$$H(0) \subset \ldots \subset H(r) \subset \ldots \subset H(n)$$

where $H(0)$ represents the non-cointegrated VAR model with $\mathbf{\Pi} = 0$ and $H(n)$ represents an unrestricted stationary $VAR(p)$ model. This nested formulation is a convenient way to develop a sequential procedure to test for the number r of co-integrating relationships. Also note that the $rank(\mathbf{\Pi})$ is equal to the non-zero eigenvalues of $\mathbf{\Pi}$ and gives the number of co-integrating relationships in $\mathbf{Y}_t$. Johansen formulates likelihood ratio (LR) statistics for the number of co-integrating relationships as LR statistics for determining the $rank(\mathbf{\Pi})$. These tests are based on the estimated eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2 > \ldots > \hat{\lambda}_n$ of the matrix $\mathbf{\Pi}$ [39]. The LR statistic is called *the trace statistic* and is given by:

$$LR_{trace}(r_0) = 2(\hat{\ell}_U - \hat{\ell}_R) = T\left[\sum_{i=1}^{r_o} \ln(1-\lambda_i) - \sum_{i=1}^{n} \ln(1-\lambda_i)\right] = -T\sum_{i=r_0+1}^{n} \ln(1-\hat{\lambda}_i)$$

and the nested hypotheses to be tested is[40]:

$$H_0(r_0): r = r_0$$
$$H_1(r_0): r > r_0$$

The asymptotic null distribution of $LR_{trace}(r_0)$ is a multivariate version of the Dickey-Fuller unit root distribution[41] which depends on the dimension $n - r_0$ and the specification of the deterministic terms which are described in section 4.4.

[39] The calculation of the eigenvalues is described in Zivot, E., Wang, J., 2006, '*Modeling Financial Time Series with S-PLUS*' Chapter 12-Appendix.

[40] See Johansen, S. (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

[41] For critical values of this distribution see Osterwald-Lenum (1992).

Johansen's Maximum Eigenvalue Statistic

The interpretation of the trace test described above indicates that it should be applied in sequential procedure in order to determine the co-integration rank, r. One starts by testing $r=0$ against $r=1$ and, if this is rejected, moves on to testing $r=1$ against $r=2$ and so on until one hits a value of r that is not rejected. The last value of r to be tested will be $n-1$. Hence the *second* way of formulating the hypothesis of co-integration rank is:

$$H_0(r_0): r = r_0$$
$$vs$$
$$H_1(r_0): r = r_0 + 1$$

The LR statistic, in this case becomes,

$$LR_{\max}(r_0) = T\left[\sum_{i=1}^{r_o} \ln(1-\lambda_i) - \sum_{i=1}^{r_0+1} \ln(1-\lambda_i)\right] = -T\sum_{i=r_0+1}^{n} \ln(1-\hat{\lambda}_{r_0+1})$$

This statistic is called the *maximum eigenvalue statistic*. In using this statistic to carry out a sequential test as described above, the null and the alternative hypotheses continually change. The sequence again stops when H_0 is not rejected. As with the trace statistic, the asymptotic null distribution of $LR_{\max}(r_0)$ is a complicated function of Brownian motion, which depends on the dimension $n-r_0$ and the specification of the deterministic terms[42].
In '*Appendix section 4.1*' we represent an extensive example of VEC models and Johansen's Methodology.

[42] For critical values of this distribution see Osterwald-Lenum (1992).

5.1 Spurious Regression

Up to now, we have discussed the *Granger-Engle Procedure* and the *Johansen's Methodology* where help as to specify if two or more time-series variables are co-integrated. Sometimes there is a probability to derive to misleading results. One important case in which the usual statistical results do not hold is *spurious regression* when all the times-series variables are $I(1)$ and not co-integrated.

5.2 Statistical Implications of Spurious Regression

Let us consider again the $n \times 1$ vector $\mathbf{Y}_t = (y_{1t}, \ldots, y_{nt})'$ of $I(1)$ time-series variables, and let us also assume that are not co-integrated. Applying the OLS regression of y_{1t} on $\mathbf{Y}_{2t}$ we obtain the fitted model:

$$y_{1t} = \hat{\boldsymbol{\gamma}}_2^{'} \mathbf{Y}_{2t} + \hat{e}_t$$

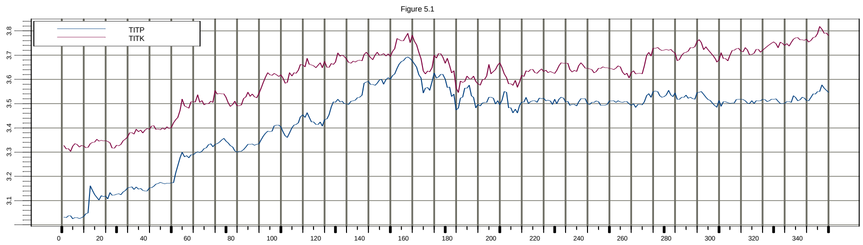
Since y_{1t} is not co-integrated with $\mathbf{Y}_{2t}$ is a spurious regression and the true value of the co-integrating vector $\boldsymbol{\gamma}_2$ should be zero. Below we outline the behavior of $\hat{\boldsymbol{\gamma}}_2^{'}$ in case we have a 'spurious regression'[43].

- A spurious regression phenomenon is that $\hat{\boldsymbol{\gamma}}_2^{'}$ does not converge in probability to zero.
- The *t-statistics* for testing that elements of $\boldsymbol{\gamma}_2$ are zero diverge to $\pm\infty$ as $\mathrm{T} \to \infty$. This implies that for large samples it will appear that $\mathbf{Y}_t$ is co-integrated when it is not.
- The R^2 from the regression converges to unity as $\mathrm{T} \to \infty$. As a result the model will loot to fit well even though it is misspecified.
- All above imply that regression with $I(1)$ time-series variables makes sense only when they are indeed co-integrated.

5.3 An example of a spurious regression

In order to explain the concept of the spurious regression and ways that we can test this phenomenon, we give an example considering a pair of stocks which seems to be co-integrated. The next figure shows the co-movements of the logarithm of these two stocks, 'TITK' and 'TITP'.

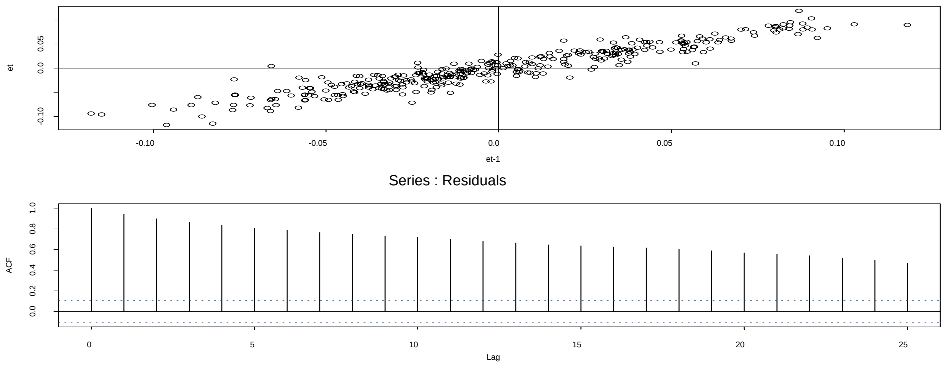
[43] Philips, P.C.B., 1986, *'Understanding Spurious Regression in Econometrics'*, Journal of Econometrics, 33, 311-340.



The correlation of these stocks is equal to 0.937, and both of them are $I(1)$. Applying the first step of the Granger-Engle procedure regressing 'TITK' on 'TITP' we get the following results:

```
Coefficients:
                 Value Std. Error t value Pr(>|t|)
 (Intercept)  0.9815  0.0528    18.5948  0.0000
        TITP  0.7655  0.0153    49.8849  0.0000

          R-Squared 0.8773
 Adjusted R-Squared 0.8770
 Durbin-Watson Stat 0.1059

                    Stat   P-Value
 Jarque-Bera    4.0538    0.1317
   Ljung-Box 4347.0802    0.0000
```

Looking at the large *t-statistic* and the *p-value* (close to zero) of the co-integrating coefficient, we have strong evidence to claim that is statistical significant and as a result different than zero. Also the regression R^2 is very high. The only suspicious statistic is the low *Durbin-Watson statistic*[44] suggesting strong residual autocorrelation. This statistics are representative of the spurious regression phenomenon with $I(1)$ that are not co-integrated. Also the very low p-value of the Ljung-Box statistic indicates that the residuals exhibit very high correlation. The next plots indicate the high autocorrelation of the residuals.

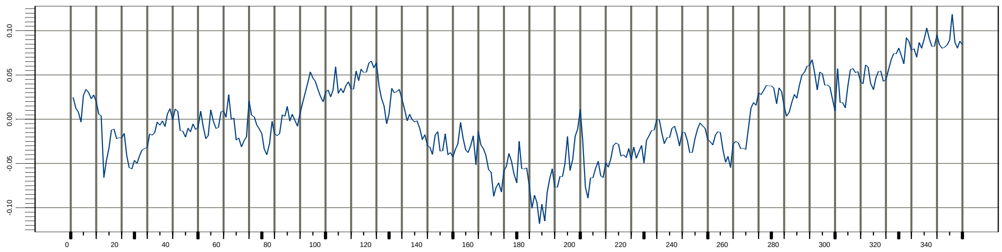


44

$$DW = \frac{\sum_{i=2}^{n}(e_t - e_{t-1})^2}{\sum_{i=1}^{n} e_t^2}$$

At this point let us see what results we will get if we implement the Granger-Engle procedure. Firs of all, applying the unit root tests discussed in *chapter 2.* we found that both time-series variables are $I(1)$. Then implementing the OLS regression and we get the co-integration coefficient and the co-integrating residuals. Now if the co-integrating residuals are $I(0)$ then we can infer that these two time-series variables are co-integrating. Below we represent the output of the unit root tests that we have applied to residuals, in order to specify if they are indeed $I(0)$ or not.

Figure 5.3 'Co-integrating residuals'



```
Test for Unit Root: Phillips-Perron Test

   Null Hypothesis: there is a unit root
      Type of Test: t test
    Test Statistic: -2.249
           P-value: 0.02385
```

```
   Test for Unit Root: Augmented DF Test

   Null Hypothesis: there is a unit root
      Type of Test: t test
    Test Statistic: -2.063
           P-value: 0.03766

   Coefficients:
          Value  Std. Error t value Pr(>|t|)
   lag1 -0.0366  0.0178    -2.0635  0.0398
   lag2 -0.1358  0.0544    -2.4950  0.0131
   lag3 -0.0863  0.0539    -1.5997  0.1106
```

From the above tables, observing the p-value (<5%) we can infer that the residuals are $I(0)$ and hence the two time series variable are co-integrated.

Let us now apply the *Johansen's Methodology* described in chapter 4. to the same pair. The LR statistics are representing below:

```
Tests for Cointegration Rank:
         Eigenvalue Trace Stat  95% CV  99% CV Max Stat  95% CV  99% CV
H(0)       0.0326    12.9025   15.4100 20.0400 11.3879  14.0700 18.6300
H(1)       0.0044     1.5146    3.7600  6.6500  1.5146   3.7600  6.6500
```

As we may observe, applying the Johansen's methodology we can infer that the two time-series variables are not co-integrated. Hence the question which arises, is which one of these two procedures should we trust. As we have mentioned above, the low value of the Durbin-Watson statistic give us some hints about the *spurious regression* phenomenon. In this case the statistical inference depends on the analyst. Modestly we strongly recommend one to take into account the Durbin-Watson statistic and to rely more on the Johansen's methodology since it can capture better the stock dynamics. Finally let us see what would happen if we had applied the statistical trading strategy for the next 99 trading days, considering as trading signals the 10% and 90% quantile of the empirical distribution of the residuals.

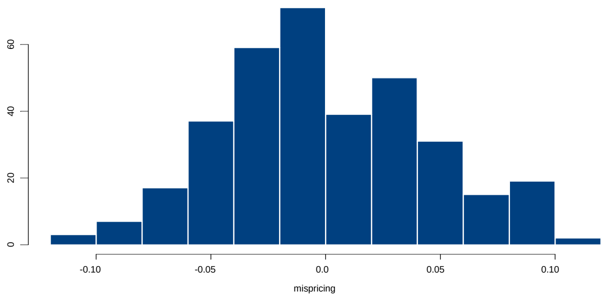
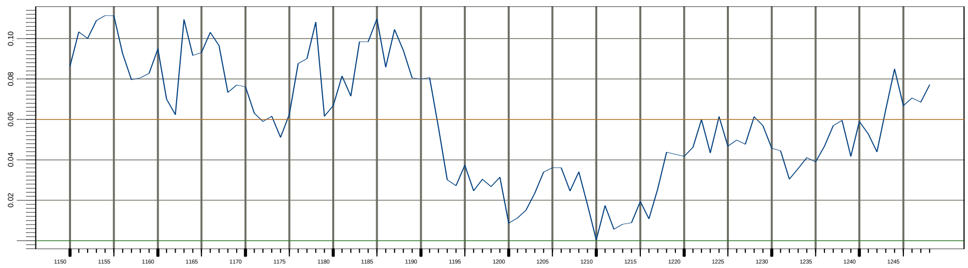


Figure 5.4 'Mispricing'



The figure above indicates that the process of the 'mispricing' in this case does not exhibit a mean-reverting property, since it does not oscillates around zero. Also the two stocks seem to not converge to their equilibrium. As a result, if we had applied the trading strategy, it would not be profitable. Hereafter, we will infer that two or more time-series variables are co-integrated if both procedures verify this.

6.1 Long Memory Time-Series

Up to now, we have discussed the concept of co-integration and the meaning of the long-run equilibrium. In order to in position to model the *mispricings* a basic requirement is the co-integrating residuals to be a stationary $I(0)$ processes. It is known that if a time-series variable x_t is $I(0)$ then its *ACF* declines at a geometric rate. As a result, $I(0)$ processes have *short memory* since observations far apart in time are essential independent. On the other hand, if x_t is $I(1)$ then its *ACF* declines at a linear rate and observations far apart in time are not independent. As a result, between $I(0)$ and $I(1)$ processes there exist the so-called *fractionally integrated* $I(d)$ process where $0 < d < 1$. In this case the *ACF* for the fractionally integrated processed declines at a hyperbolic rate. This implies that observations far apart in time may exhibit weak but non-zero correlation, where this phenomenon is often referred to as *long memory*.

Implementing the statistical analysis we have discussed to our data, we may observe that the co-integrating residuals, which can be supposed to be $I(0)$ processes, exhibit a long memory since their *ACF* declines rather slowly. Hence, sometimes the stationary and the unit root tests may fail to capture this phenomenon. For this reason, an analyst who apply these statistical arbitrage strategies should be alert, in order to avoid misleading results. It can be shown that x_t is stationary and ergodic for $0 < d < 0.5$, hence we are interesting only in processes which satisfy this restriction. Since we require the mispricing to exhibit a strong mean-reverting property, we are seeking for processes with *d* as close to zero. At this chapter, we will discuss briefly the concept of the FARIMA models, some tests for long memory and how can we estimate the value of *d*.

6.2 Statistical Tests for Long Memory

A stationary process x_t has *long memory* or *long range dependence*, if its autocorrelation function behaves like:

$$\rho(k) \to C_\rho k^{-\alpha} \text{ as } k \to \infty \qquad ,(6.2.1)$$

Where C_ρ is a positive constant, and α is a real number between 0 and 1. In this case the autocorrelation function of a long memory process decays slowly at a hyperbolic rate, so that the autocorrelations are not summable. I.e.:

$$\sum_{\kappa=-\infty}^{\infty} \rho(k) = \infty$$

and the spectral density is defined as:

$$f(\omega) = \frac{1}{2\pi} \sum_{\kappa=-\infty}^{\infty} \rho(k) e^{ik\omega}$$

For long memory process, its spectral density tends to infinity as zero frequency. Now if in equation (6.2.1) we replace α with *Hurst coefficient*[45]:

$$H = 1 - \alpha / 2 \in (0.5,\ 1)$$

we get a measure of the long memory in x_t. More specific, the larger *H* is, the longer the memory the stationary process has.

A long memory process x_t can also be modeled parametrically by extending an integrated process to a fractionally integrated process as follows:

$$(1-L)^d (x_t - \mu) = u_t \tag{6.2.2}$$

where μ is the mean value of x_t and u_t is a stationary short memory disturbance. The fractional difference filter is defined as follows, for any real $d > -1$:

$$(1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k L^k$$

6.2.1 *R/S Statistic*

The most widely-used test for long memory or long range dependence is the *R/S statistic* which is the range of partial sums of deviations of a time series from its mean, rescaled by its standard deviation and is defined as follows:

$$Q_T = \frac{1}{S_T}\left[\max_{1\le k\le T} \sum_{j=1}^{k} (x_j - \bar{x}) - \min_{1\le k\le T} \sum_{j=1}^{k} (x_j - \bar{x})\right]$$

where

$$S_T = \sqrt{\frac{\sum_{i=1}^{T} (x_i - \bar{x})^2}{T}}$$

If x_t are *i.i.d.* normal random variables, then

$$\frac{1}{\sqrt{T}} Q_T \Rightarrow V$$

This means that it converges weakly to *V,* where *V* is the range of a *Brownian Bridge* on the unit interval. Selected quantiles of *V* are given in Lo (1991).

---

[45] Hurst, H.E. (1951), *'Long Term Storage Capacity of Reservoirs'* Transactions of the American Society of Civil Engineers, 116, 770-799

6.2.2 GPH Test

Another semi-nonparametric approach to testing for long memory is the well-known *GPH Test*[46] where the spectral density of the fractionally integrated process x_t is given by:

$$f(\omega) = \left[4\sin^2(\frac{\omega}{2})\right]^{-d} f_u(\omega)$$

where ω is the Fourier frequency, and $f_u(\omega)$ is the spectral density corresponding to the stationary process u_t described in equation (6.2.2). The fractional difference parameter d can be estimated by the following regression:

$$\log\left(f_u(\omega)\right) = \beta - d\log\left[4\sin^2(\frac{\omega_j}{2})\right] + e_j$$

Under the null-hypothesis of no long memory, (i.e. *d=0)* the *t-statistic* is given by:

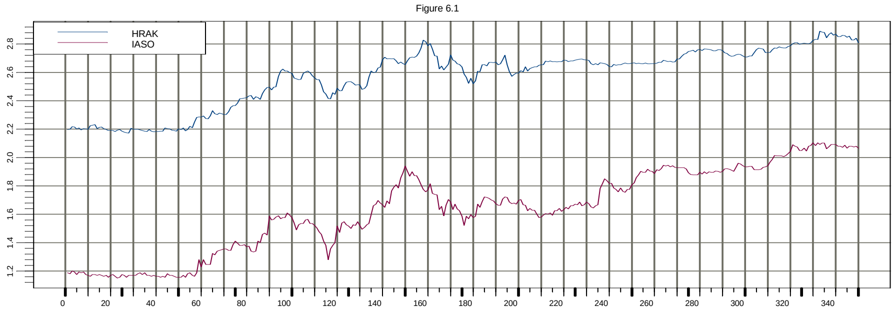
$$t_{d=0} = \hat{d}\left(\frac{\pi^2}{6\sum_{j=1}^{n_f}(U_j - \bar{U})^2}\right)^{-1/2}$$

where $U_j = \log\left[4\sin^2(\frac{\omega_j}{2})\right]$. The *t-statistic* has a limiting standard normal distribution.

6.3 Example of Long Memory Time-Series

Long memory tests, may give us some hints about spurious regression. To be more specific we outline an example below, where we consider again a pair of stocks applying the methodology which have been discussed up to now. The figure below represents the co-movements of the log price of this pair. The correlation between these two stocks is 0.954 obtained from 21/9/2005 to 12/2/2007.

Figure 6.1



[46] Geweke, J., Porter-Hudak, S., (1983), '*The Estimation of Security Price Volatility from Historical Data',* Journal of Time Series Analysis, 4, 221-237.

6.3.1 Granger-Engle Procedure

Regressing the time-series variable '*HRAK*' on '*IASO*' we obtain the co-integrating coefficient. The out-put is representing below.

```
Coefficients:
               Value Std. Error t value Pr(>|t|)
 (Intercept)  1.4583  0.0191    76.3976  0.0000
        IASO  0.6797  0.0115    59.1413  0.0000


          R-Squared 0.9095
 Adjusted R-Squared 0.9092
 Durbin-Watson Stat 0.1204

                  Stat   P-Value
 Jarque-Bera   21.8879    0.0000
   Ljung-Box 2579.5084    0.0000
```

Notice that the *Durbin-Watson* and the *Ljung-BOX* statistics warn us for spurious regression and long memory phenomenon. Continuing the Granger-Engle Procedure, we apply the unit root tests in order to specify if the co-integrating residuals are $I(1)$ or $I(0)$. The out-put of the results is representing below.

```
1)Test for Unit Root: Augmented DF Test

  Null Hypothesis: there is a unit root
     Type of Test: t-test
   Test Statistic: -3.311
          P-value: 9.74e-4

2)Test for Unit Root: Phillips-Perron Test

Null Hypothesis: there is a unit root
   Type of Test: t-test
 Test Statistic: -3.537
        P-value: 4.35e-4
```

As we may observe both tests reject the null hypothesis for unit root at 1% significant level. Notice if we apply the *stationarity test* (*KPSS*) described in *Chapter 2.4* we get different results, i.e. that the co-integrating residuals can not be considered as a $I(0)$ process. The results are illustrating below.

```
Test for Stationarity: KPSS Test

 Null Hypothesis: stationary around a constant

 Test Statistics: 0.7756**

** : significant at 1% level
```

6.3.2 Johansen Methodology

Implementing the Johansen Methodology, we have strong evidence to conclude that these two time-series variable share a common trend. The LR statistics are illustrated below.

```
Trend Specification:
H1(r): Unrestricted constant

Tests for Cointegration Rank:
         Eigenvalue Trace Stat  95% CV  99% CV Max Stat  95% CV  99% CV
H(0)+ *   0.0407    15.5114    15.4100 20.0400 14.3914  14.0700 18.6300
H(1)      0.0032     1.1199     3.7600  6.6500  1.1199   3.7600  6.6500

Cointegrating Vectors:
     coint.1
HRAK  1.0000
IASO -0.6781
```

6.3.3 Long-Memory Test

We have mentioned several times that the co-integrating residuals must be a stationary process. Since sometimes the unit root tests cannot capture perfectly the dynamics of the mispricings, we apply the two long-memory tests described in *Section 6.2*. In order to conclude that the co-integrating residuals are a stationary process, we require the fractional difference parameter *d* to be less than 0.5. The results of the long-memory tests are illustrated below:

```
1)Test for Long Memory:R/S Test

Null Hypothesis: no long-term dependence

Test Statistics:3.0911**

2)Test for Long Memory: GPH Test

Null Hypothesis: d = 0

Test Statistics:

   d 0.6031
stat 3.1045**

** : significant at 1% level
```

In this case the co-integrating residuals cannot be considered as a stationary process since the fractional difference parameter *d=0.6031* is relative high. As a result the process exhibits significant long-term dependence, and strictly speaking this pair could not be considered appropriate for the statistical arbitrage strategy, even if the Granger-Engle Procedure and the Johansen Methodology suggest us the opposite. However in this empirical study we excluded only the portfolios which have a fractional difference parameter *d* larger or equal to 0.65

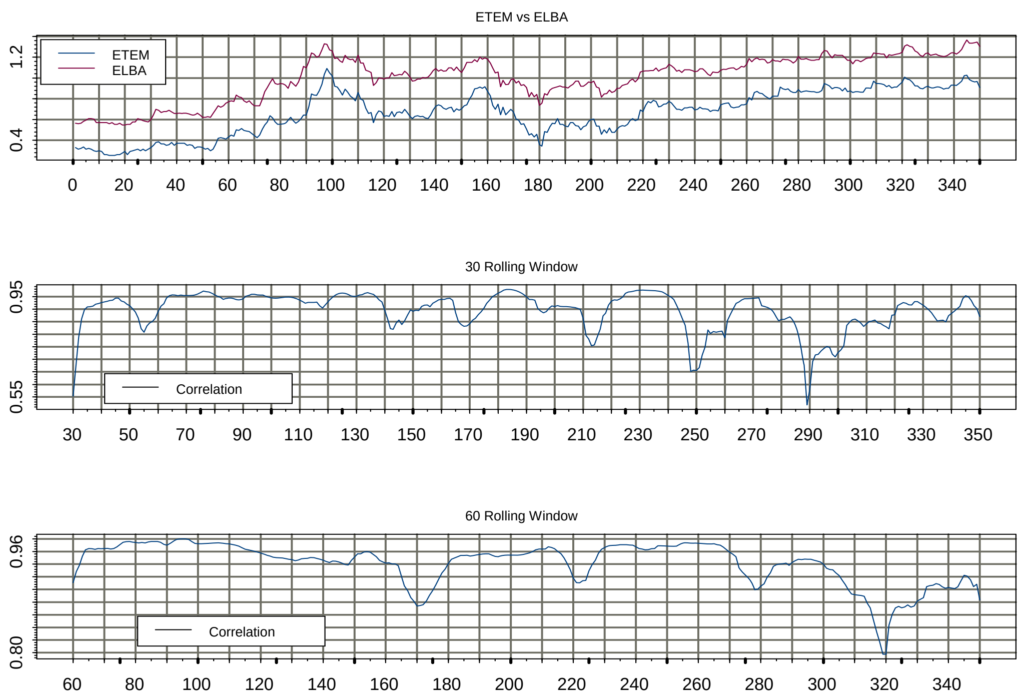
7.1 Rolling Analysis

In the discussions so far, we have mentioned some things that we have to attend in order to conclude that two or more stocks are suitable for statistical arbitrage strategy. Notice that when analyzing financial time-series data using statistical models, a basic requirement is that the estimated parameters of the model are constant over time. Since we are trying to model the mispricing dynamics, we have to be sure that the parameters can be considered time-invariant, otherwise may lead us to misleading results. Since the equilibrium relationship of two or more assets often changes considerably, it may not reasonable to assume that a model's parameters are constant. A simple technique to asses the constancy of the estimated parameters is to compute parameters estimates over a rolling window of a fixed size through the sample. If the estimates over the rolling window do not change significantly, then we can assume that are time-invariant.

7.2 Rolling Correlation

The first thing that an analyst is looking for in order to find potential co-integrated assets is their correlation. However this correlation estimate may not be constant over time. One way to test this hypothesis is to estimate the correlation over a rolling window. For convenience we give an example below considering a 30 and 60 days rolling window.

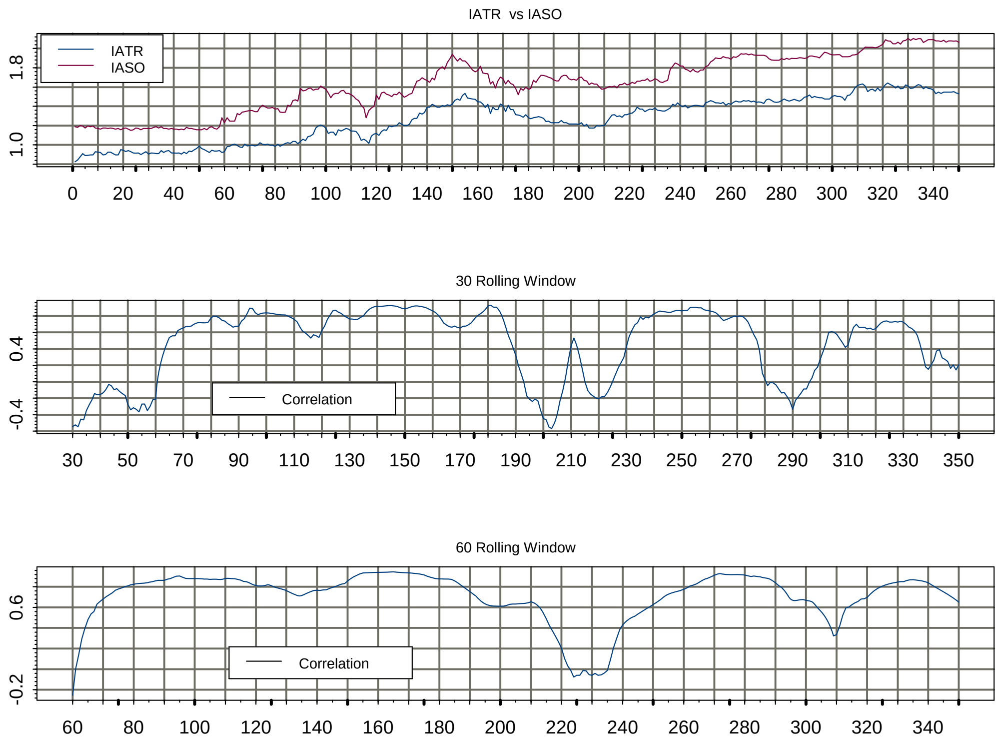
Example 7.2.1



The correlation of these two stocks, considering the full sample is equal to *0.97*. As we can observe from the two plots above, the correlation of these two stocks is around 0.96, and can be considered constant.

Below we represent another example where the correlation seems to not be constant.

Example 7.2.2



Notice that this pair has also a correlation equal to *0.97* but it seems to not be constant over time. Therefore, even if we had concluded that these two stocks are co-integrated, it would be risky to enter into a trading strategy since includes the risk that the estimated co-integration coefficient will not be the true one.

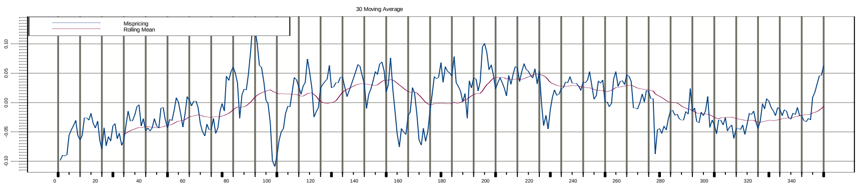
7.3 Rolling Mean

As we have mentioned earlier, a basic requirement for co-integrating residuals, is to be stationary around mean. This can be tested applying the KPPS test (*See Chapter 2*). This test could also be done by plotting a rolling mean given by:

$$\hat{\mu}_t(n) = \frac{1}{n}\sum_{i=0}^{n-1} y_{t-1}$$

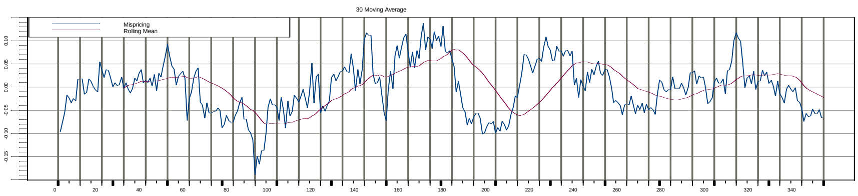
where n denotes the rolling window. Again for convenience of the reader we represent an example, considering again these two pair of stocks as above. The next figure represents the co-integrating residuals of the pair '*ELBA-ETEM*' superimposed the rolling mean, considering a rolling window of 30 observations.

Mispricing of ELBA-ETEM



The next figure represents the co-integrating residuals of the pair '*IATR-IASO*' superimposed a rolling mean, with width equal to 30

Mispricing of IATR-IASO

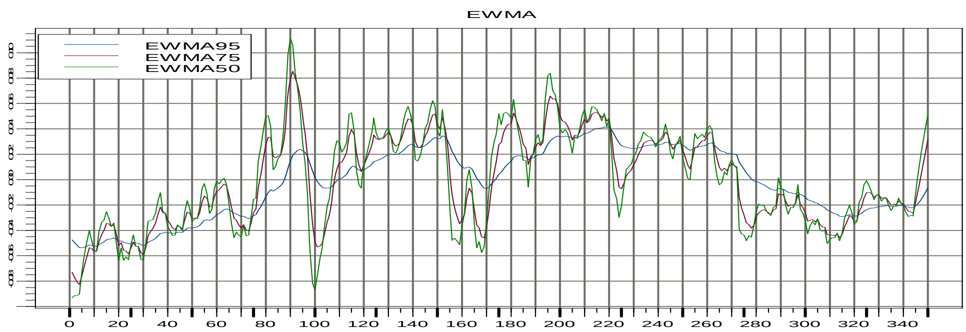


Comparing these two plots, we may argue that the first pair looks more stationary, since the moving mean does not diverge significant from 0, and as a result can be considered as constant over time. The main problem with the rolling analysis, is that the tests are rather objective, however can be a useful tool for an analyst.

Notice that the simple moving average is based on equally weighted moving averages of an observed time-series y_t. The main drawback of the rolling mean is that is very sensitive to outlier values. A common weighting model that puts more weight on the most recent observation is called an *exponentially weighted moving average* (EWMA), and has the advantage that mitigate the effects of extreme observations. An *n-period* EWMA is given by:

$$\hat{\mu}_t(n) = \sum_{i=0}^{n-1} \left(\frac{\lambda^{i-1}}{\sum_{i=0}^{n-1} \lambda^{i-1}} \cdot y_{t-1} \right)$$

where $0 < \lambda < 1$ is the decay parameter. Notice that the larger the λ the more weight is put on the previous period's estimate relative to the current period's observatrion[47]. The next figure represents a EWMA model for the co-integrating residuals of the pair '*ELBA-ETEM*' for three different values of *λ*.



[47] Zivot, E., Wang, J., 2006, '*Modeling Financial Time Series with S-PLUS*' pages 324

7.4 Rolling Regression

For the linear regression model, rolling analysis may be used to asses the stability of the co-integrating vector. Considering a rolling window *n<T,* the rolling linear regression model is as follows:

$$\mathbf{y}_t(n) = \mathbf{X}_t(n)\boldsymbol{\beta}_t(n) + \boldsymbol{\varepsilon}_t(n), \quad t = n, \ldots, T$$

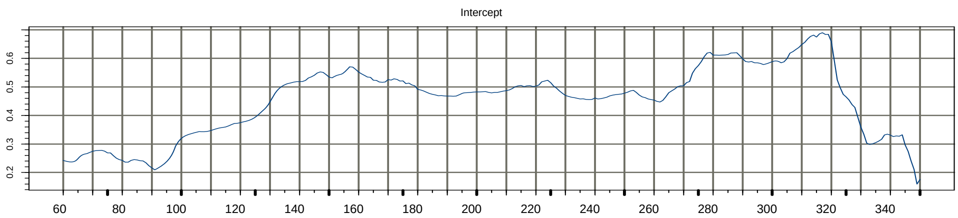
where $\mathbf{y}_t(n)$ is an $(n \times 1)$ vector of the response variable, $\mathbf{X}_t(n)$ is an $(n \times k)$ matrix of explanatory variables, and $\boldsymbol{\beta}_t(n)$ is an $(k \times 1)$ vector of regression parameters. The rolling least square coefficients are estimated by[48]:

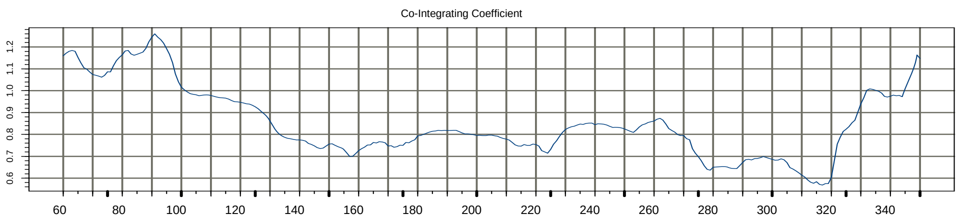
$$\hat{\boldsymbol{\beta}}_t(n) = \left[\mathbf{X}_t(n)'\mathbf{X}_t(n)\right]^{-1} \mathbf{X}_t(n)'\mathbf{y}_t(n)$$

For convenience we give an example of a rolling regression, considering a width equal to 60 observations, for the pair '*ELBA-ETEM*'. Notice that using the OLS for the full sample the co-integrating coefficient and the intercept are **0.9956** and **0.3359** respectively. The output of the rolling regression is given below, where are illustrated the average mean of the 291 rolling regression estimates:

```
Coefficients:
            (Intercept)   ETEM
       mean 0.4535      0.8506
  std. dev. 0.1214      0.1610
```

As we can observe, the parameters are quite different, but the plots of the coefficients may give us more details about the stability of the parameters.



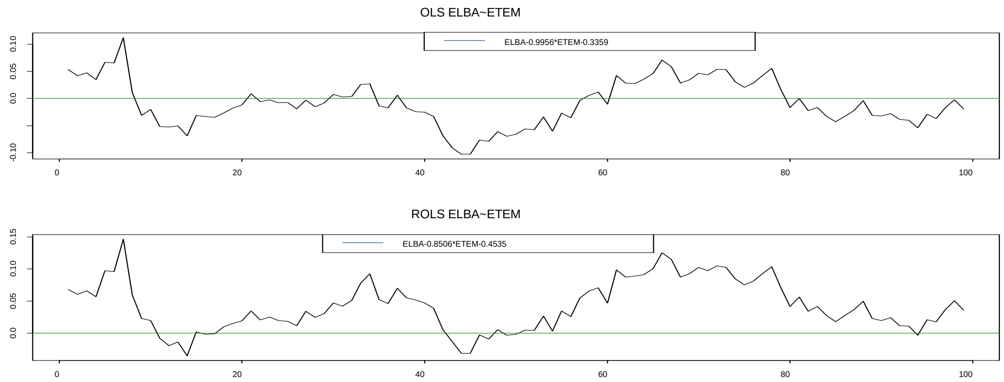


As it seems from the plots and the standard deviation, the coefficients are quite variable over time. Notice that we do not expect to observe a perfectly time-invariant estimate, and this is one reason which explains why say pairs trading is not a riskless strategy.

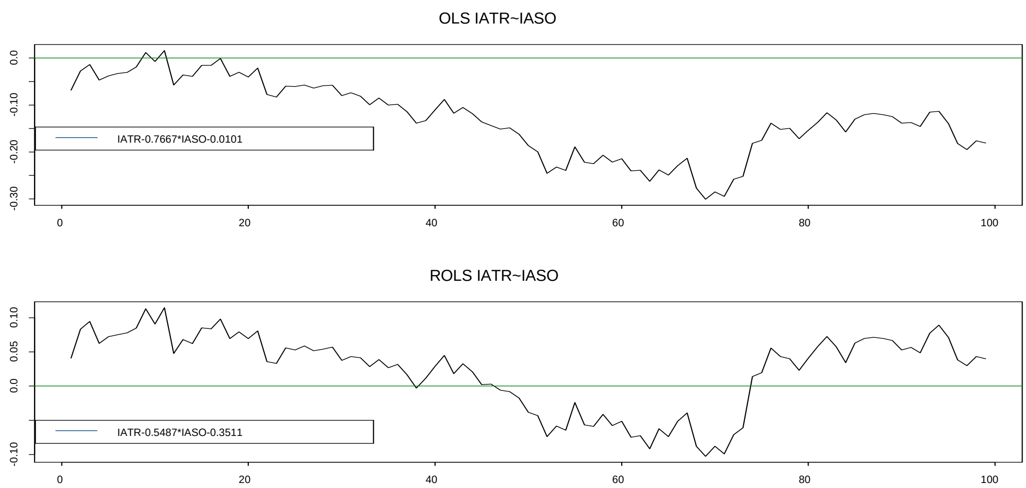
[48] Zivot, E., Wang, J., 2006, '*Modeling Financial Time Series with S-PLUS*' pages 343

7.5 Rolling Co-Integrating Vector

The aim of this project is to discuss and examine some sensible ways in order to specify the mispricing dynamics. Up to now, we have discussed three different ways of estimating the parameters of the mispricing (OLS, DOLS, Johansen's Methodology), where the obtained values are not very different. As we have discussed in the previous section, we can obtain significant different estimates by rolling OLS. As a result, we have to examine which one of these four methods gives us better results. Here we give an example of a pairs trading strategy, considering the two pairs mentioned above, comparing the OLS and the rolling OLS methods. The strategy is applied taking into account the next 99 closing prices.
From the plot below, we may infer that the OLS give us better description regarding the long run-equilibrium of these two stocks since the mispricing is oscillates around zero. On the other hand, the ROLS seems that cannot capture adequately the mispricing dynamics.



Applying the same methodology to the second pair, the results are surprising. The OLS method failed to capture the mispricing dynamics whereas the ROLS method looks much better. We just inform the reader that the second pair (IATR~IASO) cannot be considered co-integrated, since has not satisfied the desired requirements discussed in previous chapters.



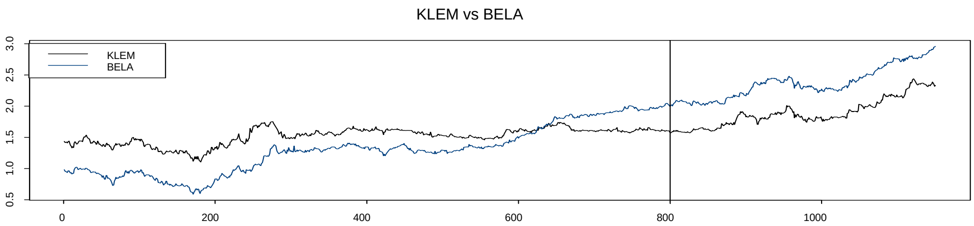
8.1 Trading Design

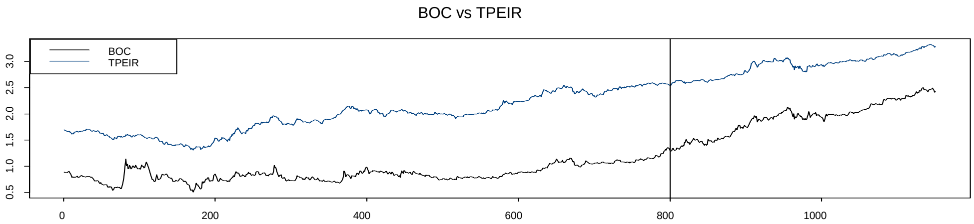
As we have discussed so far, the key requirement for entering statistical arbitrage strategies, is the existence of an equilibrium relationship between the log price series of the stocks, where this relationship is characterized by the co-integrating vector, and could be obtained by OLS, DOLS, ROLS or applying the Johansen's Methodology. Once the co-integration vector is estimated, it can be used to construct the dynamics of the mispricing. The strategy based on the mean-reverting property of the mispricing. When we make the determination that the mispricing has diverged significantly from the equilibrium value, we enter into an appropriate position, betting that the divergence will correct itself, reverting back to equilibrium. Therefore it is important to define what would qualify as a sufficient divergence of the mispricing from equilibrium for us to consider entering into a trade, setting simple rules, which we refer to as *trading signals.*

We have to keep in mind that the statistical arbitrage strategy is not a riskless strategy, and the investor who enters into a trade should eliminate the risk. In this section, we recommend a methodology for trading design, using the tools we have discussed up to now.

8.2 Sample Size

It is very important to specify how many observations we should consider in order to examine the existence of the equilibrium relationship. This empirical study has shown that a sample size of more than 1000 observations cannot capture adequately the mispricing dynamics, since the estimated co-integrating vector affected by the very past observations. In this study we get about 1150 daily closing prices of around 90 stocks and we started the analysis taking into account *only* the 350 most recent observations. Then we examine the 1150 observations again to see if this equilibrium relationship exists also in the past history data. To be more specific, we give an example below.

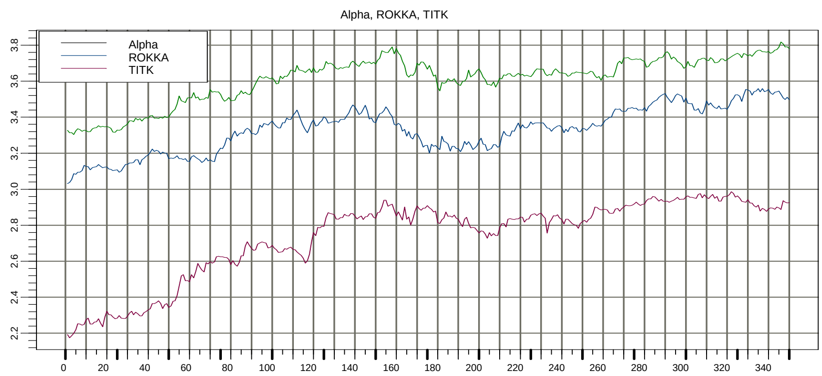




Clearly the first pair appears risky since the long-run equilibrium does not hold in the past history data, whereas the second pair appears to exhibit a long-run equilibrium.

8.3 Size of the portfolio

Up to now, we have discussed the methodology for testing stocks which are co-integrated, but for simplicity we have given examples considering only the two stocks. Analyzing extensively the data, we may infer that is very difficult to find more than seven stocks which are truly co-integrated. However adding more variables, say four stocks, the co-integrating residuals acquire a better behavior[49] and as a result we can capture better the mispricing dynamics. Of course our aim is to capture the mispricng dynamics in order to take an advantage of them. The main problem with the larger basket of assets is the transaction cost. Let us give an example below in order to explain better what we mean by the argument above. We consider the log price of the stocks 'Alpha', 'ROKKA', and 'TITK' and we apply the *Johansen's Methodology*. The corresponding plot and output of the S-plus is representing below.

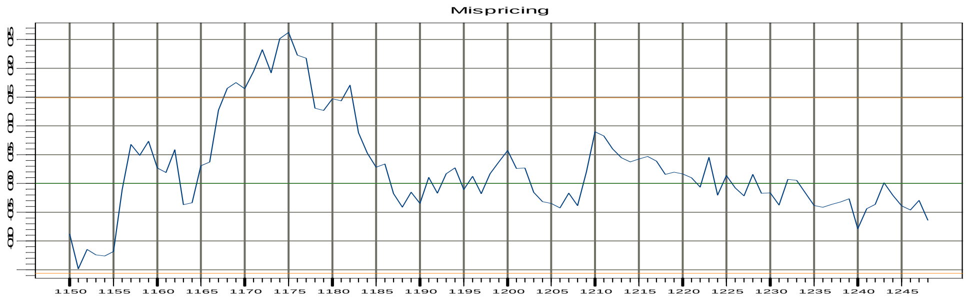



```
Tests for Cointegration Rank:
          Eigenvalue Trace Stat  95% CV  99% CV Max Stat  95% CV  99% CV
H(0)++**   0.0730     41.0650    29.6800 35.6500 26.3849  20.9700 25.5200
H(1)       0.0256     14.6801    15.4100 20.0400  9.0236  14.0700 18.6300
H(2)+ *    0.0161      5.6565     3.7600  6.6500  5.6565   3.7600  6.6500

Cointegrating Vectors:

Alpha   1.0000
ROKKA   0.9196
 TITK  -2.3850
```

From the *Trace* and *Max statistics* we can infer that there is at least one co-integrating vector. Notice that the portfolio consists of around four stocks, which means that for every trading signal we have to buy and sell four stocks. Clearly the transaction cost affects by far the profit. Let us see what would have happened if we had applied a trading strategy. Notice that every trade must be profitable, i.e. higher than the transaction cost. For that reason we set wider bands as trading signals, i.e. the 3% and

[49] I.e. they have low autocorrelation, and look like a white noise process. As a result they exhibit stronger mean-reverting property.

97% quantile of the co-integrating residuals. The strategy is applied to the next 99 trading days. The corresponding plot of the mispricing is representing below:



Notice that we fitted the model taking 350 observations (from 21/09/05-12/02/07) and we test it for the next 99 trading days (13/02/07-06/07/07). The results are as follows:
12/03/2007: Buy 2.3859 stocks of ‘TITK, short-sell 0.9169 stocks of ‘ROKKA’ and 1 stock of ‘Alpha’
10/04/2007: Close the position.
The net profit of the strategy without including the transaction cost is equal to 6.35€. The transaction cost is equal to 2.85€ which implies that the true net profit is equal to **3.5€**. We consider a transaction cost equal to 1%. The above example indicates why larger basket of assets are not profitable. This is a reason why we apply the strategy only to baskets up to four assets. Notice that in case we had assumed that there is no transaction cost, the larger basket of assets may have been more profitable and less risky.

8.4 Band Design

Let us discuss the design of trading signals. Due to the transaction cost, we should avoid the frequent trades. This implies that the trading signals should be large enough in order to achieve a substantial profit per every trade. It is very difficult determine the optimum trading signals, since we have to take into account a lot of parameters. In order to define sensible trading signals, we applied some non-parametric tests, testing the profit for different values of the trading signals. The bands which maximized the profit are as follows:

- **Portfolio of two stocks:** The trading signals could be the 8% and 92% quantiles of the co-integrating residuals.
- **Portfolio of tree stocks:** The trading signals could be the 3% and 97% quantiles of the co-integrating residuals.
- **Portfolio of four stocks:** The trading signals could be the 1% and 99% percent of the co-integrating residuals.

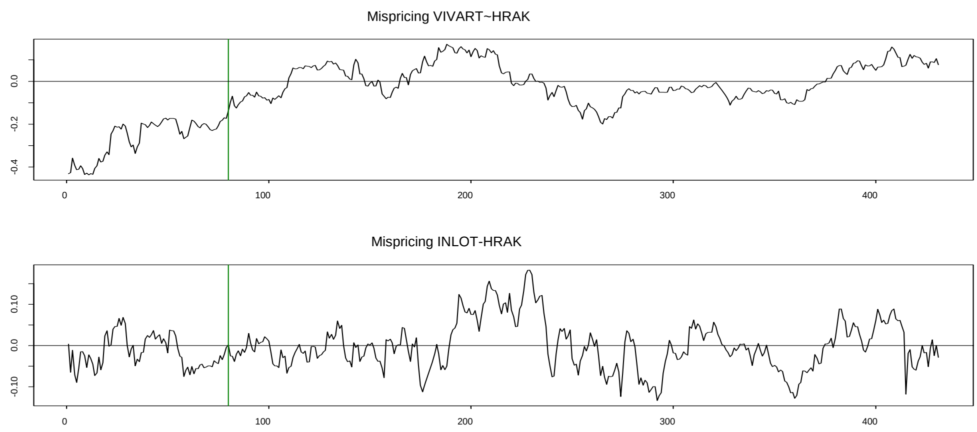
Notice that as we add more variables the trading signals are getting wider, due to the next reasons. The first one is that we want to avoid frequent trades as we constructing larger basket of assets, and the second one is because by adding more variable, we explain more of the variability of the mispricing dynamics, and as result we should consider wider bands as a sufficient divergence.

8.5 Construction of the portfolio

We have mentioned before that we should construct portfolios of up to four stocks. The number of stocks for each portfolio is give by the co-integrating vector. Since we have discussed four different ways of estimating the co-integrating vector, the question is which one of these methods we should apply. Implementing these four methods to our data, we may conclude that by applying the OLS or DOLS we are getting almost the same results. However this study has shown that by applying the Johansen's Methodology the portfolios becoming more profitable. As a result we may infer that this is the optimum way of constructing the appropriate portfolios. Quite different results we get by applying rolling OLS (ROLS), where on average we get less profit. Hence we just recommend you to avoid using ROLS. Another sensible way of estimating the co-integrating vector could be to use a *rolling Johansen's Methodology*. However during this study we have not tested how does it work in practice, and as result we are unable to evaluate it.

8.6 Enter Into Trading Strategy

Since statistical arbitrage strategies involve a risk, and especially the so-called pairs trading, we have to be very careful before we enter into the trading strategy. Up to know we have discussed all the requirements[50] that the portfolio must satisfy in order to be considered appropriate for statistical arbitrage strategy. Another final test which can be done before we enter into a strategy is to test what would have done if we had applied the strategy to the previous observations. In order to clarify the above argument we give an example below. We fitted the model taking 350 closing prices from 21/09/05-12/02/07 and we tested it for the 80 previous observations (30/05/05-20/09/05). For comparison we represent two pairs.



As we may observe from the plot, the first portfolio appears to be risky since during 80 days the mispricing dynamics never oscillates around zero. On contrary the second portfolio appears to be appropriate for a trading strategy. The above test, let us call it '*back-shifting*', is an informal way of testing the constancy of the parameters of the model and is suggested by the writer of this thesis.

[50] See Unit root tests, KPPS, ECM (for two dimensions), VECM, R/S, GPH and so on.

Discussion

The main objective of this study was to discuss a computational methodology for modeling the mispricing dynamics and to verify if the statistical arbitrage works in Greek stock market (*ASE*). Implementing the methodology taking daily closing stock prices from the *Athens Stock Exchange* we may infer that Johansen's Methodology is the most appropriate method of modeling the mispricing dynamics since the portfolios constructing by this approach found to be the most profitable. The numerical results of this survey are quite interesting and there is a brief presentation of them in '*Appendix Section 5.1*'. We managed to construct 20 portfolios of two assets, 5 portfolios of three assets and 1 portfolio of four and five assets respectively. The criterion of selecting each portfolio was to satisfy the requirements which have been discussed explicitly during this study. Applying the strategy virtually to the next 99 trading days we found that the 24 out of 28 portfolios were profitable, where the portfolios of two assets or three assets found to be the most profitable. Maybe this is one reason why pairs trading is so popular and is widely used by financial institutions. The basic motivation of this study was to examine if we could apply a trading strategy to the ASE enabling us to make money consistently. The idea was to examine if there exist assets which have long-run equilibrium relationships and if so to implement a statistical arbitrage strategy betting that the divergence of the equilibrium will correct itself. From the numerical results it appears that we won the bet since in time period of 99 trading days the 89.3% of the portfolios do not diverge permanently from their long-run equilibrium[51]. At this point it would be proper to evaluate briefly the performance of the suggested strategy mentioning some issues regarding this survey.

Market Neutral Strategy: The proposed strategy seems to be a market neutral strategy, which means that the return of the investment does not affected by the market. This is the great advantage of this strategy since we can make a profit even during a bearish market.

Short Selling and Divisible Assets: Since our scope is to propose a realistic and feasible trading strategy we have to take into account all the parameters. In this survey we have made two basic assumptions which are that the sort-selling is allowed in '*ASE*' and that all assets are divisible. Thus we have to mention that there are a lot of restrictions in Greek market regarding the short-selling which implies that the proposed strategy could not be applied in '*ASE*'. Regarding the issue of the divisible assets, it could be solved by buying a lot of stocks. Clearly this is a matter for small investors but not at all for the big financial institutions.

Optimum Trading Signals: The aim of this survey, regarding the issue of the trading signals, was to specify a sensible and mathematically consistent way of setting these bands. We tried a lot of alternative ways, and among these we chose to use the quantiles of the co-integrating residuals[52]. However, applying the strategy we observed that in some cases, we should have chosen wider bands and in some other

[51] Notice that 25 out of 28 portfolios (i.e. 89.3%) do not diverge permanently form their long-run equilibrium. However in one portfolio (TPEIR-BOC) we did not receive any trading signal since it does not diverge sufficiently from its equilibrium value within these 99 days. Only three portfolios (ETEM-Alpha , INLOT-HRAK, IASO-HRAK) did not revert back to its equilibrium value.

[52] See chapter 8

cases narrower. Since it seems rather complicated to find the optimum trading signals, we just suggest that an investor could set his/her trading signals based on what he/she considers as a substantial profit. Clearly the bands must be set in such a way, that the transaction cost will be less than the profit of the trade.

Time Horizon of the Strategy: Regarding the trading strategy an important issue is for how long should an investor follow the strategy without re-constructing his/her portfolio. In this study we considered a time horizon of 99 trading days, and we may infer that on average the behavior of the mispricing was satisfactory, since in lot of cases the portfolios revert back to their equilibrium in last 10-15 trading days[53]. Hence a time period of the strategy up to 100 days may be considered appropriate. If an investor is eager to follow the strategy for a further time period, then we suggest him/her to fit a new model, taking again the 350 most recent observations as discussed in *chapter 8*.

Concerns about Larger Portfolios: By constructing larger portfolios we reduce the risk of the permanent divergence of the long-run equilibrium value but also we reduce the possibility of trading strategy. To be more specific, let us consider a large portfolio of 10 different stocks. Clearly we enter into a trade when the portfolio has diverged sufficiently from the equilibrium value. Recall that for every signal we sell or buy all the assets simultaneously. Obviously the trading signals must be set in such a way where at least the transaction cost is less than the profit per trade. But with such large portfolios rarely can we find such a divergence where the transaction cost is less than the net profit per trade. As a result we may conclude that portfolios which consist of more than four assets are not suitable for such a trading strategy.

Evaluation of the Portfolios: In '*Appendix section 5.1*' we represent the net profit for each of the 28 portfolios. At each trading signal we were selling or buying a unit of the portfolio. Hence the net profit should be expressed as a percentage in order to be able to compare the performance of these portfolios. The problem was that many portfolios did not require any initial investment (i.e. go-short the portfolio), which implies that is difficult to express the net profit as a percentage. However, one may observe that some pairs were quite profitable (13.60€), and some others had a very low profit (0.49€). Thus we would like to mention that the portfolios which found to be more profitable were also more expensive, in other words in order to construct such portfolios we required a larger amount of money in absolute terms. We say in absolute terms because of the negative investment (i.e. short-sell the portfolio). Thus if an investor had invested the same amount of money in each portfolio, he or she would have found that the portfolios of two assets can make significant profit. However the portfolios of three assets are found to be quite profitable, whereas by adding more assets we reduced the profit margin. Hence we recommend investing in portfolios of up to three assets.

Diversification: From the numerical results one may infer that pairs trading is more risky than trading larger basket of assets. Indeed 3 out of 20 pairs suffer a loss. Since all rational investor are risk-averse it is important to propose ways of how one can diversify the risk that involves the pairs trading. Clearly one can invest only in portfolios consisting of four assets. An alternative way is to construct a large portfolio

[53] See the figures in '*Appendix section 5.1*'

consisting of many independent pairs. In this way he or she is position to reduce the risk significantly.

Issues regarding the Survey: Since we are interesting to accomplish a valid survey we are obliged to mention all the issues which may affect the results of this study. Thus the main concern is that we applied the strategy virtually having only the daily closing prices of the stocks. This implies that there is a high probability that we have missed some trading signals and consequently the numerical results could be quite different. Finally we have to mention that we fitted the models taking also the closing prices. Unfortunately we did not have the opportunity to fit and test models considering more frequent observations, such as per hour or per minute.

Risk of the Strategy: Clearly statistical arbitrage strategies are not riskless strategies. Regarding the proposed strategy an investor should be alert that the fundamentals of a company may be changed, an as a result he/she avoid including such stocks to his/her portfolio. The most typical cases where the price of the stocks change significantly and permanently is when we have an announcement of a take over bid, an increase in the dividend and an announcement of a split. Moreover some stocks never return back to their equilibrium value for unobservable reasons. As we may observe in this study we found that three pairs never converge back to their equilibrium relationship within these 99 days, and we are not able to explain why this occurred. However by investing in different portfolios we are in position to eliminate the risk of the strategy.

Feasible Strategy: One may wonder if the proposed strategy is feasible for every investor. If we look carefully to the numerical results we will realize that in order to accomplish a significant profit we have to invest or borrow a lot of money. Also taking into account that the assets are not divisible it seems to be very difficult for a small investor to enter into this strategy.

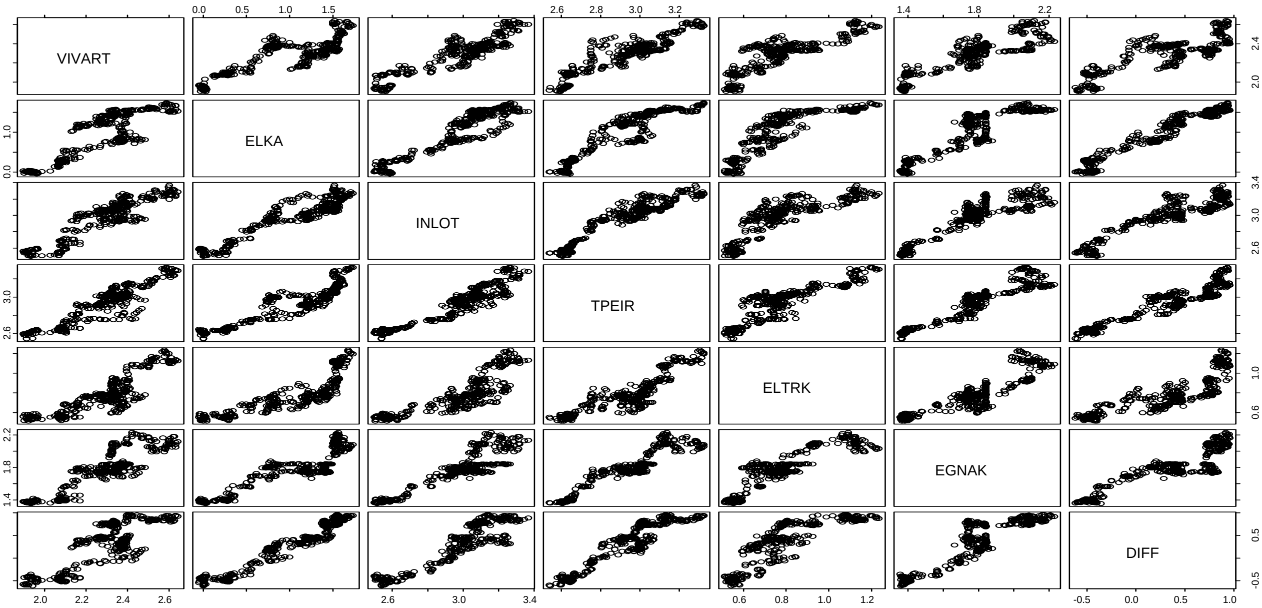
Make Money Consistently: This survey have shown that we have evidence to support that the ‘ASE’ is suitable market for statistical arbitrage strategies. It would be fulsome to argue that we know the ‘secret’ of making money consistently, but based on this empirical study we may claim that the strategy seems to work well in ‘ASE’.

APPENDIX

1.1 Graphical Illustration of co-movements

The plot below is the so-called 'Matrix Plot' or 'Pairs Plot'. This plot enables as to observe quickly which stocks seem to be strongly correlated. The prices of the stocks are transformed to the log-scale.

Figure 1

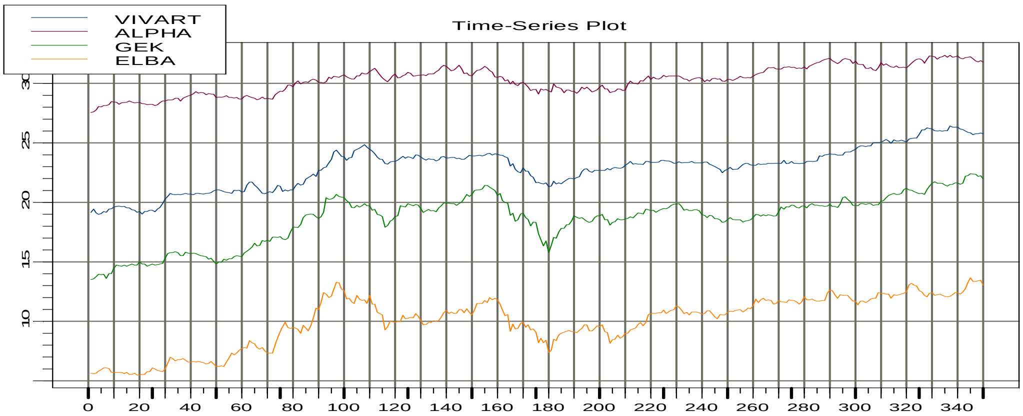


The next table represents the correlation Pearson measure between pair of stocks. The data are transformed to a logarithmic scale and we represent a small sample of them. The next figure shows the co-movements of four stocks which are strongly correlated.

Table 1

	Alpha	BIOX	BRAIN	GEK	VIVART	DIFF	EGNAK	ELTRK	ELBA	ELKA	ETEM	HRAK
Alpha	1.00	0.90	0.88	0.91	0.93	0.83	0.92	0.86	0.92	0.82	0.92	0.86
BIOX	0.90	1.00	0.81	0.93	0.89	0.74	0.84	0.77	0.97	0.77	0.96	0.84
BRAIN	0.88	0.81	1.00	0.82	0.84	0.87	0.94	0.95	0.86	0.87	0.87	0.88
GEK	0.91	0.93	0.82	1.00	0.95	0.82	0.86	0.82	0.94	0.85	0.92	0.91
VIVART	0.93	0.89	0.84	0.95	1.00	0.82	0.88	0.85	0.92	0.83	0.90	0.88
DIFF	0.83	0.74	0.87	0.82	0.82	1.00	0.91	0.84	0.83	0.97	0.80	0.94
EGNAK	0.92	0.84	0.94	0.86	0.88	0.91	1.00	0.89	0.88	0.90	0.87	0.91
ELTRK	0.86	0.77	0.95	0.82	0.85	0.84	0.89	1.00	0.84	0.84	0.86	0.85
ELBA	0.92	0.97	0.86	0.94	0.92	0.83	0.88	0.84	1.00	0.85	0.98	0.89
ELKA	0.82	0.77	0.87	0.85	0.83	0.97	0.90	0.84	0.85	1.00	0.82	0.97
ETEM	0.92	0.96	0.87	0.92	0.90	0.80	0.87	0.86	0.98	0.82	1.00	0.86
HRAK	0.86	0.84	0.88	0.91	0.88	0.94	0.91	0.85	0.89	0.97	0.86	1.00

Figure 2



1.4 Correlation of returns-Single Factor Model

The next table illustrates a sample of correlations returns which obtained applying the *Single Factor Model.*

Table 2

	HRAK	IASO	IATR	INLOT	KLEM	MARFKO	BELA	BOC	TPEIR	ROKKA	TITK	TITP
HRAK	1.00	0.26	0.29	0.31	0.28	0.20	0.24	0.31	0.41	0.27	0.27	0.28
IASO	0.26	1.00	0.22	0.23	0.21	0.15	0.18	0.23	0.31	0.20	0.20	0.21
IATR	0.29	0.22	1.00	0.26	0.23	0.17	0.20	0.26	0.34	0.23	0.22	0.23
INLOT	0.31	0.23	0.26	1.00	0.25	0.18	0.21	0.28	0.36	0.24	0.24	0.25
KLEM	0.28	0.21	0.23	0.25	1.00	0.16	0.19	0.25	0.33	0.22	0.21	0.22
MARFKO	0.20	0.15	0.17	0.18	0.16	1.00	0.14	0.18	0.23	0.16	0.15	0.16
BELA	0.24	0.18	0.20	0.21	0.19	0.14	1.00	0.22	0.28	0.19	0.19	0.19
BOC	0.31	0.23	0.26	0.28	0.25	0.18	0.22	1.00	0.37	0.24	0.24	0.25
TPEIR	0.41	0.31	0.34	0.36	0.33	0.23	0.28	0.37	1.00	0.32	0.31	0.33
ROKKA	0.27	0.20	0.23	0.24	0.22	0.16	0.19	0.24	0.32	1.00	0.21	0.22
TITK	0.27	0.20	0.22	0.24	0.21	0.15	0.19	0.24	0.31	0.21	1.00	0.21
TITP	0.28	0.21	0.23	0.25	0.22	0.16	0.19	0.25	0.33	0.22	0.21	1.00

1.4 Correlation of returns-Fundamental Factor Model

The table below represents the industry factor correlation matrix which is based on *case 2 (*based on industry classification*).*

Table 3

	Alpha	BIOX	BRAIN	GEK	VIVART	DIFF	EGNAK	ELTRK	ELBA	ELKA	ETEM	HRAK	IASO
Alpha	1.00	0.34	0.19	0.29	0.29	0.20	0.44	0.26	0.30	0.24	0.27	0.31	0.28
BIOX	0.34	1.00	0.24	0.41	0.56	0.25	0.29	0.33	0.57	0.46	0.51	0.40	0.39
BRAIN	0.19	0.24	1.00	0.21	0.21	0.18	0.16	0.23	0.21	0.17	0.19	0.28	0.20
GEK	0.29	0.41	0.21	1.00	0.36	0.22	0.25	0.29	0.36	0.29	0.33	0.35	0.54
VIVART	0.29	0.56	0.21	0.36	1.00	0.22	0.25	0.29	0.50	0.40	0.45	0.35	0.34
DIFF	0.20	0.25	0.18	0.22	0.22	1.00	0.17	0.24	0.23	0.18	0.20	0.29	0.21
EGNAK	0.44	0.29	0.16	0.25	0.25	0.17	1.00	0.22	0.26	0.21	0.23	0.27	0.24
ELTRK	0.26	0.33	0.23	0.29	0.29	0.24	0.22	1.00	0.30	0.24	0.27	0.38	0.28
ELBA	0.30	0.57	0.21	0.36	0.50	0.23	0.26	0.30	1.00	0.41	0.46	0.36	0.35
ELKA	0.24	0.46	0.17	0.29	0.40	0.18	0.21	0.24	0.41	1.00	0.37	0.29	0.28
ETEM	0.27	0.51	0.19	0.33	0.45	0.20	0.23	0.27	0.46	0.37	1.00	0.32	0.31
HRAK	0.31	0.40	0.28	0.35	0.35	0.29	0.27	0.38	0.36	0.29	0.32	1.00	0.34
IASO	0.28	0.39	0.20	0.54	0.34	0.21	0.24	0.28	0.35	0.28	0.31	0.34	1.00
IATR	0.29	0.41	0.21	0.56	0.35	0.22	0.25	0.29	0.36	0.29	0.33	0.35	0.53
INLOT	0.26	0.34	0.23	0.30	0.29	0.25	0.22	0.32	0.30	0.24	0.27	0.39	0.28
KLEM	0.29	0.55	0.21	0.35	0.48	0.22	0.25	0.29	0.49	0.39	0.44	0.34	0.33
MARFKO	0.47	0.31	0.17	0.26	0.27	0.18	0.40	0.24	0.27	0.22	0.25	0.28	0.25
BELA	0.27	0.35	0.24	0.31	0.30	0.25	0.23	0.33	0.31	0.25	0.28	0.40	0.29
BOC	0.43	0.29	0.16	0.24	0.25	0.17	0.37	0.22	0.25	0.20	0.23	0.26	0.23
TPEIR	0.52	0.34	0.19	0.29	0.30	0.20	0.45	0.26	0.31	0.25	0.28	0.32	0.28
ROKKA	0.28	0.54	0.20	0.34	0.47	0.21	0.24	0.28	0.48	0.39	0.43	0.34	0.33
TITK	0.29	0.55	0.20	0.35	0.47	0.22	0.25	0.28	0.49	0.39	0.44	0.34	0.33
TITP	0.30	0.39	0.27	0.34	0.34	0.28	0.26	0.37	0.34	0.28	0.31	0.45	0.33

	IATR	INLOT	KLEM	MARFKO	BELA	BOC	TPEIR	ROKKA	TITK	TITP
Alpha	0.29	0.26	0.29	0.47	0.27	0.43	0.52	0.28	0.29	0.30
BIOX	0.41	0.34	0.55	0.31	0.35	0.29	0.34	0.54	0.55	0.39
BRAIN	0.21	0.23	0.21	0.17	0.24	0.16	0.19	0.20	0.20	0.27
GEK	0.56	0.30	0.35	0.26	0.31	0.24	0.29	0.34	0.35	0.34
VIVART	0.35	0.29	0.48	0.27	0.30	0.25	0.30	0.47	0.47	0.34
DIFF	0.22	0.25	0.22	0.18	0.25	0.17	0.20	0.21	0.22	0.28
EGNAK	0.25	0.22	0.25	0.40	0.23	0.37	0.45	0.24	0.25	0.26
ELTRK	0.29	0.32	0.29	0.24	0.33	0.22	0.26	0.28	0.28	0.37
ELBA	0.36	0.30	0.49	0.27	0.31	0.25	0.31	0.48	0.49	0.34
ELKA	0.29	0.24	0.39	0.22	0.25	0.20	0.25	0.39	0.39	0.28
ETEM	0.33	0.27	0.44	0.25	0.28	0.23	0.28	0.43	0.44	0.31
HRAK	0.35	0.39	0.34	0.28	0.40	0.26	0.32	0.34	0.34	0.45
IASO	0.53	0.28	0.33	0.25	0.29	0.23	0.28	0.33	0.33	0.33
IATR	1.00	0.30	0.35	0.26	0.30	0.24	0.29	0.34	0.35	0.34
INLOT	0.30	1.00	0.29	0.24	0.33	0.22	0.27	0.28	0.29	0.37
KLEM	0.35	0.29	1.00	0.26	0.30	0.24	0.29	0.46	0.47	0.33
MARFKO	0.26	0.24	0.26	1.00	0.24	0.39	0.47	0.26	0.26	0.27
BELA	0.30	0.33	0.30	0.24	1.00	0.23	0.27	0.29	0.29	0.39
BOC	0.24	0.22	0.24	0.39	0.23	1.00	0.44	0.24	0.24	0.25
TPEIR	0.29	0.27	0.29	0.47	0.27	0.44	1.00	0.29	0.29	0.31
ROKKA	0.34	0.28	0.46	0.26	0.29	0.24	0.29	1.00	0.46	0.32
TITK	0.35	0.29	0.47	0.26	0.29	0.24	0.29	0.46	1.00	0.33
TITP	0.34	0.37	0.33	0.27	0.39	0.25	0.31	0.32	0.33	1.00

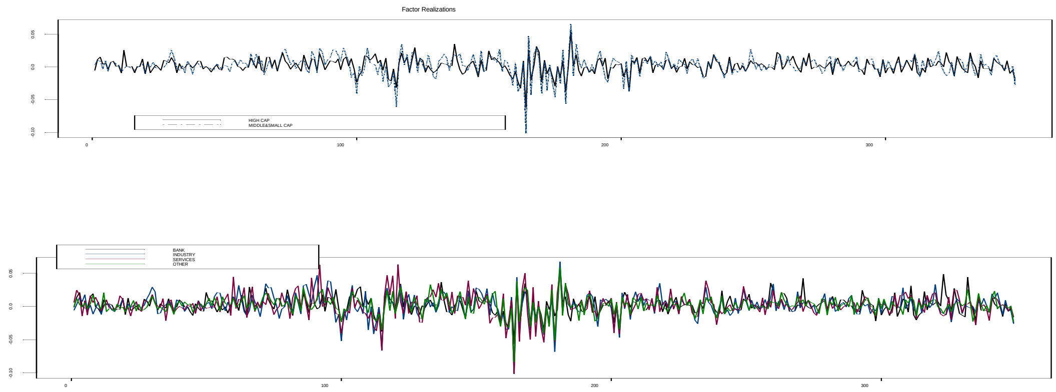
The table below represents the industry factor correlation matrix which is based on *case 1* (based on the size of the company).

Table 4

	Alpha	BIOX	BRAIN	GEK	VIVART	DIFF	EGNAK	ELTRK	ELBA	ELKA	ETEM	HRAK	IASO
Alpha	1.00	0.38	0.23	0.34	0.34	0.24	0.34	0.31	0.32	0.26	0.28	0.36	0.29
BIOX	0.38	1.00	0.19	0.29	0.29	0.20	0.29	0.26	0.27	0.22	0.24	0.31	0.24
BRAIN	0.23	0.19	1.00	0.29	0.29	0.21	0.17	0.27	0.27	0.22	0.24	0.31	0.24
GEK	0.34	0.29	0.29	1.00	0.44	0.31	0.26	0.40	0.41	0.33	0.36	0.47	0.37
VIVART	0.34	0.29	0.29	0.44	1.00	0.31	0.26	0.40	0.41	0.34	0.37	0.47	0.37
DIFF	0.24	0.20	0.21	0.31	0.31	1.00	0.18	0.29	0.29	0.24	0.26	0.33	0.26
EGNAK	0.34	0.29	0.17	0.26	0.26	0.18	1.00	0.24	0.24	0.20	0.22	0.28	0.22
ELTRK	0.31	0.26	0.27	0.40	0.40	0.29	0.24	1.00	0.37	0.31	0.33	0.43	0.34
ELBA	0.32	0.27	0.27	0.41	0.41	0.29	0.24	0.37	1.00	0.31	0.34	0.43	0.34
ELKA	0.26	0.22	0.22	0.33	0.34	0.24	0.20	0.31	0.31	1.00	0.28	0.36	0.28
ETEM	0.28	0.24	0.24	0.36	0.37	0.26	0.22	0.33	0.34	0.28	1.00	0.39	0.31
HRAK	0.36	0.31	0.31	0.47	0.47	0.33	0.28	0.43	0.43	0.36	0.39	1.00	0.39
IASO	0.29	0.24	0.24	0.37	0.37	0.26	0.22	0.34	0.34	0.28	0.31	0.39	1.00
IATR	0.33	0.28	0.28	0.42	0.42	0.30	0.25	0.39	0.39	0.32	0.35	0.45	0.35
INLOT	0.31	0.26	0.26	0.40	0.40	0.28	0.24	0.37	0.37	0.30	0.33	0.43	0.34
KLEM	0.33	0.28	0.28	0.43	0.43	0.31	0.26	0.40	0.40	0.33	0.36	0.46	0.36
MARFKO	0.36	0.30	0.18	0.27	0.28	0.20	0.27	0.25	0.25	0.21	0.23	0.29	0.23
BELA	0.32	0.27	0.27	0.41	0.41	0.29	0.24	0.38	0.38	0.31	0.34	0.44	0.34
BOC	0.35	0.30	0.18	0.27	0.27	0.19	0.27	0.25	0.25	0.20	0.22	0.29	0.23
TPEIR	0.45	0.38	0.23	0.34	0.35	0.24	0.34	0.32	0.32	0.26	0.29	0.37	0.29
ROKKA	0.37	0.32	0.19	0.29	0.29	0.20	0.29	0.26	0.27	0.22	0.24	0.31	0.24
TITK	0.42	0.36	0.21	0.32	0.33	0.23	0.32	0.30	0.30	0.25	0.27	0.35	0.27
TITP	0.45	0.38	0.22	0.34	0.34	0.24	0.34	0.31	0.32	0.26	0.28	0.36	0.29

Below there is an illustration of factor realizations obtained for each case.

Figure 3



1.4 Correlation of returns-Statistical Factor Model

Below we represent the results that we obtained from S-PLUS in order to determine the sufficient number of common factors.

```
Test of the hypothesis that 3 factors are sufficient
versus the alternative that more are required:
The chi square statistic is 248.34 on 187 degrees of freedom.
The p-value is 0.00179
```

As we see the null hypothesis for 3 factors is rejected in 1% significance level, but the null hypothesis is not rejected in case of 4 factors.

```
Test of the hypothesis that 4 factors are sufficient
versus the alternative that more are required:
The chi square statistic is 178.84 on 167 degrees of freedom.
The p-value is 0.252
```

Below there is a presentation of the factor model correlation matrix.

Table 5

	BIOX	GEK	DIFF	ELTRK	ELKA	HRAK	IATR	KLEM	BELA	TPEIR	TITK
BIOX	1.00	0.49	0.37	0.402	0.65	0.45	0.39	0.40	0.26	0.44	0.254
GEK	0.49	1.00	0.34	0.359	0.48	0.44	0.38	0.39	0.28	0.44	0.273
DIFF	0.37	0.34	1.00	0.276	0.36	0.33	0.28	0.29	0.20	0.32	0.154
ELTRK	0.40	0.36	0.28	1.000	0.39	0.35	0.31	0.32	0.20	0.35	0.099
ELKA	0.65	0.48	0.36	0.392	1.00	0.44	0.38	0.39	0.25	0.42	0.227
HRAK	0.45	0.44	0.33	0.349	0.44	1.00	0.37	0.38	0.27	0.43	0.247
IATR	0.39	0.38	0.28	0.313	0.38	0.37	1.00	0.33	0.22	0.36	0.129
KLEM	0.40	0.39	0.29	0.316	0.39	0.38	0.33	1.00	0.23	0.38	0.155
BELA	0.26	0.28	0.20	0.203	0.25	0.27	0.22	0.23	1.00	0.27	0.216
TPEIR	0.44	0.44	0.32	0.345	0.42	0.43	0.36	0.38	0.27	1.00	0.273
TITK	0.25	0.27	0.15	0.099	0.23	0.25	0.13	0.15	0.22	0.27	1.000

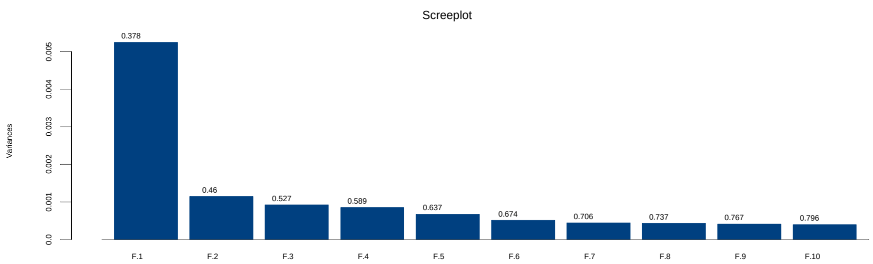
1.5 Correlation of returns-Principal Component Analysis

Below there is a presentation of the correlation matrix returns estimated via PCA.

Table 6

	BIOX	GEK	DIFF	EGNAK	ELTRK	ELKA	HRAK	BELA	TPEIR
BIOX	1.00	0.53	0.46	0.22	0.43	0.64	0.47	0.28	0.47
GEK	0.53	1.00	0.40	0.19	0.37	0.56	0.41	0.24	0.41
DIFF	0.46	0.40	1.00	0.17	0.32	0.48	0.36	0.21	0.35
EGNAK	0.22	0.19	0.17	1.00	0.16	0.24	0.17	0.10	0.17
ELTRK	0.43	0.37	0.32	0.16	1.00	0.45	0.34	0.20	0.33
ELKA	0.64	0.56	0.48	0.24	0.45	1.00	0.50	0.29	0.50
HRAK	0.47	0.41	0.36	0.17	0.34	0.50	1.00	0.22	0.37
BELA	0.28	0.24	0.21	0.10	0.20	0.29	0.22	1.00	0.21
TPEIR	0.47	0.41	0.35	0.17	0.33	0.50	0.37	0.21	1.00

Figure 4
Screeplot of eigenvalues from PCA of ASE returns



The next two figures represent the loadings of the third and fourth principal component respectively.

Figure 5

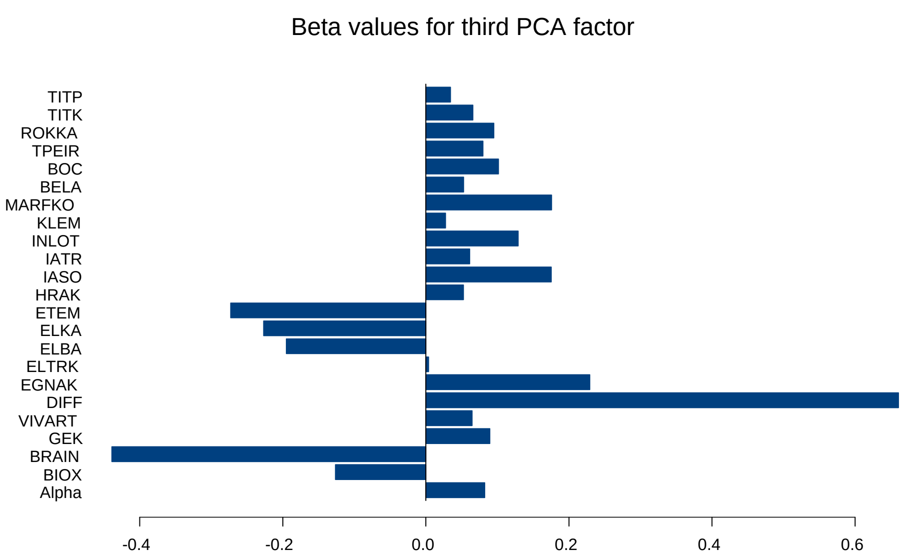
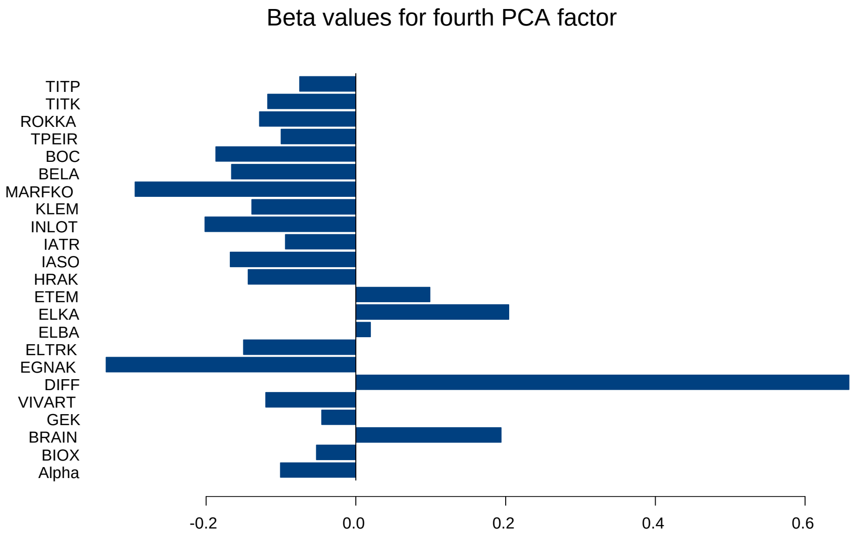


Figure 6

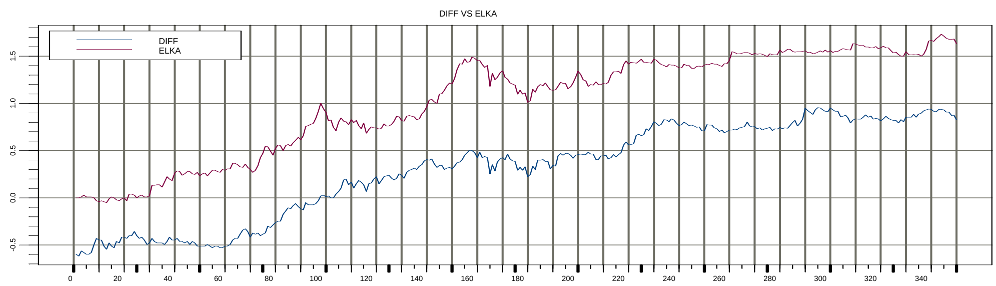


2.1 Testing for Co-Integration in Practice

Up to know we have discussed the tools which give us hints for co-integrated time-series variables, and the *Engle-Granger procedure* which is a tool to test for co-integration. At this point we give an example following the Engle-Granger procedure step by step.

From the statistical analysis in *chapter 1* there is a suspicion that the stocks 'DIFF' and 'ELKA' share a common trend which implies that they may be co-integrated. Transforming the time-series variables to the log-scale and considering the daily closing prices from 21/9/2005 up to 12/2/2007 we computed the correlation equal to 0.972 and the corresponding plot is the following.

Figure 7

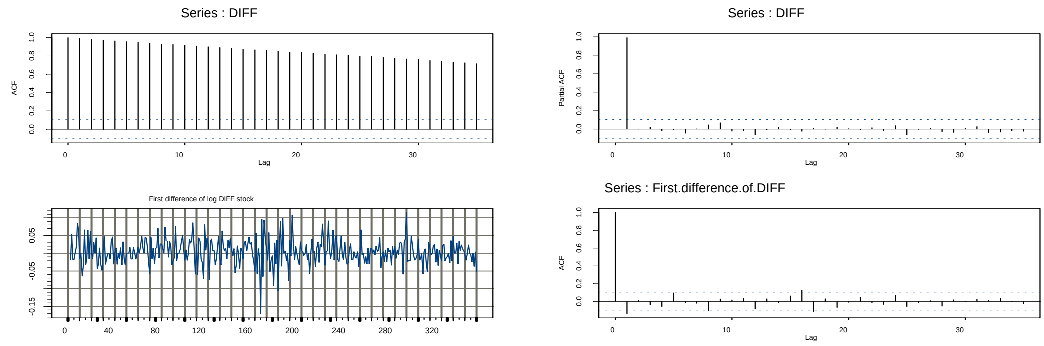


Now the first step is to determine if the time-series variables are both $I(1)$ applying the unit root test which discussed above. Before implementing the unit root test, it will be a proper to represents some plots which give us hints about the time-series variables. We start the analysis taking the logarithm of the stock 'DIFF' representing the autocorrelation plot, the partial autocorrelation plot[1], the time-series plot taking

[1] Notice that the PACF is used to give us hints about the order of an AR model.

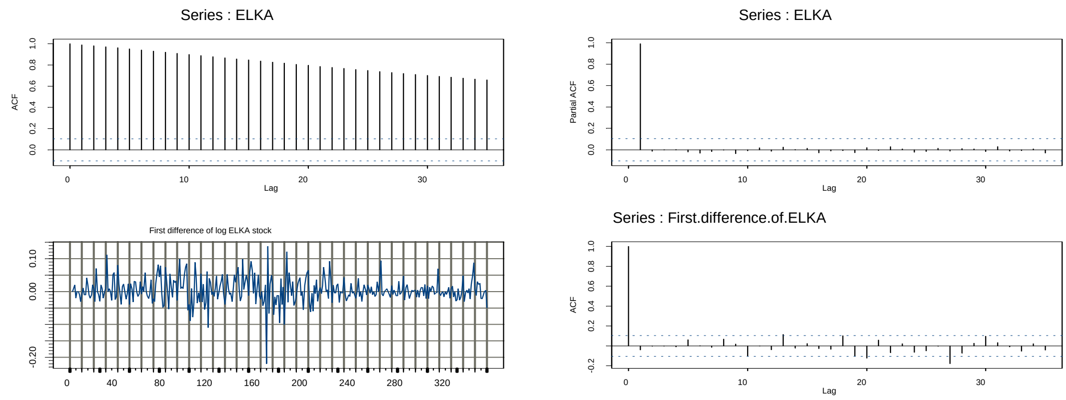
the first difference and finally the autocorrelation plot of the first difference. We follow the same procedure for the logarithm of the stock ‘ELKA’.

Figure 8



From the autocorrelation plot of the log ‘DIFF’ we can observe that the autocorrelations $\hat{\rho}_k$ decay very slowly which is an evidence of a non-stationary process. Also the partial autocorrelation plot indicates that an $AR(1)$ model may explain the time-series variable. Moreover taking the first difference of the logarithm of the stock ‘DIFF’, it seems that we obtain a stationary process, which looks like a *white-noise* process[2]. As we can observe from the figure below, we lead to the same arguments considering the corresponding plots of the stock ‘ELKA’

Figure 9



The next step now is to apply the ADF and the PP tets. The next table represents the results that we obtained from the S-Plus, applying the ADF test to the logarithm of the stock “DIFF”. The null hypothesis to be tested is that $\log(DIFF) \sim I(1)$ with drift, against the alternative that $\log(DIFF) \sim I(0)$ about a deterministic time trend. The ADF t-statistic to test these hypotheses is computed with constant and time trend in the test regression and five lags selected applying the Ng-Perron backward selection.

[2] For details about these time-series concepts, a perfect text book is that of Box, G.E.P. and Jenkins, G.M. 1976. Time Series Analysis. Holden-Day

Table 7

```
Test for Unit Root: Augmented DF Test

Null Hypothesis: there is a unit root
   Type of Test: t test
 Test Statistic: -1.311
        P-value: 0.8835

Coefficients:
           Value Std. Error t value Pr(>|t|)
    lag1 -0.0187  0.0143    -1.3110  0.1908
    lag2 -0.1317  0.0551    -2.3918  0.0173
    lag3 -0.0064  0.0554    -0.1165  0.9074
    lag4 -0.0440  0.0553    -0.7956  0.4268
    lag5 -0.0652  0.0546    -1.1955  0.2327
constant  0.0110  0.0047     2.3338  0.0202
    time  0.0001  0.0001     0.9076  0.3647
```

From the table above we observe that the $ADF_t = -1.311$ and has a *p-value* equal to 0.8835, hence one clearly does not reject the null hypothesis. Below we represent the results of the PP test for the same data.

Table 8

```
Test for Unit Root: Phillips-Perron Test

Null Hypothesis: there is a unit root
   Type of Test: t test
 Test Statistic: -1.607
        P-value: 0.7884
using bartlett window with bandwidth 5

Coefficients:
           Value Std. Error t value Pr(>|t|)
    lag1 -0.0264  0.0138    -1.9148  0.0563
constant  0.0120  0.0046     2.6430  0.0086
    time  0.0001  0.0001     1.5630  0.1190
```

Again we do not have strong evidence against the null hypothesis (p-value equal to 0.7784).
In sequel we represent the corresponding tests for the logarithm of the 'ELKA' stock.

Table 9

```
Test for Unit Root: Augmented DF Test

Null Hypothesis: there is a unit root
   Type of Test: t test
 Test Statistic: -1.748
        P-value: 0.7276

Coefficients:
           Value Std. Error t value Pr(>|t|)
    lag1 -0.0196  0.0112    -1.7483  0.0813
constant  0.0247  0.0116     2.1269  0.0341
    time  0.0001  0.0001     1.3119  0.1904
```

Table 10

```
Test for Unit Root: Phillips-Perron Test

Null Hypothesis: there is a unit root
   Type of Test: t test
 Test Statistic: -1.725
        P-value: 0.7385
using bartlett window with bandwidth 5

Coefficients:
           Value Std. Error t value Pr(>|t|)
    lag1 -0.0196  0.0112    -1.7483  0.0813
constant  0.0247  0.0116     2.1269  0.0341
    time  0.0001  0.0001     1.3119  0.1904
```

From these test above we can conclude that both time-series variables are $I(1)$. Hence we can continue the Engle-Granger procedure estimating the co-integrating vector $\boldsymbol{\gamma} = (1, -\gamma_2)'$ using the OLS approach[3]. The model that we fit is as follows.

$$\log(\text{DIFF}) = \mu + \gamma_2 \log(\text{ELKA}) + \varepsilon_t$$

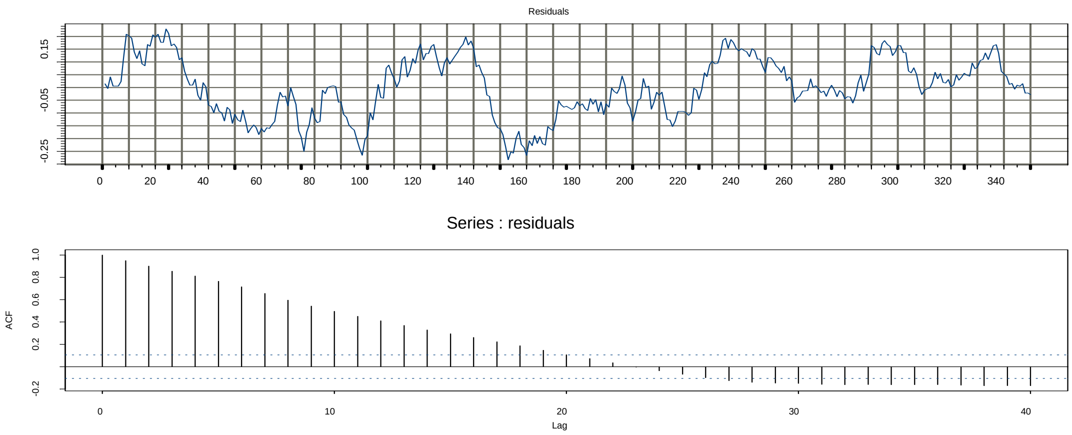
Applying the OLS method we get the following results:

```
Coefficients:
                Value Std. Error  t value Pr(>|t|)
(Intercept)  -0.6123   0.0133   -45.9923   0.0000
       ELKA   0.8968   0.0115    77.8581   0.0000
```

[3] Later on we will consider the Dynamic Ordinary Least Square (DOLS) approach.

Hence $\boldsymbol{\gamma} = (1, -\gamma_2)' = (1, -0.8968)$ and $\boldsymbol{\gamma}'\mathbf{X}_t = \log(\mathrm{DIFF})\text{-}\gamma_2\log(\mathrm{ELKA}) = -0.6123 + \varepsilon_t$. Now we would like to check if the $spread = -0.6123 + \varepsilon_t$ is a $I(0)$ and stationary or equivalently to examine just the mispricing ε_t. Of course the ε_t are unknown and for that reason we consider the estimated residuals. Below we represent the results we obtain applying the unit root test as well as the KPSS stationarity test, illustrating a time-series plot of the residuals.

Figure 10




```
Test for Unit Root: Augmented DF Test

Null Hypothesis: there is a unit root
   Type of Test: t test
 Test Statistic: -2.936
        P-value: 0.003362

Coefficients:
       Value Std. Error t value Pr(>|t|)
lag1 -0.0505  0.0172    -2.9361  0.0035
lag2  0.0212  0.0538     0.3943  0.6936
lag3 -0.0054  0.0538    -0.1001  0.9203
```

```
Test for Unit Root: Phillips-Perron Test

Null Hypothesis: there is a unit root
   Type of Test: t test
 Test Statistic: -3.075
        P-value: 0.002151

Coefficients:
       Value Std. Error t value Pr(>|t|)
lag1 -0.0498  0.0167    -2.9788  0.0031
```

```
Test for Stationarity: KPSS Test

 Null Hypothesis: stationary around a constant

 Test Statistics:
  residuals
  0.5378*

  * : significant at 5% level
 ** : significant at 1% level
```

Applying the ADF and PP tests we reject the null hypothesis that the residuals are $I(1)$ against the alternative against the alternative that the residuals are $I(0)$. Hence considering the Engle-Granger procedure we can conclude that the time-series variables are co-integrated. But the problem which arises is that the KPSS test reject the null hypothesis that the residuals are a stationary process around zero, which means that we reject the null hypothesis that they are $I(0)$ at 5% significance level. Thus it depends on the analyst if he/she can consider that these time series are co-integrated or not, but strictly speaking can not be considered co-integrated.

Let us examine what would have happened if we had concluded that these two stocks are co-integrated, applying a pairs trading strategy for 50 trading days. The trading strategy is simply. As we have mentioned in '*Chapter 1.*', if two or more assets are co-integrated, means that they have long-run equilibrium relationships and can not diverge too far apart from the equilibrium. Hence economic forces will act to restore the equilibrium relationship. Based on this, when the *mispricing* has diverged significantly above its mean (which is zero by construction) then we sell the portfolio (spread), and when the *mispricing* has diverged significantly below its mean then we buy the spread. In each case we close our position when the *mispricing* converges back to its equilibrium. We consider a significant divergence a $\pm SE$ of the residuals, where in our case is equal to 0.1159. [4]

Thus when

- $\log(\text{DIFF})-0.8968\times\log(\text{ELKA})+0.6123>0.1159$

then we sell 1 stock of DIFF and we buy 0.8968 stocks of ELKA (or equivalently we sell a *spread*), and when

- $\log(\text{DIFF})-0.8968\times\log(\text{ELKA})+0.6123<-0.1159$

then we buy one stock of DIFF and we sell 0.8968 stock of ELKA (or equivalently we buy a *spread*).

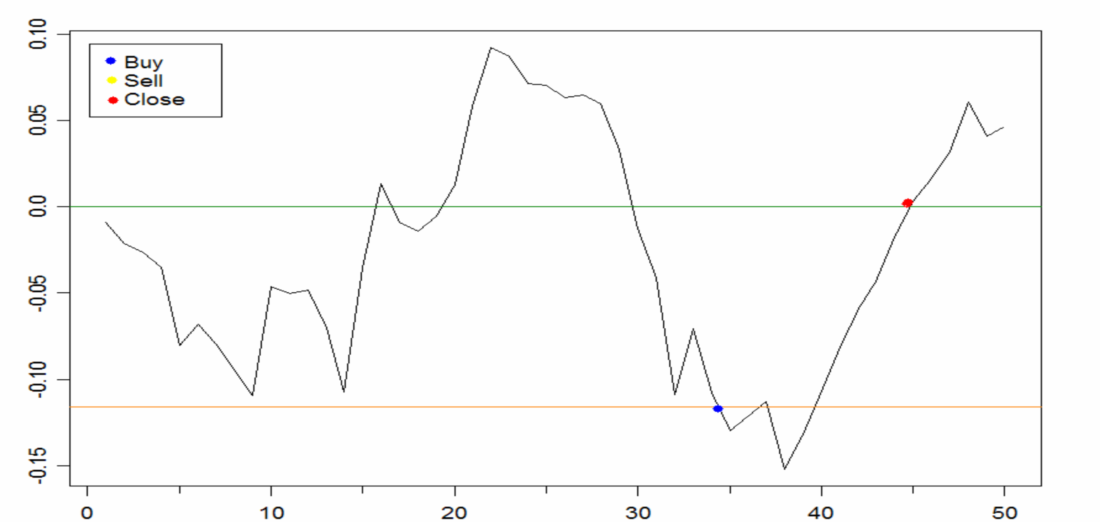
Applying the strategy for the 50 next trading days (13/2/07-26/4/07) we found that the mispricing never diverged above the threshold 0.1159 but it diverged below the

---

[4] The standard error is given by the formula $S.E=\sqrt{\dfrac{\sum_{i=1}^{n} e_i^2}{n-(k+1)}}$ where in our case k=1(explanatory variable) and n=350

threshold -0.1159. The strategy is illustrated below graphically. The results are referred to the time period 13/2/2007-26/4/2007.

Figure 11



The strategy is as follows:

3/4/2007: We buy one stock of ‘DIFF’ at **2.31€** and we go-short 0.8968 stocks of ‘ELKA’ at **5.82€** ($5.82 \times 0.8968 = 5.219$), hence the portfolio cost us **-2.909376€**

19/4/2007: We sell the stock of ‘DIFF’ at **2.5€** and we return back the 0.8968 stocks of ‘ELKA’ at **5.48€** ($5.48 \times 0.8968 = 4.9144$). Thus we make a total profit equal to **0.4949€.**

Notice that in order to enter this strategy we did not need any initial capital. This is one reason why it is referred as Statistical Arbitrage Strategy. Of course we did not take it into account the transaction cost as well as we made the assumptions that the assets are divisible. Later on we will try to explain better the strategy, discussing other sensible thresholds, and taking into account the transaction cost. Still now, we focus on co-integrating modeling, and the above example is to give an idea of how the pairs trading strategy could work.

3.1 Dynamic OLS and ECM

Below there is a presentation of the output that we received form S-Plus implementing the Error Correction Model considering the logarithm of the stock ‘BIOX’ and ‘ELBA’.

Table 11

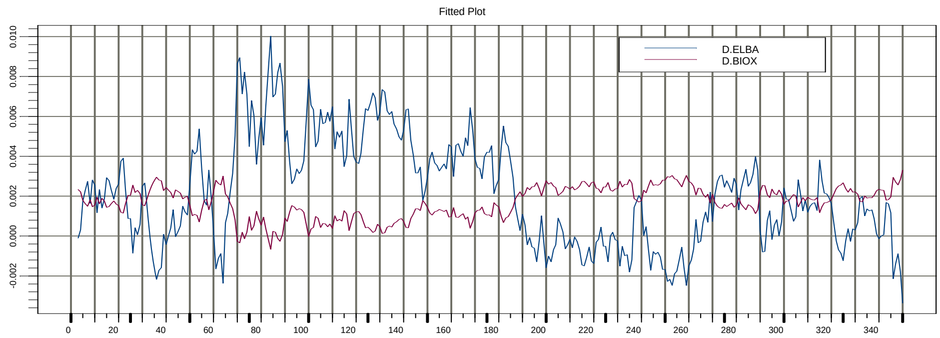
```
D.BIOX ~ tslag(U.HAT) + tslag(D.BIOX) +tslag(D.ELBA)

 Coefficients:
                  Value Std. Error t value Pr(>|t|)
   (Intercept)  0.0015  0.0012     1.2674  0.2059
  tslag(U.HAT) -0.0253  0.0286    -0.8848  0.3769
 tslag(D.BIOX)  0.1555  0.0846     1.8383  0.0669
 tslag(D.ELBA) -0.0568  0.0632    -0.9000  0.3687

Regression Diagnostics:

         R-Squared 0.0129
Adjusted R-Squared 0.0043
Durbin-Watson Stat 1.9948
```

```
D.ELBA ~ tslag(U.HAT) + tslag(D.BIOX) + tslag(D.ELBA)

Coefficients:
                Value Std. Error t value Pr(>|t|)
  (Intercept)  0.0022  0.0016     1.3400  0.1811
 tslag(U.HAT)  0.0569  0.0387     1.4708  0.1423
tslag(D.BIOX)  0.0960  0.1143     0.8400  0.4015
tslag(D.ELBA) -0.0540  0.0853    -0.6332  0.5270

Regression Diagnostics:

         R-Squared 0.0099
Adjusted R-Squared 0.0013
Durbin-Watson Stat 2.0021
```

```
D.BIOX ~ tslag(U.HAT)
Coefficients:
               Value Std. Error t value Pr(>|t|)
 (Intercept)  0.0017  0.0012     1.3892  0.1657
tslag(U.HAT) -0.0185  0.0281    -0.6599  0.5098
=================================================
D.ELBA ~ tslag(U.HAT)
Coefficients:
              Value Std. Error t value Pr(>|t|)
 (Intercept) 0.0022 0.0016     1.3771  0.1694
 tslag(U.HAT) 0.0625 0.0378     1.6566  0.0985
```

From the above output we may conclude that the specific model has not very good behavior since the p-value of the coefficients are larger than 5%. Below we represent the fitted values of the estimated *Error Correction Model*, where it indicated that the model is not the ideal one. However as we have observed (*Chapter 3.3*) the *Statistical Arbitrage strategy* works well.

Figure 12



3.2 Application of an ECM

At this point we give an example of a pair of stocks which seems to work better. The tickers of the stocks are ‘*BOC*’ and ‘*TPEIR*’. The first step is to estimate the co-integrating vector applying the DOLS, the next step is to test if the residuals are indeed a stationary process applying the unit root and stationarity tests, then we estimate the Error Correction Model, and finally we apply the strategy. Below we represent the outputs obtained from S-Plus.

Figure 13

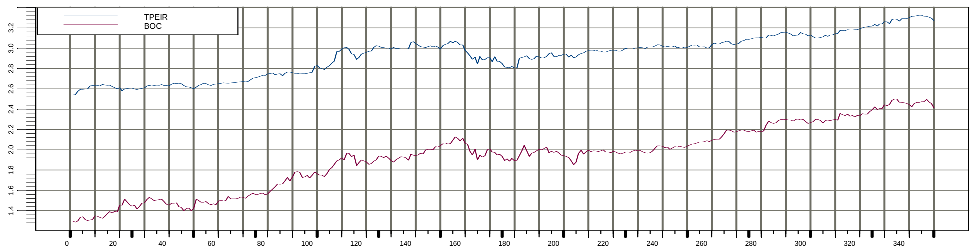


Table 12-Dynamic OLS

Coefficients:

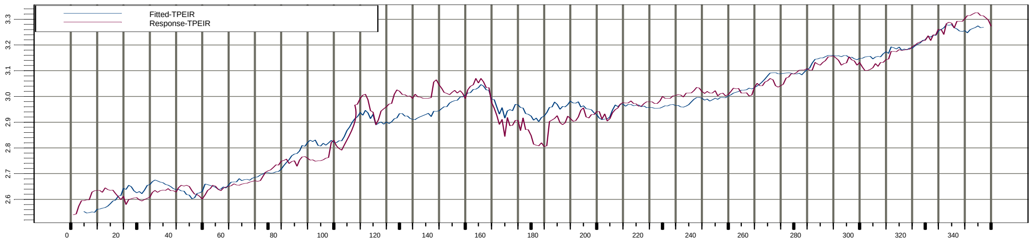
	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	**1.7251**	0.0282	61.1742	0.0000
BOC	**0.6243**	0.0140	44.6280	0.0000
D.BOC.lead3	0.1120	0.0967	1.1579	0.2477
D.BOC.lead2	0.2506	0.0954	2.6258	0.0090
D.BOC.lead1	0.2444	0.0916	2.6668	0.0080
D.BOC.lag0	-0.0192	0.1006	-0.1911	0.8485
D.BOC.lag1	0.0019	0.0956	0.0202	0.9839
D.BOC.lag2	0.0329	0.1082	0.3039	0.7614
D.BOC.lag3	0.0657	0.1081	0.6076	0.5439

Regression Diagnostics:

R-Squared 0.9538

The next plot represents the fitted values and the real values of the stock *‘TPEIR’* obtained applying the DOLS model above.

Figure14



The next plot is an illustration of the ‘*mispricing*’.

Figure 15

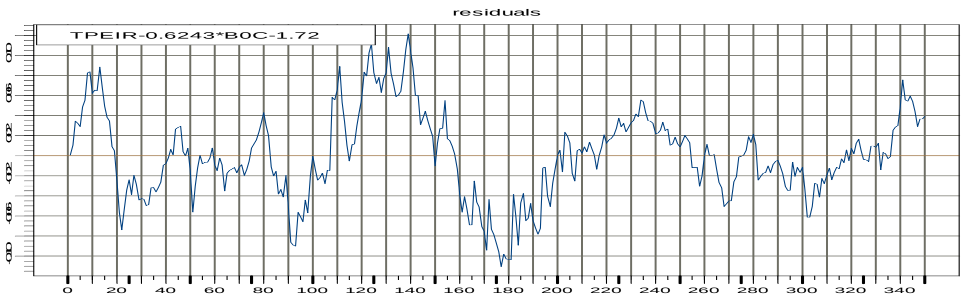


Table 13-Stationarity and Unit Root tests

Test for Unit Root: Augmented DF ***Null Hypothesis:*** **there is a unit root** Type of Test: t-test Test Statistic: -2.861 ***P-value: 0.004245***	***Test for Unit Root: Phillips-Perron*** ***Null Hypothesis:*** **there is a unit root** Type of Test: t-test Test Statistic: -3.532 ***P-value: 4.433e-4***

Test for Stationarity: KPSS Test

Null Hypothesis: stationary around a constant

Test Statistics:

0.1312

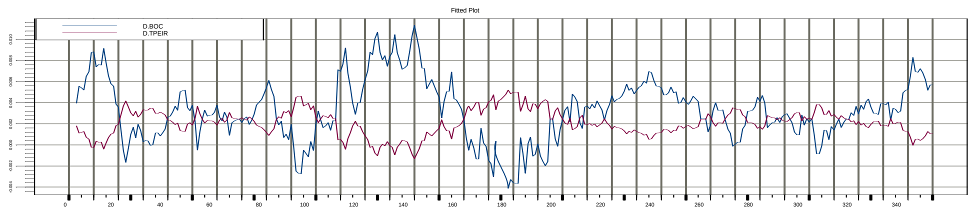
* : significant at 5% level
** : significant at 1% level

Table 14: Error Correction Models

```
                  D.TPEIR ~ tslag(U.HAT)

Coefficients:
               Value Std. Error t value Pr(>|t|)
 (Intercept)  0.0021  0.0009      2.2865  0.0228
tslag(U.HAT) -0.0281  0.0213     -1.3181  0.1883
```

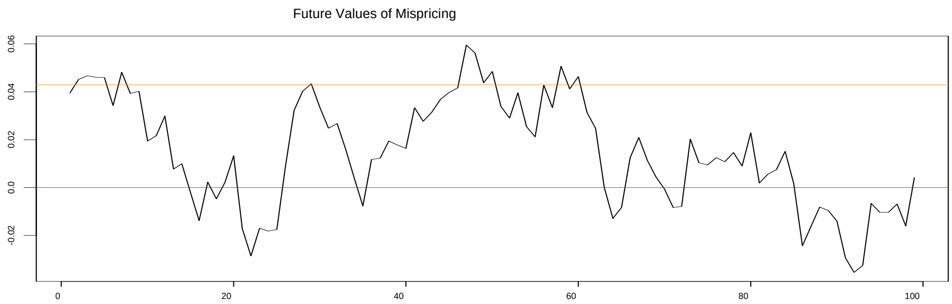
```
                  D.BOC ~ tslag(U.HAT)

Coefficients:
              Value Std. Error t value Pr(>|t|)
 (Intercept) 0.0033 0.0013      2.5410  0.0115
tslag(U.HAT) 0.0666 0.0299      2.2304  0.0264
```

Figure 16:Fitted Values of the ECM



From the above statistical analysis we may conclude that these time series variables are co-integrated. Below we adapt the trading strategy for the next 99 trading days (13/2/2007-6/7/2007) and the results are illustrated below.

Figure 17



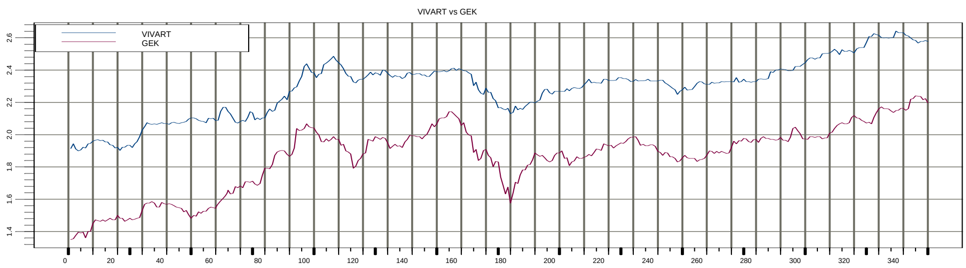
14/02/2007: Short-sell the spread at 19.86€
06/03/2007: Close the position paying back 18.23€
26/03/2007: Short-sell the spread at 19.61€
03/04/2007: Close the position paying back 18.53€
23/04/2007: Short-Sell the spread at 20.30€
17/05/2007: Close the position paying back 19.68€

The net profit of this strategy is equal to **3.31€.** Notice that the initial investment is 0.

4.1 Implementation of Johansen's Methodology

At this point we will give an extended example of VAR and VEC models, in order to show clearly how they work these models in practice. As in all examples up to now, we consider a bivariate case of two stocks from ASE which seems to be co-integrated. The values are referred to same period of time as before, and there are transformed to the logarithmic scale.

Figure 18



VAR MODEL

The VAR model may be estimated with the lag length *p* determined using a specified information criterion. In our case we applied the AIC test and the results are illustrated below.

	ar(1)	ar(2)	ar(3)	**ar(4)**	ar(5)	ar(6)
AIC	-3474.389	-3475.669	-3481.829	***-3483.91***	-3477.737	-3474.165

As we can see from the table above, for *p=4* we obtained the minimum AIC. Since we have specified the lag length of the VAR model we are in position to estimate the parameters. The results are representing below:

Coefficients:

	VIVART	GEK
(Intercept)	0.0370	-0.0177
VIVART.lag1	1.0597	-0.0269
GEK.lag1	0.0441	1.0306
VIVART.lag2	0.0469	0.2158
GEK.lag2	-0.0009	0.0055
VIVART.lag3	-0.2029	0.0160
GEK.lag3	0.0371	-0.0303
VIVART.lag4	0.0607	-0.1499
GEK.lag4	-0.0557	-0.0632

Std. Errors of Residuals:

VIVART	GEK
0.0161	0.0236

Information Criteria:

logL	**AIC**	**BIC**	**HQ**
1771.271	**-3506.542**	**-3437.306**	**-3478.972**

Hence the VAR(4) model is written as follows:

$$V_t = 0.0370 + 1.0597V_{t-1} + 0.0441G_{t-1} + 0.0469V_{t-2} - 0.0009G_{t-2} - \\ - 0.2029V_{t-3} + 0.0371G_{t-3} + 0.0607V_{t-4} - 0.0557G_{t-4}$$

and

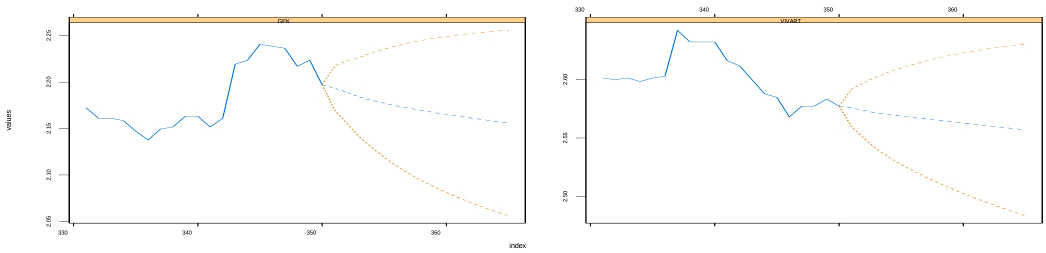
$$G_t = -0.0177 - 0.0269V_{t-1} + 1.0306G_{t-1} + 0.2158V_{t-2} - 0.0055G_{t-2} + \\ +0.0160V_{t-3} - 0.0303G_{t-3} - 0.1499V_{t-4} - 0.0632G_{t-4}$$

where V_t and G_t are referred to the logarithm of the stocks ‘VIVART’ and ‘GEK’ respectively.

FORECASTING

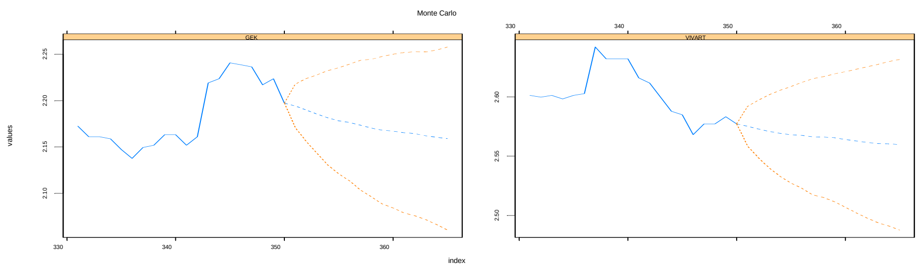
Forecasting from a *VAR* model is similar to forecasting from univaritate *AR* model. Below we represent the 15-step predicted values as well as the 95% CI of both transformed stocks applying the ‘*Traditional Forecasting Algorithm*[5]’.

Figure 19



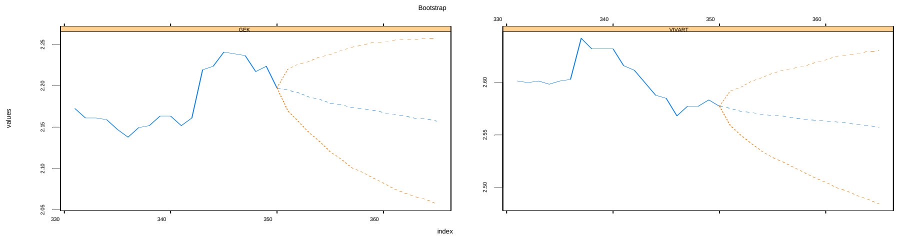
The *h-step* forecasts for these two stocks may be calculated using the *Monte Carlo simulation* or the *Bootstrap simulation method*[6]. The corresponding plots are illustrated below.

Figure 20



[5] See Zivot, E., Wang, J., 2006, ‘*Modeling Financial Time Series with S-PLUS*’ Chapter 11.3.1.

[6] See Zivot, E., Wang, J., 2006, ‘*Modeling Financial Time Series with S-PLUS*’ Chapter 11.3.2



TESTING HYPOTHESIS ON CO-INTEGRATING VECTOR

Now consider testing for the number of co-integrating relations among the logarithms of the stocks "*GEK*" and "*VIVART*". The first step is to determine the number of lags to use in the *VECM*. Before, using the *AIC* test, we found that the corresponding *VAR* model is of order *p=4*, which implies that the lag length for the *VECM* is *p-1=3*. Since the stock prices are trending, the Johansen *LR* tests are computed assuming the *unrestricted constant* case III discussed in *section 4.4.* The results of the tests are illustrated in the next table:

Table 15

```
Trend Specification:
H1(r): Unrestricted constant

Trace tests significant at the 5% level are flagged by ' +'.
Trace tests significant at the 1% level are flagged by '++'.
Max Eigenvalue tests significant at the 5% level are flagged by ' *'.
Max Eigenvalue tests significant at the 1% level are flagged by '**'.

Tests for Cointegration Rank:
         Eigenvalue Trace Stat  99% CV   95% CV Max Stat  99% CV  95% CV
H(0)+ *    0.0427     18.3113   15.4100 20.0400 15.0851  14.0700 18.6300
H(1)       0.0093      3.2262    3.7600  6.6500  3.2262   3.7600  6.6500

Long Run Impact Matrix: Π
        VIVART      GEK
VIVART -0.0356   0.0247
   GEK  0.0550  -0.0574

Unnormalized Cointegrating Vectors:
          VIVART      GEK
coint.1 -18.5404  16.5343
coint.2   5.1308   0.6013

Unnormalized Adjustment Coefficients:
       coint.1 coint.2
VIVART  0.0015 -0.0014
   GEK -0.0034 -0.0016
```

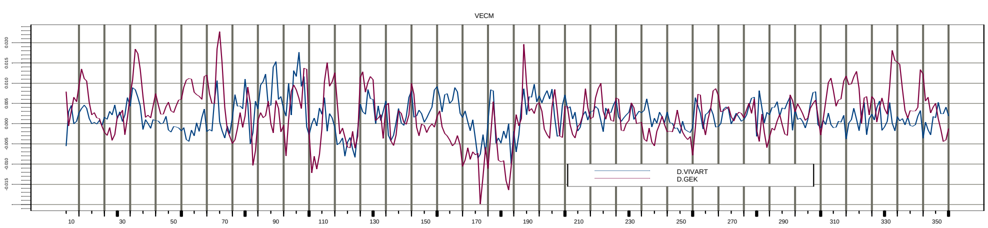
Recall that the number of co-integrating vectors is equal to the number of non-zero eigenvalues of Π. The first row in table gives $LR_{trace}(0)$ and $LR_{\max}(0)$ for testing the null of $r_0 = 0$ co-integrating vectors as well as the 95% and 99% quantiles of the appropriate asymptotic distributions taken from the tables in Osterwald-Lenum (1992). As we can see both the trace and maximum eigenvalue statistics reject the null hypothesis at 5% significance level. The second row in the table gives $LR_{\max}(1)$ and $LR_{trace}(1)$ for testing the null hypothesis $r_0 = 1$, where neither statistic rejects it. Finally in *table 4.1* there is a presentation of the unnormilized co-integrating vectors. For interpretation, it is often convenient to normalized or identify the co-integrating vectors by choosing a specific coordinate system in which to express the variables. One arbitrary way to do this, suggested by Johansen, is to solve for the triangular representation of the co-integrated system. Applying this method we obtained the normalized co-integrating vector which is representing below:

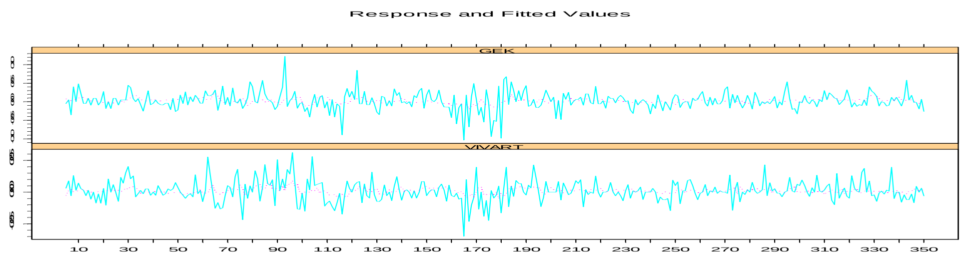
```
 Cointegrating Vector:

VIVART  1.0000
   GEK -0.8918
```

Once the number of co-integrating vectors is determined, the maximum likelihood estimates of the full VECM may be obtained. The results are representing below:

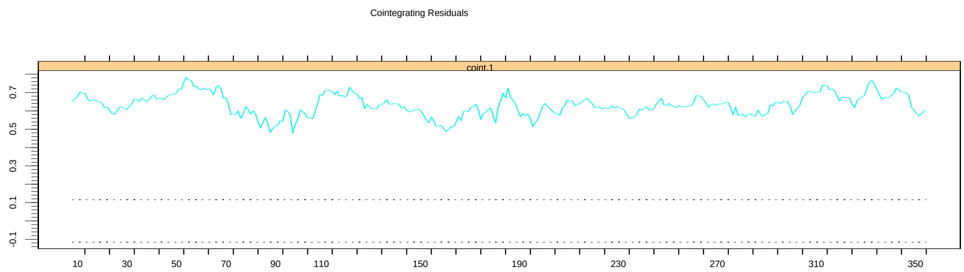
```
Cointegrating Vectors:
       coint.1
VIVART  1.0000
   GEK -0.8918

VECM Coefficients:
              VIVART     GEK
    coint.1 -0.0286  0.0633
VIVART.lag1  0.0940 -0.0834
   GEK.lag1  0.0210  0.0899
VIVART.lag2  0.1388  0.1300
   GEK.lag2  0.0201  0.0954
VIVART.lag3 -0.0643  0.1456
   GEK.lag3  0.0566  0.0643
  Intercept  0.0195 -0.0387

Std. Errors of Residuals:
 VIVART    GEK
 0.0161 0.0236
```

The next two figure is referred to the observed and fitted first-differences of the two stocks, obtained by the VECM(3) above.

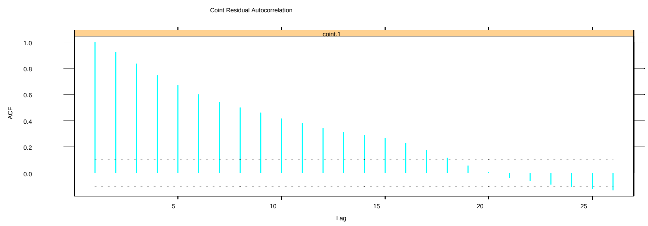




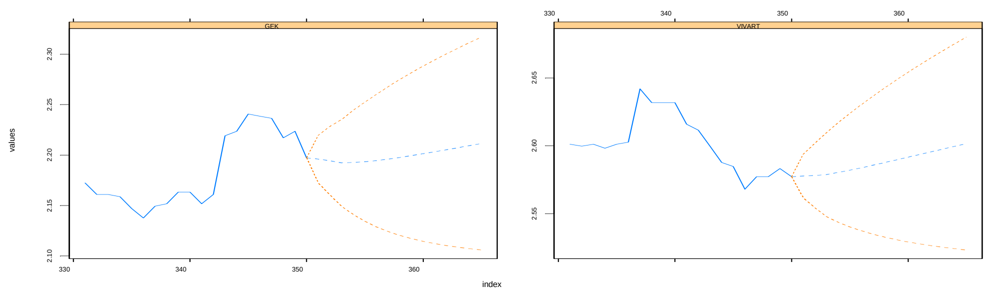
The next figure is referred to the co-integrated residuals. As we see the process looks stationary. We may test it this applying the stationary and unit root tests described in *chapter 2,* or alternative by looking at the ACF plot.



ACF



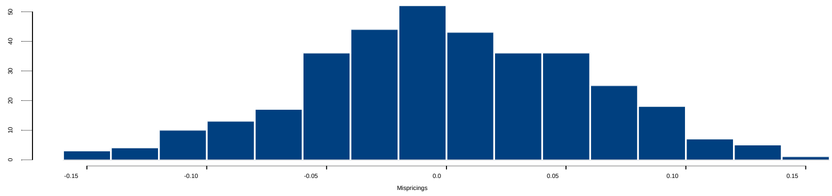
Finally the next plot represents the 15-step predicted values for each stock. Notice that the predicted values obtained by the VECM are slightly different than those obtained by the VAR model before.



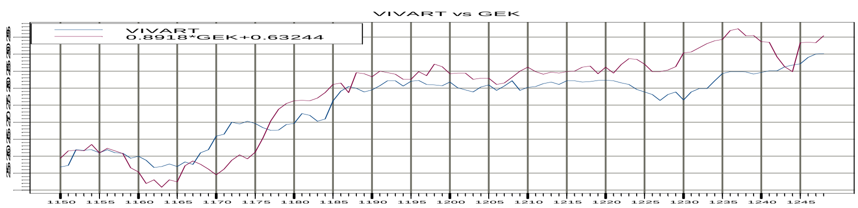
TRADING STRATEGY

Since we have strong evidence that the two time-series variables are co-integrated, we can apply the trading strategy which we have discussed in *chapter 2* trying to exploit any significant deviation from the long-run equilibrium. Up to now, we have

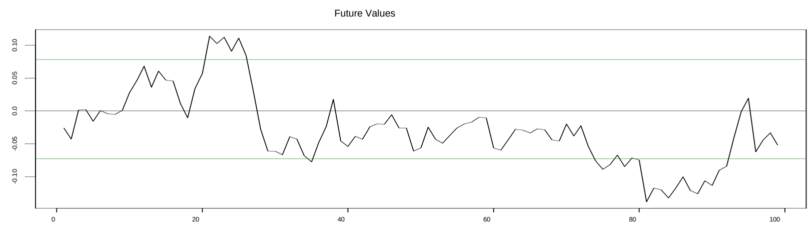
considered as a trading signal, a deviation equal to one standard error of the estimated residuals. The main problem is that in previous examples we have not taken into account the transaction cost. Clearly an active trading strategy implies a high transaction cost, which should be avoided it. Hence we need wider bounds for the trading signals. Also it is very important to take into account any asymmetry of the estimated distribution of the 'mispricings' i.e. the estimated residuals, since the distribution of them is often skewed. Thus we may consider as trading signals say the 20% and 80% or the 10% and 90% quantiles of the empirical distribution of the co-integrated residuals.
Below we represent the histogram of the co-integrating residuals after subtracting the mean. The 10% and the 90% quantiles are *-0.0727* and *0.0782* respectively.



The next figures represent the 'mispricing' of the portfolio for the next 99 trading days. The 'green' horizontal lines represent the trading signals.



Mispricing



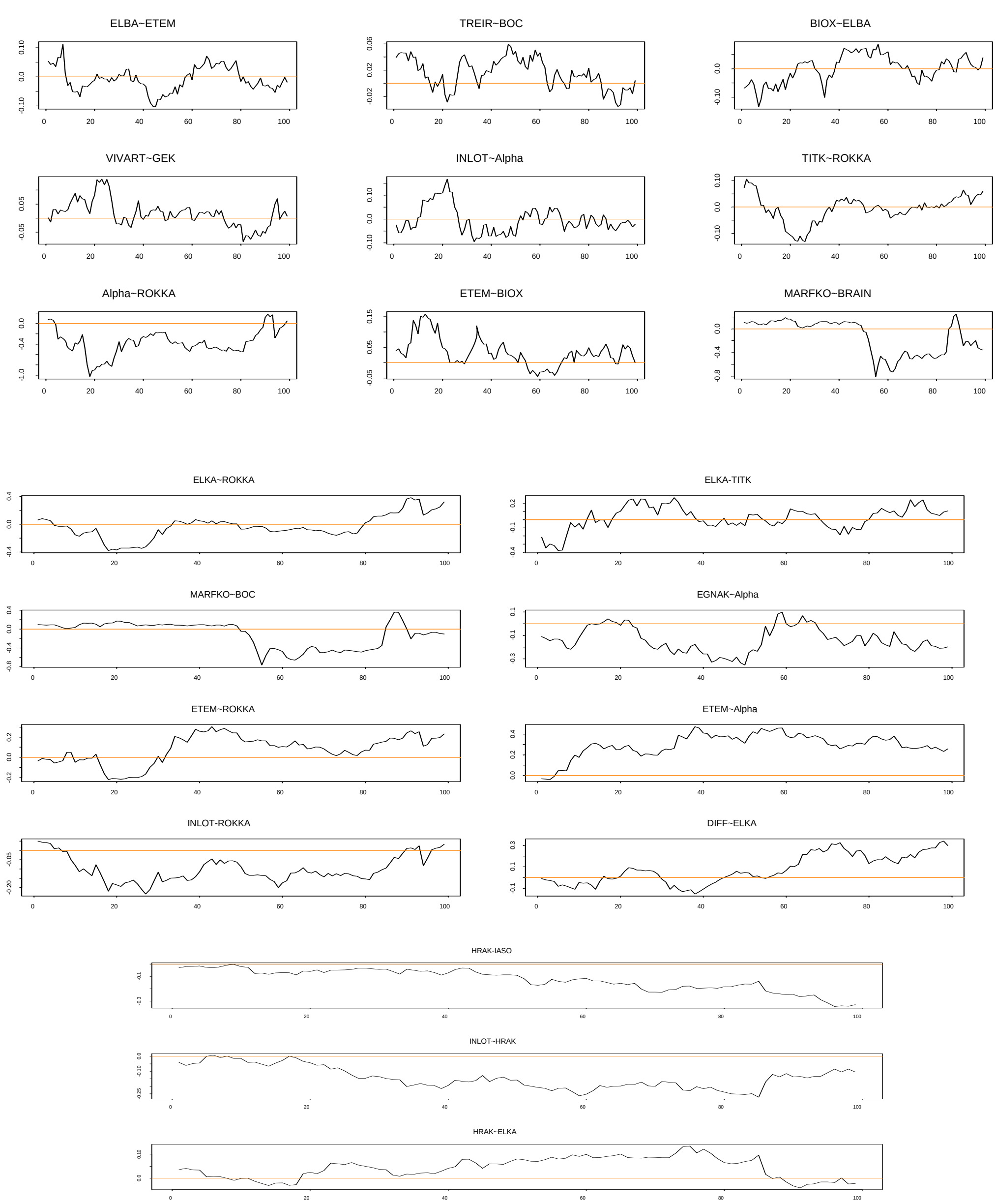
14/03/2007: Short-Sell the spread at 6.664€
22/03/2007: Close the position paying back 6.043€
03/04/2007: Buy the spread at 5.03184€
10/04/2007: Sell back the spread at 6.53184€
01/06/2007: Buy the spread at 5.45466€
02/07/2007: Sell back the spread at 6.8984€

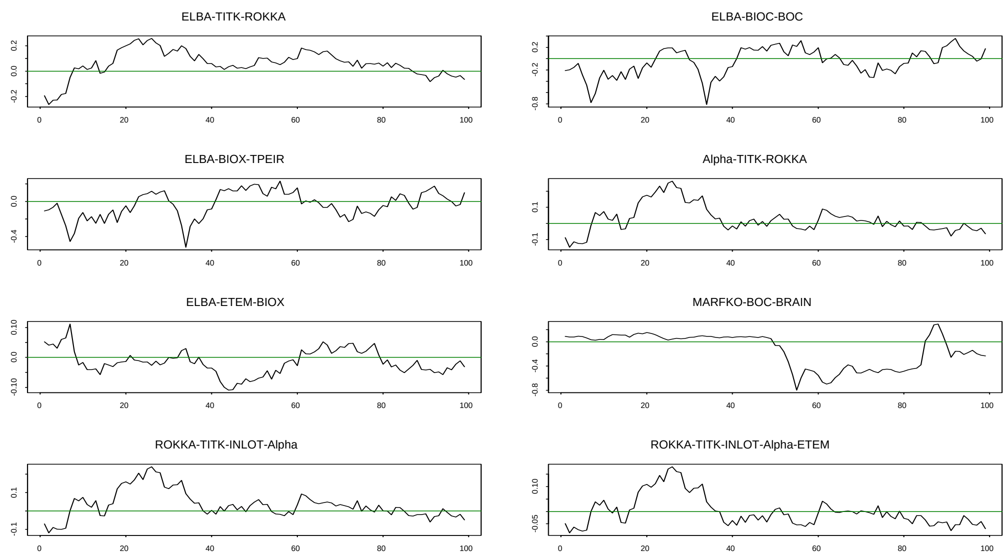
Applying this strategy we calculated that the net profit was equal to **3.565€**. We derive to this result without calculate the transaction cost. If we assume that the transaction cost is equal to 1% of the price of the stock (this is a true value for the '*ASE*') then the net profit becomes equal to **2.033€**. Notice that in 99 days we made 12 trades and as a result the transaction cost was quite high.

5.1 Numerical Results

At this point we represent briefly the numerical results that we obtained applying the analysis we have discussed up now. The co-integration coefficient in each case is obtained by applying the Johansen's Methodology, and we fitted the models taking 350 daily closing stock prices from *Athens Stock Exchange* (21/09/05-12/02/07). Then we tested the model for the next 99 trading days (13/02/07-06/07/07). As trading signals in bivariate case we considered the 3% and 97% quantiles of the co-integrating residuals, in trivariate case the 2% and 98% quantiles and in higher dimensions the 1% and 99% quantiles of the co-integrating residuals. The next plots illustrate the portfolios that we applied the strategy, where the horizontal orange line is referred to the equilibrium value i.e. 0.

Portfolios





The net profit per portfolio including the transaction cost (1%) is as follows:

1) VIVART-GEK⇒2.03€
2) ELBA-ETEM⇒0.49€
3) BOC-TPEIR⇒**0€ (Does not diverge sufficiently)**
4) BIOX-ELBA⇒2.88€
5) INLOT-Alpha⇒3.18€
6) TITK-ROKKA⇒2.77€
7) Alpha-ROKKA⇒3.24€
8) ETEM-BIOX⇒2.08€
9) MARFKO-BRAIN⇒9.24€
10) INLOT-HRAK⇒**-0.10€(Negative Profit)**
11) IASO-HRAK⇒**-2.99€(Negative Profit)**
12) ELKA-ROKKA⇒5.85€
13) ELKA-TITK⇒13.86€
14) HRAK-ELKA⇒1.05€
15) MARFKO-BOC⇒2.59€
16) EGNAL-Alpha⇒5.94€
17) ETEM-ROKKA⇒3.82€
18) ETEM-Alpha⇒**-2.06€ (Negative Profit)**
19) INLOT-ROKKA⇒2.59€
20) DIFF-ELKA⇒2.21€
21) TITK-ROKKA-Alpha⇒3.5€
22) ELBA-TITK-ROKKA⇒4.89€
23) BOC-BIOX-ELBA⇒8.14€
24) TPEIR-BIOX-ELBA⇒3.53€
25) ELBA-ETEM-BIOX⇒0.60€
26) MARFKO-BOC-BRAIN⇒2.74€
27) ROKKA-TITK-INLOT-Alpha⇒1.04€
28) ROKKA-TITK-INLOT-Alpha-ETEM⇒0.36€

References

Burgess, A.N., 1999, '*A Computational Methodology for Modelling the Dynamics of Statistical Arbitrage*'

Dickey, D. Fuller, W., 1979, '*Distribution of the Estimators for Autoregressive Time Series with a Unit Root*', Econometrica, 49, 1057-1072

Engle, R.F., Granger, C.W.J., *1987, 'Co-Integration and Error Correction: Representation, Estimation and Testing,'*, Econometrica, 55, 251-276

Fama, E., 1965, *'The Behavior of Stock Market Prices'*

Fama, E., 1970, *'Efficient Market Theory'*

Fama, E. , French, K.R., 1992, *'The Cross-Section of Expected Stock Returns'*

Gatev et al 1999, '*Pairs Trading: Performance of a Relative Value Arbitrage Rule*'

Geweke, J., Porter-Hudak, S., (1983), '*The Estimation of Security Price Volatility from Historical Data',* Journal of Time Series Analysis, 4, 221-237.

Hamilton, J., 1994, *'Time Series Analysis'* Princeton University Press

Hurst, H.E. (1951), *'Long Term Storage Capacity of Reservoirs'* Transactions of the American Society of Civil Engineers, 116, 770-799

Johansen, S. (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Johnson and Wichern (1998), '*Factor Analysis using S-PLUS'*

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992, '*Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root*' , Journal of Econometrics, 54, 159-178

Ng, S., Perron, P., 1995, '*Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag*', Journal of the American Statistical Association, 90, 268-281

Philips, P.C.B., 1986, *'Understanding Spurious Regression in Econometrics'*, Journal of Econometrics, 33, 311-340.

Phillips, P.C.B., 1987, '*Time Series Regression with a Unit Root*', Econometrica, 55, 227-301

Phillips, P.C.B., Perron, P., 1988, '*Testing for Unit Roots in Time Series Regression'*, Biometrika, 75, 335-346

Said, S.E., Dickey, D., 1984, 'Testing for Unit Roots in Autoregressive Moving-Average Models with Unknown Order', Biometrika, 71, 599-607

Sheikh, A., 1995, '*BARRA's Risk Models*'

Stock, J.H., Watson, M.W., 1993, '*A Simple Estimator of Co-integrating Vectors in Higher Order Integrated Systems*', Econometrica, 61, 783-820.

Vidyamurthy G., 2004, '*Pairs Trading: Quantitative Methods and Analysis'*

Wong, M.A., 1993, '*Fixed-Income Arbitrage: Analytical Techniques and Strategies'*

Zivot, E., Wang, J.2006, '*Modeling Financial Time Series with S-PLUS*'