Моделирование хаотических процессов на рынках: краткий обзор и критический анализ

А. А. Мусаев[1], И. В. Ананченко[2]
ФГБОУ ВО «Санкт-Петербургский государственный технологический институт (технический университет)»
[1]amusaev@technolog.edu.ru, [2]anantchenko@yandex.ru

В. В. Трофимов[1], С. М. Газуль[2]
ФГБОУ ВО «Санкт-Петербургский государственный экономический университет»
[1]tvv@mail.ru, [2] stanislav@gazul.ru

Аннотация. **Рассмотрена задача моделирования изменения цен торговых активов и валютных котировок на основе концепций статистической и хаотической динамики. Приведен исторический обзор эволюции моделей процесса котировок и критический анализ основных подходов к решению данной задачи. Представлены материалы численных исследований, уточняющих структуру процесса ценообразования.**

Ключевые слова: моделирование; статистический анализ; нестационарный процесс; хаотический процесс; котировки активов; валютный рынок; Forex

I. Введение

Задачи моделирования динамических систем в большинстве прикладных задач ориентированы на формирование алгоритма или генератора временных последовательностей, свойства которых в каком-то заранее оговоренном смысле близки к свойствам наблюдений за изменением состояния моделируемой системы. Моделирование динамики котировок является важным элементом проведения исследовательских работ, ориентированных на построение эффективных торговых стратегий. Критический обзор существующих подходов к задаче моделирования динамики котировок и некоторые результаты численного анализа, ориентированного на построение таких моделей, рассмотрены в настоящей статье.

II. Моделирование динамики котировок: от броуновского движения до модели Самуэльсона

По-видимому, первым значимым подходом к формализации процессов ценообразования, протекающих на фондовых и иных биржах, была диссертационная работа французского математика Луи Башелье «Theorie de la speculation», представленная к защите в 1900 г. [1]. В основу его концепции лежала гипотеза о полной статистической независимости рядов котировок биржевых активов. Иными словами, корреляция между двумя следующими друг за другом наблюдения равна нулю, а автокорреляционная функция вырождается в $\delta-$*функцию* Дирака. Природным аналогом такого процесса является *броуновское движение*, открытое шотландским ботаником Робертом Брауном в 1827 г.

С математической точки зрения броуновское движение представляет собой *случайный процесс с независимыми приращениями* $\{X_t\}_{t\in T\subset[0,\infty)}$, у которого для любой возрастающей временной последовательности $t_0=0, t_1, ..., t_n$ случайные величины $X_{t_0}, X_{t_1}-X_{t_0}, ...,$ $X_{t_n}-X_{t_{n-2}}$ являются независимыми. В дальнейшем будем рассматривать простейшую равномерную дискретную временную последовательность $T=\{0, 1, ..., N\}$.

Одним из наиболее известных вариантов процесса с независимыми приращениями является процесс одномерных *случайных блужданий* $X_t = X_0+\sum_{k=1}^{t}\delta X_k$, где

$$\delta X_t = X_t - X_{t-1} = \begin{cases} 1, & \textit{с вероятностью } p_{t,} \quad 0<p_t<1, \\ -1, & \textit{с вероятностью } q_t = p_t-1, \end{cases}$$

последовательность независимых случайных величин. Применение этого процесса к моделированию ценообразования приведено, например, в [2].

Естественным обобщением случайного блуждания для моделирования динамики котировок явился винеровский процесс $W_t=\{X_t\}_{t\in T\subset[0,\infty)}$ [3], у которого независимые приращения ΔX_t представляет собой гауссовский процесс, т.е. для $\forall t, s\in[0,\infty)$, $s<t$, $\sigma = const$ $\Delta X_{ts}\in N\{0, \sigma^2\cdot|t-s|\}$.

Обозначим через v – случайную величину, подчиненную нормальному закону с параметрами $N\{0,1\}$. Тогда для винеровского процесса $\Delta X_{\Delta t}=v\sqrt{\Delta t}$ или, переходя в пределе к непрерывному случаю, $dX=v\sqrt{dt}$. Винеровский процесс и до сих пор остается базовой моделью динамики котировок биржевых активов. Типовыми недостатками модели, отражающими ее прикладную неадекватность процессам ценообразования торговых активов, считаются:

1. Наличие отрицательных значений котировок. Башелье для устранения этого недостатка предложил использовать логарифмическое преобразование, что существенно снизило наглядность и интерпретируемость процессов.

2. Существенные отклонения формы эмпирического распределения приращений ΔX_{ts} от гауссовой кривой $N\{0, \sigma^2 \cdot |t-s|\}$. В частности, применение критериев нормальности типа Харке-Бера (*JB*), основанного на равенстве нулю коэффициентов асимметрии и эксцесса, или непараметрического критерия Колмогорова [4, 5] подтверждает негауссовскую природу данных. 3. Существенная нестационарность рядов наблюдений.

3. Наличие хаотической составляющей в рядах наблюдений [6, 7].

Перечисленные несоответствия модели привели к различным обобщениям, усложняющим базовую модель.

Одним из направлений обобщения винеровской модели является добавление составляющей, имитирующей системный тренд αdt. Для непрерывного случая получаем соотношение, представляющее собой стохастическое дифференциальное уравнение Ито [8]: $dX = \alpha dt + \sigma v\sqrt{dt}$. В качестве σ обычно используется оценка параметра рассеянья (например, *среднеквадратическое отклонение* (ско)). Оценка ско осуществляется по ряду котировок на предшествующем участке наблюдения. Для дискретного времени данная модель (известная как модель Башелье) имеет вид $\Delta X = \alpha\Delta t + \sigma v\sqrt{\Delta t} \in N\{\alpha\Delta t, \sigma\sqrt{\Delta t}\}$.

Обобщением модели Башелье явилось стохастическое дифференциальное уравнение, в котором параметры тренда α и волатильности σ являются функциями от цены актива: $dX = \alpha(X,t)dt + \sigma(X,t)\varepsilon\sqrt{dt}$. Используя понятие среднего приращения цены μ и абсолютного дохода μX, получим дифференциальное соотношение $dX = \mu X dt$. С учетом винеровской компоненты можно записать уравнение в терминах ставки доходности dX / X

$$\frac{dX}{X} = \mu dt + \sigma\varepsilon\sqrt{dt} \in N\{\mu dt, \sigma\sqrt{dt}\} .$$

Приведенное уравнение изменения цены актива называют моделью Самуэльсона [9]. Решение этого уравнения имеет вид $X_t = X_0 \exp\left(\mu t + \sigma W_t - \frac{\sigma^2}{2}t\right)$.

где W_t – стандартный винеровский процесс.

III. Моделирование динамики котировок: временные ряды

Второе направление моделирование траекторий биржевых активов связано с применением моделей временных рядов. При этом подходе базовыми математическими конструкциями являются модели авторегрессии, скользящего среднего и их производные.

*Модель авторегрессии (*Autoregressive *process,* $AR(p)$*)* задается соотношением $X_t = \sum_{j=1}^{p} a_j X_{t-j} + v_t$,

где X_t, X_{t-i}, $i = 1,..., p$, – значения переменной x в соответствующие моменты времени; p – порядок модели; $\alpha_1,...,\alpha_p$ – параметры модели; v_t – случайная составляющая с нулевым математическим ожиданием, конечной дисперсией и единичной автокорреляционной матрицей, свидетельствующей об отсутствии автокорреляционной связи в рядах наблюдений, т. е. $v_t \in N\{0, \sigma_v^2\}$, $\mathrm{cov}(v) = \sigma_v^2 E$, E – единичная матрица. В некоторых случаях используется нотация на основе оператора лага $L^d X_t = X_{t-d}$. В этом случае процесс $AR(p)$ может быть записан в виде $(1 - \sum_{j=1}^{p} a_j L^j) X_t = v_t$.

Статистическая подгонка модели *AR(p)*, как и любой другой модели временного ряда, предполагает использование двухэтапной процедуры.

Модель скользящего среднего (Moving Average, MA(q)) определяется в виде линейной комбинации текущего и прошедших значений случайной составляющей $v_t, v_{t-1},..., v_{t-q}$, соответствующей «белому шуму»: $X_t = \sum_{j=1}^{q} \beta_j v_{t-j} + v_t$ где $\beta_1,...,\beta_q$ – параметры модели. При использовании оператора лага эту модель можно записать в виде $X_t = (1 - \sum_{j=1}^{q} \beta_j L^j) v_t = \beta(L) v_t$.

Модель авторегрессии-скользящего среднего (Autoregressive-Moving average, ARMA(p, q)) определяется как аддитивная комбинация выше рассмотренных моделей *AR(p)* и *MA(q):*

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + v_t - \sum_{j=1}^{q} \beta_j v_{t-j} + v_t ,$$

где $\alpha_1,...,\alpha_p$, $\beta_1,...,\beta_q$ – параметры модели; p – порядок авторегрессии; q – порядок скользящего среднего. Другая запись этой модели, использующая оператор лага, имеет вид $\alpha(L) X_t = \beta(L) v_t$.

Процесс, отвечающий модели *ARMA(p, q)* является стационарным, если корни уравнения $\alpha(L) = 0$ лежат вне единичного круга. Считается, что данная модель наиболее адекватна наблюдениям процесса, в котором «собственная» стохастика усугублена рядами внешних неконтролируемых возмущений, формируемых новостной информацией и групповой психологией участников торгов [10].

Модель *ARMA* оказалась весьма конструктивной и получила целый ряд обобщений. Очевидным обобщением описанных выше временных рядов являются модели с нелинейной зависимостью от предшествующих наблюдений и шумовой компоненты. Соответствующие нелинейные аналоги получили наименование нелинейной авторегрессии (*Nonlinear Autoregressive, NAR*), нелинейного скользящего среднего (*Nonlinear Moving Average, NMA*) и нелинейной авторегрессии-скользящего среднего (*NARMA*).

Важным обобщением *ARMA(p, q)* процесса является *модель Бокса-Дженкинса*, известная также как модель авторегрессии-интегрированного скользящего среднего *(Autoregressive Integrated Moving Average, ARIMA(p, d, q)* [11]. Данная модель используется для описания нестационарных процессов, обладающих некоторой однородностью. Однородность проявляется в том, что, если не учитывать локальных трендов, различные участки этого процесса до определенной степени подобны. Указанное свойство характерно для процессов, у которых конечная разность некоторого порядка d является стационарным процессом. Модель Бокса-Дженкинса обычно задается уравнением вида

$$\alpha(L)\nabla^d X_t = \beta(L)v_t \quad (1)$$

где ∇^d – оператор взятия конечной разности порядка *d*, *L* – оператор лага. Уравнение (1) может быть записано в виде:

$$(1-\sum_{i=1}^{p}\alpha_i L^i)(1-L)^d X_t = (1+\sum_{j=1}^{q}\beta_i L^i)v_t \quad (2)$$

Более лаконичную запись (2) можно получить, используя матричную нотацию: $\mathrm{A}(L)(1-L)^d X_t = \mathrm{B}(L)v_t$, где $L^d X_t = X_{t-d}$, $A(L) = 1-\alpha_1 L - ... - \alpha_p L^p$, $\mathrm{B}(L) = 1-\beta_1 L - ... - \beta_q L^q$. В качестве наглядного примера можно привести модель *ARIMA (0, d, 0),* для которой соотношение (2) будет иметь вид $(1-L)^d X_t = v_t$ или, в развернутом виде: $X_t - dX_t + \dfrac{d(d-1)}{2}X_{t-2} - ... = v_t$.

Как указывается в [11], модель (1, 2) соответствуют процессу на выходе неустойчивого (при $d \neq 0$) линейного фильтра, на входе которого – белый шум v_t.

Следующим обобщением данной линии моделей временных рядов является модель авторегрессии-дробно интегрированного скользящего среднего (Autoregressive Fractionally Integrated Moving Average, ARFIMA (p, d, q)), допускающая использование дробных значений дифференциального параметра d [10–14]. При этом в дробной модели для оператора лага используется биномиальное представление вида: $(1-L)^d = \sum_{k=1}^{\infty} C_d^k(-L)^k = 1 - dL + \dfrac{d(d-1)}{2}L^2 - ...$, что рассматривается, как математическая интерпретация долговременной памяти модели. Модель ARFIMA (p, d, q) описывается тем же соотношением (2), что и модель ARIMA (p, d, q), однако величина d, как уже указывалось, является дробной.

В случае если модели временного ряда типа *ARMA* учитывает сезонные или периодические эффекты, то соответствующие обобщения образуют класс сезонных (*seasonal*) моделей типа *SAR*, *SARMA*, *или SARIMA* [15]. Поиск сезонных процессов в хаотических рядах наблюдений, содержащих явно выраженную непериодическую компоненту, представляется мало перспективным занятием. Расширение этого же класса моделей на случай многомерных (*multivariate*) временных рядов [16] приводит к классу векторных моделей типа *VAR (Vector AR), VMA (Vector MA)* или *VARMA (Vector ARMA).* Многомерное направление моделирование позволяет осуществлять совместное исследование динамики различных инструментов с учетом их взаимосвязи, однако алгоритмы идентификации и прогноза становятся существенно более сложными. В ситуациях, связанных с *нелинейными* (nonlinear) моделями временных рядов используется спецификация вида *NAR, NMA, или NARMA.* Для проверки исходного ряда на нелинейность можно использовать тест *BDS* (*Brock-Dechert-Scheinkman*), разработанный для эконометрических моделей [17].

Говорить о повышенной эффективности нелинейных моделей можно только в контексте конкретной задачи с определенными критериями качества ее решения и совокупностью ограничений. Модель *ARCH* (*Autoregressive Conditional Heteroskedasticity*), рассмотренная в [18], использует стандартное представление случайной составляющей наблюдений в виде $v_t = \sigma_t \cdot n_t(0,1)$, где $n_t(0,1)$ – ряд наблюдений, моделируемый независимой нормально распределенной случайной величиной с параметрами (0, 1), а σ_t задается соотношением $\sigma_t^2 = \gamma_0 + \sum_{i=1}^{m}\gamma_i v_{t-i}^2, \quad \gamma_0 > 0, \quad \gamma_i \geq 0, \quad \forall i > 0.$

Развитие этого направления привело к возникновению таких моделей, как *GARCH*, *NGARCH*, *IGARCH*, *EGARCH*, *GARCH-M*, *QGARCH*, *GJR-GARCH*, с которыми можно познакомиться, например, в [19] или на сайте [20]. Другой пример нелинейных моделей связан с учетом группы детерминированных экзогенных факторов $U_{<t,\,t-s>} = (u_t, ..., u_{t-s})$ [21]. Общий вид нелинейной модели авторегрессии с экзогенными факторами (*nonlinear autoregressive exogenous model*, *NARX*) можно записать в виде $X_t = f(X_{<t,\,t-p>},\ U_{<t,\,t-s>}) + v_t$, где $X_{<t,\,t-p>} = (x_t, ..., x_{t-p})$, f – нелинейная функция времени, в качестве которой часто используют полиномы.

Следует упомянуть процессы с двойной стохастичностью, когда один или несколько параметров в исходной модели $X_t = f(X_{<t,\,t-p>}) + v_t$ сами являются случайными величинами. Для одномерной задачи данный подход известен как концепция смешанных распределений (*compounded distributions*). Данная технология моделирования аналогична моделям со скрытыми (латентными) переменными [22].

Приведенный обзор представляет собой неполный перечень основных направления в моделировании динамики котировок биржевых активов.

IV. Критический анализ

Зачем нужно моделировать котировки? Прежде всего, для создания полигона данных, необходимого для тестирования торговых стратегий, советников, торговых роботов и других программно-алгоритмических средств, ориентированных на управление торговыми операциями.

Вторая задача моделирования котировок состоит в использовании модели в качестве базового элемента торговой стратегии. Ее структура должна определяться лишь терминальными показателями эффективности торговой стратегии и соответствующим ей торговым операциям [6]. Трейдерская практика показывает, что простейшее применение экстраполяционной прогностической модели не позволяет построить эффективную торговую стратегию. Это связано с существенной нестационарностью случайной составляющей динамики котировок, обусловленной, как будет показано ниже, наличием хаотической компоненты рядов наблюдений [23].

В качестве третьей задачи моделирования можно выделить задачу обнаружения и идентификации статистических свойств «квазистационарных» участков случайной составляющей динамики котировок.

Гипотеза о принципиальной возможности построения эффективной стратегии имеет право на существование, о чем свидетельствует успешная работа. Возможным вариантов реализации такой стратегии являются робастифицированные алгоритмы, обладающие пониженной чувствительностью к вариациям статистической структуры случайной составляющей динамики котировок [24].

V. Заключение

Следует заметить, что на практике, критерий нормальности может не выполняться, так как сглаживающая технология, использованная для идентификации квазирегулярной модели, приводит, как уже отмечалось выше, к возникновению заметного эксцесса. Главным выводом из приведенных исследований является заключение о целесообразности применения трехкомпонентной аддитивной модели динамики процессов изменения котировок, принципиально отличающейся от традиционной модели наблюдений с аддитивными шумами:

- спецификой данных, полученных в процессе мониторинга за торговой ситуацией, и практически не содержащих погрешностей наблюдений;
- наличием субъективной составляющей, формируемой в процессе выбора торговой стратегии и вычислительной схемы формирования системной составляющей;
- наличием хаотической компоненты в форме колебательного непериодического процесса.

Список литературы

[1] Bachelier L. Theorie de la speculation. - Annales scientifiques de l'E.N.S. 1900. Ser. 3. Tome 17. P. 21–86.

[2] The Random Character of Stock Market Prices. - Ed. by P.H. Cootner. Cambridge: MIT Press. 1964.

[3] Miller S., Childers D. Probability and random processes with applications to signal processing. 3rd edition. New Jersey: Prentice Hall. 2002. 552p.

[4] Doane D., Seward L. Applied Statistics in Business and Economics Applied Statistics in Business & Economics. 3rd Edition. — McGraw-Hill/IrwinInc. 2010. 864 p.

[5] Ullah A. Handbook of Applied Economic Statistics. CRC Press Inc. 1998. 640p.

[6] Musaev A.A. Quod est veritas. Point of view transformation on a system component of observed process. SPIIRAS Proc., 2010. №15. P. 53–74.

[7] Peters E.E. Chaos and order in the capital markets: a new view of cycles, prices, and market volatility (2nd ed.) / NY: John Wiley & Sons, 1996. 288p.

[8] Ito K. Essentials of stochastic processess Providence: AMS, 2006. 170 p.

[9] Samuelson P. Foundations of Economic Analysis / Cambridge, MA: Harvard University Press, 1947. 850p.

[10] Autoregressive-moving average model [электронный ресурс http://en.wikipedia.org/wiki /Autoregressive_ moving_average_model].

[11] Box G.E.P., Jenkins G.M., Reinsel G. Time series analysis: Forecasting and control. Wiley: 2008. 784p.

[12] Autoregressive fractionally integrated moving average [электронный ресурс http:// en.wikipedia.org/wiki/Autoregressive_fractionally_integrated_moving_average].

[13] Granger C.W.J., Joyeux R. An introduction to long-memory time series and fractional differencing // Journal of Time Series Analysis, 1980. V.1. P. 15–29.

[14] Hosking J.R.M. Fractional differencing. // Biometrika. 1981. V. 68(1). P. 165–176.

[15] Cox D.R. Some statistical methods connected with series of events // J. R. Statist. Soc. Ser. B. 1955. V.17. P.129–164.

[16] Reinsel G.C. Elements of Multivariate Time Series Analysis (Springer Series in Statistics). Berlin: Springer. 2001. 358 p.

[17] Brock W.A., Scheinkman J.A., Dechert W.D., LeBaron B. A Test for independence based on the correlation dimension // Econometric Reviews. 1996. V.15. P. 197–235.

[18] Engle R.F., Ng. V.K. Measuring and testing the impact of news on volatility // Journal of Finance. 1982. v.48 (5). P. 1749–1778.

[19] Ljung L. System Identification: Theory for the User. (2nd Edition). Prentice Hall. 1999. 609p.

[20] Autoregressive conditional heteroskedasticity [электронный ресурс http://en.wikipedia.org/wiki /Autoregressive_ conditional_heteroskedasticity].

[21] Nelles O. Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models Book Description / Berlin: Springer Berlin, 2000. 785 p.

[22] Doubly stochastic model [электронный ресурс http://en.wikipedia.org/wiki/Doubly_ stochastic_model#cite_note-0#cite_note-0].

[23] Musaev A.A. Numerical analysis of chaotic processes persistence. SPIIRAS Proc., 2014. №2(33). P. 48–59.

[24] Huber P.J. Robust Statistics. NY: John Wiley & Sons, 1981. - 320p.

[25] Luenberger D.J. Introduction to Dynamic Systems. NY: Wiley, 1979. 446 pp.

[26] Kalman, P. L. Falb, M. A. Arbib. Topics in mathematical system theory. NY: McGraw Hill Book Co., 1969. 355p.