DIMACS Technical Report 2004-47
October 2004

Principles of nonstationary regression estimation: A new approach to dynamic multi-factor models in finance

by

Michael Markov

Markov Processes International
428 Springfield Ave., Summit NJ 07901
michael.markov@markovprocesses.com

Vadim Mottl

Scientific Council "Cybernetics"
at the Russian Academy of Sciences
Vavilov St. 40, 117333 Moscow, Russia
vmottl@yahoo.com

Ilya Muchnik

DIMACS
Rutgers University
P.O. Box 8018, Piscataway, NJ 08855
muchnik@dimacs.rutgers.edu

Abstract

A new class of signal analysis problems is considered, which appear in finance when it is required to detect the hidden dynamics of an investment instrument or a portfolio in respect to certain market or economic factors. Such problems can be naturally formulated as a complex of models concerned with estimating a nonstationary linear regression under additional constraints and requirements which have been not considered in the classical methodology of signal analysis. These models are also applicable to many other engineering and scientific problems.

We review existing financial multi-factor models from the standpoint of their performance in detecting hidden investment portfolio dynamics. Using practical examples, we present and analyze the shortcomings of these models in detecting both gradual and rapid changes in investment portfolio structure. We then lay the groundwork for a new approach, which we call *Dynamic Style Analysis* (DSA), representing a true time-series multi-factor portfolio analysis model. At the core of the methodology, we present a new dynamic regression model, which we call *Constrained Flexible Least Squares* (CFLS). One of the most important features of the DSA model is that it is fully adaptive, i.e., all model parameters are determined from data. The major concepts of the new methodology are gradually introduced and applied to analyses of both model portfolios and well-known public US mutual funds. By comparing publicly available holdings data with results obtained with DSA, we demonstrate both the superiority of the new model and its remarkable accuracy in detecting portfolio dynamics. We also address issues such as the computational complexity of DSA and its practical applications in the areas of risk management, performance measurement and investment research. One of the major applications of the new methodology lies in hedge fund due diligence and risk monitoring, where the importance of uncovering and controlling hidden factor dynamics is especially valuable given the limited transparency of hedge funds.

Contents

1 Introduction

There exists a wide class of signal analysis problems in which it is required, for the given signal $Y=(y_t,\, t=1,...,N)$ on the axis of a discrete argument (usually time), to estimate the values of a sufficiently smoothly changing parameter $B=(\beta_t,\, t=1,...,N)$ that takes values from some set $\beta_t \in \mathbb{B}$ and forms the hidden process usually considered as being random.

Given rapid increases in computer power and the emergence of new computational techniques, the process of computerization is being applied to new practical problems for which existing methods are not totally adequate. In the current work, we consider a new class of signal analysis problems in which the hidden process to be estimated can be described as a dynamically changing allocation of some resource over a finite set of positions

$$\boldsymbol{\beta}_t=(\beta_t^{(1)},...,\beta_t^{(n)}),\ \sum\nolimits_{i=1}^{n}\beta_t^{(i)}=1,\ \ \beta_t^{(i)}\geq 0\,. \tag{1}$$

The stated non-negativity of the hidden process prevents using linear algorithms such as the Kalman-Bucy filter and interpolator [1,2,3,4].

A typical example of such a problem is that of determining major market factors affecting performance of an investment portfolio, which plays an important role in modern investment analysis. We consider here a dynamic generalization of this problem, the original static formulation of which belongs to William Sharpe, 1990 Nobel Prize winner in Economics [12].

Multi-factor models play an important practical role in Finance. They are used extensively in both the portfolio construction process and in analysis of existing portfolios. When creating or adjusting their portfolios, money managers have to determine which asset(s)[1] (stocks, bonds, currencies, etc.) and in what quantities they have to buy or sell to achieve the desired characteristics of their portfolio. For example, a typical task would be to hedge (neutralize) certain factor sensitivities of the portfolio, while leveraging (magnifying) sensitivity to other factors. Frequently, a multi-factor model is used in the process and the trade list of assets is determined by a multi-factor optimization.

For investors analyzing financial products such as mutual funds or hedge funds, the interest lies primarily in the sensitivity of these products to various market factors, in order to determine what performance to expect from the product in certain market conditions. The latter application is of even greater importance for risk managers of trading institutions such as banks, brokerages and securities trading firms, where the effort of sometimes hundreds of employees is geared toward determining a single confidence interval for a daily potential loss of the bank's investment portfolio. Such an interval is called Value-at-Risk, and the methodology frequently involves multi-factor modeling.

In addition, since factor models can tell what kind of performance was expected from a portfolio given its factor sensitivity in the past, they are extensively used in the area of performance measurement. The portfolio management team's skill assessment is based on such evaluation. Such an approach is also used to understand the source of good or poor performance – whether it was related to buying a certain industry sector or certain assets within the sector. This area of Finance is called performance attribution (where performance is attributed to various factors).

---

[1] An individual asset such as a stock or a bond is frequently called a *security*. By *portfolio* we usually mean an investment in one or more individual securities. The term *financial instrument* usually has a broader meaning than that of a security. When we talk about financial instruments, we refer to investments ranging from individual securities to portfolios, and to complex contracts on such securities called *derivatives*.

It is also important to note that most such real-life models of financial instruments (or portfolios of such instruments) are dynamic in nature. This relates to the fact that either the relationships between factors change over time due to economic or political forces, or the structure of instruments and portfolios change. For example, if managers are engaging in active trading, a portfolio's factor exposures can be changed dramatically within days.

The focus of the current paper is estimation of time-varying parameters of a certain class of practical multi-factor financial models. These models are based on performance data and are often called performance-based or returns-based models, where the observed time series characteristic of the instrument (the signal) is its price[1]. It is common practice in Finance to work not with the price of a security (or the market value of a portfolio) but rather with its relative change over a certain time period (i.e., day, week, month) which is also called investment return or daily return, weekly return or monthly return, respectively[2]. Note that prices of instruments change very frequently throughout the day, and in theory the discussed material is applicable to such intra-day price changes as well.

In the current paper we show limitations of existing methods of modeling dynamic financial instruments and propose a new approach that addresses these limitations. We show that the proposed model represents a special filtering technique, which is applied to financial problems of this kind for the first time. We extend this filter to reflect the practicalities of the investment process and make it fully adaptive – based on the suggested cross-validation technique, the model's parameters are determined from the observed data automatically.

2 An overview of multi-factor models in finance and existing methods of their estimation

2.1 Basic multi-factor models

2.1.1 Capital Asset Pricing Model (CAPM)

In his landmark paper in 1964[5] Sharpe suggested a single-factor model to determine efficient return of a security or any market asset. According to this model, the return of the asset in efficient markets is determined by the formula

$$r^{(a)} = r^{(f)} + \beta(r^{(m)} - r^{(f)}), \tag{2}$$

---

[1] For an investment portfolio consisting of a number of such instruments, its price is often called the *market value*.

[2] Although prices are recorded many times during the day, daily price most of the time refers to closing market price for the day, i.e., the last price at which security was traded. For weekly and monthly prices the closing daily prices are taken on the last day of the week and month, respectively. For example, in order to compute weekly return, the week's Friday's closing price of the instrument is divided by the price of the instrument as recorded on the previous Friday market closing, and then one is subtracted from the ratio.

Because it is commonly accepted that individual securities prices follow a geometric Markov process, very often, the differences of price logarithms, rather than returns, are considered. Paul Samuelson, a Nobel Prize laureate in Economics, was the first to suggest that securities prices have to be modeled by a log-normal process. Refer to Hull [6] for a discussion of log-normal nature of securities returns. The methodology described in the current paper is directly applicable to such logarithmic transformations of prices.

where $r^{(a)}, r^{(f)}$ and $r^{(m)}$ represent periodic returns on the asset, a risk-free asset and the market portfolio, respectively. This approach laid the foundation of the Capital Asset Price Model, which is attributed to Sharpe, Lintner and Mossin, who independently developed the concept. Single index models of asset returns were suggested even earlier in the 1950's. Harry Markowitz suggested a similar single index model several years earlier [7]. The focus of the Sharpe-Lintner-Mossin CAPM theory was much broader, and included the concept of market equilibrium and efficient prices of assets.

The market portfolio here is an abstract that is defined as a portfolio containing all traded securities with their weights in the portfolio proportional to their respective capitalizations[1]. A broad index such as the S&P 500 Index or the Russell 3000 Index is usually being used as a proxy for the equity market portfolio[2]. The risk-free asset in CAPM is another abstract, which has guaranteed non-negative return and can be thought of as a bank deposit in a 100% insured bank. A typical proxy for the risk-free asset is the 90-day US Treasury Bill that is regularly auctioned by the U.S. Government. For individual stocks, the regression coefficient β is an important indicator of the stock risk which measures its sensitivity to the market movements. In other words, it provides important information about what behavior to expect from the stock when the market falls or rises. Stocks with $\beta > 1$, such as, for example, airline companies, are more sensitive to market moves. Utility companies are less sensitive to market moves, and their betas are usually less than one.

The CAPM model was also applied to portfolio evaluation, since a portfolio can be thought of as a linear combination of individual assets in (2). In the portfolio evaluation framework, the single asset CAPM model can be rewritten as follows:

$$r^{(p)} - r^{(f)} \cong \alpha + \beta(r^{(m)} - r^{(f)}), \qquad (3)$$

where $r^{(p)}$ represents the return of a portfolio. In application of CAPM to investment portfolios, the regression intercept term in (2), also called *Jensen's alpha*, is used as a measure of the value added by actively selecting individual securities for the individual portfolio vs. a passive investment in a combination of the market portfolio (m) and the risk-free asset (f). The regression coefficient β measures the sensitivity of the portfolio to market moves and is usually described as a measure of a portfolio's risk. To estimate single index betas in (3), an *Ordinary Least Squares* (OLS) is usually applied to daily or monthly return observations. Such beta estimates have very significant practical value because they convey the level of the asset risk to investors.

The dynamic nature of beta was recognized, and several methods for adjusting betas were suggested, including the most popular ones by Vasicek [8] and Blume [9]. The work by Wells [4] contains an extensive comparison of methods of estimation of time-varying betas and also compares various single-factor filtering techniques commonly used in econometrics applied to estimation of individual stock betas.

Later, multi-factor CAPM models appeared that recognized the fact that both individual securities and portfolios are multi-factor in nature. One of the first works that measured the effect of industry factors on securities prices was done by King [10]. The multi-factor CAPM can be represented by the equation

---

[1] Capitalization of a company is defined as the stock price multiplied by the total number of shares on the market.

[2] These two indices represent the largest 500 and 3,000 US stocks, respectively, based on their capitalization. A special committee includes smaller stocks in the S&P 500 Index to achieve better representation of all major economic sectors. Prices of these indices (also called index values) are quoted daily in financial media.

$$r^{(p)} - r^{(f)} \cong \alpha + \beta^{(1)}(r^{(1)} - r^{(f)}) + \beta^{(2)}(r^{(2)} - r^{(f)}) + \ldots + \beta^{(n)}(r^{(n)} - r^{(f)}), \tag{4}$$

where $r^{(p)}$ is the portfolio investment return, $^{(i)}$ are returns on the market portfolio as well as changes in other factors (e.g., industry sectors,) and $^{(f)}$ is return on a risk-free instrument.

2.1.2 Arbitrage Pricing Theory (APT)

One of the multi-factor models most widely used in research and in practice is the APT model described in [11]:

$$r^{(p)} \cong \alpha + \beta^{(1)} I^{(1)} + \beta^{(2)} I^{(2)} + \ldots + \beta^{(n)} I^{(n)}, \tag{5}$$

where the factors $I^{(i)}$ are typically chosen to be the major external economic factors that influence asset returns, such as industrial production, inflation, interest rates, business cycle, etc. Such models attempt to define what securities prices to expect given certain economic conditions. The same model can also be applied to an investment portfolio and help investors to determine factor sensitivity of the entire portfolio rather than individual security.

Coefficients $\beta^{(1)},\ldots,\beta^{(n)}$ in the CAPM (4) and APT (5) models are called *factor exposures*. Along with the constant α, the factor exposures make the vector of model parameters $(\alpha,\beta^{(1)},\ldots,\beta^{(n)})$, which is typically estimated by applying a linear regression technique to the time series of security/portfolio returns $r_t^{(p)}$ and economic and market factors $r_t^{(i)}$ or $I_t^{(i)}$ over a certain estimation window $t = 1,\ldots,N$ of time intervals, typically months or quarters:

$$(\hat{\alpha},\hat{\beta}^{(1)},\ldots,\hat{\beta}^{(n)}) = \underset{\alpha,\beta^{(1)},\ldots,\beta^{(n)}}{\arg\min} \sum_{t=1}^{N} (r_t^{(p)} - \alpha - \beta^{(1)} I_t^{(1)} - \ldots - \beta^{(n)} I_t^{(n)})^2 . \tag{6}$$

Although both APT and CAPM may seem similar in the multi-factor approach to explaining asset returns, in the CAPM the factors represent in most cases investible assets such as, for example, market indices, while in the APT approach the factors can be both the market and external factors – anything that influences asset prices.

2.1.3 Returns-Based Style Analysis (RBSA)

One of the most effective and practical multi-factor models for analyses of investment portfolios, called the Returns-Based Style Analysis (RBSA), was suggested by Sharpe in [12] and [13]. In the RBSA model, the periodic return $r^{(p)}$ of a portfolio consisting of n kinds of assets is approximately represented by a linear combination of single factors $(r^{(1)},\ldots,r^{(n)})$ whose role is played by periodic returns of generic market indices for the respective classes of assets. To enhance the quality of parameter estimation, a set of linear constraints is added to the basic equation:

$$\begin{cases} r^{(p)} \cong \alpha + \beta^{(1)} r^{(1)} + \beta^{(2)} r^{(2)} + \ldots + \beta^{(n)} r^{(n)}, \\ \sum_{i=1}^{n} \beta^{(i)} = 1,\ \beta^{(i)} \ge 0,\ i = 1,\ldots,n. \end{cases} \tag{7}$$

In such a model, $r^{(i)}$, $i = 1,\ldots,n$, represent periodic returns (for example, daily, weekly or monthly) of generic market indices such as bonds, equities, economic sectors, country indices, currencies, etc. For example, as described in [13], twelve such generic asset indices are used to represent possible areas of investment. The periodic portfolio return $r_t^{(p)}$ for a period is approximated by the return of a portfolio consisting of n assets which is theoretically equal to the linear combination of periodic returns of generic market indices for the respective classes of assets $(r^{(1)},\ldots,r^{(n)})$ with coefficients $(\beta^{(1)},\ldots,\beta^{(n)})$ having the meaning of the fractions invested in each asset class at the beginning of the period. The fact that the sum of all coefficients in the RBSA model is equal to one is equivalent to having a fully invested portfolio[1].

The non-negativity constraint in the RBSA reflects a very important practical requirement. As it was noted in Markov [14], the RBSA model can be thought of as a *practical extension* of the single-factor CAPM model developed for individual securities. To demonstrate this fact, let's rewrite the CAPM model (3) as follows:

$$r^{(p)} \cong \alpha + (1-\beta) r^{(f)} + \beta r^{(m)} = \alpha + r^{(p')}, \tag{8}$$

where

$$r^{(p')} = (1-\beta) r^{(f)} + \beta r^{(m)}. \tag{9}$$

The return on the given portfolio $r^{(p)}$ is compared with the performance of a portfolio (p') – a benchmark, consisting of the market portfolio and a risk-free asset. Although such a framework for evaluating performance of active investment managers has been used in practice and academic research for 40 years, its limitations are immediately apparent: (a) it allows for unlimited borrowing[2] while the majority of institutional money managers and most mutual funds are allowed to hold long (non-negative) positions only, and (b) it doesn't take into consideration valuable prior information about the portfolio manager's investment options by effectively limiting its benchmark portfolio P' to hold a single market portfolio only. Given these considerations, any use of both *alpha* and *beta* in practice is clearly of limited value.

As the first step in his departure from the classic CAPM, Sharpe constrained the portfolio beta to include values between zero and one, thus making the model more practical and addressing the unlimited borrowing of assets issue cited above. Next, to address the problem with the model being less risky as the result of the first step ($\beta \le 1$), he allowed for inclusion into the model many assets representing various levels of risk. Although multi-factor models had been applied to portfolio analysis before, and CAPM with restricted borrowing models had been studied as well, Sharpe's approach combined both in an elegant and intuitive way by using prior information about the portfolio in parameter estimation.

To estimate the parameters of equation (7), Sharpe used the *Constrained Least Squares Technique*, i.e., the parameters are found by solving the constrained quadratic optimization problem in a window of $t = 1,...,N$ time periods in contrast to the unconstrained one (6):

$$\begin{cases} (\hat{\alpha}, \hat{\beta}^{(1)}, ..., \hat{\beta}^{(n)}) = \underset{\alpha, \beta^{(1)}, ..., \beta^{(n)}}{\arg\min} \sum_{t=1}^{N} (r_t^{(p)} - \alpha - \beta^{(1)} r_t^{(1)} - ... - \beta^{(n)} r_t^{(n)})^2, \\ \text{subject to } \sum_{i=1}^{n} \beta^{(i)} = 1, \ \beta^{(i)} \ge 0, \ i = 1,...,n. \end{cases} \tag{10}$$

Model parameters $(\alpha, \beta^{(1)}, ..., \beta^{(n)})$ estimated using unconstrained (6) and constrained least squares techniques (10) represent average factor exposures in the estimation window – time in-

---

[1] The periodic return of a portfolio is equal to the linear combinations of periodic returns of the portfolio's assets for the same period with weights taken at the beginning of the period $r^{(p)} = \sum_{i=1}^{n} \beta^{(i)} r^{(i)}$ under the assumption that the portfolio hasn't been traded during the period (i.e., shares invested in each asset remain unchanged). This simple fact is demonstrated in Appendix 1. Such linear constraint is often called the *budget constraint*, reflecting the fact that the entire budget is spent on the investment in the model's assets.

[2] Unlimited borrowing here means negative weight of the risk-free asset $1-\beta$ in the CAPM model corresponding to values of $\beta > 1$.

terval $t = 1,...,N$. However, the factor exposures typically change in time. For example, an active trading of a portfolio of securities can lead to significant changes in its exposures to market indices within the interval. Detecting such dynamic changes, even though they happened in the past, represents a very important task.

2.2 The moving window technique

The multi-factor RBSA model (7), as well as the CAPM (4) and APT (5), are, in their essence, linear regression models with constant regression coefficients $(\alpha, \beta^{(1)}, ..., \beta^{(n)})$.

In order to estimate dynamic changes in factor exposures, a moving window technique is typically applied. For example, in RBSA model (7), the exposures at any moment of time t are determined on the basis of solving (10) using a window of K portfolio returns $\left[t-(K-1), ..., t\right]$ and the returns on asset class indices over the same time period, as described, for example, in [13]:

$$\begin{cases} (\hat{\alpha}_t, \hat{\beta}_t^{(1)}, ..., \hat{\beta}_t^{(n)}) = \arg\min\limits_{\alpha, \beta^{(1)}, ..., \beta^{(n)}} \sum\limits_{\tau=0}^{K-1} (r_{t-\tau}^{(p)} - \alpha - \beta^{(1)} r_{t-\tau}^{(1)} - ... - \beta^{(n)} r_{t-\tau}^{(n)})^2, \\ \text{subject to } \sum\limits_{i=1}^{n} \beta^{(i)} = 1, \ \beta^{(i)} \geq 0, \ i = 1, ..., n, \end{cases} \tag{11}$$

By moving the estimation window forward period-by-period, dynamic changes in factor exposures can be approximately estimated.

The moving window technique described above has its limitations and deficiencies. The problem setup assumes that exposures are constant within the window, yet it is used to estimate their changes. Reliable estimates of model parameters can be obtained only if the window is sufficiently large, which makes it impossible to detect changes that occurred within a day or a month, and therefore, such a technique can be applied only in cases where parameters do not show marked changes within the estimation window: $(\alpha_s, \beta_s^{(1)}, ..., \beta_s^{(n)}) \cong \text{const}$, $t-(K-1) \leq s \leq t$. In addition, such an approach fails to identify very quick, abrupt changes in investment portfolio exposures that can occur due to trading.

To illustrate the effectiveness of the moving-window method in analyzing active portfolios, a simple two-asset model was constructed made up of US stocks and bonds represented by the S&P 500 Index and the Lehman Aggregate Bond Index, respectively. An initial 50/50 percentage mix was changed over time in a sine wave pattern as shown in Figure 1, and composite weighted returns were calculated in each monthly data point based on monthly returns of the two indices. The annual turnover of this two-asset portfolio was set at 50%, which means that in every year 50% of the portfolio was traded and the change from a 100% stock portfolio to a 100% bond portfolio occurred in a two-year term. Such turnover is considered reasonable for an average portfolio.

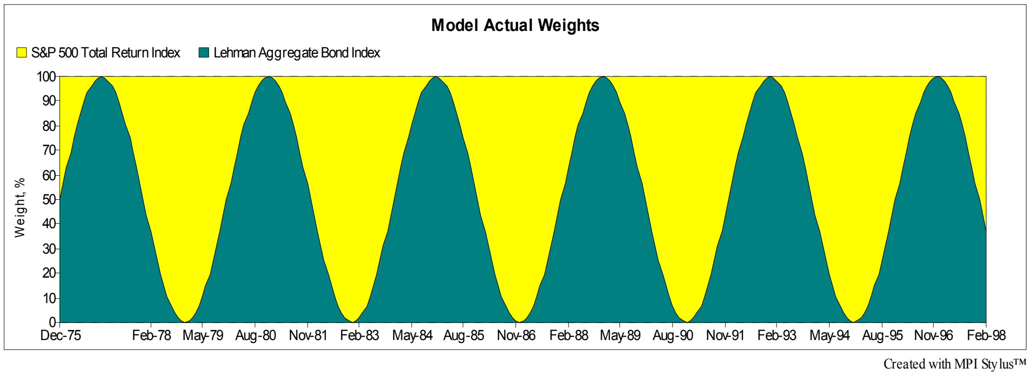


Figure 1. Two-asset model portfolio.

The composite return of the two-asset portfolio above

$$r_t^{(p)} = \beta^{(1)} r_t^{(1)} + \beta^{(2)} r_t^{(2)} \tag{12}$$

was then used as input to the RBSA model (11). Both indices that had been used to construct the model portfolio were used as the only assets in the model. The chart in Figure 2 shows the results of estimation using the trailing 24-month moving window method[1].

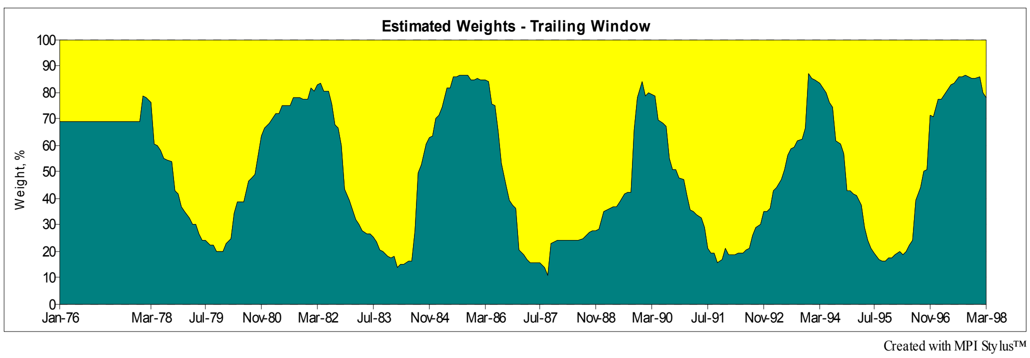


Figure 2. 24-month trailing window regression.

2.3 Exponential weighting

The moving window technique described above has its limitations and deficiencies. The problem setup assumes that exposures are constant within the window, yet it is used to estimate their changes. Reliable estimates of model parameters can be obtained only if the window is sufficiently large which makes it impossible to sense changes that occurred within a day or a month, and, therefore, such a technique can be applied only in cases where parameters do not show marked changes within it: $(\alpha_s, \beta_s^{(1)}, ..., \beta_s^{(n)}) \cong \text{const}$, $t-(K-1) \le s \le t$. In addition, such an approach fails to identify very quick, abrupt changes in investment portfolio exposures that can occur due to trading.

An attempt to alleviate the problem with the moving window technique when applying it to the model (11) was made by Sharpe in [15]. He suggested assigning weights to the historical returns used in constrained regression, so that the more recent returns had a larger effect on the effective asset mix. According to this method, the weights would gradually decrease within the es-

[1] Practitioners in such analyses typically use 36 or 24 monthly return observations, representing three or two years of performance history respectively.

timation window from the point of estimate (being the rightmost or the most recent data point) to the end of the window. The weight would decrease with "exponential decay", i.e., each weight would be equal to the preceding weight times a certain number $0 < \delta < 1$, usually called a decay factor. Though this method somewhat decreases the delay in detecting changes in exposure, the improvement is small and it is largely neutralized by an increase in the noise – random errors in evaluating the effective asset mix. Also, the same contradiction remained: when performing estimation, the method assumed that the effective mix was constant within the window.

2.4 Locally Weighted Regression (LWR) approach

A better approach to deal with the problem of static coefficients in solving problem (11) was described in [16]. According to this methodology, the regression window that is used to estimate the effective asset mix was formed in each point of estimate based on the k-nearest neighbor rule. The method recognizes the fact that, for the purpose of estimating exposures at any such point in the past, observations on both sides of the point of estimate are equally important (if there's no other prior information). Therefore, the regression window would always include k observations that are closest in time to the point of estimate. Each data point within the window is assigned a weight decreasing exponentially from the estimation point to both edges of the window. For the most part, such a window would be centered around the estimation point. In the last available point all k nearest returns are the returns that immediately precede the point of estimate.

This is the only point where the effective asset mix estimated by this method coincides with the effective asset mix determined using Sharpe's method, assuming, of course, that weights are assigned in a similar manner. In all other points the result is different because, according to the proposed method, the k returns include both returns that precede and returns that follow the point of estimate. The maximum accuracy of the proposed method is achieved in the points where approximately half of the k returns precede and the rest follow the point of estimate. In this case the window is "centered" around the point of estimate. The combination of the k-nearest neighbor algorithm with weights substantially improves the ability of the model to suppress noise and alleviate the lag in sensing changes in exposures.

The method described above represents a class of local estimation methods and is called *Locally Weighed Regression*. For overview of such methods, see [17] and [18]. In our case, the "locality" of the model is time, and an exponent is used as the weight-function. In a more general case, different weight functions such as Gaussian can be used, depending on the nature of the relationship of observations in adjacent points.

The results of the analysis of the two-asset model portfolio (12) from the previous Section are shown in Figure 3. Note that the lag in exposure detection disappeared, and the results are much closer to the underlying weights than the ones produced by Sharpe's method. The decay factor 0.86 we applied is the typical exponential decay that is used for analyzing (for example) mutual funds. A higher decay would lead to a better result in this specific two-asset case.

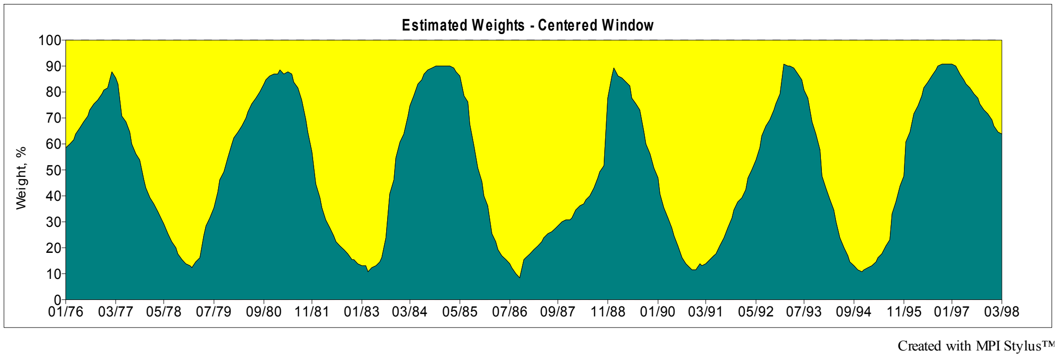


Figure 3. 4-month Locally Weighted Regression; decay factor = 0.86.

2.5 Limitations in hedge fund analyses

As was mentioned earlier, adding non-negativity constraints in the RBSA model (7) addresses practical prior information, reflecting the fact that most of the investment managers don't take negative positions[1] in assets, is crucial and two-fold: they represent valuable prior information about parameter distribution which, in turn, also ensures stability of the dynamic effective mix when applied in the moving window approach. In the analysis of long-short strategies and most hedge funds, where money managers engage in significant leverage and often take short positions across all assets, the non-negativity bounds have to be either removed or relaxed. Typically, when the constraints are removed and the method is reduced to series of rolling ordinary regressions, the effective mix shows unexplainable swings in exposures from large negative position in one period to large positive position in the next period.

The limitations of the existing RBSA model described in this Section are often blamed on noise in the data. It should be clear by now that it's not the noise in the data that is responsible for poor estimation results, but the model itself. The moving window regression technique is just a trick, while a true time series estimation model is required for the portfolio exposure estimation. It is also desirable for such a model to incorporate some prior information about the portfolio structure, other than constraints on parameters. This will make such a model applicable to analysis of portfolios with short positions, such as hedge funds.

3 The general dynamic approach to investment portfolio analysis

3.1 The dynamic multi-factor model

For monitoring a portfolio for quick changes in investment allocation or investment style, deviations from investment mandate, etc., a time-varying multi-factor model, in particular, the dynamic RBSA model, is needed to represent the time series of portfolio returns. This can be achieved, for instance, by considering $r_t^{(p)}$ in (7), as dynamically changing linear combination of

[1] Negative positions are also called "short" positions. To engage in a short position in an asset, an investor borrows the asset from a broker and sells it in the market. Because the investor has to buy back the asset and return it to the broker, the amount of investment in the asset has a negative sign attached to it: if the asset increases in value, the investor loses. If, on contrary, the asset decreases in value, the investor's return will be positive in this investment.

a finite number n of time series of basic factors $\mathbf{r}_t = (r_t^{(1)}, ..., r_t^{(n)})^T$ with unknown real-valued factor exposures $\boldsymbol{\beta}_t = (\beta_t^{(1)}, ..., \beta_t^{(n)})^T$ and an unknown intercept term α_t. However, in the RBSA model, both factor exposures and intercepts are subject to appropriate constraints $(\alpha_t, \boldsymbol{\beta}_t) \in Z_t$, in the simplest case, the linear ones $\sum_{i=1}^n \beta_t^{(i)} = 1$, $\beta_t^{(i)} \geq 0$:

$$\begin{cases} r_t^{(p)} = \alpha_t + \sum_{i=1}^{n} \beta_t^{(i)} r_t^{(i)} + e_t = \alpha_t + \boldsymbol{\beta}_t^T \mathbf{r}_t + e_t, \\ (\alpha_t, \boldsymbol{\beta}_t) \in Z_t, \end{cases} \quad (13)$$

where e_t is the residual model inaccuracy treated as white noise[1].

It is enough to consider the simplified regression model without the intercept term

$$\begin{cases} r_t^{(p)} = \sum_{i=1}^{n+1} \beta_t^{(i)} r_t^{(i)} + e_t = \boldsymbol{\beta}_t^T \mathbf{r}_t + e_t, \\ \boldsymbol{\beta}_t \in Z_t, \end{cases} \quad (14)$$

in which α_t is incorporated into the regression coefficient vector and, respectively, the vector of single factors is extended by an additional element equal to unity:

$$\begin{cases} \boldsymbol{\beta}_t = (\beta_t^{(1)}, ..., \beta_t^{(n)}, \alpha_t)^T = (\beta_t^{(1)}, ..., \beta_t^{(n)}, \beta_t^{(n+1)})^T, \\ \mathbf{r}_t = (r_t^{(1)}, ..., r_t^{(n)}, 1)^T = (r_t^{(1)}, ..., r_t^{(n)}, r_t^{(n+1)})^T. \end{cases} \quad (15)$$

The key to the new approach in estimating parameters of a time-varying multi-factor model lies in understanding its multi-criteria nature. On the one hand, the goal is to create a portfolio of generic factors that closely tracks the analyzed portfolio returns, what can be expressed as the intent to maximize or minimize some numerical measure of consistency $\Phi_{cons}(R, B) \to \max(B)$ or, respectively, inconsistency $\Phi_{incons}(R, B) \to \min(B)$ of the factor exposures $B = (\boldsymbol{\beta}_t, t \in T)$, where T is the observation interval, with the observed portfolio returns $R = (r_t^{(p)},\ t \in T)$.

On the other hand, there always exists some *a priori* information on the expected dynamics of the factor exposures $B = (\boldsymbol{\beta}_t, t \in T)$, which can be expressed as a numerical measure quantitatively assessing the likelihood $\Psi_{like}(B)$ or unlikelihood $\Psi_{unlike}(B)$ of any supposed path B, so that the decision should be made by the condition $\Psi_{like}(B) \to \max(B)$ or, $\Psi_{unlike}(B) \to \min(B)$, respectively. The constraints of a much more general kind than $\boldsymbol{\beta}_t \in Z_t$ can be incorporated into the likelihood or unlikelihood measure as

$$\left. \begin{array}{r} \Psi_{like}(B) = \text{very small value} \\ \text{or} \ \ \Psi_{unlike}(B) = \text{very large value} \end{array} \right\} \ \text{for inadmissible realizations.}$$

Thus, this approach can be formulated as the multi-criteria optimization problem – maximizing (minimizing) the consistency (inconsistency) of the factor exposures dynamics with the observations and, at the same time, maximizing (minimizing) the *a priori* likelihood (unlikelihood) of it:

---

[1] If periodic returns of assets or asset classes $\mathbf{r}_t = (r_t^{(1)}, ..., r_t^{(n)})^T$ for the succession of time periods $t = 1, 2, 3, ...$ are taken as the basic factors, the factor exposures $\beta_t^{(i)}$ have the sense of dynamic allocation of the portfolio cost over these assets or asset classes, computed at the beginning of each period t.

$$\left.\begin{array}{l}\Phi_{cons}(R,B)\\ \Psi_{like}(B)\end{array}\right\}\to\max(B)\quad\text{or}\quad\left.\begin{array}{l}\Phi_{incons}(R,B)\\ \Psi_{unlike}(B)\end{array}\right\}\to\min(B)\,. \tag{16}$$

The main departure from the traditional approach is that the solution of this model containing $N(n+1)$ unknowns has to be obtained at once. In contrast to the moving window approach, no assumptions are made here on the constancy or locality of the solution.

However, the multi-criteria optimization is a problematic question in the optimization theory [19]. Actually, the notion of the Pareto optimal set is the only mathematically sound idea in solving that challenge, but it only excludes unacceptable versions without specifying the best decision. All other methods boil down to forming a linear combination of the given criteria and solving the resulting single-criterion problem. The question is how to choose the weights at each of the criteria when assembling the combined one.

In the next Section, we propose a probabilistic view on the problem of investment portfolio analysis that gives a sound concept for balancing both parts of the combined criterion (16), and, what is especially important, that concept is formulated in the commonly adopted mathematical language of the probability theory and statistical decision-making.

3.2 The problem of investment portfolio analysis as that of estimating the hidden component of a two-component random process

We shall consider the unknown path of factor exposures $B=(\beta_t, t\in T)$ as a realization of an unobservable random process. The information on the difference between "likely" and "hardly likely" paths, that is available *before* the time series of portfolio returns $R=(r_t^{(p)},\ t\in T)$ is observed, will be expressed in the mathematical form of an *a priori* probability density $\Psi(B)$, $B\in\mathcal{B}$, where $\mathcal{B}$ is the set of all successions $B=(\boldsymbol{\beta}_t,\ t\in T)$. To specify the *a priori* constraints it is enough to take the density $\Psi(B)$ which is positive only inside the admissible area $B\in\mathcal{Z}\subseteq\mathcal{B}$ and equals zero out of it:

$$\Psi(B)\begin{cases}>0,\ B\in\mathcal{Z},\\ =0,\ B\notin\mathcal{Z},\end{cases}\qquad \int\limits_{B\in\mathcal{Z}}\Psi(B)dB=\int\limits_{B\in\mathcal{B}}\Psi(B)dB=1\,. \tag{17}$$

Analogously, the model of forming the values of portfolio returns $R=(r_t^{(p)},\ t\in T)$ for the given succession of factor exposures $B=(\beta_t, t\in T)$ is assumed to be known in the form of a conditional probability density $\Phi(R\,|\,B)$ over the set $Y\in\mathcal{Y}$ of all numerical successions $R=(r_t^{(p)},\ t\in T)$:

$$\Phi(R\,|\,B)\geq 0,\ R\in\mathcal{R}\,,\qquad \int\limits_{R\in\mathcal{R}}\Phi(R\,|\,B)dR=1\ \text{ for all }\ B\in\mathcal{B}\,. \tag{18}$$

We shall call the densities $\left[\Psi(B),\Phi(R\,|\,B)\right]$ considered jointly the probabilistic portfolio model, as they express the available knowledge on the variety and likelihood of possible changes in the portfolio composition and its influence on the portfolio return.

So far, the distinction between the portfolio model (17) and (18) and the general multi-criteria approach (16) is no more than a terminological one, because the densities $\Psi(B)$ and $\Phi(R\,|\,B)$ are, by their nature, the likelihood measures $\Psi_{like}(B)$ and $\Phi_{cons}(R,B)$. If considered separately, they lead to the left version of the condition (16).

The main idea of this work is to fuse the two single criteria into a combined one on the mathematically strict statistical basis of treating $B=(\boldsymbol{\beta}_t, t\in T)$ and $R=(r_t^{(p)},\ t\in T)$ jointly as a two-component random process (B,R) completely defined by probability densities (17) and (18). Only the component Y is assumed to be observed, and it is required to find the best estimation operator $\hat{B}(R):\mathscr{R}\to\mathscr{B}$ which provides the minimum error risk, i.e. minimum average distinction between the actual hidden value B and its estimate $\hat{B}(R)$ measured by the given loss function $\eta(B,\hat{B})$.

For each estimation operator $\hat{B}(R)$, the probabilistic model (17) and (18) completely determines the random variable $\eta[B,\hat{B}(R)]$. In the theory of statistical decision making, the mathematical expectation of the loss function is called the average error risk [35]. Let $\mathscr{E}$ be the set of all operators $\mathscr{E}=\left\{\hat{B}(\cdot)\right\}=\left\{[\hat{B}(R),\ R\in\mathscr{R}]\right\}$, then the error risk can be defined as a function on this set:

$$r[\hat{B}(\cdot)]=E\left\{\lambda[B,\hat{B}(\cdot)]\right\}=\int\limits_{R\in\mathscr{R}}\int\limits_{B\in\mathscr{B}}\eta[B,\hat{B}(R)]\Psi(B)\Phi(R\,|\,B)dBdR,\ \ \hat{B}(\cdot)\in\mathscr{E}. \tag{19}$$

It appears natural to choose the so-called Bayesian decision; namely, the estimation operator that affords the minimum risk:

$$\hat{B}^*(\cdot)=\underset{\hat{B}(\cdot)\in\mathscr{E}}{\arg\min}\, r[\hat{B}(\cdot)]. \tag{20}$$

Once this operator is found, it specifies the estimate of the hidden factor exposures dynamics for any observed succession of portfolio returns $\hat{B}^*(R)$, $R\in\mathscr{R}$. It remains only to solve the optimization problem (19-20) for the given model (17-18) and loss function $\eta(B,\hat{B})$.

The product $\Psi(B)\Phi(R\,|\,B)$ in (19) is the joint probability density over the set of all realizations of the two-component random process (B,R). This density can be represented in two equivalent forms

$$Q(B,R)=\Psi(B)\Phi(R\,|\,B)=F(R)P(B\,|\,R), \tag{21}$$

where $F(R)$ and $P(B\,|\,R)$ are, respectively, the marginal probability density over the set $\mathscr{R}$ of all realizations of the observable portfolio returns $R=(r_t^{(p)},\ t\in T)$ and the *a posteriori* probability density over the set $\mathscr{B}$ of all versions of the hidden factor exposures paths $B=(\boldsymbol{\beta}_t, t\in T)$:

$$F(R)=\int\limits_{B\in\mathscr{B}}\Psi(B)\Phi(R\,|\,B)dB,\quad P(B\,|\,R)=\frac{\Psi(B)\Phi(R\,|\,B)}{F(R)}. \tag{22}$$

Substitution of the second version of the joint density (21) into (19) gives the equivalent form of the average risk:

$$r[\hat{B}(\cdot)]=\int\limits_{R\in\mathscr{R}}F(R)\left(\int\limits_{B\in\mathscr{B}}\eta[B,\hat{B}(R)]P(B\,|\,R)dB\right)dR.$$

To solve the problem (20) means to assign the best value $\hat{B}^*(R)$ to each $R\in\mathscr{R}$. It is clear that if R is fixed, $F(R)$ is also fixed, and the minimum error risk will be provided by minimizing the inner integral

$$\hat{B}^*(R)=\underset{\hat{B}\in\mathscr{B}}{\arg\min}\int\limits_{B\in\mathscr{B}}\eta(B,\hat{B})P(B\,|\,R)dB. \tag{23}$$

This is the general form of the optimal estimation operator (20) for an arbitrary loss function $\eta(B,\hat{B})$. Further specifications are possible only for particular kinds of loss functions and portfolio models $\left[\Psi(B),\Phi(R\,|\,B)\right]$.

3.3 The structure of the estimation operator for the singular loss function

The structure of the simplest loss function is based on the notion of the Dirac delta-function $\delta(B,\hat{B})$:

$$\eta(B,\hat{B}) = 1 - \delta(B,\hat{B}),$$
$$\delta(B,\hat{B}) = 0 \text{ if } B \neq \hat{B}, \int_{\hat{\mathscr{B}}} \delta(B,\hat{B})dB = 1 \text{ for any set } \hat{\mathscr{B}} \subseteq \mathscr{B} \text{ such that } \hat{B} \in \hat{\mathscr{B}}. \tag{24}$$

We shall call this loss function the *singular loss function*. It corresponds to the “naive” desire to guess the hidden path $B = (\boldsymbol{\beta}_t, t \in T)$ absolutely exactly.

We have for the singular loss function

$$\int_{\mathscr{B}} \eta(B,\hat{B})P(B\,|\,R)dB = \int_{\mathscr{B}} P(B\,|\,R)dB - \int_{\mathscr{B}} \delta(B,\hat{B})P(B\,|\,R)dB = 1 - P(\hat{B}\,|\,R),$$

hence, it follows from (23) that the estimation operator is the optimal one if it maximizes the *a posteriori* probability density:

$$\hat{B}^*(R) = \arg\max_{B\in\mathscr{B}} P(B\,|\,R).$$

In accordance with the Bayes' formula (21), it follows that $P(B\,|\,R) = \Psi(B)\Phi(R\,|\,B)/F(R)$, where the denominator does not depend on the sought variable B. Thus, the estimation operator can be put as

$$\hat{B}^*(R) = \arg\max_{B\in\mathscr{B}} \Psi(B)\Phi(B\,|\,R).$$

In the logarithmic form, the same criterion will have the equivalent form

$$\begin{cases} \hat{B}^*(R) = \arg\max\limits_{B\in\mathscr{B}} J(B), \\ J(B) = \log\Psi(B) + \log\Phi(B\,|\,R), \end{cases} \quad \text{or} \quad \begin{cases} \hat{B}^*(R) = \arg\min\limits_{B\in\mathscr{B}} J(B), \\ J(B) = -\log\Psi(B) - \log\Phi(B\,|\,R). \end{cases} \tag{25}$$

We have come to a particular version of two parts in the combined criterion (16) and a statistically sound way of balancing them. Despite the fact that the general form of the optimal estimation operator (23) completely rests on the posterior probabilistic properties of the factor exposures $B = (\boldsymbol{\beta}_t, t \in T)$, there is no need to compute the posterior joint probability density $P(B\,|\,R)$; it is enough to sum the logarithms of the *a priori* density $\Psi(B)$ and the conditional density $\Phi(B\,|\,R)$.

The singular loss function evaluates the non-coincidence of the actual path $B = (\boldsymbol{\beta}_t, t \in T)$ and its estimate $\hat{B} = (\hat{\boldsymbol{\beta}}_t, t \in T)$ over the entire time interval T, so it draws no distinction between errors at different time moments.

Thus, the main computational difficulty of the portfolio analysis problem falls onto solving the optimization problem (25) with the sum of the model densities $\Psi(B)$ and $\Phi(B\,|\,R)$. This problem hardly lends itself to an easy computational solution, except special cases when the probabilistic model of the portfolio is assumed to possess some special properties. One of mathematics’ most studied classes of hidden processes (which offer a compromise between the simplicity of estimation procedures and the ability to represent real-world phenomena) is the class of Makov random processes [20]. The structure of the hidden Markov model of the portfolio is presented in the next Section.

4 Hidden Markov models of the portfolio

4.1 The first-order model

An exhaustive representation of the probability density $P(X)$ over all realizations of any random process $X=(x_t, t=1,...,N)$ can be given by the product of conditional probability densities

$$P(X)=P(x_t,\ t=1,...,N)=p_1(x_1)\prod_{t=2}^{N}p_t(x_t\,|\,x_1,...,x_{t-1})\,. \tag{26}$$

The random process is said to possess the Markov property if its value at only one immediately precedent time moment occurs as an argument in each of the instantaneous conditional densities:

$$p_t(x_t\,|\,x_1,...,x_{t-1})=p_t(x_t\,|\,x_{t-1})\,. \tag{27}$$

The conditional probability densities $p_t(x_t\,|\,x_{t-1})$ are called the transition densities.

The definition (27) is generic for the notion of a Markov random process. It places on the current value of the process the requirement to depend only on one immediately precedent value; therefore, random processes of such a kind are often called Markov processes of the first order.

We define the first-order hidden Markov model of the portfolio as complex of the following two assumptions.

Assumption 1. *The unobservable random succession of factor exposures* $B=(\beta_t, t=0,1,...,N)$ *that includes the initial value* β_0 *is considered a priori as a Markov random process with known transition densities* $\psi_t(\beta_t\,|\,\beta_{t-1})$*:*

$$\Psi(B)=\psi_0(\beta_0)\Psi(B\,|\,\beta_0)\,,\ \ \Psi(B\,|\,\beta_0)=\prod_{t=1}^{N}\psi_t(\beta_t\,|\,\beta_{t-1})\,. \tag{28}$$

Assumption 2. *The observable portfolio returns form a succession of conditionally independent random variables* $R=(r_t^{(p)},\ t\in T)$ *each of which depends on only one current vector of factor exposures* $\varphi_t(r_t^{(p)}\,|\,\beta_t)$*:*

$$\Phi(R\,|\,B)=\prod_{t=1}^{N}\varphi_t(r_t^{(p)}\,|\,\beta_t)\,. \tag{29}$$

Below, in Section 5, we shall show that the assumptions 1 and 2 immediately effect the simplicity of the estimation operator for the singular loss function.

4.2 Higher-order models

Along with the notion of the Markov process of the first order (27), a more general definition is often used, especially in applied statistics, in accordance with which a random process is called a Markov process of the mth order if each current value of it depends on m immediately precedent values:

$$p_t(x_t\,|\,x_1,...,x_{t-1})=p_t(x_t\,|\,x_{t-1},...,x_{t-m})\,. \tag{30}$$

By analogy with the first-order hidden Markov model of the portfolio defined by assumptions 1 and 2, we shall use the notion of the mth-order hidden Markov model as that satisfying the following assumption:

Assumption 1a. *The unobservable random succession of factor exposures* $B=(\beta_t, t=-m+1,...,0,1,...,N)$*, that includes* n *initial values* $(\beta_0,...,\beta_{-m+1})$ *immediately prece-*

dent to the observation interval $t=1,\dots,N$, is considered a priori as a Markov random process of the m th order with known transition densities $\psi_t(\boldsymbol{\beta}_t \mid \boldsymbol{\beta}_{t-1},\dots,\boldsymbol{\beta}_{t-m})$, so that

$$\Psi(B)=\psi_0(\boldsymbol{\beta}_0,\dots,\boldsymbol{\beta}_{-m+1})\prod_{t=1}^{N}\psi_t(\boldsymbol{\beta}_t \mid \boldsymbol{\beta}_{t-1},\dots,\boldsymbol{\beta}_{t-m}). \quad (31)$$

As to the assumption 2, there is no need to generalize it from the viewpoint of practical problems concerned with portfolio analysis.

Assumptions 1a and 2 set off a much wider class of portfolio models than assumptions 1 and 2. Of course, the optimal decision rules of portfolio estimation considered below in Section 5 for the singular loss function will be somewhat more complicated in the case of higher-order hidden Markov models in comparison with the first-order model.

4.3 Linear and nonlinear normal hidden Markov models

4.3.1 First-order models

4.3.1.1 The linear model of general kind

The assumptions formulated in the previous Section state only the principal structure of the hidden Markov model of the portfolio (28)-(29), but the conditional densities $\psi_t(\boldsymbol{\beta}_t \mid \boldsymbol{\beta}_{t-1})$ and $\varphi_t(r_t^{(p)} \mid \boldsymbol{\beta}_t)$ remain free. For it to be possible to put down the respective optimal estimation operator in a computable form, these densities have to be defined concretely.

The simplest form of a conditional distribution commonly adopted in mathematical statistics is the so-called linear normal model [21]. In this model, only the mathematical expectation of the random variable, which is assumed to be normally distributed, depends on the value of the argument, whereas the variance or, in the multidimensional case, the covariance matrix of it remains constant by the argument:

$$\boldsymbol{\beta}_t=\mathrm{M}(\boldsymbol{\beta}_t \mid \boldsymbol{\beta}_{t-1})+\boldsymbol{\varepsilon}_t=\mathbf{V}_t\boldsymbol{\beta}_{t-1}+\boldsymbol{\varepsilon}_t, \quad (32)$$

where $\boldsymbol{\varepsilon}_t$ is the vector white noise with covariance matrices $\mathbf{Q}_t$, and $\mathbf{V}_t$ is a square matrix that determines the dynamics of the hidden process. The first-order portfolio model of such a kind will have the structure

$$\begin{cases}\psi_t(\boldsymbol{\beta}_t \mid \boldsymbol{\beta}_{t-1})=\dfrac{1}{|\mathbf{Q}_t|^{1/2}(2\pi)^{n/2}}\exp\left[-\dfrac{1}{2}(\boldsymbol{\beta}_t-\mathbf{V}_t\boldsymbol{\beta}_{t-1})^T\mathbf{Q}_t^{-1}(\boldsymbol{\beta}_t-\mathbf{V}_t\boldsymbol{\beta}_{t-1})\right],\\ \varphi_t(r_t^{(p)} \mid \boldsymbol{\beta}_t)=\dfrac{1}{\lambda^{1/2}(2\pi)^{1/2}}\exp\left[-\dfrac{1}{2\lambda}(r_t^{(p)}-\mathbf{r}_t^T\boldsymbol{\beta}_t)^2\right],\end{cases} \quad (33)$$

where, in accordance with (15), $\mathbf{r}_t$ and $\boldsymbol{\beta}_t$ are successions of $(n+1)$-dimensional vectors of known factors and unknown factor exposures, $r_t^{(p)}$ is succession of known portfolio returns. As predefined parameters of this model serve the square matrices $\mathbf{V}_t\ [(n+1)\times(n+1)]$ that determine the assumed, generally speaking, nonstationary linear dynamics of factor exposures, which is stationary if $\mathbf{V}_t=\mathbf{V}=\mathrm{const}$, the covariance matrices $\mathbf{Q}_t\ [(n+1)\times(n+1)]$ controlling the *a priori* variability of factor exposures, which, in the particular case, may be assumed to be constant $\mathbf{Q}_t=\mathbf{Q}=\mathrm{const}$, and the assumed variance λ of white noise e_t in (14).

The choice of matrix parameters $\mathbf{V}_t$ and $\mathbf{Q}_t$ expresses the *a priori* view on the hidden dynamics of factor exposures, whereas the assumed value of the observation noise variance σ^2 controls the balance between the *a priori* and observation-based information which will be fused in the estimate of the factor exposures $\hat{B} = (\hat{\boldsymbol{\beta}}_t, t \in T)$: the less λ, the more reliance is placed on the immediate observation, i.e., the time series $\left[(r_t^{(p)}, \mathbf{r}_t); t = 1,...,N\right]$, and vice-versa – greater values of λ shift the center of reliance to the *a priori* assumptions.

In particular, the absence of *a priori* information on the interdependence between, on the one hand, the single hidden factor exposures $(\beta_t^{(1)},...,\beta_t^{(n)})$ and, on the other hand, the intercept term $\alpha_t = \beta_t^{(n+1)}$ in (15) will be expressed as block-diagonal form of dynamic matrix $\mathbf{V}_t\left[(n+1)\times(n+1)\right]$:

$$\mathbf{V}_t = \begin{pmatrix} \mathbf{W}_t & \mathbf{0} \\ \mathbf{0}^T & w \end{pmatrix}. \tag{34}$$

It is hardly reasonable to take any other value of the scalar parameter w than $w = 1$, so, only matrix $\mathbf{W}_t \ (n \times n)$ is to be chosen here. Most likely, it will be enough to take a diagonal dynamics matrix as $\mathbf{W}_t = \text{Diag}(w_t^{(1)},...,w_t^{(n)})$, then

$$\mathbf{V}_t = \begin{pmatrix} w_t^{(1)} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & w_t^{(n)} & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}. \tag{35}$$

4.3.1.2 The general nonlinear model

A more general kind of a normal conditional distribution is the nonlinear normal model, in which, just as in the linear normal model, only the mathematical expectation of the normally distributed variable depends on the argument, but this dependence is expressed by an arbitrary function. This kind of model is appropriate for the *a priori* Markov succession of hidden factor exposures $\psi_t(\boldsymbol{\beta}_t \,|\, \boldsymbol{\beta}_{t-1})$, but is hardly necessary for the observation model $\varphi_t(r_t^{(p)} \,|\, \boldsymbol{\beta}_t)$:

$$\begin{cases} \psi_t(\boldsymbol{\beta}_t \,|\, \boldsymbol{\beta}_{t-1}) = \dfrac{1}{|\mathbf{Q}_t|^{1/2}\,(2\pi)^{n/2}} \exp\left\{-\dfrac{1}{2}\left[\boldsymbol{\beta}_t - \mathbf{f}_t(\boldsymbol{\beta}_{t-1})\right]^T \mathbf{Q}_t^{-1}\left[\boldsymbol{\beta}_t - \mathbf{f}_t(\boldsymbol{\beta}_{t-1})\right]\right\}, \\ \varphi_t(r_t^{(p)} \,|\, \boldsymbol{\beta}_t) = \dfrac{1}{\lambda^{1/2}(2\pi)^{1/2}} \exp\left[-\dfrac{1}{2\lambda}(r_t^{(p)} - \mathbf{r}_t^T\boldsymbol{\beta}_t)^2\right], \end{cases} \tag{36}$$

where $\mathbf{f}_t(\boldsymbol{\beta}_{t-1}): \mathsf{R}^{n+1} \to \mathsf{R}^{n+1}$ are appropriate vector functions. It is clear that the linear model (33) is a particular case of (36) with $\mathbf{f}_t(\boldsymbol{\beta}_{t-1}) = \mathbf{V}_t\boldsymbol{\beta}_{t-1}$.

Like the diagonal version of the matrix of linear dynamics (35), the simplest version of function $\mathbf{f}_t(\boldsymbol{\beta}_{t-1})$ in the nonlinear model is a collection of individual functions for each factor exposure $(\beta_t^{(1)},...,\beta_t^{(n)})$, whereas it is hardly reasonable to take a more complex model of the intercept term $\alpha_t = \beta_t^{(n+1)}$ than the trivial equality $\beta_t^{(n+1)} = \beta_{t-1}^{(n+1)}$:

$$\mathbf{f}_t(\boldsymbol{\beta}_{t-1}) = \left(f_t^{(1)}(\beta_{t-1}^{(1)}), \ldots, f_t^{(n)}(\beta_{t-1}^{(n)}),\ \beta_{t-1}^{(n+1)}\right)^T. \tag{37}$$

4.3.1.3 Fundamental nonlinearity of the dynamic RBSA model and its linear approximation

In Sharpe's original RBSA model (7), periodic return $r_t^{(p)}$ is approximated by a return on a portfolio consisting of n kinds of market assets. The latter return is represented by a linear combination of periodic returns of the assets $(r^{(1)},...,r^{(n)})$ under the assumption that such a portfolio remains unchanged within the estimation window. This fact is explained in Appendix 1.

In the dynamic RBSA model (13), in contrast to Sharpe's static model (7), the approximation portfolio's allocations are assumed dynamic, where the changes in allocations $\beta_t^{(i)} - \beta_{t-1}^{(i)}$ can be explained by the following two factors:

(a) changes in asset market values (or market prices); and

(b) buying and selling of assets in the approximation portfolio.

Although the approximation portfolio may have little in common with the portfolio or the instrument that it's designed to approximate[1], we can still think of the approximation portfolio as an active portfolio, in which assets' quantities are bought and sold to get close approximation of the given return series. Let's denote prices and amounts (quantities) of each asset as $x_t^{(i)}$ and $m_t^{(i)}$. Even if the asset quantities $m_t^{(i)}$ are changing very little (almost no trading or active changes in allocations of assets)

$$m_t^{(i)} \cong m_{t-1}^{(i)}, \tag{38}$$

the approximating portfolio's asset fractional weights

$$\beta_t^{(i)} = \frac{m_t^{(i)} x_t^{(i)}}{\sum_{k=1}^{n} m_t^{(k)} x_t^{(k)}}$$

will not be constant $\beta_t^{(i)} \neq \beta_{t-1}^{(i)}$ and will "drift" in time even if $m_t^{(i)} = m_{t-1}^{(i)}$, because the prices of assets are not constant $x_t^{(k)} \neq x_{t-1}^{(k)}$.

It is shown in Appendix 2 that $\beta_t^{(i)} = (1 + r_{t-1}^{(i)}) \Big/ \left(1 + \sum_{k=1}^{n} \beta_{t-1}^{(k)} r_{t-1}^{(k)}\right) \beta_{t-1}^{(i)}$ if $m_t^{(i)} = m_{t-1}^{(i)}$. Such a portfolio that doesn't involve any trading and the quantities of assets that are being held constant is called a "buy-and-hold" portfolio. Any portfolio that is not of the "buy-and-hold" kind will result in fluctuations in the differences $\beta_t^{(i)} - \left[(1 + r_{t-1}^{(i)}) \Big/ \left(1 + \sum_{k=1}^{n} \beta_{t-1}^{(k)} r_{t-1}^{(k)}\right)\right] \beta_{t-1}^{(i)}$, which we will assume as random noise with zero mathematical expectation. Such approximation can be written as the first-order hidden Markov model of asset fractions dynamics:

$$\beta_t^{(i)} = \frac{1 + r_{t-1}^{(i)}}{1 + \sum_{k=1}^{n} \beta_{t-1}^{(k)} r_{t-1}^{(k)}} \beta_{t-1}^{(i)} + \varepsilon_t^{(i)}. \tag{39}$$

Thus, nonlinear models of hidden dynamics like (36) are the only acceptable class of models for factor exposures $(\beta_t^{(1)},...,\beta_t^{(n)})$ having the meaning of a portfolio's cost distribution over the assumed asset classes.

If, in its turn, the noise is considered as a succession of normally distributed random variables, independent by t as well as by i, and if the *a priori* dynamics of intercept α_t in $\boldsymbol{\beta}_t = (\beta_t^{(1)},...,\beta_t^{(n)},\alpha_t)^T = (\beta_t^{(1)},...,\beta_t^{(n)},\ \beta_t^{(n+1)})^T$ in accordance with (13)-(14) is accepted as

$$\beta_t^{(n+1)} = \beta_{t-1}^{(n+1)} + \varepsilon_t^{(n+1)},$$

we obtain the nonlinear portfolio model (36) with

[1] The number of assets or factors is usually relatively small -- much smaller than the number of actual securities in the portfolio, which can number thousands.

$$\mathbf{f}_t(\boldsymbol{\beta}_{t-1}) = \begin{pmatrix} f_t^{(1)}(\boldsymbol{\beta}_{t-1}) \\ \vdots \\ f_t^{(n)}(\boldsymbol{\beta}_{t-1}) \\ f_t^{(n+1)}(\beta_{t-1}^{(n+1)}) \end{pmatrix} = \begin{pmatrix} \left[w_t^{(1)}(\boldsymbol{\beta}_{t-1})\right]\beta_{t-1}^{(1)} \\ \vdots \\ \left[w_t^{(1)}(\boldsymbol{\beta}_{t-1})\right]\beta_{t-1}^{(n)} \\ 1\cdot\beta_{t-1}^{(n+1)} \end{pmatrix}, \quad \begin{array}{l} w_t^{(i)}(\boldsymbol{\beta}_{t-1}) = \dfrac{1+r_{t-1}^{(i)}}{1+\sum_{k=1}^{n}\beta_{t-1}^{(k)}r_{t-1}^{(k)}}, \; i=1,\ldots n, \\ w_t^{(n+1)}(\boldsymbol{\beta}_{t-1}) = 1, \end{array} \tag{40}$$

and with an appropriate diagonal covariance matrix $\mathbf{Q}_t = \mathrm{Diag}(q_t^{(1)},\ldots,q_t^{(n+1)})$.

With matrix denotations

$$\mathbf{V}(\boldsymbol{\beta}_{t-1}) = \begin{pmatrix} \mathbf{W}(\boldsymbol{\beta}_{t-1}) & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{W}(\boldsymbol{\beta}_{t-1}) = \begin{pmatrix} w_t^{(1)}(\boldsymbol{\beta}_{t-1}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_t^{(n)}(\boldsymbol{\beta}_{t-1}) \end{pmatrix},$$

the dynamic RBSA model will obtain the form:

$$\begin{cases} \psi_t(\boldsymbol{\beta}_t \mid \boldsymbol{\beta}_{t-1}) = \dfrac{1}{|\mathbf{Q}_t|^{1/2}(2\pi)^{n/2}} \exp\left\{-\dfrac{1}{2}\left[\boldsymbol{\beta}_t - [\mathbf{V}_t(\boldsymbol{\beta}_{t-1})]\boldsymbol{\beta}_{t-1}\right]^T \mathbf{Q}_t^{-1}\left[\boldsymbol{\beta}_t - [\mathbf{V}_t(\boldsymbol{\beta}_{t-1})]\boldsymbol{\beta}_{t-1}\right]\right\}, \\ \varphi_t(r_t^{(p)} \mid \boldsymbol{\beta}_t) = \dfrac{1}{\lambda^{1/2}(2\pi)^{1/2}} \exp\left[-\dfrac{1}{2\lambda}(r_t^{(p)} - \mathbf{r}_t^T\boldsymbol{\beta}_t)^2\right]. \end{cases} \tag{41}$$

This is a nonlinear normal Markov model of the first order. The nonlinearity of this model consists in that matrices $\mathbf{V}_t(\boldsymbol{\beta}_{t-1})$ depend on the variables whose dynamics they express.

From the computational point of view, linear models are always preferable because they lead to much simpler estimation operators than nonlinear ones. The nonlinearity of the dynamic RBSA model originates from the model of asset fractions dynamics (39). If, as first approximation, we replace $\sum_{k=1}^{n}\beta_{t-1}^{(k)}r_{t-1}^{(k)}$ by $r_{t-1}^{(p)}$, the linear combination $\sum_{k=1}^{n}\beta_{t-1}^{(k)}r_{t-1}^{(k)}$ will differ from the current portfolio return $r_{t-1}^{(p)}$ (13) by the intercept term α_{t-1} and the noise term e_{t-1}. If we choose to ignore this inaccuracy, we can approximate the asset fractions dynamics by the equation

$$\beta_t^{(i)} \cong \frac{1+r_{t-1}^{(i)}}{1+r_{t-1}^{(p)}}\beta_{t-1}^{(i)} + \varepsilon_t^{(i)}.$$

As a result, the coefficients in (40) $w_t^{(i)} = (1+r_{t-1}^{(i)})/(1+r_{t-1}^{(p)})$ will no longer depend on the variables $\boldsymbol{\beta}_{t-1}$, and the model as a whole will become linear (33) with

$$\mathbf{V}_t = \begin{pmatrix} \mathbf{W}_t & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{W}_t = \mathrm{Diag}\left(\frac{1+r_{t-1}^{(1)}}{1+r_{t-1}^{(p)}},\ldots,\frac{1+r_{t-1}^{(n)}}{1+r_{t-1}^{(p)}}\right). \tag{42}$$

If such an approximation of the nonlinear model is not sufficient, the problem can be solved in a series of iterations of linear models, where the asset weights obtained as a solution on the previous step are used in the denominator of (39) on the next step.

4.3.2 Higher-order models

In accordance with assumptions 1a and 2 (Section 4.2), the linear equation of the hidden Markov process of order m, in contrast to (32), will have the form

$$\boldsymbol{\beta}_t = \mathrm{M}(\boldsymbol{\beta}_t \mid \boldsymbol{\beta}_{t-1},\ldots,\boldsymbol{\beta}_{t-m}) + \boldsymbol{\varepsilon}_t = \mathbf{V}_{t,1}\boldsymbol{\beta}_{t-1} + \ldots + \mathbf{V}_{t,m}\boldsymbol{\beta}_{t-m} + \boldsymbol{\varepsilon}_t, \tag{43}$$

in which the higher-order dynamics of the hidden process is determined by m square matrices $\mathbf{V}_{t,j}$, whereas $\boldsymbol{\varepsilon}_t$ remains to be the normal vector white noise with covariance matrices $\mathbf{Q}_t$, just as in (32). Hence, we can put down the linear normal Markov model of the mth order as

$$\begin{cases} \psi_t(\boldsymbol{\beta}_t \,|\, \boldsymbol{\beta}_{t-1},...,\boldsymbol{\beta}_{t-m}) = \dfrac{1}{|\mathbf{Q}_t|^{1/2}\,(2\pi)^{n/2}} \exp\left[-\dfrac{1}{2}\left(\boldsymbol{\beta}_t - \sum\nolimits_{j=1}^{m} \mathbf{V}_{t,j}\boldsymbol{\beta}_{t-j}\right)^T \mathbf{Q}_t^{-1} \left(\boldsymbol{\beta}_t - \sum\nolimits_{j=1}^{m} \mathbf{V}_{t,j}\boldsymbol{\beta}_{t-j}\right)\right], \\ \varphi_t(r_t^{(p)} \,|\, \boldsymbol{\beta}_t) = \dfrac{1}{\lambda^{1/2}(2\pi)^{1/2}} \exp\left[-\dfrac{1}{2\lambda}(r_t^{(p)} - \mathbf{r}_t^T \boldsymbol{\beta}_t)^2 \right]. \end{cases} \quad (44)$$

Analogously, the nonlinear normal model of the m th order will be expressed by the notation

$$\begin{cases} \psi_t(\boldsymbol{\beta}_t \,|\boldsymbol{\beta}_{t-1},...,\boldsymbol{\beta}_{t-m}) = \dfrac{1}{|\mathbf{Q}_t|^{1/2}(2\pi)^{n/2}} \exp\left\{ -\dfrac{1}{2}\left[\boldsymbol{\beta}_t - \mathbf{f}_t(\boldsymbol{\beta}_{t-1},...,\boldsymbol{\beta}_{t-m})\right]^T \mathbf{Q}_t^{-1} \left[\boldsymbol{\beta}_t - \mathbf{f}_t(\boldsymbol{\beta}_{t-1},...,\boldsymbol{\beta}_{t-m})\right] \right\}, \\ \varphi_t(r_t^{(p)} \,|\, \boldsymbol{\beta}_t) = \dfrac{1}{\lambda^{1/2}(2\pi)^{1/2}} \exp\left[-\dfrac{1}{2\lambda}(r_t^{(p)} - \mathbf{r}_t^T \boldsymbol{\beta}_t)^2 \right]. \end{cases} \quad (45)$$

These two formulations explicitly show the complication of a higher-order model in comparison with the first-order one (33) and (36).

4.4 Constrained normal hidden Markov models

4.4.1 The basic principle of forming constrained hidden Markov models

In accordance with (17), to impose some *a priori* constraint $B \in \mathcal{Z} \subset \mathcal{B}$ upon the hidden random process originally defined by a probability density over the whole range $\mathcal{B}$

$$\Psi(B),\ B \in \mathcal{B},\ \int\limits_{B \in \mathcal{B}} \Psi(B)dB = 1,$$

we have to truncate this density by setting to zero its values outside the admissible area:

$$\Psi_{\mathcal{Z}}(B) = \begin{cases} (1/c)\Psi(B),\ B \in \mathcal{Z}, \\ 0,\ B \neq \mathcal{Z}, \end{cases} \qquad c = \int\limits_{\mathcal{Z}} \Psi(B)dB. \quad (46)$$

It is often enough to use the partial definition of the probabilistic model

$$\Psi_{\mathcal{Z}}(B) \begin{cases} \propto \Psi(B),\ B \in \mathcal{Z}, \\ = 0,\ B \neq \mathcal{Z}, \end{cases} \quad (46a)$$

where $\propto$ is the sign of proportionality, and thereby spare computing the normalization coefficient $1/c$, which may turn out to be problematic, especially when the admissible area is a set of measure null and the integral in (46) equals zero.

In contrast to the full definition of a constrained model (46)-(46a), Markov models define the hidden random process in an indirect way through a series of conditional distributions of its instantaneous values (28). Introduction of the *a priori* constraint $B \in \mathcal{Z} \subset \mathcal{B}$ into an already defined full Markov model will be especially simple if the set $\mathcal{Z}$ of admissible realizations of the random process $B = (\boldsymbol{\beta}_t, t = 0,1,...,N)$ is completely determined by the sets Z_t of admissible instantaneous values:

$$B = (\boldsymbol{\beta}_t, t = 0,1,...,N) \in \mathcal{Z} \ \text{ if } \ \boldsymbol{\beta}_t \in \mathrm{Z}_t,\ t = 0,1,...,N.$$

In this case, it is enough to truncate the transition densities:

$$\overline{\psi}_t(\boldsymbol{\beta}_t \,|\, \boldsymbol{\beta}_{t-1}) \begin{cases} \propto \psi_t(\boldsymbol{\beta}_t \,|\, \boldsymbol{\beta}_{t-1}),\ \boldsymbol{\beta}_t \in \mathrm{Z}_t, \\ = 0,\ \boldsymbol{\beta}_t \notin \mathrm{Z}_t. \end{cases} \quad (47)$$

In this work, we shall use just this way of introducing constraints into hidden Markov models.

4.4.2 Equality and inequality constraints on factor exposures: Truncated normal hidden Markov models

The most natural interpretation of factor exposures $(\beta_t^{(1)},...,\beta_t^{(n)})$ in the combined vector $\boldsymbol{\beta}_t=(\beta_t^{(1)},...,\beta_t^{(n)},\alpha_t)^T=(\beta_t^{(1)},...,\beta_t^{(n)},\ \beta_t^{(n+1)})^T$ (15) is the distribution of the portfolio's cost over the assets classes. If the portfolio is fully invested, the budget constraint must hold true at all time moments:

$$\sum_{i=1}^{n}\beta_t^{(i)}=1\,,\quad t\in T. \tag{48}$$

Other types of constraints typical for portfolio analysis are non-negativity bounds that represent the assumption that the portfolio manager is not involved in any significant short selling or hedging:

$$\beta_t^{(i)}\geq 0\,,\ i=1,...,n\,. \tag{49}$$

The constraints (48) and (49) imposed upon the unknown values of factor exposures express the major portion of *a priori* information about the sought path $B=(\boldsymbol{\beta}_t,\ t\in T)$ and are an integral part of the portfolio model.

There is no specificity in the budget constraint (48), which is a particular case of a system of $k<n+1$ linear equality constraints

$$\mathbf{F}_t\boldsymbol{\beta}_t+\mathbf{c}_t=\mathbf{0}\,,\ \ \mathbf{F}_t\left[k\times(n+1)\right],\ \ \mathbf{c}_t\in\mathbb{R}^k\,. \tag{50}$$

Here $\text{Rank}(\mathbf{F}_t)=k$, i.e. $\mathbf{F}_t$ are full-rank matrices, because otherwise the number of constraints could be decreased. The budget constraint (48) results from (50) with $k=1$ and

$$\mathbf{F}_t=\mathbf{F}=(\underbrace{1,...,1}_{n},0)\ \left[1\times(n+1)\right],\ \ \mathbf{c}_t=c=-1\in\mathbb{R}\,.$$

Analogously, in the general case, a system of an arbitrary number l of linear inequality constraints may be imposed upon the vector of hidden variables $\boldsymbol{\beta}_t=(\beta_t^{(1)},...,\beta_t^{(n)},\ \beta_t^{(n+1)})^T$:

$$\mathbf{G}_t\boldsymbol{\beta}_t+\mathbf{h}_t\geq\mathbf{0}\,,\ \ \mathbf{G}_t\left[l\times(n+1)\right],\ \ \mathbf{h}_t\in\mathbb{R}^l\,. \tag{51}$$

The non-negativity constraints (49) are a particular case of (51) with $l=n$ and

$$\mathbf{G}_t=\mathbf{G}=\left(\text{Diag}(\underbrace{1,...,1}_{n}),\mathbf{0}\right)\left[n\times(n+1)\right],\ \mathbf{0}\in\mathbb{R}^n\,,\ \ \mathbf{h}_t=\mathbf{h}=\mathbf{0}\in\mathbb{R}^n\,.$$

In the general case, constraints (50) and (51) cut off a succession of subsets $Z_t\subset\mathbb{R}^{n+1}$ which have measure null if equality constraints are present. Application of the general principle (47) to (33) gives the truncated first-order linear normal model:

$$\begin{cases}\overline{\psi}_t(\boldsymbol{\beta}_t\,|\,\boldsymbol{\beta}_{t-1})\begin{cases}\propto\exp\left[-\dfrac{1}{2}(\boldsymbol{\beta}_t-\mathbf{V}_t\boldsymbol{\beta}_{t-1})^T\mathbf{Q}_t^{-1}(\boldsymbol{\beta}_t-\mathbf{V}_t\boldsymbol{\beta}_{t-1})\right],\ \mathbf{F}_t\boldsymbol{\beta}_t+\mathbf{c}_t=\mathbf{0}\ \text{and}\ \mathbf{G}_t\boldsymbol{\beta}_t+\mathbf{h}_t\geq\mathbf{0},\\ 0,\ \text{otherwise}.\end{cases}\\ \varphi_t(r_t^{(p)}\,|\,\boldsymbol{\beta}_t)=\dfrac{1}{\lambda^{1/2}(2\pi)^{1/2}}\exp\left[-\dfrac{1}{2\lambda}(r_t^{(p)}-\mathbf{r}_t^T\boldsymbol{\beta}_t)^2\right].\end{cases} \tag{52}$$

Respectively, the truncated nonlinear normal model results from (36):

$$\begin{cases} \overline{\psi}_t(\boldsymbol{\beta}_t \mid \boldsymbol{\beta}_{t-1}) \begin{cases} \propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}_t - \mathbf{f}_t(\boldsymbol{\beta}_{t-1})\right]^T \mathbf{Q}_t^{-1}\left[\boldsymbol{\beta}_t - \mathbf{f}_t(\boldsymbol{\beta}_{t-1})\right]\right\}, \ \mathbf{F}_t\boldsymbol{\beta}_t + \mathbf{c}_t = \mathbf{0} \text{ and } \mathbf{G}_t\boldsymbol{\beta}_t + \mathbf{h}_t \geq \mathbf{0}, \\ 0, \text{ otherwise.} \end{cases} \\ \varphi_t(r_t^{(p)} \mid \boldsymbol{\beta}_t) = \frac{1}{\lambda^{1/2}(2\pi)^{1/2}} \exp\left[-\frac{1}{2\lambda}(r_t^{(p)} - \mathbf{r}_t^T\boldsymbol{\beta}_t)^2\right]. \end{cases} \tag{53}$$

5 Minimization of the error risk for the singular loss function: The optimization-based principle of estimating dynamic multi-factor models

5.1 The general criterion of decision making

For the Markov model of the portfolio (28)-(29), the optimal estimation operator (25) is defined as the solution of the optimization problem which can be put down in two equivalent versions – either in the maximization or in the minimization form. In the statistical decision-making theory, the latter version is usually preferred:

$$\begin{cases} \hat{B}^*(R) = \arg\min_B J(B \mid R), \\ J(B \mid R) = J_Y(\boldsymbol{\beta}_t, t = 0,1,...,N) = \sum_{t=1}^{N}\left[-\log\varphi_t(r_t^{(p)} \mid \boldsymbol{\beta}_t)\right] + \sum_{t=1}^{N}\left[-\log\psi_t(\boldsymbol{\beta}_t \mid \boldsymbol{\beta}_{t-1})\right] + \left[-\log\psi_0(\boldsymbol{\beta}_0)\right]. \end{cases} \tag{54}$$

The first item of the criterion $J(B \mid R)$ is responsible for matching of the factor exposures and portfolio return at each time moment individually, whereas the second and the third items express the *a priori* information on the unknown succession of factor exposures.

The specificity of the Markov model displays itself in that the objective function is pair-wise separable, i.e., is the sum of partial functions, each of which depends on not more than two immediately consecutive vectors of factor exposures. Optimization problems of such a kind invite for dynamic programming as the most appropriate computational principle of solving them [22,23]. The computational complexity of the dynamic programming procedure is proportional to the number of immediately adjacent variables, namely, $N+1$ in (54), and is provided by decomposition of the original objective function of the full number of variables into a succession of intervening function each of only one variable, that are to be minimized sequentially. However, in its classical form, the dynamic programming procedure is immediately applicable only if the goal variables take values from a finite set, which is not the case in the problems of portfolio analysis, in which the factor exposures $\boldsymbol{\beta}_t = (\beta_t^{(1)},...,\beta_t^{(n)},\alpha_t)^T = (\beta_t^{(1)},...,\beta_t^{(n)}, \beta_t^{(n+1)})^T$ are continuous variables in principle.

In any case, the numerical solution of the optimization problem (54) is impossible until the conditional probability densities $\psi_0(\boldsymbol{\beta}_0)$, $\psi_t(\boldsymbol{\beta}_t \mid \boldsymbol{\beta}_{t-1})$ and $\varphi_t(r_t^{(p)} \mid \boldsymbol{\beta}_t)$ are defined explicitly. In particular, the normality assumption on the hidden factor exposures within the bounds of the Markov model of the portfolio leads to the class of quadratic pair-wise separable objective functions, for which, as it is shown in [24], there exists a parametric version of the dynamic programming procedure that reconciles the exceptionally low computational complexity of the classical dynamic programming procedure with the continuous nature of factor exposures.

5.2 Normality assumption: Flexible least squares and generalized flexible least squares

Under the normality and linearity assumption on the hidden Markov model of factor exposures (33) and the additional assumption on the normality of the initial value $\boldsymbol{\beta}_0$

$$\psi_0(\boldsymbol{\beta}_0) = \frac{1}{|\mathbf{Q}_0|^{1/2}\,(2\pi)^{n/2}} \exp\left[-\frac{1}{2}(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0^*)^T \mathbf{Q}_0^{-1}(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0^*)\right],$$

where $\boldsymbol{\beta}_0^*$ and $\mathbf{Q}_0$ are, respectively, the *a priori* mathematical expectation and *a priori* covariance matrix of $\boldsymbol{\beta}_0$, the summands in the general criterion (54) will become quadratic functions. If we double the natural logarithm of the right part of (33), omit summands that don't depend on $\boldsymbol{\beta}_t = (\beta_t^{(1)},...,\beta_t^{(n)},\ \beta_t^{(n+1)})^T$, and multiply the result by λ, we shall come to the following objective function:

$$\begin{cases} (\hat{\boldsymbol{\beta}}_t, t = 0,1,...,N) = \arg\min J(\boldsymbol{\beta}_t, t = 0,1,...,N), \\ J(\boldsymbol{\beta}_t, t = 0,1,...,N) = \sum_{t=1}^{N} (r_t^{(p)} - \mathbf{r}_t^T \boldsymbol{\beta}_t)^2 + \lambda\left[\sum_{t=1}^{N} (\boldsymbol{\beta}_t - \mathbf{V}_t\boldsymbol{\beta}_{t-1})^T \mathbf{U}_t (\boldsymbol{\beta}_t - \mathbf{V}_t\boldsymbol{\beta}_{t-1}) + (\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0^*)^T \mathbf{U}_0 (\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0^*)\right], \end{cases} \quad (55)$$

where $\mathbf{U}_t = \mathbf{Q}_t^{-1}$ are the inverse covariance matrices in the normal hidden Markov model (33). Here the term $(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0^*)^T \mathbf{U}_0 (\boldsymbol{\beta}_0 - \boldsymbol{\beta}_0^*)$ expresses the eventual *a priori* knowledge on the initial value of the series of regression coefficients $\boldsymbol{\beta}_0 = (\beta_0^{(1)},...,\beta_0^{(n)},\beta_0^{(n+1)})^T$ at the time moment $t = 0$ immediately preceding the start of observation. If no *a priori* assumption on this vector is available, we have to put $\mathbf{U}_0 = \mathbf{0}$, or, in the original probabilistic formulation of the model, to assume the *a priori* variances of independent elements $\beta_0^{(i)}$ to be infinite $\mathbf{Q}_0 = \mathrm{Diag}(\infty,...,\infty)$, which is equivalent to the absence of this term:

$$\begin{cases} (\hat{\boldsymbol{\beta}}_t, t = 0,1,...,N) = \arg\min J(\boldsymbol{\beta}_t, t = 0,1,...,N), \\ J(\boldsymbol{\beta}_t, t = 0,1,...,N) = \sum_{t=1}^{N} (r_t^{(p)} - \mathbf{r}_t^T \boldsymbol{\beta}_t)^2 + \lambda \sum_{t=1}^{N} (\boldsymbol{\beta}_t - \mathbf{V}_t\boldsymbol{\beta}_{t-1})^T \mathbf{U}_t (\boldsymbol{\beta}_t - \mathbf{V}_t\boldsymbol{\beta}_{t-1}). \end{cases} \quad (56)$$

In the opposite extreme case, when the initial vector of regression coefficients is completely known $\boldsymbol{\beta}_0 = \boldsymbol{\beta}_0^*$, all the *a priori* variances equal zero, and, from the formal viewpoint, $\mathbf{U}_0 = \mathrm{Diag}(\infty,...,\infty)$. But it is much more convenient to remove the third term from the criterion, consider $\boldsymbol{\beta}_0 = \boldsymbol{\beta}_0^*$ as a fixed parameter, cancel it from the set of variables, and minimize the criterion with respect to the remaining regression coefficients $J(\boldsymbol{\beta}_t, t = 1,...,N) \to \min$.

The criterion of decision making (56) is known under the name of the *Flexible Least Squares* (FLS) method proposed by Kalaba and Tesfatsion [25,26,27] as a means of parameter estimation in dynamic linear regression models. The structure of this criterion explicitly displays the essence of the FLS approach to dynamic regression estimation as a multi-objective optimization problem. The first term is the squared Euclidean norm of the linear regression residuals $\| e_{[1,...,N]} \|^2$, $e_t = r_t^{(p)} - \mathbf{r}_t^T \boldsymbol{\beta}_t$, responsible for the model fit (14), the second term is a specific squared Euclidean norm of the deviation of regression coefficients from the assumed dynamics $\| \boldsymbol{\varepsilon}_{[2,...,N]} \|^2 = \sum_{t=1}^{N} \boldsymbol{\varepsilon}_t^T \mathbf{U}_t \boldsymbol{\varepsilon}_t$, $\boldsymbol{\varepsilon}_t = \boldsymbol{\beta}_t - \mathbf{V}_t \boldsymbol{\beta}_{t-1}$ (32), which is determined by the choice of the positive semidefinite matrices $\mathbf{U}_t$, whereas the positive weighting coefficient λ is to be chosen to balance the relative weights between these two particular objective functions. If $\lambda \to \infty$, the solution of (56) becomes very smooth, and approaches the ordinary least squares solution, while selecting λ close to zero makes the parameters very volatile. Typically, the problem (56) is solved and presented for different values of parameter λ.

The unconstrained quadratic optimization problem (56) leads to a system of linear equations with respect to $(n+1)(N+1)$ unknown regression coefficients $B=(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_N)$. In contrast to quadratic optimization problems of general kind, all the summands of the pair-wise separable objective function are functions of not more than two vector variables $\boldsymbol{\beta}_{t-1}$ and $\boldsymbol{\beta}_t$. The specificity of the linear equation system resulting from a pair-wise separable quadratic objective function is that its matrix is block-wise three-diagonal and, so, can be easily solved by the double-sweep method that has the linear computational complexity with respect to the number $N+1$ of vector variables, i.e. the length of the time series being analyzed. The size of the square blocks $\left[(n+1)\times(n+1)\right]$ to be inverted when solving this linear equation system is determined by the dimensionality of vector arguments $\boldsymbol{\beta}_t \in \mathbb{R}^{n+1}$. Since the number of operations required for inverting a square matrix is proportional to the cube of its size, the computational complexity of the double-sweep method with respect to the number of elements in each vector $\boldsymbol{\beta}_t$ will be cubic $(n+1)^3$.

If a more general nonlinear normal Markov model (36) is accepted, the FLS criterion will no longer be quadratic because of the nonlinear function $\mathbf{f}_t(\boldsymbol{\beta}_{t-1})$ occurring in the dynamics term:

$$J(\boldsymbol{\beta}_t, t=0,1,...,N) = \sum_{t=1}^{N}(r_t^{(p)} - \mathbf{r}_t^T\boldsymbol{\beta}_t)^2 + \lambda\sum_{t=1}^{N}\left[\boldsymbol{\beta}_t - \mathbf{f}_t(\boldsymbol{\beta}_{t-1})\right]^T\mathbf{U}_t\left[\boldsymbol{\beta}_t - \mathbf{f}_t(\boldsymbol{\beta}_{t-1})\right] \to \min. \qquad (57)$$

In accordance with the above reasoning, we have omitted here the term responsible for the *a priori* knowledge on the initial value $\boldsymbol{\beta}_0$. To obtain the linear FLS criterion (56), it is enough to take the linear model of the hidden dynamics $\mathbf{f}_t(\boldsymbol{\beta}_{t-1}) = \mathbf{V}_t\boldsymbol{\beta}_{t-1}$.

In the case of the normal linear Markov model of m th order (44), the FLS criterion (55) will obtain a more complicated form. First of all, it will contain, as variables, the vectors of regression coefficients at m time moments preceding the start of observation $(\boldsymbol{\beta}_{-m+1},...,\boldsymbol{\beta}_0)$ instead of only one vector $\boldsymbol{\beta}_0$. Secondly, the *a priori* information on these values needs a much more complicated probabilistic description than in the first-order case. We assume that this information is unavailable and omit here this description:

$$J(\boldsymbol{\beta}_t, t=0,1,...,N) = \sum_{t=1}^{N}(r_t^{(p)} - \mathbf{r}_t^T\boldsymbol{\beta}_t)^2 + \lambda\sum_{t=1}^{N}\left(\boldsymbol{\beta}_t - \sum\nolimits_{j=1}^{m}\mathbf{V}_{t,j}\boldsymbol{\beta}_{t-j}\right)^T\mathbf{U}_t\left(\boldsymbol{\beta}_t - \sum\nolimits_{j=1}^{m}\mathbf{V}_{t,j}\boldsymbol{\beta}_{t-j}\right) \to \min. \qquad (58)$$

In the opposite case, when the initial values $(\boldsymbol{\beta}_{-m+1},...,\boldsymbol{\beta}_0)$ are completely known, the criterion is to be minimized only by the regression coefficients within the observation interval $J(\boldsymbol{\beta}_t, t=1,...,N) \to \min$ under the assumptions $\boldsymbol{\beta}_{-m+1} = \boldsymbol{\beta}^*_{-m+1},...,\boldsymbol{\beta}_0 = \boldsymbol{\beta}^*_0$.

The computational complexity of this problem with respect to the length of the time series to be analyzed remains linear (namely, proportional to $N+1$,) but it is cubic with respect to the order m of the Markov model (becomes proportional to $\left[(n+1)m\right]^3$ instead of $(n+1)^3$ in the first-order case.)

A particular case of the criterion (58) was suggested by Lütkepohl and Herwartz under the name of *Generalized Flexible Least Squares* (GFLS) method [28]:

$$J(\boldsymbol{\beta}_t, t=1,...,N) = \sum_{t=1}^{N} (r_t^{(p)} - \mathbf{r}_t^T \boldsymbol{\beta}_t)^2 + \lambda_1 \sum_{t=m+1}^{N} \left(\boldsymbol{\beta}_t - \sum\nolimits_{j=1}^{m} \mathbf{V}_{1,j} \boldsymbol{\beta}_{t-j}\right)^T \mathbf{U}_1 \left(\boldsymbol{\beta}_t - \sum\nolimits_{j=1}^{m} \mathbf{V}_{1,j} \boldsymbol{\beta}_{t-j}\right) + \lambda_2 \sum_{t=s+1}^{N} (\boldsymbol{\beta}_t - \mathbf{V}_2 \boldsymbol{\beta}_{t-s})^T \mathbf{U}_2 (\boldsymbol{\beta}_t - \mathbf{V}_2 \boldsymbol{\beta}_{t-s}) \to \min. \quad (59)$$

In this specific version of the multi-objective criterion, two different norms of the regression parameter variation are fused – namely, the norm based on a higher-order model of regression dynamics $\| \boldsymbol{\varepsilon}_{1,[m+1,...,N]} \|^2 = \sum\nolimits_{t=1}^{N} \boldsymbol{\varepsilon}_{1,t}^T \mathbf{U}_1 \boldsymbol{\varepsilon}_{1,t}$, $\boldsymbol{\varepsilon}_{1,t} = \boldsymbol{\beta}_t - \sum\nolimits_{j=1}^{m} \mathbf{V}_{1,j} \boldsymbol{\beta}_{t-j}$, and that representing the variation at a single predefined value of the time lag $\| \varepsilon_{2,[s+1,...,N]} \|^2 = \sum\nolimits_{t=1}^{N} \boldsymbol{\varepsilon}_{2,t}^T \mathbf{U}_2 \boldsymbol{\varepsilon}_{2,t}$, $\varepsilon_{2,t} = \boldsymbol{\beta}_t - \mathbf{V}_2 \boldsymbol{\beta}_{t-s}$. Each of these norms is defined by the choice of the respective positive semidefinite matrix, respectively, $\mathbf{U}_1$ and $\mathbf{U}_2$, and weighting parameters λ_1 and λ_2.

However, the FLS and GFLS methods discussed above are insufficient for estimating dynamic multi-factor models adequate for financial applications – first of all, because of the presence of constraints $\boldsymbol{\beta}_t \in Z_t$ in the RBSA model (14). In the next Section, we consider modifications of these methods with respect to linear equality and inequality constraints, respectively, (50) and (51).

5.3 Truncated normality assumption: The Constrained Flexible Least Squares (CFLS) method for Markov hidden models of the portfolio dynamics

5.3.1 The first-order Markov CFLS

To take into account the equality (50) and inequality constraints (51), the truncated normal Markov model (52) is to be used. Substitution of $\overline{\psi}_t(\boldsymbol{\beta}_t \,|\, \boldsymbol{\beta}_{t-1})$ into the general criterion (54) gives the same quadratic FLS criterion (56) additionally constrained by linear equalities and inequalities:

$$\begin{cases} (\hat{\boldsymbol{\beta}}_t, t=0,1,...,N) = \arg\min J(\boldsymbol{\beta}_t, t=0,1,...,N), \\ J(\boldsymbol{\beta}_t, t=0,1,...,N) = \sum_{t=1}^{N} (r_t^{(p)} - \mathbf{r}_t^T \boldsymbol{\beta}_t)^2 + \lambda \sum_{t=1}^{N} (\boldsymbol{\beta}_t - \mathbf{V}_t \boldsymbol{\beta}_{t-1})^T \mathbf{U}_t (\boldsymbol{\beta}_t - \mathbf{V}_t \boldsymbol{\beta}_{t-1}) \to \min, \\ \mathbf{F}_t \boldsymbol{\beta}_t + \mathbf{c}_t = \mathbf{0}, \ \mathbf{G}_t \boldsymbol{\beta}_t + \mathbf{h}_t \geq \mathbf{0} \end{cases} \quad (60)$$

where $\mathbf{U}_t = \mathbf{Q}_t^{-1}$. We call this criterion of decision making the *Constrained Flexible Least Squares* (CFLS) method in contrast to the unconstrained *Flexible Least Squares* (FLS) criterion (56).

The presence of linear constraints in the CFLS criterion (60) transfers the originally quadratic optimization problem (56) into a much wider class of quadratic programming problems [29,30]. This problem cannot be solved by the computationally effective double-sweep method that essentially exploits the pair-wise separability of the quadratic objective function. The forced application of quadratic programming procedures of general kind to the quadratic programming problem (56) raises the computational complexity of the problem to the third power of the length of the time series $(N+1)^3$ in contrast to the linear computational complexity $N+1$ in the case of the unconstrained FLS criterion. All in all, the total computational complexity of the quadratic programming problem, if traditional quadratic programming procedures are applied, is proportional to $(N+1)^3 (n+1)^3$.

5.3.2 Higher-order Markov CFLS

The structure of the CFLS criterion (60) remains, actually, the same for Markov models of higher orders:

$$\begin{cases} J(\boldsymbol{\beta}_t, t=0,1,...,N) = \sum_{t=1}^{N}(r_t^{(p)} - \mathbf{r}_t^T \boldsymbol{\beta}_t)^2 + \lambda \sum_{t=1}^{N}\left(\boldsymbol{\beta}_t - \sum_{j=1}^{m} \mathbf{V}_{t,j}\boldsymbol{\beta}_{t-j}\right)^T \mathbf{U}_t \left(\boldsymbol{\beta}_t - \sum_{j=1}^{m} \mathbf{V}_{t,j}\boldsymbol{\beta}_{t-j}\right) \to \min, \\ \mathbf{F}_t\boldsymbol{\beta}_t + \mathbf{c}_t = \mathbf{0}, \ \mathbf{G}_t\boldsymbol{\beta}_t + \mathbf{h}_t \geq \mathbf{0}. \end{cases}$$

What is different here is only the computational complexity of the quadratic programming problem, which, in addition to the cubic complexity with respect to the length of the time series and the number of factors $(N+1)^3(n+1)^3$, is cubic also with respect to the order of the Markov model m^3.

6 Parameter adjustment in the FLS and CFLS model

6.1 Parameters of the FLS and CFLS criteria of decision making

The FLS (56) and CFLS (60) criteria of decision making, based on the linear normal hidden Markov model of the portfolio (33), contain, as parameters, matrices of the assumed linear dynamics $\mathbf{V}_t\left[(n+1)\times(n+1)\right]$, inverse covariance matrices of the normal Markov model $\mathbf{U}_t = \mathbf{Q}_t^{-1}\left[(n+1)\times(n+1)\right]$ and the variance of observation noise σ^2. Whereas the dynamics matrices $\mathbf{V}_t$ are, as a rule, determined by some principal assumptions on the portfolio, as is the case with the RBSA model (42), it is hard to find firm *a priori* reasons for the choice of covariance matrices controlling the *a priori* variability of factor exposures and the intensity of the observation noise responsible for the inaccuracy of the multi-factor model (13). On the force of this circumstance, positive semidefinite matrices $\mathbf{U}_t$ and positive number σ^2 are to be considered as free parameters of the decision rule which are to be chosen by the user.

The structure of the objective function in (56) and (60) explicitly displays its two-criteria nature: it is required, on the one hand, to find a portfolio $(\boldsymbol{\beta}_t, t=1,...,N)$ that tracks the analyzed returns as closely as possible $\sum_{t=1}^{N}(r_t^{(p)} - \mathbf{r}_t^T\boldsymbol{\beta}_t)^2 \to \min$, and, on the other hand, to minimize the *a priori* unlikelihood of the portfolio dynamics $\sum_{t=1}^{N}(\boldsymbol{\beta}_t - \mathbf{V}_t\boldsymbol{\beta}_{t-1})^T \mathbf{U}_t(\boldsymbol{\beta}_t - \mathbf{V}_t\boldsymbol{\beta}_{t-1}) \to \min$. The balance between these two mutually contradictory requirements is controlled by the parameter $\lambda > 0$: the greater λ, the more preference is given to the latter goal at the price of partially sacrificing the former one.

It is hardly reasonable to take other kinds of matrices $\mathbf{U}_t$ than diagonal ones $\mathbf{U}_t = \mathrm{Diag}(u_t^{(1)},...,u_t^{(n)},u_t^{(n+1)})$, $u_t^{(i)} \geq 0$:

$$\mathbf{U}_t = \begin{pmatrix} u_t^{(1)} & 0 & \cdots & 0 & 0 \\ 0 & u_t^{(2)} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & u_t^{(n)} & 0 \\ 0 & 0 & \cdots & 0 & u_t^{(n+1)} \end{pmatrix}. \quad (61)$$

It is well seen from (56) and (60) that the minimum point of the objective function depends on the products $\lambda\mathbf{U}_t$ but not on the individual values of λ and $\mathbf{U}_t$; therefore, only ratios between values of $u_t^{(i)}$ in each matrix are of significance. As a rule, it is enough to retain the same matrix $\mathbf{U}_t=\mathbf{U}$ over the entire time interval $u_t^{(i)}=u^{(i)}$. An exception is the need to determine structural breaks in factor exposures, which is considered below in Section 7.

When matrices $\mathbf{U}_t$ are fixed, the value of λ in (60) determines a class of models in which a specific model $(\hat{\boldsymbol{\beta}}_0,\hat{\boldsymbol{\beta}}_1,...,\hat{\boldsymbol{\beta}}_N)$ is to be chosen by minimizing the respective objective function. From this point of view, the choice of λ is the choice of a class of models, such that the smoothness degree of the succession of factor exposures would be most adequate to the observed time series $\left((r_1^{(p)},\mathbf{r}_1),...,(r_N^{(p)},\mathbf{r}_N)\right)$.

If the value of smoothness parameter λ is taken too small, each local final estimate of the hidden vector of factor exposures $\hat{\boldsymbol{\beta}}_t$ will be inferred from a too-small number of neighboring observations $(r_{t\pm\tau}^{(p)},\mathbf{r}_{t\pm\tau})$ at the left and the right of the current time moment, which will result in a noisy succession of estimated vectors of factor exposures $(\hat{\boldsymbol{\beta}}_0,\hat{\boldsymbol{\beta}}_1,...,\hat{\boldsymbol{\beta}}_N)$, because local disturbances will be insufficiently suppressed. On the other hand, too large values of the smoothing parameters will suppress natural oscillations of the hidden succession of factor exposures along with suppressing the observation noise.

Typically, researchers present solutions of a large number of optimizations for various values of parameters, where such parameters belong to a multi-dimensional grid. The results $(\hat{\boldsymbol{\beta}}_0,\hat{\boldsymbol{\beta}}_1,...,\hat{\boldsymbol{\beta}}_N)$ are then visually evaluated for consistency. For example, in [28] the authors take three values 10^{-3}, 1 and 10^3 for adjusting one smoothing parameter.

The most commonly used measure of regression model fit is its coefficient of determination R^2. This coefficient was computed by Sharpe in his work [13] as the proportion of the portfolio volatility explained by systematic exposures using the moving window technique (Section 2.2). In the terms of FLS (56) and CFLS (60) criteria, the coefficient of determination will be expressed as the ratio

$$R^2=\frac{\sum_{t=1}^{N}(r_t^{(p)})^2-\sum_{t=1}^{N}(r_t^{(p)}-\mathbf{r}_t^T\hat{\boldsymbol{\beta}}_t)^2}{\sum_{t=1}^{N}(r_t^{(p)})^2}=1-\frac{\sum_{t=1}^{N}(r_t^{(p)}-\mathbf{r}_t^T\hat{\boldsymbol{\beta}}_t)^2}{\sum_{t=1}^{N}(r_t^{(p)})^2}. \tag{62}$$

However, by shortening the window size it is easy to drive R^2 to 100% and, at the same time, obtain highly volatile, meaningless exposure estimates. Likewise, in the FLS and CFLS approach, the more weight that is put on the fit-estimation term, the higher is R^2, so that R^2=100% can be obtained easily by assigning a very small value to parameter λ, which will result in perfect model fit. Therefore, R^2 is not an appropriate measure of the model adequacy.

The major reason for such inadequacy of the R^2 statistic is that it uses the same data set for both estimation and verification of the model, whereas for the purpose of evaluating the ability of a model to predict (forecast), the observation data sample has to be split into two sets – the estimation set and a test set. However, this way is applicable only when the amount of available data is twice as large as is required for reliable estimation and reliable verification.

In the next Section, we consider an alternative method of verification of the model and choosing its parameters that is based on a trade-off between two mutually contradictory requirements to the size of the data set – to be large enough as for estimation as well as for testing.

6.2 The Cross Validation (CV) principle of model adequacy verification

One of the most commonly used versions of the *Cross Validation* method was suggested by Allen [31] under the name of *Prediction Sum of Squares* (PRESS) statistic. According to this method, an observation is removed from the sample, the model is evaluated on the remaining observations, and the prediction error is calculated on the removed observation. This procedure is then repeated for each observation in the sample, and the sum of squared errors is computed.

The Cross Validation principle is widely adopted in data analysis [32,33,34], including pattern recognition, where it is known under the name of the "leave-one-out" procedure [35].

As applied to verification of the accuracy of an FLS or CFLS model, the essence of the Cross Validation principle can be explained as the idea to assess the adequacy of the given model by estimating the variance of the residual noise $D(e)$ in (13)-(14) and comparing it with the full variance of the goal variable $D(r) = (1/N)\sum_{t=1}^{N}(r_t^{(p)})^2$. But, when computing the error at a time moment t, it is incorrect to use the estimate $\hat{\boldsymbol{\beta}}_t$ obtained by minimizing the criterion (56) or (60) with participation of the observation at this time moment $(r_t^{(p)}, \mathbf{r}_t)$, because the resulting value

$$\hat{D}(e) = \frac{1}{N}\sum_{t=1}^{N}(r_t^{(p)} - \mathbf{r}_t^T \hat{\boldsymbol{\beta}}_t)^2$$

will inevitably be less than the actual noise variance $D(e)$. The CV principle leads to the following procedure that provides a correct estimate of the observation noise.

In the full time series $\big((r_1^{(p)}, \mathbf{r}_1), \ldots, (r_N^{(p)}, \mathbf{r}_N)\big)$, single elements $t = 1, \ldots, N$ are skipped one by one $\big((r_1^{(p)}, \mathbf{r}_1), \ldots, (r_{t-1}^{(p)}, \mathbf{r}_{t-1}),\ (r_{t+1}^{(p)}, \mathbf{r}_{t+1}), \ldots, (r_N^{(p)}, \mathbf{r}_N)\big)$, each time by replacing the sum $\sum_{t=1}^{N}(r_t^{(p)} - \mathbf{r}_t^T \boldsymbol{\beta}_t)^2$ in (56) or (60) by the truncated sum $\sum_{s=1,\, s\neq t}^{N}(r_s^{(p)} - \mathbf{r}_s^T \boldsymbol{\beta}_s)^2$, and the optimal vector successions $(\hat{\boldsymbol{\beta}}_1^{(t)}, \ldots, \hat{\boldsymbol{\beta}}_N^{(t)})$ are found, where the upper index (t) means that the observation $(r_t^{(p)}, \mathbf{r}_t)$ was omitted when computing the respective estimate. For each t, the instantaneous squared prediction error is calculated using the respective single estimate $\big(r_t^{(p)} - (\hat{\boldsymbol{\beta}}_t^{(t)})^T \mathbf{r}_t\big)^2$. The cross-validation estimate of the noise variance is found as the average over all the local squared prediction errors

$$\hat{D}_{CV}(e) = \frac{1}{N}\sum_{t=1}^{N}\Big(r_t^{(p)} - (\hat{\boldsymbol{\beta}}_t^{(t)})^T \mathbf{r}_t\Big)^2. \tag{63}$$

The less $\hat{D}_{CV}(e)$, the more adequate is the model with the given succession of smoothing matrices $(\mathbf{U}_1, \ldots, \mathbf{U}_N)$ to the observed time series $\big((r_1^{(p)}, \mathbf{r}_1), \ldots, (r_N^{(p)}, \mathbf{r}_N)\big)$.

The cross-validation estimate of the residual noise variance $\hat{D}_{CV}(e)$ can be further scaled to make it comparable across different analyzed portfolios or instruments. We suggest the cross-validation statistic

$$PR^2 = \frac{D(r) - \hat{D}^{(1,\ldots,N)}(e)}{D(r)} = 1 - \frac{\hat{D}^{(1,\ldots,N)}(e)}{D(r)} \tag{64}$$

which we call *Predicted R-squared*. Note that it is computed similarly to the regression R-squared statistic (62).

6.3 Adjustment of the smoothness level

Note that the cross validation statistic $\hat{D}_{CV}(e)$ discussed above is a function of smoothing parameter λ, such that $\hat{D}_{CV}(e)=\hat{D}_{CV}(e\,|\,\lambda)$. Choosing different values of λ for the objective function (56) or (60) results, in general, in different solutions, different predictions and therefore, different prediction errors.

We suggest a method of determining optimal model parameters that consists in processing the given time series $\left((y_1,\mathbf{z}_1),...,(y_N,\mathbf{z}_N)\right)$ several times with different tentative values of λ. Each time, the model adequacy is assessed by the averaged squared prediction error (63) estimated by the cross validation procedure. The value λ^* that yields the minimum prediction error is taken as the smoothing parameter recommended for the given time series:

$$(\lambda^*)=\arg\min_{\lambda}\hat{D}_{CV}(e\,|\,\lambda).$$

It should be noted that the selection of model parameters through minimizing the prediction error makes this method a version of the James-Stein estimator producing the smallest prediction error [32].

6.4 Application to the two-asset model portfolio

We will now apply the methodology developed above to the model two-asset portfolio (12) described in Section 2 and shown in Figure 1. The analysis was performed using three different values of the smoothness parameter λ= 10, 1, 0.1. The results are shown in Figures 4, 5, and 6. Note that the last analysis with $\lambda=0.1$ provides a nearly identical replication of the model portfolio.

Below in Table 1 we present the levels of both R-squareds: the standard coefficient of determination and the Predicted R-squared for the sine-wave two-asset test. The value of smoothness equal to 0.1 provides the best solution based on the Predicted R-squared value, which corresponds to what we observed from Figure 6.

Cross validation methodology described in the current Section provides the framework for selecting optimal smoothness parameters. We have defined the optimal smoothing parameter λ for a given model as the one providing the smallest prediction error or, equivalently, the highest Predicted R^2. Note that in the example above the most accurate result corresponded to the model with the highest Predicted R^2. At the same time, the number of computations performed for different values of λ was relatively small. The simplest way to find a global optimal parameter value is to perform a large number of such computations over a certain grid of values of parameter λ and select the one producing the highest Predicted R^2. An alternative would be to use an iterative gradient method to find the optimal smoothness parameter.

In Figure 7, we show a chart with Predicted R^2 values for 100 solutions of the two-asset dynamic model problem in Figure 1 obtained for 100 values of the smoothness parameter λ. We plot parameter values along the logarithmic axis μ, so that $\lambda=10^{(10\mu-5)}$ and $0\le\mu\le1$. The maximum Predicted R^2 value is attained at the log-smoothness value $\mu=0.3$ corresponding to $\lambda=0.01$. Note that the global maximum is not very pronounced, and there is a range of smoothness values between 0.2 and 0.4 producing about the same result.

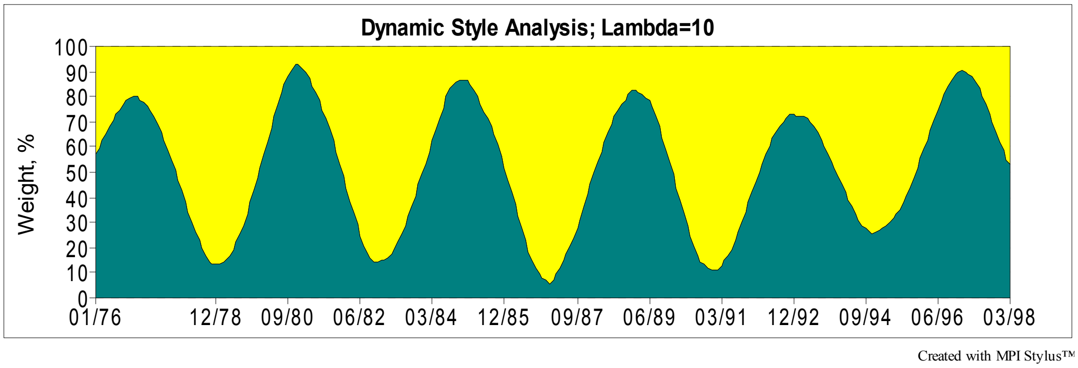


Figure 4. Two-asset model: λ=10.

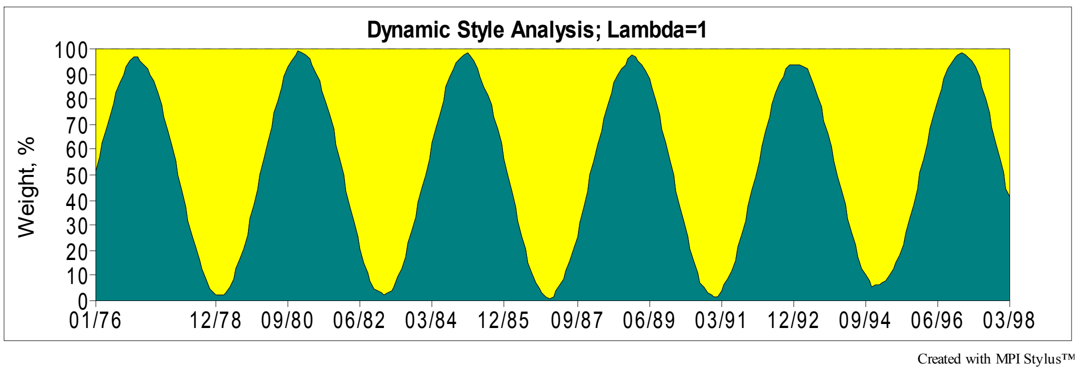


Figure 5. Two-asset model: λ=1.

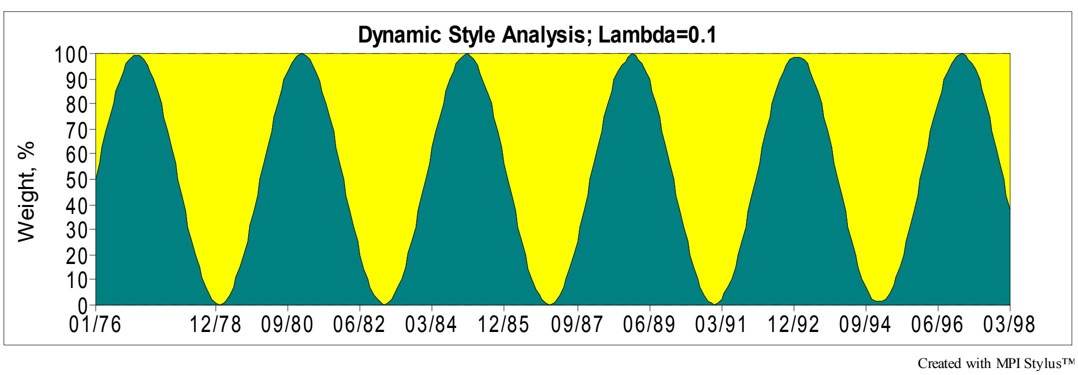


Figure 6. Two-asset model: λ=0.1.

Table 1. Two-asset model. Predicted R-squared values.

Smoothness	R-squared	Predicted R-squared
10	98.89%	97.74%
1	99.97%	99.73%
0.1	100.00%	99.95%

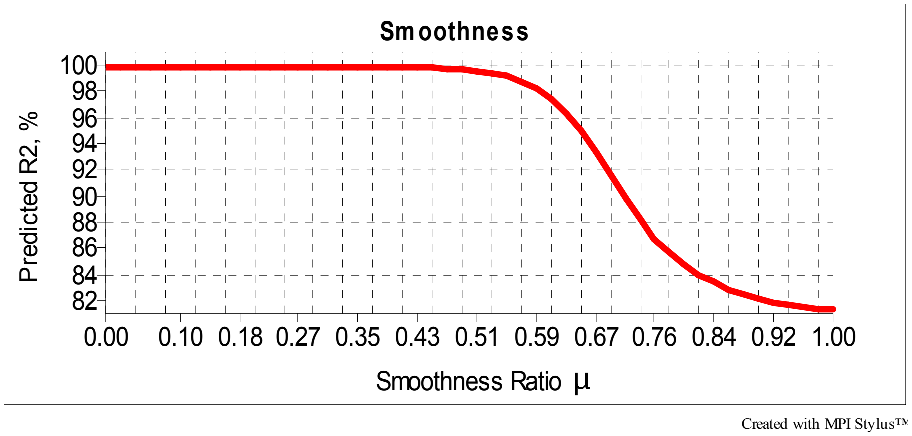


Figure 7. Two-asset model: Smoothness parameter selection.

7 Determining structural breaks in factor exposures

7.1 Structural changes in dynamic models

One of the most important applications of dynamic models lies in determining points of abrupt changes in factor exposures. For example, for anyone analyzing investment portfolios, the knowledge about timing and the scale of such changes is very important because such breaks can be caused by massive and rapid changes in positions due to a change in strategy or in management. It is especially important in analyses of hedge funds because the latter, not having any restrictions on using derivatives, can change the entire portfolio virtually overnight.

For the reasons outlined earlier, the traditional window-based methodology is unable to detect abrupt changes in investment style and/or asset allocation. When such a change in portfolio structure occurs within a few data points, the current estimation technique indicates a gradual change spanning all the estimation windows containing the points of the structural change. To illustrate this concept we use the same example as the one presented in [36] where the authors created a portfolio of two assets: Russell 1000 Growth Index and Russell 1000 Value Index, with their weights representing a "step" function as shown in Figure 8.

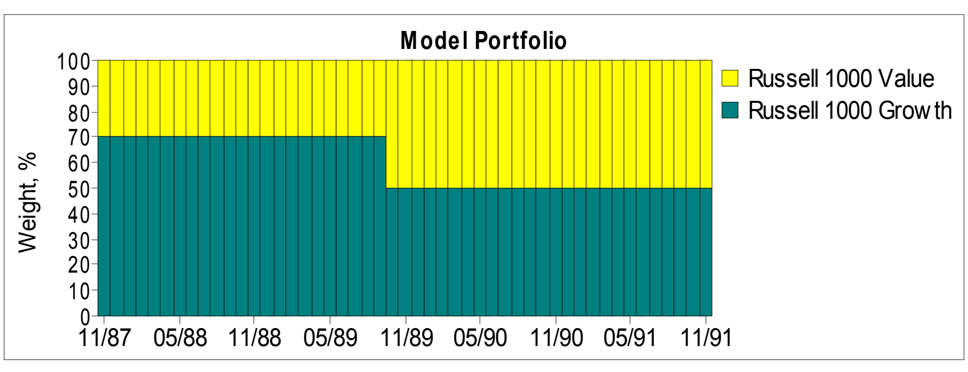


Figure 8. Two-asset model with a "step".

The weights of the model portfolio are defined as follows: first 23 months from Nov-87 through Sep-89 the portfolio maintains a constant 70/30 percentage breakdown between Growth and Value indices, and starting from Oct-89 through Nov-91 maintains a 50/50 allocation to the indices. Similar to the example in the previous Section, a composite return series was created using the given weights and monthly index returns and then used Sharpe's RBSA with the same Growth and Value indices that were used in its construction to estimate original weights. Noise was added neither to the weights nor to the composite return series. The resulting effective mix for a 24-month estimation window is shown in Figure 9.

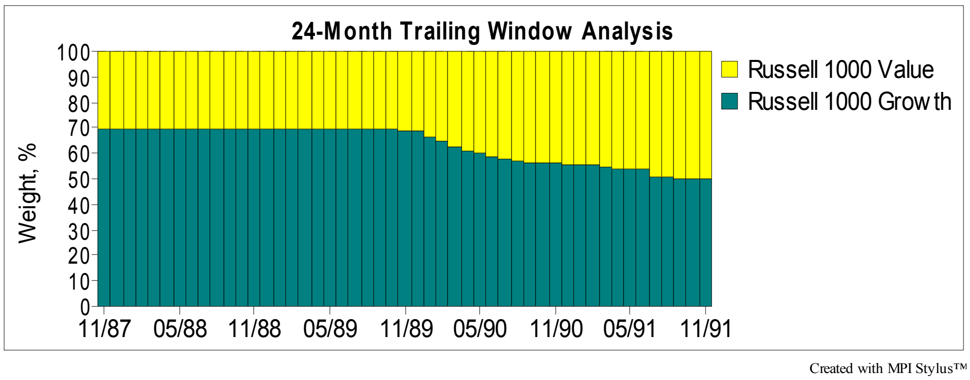


Figure 9. Two-asset "step" model: 24-month RBSA result.

The result is indicative of the problems inherent in the moving window RBSA: the abrupt change in the strategy is depicted as gradually changing allocations with the correct 50/50 split happening only 2 years after the fact, when the entire estimation window is eventually located in the 50/50 portion of the composite return history.

7.2 The jump map and Structural Breakpoint Ratio (SBR)

The method of preserving structural breaks in the primarily smooth path of factor exposures, which is proposed in this work, is based on the fact that if the parameters $u_t^{(i)}$ of the smoothness matrices $\mathbf{U}_t$ (61) take essentially smaller values at a single time moment than over the surrounding interval

$$u_t^{(i)} \ll u_s^{(i)} \text{ at } s<t \text{ and } s>t ,$$

the factor exposures $\boldsymbol{\beta}_{t-1}$ and $\boldsymbol{\beta}_t$ will be estimated almost independently of each other. In particular, they will be estimated completely independently and completely preserve, thereby, the eventual break, if $u_t^{(i)}=0$.

Let the positions of breaks $T^*=\{t_j^*,\ j=1,...,k\}$ be called the jump map. If we knew the jump map, we should take the succession of smoothness matrices $\mathbf{U}_t=\mathrm{Diag}(u_t^{(1)},...,u_t^{(n)},u_t^{(n+1)})$ with much lesser values $u_t^{(i)}$ at $t\in T^*$ than at other points $t\notin T^*$, and solve the problem (56) or (60). At the moments of breaks legitimated in the jump map, the relatively small diagonal elements of $\mathbf{U}_{t^*}$ will prevent, at least partially, the averaging of the factor exposures. But the number and positions of eventual jumps are unknown in the general case, and the jump map is subject to estimation along with the factor exposures.

Let $u^{(i)}$ be the basic parameters of the smoothness matrices that guarantee the "normal" relatively high degree of smoothing the factor exposures, and $u^{*(i)}<u^{(i)}$ be sufficiently lower values

that are meant to provide preserving eventual breaks. Further, let T^* be the assumed jump map, and

$$\mathbb{U}_{T^*} = \left(\mathbf{U}_t = \text{Diag}(u_t^{(1)},...,u_t^{(n)},u_t^{(n+1)}) \text{ at } t \notin T^*;\ \mathbf{U}_t = \text{Diag}(u_t^{*(1)},...,u_t^{*(n)},u_t^{*(n+1)}) \text{ at } t \in T^* \right)$$

be the respective succession of smoothness matrices. If we extend the jump map by an extra break $\tilde{T}_t^* = T^* \bigcup \{t\}$ at some point $t \notin T^*$, the minimum of the criterion (56) or (60) $J(\hat{\boldsymbol{\beta}}_t, t = 0,1,...,N \mid \mathbb{U}_{\tilde{T}_t^*})$ computed with the extended jump map will be less that the minimum of the initial criterion $J(\hat{\boldsymbol{\beta}}_t, t = 0,1,...,N \mid \mathbb{U}_{T^*})$. We shall call the value

$$\rho_t = \frac{\min\limits_{\boldsymbol{\beta}_t,\, t=0,1,...,N} J(\boldsymbol{\beta}_t, t = 0,1,...,N \mid \mathbb{U}_{T^*})}{\min\limits_{\boldsymbol{\beta}_t,\, t=0,1,...,N} J(\boldsymbol{\beta}_t, t = 0,1,...,N \mid \mathbb{U}_{\tilde{T}^*})} \geq 1 \qquad (65)$$

the *Structural Breakpoint Ratio* (SBR). The greater ρ_t, the more likely is the hypothesis of the presence of a not yet legitimated break at the pair of adjacent points $(t-1, t)$.

The values ρ_t appear to be appropriate measures of the local "strain" in the estimated hidden process when the model is attempting to adjust it to the observed time series without jump at the current pair of points. A visual or analytical analysis of SBR values can be used to determine one or several possible breaks. Specifically, the most stressed pair of adjacent points (t^*-1, t^*) should be considered as the position of a possible structural shift if the maximal SBR value exceeds a preset threshold $h > 1$:

$$t^* = \arg\max_{t=1,...,N,\, \rho_t > h} \rho_t . \qquad (66)$$

If the condition $\rho_t > h$ is not met at any time moment, then there are no additional breakpoints in the succession of factor exposures.

7.3 Multistage estimation of the jump map

In this work, we propose a method of multistage estimation of the jump map. Let T^* be known up to a number of breaks which still remain to be found. Let us compute SBR ρ_t (65) sequentially for all the points $t = 1,...,N$ and check the condition (66). Note that this condition may be met only at a point where no jump had been found at the previous stages of processing, so, $t \notin T^*$, and the matrix $\mathbf{U}_t$ in $\mathbb{U}_{T^*}$ is still a matrix with large diagonal elements $\mathbf{U}_t = \text{Diag}(u^{\beta},...,u^{\beta},u^{\alpha})$ which present the normal smoothing degree of the hidden process. This point is to be added to T^*, i.e. the "old" matrix $\mathbf{U}_t$ is to be replaced in $\mathbb{U}_{T^*}$ by the matrix $\mathbf{U}_t = \text{Diag}(u^{*\beta},...,u^{*\beta},u^{*\alpha})$ with small diagonal values to mark the position where the smoothness is to be broken.

The procedure starts with the initial empty jump map T^* which presupposes no jumps, and is to be repeated again and again, each time with renewed T^* and $\mathbb{U}_{T^*}$, while the most stressed point satisfies the condition (66).

7.4 Application to the model "step" portfolio

As an example of application of the breakpoint methodology developed in the current Section, we present below results of the analysis for the model two-asset step portfolio shown in Figure 8. First, we perform analysis of the composite return series and determine the optimal smoothness, using an algorithm described in previous Section. The result of this analysis is shown in Figure 10.

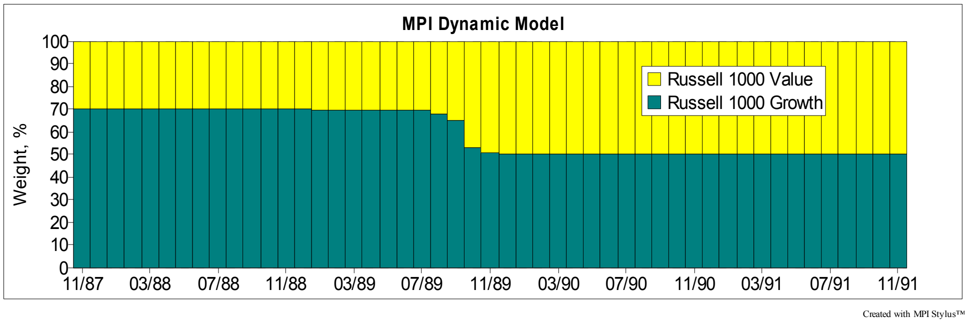


Figure 10. The "step" portfolio: CFLS analysis with optimal smoothness.

Note that we already get significant improvement as compared to the moving window RBSA – the timing of the shift is placed correctly in Oct-89 and the transition takes only a few months vs. years for RBSA (compare to Figure 9). Next, we will run a breakpoint analysis for the entire time interval and present results in Figure 11.

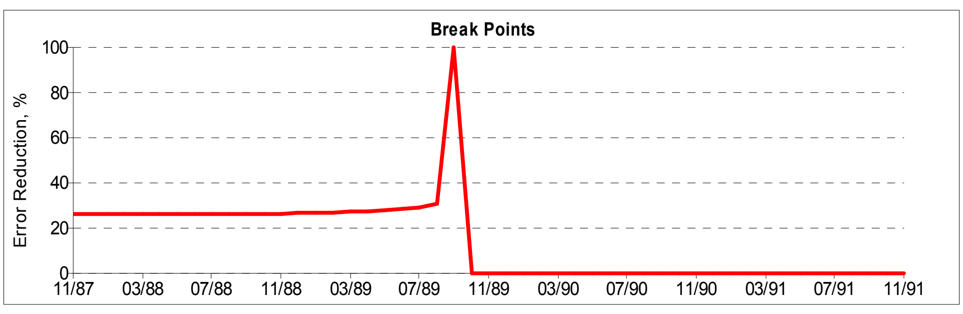


Figure 11. The "step" portfolio: breakpoint analysis.

On the Y-axis of this chart we plot the percentage of the objective improvement relative to the original CFLS performed on the first step. The improvement in the objective is measured by relaxing the smoothness penalty in each month shown on the X-axis and performing the CFLS with the new objective. The breakpoint function attains maximum in Oct-89 where the objective is decreased by 100%. Evidently, this means that by allowing unlimited turnover in this point we attain the perfect fit, where the corresponding "relaxed" CFLS results in the following effective mix (Figure 12) which is a perfect replication of the original model.

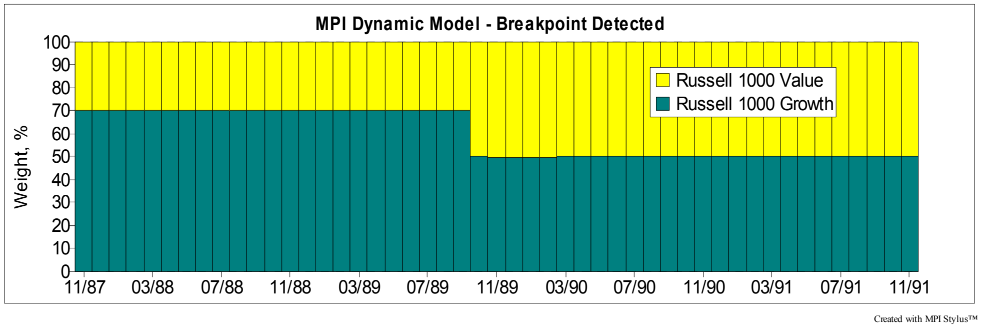


Figure 12. The "step" portfolio: CFLS analysis with relaxed fit.

8 Applications

Below we present several examples of applications of the dynamic multi-factor methodology developed in previous sections to real-life portfolios using public data from U.S. financial markets. The analyses presented below involve mutual funds, representing portfolios of financial instruments that are publicly available for investment by individuals as well as institutions such as public and corporate pension funds, foundations and endowments. As of February 2004, there were over 9,000 mutual funds with about 86 million investors and $7.6 trillion in assets[1]. Because of their huge importance for the US investment industry and pensions, mutual funds are tightly regulated by the U.S. government bodies such as the Security and Exchange Commission (SEC)[2]. One of the aspects of this regulation is the requirement for mutual funds to disclose their holdings, i.e., listings of securities held in portfolios and their respective market values to the SEC on a regular basis – quarterly and/or semiannually. A historical collection of such reports is then provided by the SEC to the public in the form of a searchable database called EDGAR. Mutual fund prices are publicly available from financial publications such as *The Wall Street Journal*, *Barron's*, etc., and online resources such as Yahoo, Inc. at http://finance.yahoo.com .

Our analysis below is presented as follows: we are using monthly or daily mutual fund returns, i.e., relative periodic price changes, as inputs for our model to recover hidden information about dynamics of the portfolio structure, e.g., changes in its asset and/or economic sector allocations. These returns are the only piece of information pertaining to mutual funds that is used in our analyses. We then compare the results of our analysis for accuracy with the actual allocation information obtained from the SEC and other public sources.

8.1 Fidelity Magellan Fund

8.1.1 RBSA with monthly data 1995-1996

Fidelity Magellan Fund is the largest and the most prominent fund in the U.S. mutual fund industry. Most of its success is attributed to Peter Lynch, a legendary money manager and stock picker. In October 1995, Jeffrey Vinnik, then current manager of the largest $50B fund, sharply

[1] Source: "Trends in Mutual Funds Investing February 2004", *Investment Company Institute* (ICI), Washington, DC, March 30, 2004, www.ici.org. The Investment Company Institute (ICI) is the national association of the U.S. investment company industry.

[2] U.S. Securities and Exchange Commission, www.sec.gov .

cut its holdings in stocks – technology stocks in particular[1]. More stock sell-off continued through mid-1996, and the fund performance suffered as a result because the market didn't fall as the manager had expected, and in June 1996 Vinnik resigned. The new fund manager Robert Stansky quickly liquidated bond positions and by the end of 1996, Magellan Fund was back over 90% in stocks.

Allocation history of the fund obtained from public sources[2] is presented in Table 2 and in Figure 13 below:

Table 2. Fidelity Magellan: Publicly Available Allocation Data.

	Sep-94	Mar-95	Sep-95	Nov-95	Feb-96	Mar-96	Apr-96	Jul-96	Aug-96	Sep-96	Mar-97
Cash-other	2.6%	1.3%	3.2%	6.7%	15.6%	10.7%	9.90%	1.6%	4.3%	1.2%	3.9%
Bonds	0.9%	0.4%	2.9%	11.6%	19.4%	19.0%	19.20%	15.6%	11.8%	9.8%	0.1%
Stocks	96.5%	98.3%	93.9%	81.7%	65.0%	70.3%	70.80%	82.8%	83.9%	89.0%	96.0%

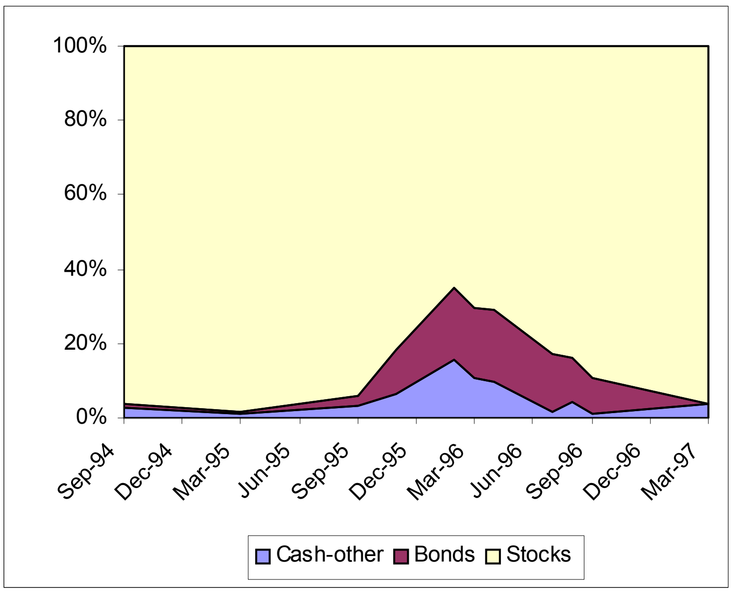


Figure 13. Fidelity Magellan: Public Data.

The chart provides a convenient view of the allocation history of the fund with the quarterly time-line represented by the X-axis. The asset allocations of the fund for each time period are "stacked" over each other along the Y-axis adding up to 100% and, therefore, filling the entire area of the chart.

In Table 33 we show the list of publicly available generic asset indices[3], which we use for the analysis of the Magellan fund. For example, Russell Indices represent the top 3000 largest US stocks divided into three groups according to their size (sorted by market capitalization): top 200,

---

[1] "Market Place: Magellan Shifted From Technology in November", The New York Times, January 12, 1996.

[2] Source: www.fidelity.com in 1995-1997. EDGAR database at www.sec.gov.

[3] Source of monthly returns: Fidelity Magellan Fund – Morningstar, Inc.; Russell Equity Indices - Frank Russell Company, www.russell.com; MSCI EAFE Index – Morgan Stanley Capital International, www.msci.com; Lehman Aggregate Bond Index – Lehman Brothers, www.lehman.com; Cash Index – 90-day T-Bill Index (Merrill Lynch Fixed Income Research), www.ml.com.

mid 800 and smaller 2,000 stocks. Each of these groups is then divided according to certain criteria into two groups of so-called growth and value stocks and the resulting 6 portfolios are then "capitalization-weighted", i.e., each stock in each portfolio is getting weight proportional to its market size, or capitalization.

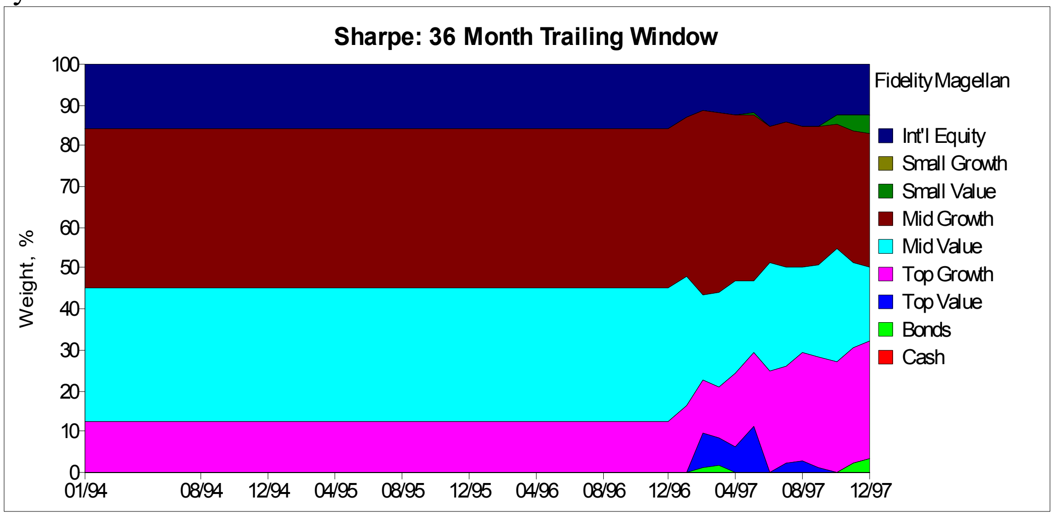
Table 3. Asset Indexes.

	Asset	Index	Description
1	Large Growth U.S. Stocks	Russell Top 200 Growth Index	An index of growth oriented stocks among the 200 largest U.S. stocks.
2	Large Value U.S. Stocks	Russell Top 200 Value Index	An index of value oriented stocks among the 200 largest U.S. stocks.
3	Mid Cap Growth U.S. Stocks	Russell Midcap Growth Index	An index of growth oriented stocks among the next 800 largest U.S. stocks.
4	Mid Cap Value U.S. Stocks	Russell Midcap Value Index	An index of value oriented stocks among the next 800 largest U.S. stocks
5	Small Cap Growth U.S. Stocks	Russell 2000 Growth Index	An index of growth oriented stocks among the next 2000 largest U.S. stocks.
6	Small Cap Value U.S. Stocks	Russell 2000 Value Index	An index of value oriented stocks among the next 2000 largest U.S. stocks.
7	Foreign Stocks	MSCI EAFE Index	A broad equity market benchmark of developed countries outside U.S. Covers hundreds of issues in 20 European and Asian country markets.
8	U.S. Bonds	Lehman Aggregate Bond Index	A broad US Fixed Income benchmark consisting of about 7,000 issues: US Treasuries, Corporate Bonds, Asset and Mortgage-backed securities.
9	Cash – Liquidity	Cash Index	A 90-day Treasury Bill (T-Bill) Index.

Note that these indices represent thousands of securities, while the Magellan portfolio held only several hundred of these. For example, the fund could have held only several bond issues out of 7,000 issues covered by the Lehman Aggregate Bond Index. Moreover, for the same securities their weights in the fund portfolio could have been quite different from their weights in the corresponding index portfolio. Therefore, when trying to replicate the performance of the fund with just nine generic indices above, we are looking to get a pattern close to the one in the table and the chart above and not necessarily the exact match.

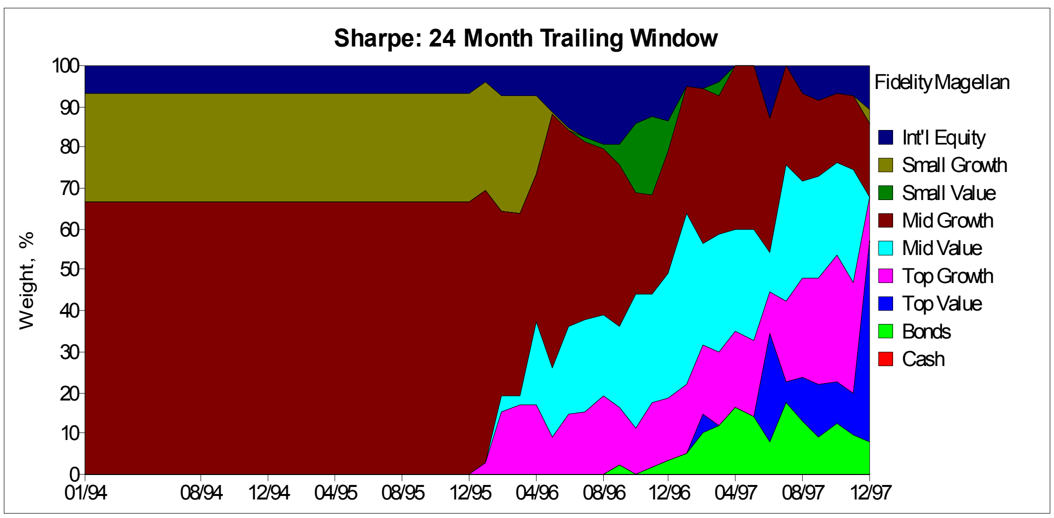
As the first step, we use Sharpe's trailing window RBSA model to analyze the dynamics of the Magellan portfolio in 1994-1997. The inputs to the model will be monthly returns of the Magellan fund and of the indices listed in Table 3 computed over 48 months from January 1994 to December 1997.

In practical applications the most typical window size for such analyses is usually taken equal to 36 or 24 months. It is intuitively clear that to detect such a short-time change in allocation that happened over 6-8 months, a shorter window has to be used. The three charts below in Figures 14, 15 and 16 represent results of such analysis using 36, 24 and 12-month moving windows respectively. Note constant allocations at the beginning of each analysis. They appear because the result of analysis in the first window is assigned to the rightmost (most recent) edge of the window and, therefore, no information is implied about the rest of the first window. By setting all values in the first estimation window to be constant, we allow readers to compare our further analyses over the same time scale.



Created with MPI Stylus™ (Data: Morningstar®)

Figure 14. Magellan RBSA, 36-month window.



Created with MPI Stylus™ (Data: Morningstar®)

Figure 15. Magellan RBSA, 24-month window.

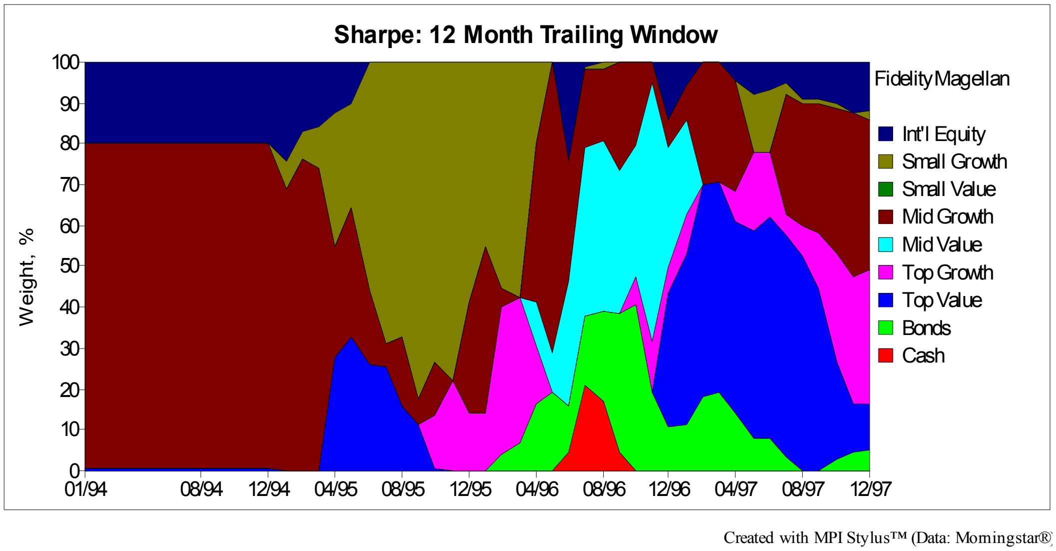


Figure 16. Magellan RBSA, 12-month window.

Note that allocations to indices are unrealistically volatile and the bond allocation is shifted about one-half of the window size to the right. For example, in the analysis with the 24-month window in Figure 15, the bond accumulation is shown in early 1997, about 12 months after the fact. The 36-month window analysis in Figure 14 doesn't show any signs of the 1995-1996 portfolio dynamics. The smaller the window size becomes, the closer is the result of the bond plus cash allocation to the public data, although due to high short-term correlation of equity indices, the results become very noisy.

Next, we will perform analysis using the same data and the CFLS model (60) with the budget and non-negativity constraints for different exponentially increasing values of the smoothness parameter λ=1, 10, 100, and 1000. The results of such analyses are presented in Figures 17-20 below.

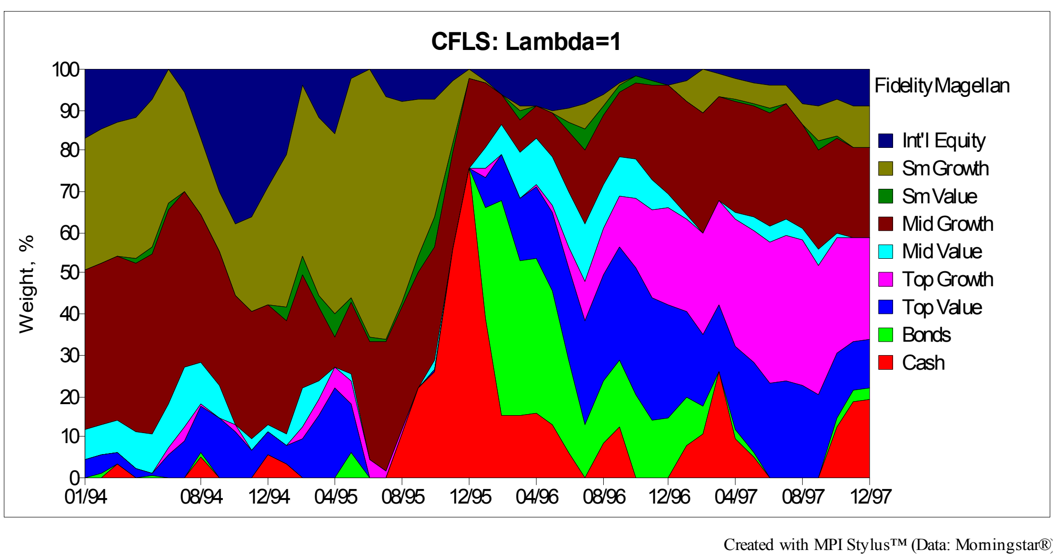


Figure 17. Magellan CFLS, λ=1.

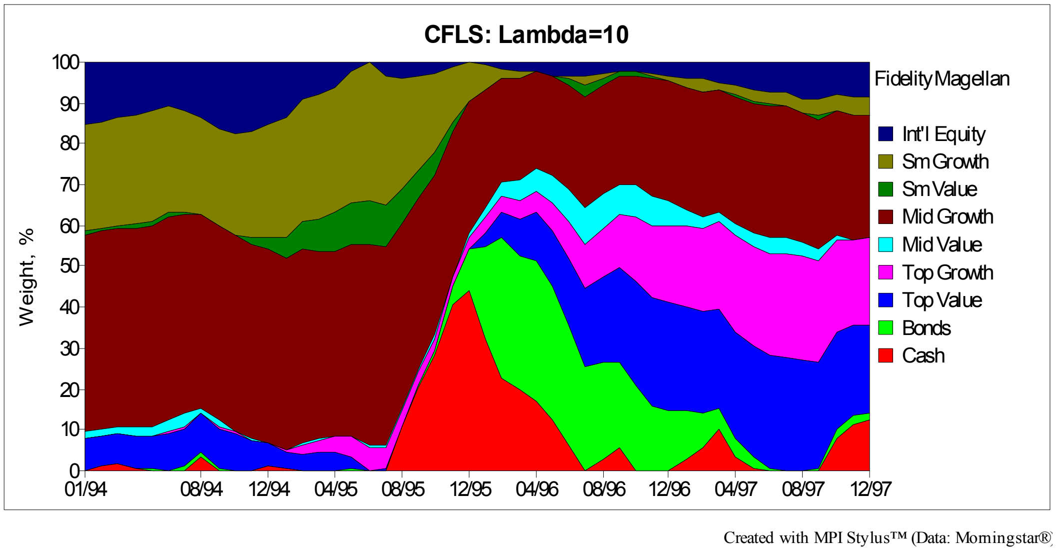


Figure 18. Magellan: CFLS, λ=10.

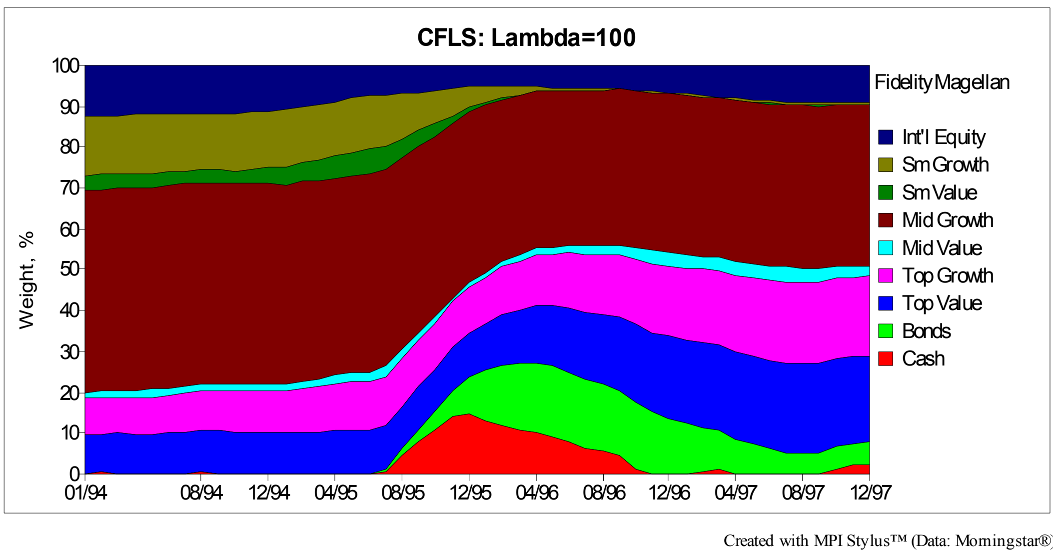


Figure 19. Magellan: CFLS, λ=100.

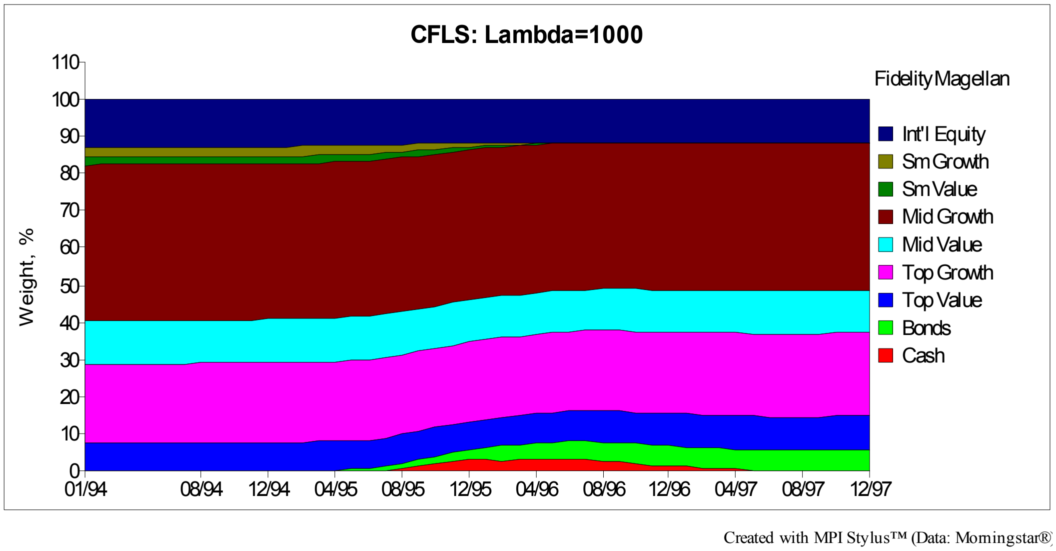


Figure 20. Magellan: CFLS, λ=1000.

Note that the quality of the analyses is notably higher for the CFLS model: the patterns are smoother and the timing of the bond accumulation is accurately placed at the end of 1995 with the peak bond plus cash position in the first quarter of 1996.

In order to determine the optimal value of the smoothness parameter, we performed 100 CFLS optimizations with varying values of the smoothness parameter, and for each result we computed the cross-validation statistic defined in Section 6.2 – the *Predicted R-squared* (64). The chart in Figure 21 shows the results of these computations, where the smoothness ratio μ provides a logarithmical scale for the smoothness parameter λ and is computed as $\lambda = 10^{10\mu - 5}$. The Predicted R-squared function has a very clear global maximum, which is attained in the point $\mu = 0.65$, what is equivalent to $\lambda = 31.62$.

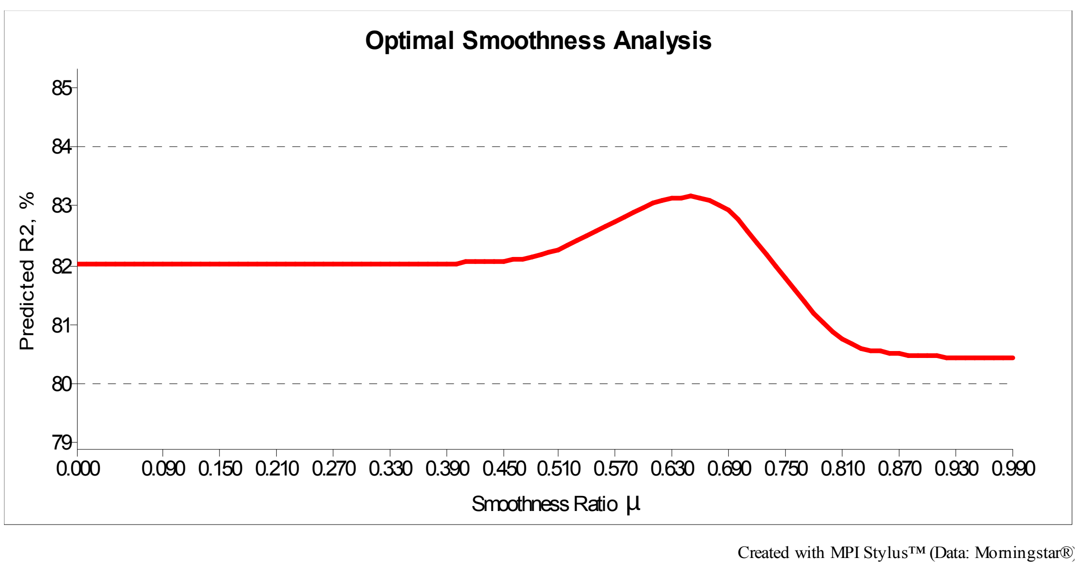


Figure 21. Magellan: CFLS, optimal smoothness analysis.

The results of the CFLS model corresponding to the point of the optimal smoothness are presented in Figure 22. There is a clear match of the bond plus cash position dynamics with the publicly available information presented above in Table 2 and Figure 13.

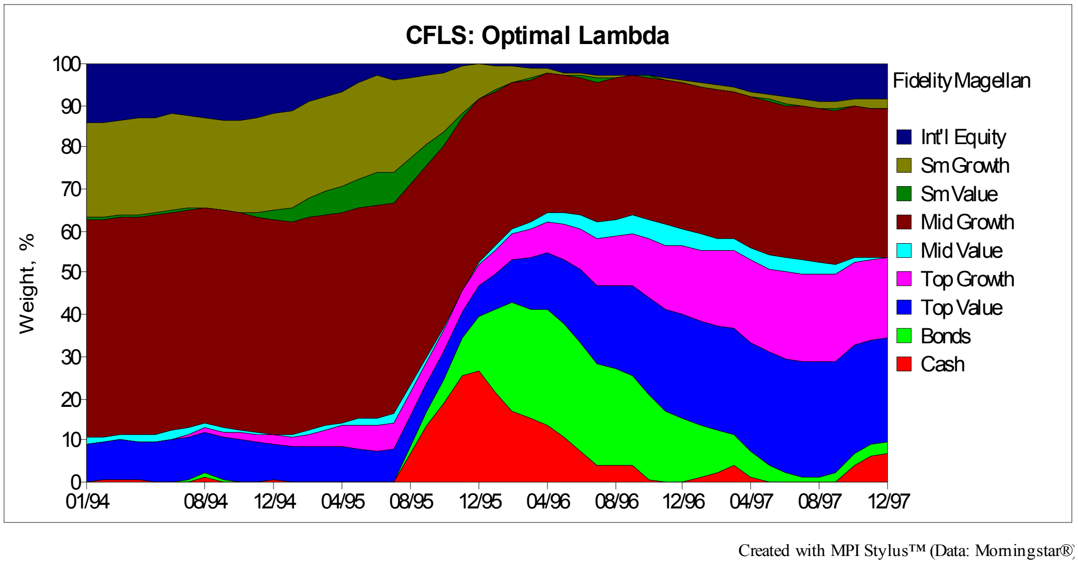


Figure 22. Magellan: CFLS with optimal smoothness.

8.1.2 RBSA with daily data

Our analysis of the Magellan fund in the previous Section was performed using monthly returns. To analyze short-term changes in fund allocations spanning one or two months, higher frequencies of data (daily or weekly) has to be used. Compared to monthly data, daily data provides over twenty times more observations within the same time interval and, therefore, allows one to use more asset indices. Below we are using the 13 industry sector indices listed in Table 4 rather than 9 size/style indices used in the previous section[1].

Table 4. Magellan analysis: Daily Indices.

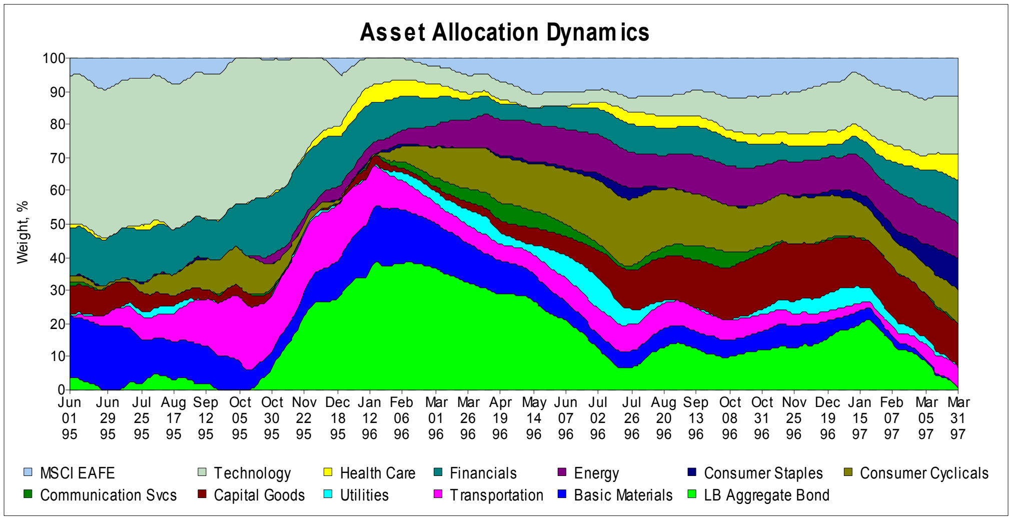
	Asset Index Name
1	S&P 500 Basic Materials Index
2	S&P 500 Transportation Index
3	S&P 500 Utilities Index
4	S&P 500 Capital Goods Index
5	S&P 500 Communication Services Index
6	S&P 500 Consumer Cyclicals Index
7	S&P 500 Consumer Staples Index
8	S&P 500 Energy Index
9	S&P 500 Financials Index
10	S&P 500 Health Care Index
11	S&P 500 Technology Index
12	Int'l Stocks – MSCI EAFE Index
13	U.S. Bonds – Lehman Aggregate Bond Index

Below we perform analysis of the same fund using daily data in order to determine the dynamics of Magellan's position in Technology stocks in October – November 1995 and compare the results with the official public data released almost 1.5 months after the fact. On January 11, 1996, Fidelity Investments announced[2] that the fund "cut its stake in technology stocks to 24.5 percent of its holdings at the end of November from 43.2 percent at the end of October." We are going to verify this fact by using 463 daily returns for the period June 1995 – March 1997. Given that the number of assets is 13, the total number of variables for a single CFLS optimization is equal to $463 \times 13 = 6{,}019$. The results of the optimization are shown in Figure 23. Note that both magnitude and timing of the bond positions match public data.

---

[1] Source of daily returns: Fidelity Magellan Fund – S&P/Micropal www.sandp.com; S&P Sector Indices – Vestek Systems, Inc. www.vestek.com (a Thomson Company); MSCI EAFE Index – Morgan Stanley Capital International, www.msci.com; Lehman Aggregate Bond Index – Lehman Brothers, www.lehman.com.

[2] The Reuters European Business Report: Fidelity Investment's Magellan Fund. *Reuters America Inc.*, January 11, 1996.



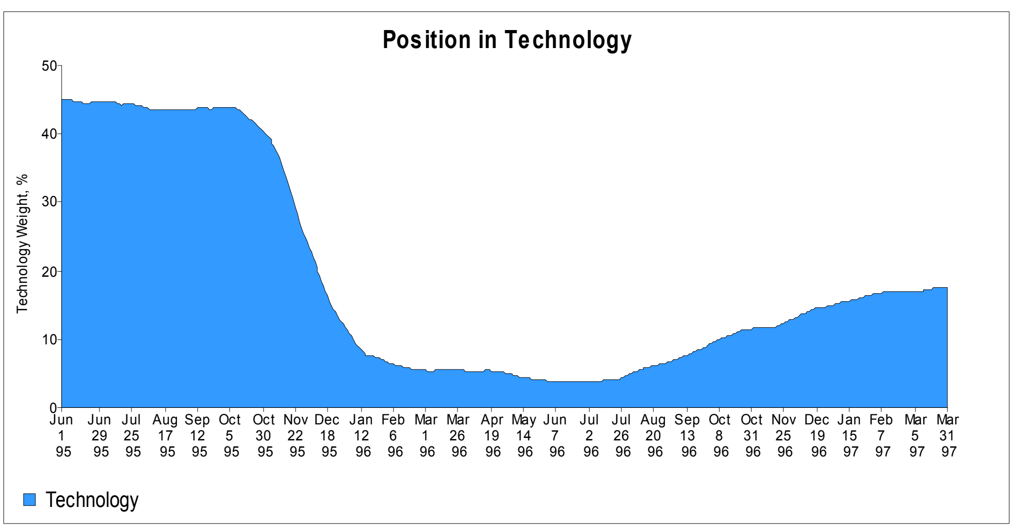
Created with MPI Stylus™ (Source: Micropal™ - www.micropal.com)

Figure 23. Magellan: Daily CFLS.

In Figure 24 we show the dynamics of the technology position determined by CFLS analysis. Note that the October-November change in this sector is captured very well.

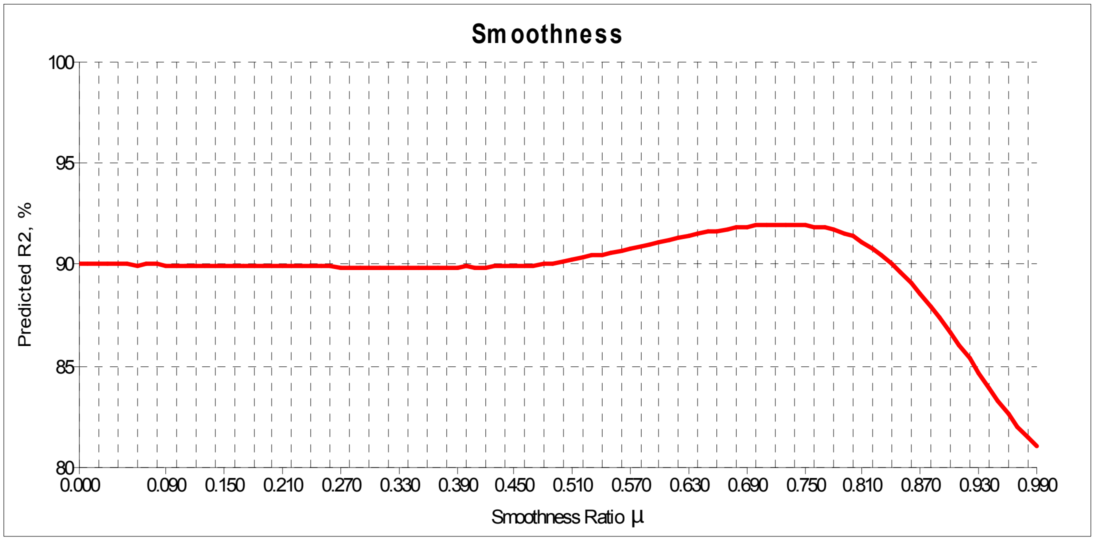
The Figure 25 shows the analysis of optimal smoothness, which is a result of 100×463=4,630 cross-validation CFLS optimizations. The global optimum is achieved at the value of $\mu = 0.72$, i.e. $\lambda = 158.5$.

The following two Figures 26 and 27 represent the result of the dynamic analysis described in the section 4.3.1.3. Note that while the overall trends remain unchanged, the paths of factor (sector) exposures became more volatile. This is due to the fact that the nonlinear model (41) is not imposing a penalty on the changes of weights caused by market prices fluctuations. Such an approach is clearly presenting a more realistic reflection of the portfolio structure.



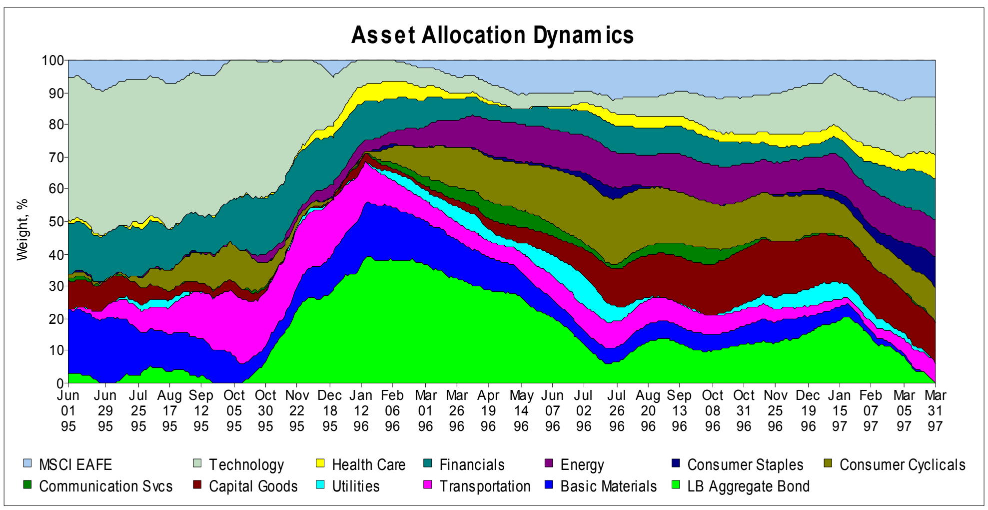
Created with MPI Stylus™ (Source: Micropal™ - www.micropal.com)

Figure 24. Magellan: Daily CFLS, technology position.



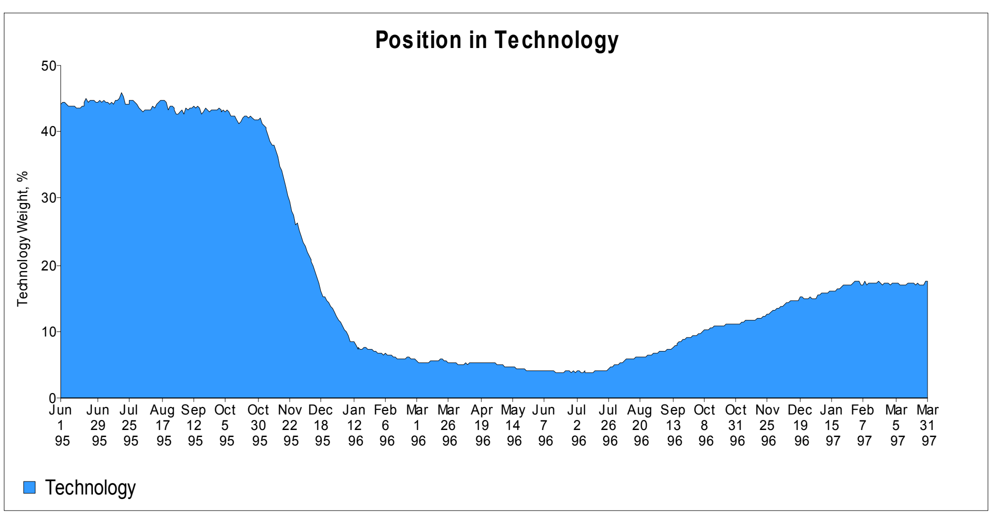
Created with MPI Stylus™ (Source: Micropal™ - www.micropal.com)

Figure 25. Magellan: Daily CFLS, smoothness analysis.



Created with MPI Stylus™ (Source: Micropal™ - www.micropal.com)

Figure 26. Magellan: Daily nonlinear CFLS.



Created with MPI Stylus™ (Source: Micropal™ - www.micropal.com)

Figure 27. Magellan: Daily nonlinear CFLS, technology position.

8.2 Dynamic Analysis of the Market Beta

One of the most important applications of the methodology developed in this paper lies in its ability to provide more accurate value of the CAPM market beta for both individual securities and investment portfolios. Although multi-factor models are much superior to single-factor CAPM, the latter is still being used in risk and performance reporting and evaluation[1]. There exist a number of methods to adjust security betas for their dynamic tendency such as, for example, Blume [9] and Vasicek [8] although we believe that our framework including both a cross-validation test for adjusting parameters and breakpoint detection technique presents a superior alternative. Below we present an application of the methodology developed in this paper to compute dynamic betas for individual stocks.

Market beta is usually computed vs. a diversified market index such as the S&P 500 Index and is regarded as a measure of relative risk of the stock. The higher the beta, the more sensitive is the stock to market fluctuations and the higher is the risk. Stocks with beta significantly lower than 1.0 are considered lower risk stocks, while stocks with beta significantly higher than 1.0 are considered as higher risk stocks. Accurate assessment of market beta is, therefore, an important factor influencing buying and selling decisions.

For our analysis we use Tyco International Ltd (ticker: TYC, www.tyco.com) common stock. Reuters www.reuters.com provides the following description of the company:

> "Tyco International Ltd. is a diversified manufacturing and service company that organizes its businesses into five segments. The Fire and Security segment designs, manufactures, installs, monitors and services electronic security and fire protection systems. The Electronics segment designs, manufactures and distributes electrical and electronic components, and designs, manufactures, installs, operates and maintains undersea fiber-optic cable communications systems. The Healthcare segment designs, manufactures and distributes medical devices and supplies, imaging agents, pharmaceuticals and adult incontinence and infant care products. The Engineered Products and Services segment designs, manufactures, distributes and services engineered products and provides environmental and other industrial consulting services. The Plastics and Adhesives segment designs, manufactures and distributes plastic products, adhesives and films."

For the analysis we use monthly total returns for the period of 1992-2003 presented in Figure 28 (Tyco returns are in red).

We perform a two-index analysis with the non-negativity constraint removed. The budget constraint makes such analysis equivalent to the single index CAPM, albeit dynamic. We demonstrate the results of our analysis on the three charts below. In Figure 29 we show TYC beta changing over time where the optimal smoothness of the analysis was determined using the point of global optimum in Figure 30. Note that we show beta multiplied by 100, as if it is a percentage weight in the S&P500 and T-Bills portfolio. In more common fractional terms, the analysis shows that TYC beta starts with 0.7 and reaches 1.9 plateau by the end of the data period in 2003.

The breakpoint analysis chart in Figure 31 is of the most interest. Two points of possible shift of the market perception of the Tyco stock stand out prominently: January 2000 and July 2002. Interestingly, both of them correspond to the most significant events in the company history. In November 1999, the SEC launched an investigation of the company's accounting practices, and in January 2000 Tyco management announced changes in its accounting of acquisitions, followed by a surge in revenues and the decision by the board to repurchase $2 billion of stock. Although the company's core business hasn't changed, these events created a boost in

[1] There exist a number of financial analytical services companies providing individual stock beta forecasts for a fee. One of such firms is BARRA, Inc. www.barra.com.

speculative trading, which potentially explains the first breakpoint. In July 2002, the Tyco board named a new chief executive to replace former CEO Dennis Kozlowski, who was indicted on charges of sales-tax evasion. The news came as a relief to wary investors and boosted their confidence, which clearly reflected on the stock beta.

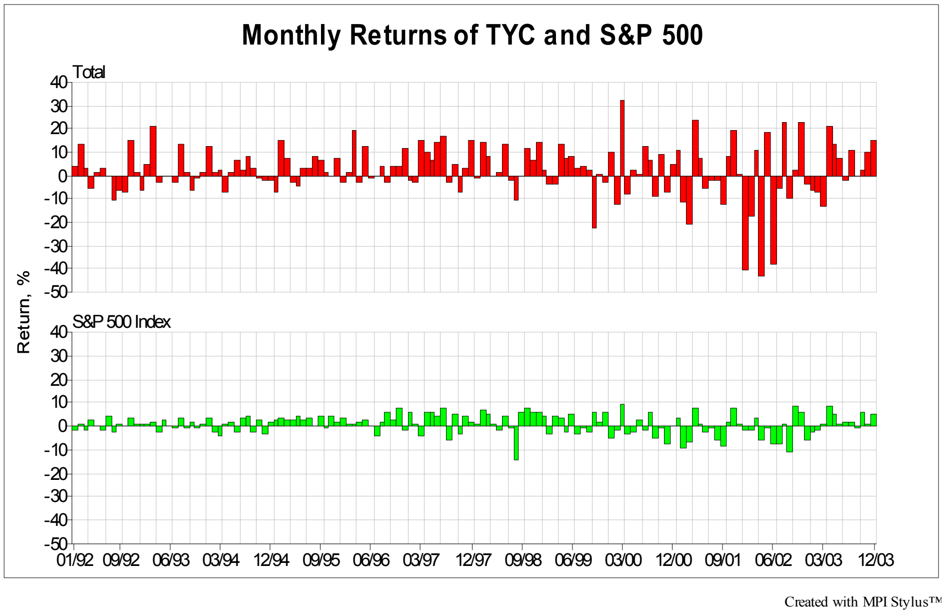


Figure 28. TYC and S&P 500 Monthly Returns.

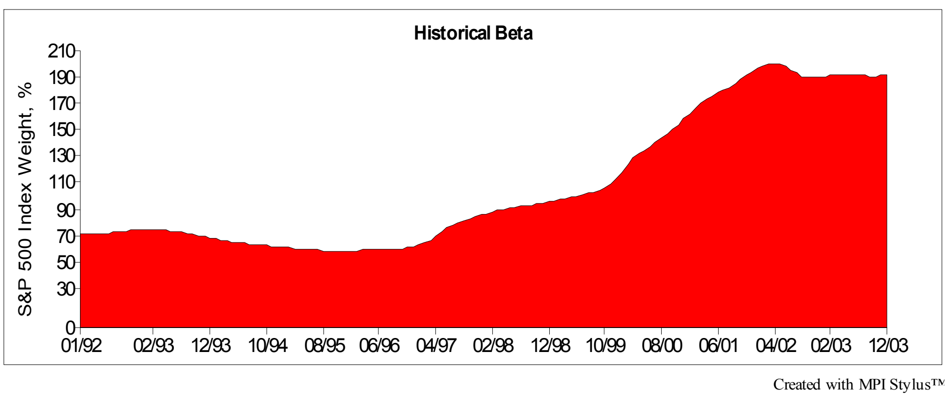


Figure 29. Tyco: Beta vs. SP500.

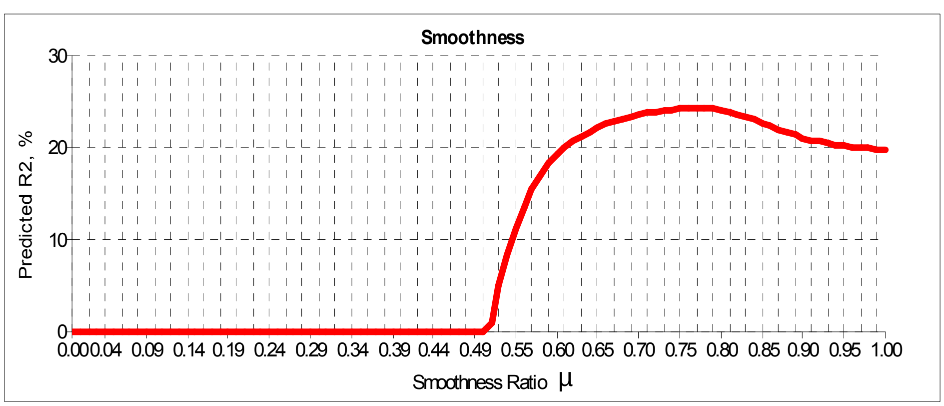


Figure 30. Tyco: Beta smoothness analysis.

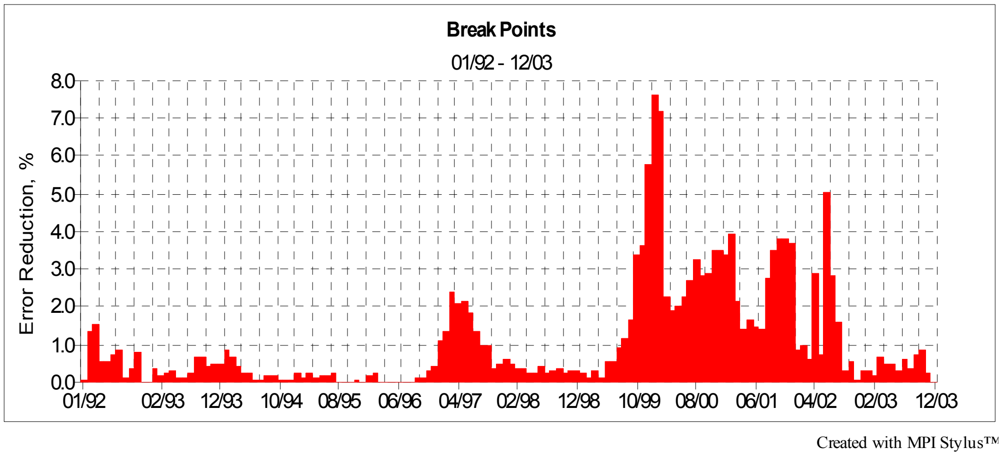


Figure 31. Tyco: Beta breakpoint analysis.

We then release non-smoothness penalty for the two breakpoints in the final dynamic analysis. The results are presented in Figure 32. We observe two changes in TYC beta in that chart: a significant increase at the first breakpoint and decrease at the second. Given that beta is a measure of a relative market risk, the first news increased the risk of the security, making it too high given the company's operations. The second event apparently pacified investors and the stock's beta dropped significantly, still being on a relatively higher side for the industry.

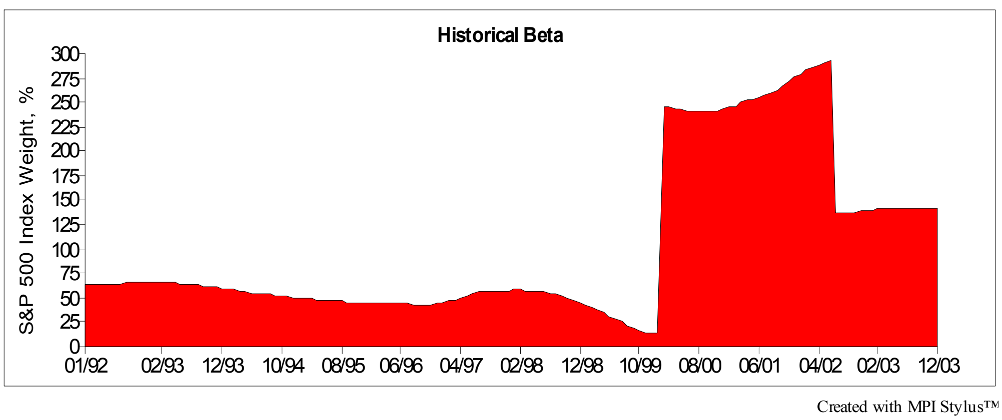


Figure 32. Tyco: Beta adjusted for two breakpoints.

9 Conclusions

The dynamic multi-factor methodology developed in this paper presents a major improvement over the existing RBSA methods currently employed in Finance. It represents a true time-series model, thus providing superior quality recognition of the hidden dynamics of investment portfolios, supported by numerous practical applications presented in the text. Incorporation of the parameter smoothness requirement represents the major innovation of the proposed methodology. Because such a requirement reflects the fundamental rules of the investment process, it ensures that the obtained results are practical. It also ensures the stability of the RBSA results when the non-negativity constraints are relaxed or removed, which proves to be crucial for analysis of such risky investments as hedge funds.

One of the most innovative features of the proposed methodology is that it provides a framework for optimal selection of model parameters. The algorithms for optimal parameter selection presented in this paper are very intuitive and accurate, as has been proved in both practical and model examples. Most of the existing methodologies rely on the researcher and external practical knowledge for manual adjustment of model parameters. The suggested methodology simplifies such a parameter selection process and, in cases with no additional knowledge about the underlying process, presents the only possible method of parameter selection.

Another important feature of the proposed methodology lies in providing a framework for detection of structural changes in parameters. Such changes could represent certain strategy changes or very rapid sell off of the portfolio, which in many cases is hidden from the investors. The proposed methodology allows investors to identify such cases and take immediate action, something that they were not able to do before.

The proposed method is very practical and can be implemented using existing quadratic optimization algorithms and can be executed on a personal computer. Therefore, it can be widely used by investment practitioners and researchers alike.

Some of the most important areas of applications lie in risk management, portfolio construction, and investment research. It can be used to compute market betas on individual assets. It can be used to detect changes in investment products such as mutual funds and hedge funds that are hidden from public. Analysis of hedge funds is the most promising application of the proposed methodology because of the very dynamic and active nature of their investment process and the lack of disclosure of their holdings, which makes the proposed methodology invaluable.

10 Appendices

Appendix 1. Periodic return of a portfolio

Let $m^{(i)}$, $i=1,...,n$, be the constant quantities of assets (for example, number of shares of stock) contained in the portfolio during a holding period, for instance, day, week, month or year, $x'^{(i)}$ and $x''^{(i)}$ be, respectively, the opening and the closing price of a unit of the respective asset at this period, where the return of the ith asset is the ratio $r^{(i)}=(x''^{(i)}-x'^{(i)})/x'^{(i)}$. The total opening price (usually called the market value) of the portfolio is the sum $\sum_{i=1}^{n} m^{(i)}x'^{(i)}$, and its closing price equals $\sum_{i=1}^{n} m^{(i)}x''^{(i)}$. Then, the return of the portfolio at this period is the ratio $r=\left(\sum_{i=1}^{n} m^{(i)}x''^{(i)}-\sum_{i=1}^{n} m^{(i)}x'^{(i)}\right)\Big/\sum_{i=1}^{n} m^{(i)}x'^{(i)}$. Pooling of the two sums in the nominator gives

$$r=\frac{\sum_{i=1}^{n} m^{(i)}(x''^{(i)}-x'^{(i)})}{\sum_{k=1}^{n} m^{(k)}x'^{(k)}}=\frac{\sum_{i=1}^{n} m^{(i)}x'^{(i)}\frac{(x''^{(i)}-x'^{(i)})}{x'^{(i)}}}{\sum_{k=1}^{n} m^{(k)}x'^{(k)}}=$$
$$\sum_{i=1}^{n}\left(\frac{m^{(i)}x'^{(i)}}{\sum_{k=1}^{n} m^{(k)}x'^{(k)}}\frac{(x''^{(i)}-x'^{(i)})}{x'^{(i)}}\right)=\sum_{i=1}^{n}\frac{m^{(i)}x'^{(i)}}{\sum_{k=1}^{n} m^{(k)}x'^{(k)}}r^{(i)}. \qquad (67)$$

Here the ratios in the summands are the fractions representing the weights of each respective ith asset in the total market value of the portfolio, which can be rewritten in the following form using the budget constraint:

$$\beta^{(i)}=(m^{(i)}x'^{(i)})\Big/\sum_{k=1}^{n} m^{(k)}x'^{(k)},\ \sum_{i=1}^{n}\beta^{(i)}=1. \qquad (68)$$

Thus, the return of the portfolio for a period is the linear combination of the returns of its assets for this period with their weights in the portfolio's market value taken at the beginning of the period (or, equivalently, at the end of the previous period) as coefficients:

$$r=\sum_{i=1}^{n}\beta^{(i)}r^{(i)}. \qquad (69)$$

Note that two important assumptions were made to derive the above formula: (a) the quantities of assets $m^{(i)}$ remain constant between the opening and closing of the respective period, while only their prices change and (b) the assets represent *all* portfolio assets, meaning that only the assets in consideration determine the market value of the portfolio at each time period.

Appendix 2. The nonlinear dynamics of the hidden portfolio

The *a priori* assumption on the smoothness of asset amounts $m_t^{(i)}\cong m_{t-1}^{(i)}$ (38) is generic in the dynamic RBSA model, but it is clear that changes of market prices of assets affect their respective weights in the portfolio $\beta_t^{(i)}$ in (13). Assuming that the amounts of assets didn't change during the period *t* i.e., their ending values for the period *t* and *t-1* remain the same $m_t^{(i)}=m_{t-1}^{(i)}$, we have:

$$\beta_{t+1}^{(i)} = \frac{m_t^{(i)} x_t^{(i)}}{\sum_{k=1}^{n} m_t^{(k)} x_t^{(k)}} = \frac{m_{t-1}^{(i)} x_t^{(i)}}{\sum_{k=1}^{n} m_{t-1}^{(k)} x_t^{(k)}} = \frac{m_{t-1}^{(i)} x_{t-1}^{(i)} + m_{t-1}^{(i)} (x_t^{(i)} - x_{t-1}^{(i)})}{\sum_{k=1}^{n} \left[m_{t-1}^{(k)} x_{t-1}^{(k)} + m_{t-1}^{(k)} (x_t^{(k)} - x_{t-1}^{(k)}) \right]} =$$

$$\frac{\left[1 + (x_t^{(i)} - x_{t-1}^{(i)}) \big/ x_{t-1}^{(i)} \right] m_{t-1}^{(i)} x_{t-1}^{(i)}}{\sum_{k=1}^{n} \left[1 + (x_t^{(k)} - x_{t-1}^{(k)}) \big/ x_{t-1}^{(k)} \right] m_{t-1}^{(i)} x_{t-1}^{(i)}}.$$

Here $(x_t^{(i)} - x_{t-1}^{(i)}) \big/ x_{t-1}^{(i)} = z_t^{(i)}$ are the asset returns, therefore,

$$\beta_{t+1}^{(i)} = \frac{(1 + r_t^{(i)}) m_{t-1}^{(i)} x_{t-1}^{(i)}}{\sum_{k=1}^{n} (1 + r_t^{(k)}) m_{t-1}^{(k)} x_{t-1}^{(k)}} = \frac{1 + r_t^{(i)}}{\sum_{k=1}^{n} (1 + r_t^{(k)}) \frac{m_{t-1}^{(k)} x_{t-1}^{(k)}}{\sum_{l=1}^{n} m_{t-1}^{(l)} x_{t-1}^{(l)}}} \times \frac{m_{t-1}^{(i)} x_{t-1}^{(i)}}{\sum_{l=1}^{n} m_{t-1}^{(l)} x_{t-1}^{(l)}} = \frac{1 + r_t^{(i)}}{\sum_{k=1}^{n} (1 + r_t^{(k)}) \beta_t^{(k)}} \beta_t^{(i)}.$$

Here $m_{t-1}^{(i)} x_{t-1}^{(i)} \Big/ \sum_{l=1}^{n} m_{t-1}^{(l)} x_{t-1}^{(l)} = \beta_t^{(i)}$, whence it follows that

$$\beta_{t+1}^{(i)} = \frac{1 + r_t^{(i)}}{1 + \sum_{k=1}^{n} \beta_t^{(k)} r_t^{(k)}} \beta_t^{(i)}. \tag{70}$$

Note that to be consistent with the RBSA notation, where the returns for the period are multiplied by the weights at the beginning of the period, we consider β_{t+1} as reflecting asset weights at the end of the period t or, equivalently, the beginning of the period $t+1$.

References

1. Bucy R.S. *Optimum Finite Time Filters for a Special Non-stationary Class of Inputs*. Internal Report Nr. BBD-600. Jons Hopkins University, Applied Physics Lab, 1959.
2. Kalman R.E. A new approach to linear filtering and prediction problems. *Trans ASME, Series D, J. Basic Eng*, 1960, Vol. 82, pp. 35-45.
3. K. Brammer and G. Siffling, *Kalman-Bucy Filter*. Oldenbourg, Munchen, 1975.
4. Wells C. *The Kalman Filter in Finance*. Kluwer Academic Publishers, 1996.
5. Sharpe W.F. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, Sept. 1964, pp. 425-442.
6. Hull J. *Options, Futures and Other Derivative Securities*, 4 th edition, 2000, ISBN 0-13-022444-8.
7. Markowitz H. *Portfolio Selection: Efficient Diversification of Investments*. Wiley, 1959.
8. Vasicek O. A Note on Using Cross-Sectional Information in Bayesian Estimation of Security Betas. *Journal of Finance*, VIII, No.5 (Dec. 1973), pp. 1233-1239.
9. Blume M. Betas and their regression tendencies. *Journal of Finance*, X, No.3, June 1975, pp. 785-795.
10. King B. Market and industry factors in stock price behavior. *Journal of Business*, 39 (1966), pp. 139-140.
11. Chen N., Richard R., Stephen A. R. Economic forces and the stock market. *Journal of Business*, 59, July 1986, pp. 383-403.
12. Sharpe W.F. Determining a fund's effective asset mix. *Investment Management Review*, Vol. 2, 1988, No. 6 (December), pp. 59-69.
13. Sharpe W.F., Asset Allocation: Management Style and Performance Measurement. *The Journal of Portfolio Management*, Winter 1992, pp. 7-19.
14. Markov M. Style trends in institutional investment. *Equity Style Management* (Klein and Lederman, ed.), Probus, 1995.
15. Sharpe W.F. Macro-Investment Analysis. www.stanford.edu/~wfsharpe, 1995-1998.
16. Zurakhinsky V. Capturing changes in style exposure. *The Journal of Performance Measurement*, Winter 1997/1998, pp.48-50.
17. Cleveland W.S., Loader C. Smoothing by local regression: Principles and methods. *Statistical Theory and Computational Aspects of Smoothing*. (W. Hardle and M. G. Schimek, ed.). Physica Verlag, Heidelberg, 1996.
18. Loader C. *Local Regression and Likelihood*. Springer Series in Statistic, 1999.
19. Chankong V., Yaimes Y. *Multiobjective Decision Making*. North-Holland, New York, 1983.
20. Mottl V.V., Muchnik I.B. *Hidden Markov Models for structural signal analysis*. Nauka, Moscow, 1999 (in Russian).
21. DeGroot M. *Optimal Statistical Decisions*. McGraw-Hill Company, 1970.
22. Bellman, R. *Dynamic Programming*. Princeton University Press, Princeton, N.J., 1957.
23. Sniedovich M. *Dynamic Programming*. Marcel Dekker, NY, 1991.

24. Muchnik I., Mottl V. *Bellman Functions on Trees for Segmentation, Generalized Smoothing and Multi-Alignment in Massive Data Sets*. DIMACS Technical Report 98-15, February, 1998, 63 p.
ftp://dimacs.rutgers.edu/pub/dimacs/TechnicalReports/TechReports/1998/98-15.ps.gz
25. Kalaba R., Tesfatsion L. Time-varying linear regression via flexible least squares. *Computers and Mathematics with Applications*, 17, 1989, pp. 1215-1245.
26. Kalaba R., Tesfatsion L. Flexible least squares for approximately linear systems. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-5, 1990, 978-989.
27. Tesfatsion L. GFLS implementation in FORTRAN and the algorithm.
http://www.econ.iastate.edu/tesfatsi/gflshelp.htm (1997).
28. Lütkepohl H., Herwartz H. Specification of varying coefficient time series models via generalized flexible least squares. *Journal of Econometrics*, 70, 1996, pp.261-290.
29. Powell M.J.D. *On the Quadratic Programming Algorithm of Goldfarb and Idnani*. Report DAMTP, 1983, No. 19, University of Cambridge, Cambridge.
30. Wright S. *Primal-Dual Interior-Point Methods*. SIAM, 1997.
31. Allen, D., "Mean square error of prediction as a criterion for selecting variables," Technometrics, 13, 469-475; 477-481.
32. Stone M. Cross-validatory choice and assessment of statistical predictions. *Journal of Royal Statistical Soc*., 1974, B, 36, pp.111-133.
33. Conway D.A., Marc R.R. Stable factors in security returns: Identification through cross validation. *Journal of Business & Economic Statistics*, 6, January 1988, pp. 1-15.
34. Hjorth, Urban. *Computer Intensive Statistical Methods*. Chapman & Hall, 1994.
35. Vapnik V. *Statistical Learning Theory*. John Wiley & Sons, Inc., 1998.
36. Christopherson J.A., Sabin F.C. How effective is effective mix. *The Journal of Investment Consulting*, December 1998, pp.39-50.