Doklady Mathematics, Vol. 66, No. 3, 2002, pp. 468–470. Translated from Doklady Akademii Nauk, Vol. 387, No. 6, 2002, pp. 754–756.

COMPUTER SCIENCE

Short-Term Prediction in a Securities Market

Corresponding Member of the RAS **B. S. Kashin and S. V. Pastukhov**

Received September 27, 2002

The problem of short-term (in particular, single-day) prediction of prices in actual securities markets is considered using the apparatus of function approximation theory. As an example, we analyze the behavior of the prices for SP500 and Nasdaq-100 futures traded on the Chicago Mercantile Exchange (CME).

One of the most widely used financial-market models is that introduced by P. Samuelson (see [2, p. 344] for more detail), in which the price $S(t)$ of a stock, futures, etc., at time t is described by a random process

$$S(t) = S(0)e^{\mu t + \sigma W_t - \frac{\sigma^2}{2}t}, \qquad (1)$$

where W_t is a standard Wiener process.

It is easy to understand that the direction of price changes cannot be effectively predicted within the framework of this and other similar models [the lack of "arbitrage" in the market (1)]. At the same time, it has long been recognized by experts in applied financial mathematics that model (1) is not entirely adequate for describing actual market processes, while the problem of effective prediction on actual securities markets is quite meaningful.

One approach to this problem is a technical analysis that constructs predictions by using in some way the statistics of historical data on financial asset prices. The approach described here can also be considered a technical analysis, but it has the feature of using an adequate (in our view) mathematical apparatus, the so-called piecewise monotone approximations. In particular, this apparatus allows us to obtain statistical analysis results in a natural and visual form.

For a function $F(t)$ defined on a time interval $[T_0, T_1]$ at a given $n = 1, 2, \ldots$, we consider its best approximation by elements of families Σ_n of functions with no more than n monotonicity intervals. More precisely, $f \in \Sigma_n \leftrightarrow \exists(t_0, t_1, \ldots, t_n)$: $T_0 = t_0 < t_1 < \ldots < t_n = T_1$, and f is monotone on each interval $\Delta_i = [t_{i-1}, t_i]$ for $i = 1, 2, \ldots, n$.

The best piecewise monotone approximation of order n for $F(t)$ is defined by the relation

$$M_n(F) = \inf_{f_n \in \Sigma_n} \sup_{t \in [T_0, T_1]} |F(t) - f_n(t)|. \qquad (2)$$

The values in (2) have been actively studied in function approximation theory [in particular, in connection with the fact that $M_n(F)$ estimates from below the corresponding best polynomial and rational approximations of F]. In [1] an algorithm was suggested for calculating values (2) and for finding best-approximation elements $f_n \in \Sigma_n$ ($n = 1, 2, \ldots$) for an arbitrary function F continuous on $[T_0, T_1]$. Note that f_n is generally defined nonuniquely.

For our purposes, we need a function that is (in a sense) inverse to $M_n(F)$:

$$N(\varepsilon, F) = \min\left\{n = 1, 2, \ldots: M_n(F) < \frac{\varepsilon}{2}\right\}. \qquad (3)$$

Knowledge of (3) makes it possible to answer the question as to how many times the direction of variation in F changes on the time interval $[T_0, T_1]$ if we neglect variations by values smaller than ε. This characteristic is quite natural when F is the price of some financial assets. It can be interpreted as a characteristic of the variability of F (by a value $\geq\varepsilon$), an analogue of the volatility σ in model (1).

It is especially important for us that the function $f_n \in \Sigma_n$, $n = N(\varepsilon, F) > 1$, providing the best approximation of F on $[T_0, T_1]$ can be constructed dynamically; i.e., the monotonicity intervals $\Delta_i = [t_{i-1}, t_i]$ of f_n for which $t_i \leq t$ and the values of f_n on these intervals can be determined from the values of $F(x)_{T_0 \leq x \leq t}$. This algorithm is also applicable when F is given on a discrete set $\Lambda \in [T_0, T_1]$. Moreover, it is easy to verify that the run time of the algorithm for determining, given ε, the number $n = N(\varepsilon, F)$ and the corresponding best-approximation element f_n is then a linear function of the numbers of elements in Λ.

Before analyzing financial data, it is necessary to refine the concept of the price of a stock, futures, etc. We consider only those situations where price data are received from the Electronic Communications Network, because the corresponding quotations, deal prices, etc., are then specified sufficiently accurately.

Steklov Mathematical Institute, Russian Academy of Sciences, ul. Gubkina 8, Moscow, 119991 Russia

Table 1

Emini–Nasdaq-100 futures				Emini–SP 500 futures			
ε	$K_\varepsilon(1)$, $K_\varepsilon(2)$, $K_\varepsilon(3)$, $K_\varepsilon(4)$	N^+	N^-	ε	$K_\varepsilon(1)$, $K_\varepsilon(2)$, $K_\varepsilon(3)$, $K_\varepsilon(4)$	N^+	N^-
2	+4, –2, +3, –1	6	20	1	+5, –1, +1, –1	22	40
	+2, –4, +3, –1	13	2		+2, –3, +2, –1	56	27
	+1, –1, +4, –1	177	121		+1, –2, +3, –1	137	86
	+1, –2, +5, –1	36	13		–1, +2, –3, +1	95	138
	–2, +4, –2, +1	31	16		–1, +1, –4, +1	71	141
	–5, +1, –1, +1	36	57		–2, +1, –3, +1	85	140
3	+3, –3, +2, –1	13	1	1.5	+3, –1, +3, –1	25	6
	+1, –2, +3, –1	92	63		+1, –1, +4, –1	50	30
	+1, –2, +2, –1	212	141		+1, –2, +3, –1	56	29
	–1, +2, –2, +1	131	214		–1, +2, –3, +1	33	49
	–1, +2, –3, +1	51	84		–1, +1, –4, +1	36	69
	–3, +2, –1, +1	70	46		–3, +1, –3, +1	11	22

We assume that time t changes discretely at intervals of 1 s. In an actual market, for liquid financial assets, such as SP500 or Nasdaq-100 futures, a few deals can occur and quotations can vary several times per second. To be definite, the price $F(t)$ is set equal to the price of the last deal at time t (if there were no deals at time t, we take the last deal on the interval $[T_0, t]$). The differences arising when predictions are constructed with another definition of $F(t)$ are not crucial, although they can have a large effect on the possibility of the practical use of predictions.

Before describing the prediction method, we have to introduce the concept of the similarity of two piecewise monotone functions for a given error ε. Suppose that we are given two piecewise monotone functions: f_n^1 on an interval $[T_0^1, T_1^1]$ and f_n^2 on an interval $[T_0^2, T_1^2]$, with the respective monotonicity intervals $[t_{i-1}^1, t_i^1]$ and $[t_{i-1}^2, t_i^2]$ for $i = 1, 2, \ldots, n$, where $T_0^1 = t_0^1 < t_1^1 < \ldots < t_n^1 = T_1^1$ and $T_0^2 = t_0^2 < t_1^2 < \ldots < t_n^2 = T_1^2$. For $i = 1, 2, \ldots, n$, we define $\alpha_i^1 = f_n^1(t_i^1) - f_n^1(t_{i-1}^1)$ and $\alpha_i^2 = f_n^2(t_i^2) - f_n^2(t_{i-1}^2)$. It is clear that, by definition, the piecewise monotone function $|\alpha_i^1|$ is the amplitude of oscillations in f_n^1 on the interval $[t_{i-1}^1, t_i^1]$, and a similar situation occurs with $|\alpha_i^2|$. The functions f_n^1 and f_n^2 are said to belong to the same similarity class $\mathcal{K}(\varepsilon, n)$ for given ε and n if for a set of integers $K_\varepsilon(i)$ ($i = 1, 2, \ldots, n$) the following inequalities are fulfilled simultaneously:

$$\operatorname{sgn}(\alpha_i^1)K_\varepsilon(i)\varepsilon \le |\alpha_i^1| < \operatorname{sgn}(\alpha_i^1)(K_\varepsilon(i) + \operatorname{sgn}(\alpha_i^1))\varepsilon,$$

$$\operatorname{sgn}(\alpha_i^2)K_\varepsilon(i)\varepsilon \le |\alpha_i^2| < \operatorname{sgn}(\alpha_i^2)(K_\varepsilon(i) + \operatorname{sgn}(\alpha_i^2))\varepsilon.$$

The similarity classes $\mathcal{K}(\varepsilon, n)$ are denoted by $\{K_\varepsilon(1), K_\varepsilon(2), \ldots, K_\varepsilon(n)\}$ and are called patterns. Note that $K_\varepsilon(i) = 0$ only if $n = 1$ and $\Omega_{[T_0, T_1]}f_n < \varepsilon$, where $\Omega_{[T_0, T_1]}f_n = \max\limits_{[T_0, T_1]} f_n - \min\limits_{[T_0, T_1]} f_n$.

From a practical point of view, the concept of a pattern is fairly illustrative, because it shows (with an error less than ε) the number of multiples of ε by which the financial asset price changes on each monotonicity interval.

Consider arbitrary positive numbers A^+ and A^- and an instant of time t. We are interested in the random variable

$$\tau(t, A^+, A^-) = \inf\{u \ge t\colon F(u) \notin (F(t) - A^-, F(t) + A^+)\}.$$

The goal of a prediction is to estimate the probabilities P^+ and P^- of the events

$$\Omega^+ = \{F(\tau(t, A^+, A^-)) \ge F(t) + A^+\},$$

$$\Omega^- = \{F(\tau(t, A^+, A^-)) \le F(t) - A^-\}.$$

To this end, a best-approximation element f_n for $F(x)_{|T_0 \le x \le t|}$ at $n = N(\varepsilon, F)$ is dynamically constructed on $[T_0, t]$. This means that we sequentially build the "tendency-variation" points $T_0 = t_0 < t_1 < \ldots < t_n \le t$. It should be noted that, for practical reasons, we consider only those times t that satisfy $t = \inf\{u \ge t_{n-1}: |F(u) - F(t_{n-1})| \ge \mu\}$ for prescribed $\mu \ge \varepsilon$. Next, given a positive integer $Q < n$, we determine the pattern $\{K_\varepsilon(1), K_\varepsilon(2), \ldots, K_\varepsilon(Q)\}$, which contains the restriction of f_n to the interval $[t_{n-Q}, t]$. Analyzing historical price data, we find all such cases where a piecewise monotone approximation belongs to the similarity class $\{K_\varepsilon(1), K_\varepsilon(2), \ldots, K_\varepsilon(Q)\}$ and, based on the statistics of a further price behavior, we determine the required probabilities P^+ and P^-. For example, if we buy financial assets at the price $F(t)$ at time t, then our mean profit (loss) is $A^+P^+ - A^-P^-$. Thus, a sufficiently large value of $|A^+P^+ - A^-P^-|$ can be a signal for the purchase or sale of financial assets at time t. In other words, our goal is to find such patterns for which $|A^+P^+ - A^-P^-|$ is sufficiently large (greater than a threshold value θ). In practice θ can be chosen based on the commission value and other market characteristics.

It was found that an effective prediction is possible in many cases (this is the main result of this paper). Apparently, this can be explained by the behavioral psychology of the market participants. The conclusion about effective prediction is confirmed by statistical data on P^+ and P^- for given ε, Q, μ, A^+, and A^-. To calculate statistics, we used data about within-day prices for SP500 futures (Emini–SP500 futures) and Nasdaq-100 futures (Emini–Nasdaq-100 futures) over one year, namely, from October 1, 2001 until September 9, 2002 (this period contains 234 trading days). On each day from the period indicated, we selected the most liquid futures contract (i.e., either the futures with the nearest date or the next futures). In all cases, $\mu = A^+ = A^- = \varepsilon$, and $Q = 4$. For better insight into the capacity of the statistical sample, we present not P^+ and P^- but rather N^+ and N^-, which are the numbers of elements in the sets Ω^+ and Ω^-, respectively. It is clear that $P^+ = \frac{N^+}{N^+ + N^-}$ and $P^- = \frac{N^-}{N^+ + N^-}$. The value of ε is measured in items of points of the corresponding futures (a one-point change in the futures price corresponds to a \$50 change for Emini–SP500 futures and to a \$20 change for Emini–Nasdaq-100 futures). The values of N^+ and N^- for some patterns are listed in Table 1. The "distortions" noted in the table are persistent and are exhibited on other time intervals.

In our opinion, the statistical analysis results presented here should be taken into account in the construction of adequate financial-market models.

REFERENCES

1. Sevast'yanov, E.A., *Anal. Math.*, 1975, vol. 1, no. 2, pp. 141–164.

2. Shiryaev, A.N., *Osnovy stokhasticheskoi finansovoi matematiki* (Fundamentals of Stochastic Financial Mathematics), Moscow: Fazis, 1998, vol. 1.