BAYESIAN ESTIMATION OF TIME-VARYING REGRESSION WITH CHANGING TIME-VOLATILITY FOR DETECTION OF HIDDEN EVENTS IN NONSTATIONARY SIGNALS

O. Krasotkina, A. Kopylov
Tula State University, Tula, Russia
ko180177@yandex.ru

V. Mottl
Computing Center of the Russian Academy of Sciences, Moscow, Russia
vmottl@yandex.ru

M. Markov
Markov Processes Int., New Jersey, USA
michael.markov@markovprocesses.com

ABSTRACT[1]
Problems of signal analysis may be practically always considered as those of recovering some hidden dependences, which are time-varying in the general case. In many situations, the nostationarity mode of the dependence should be expected to change in the observation interval, maybe with some jumps or spikes. In this work, we consider a statistical framework and a family of respective algorithm for time-varying linear regression estimation which preserve essential peculiarities in basically smoothly changing regression coefficients. The method being proposed is simple in tuning and has linear computational complexity with respect to the signal length. In particular, we show how this technique allows to watch the dynamics of the hidden asset composition of an investment portfolio from publicly available data with the purpose of detecting sharp changes in its investment strategy.

KEYWORDS
Nonstationary regression, parameterized a priori model, quadratic dynamic programming

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 vector parameter $B=(\beta_t, t=1,...,N)$ of a nonstationary local model.

One of the simplest examples of such a model is that of time-varying regression. The vector signal to be analyzed is considered as consisting of two components $(Y,X)=\big((y_t,\mathbf{x}_t), t=1,...,N\big)$, namely, a vector sequence $\mathbf{x}_t\in\mathbb{R}^n$ called regressors and a numerical sequence $y_t\in\mathbb{R}$. The main assumption is that the vector sequence $\mathbf{x}_t$ is recorded as a manifestation of an observable process whose properties are not studied, and the numerical component y_t is obtained at each point of observation as a noisy linear function of the vector $\mathbf{x}_t$:

$$y_t=\sum_{i=1}^n \beta_t^{(i)}x_t^{(i)}+e_t=\mathbf{x}_t^T\beta_t+e_t. \quad (1)$$

The number of variables nN to be estimated in this model $(\beta_t^{(i)}, i=1,...,n, t=1,...,N)$ ever exceeds the number of observations N. Thus, it is impossible to estimate the time-varying regression model without additional regularization, namely, without making a priori assumptions of sufficiently smooth mode of changing the regression coefficients in time.

Under this assumption, the problem of estimating the sequence of time-varying regression coefficients $B=(\beta_t, t=1,...,N)$ has been the subject of intensive study in statistical literature during the last at least fifteen years. The two-criteria approach developed in the pioneering paper [1] has lead to the principle of Flexible Least Squares (FLS) which, in a slightly generalized form, may be put as

$$\begin{cases}\hat{B}(\lambda)=\arg\min_B J(B\,|\,Y,X,\lambda),\\ J(B\,|\,Y,X,\lambda)=\\ \sum_{t=1}^N (y_t-\beta_t^T\mathbf{x}_t)^2+\frac{1}{\lambda}\sum_{t=2}^N(\beta_t-\mathbf{V}_t\beta_{t-1})^T(\beta_t-\mathbf{V}_t\beta_{t-1}).\end{cases} \quad (2)$$

Here matrices $\mathbf{V}_t$ determine the assumed hidden dynamics of the regression coefficients, and the parameter $\lambda>0$ is responsible for the trade-off between two mutually contradictory requirements – to provide close approximation of the observed signal and, at the same time, to control the smoothness of its time-varying model. The smaller λ, the lesser the time-volatility of the estimated sequence of regression coefficients.

The FLS criterion is justified in [2,3] as realization of the Bayesian approach to time-varying regression estimation under the a priori assumption that the hidden sequence of regression coefficients is a random Markov process

$$\beta_t=\mathbf{V}_t\beta_{t-1}+\xi_t. \quad (3)$$

Its Bayesian estimate is given by (2) with time-volatility λ defined as the ratio between the assumed variances of the renewal white noise ξ_t in the a priori Markov process (3) and of the observation white noise e_t (1).

The time-varying regression model (1) coupled with the FLS criterion (2) offers an adequate formalization of such classical signal-processing problems as smoothing and time-frequency analysis. In particular, this criterion is shown in [3,4] to cover the problem of Dynamic Returns Based Style Analysis of investment portfolios. This technique is a generalization of the classical static Returns Based Style Analysis of investments [5], introduced by William Sharpe, 1990 Nobel Prize winner in Economics, as instrument of estimating the hidden asset composition of a portfolio from publicly available data.

It is just the latter application which is primarily addressed in this paper. In Dynamic Style Analysis, just as in many others signal-analysis problems, it is hardly likely to

[1] This work is supported by the Russian Foundation for Basic Research, Grant 09-07-90700, 09-01-12085

expect that the time-volatility of the hidden process to be estimated remains the same over the entire observation interval. If economic situation changes, the investment company being monitored may quickly adjust its management style to new conditions. This practical reality gets in conflict with the traditional form of the FLS criterion which presupposes the constant time-volatility of the hidden regression coefficients λ.

In the literature, the problems concerned with preserving quick discontinuities when estimating a basically smooth hidden signal from noisy observations have been considered only for plain signal smoothing or nonparametric regression estimation and called edge-preserving smoothing.

Methods for edge-preserving nonstationary regression estimation are studied considerably worse. In fact, only two methods of generalized edge-preserving smoothing are known – one of them is based on competitive Kalman filtration [6], and the other one exploits the iterative dynamic programming procedure [7]. However, both of these methods lean upon the tacit assumption that the jumps to be revealed are abrupt and spaced at a distance not smaller than the minimum acceptable effective width of the time-window sufficient for achieving the desired smoothing degree between them.

In this paper, we present a generalization of the Bayesian approach to time-varying regression estimation [2,3], which involves automatic adjustment of the changing time-volatility to the given signal $(Y,X)=\big((y_t,\mathbf{x}_t),t=1,...,N\big)$. As previously, we assume the observation noise variance $E(e_t^2)=\delta$ in (1) to be unknown. The only novelty in comparison with [2,3] is the assumption that the variance of the white noise in the Markov model of the hidden regression coefficients (3) is changing in time $E(\xi_t\xi_t^T)=r_t\mathbf{I}$ and unknown. The seemingly incorrect intent to jointly estimate both sequences of regression coefficients $(\beta_t\in\mathbb{R}^n,\ t=1,...,N)$ and of variances $(r_t,t=1,...,N)$ is compensated by a priori constraints on the latter ones as independent identical gamma distributions of the time-volatility factors $\lambda_t=r_t/\delta$ with the same mathematical expectation λ and sufficiently small variance μ considered as two structural parameters of the signal model. Parameter λ presets the basic average time volatility of the regression coefficients and μ controls the rigidity of adherence to it.

The proposed technique is illustrated by applying it to the time series of daily returns of George Soros' Quantum hedge fund in year 1992 along with those of the asset classes in which, by Soros' declaration, the capital was invested in that period. We show that the edge-preserving time-varying regression analysis clearly detects the event of the "Black Wednesday", when a quick deliberate change in the fund's portfolio broke down the Britain Pound.

2. Linear normal-gamma model of time-varying regression with changing time-volatility

Let the real-valued vector time series (Y,X) to be processed have the length N and consist of $n+1$ components $(Y,X)=\big((y_t,\mathbf{x}_t),t=1,...,N\big)$, $y_t\in\mathbb{R}$, $\mathbf{x}_t\in\mathbb{R}^n$. We shall consider it as the observable part of a more-dimensional vector signal (Y,X,B) whose additional component $B=(\beta_t,t=1,...,N)$, $\beta_t\in\mathbb{R}^n$, is hidden from observation.

It is assumed that the observable component (Y,X) and the hidden one B are related to each other in accordance with the model of linear regression (1) in which the zero-mean normal white noise of observation has some variance $E(e_t^2)=\delta$ which is considered as unknown. Thus, the joint conditional probability density of the scalar component of the signal $Y=(y_t,t=1,...,N)$ with respect to the vector regressors $X=(\mathbf{x}_t,t=1,...,N)$ and regression coefficients $B=(\beta_t,t=1,...,N)$ will be normal:

$$\Phi(Y\,|\,B,X,\delta)=\frac{1}{\delta^{N/2}(2\pi)^{N/2}}\exp\left(-\frac{1}{2\delta}\sum_{t=1}^{N}(y_t-\mathbf{x}_t^T\beta_t)^2\right).\quad(4)$$

Let the aim of processing the observed time series (Y,X) be estimating the hidden signal B, which occurs in the model (1) as the sequence of time-varying regression coefficients. Estimation of B by plain maximization of the likelihood function (4) $\sum_{t=1}^{N}(y_t-\mathbf{x}_t^T\beta_t)^2\to\ \min(B)$ is senseless because of too large number of variables.

Just as in [2,3], we a priori consider the sequence of regression coefficients as Markov random process (3) which starts with the normally distributed first value $E(\beta_1)=\mathbf{0}$, $E(\beta_1\beta_1^T)=\rho\mathbf{I}$, and is disturbed by a normal vector white noise $E(\xi_t\xi_s^T)=0$. But, in contrast to [2,3], we do not assume the common variance of its independent components to remain the same in time, and suppose $E(\xi_t\xi_t^T)=r_t\,\mathbf{I}$. The unknown variances $(r_t,t=1,...,N)$ are considered as proportional to the variance of the observation noise $r_t=\lambda_t\delta$, unknown as well, with the proportionality coefficients λ_t acting as factors of the unknown instantaneous time-volatility of regression coefficients.

Under this assumption, the a priori joint distribution of the hidden sequence of regression coefficients $B=(\beta_t,t=1,...,N)$ is conditionally normal with respect to the sequence of the time-volatility factors $\Lambda=(\lambda_t,t=2,...,N)$. In practice, a priori information on β_1 is hardly available, therefore, we put the variance ρ equal to a sufficiently large number. So, we come to the improper a priori density

$$\Psi(B\,|\,\Lambda,\delta)\propto\frac{1}{\left(\prod_{t=2}^{N}\delta\lambda_t\right)^{1/2}(2\pi)^{k(N-1)/2}}\times \exp\left(-\frac{1}{2}\sum_{t=2}^{N}\frac{1}{\delta\lambda_t}(\beta_t-\mathbf{V}_t\beta_{t-1})^T(\beta_t-\mathbf{V}_t\beta_{t-1})\right).\quad(5)$$

Finally, we assume the inverse time-volatility factors $1/\lambda_t$ to be a priori independent and identically gamma-distributed $\gamma\left(1/\lambda_t|\alpha,\vartheta\right)\propto(1/\lambda_t)^{\alpha-1}\exp\left(-\vartheta(1/\lambda_t)\right)$ on the positive half-axis $\lambda_t\ge 0$. The mathematical expectation and variance of gamma distribution are ratios α/ϑ, and α/ϑ^2.

The a priori distribution density of the entire sequence of time-volatility factors is the product

$$G(\Lambda|\alpha,\vartheta)=\prod_{t=2}^{N}\gamma(\lambda_t|\alpha,\vartheta)\propto\left(\prod_{t=2}^{N}\frac{1}{\lambda_t}\right)^{\alpha-1}\exp\left(-\vartheta\sum_{t=2}^{N}\frac{1}{\lambda_t}\right)= \exp\left[-(\alpha-1)\sum_{t=2}^{N}\ln\lambda_t-\vartheta\sum_{t=2}^{N}\frac{1}{\lambda_t}\right]. \quad (6)$$

We redefine the parameters through new parameters λ and μ

$$\alpha=\frac{1}{2}\left[\frac{1}{\delta}\left(1+\frac{1}{\mu}\right)+1\right],\ \vartheta=\frac{\lambda}{2\delta\mu}, \quad (7)$$

assuming thereby the parametric family of gamma distributions of random variables $1/\lambda_t$

$$\gamma\left(1/\lambda_t|\delta,\lambda,\mu\right)\propto(1/\lambda_t)^{\frac{2\mu+1}{2\delta\mu}}\exp\left(-\frac{\lambda}{2\delta\mu}(1/\lambda_t)\right)$$

with mathematical expectations and variances

$$E(1/\lambda_t)=\frac{(1+\delta)\mu+1}{\lambda},\quad Var(1/\lambda_t)=2\delta\mu\frac{(1+\delta)\mu+1}{\lambda^2}.$$

In terms of this parameterization, the independent a priori distribution of each instantaneous inverse time-volatility factor $1/\lambda_t$ is almost completely concentrated around the mathematical expectation $1/\lambda$ if $\mu\to 0$. On the contrary, with $\mu\to\infty$ it tends to the almost uniform distribution. It is well seen from the plots in Fig. 1.

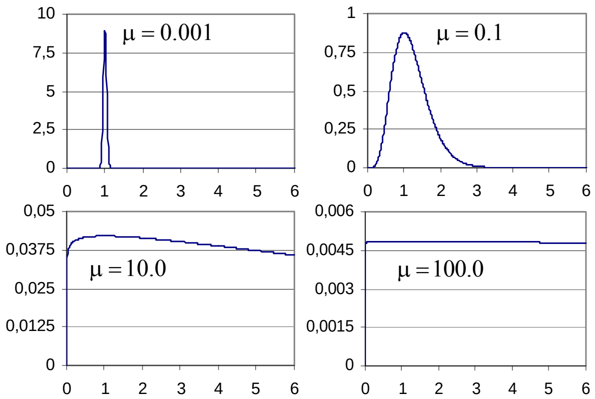

Figure 1. A priori gamma distribution of the instantaneous time-volatility λ_t with fixed parameter $\lambda=\delta=1$ and four different values of μ.

We come to the a priori density

$$G(\Lambda|\delta,\lambda,\mu)=\exp\left[-\frac{1}{2\delta\mu}\sum_{t=2}^{N}\left(\lambda\frac{1}{\lambda_t}+\frac{1}{\lambda}\ln\lambda_t\right)\right], \quad (8)$$

and, so, have completely defined the joint a priori normal-gamma distribution of both hidden sequences $B=(\beta_t, t=1,...,N)$ and $\Lambda=(\lambda_t, t=2,...,N)$:

$$H(B,\Lambda\,|\,\delta,\lambda,\mu)=\Psi(B\,|\,\Lambda,\delta)G(\Lambda\,|\,\delta,\lambda,\mu). \quad (9)$$

Coupled with the conditional density of the observable time series (4), it makes basis for Bayesian estimation of the time-varying regression coefficients.

3. The Bayesian estimate of the time-varying regression with changing time-volatility

Given a time series $(Y,X)=\left((y_t,\mathbf{x}_t),\, t=1,...,N\right)$, the joint a posteriori distribution of hidden sequences, namely, those of regression coefficients and their instantaneous time-volatility factors, is completely defined by (4), (5) and, in terms of the original parameters (α,ϑ), by (6):

$$P(B,\Lambda\,|\,Y,X,\delta,\alpha,\vartheta)=\frac{\Phi(Y|B,X,\delta)\Psi(B\,|\,\Lambda,\delta)G(\Lambda\,|\,\alpha,\vartheta)}{\int\int\Phi(Y|B',X,\delta)\Psi(B'\,|\,\Lambda',\delta)G(\Lambda'\,|\,\alpha,\vartheta)dB'd\Lambda'}.$$

The Bayesian estimate of (B,Λ) is the maximum point of the numerator

$$\begin{cases}(\hat{B},\hat{\Lambda}\,|\,\delta,\alpha,\vartheta)=\\ \arg\max_{B,\Lambda}\left[\ln\Phi(Y|B,X,\delta)+\ln\Psi(B|\Lambda,\delta)+\ln G(\Lambda|\alpha,\vartheta)\right]=\\ \arg\max_{B,\Lambda}\left(-\frac{1}{2\delta}\sum_{t=1}^{N}(y_t-\mathbf{x}_t^T\beta_t)^2-\frac{1}{2}\sum_{t=2}^{N}\ln\delta\lambda_t-\right.\\ \left.\frac{1}{2\delta}\sum_{t=2}^{N}\frac{1}{\lambda_t}(\beta_t-\mathbf{V}_t\beta_{t-1})^T(\beta_t-\mathbf{V}_t\beta_{t-1})-(\alpha-1)\sum_{t=2}^{N}\ln\lambda_t-\vartheta\sum_{t=2}^{N}\frac{1}{\lambda_t}\right),\end{cases}$$

or, what is equivalent,

$$\begin{cases}(\hat{B},\hat{\Lambda}|\delta,\alpha,\vartheta)=\arg\min_{B,\Lambda}J(B,\Lambda\,|\,Y,X,\delta,\alpha,\vartheta),\\ J(B,\Lambda\,|\,Y,X,\delta,\alpha,\vartheta)=\sum_{t=1}^{N}(y_t-\mathbf{x}_t^T\beta_t)^2+\\ \sum_{t=2}^{N}\left\{\frac{1}{\lambda_t}\left[(\beta_t-\mathbf{V}_t\beta_{t-1})^T(\beta_t-\mathbf{V}_t\beta_{t-1})+2\delta\vartheta\right]+\delta(2\alpha-1)\ln\lambda_t\right\}.\end{cases}$$

The substitution of the new parameters (7) makes the Bayesian estimate independent of the observation noise variance δ:

$$\begin{cases}(\hat{B},\hat{\Lambda}|\lambda,\mu)=\arg\min_{B,\Lambda}J(B,\Lambda\,|\,Y,X,\lambda,\mu),\\ J(B,\Lambda\,|\,Y,X,\lambda,\mu)=\sum_{t=1}^{N}(y_t-\mathbf{x}_t^T\beta_t)^2+\\ \sum_{t=2}^{N}\left\{\frac{1}{\lambda_t}\left[(\beta_t-\mathbf{V}_t\beta_{t-1})^T(\beta_t-\mathbf{V}_t\beta_{t-1})+\lambda/\mu\right]+(1+1/\mu)\ln\lambda_t\right\}.\end{cases} \quad (10)$$

We call this principle of time-varying regression estimation Flexible Least Squares with Changing Volatility (FLSCV).

4. Properties of Flexible Least Squares with changing time-volatility

Let the sequence of regression coefficients be fixed in (10), then the conditionally optimal time-volatility factors

$\hat{\Lambda}(B,\lambda,\mu)=\left[\hat{\lambda}_t(B,\lambda,\mu),\, t=2,...,N\right]$ are defined independently of each other:

$$\hat{\Lambda}(Y,X,B,\lambda,\mu)=\arg\min_{\Lambda} J(B,\Lambda\,|\,Y,X,\lambda,\mu):$$

$$\frac{\partial}{\partial\lambda_t}\left\{\frac{1}{\lambda_t}\left[(\beta_t-\mathbf{V}_t\beta_{t-1})^T(\beta_t-\mathbf{V}_t\beta_{t-1})+\lambda/\mu\right]+(1+1/\mu)\ln\lambda_t\right\}=0. \quad (11)$$

The zero conditions for the derivatives, excluding the trivial solutions $\lambda_t\to\infty$, lead to the equalities

$$\frac{1}{\lambda_t}\left[(\beta_t-\mathbf{V}_t\beta_{t-1})^T(\beta_t-\mathbf{V}_t\beta_{t-1})+\lambda/\mu\right]=(1+1/\mu),$$

and, hence,

$$\hat{\lambda}_t(B,\lambda,\mu)=\lambda\frac{(1/\lambda)(\beta_t-\mathbf{V}_t\beta_{t-1})^T(\beta_t-\mathbf{V}_t\beta_{t-1})+1/\mu}{1+1/\mu}. \quad (12)$$

Substitution of (12) into (10) gives the equivalent form of the FLSCV, which avoids immediate finding the time-volatility factors themselves:

$$\begin{cases}(\hat{B},\hat{\Lambda}\,|\,\lambda,\mu)=\arg\min_{B,\Lambda} J(B,\Lambda\,|\,Y,X,\lambda,\mu),\\ J(B,\Lambda\,|\,Y,X,\lambda,\mu)=\sum_{t=1}^{N}(y_t-\mathbf{x}_t^T\beta_t)^2+\\ \sum_{t=2}^{N}\left\{(1+1/\mu)\ln\frac{(1/\lambda)(\beta_t-\mathbf{V}_t\beta_{t-1})^T(\beta_t-\mathbf{V}_t\beta_{t-1})+1/\mu}{1+1/\mu}\right\}\end{cases} \quad (13)$$

The classical FLS principle (2) is a particular case of (13) with $\mu=0$. Indeed, if $\mu\to 0$, we have $\hat{\lambda}_t(B,\lambda,\mu)\to\lambda$ in (12), and substitution of this limit value turns (10) it into the function with constant time-volatility

$$J(B\,|\,Y,X,\lambda,\mu\to 0)=\sum_{t=1}^{N}(y_t-\mathbf{x}_t^T\beta_t)^2+$$
$$\frac{1}{\lambda}\sum_{t=2}^{N}(\beta_t-\mathbf{V}_t\beta_{t-1})^T(\beta_t-\mathbf{V}_t\beta_{t-1})+(N-1)\left[\frac{1}{\mu}+\left(1+\frac{1}{\mu}\right)\ln\lambda\right],$$

which is equivalent to (2).

On the contrary, if $\mu\to\infty$, the limit form of the criterion (13) is the function

$$J(B\,|\,Y,X,\lambda,\mu\to\infty)=\sum_{t=1}^{N}(y_t-\mathbf{x}_t^T\beta_t)^2+$$
$$\sum_{t=2}^{N}\ln\left[\frac{1}{\lambda}(\beta_t-\mathbf{V}_t\beta_{t-1})^T(\beta_t-\mathbf{V}_t\beta_{t-1})\right].$$

Fig. 2 shows the dependence of the unsmoothness penalty in (13) from the norm of the instantaneous unsmoothness of the regression coefficients $\left((\beta_t-\mathbf{V}_t\beta_{t-1})^T(\beta_t-\mathbf{V}_t\beta_{t-1})\right)^{1/2}$ for different values of the parameter μ. It is almost quadratic in a vicinity of the zero point $\beta_t-\mathbf{V}_t\beta_{t-1}=\mathbf{0}$ and remains being so practically over the entire number axis if μ is small. But as μ grows, the originally quadratic penalty undergoes more and more marked effect of saturation at some distance from zero. This means that the criterion strongly penalizes slight uncouthness of regression coefficients but becomes more and more indulgent to sharp discontinuities.

In contrast to the classical FLS (2), the FLSCV (13) is not quadratic. It is not convex either, but it remains quasi-convex for any $\mu\geq 0$ [2]. This fact is proved in Appendix 1.

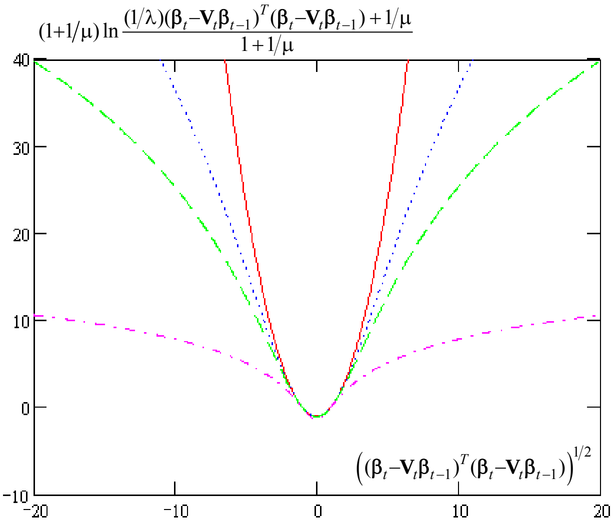

Figure 2. The "saturation effect" of the unsmoothness penalty for sufficiently large values of the parameter μ and fixed value $\lambda=1$.

Further, the quasi-convex function of B (13) is everywhere differentiable, thus, the sequence of regression coefficients $B=(\beta_t, t=1,...,N)$ is its minimum point if and only if

$$\nabla_B J(B\,|\,Y,X,\lambda,\mu)=\mathbf{0}. \quad (14)$$

This condition is applicable to (13), but can be written in terms of the equivalent form of the FLSCV criterion (10).

Let $\hat{B}(\Lambda,\lambda,\mu)$ be the optimal sequence of regression coefficients (10) with respect to a given sequence of time-volatility factors $\Lambda=(\lambda_t, t=2,...,N)$:

$$\hat{B}(\Lambda,\lambda,\mu)=\left(\hat{\beta}_t(\Lambda), t=1,...,N\right)=\arg\min_{B} J(B,\Lambda|Y,X,\lambda,\mu). \quad (15)$$

It is proved in Appendix 2 that (14) is equivalent to the condition $\lambda_t=\hat{\lambda}_t(B,\lambda,\mu)$, or, in accordance with (12),

$$\lambda_t=\frac{\left(\hat{\beta}_t(\Lambda)-\mathbf{V}_t\hat{\beta}_{t-1}(\Lambda)\right)^T\left(\hat{\beta}_t(\Lambda)-\mathbf{V}_t\hat{\beta}_{t-1}(\Lambda)\right)+\lambda/\mu}{1+1/\mu}$$

This fact suggests an iterative procedure of finding the solution of both equivalent FLSCV criterions with changing volatility (10) and (13).

5. The Gauss-Seidel procedure of estimating the time-varying regression coefficients

For finding the minimum point of the modified FLSCV criterion (10) with fixed structural parameters μ and λ, we apply the Gauss-Seidel iteration to both groups of variables

[2] Function $f(\mathbf{x}):\mathbb{R}^m\to\mathbb{R}$ is said to be quasi-convex if it meets the inequality $f\left(c\mathbf{x}'+(1-c\mathbf{x}'')\right)<\max\left(f(\mathbf{x}'),f(\mathbf{x}'')\right)$ for any $\mathbf{x}',\mathbf{x}''\in\mathbb{R}^m$ and $0\leq c\leq 1$ [8]. It is clear that, in particular, any convex function $f\left(c\mathbf{x}'+(1-c\mathbf{x}'')\right)<cf(\mathbf{x}')+(1-c)f(\mathbf{x}'')$ is quasi-convex.

$B=(\beta_t, t=1,...,N)$ and $\Lambda=(\lambda_t, t{=}2,...,N)$ in (10) starting with the initial values $\hat{\Lambda}^0=(\hat{\lambda}_t^0=\lambda, t=2,...,N)$.

At each iteration, the current sequence $\hat{\Lambda}^k=(\hat{\lambda}_t^k=1, t=1,...,N)$ turns (10) through (15) into a quadratic function of $B=(\beta_t, t=1,...,N)$, whose minimum point gives the new approximation to the estimate of the sequence of regression coefficients:

$$\hat{B}^k=(\hat{\beta}_t^k, t=1,...,N)=\arg\min_B J(B,\Lambda^k|Y,X,\lambda,\mu)=$$
$$\arg\min_B\left\{\sum_{t=1}^N(y_t-\mathbf{x}_t^T\beta_t)^2+\sum_{t=2}^N\frac{1}{\lambda_t}(\beta_t-\mathbf{V}_t\beta_{t-1})^T(\beta_t-\mathbf{V}_t\beta_{t-1})\right\}. \quad (16)$$

This optimization problem has linear computational complexity with respect to the signal length N and can be easily solved by the Kalman-Bucy filter and smoother [2,3].

Once the regression coefficients $\hat{B}^k=(\hat{\beta}_t^k, t=1,...,N)$ are found, the next approximation to the estimates of time-volatility factors $\hat{\Lambda}^{k+1}=(\hat{\lambda}_t^{k+1}=1, t=1,...,N)$ is defined by the rule (12)

$$\hat{\lambda}_t^{k+1}=\frac{(\hat{\beta}_t^k-\mathbf{V}_t\hat{\beta}_{t-1}^k)^T(\hat{\beta}_t^k-\mathbf{V}_t\hat{\beta}_{t-1}^k)+\lambda/\mu}{1+1/\mu},\ t=2,...,N, \quad (17)$$

which, in accordance with (11), gives the solution of the conditional optimization problem

$$\hat{\Lambda}^{k+1}=\arg\min_\Lambda J(\hat{B}^k,\Lambda\,|\,Y,X,\lambda,\mu)=$$
$$\arg\min_B\left\{\sum_{t=1}^N(y_t-\mathbf{x}_t^T\hat{\beta}_t^k)^2+\sum_{t=2}^N\frac{1}{\lambda_t}(\hat{\beta}_t^k-\mathbf{V}_t\hat{\beta}_{t-1}^k)^T(\hat{\beta}_t^k-\mathbf{V}_t\hat{\beta}_{t-1}^k)\right\}.$$

The very structure of the iterative procedure (16)-(17) provides strong diminishing of the FLSCV criterion (10) until the stationary point (14) is achieved

$$J(\hat{B}^{k+1},\hat{\Lambda}^{k+1}|Y,X,\lambda,\mu)<J(\hat{B}^k,\hat{\Lambda}^k|Y,X,\lambda,\mu),$$

because, as it is shown in Section 4, the equality holds only at the stationary point.

It remains only to specify the way of choosing the values of the structural parameters λ and μ, which control, respectively, the basic average time-volatility of regression coefficients and the ability of instantaneous volatility factors to change in time.

6. The leave-one out principle of choosing the structural parameters

7. Case study: Dynamic Style Analysis of investment portfolios

8. Detecting hidden events in the investment strategy of a hedge fund using only performance data

9. Appendix 1. Quasi convexity if the FLS criterion with changing volatility

10. Appendix 2. The stationary point of the FLS criterion with changing volatility

References

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

[2] M. Markov, O. Krasotkina, V. Mottl, I. Muchnik. 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.

[3] M. Markov, I. Muchnik, V. Mottl, O. Krasotkina. Dynamic analysis of hedge funds. *Proceedings of the 8th IASTED International Conference on Financial Engineering and Applications*. MIT, Cambridge, Massachusetts, USA, October 9-11, 2006.

[4] L. Swinkels, P. Van der Sluis. Return-based style analysis with time-varying exposures. *European Journal of Finance*, 2006, Vol. 12, Issue 6-7, pp. 529-552.

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

[6] M. Niedzwiecki, W.A. Sethares. Smoothing of discontinuous signals: The competitive approach. *IEEE Trans. on Signal Processing*, Vol. 43, No. 1, January 1995, pp. 1-13.

[7] V.V. Mottl, I.B. Muchnik, A.A. Kostin. Generalized edge-preserving smoothing for signal analysis. *Proceedings of the IEEE Workshop on Nonlinear Signal and Image Analysis*, Mackinac Island, Michigan, USA, September 7-11, 1997.

[8] Я.И. Заботин, А.И. Кораблев. Псевдовыпуклые функционалы и их экстремальные свойства. *Изв. вузов. Математика*, 1974, №4 (143), с. 27-31. (найти англоязычную аналогичную ссылку)