MACHINE-LEARNING FOR DYNAMIC REVERSE ENGINEERING OF HEDGE FUNDS

MICHAEL MARKOV [1], ILYA MUCHNIK[2], VADIM MOTTL[3], OLGA KRASOTKINA[4]

[1] Markov Processes International, 428 Springfield Ave., Summit, NJ 07901, USA
[2] DIMACS, Rutgers University, P.O. Box 8018, Piscataway, NJ 08854, USA
[3] Computing Center of the Russian Academy of Sciences, Vavilov St. 40, Moscow, 119991, Russian Federation
[4] Tula State University, Lenin Ave. 92, Tula, 300600, Russian Federation
E-MAIL: michael.markov@markovprocesses.com, muchnik@dimacs.rutgers.edu, vmottl@yandex.ru, krasotkina@uic.tula.ru

Abstract:

The leave-one-out cross-validation in nested sets of data models is traditionally considered in Machine Learning as the basic instrument of finding the most appropriate subset of features or regressors in pattern recognition and regression estimation. We extend the notion of a nested set of models onto the problem of time-varying regression estimation, which implies, in addition to the generic challenge of choosing the subset of regressors, also the inevitable necessity to choose the appropriate level of model volatility, ranging from the full stationarity of instant models in time to their absolute independence of each other. So, there are, at least, two axes of model nesting in the problem of nonstationary regression estimation, first, the relevant size of the set of regressors and, second, the level of model volatility in time. We use the leave-one-out measure of the model fit as quality indicator along both nesting axes. We apply the proposed technique to analysis of a hedge fund's returns and reverse-engineering its strategies.

Keywords:

Time-varying regression, subset of regressors, time volatility level, leave-one-out procedure, investment portfolio, hedge fund, dynamic style analysis

1. Introduction

The problem of finding numerical regression dependences $y_t : T \to R$ in sets of entities of arbitrary kind $t \in T$ is commonly adopted by the Machine-Learning community as a glowing problem which is far of being completely solved. Just as in much more intensively studied problem of pattern recognition $y_t : T \to \{1,...,M\}$ which differs from that of regression estimation only by the discrete range of the goal variable, the choice of an appropriate subset of the available set of features is usually considered as the most challenging aspect of this problem [1]: $y_t \cong f(x_t^{(i)}, i=1,...,n)$. The kernel-based approach to estimating dependences, which embeds a set of entities of arbitrary kind into a hypothetical linear space, wipes out the difference between linear and nonlinear models [2], so that it is enough to consider only the former of them:

$$y_t = \sum_{i=1}^{n} \beta^{(i)} x_t^{(i)} + e_t = \mathbf{x}_t^T \boldsymbol{\beta} + e_t, \qquad (1)$$

However, in many applications, when the stationary regression model (1) turns out to be insufficient, the set of observations $t \in T$ has to be treated rather as a succession $T = \{1,...,N\}$ than a plain set. Therefore, the problem of estimating a time-varying regression model

$$y_t = \sum_{i=1}^{n} \beta_t^{(i)} x_t^{(i)} + e_t = \boldsymbol{\beta}_t^T \mathbf{x}_t + e_t \qquad (2)$$

in which it is required to estimate the succession of regression coefficients $(\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_N)$, has been subject of intensive study in statistical literature [3,4].

The challenge of regressor selection remains actual for the nonstationary regression model, and, for this reason alone, the problem of time-varying regression estimation falls into the competence area of Machine Learning. In addition to this challenge, in this paper, we focus also on another Machine-Learning-related aspect of nonstationary regression estimation, which is concerned with the inevitable necessity to choose the volatility of time-varying regression coefficients to be estimated. We showed in [5] that the increasing values of the assumed time-volatility factor, even if the set of n regressors in (2) is fixed, determine a succession of "almost" nested classes of time-varying models of the time series $(y_t, \mathbf{x}_t, t=1,...,N)$ under processing.

As an appropriate instrument of data-dependent choice of both the relevant size of the subset of regressors and the time-volatility factor, we use the principle of leave-one-out cross-validation [6,7], which was originally developed for processing time-independent data sets.

We address here primarily the reverse problems that appear in Finance, when it is required to determine the time-varying exposure of an investment portfolio such as hedge fund to market or economic factors using only periodic performance of such portfolio [8]. We illustrate the proposed approach to portfolio analysis by applying it to analysis of some of the well-known hedge funds.

2. Case study: Style Analysis of investment portfolios

A typical example of such a time-varying regression estimation problem is that of determining major market factors affecting performance of an investment portfolio. 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 [8].

2.1. Sharpe's Returns Based Style Analysis (RBSA)

Let the capital of an investment company be fully invested in $n+1$ financial instruments or securities (stocks, bonds, currencies, etc.) in proportions denoted by coefficients $(\beta^{(0)},\beta^{(1)},...,\beta^{(n)})$, $\sum_{i=0}^{n}\beta^{(i)}=1$. The set of such instruments with $\beta^{(i)}\neq 0$ is called an *investment portfolio*. The notation $\beta^{(0)}$ is reserved here for a short-term instrument, such as bank deposit in an interest bearing account.

Let $z_{beg}^{(i)}$ and $z_{end}^{(i)}$ be the prices of securities at the beginning and the end of some time interval called the holding period, respectively, $z_{beg}^{(p)}$ and $z_{end}^{(p)}$ will be the market value of the portfolio as a whole. The ratio $r^{(p)}=(z_{end}^{(p)}-z_{beg}^{(p)})/z_{beg}^{(p)}$ is referred to as the portfolio return for the holding period, and $r^{(i)}=(z_{end}^{(i)}-z_{beg}^{(i)})/z_{beg}^{(i)}$ are securities returns. Since an actual portfolio may contain hundreds and even thousands of instruments, Sharpe proposed to approximate the resulting portfolio return by a small number of market indexes representing certain asset classes and investment styles:

$$r^{(p)}\cong\alpha+\sum_{i=0}^{n}\beta^{(i)}r^{(i)}. \qquad (3)$$

Asset classes, such as US stocks and bonds, International stocks and bonds, are represented by market indexes – generic portfolios comprising large number of securities and computed by various research companies such as Dow Jones, Standard & Poor's, MSCI, etc. Periodic returns $r^{(i)}$ on such market indexes are readily available in financial media and through various subscription services such as Reuters and Bloomberg.

The asset with $\beta^{(0)}$ usually denotes a short-term instrument in a typical portfolio. A typical proxy for such risk-free asset is the 90-day US Treasury Bill that is regularly auctioned by the U.S. Government. In [8] Sharpe used model (3) to analyze performance of a group of US mutual funds and determined that a significant portion of a fund return can be explained by small number of assets. In order to estimate parameters of the model (3), Sharpe would take monthly returns on both the portfolio $\{r_t^{(p)}\}$ and asset indexes $\{r_t^{(i)}\}$ for consecutive months $t=1,2,3,...$ and solve the following constrained quadratic optimization problem:

$$\begin{cases}(\hat{\alpha},\hat{\beta}^{(0)},...,\hat{\beta}^{(n)}):\sum_{t=1}^{N}\left(r_t^{(p)}-\alpha-\sum_{i=0}^{n}\beta^{(i)}r_t^{(i)}\right)^2\to\min,\\ \beta^{(i)}\geq 0,\sum_{i=0}^{n}\beta^{(i)}=1.\end{cases} \qquad (4)$$

The resulting coefficients $(\beta^{(0)},\beta^{(1)},...,\beta^{(n)})$ help to identify the major factors determining portfolio performance.

Further, recognizing that portfolio structure changes over time, Sharpe used a series of optimizations (5) in moving windows of a smaller length K

$$(\hat{\alpha}_t,\hat{\beta}_t^{(0)},...,\hat{\beta}_t^{(n)})=\arg\min_{\alpha,\beta^{(i)}}\sum_{\tau=0}^{K-1}\left(r_t^{(p)}-\alpha-\sum_{i=0}^{n}\beta^{(i)}r_{t-\tau}^{(i)}\right)^2 \qquad (5)$$

to determine the dynamics of portfolio factor exposures.

The model (4) became commonly adopted in the modern Finance under the name of Returns Based Style Analysis (RBSA). The main appeal of this method for practitioners is that it is based solely on analysis of portfolio returns and does not require any other, very often proprietary information about the portfolio composition.

2.2. Limitations of Sharpe's RBSA

Since its introduction, Sharpe's model (3) has been criticized for its inability to capture active portfolio's dynamics. Thus, because portfolio structure is assumed constant within estimation window, the moving window technique (5) appears to be inadequate to capture rapid changes in portfolio structure.

In addition, model (3) loses most of its advantages when it is applied to analysis of portfolios that are allowed to take short (negative) positions. In such cases, non-negativity constraints $\beta^{(i)}\geq 0$ have to be dropped from (4), and the problem is reduced to a simple linear regression. In most such cases, due to multicolinearity effect, moving window method (5) produces highly unstable, meaningless results.

The two limitations above make RBSA inapplicable for analysis of hedge funds because, unlike traditional investment vehicles such as mutual funds, hedge funds are extremely dynamic and take significant short positions. Most attempts to overcome these shortcomings of RBSA are limited to introduction of additional indices into the static model (3) to capture the specifics of a generic hedge fund strategy [9]. None of the methods available to date represent true dynamic model and, therefore, their explanatory power remains low.

In this paper, we introduce a dynamic model that eliminates both shortcomings of Sharpe's RBSA and thus, makes it applicable to long-short strategies and hedge funds.

2.3. Dynamic Style Analysis (DSA)

In contrast to Sharpe's static model (3), we propose a dynamic model in which the fractional structure of the portfolio is changing in time $(\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_N)$, $\boldsymbol{\beta}_t=(\beta_t^{(1)},...,\beta_t^{(n)})$. The new dynamic model of portfolio returns can be written as follows:

$$y_t=(r_t^{(p)}-r^{(0)})=\sum_{i=1}^{n}\beta_t^{(i)}(r_t^{(i)}-r^{(0)})+e_t=\boldsymbol{\beta}_t^T\mathbf{x}_t+e_t, \qquad (6)$$

where $y_t=(r_t^{(p)}-r^{(0)})$ is excess return of the portfolio, $\mathbf{x}_t=\left[(r_t^{(p)}-r^{(0)}),i=1,...,n\right]\in R^n$ is vector of observed excess returns of assets, $\boldsymbol{\beta}_t=(\beta_t^{(1)},...,\beta_t^{(n)})\in R^n$ are vectors of time-varying fractional asset weights that are being estimated, and e_t is observation white noise. Taking excess returns on the portfolio $(r_t^{(p)}-r_t^{(0)})$ and assets $(r_t^{(i)}-r_t^{(0)})$ with respect to the risk-free asset $r_t^{(0)}$ effectively eliminates the need for the budget constraint $\beta_t^{(0)}+\sum_{i=1}^{n}\beta_t^{(i)}=1$ in (3).

The key element of the proposed Dynamic Style Analysis (DSA) is treating asset weights as a hidden process assumed *a priori* to possess the Markov property:

$$\boldsymbol{\beta}_t=\mathbf{V}_t\boldsymbol{\beta}_{t-1}+\boldsymbol{\xi}_t, \qquad (7)$$

where matrices $\mathbf{V}_t$ determine the assumed hidden dynamics of the portfolio structure, and $\boldsymbol{\xi}_t$ is the vector white noise, nonstationary in the general case.

Equation (7) determines, actually, the state-space model of a dynamic system, whereas (6) plays the role of its observation model. In these terms, the DSA problem is that of estimating the time-varying state of the respective dynamic system $(\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_N)$ from observations $(y_t, \mathbf{x}_t, t=1,...,N)$.

The standard means of estimating time-varying models of kind (6)-(7) is the Flexible Least Squares approach (FLS) first introduced in [3]. As applied to the DSA problem, the FLS criterion will have the form

$$J(\beta_1,...,\beta_N)=\sum_{t=1}^{N}\left(y_t-\sum_{i=1}^{n}\beta_t^{(i)}x_t^{(i)}\right)^2+ (1/\lambda)\sum_{i=1}^{n}\sum_{t=1}^{N}(\beta_t^{(i)}-v_t^{(i)}\beta_{t-1}^{(i)})^2\to\min. \tag{8}$$

The matrices $\mathbf{V}_t=\left(Diag(v_t^{(1)},...,v_t^{(n)})\right)$ (7) are assumed here to be diagonal.

The first term $\sum_{t=1}^{N}\left(y_t-\sum_{i=1}^{n}\beta_t^{(i)}x_t^{(i)}\right)^2$ of the FLS criterion (8) stands for the approximation of the excess returns of the portfolio $y_t=(r_t^{(p)}-r^{(0)})$ by a linear combination of those of the set of market indexes $x_t^{(i)}=(r_t^{(i)}-r^{(0)})$ chosen as the basis of Style Analysis $i\in I=\{1,...,n\}$. The greater the size n of the set of generic indexes $I=\{1,...,n\}$, the closer the seeming approximation.

As to the second term $(1/\lambda)\sum_{i=1}^{n}\sum_{t=1}^{N}(\beta_t^{(i)}-v_t^{(i)}\beta_{t-1}^{(i)})^2$ in (8), it is responsible for the time-volatility of regression coefficients. The smaller coefficient $\lambda>0$ corresponds to the "smoother" the estimated asset weights $\boldsymbol{\beta}_t$ in time.

The set of regressors $I=\{1,...,n\}$ and time-volatility coefficient λ define the structure of the model, and a straightforward application of the FLS criterion (8) implies that these parameters are to be chosen by the user. However, it is extremely problematic, on the one hand, to preset the structure of the model a priori, and, on the other, to find an "appropriate" model structure by attempting to additionally minimize the residual squares sum in (8).

The problem of estimating the structural parameters of a model to be fitted to experimental data makes one of the main challenges of Machine Learning. In this paper, we propose a revision of the FLS criterion (8) with the purpose to endow it with the ability to find the relevant subset of regressors $\hat{I}\subset I=\{1,...,n\}$ and the appropriate value of the time-volatility coefficient $\hat{\lambda}$.

3. Two axes of model nesting

As regards to the problem of finding the subset of features or regressors $\hat{I}\subset I=\{1,...,n\}$, the classical Machine Learning theory [7] suggests that the entire set I is a priori represented as a succession of nested subsets $I_1\subset...\subset I=I_n$ of increasing size $m=|I_m|=1,...,n$. Once this is done, the most appropriate subset $\hat{I}=I_{\hat{m}}$ is to be chosen by the criterion of minimizing the average error of prediction measured in the universe, in our case, the mathematical expectation $E\left\{\left(y_t-\sum_{i\in I_m}\beta_t^{(i)}x_t^{(i)}\right)^2\right\}$ (6). However, the universe is unavailable at the stage of model inferring, and, as a substitute for the average risk, the data-dependent leave-one-out estimate of it is applied. The leave-one-out principle was proposed first for estimating the misclassification probability in discriminant analysis and pattern recognition [6], and was used later for measuring the average risk in both pattern recognition and function estimation problems.

As to the time-volatility parameter λ in the problem of nonstationary regression estimation, it is shown in [5] that increasing values of it also form a succession of "almost nested" model classes, and the leave-one-out principle remains completely applicable to finding the appropriate value $\hat{\lambda}$.

So, instead of one model-nesting axis considered in the classical Machine Learning problems, which traditionally deal with data acquired from a stationary universe, the problem of nonstationary regression estimation leads to the necessity to take into account two independent nesting axes related, first, to the size $1\le m\le n$ of the actual subset of regressors, and, second, to the value $0<\lambda<\infty$ of the time-volatility parameter.

However, there is a huge variety of possible nestings $I_1\subset...\subset I=I_n$ of regressors subsets, because there are very many subsets $I_m\subset I$ of each specific size $|I_m|=m<n$ in the original full set $|I|=n$. The classical Machine Learning theory does not suggest any way of choosing the relevant subset of the specific size m except full enumeration.

In this paper, we propose a novel statistical framework for automatically finding the best regressors subset of the fixed size. To be strict, we fix not the tentative size $1\le m\le n$ of the subset $I_m\subseteq I$ we are looking for, but a selectivity parameter $0<\mu<\infty$ of a probabilistic data model that underlies the estimation technique.

4. A Bayesian approach to estimating the relevant subset of regressors with controlled selectivity

Let $(\mathbf{x}_t, t=1,...,N)$, $\mathbf{x}_t=(x_t^{(i)}, i=1,...,n)$, be a given sequence of regressors not subject to any probabilistic modeling. In terms of Dynamic Style Analysis (Section 2.3), these are excess returns of the basic set of generic market indexes $x_t^{(i)}=r_t^{(i)}-r^{(0)}$.

We consider the time series to be processed $(y_t, t=1,...,N)$ (6), in our case, the excess returns of the portfolio $y_t=r_t^{(p)}-r^{(0)}$ under monitoring, as the observable part of a two-component random process, whose hidden part is the unknown sequence of time-varying regression coefficients $\left(\boldsymbol{\beta}_t=(\beta_t^{(i)}, i=1,...,n), t=0,1,...,N\right)$, i.e. fractional weights of the portfolio composition.

The main point of the regressors selection technique we propose here is the a priori probabilistic model of the hidden process of the regression coefficients $\boldsymbol{\beta}_t = (\beta_t^{(i)},\ i=1,...,n)$. First, its components are considered as a priori independent. Second, the initial values $\beta_0^{(i)}$ are assumed to be normally distributed with zero mathematical expectations $E(\beta_0^{(i)})=0$ and variances $E\left((\beta_0^{(i)})^2\right)=\rho\delta^{(i)}$ proportional to some unknown values whose product equals to unity $\prod_{i=1}^n \delta^{(i)}=1$. Third, the following values of the regression coefficients are formed by the identical autoregression models $\beta_t^{(i)}=v_t^{(i)}\beta_{t-1}^{(i)}+\xi_t^{(i)}$, where $\xi_t^{(i)}$ are independent normal white noises with zero mathematical expectations $E(\xi_t^{(i)})=0$ and variances $E\left((\xi_t^{(i)})^2\right)=(1/\lambda)\rho\delta^{(i)}$. The autoregression coefficients $v_t^{(i)}$ make the diagonal matrix $\mathbf{V}_t$ in (7).

The proportionality factor ρ is interpreted as the observation noise variance in the model of nonstationary regression (6) $E(e_t)=0$, $E\left((e_t)^2\right)=\rho$. It is clear that, if we assume the values $\left(\delta^{(i)}\ge 0, \prod_{i=1}^n \delta^{(i)}=1\right)$ to be known, the Bayesian estimates of the regression coefficients will not depend on the common factor ρ.

It is clear that if some of the a priori variances $\delta^{(i)}$ equal zero, the respective regressors $x_t^{(i)}$ will be excluded from the nonstationary regression model (6), i.e., $\hat{I}=\left\{i:\delta^{(i)}>0\right\}$.

But we do not assume the values $(\delta^{(i)},\ i=1,...,n)$ to be known. Let us preliminarily consider the collection of independent a priori gamma distributions of inverse variances $\gamma\left((1/\delta^{(i)})|\alpha,\vartheta\right)\propto(1/\delta^{(i)})^{\alpha-1}\exp\left(-\vartheta(1/\delta^{(i)})\right)$, whose mathematical expectations $E(1/\delta^{(i)})=\alpha/\vartheta$ and variances $E\left((1/\delta^{(i)})^2\right)=\alpha/\vartheta^2$ are equal to each other, and put the parameters $\alpha=\vartheta=1/2\rho\mu$. We shall have a parametric family of distributions with the only parameter $\mu\ge 0$, such that $E(1/\delta^{(i)})=1$ and $E\left((1/\delta^{(i)})^2\right)=2\rho\mu$. If $\mu\to 0$, the practically nonrandom values $1/\delta^{(i)}$ will be close to each other $1/\delta^{(1)}\cong...\cong 1/\delta^{(n)}\cong 1$, on the contrary, if $\mu\to\infty$, the independent nonnegative values $1/\delta^{(i)}$ may be, a priori, arbitrarily different. We call μ the model selectivity parameter.

The final assumed a priori distribution of inverse variances is cut from this collection of independent distributions by the above condition $\prod_{i=1}^n 1/\delta^{(i)}=1$. It is easy to see that the joint a priori distribution density of variances $(\delta^{(1)},...,\delta^{(n)})$ will be expressed as

$$G\left(\delta^{(1)},...,\delta^{(n)}\,|\,\mu\right)\propto\exp\left(-(1/2\rho\mu)\sum_{i=1}^n(1/\delta^{(i)})\right),\ \prod_{i=1}^n\delta^{(i)}=1,$$
$$G\left(\delta^{(1)},...,\delta^{(n)}\,|\,\mu\right)=0,\ \prod_{i=1}^n\delta^{(i)}\ne 1$$

So, we have defined, first, the joint a priori distribution of variances $G\left(\delta^{(1)},...,\delta^{(n)}\,|\,\mu\right)$, second, the conditional a priori distribution of the hidden sequence of regression coefficients $\Psi(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_N\,|\,\delta^{(1)},...,\delta^{(n)})$, and, third, the conditional distribution of the observed time series $\Phi(y_1,...,y_N\,|\,\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_N)$. It is clear that the a posteriori joint distribution density of the hidden vector sequence of regression coefficients and variances of their components related to single regressors will be proportional to the product

$$P(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_N,\delta^{(1)},...,\delta^{(n)}\,|\,y_1,...,y_N,\mu,\lambda)\propto$$
$$\Phi(y_1,...,y_N\,|\,\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_N)\Psi(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_N\,|\,\delta^{(1)},...,\delta^{(n)},\lambda)\times G\left(\delta^{(1)},...,\delta^{(n)}\,|\,\mu\right).$$

It appears natural to take the maximum point of this a posteriori density as the estimate of the sequence of time-varying regression coefficients along with the variances indicating participation of each of regressors in the model:

$$(\hat{\boldsymbol{\beta}}_0,\hat{\boldsymbol{\beta}}_1,...,\hat{\boldsymbol{\beta}}_N,\hat{\delta}^{(1)},...,\hat{\delta}^{(n)}\,|\,\mu,\lambda)=\arg\max P(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_N,\delta^{(1)},...,\delta^{(n)}\,|\,y_1,...,y_N,\mu,\lambda). \quad (9)$$

Theorem. The maximum point of the a posteriori density (9) by $(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_N,\delta^{(1)},...,\delta^{(n)})$ is the minimum point of the criterion

$$\begin{cases} J(\boldsymbol{\beta}_0,\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_N,\delta^{(1)},...,\delta^{(n)}\,|\,\mu,\lambda)=\sum_{t=1}^N\left(y_t-\sum_{i=1}^n\beta_t^{(i)}x_t^{(i)}\right)^2+ \\ \sum_{i=1}^n\frac{1}{\delta^{(i)}}\left((\beta_0^{(i)})^2+\lambda\sum_{t=1}^N(\beta_t^{(i)}-v_t^{(i)}\beta_{t-1}^{(i)})^2+(1/\mu)\right)\to\min, \\ \prod_{i=1}^n\delta^{(i)}=1. \end{cases} \quad (10)$$

As we see, if the a priori variances of regression coefficients at single regressors $(\delta^{(1)},...,\delta^{(n)})$ are fixed, the resulting criterion practically coincides with the FLS criterion (8). When being minimized by all the variables with almost zero value of the selectivity parameter $\mu\to 0$, its minimum point does not differ from that of FLS.

However, the presence of the additional variables $(\delta^{(1)},...,\delta^{(n)})$ is extremely important. If some $\delta^{(i)}\to 0$, the criterion drastically penalizes the deflection of the entire sequence of the respective time-varying regression coefficient $(\beta_0^{(i)},\beta_1^{(i)},...,\beta_N^{(i)})$ from zero, and practically excludes the i th regressor from the model. It is just this tendency which the criterion (10) pronouncedly displays if the selectivity parameter essentially differs from zero $\mu>0$. If $\mu\to\infty$, the criterion has full freedom of eliminating evidently redundant regressors, but never suppresses all of them.

So, the modified FLS criterion yields a subset of regressors $\hat{I}=\left\{i:\delta^{(i)}>0\right\}\subseteq I$ of the size $m=|\hat{I}|$ which monotonely decreases as the selectivity parameter grows $\mu\to\infty$.

5. The iterative minimization procedure

For finding the minimum point of the modified FLS criterion (10) with fixed structural parameters μ and λ, we apply the Gauss-Seidel iteration to both groups of variables $(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_N)$ and $(\delta^{(1)}, ..., \delta^{(n)})$ starting with the initial values $(\delta^{(i),0} = 1,\ i = 1, ..., n)$.

At each iteration, the current approximations $(\delta^{(i),k} = 1,\ i = 1, ..., n)$ turn (10) into the usual FLS criterion (8) with respect to regression coefficients $(\boldsymbol{\beta}_0^k, \boldsymbol{\beta}_1^k, ..., \boldsymbol{\beta}_N^k)$, which can be easily minimized by the standard Kalman-Bucy filter and smoother [5]. Once the regression coefficients are found, the next values of the variances $(\delta^{(i),k+1},\ i = 1, ..., n)$ are defined by the following expression which is easy to prove:

$$\delta^{(i),k+1} = \frac{(\beta_0^{(i),k})^2 + \lambda\sum_{t=1}^{N}(\beta_t^{(i),k} - v_t^{(i)}\beta_{t-1}^{(i),k})^2 + 1/\mu}{\sqrt[n]{\prod_{l=1}^{n}\left((\beta_0^{(l),k})^2 + \lambda\sum_{t=1}^{N}(\beta_t^{(l),k} - v_t^{(l)}\beta_{t-1}^{(l),k})^2 + 1/\mu\right)}}.$$

It is well seen that if $\mu \to 0$ all the variances equal each other.

For finding the appropriate values of the two structural parameters μ and λ, the leave-one-out cross-validation technique described in [5] is to be applied.

6. Reverse-engineering hedge fund strategies using only performance data

In this Section, we present an example of application of the dynamic multi-factor methodology to a real-life portfolio, namely, Long Term Capital Management (LTCM).

The collapse of this highly-leveraged fund in 1998 is by far the most dramatic hedge fund story to date. At the beginning of 1998, the $5B fund maintained a leverage ratio of about 28:1 [10]. LTCM's troubles began in May-June 1998. By the end of August, the fund had lost 52%. By the end of September, 1998, a month after the Russian crisis, the fund had lost 92% of its December 1997 assets. Fearing the destabilizing impact of the highly leveraged fund on global financial markets, on September 23rd, the Federal Reserve Bank of New York orchestrated a bailout of the fund by a group of 14 major banks.

The investment strategy of the fund was based on spread bets between thousands of closely related securities in various global markets. Such bets are based on the assumption that the securities' prices will eventually converge and the arbitrage position will result in a profit. In fact, spreads continued to increase, which eventually led to the collapse of the fund. It took several major investigations to determine that the major losses sustained by LTCM came from bets on global credit spreads.

Applying DSA to monthly returns of the fund shown in Figure 1, we attempted to determine major factors explaining the fall of LTCM. For our analysis we used the indexes provided by Lehman Brothers and Merrill Lynch and shown in Table 1.

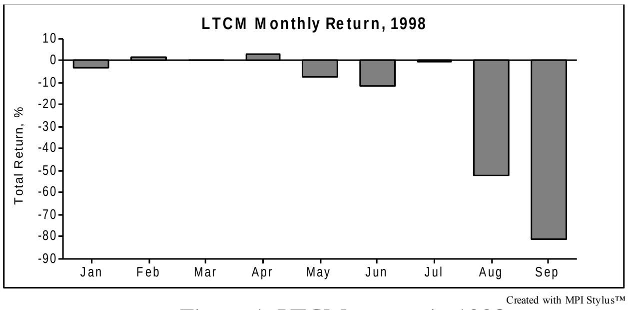

Figure 1. LTCM returns in 1998.

Table 1. Generic indexes and model selection

	Index name	Essential occurrence of generic indexes in the result of Style Analysis		
		no selectivity $\mu \cong 0$	cross-validated selectivity $\mu = 100$	unlimited selectivity $\mu \to \infty$
1	US Gov Long Bonds	■		
2	US Corp Long Bonds	■		
3	High Yield Bonds	■		
4	Mortgages	■		
5	Russell 2000 Value	■		
6	ML Emerging Bond Asia	■		
7	ML EMU Direct Govt, 5-10Y	■		
8	EMU Corp Bonds	■	■	
9	ML EMU Broad Market Index	■	■	
10	Russell 1000 Growth	■		
11	Corporate Int Bond	■		
12	Gov Int Bond	■		
13	Russell 1000 Value	■	■	
14	Russell 2000 Growth	■	■	
15	ML EMU Corp, 5-10Y	■	■	
16	ML Emerging Bond Latin Amer.	■		
17	EMU Govt Bonds	■	■	
18	ML Emerging Bond Eur/ME/Afr	■	■	■

For a preliminary analysis, we used the tentative value of the selectivity parameter $\mu = 50$. Application of the leave-one-out cross-validation technique of finding the optimal smoothness parameter λ gave the value $\hat{\lambda} = 10$.

Then, with the found smoothness $\hat{\lambda} = 10$, we varied the selectivity parameter μ in the interval from the minimum value $\mu = 10^{-6} \cong 0$, which is equivalent to the application of the classical FLS technique without any selection of regressors, to the maximum value $\mu = 10^6$ providing full selectivity.

The results of Style Analysis for the two extreme values $\mu \cong 0$ and $\mu \to \infty$ are shown in Figure 2, (a) and (c), respectively. Asset exposures $(\beta_t^{(1)}, ..., \beta_t^{(n)})$ for each time period are "stacked" along the vertical axis with respect to the sign with the sum equal to 100%. As we see, all the generic indexes participate in the straightforward approximation of the portfolio dynamics in the former case, whereas only one of them tries to explain it when the selectivity is unlimited.

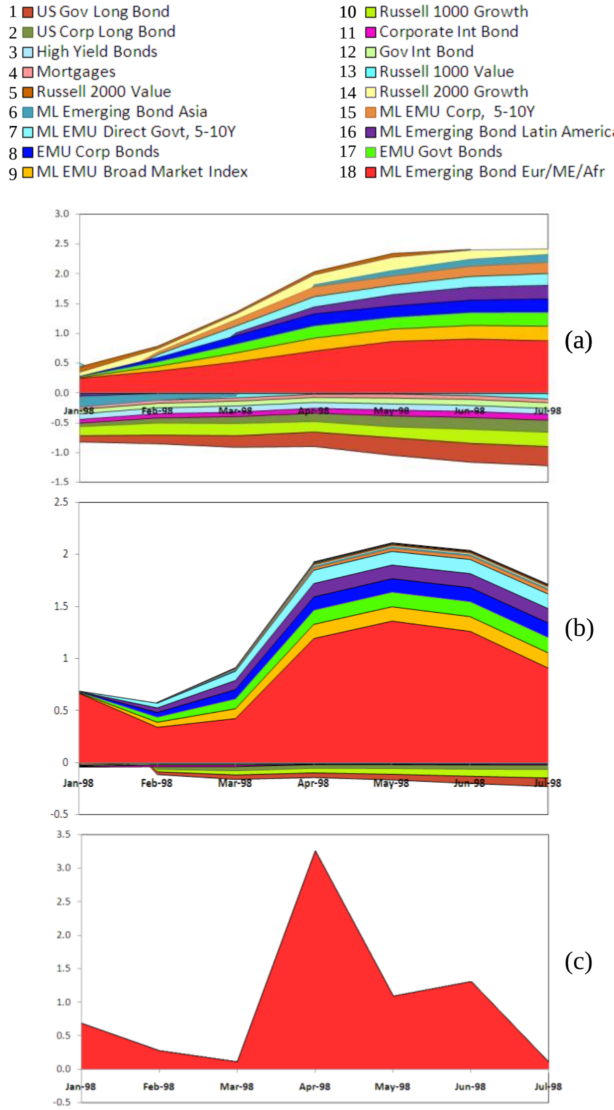

Figure 2. Style Analysis of the LTCM fund: (a) – no selectivity, (b) – cross-validated selectivity, (c) – unlimited selectivity.

Application of the leave-one-out technique to the tentative series of the selectivity values gave $\hat{\mu}=100$ as the most appropriate selectivity level. The respective result of Style Analysis is shown in Figure 2 (b). In this case, only part of the indexes essentially participate in the portfolio model. In Table 1, we ticked off the indexes whose occurrence in the model computed as $\sum_{t=1}^{N}(\beta_t^{(i)})^2$ exceeds the threshold 0.001.

7. Conclusions

We apply the Machine Learning methodology to provide a framework for truly dynamic analysis of hedge funds. The proposed Dynamic Style Analysis approach is implemented as a parametric quadratic programming model with two parameters controlling the selection of most appropriate generic indexes and the smoothness of their estimated weights in time.

We have shown that continuously increasing unknown values of model selectivity and time-volatility form two continuous axes of nested model classes. Just as in classical applications, the average risk defined by the given values of structural parameters cannot be measured immediately. By analogy, we use the traditional leave-one-out cross-validation principle when solving the problem of time-varying regression estimation.

The proposed DSA approach has made it possible to fundamentally improve the existing RBSA methodology currently employed in Finance. It results in increased transparency and better hedge fund due diligence which is of crucial importance for financial institutions today.

References

[1] Jain A., Zongker D. Feature selection: Evaluation, application, and small sample performance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, February 1997, Vol. 19, no. 2, pp. 153-158.

[2] Mottl V., Krasotkina O., Seredin O., Muchnik I. Kernel fusion and feature selection in machine learning. *Proceedings of the 8th IASTED International Conference on Intelligent Systems and Control*. Cambridge, USA, October 31 – November 2, 2005.

[3] Kalaba R., Tesfatsion L. Time-varying linear regression via flexible least squares. *International Journal on Computers and Mathematics with Applications*, 1989, Vol. 17, pp. 1215-1245.

[4] Wells C. *The Kalman Filter in Finance*. Kluwer Academic Publishers, 1996.

[5] Markov M., Krasotkina O., Mottl V., Muchnik I. Time-varying regression model with unknown time-volatility for nonstationary signal analysis. Proceedings of the 8th IASTED International Conference on Signal and Image Processing. Honolulu, Hawaii, USA, August 14-16, 2006.

[6] Lachenbruch P. An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. *Biometrics*, 1967, Vol.23, No. 4, pp.639-645.

[7] Vapnik V. *Statistical Learning Theory*. John-Wiley & Sons, Inc. 1998.

[8] Sharpe W.F. Asset allocation: Management style and performance measurement. *The Journal of Portfolio Management*, Winter 1992, pp. 7-19.

[9] Fung W.K.H., Hsieh D.A. Empirical characteristics of dynamic trading strategies: The case of hedge funds. *Review of Financial Studies*, 1997, V.10, pp. 275–302.

[10] Ph. Jorion. Risk management lessons from long-term capital management. *European Financial Management*, September, 2000.