



Босов А.Д., [Орлов Ю.Н.](#)

Моделирование
нестационарных временных
рядов с помощью
эмпирического уравнения
Лиувилля и уравнений
эволюции моментов

Рекомендуемая форма библиографической ссылки: Босов А.Д., Орлов Ю.Н. Моделирование нестационарных временных рядов с помощью эмпирического уравнения Лиувилля и уравнений эволюции моментов // Препринты ИПМ им. М.В.Келдыша. 2011. № 52. 28 с. URL: <http://library.keldysh.ru/preprint.asp?id=2011-52>

РОССИЙСКАЯ АКАДЕМИЯ НАУК

Ордена Ленина

ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ

им. М.В. Келдыша

А.Д. Босов, Ю.Н. Орлов

МОДЕЛИРОВАНИЕ НЕСТАЦИОНАРНЫХ ВРЕМЕННЫХ
РЯДОВ С ПОМОЩЬЮ ЭМПИРИЧЕСКОГО УРАВНЕНИЯ
ЛИУВИЛЛЯ И УРАВНЕНИЙ ЭВОЛЮЦИИ МОМЕНТОВ

Москва, 2011

А.Д. Босов¹, Ю.Н. Орлов²

¹МФТИ (ГУ)

²ИПМ им. М.В. Келдыша РАН

Моделирование нестационарных временных рядов с помощью эмпирического уравнения Лиувилля и уравнений эволюции моментов

Строятся гидродинамические модели нестационарных временных рядов на основе горизонтной статистики. Они включают систему эволюционных уравнений для эмпирических функций распределения и ее моментов. Приведен пример моделирования эволюции нестационарного распределения и собственно ряда.

A.D. Bosov, Yu.N. Orlov

The non-stationary time series modeling with the use of empirical Liouville equation and evolution equations for momentum

The hydrodynamic models for non-stationary time series are constructed with the use of horizon statistics. The system contains the Liouville equation for empirical distribution function and the evolution equations for the momentum. The example of time series modeling is given.

Содержание

1. ВВЕДЕНИЕ.....	4
1.1. Цель и направление исследований.....	4
1.2. Актуальность темы	5
1.3. Новизна разработки.....	5
2. ЭМПИРИЧЕСКОЕ УРАВНЕНИЕ ЛИУВИЛЛЯ.....	6
2.1. Нестационарные временные ряды.....	6
2.2. Иерархия зацепляющихся уравнений эволюции	7
2.3. Уравнение Лиувилля для ВПФР	8
3. МОДЕЛИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ С ПОМОЩЬЮ КИНЕТИЧЕСКИХ УРАВНЕНИЙ.....	11
3.1. Вывод уравнений эволюции многомерных ВПФР	11
3.2. Модели временных рядов.....	14
4. МОДЕЛИ ГИДРОДИНАМИЧЕСКОГО ТИПА	17
4.1. Концепция моделирования временного ряда	17
4.2. Методы замыкания моментной системы.....	19
4.3. Уточнение способов замыкания	22
5. ПРИМЕР МОДЕЛИРОВАНИЯ ВРЕМЕННОГО РЯДА.....	24
5.1. Временной ряд курса доллара к рублю	24
5.2. Гидродинамическая модель эволюции ВПФР.....	26
ИСТОЧНИКИ.....	28

1. ВВЕДЕНИЕ

1.1. Цель и направление исследований

Цель работы состоит в том, чтобы использовать метод Н.Н. Боголюбова [1, 2] вывода уравнений гидродинамики из системы зацепляющихся уравнений для частичных функций распределения (цепочки Боголюбова) для построения моделей нестационарных временных рядов.

Идея предлагаемого подхода опирается на представление о случайном процессе как о нестационарной динамической системе. Если динамическая система известна, то можно выписать систему зацепляющихся уравнений эволюции относительно частичных функций распределения. Уравнения эволюции моментов этих распределений прямо следуют из цепочки. Если из некоторых априорных соображений оборвать цепочку – либо для функций распределения, либо для моментов, – то получится замкнутое статистическое описание макроскопических свойств изучаемой динамической системы. Обычно метод Боголюбова применяют к гамильтоновым системам с большим (в пределе – бесконечным) числом степеней свободы.

В математической статистике для прогноза значений случайного процесса также используют представление об этом процессе как о реализации некоторой динамической системы. Например, вычислив среднее значение стационарного временного ряда по достаточно большой выборке, полагают затем прогнозное значение ряда равным этому среднему. Это – простейшая динамическая система. Можно найти автокорреляцию между соседними значениями временного ряда и построить авторегрессионную модель первого порядка, которая также представляет собой динамическую систему: это рекуррентно задаваемая функция от дискретного аргумента.

Существует большое число различных моделей для стационарных временных рядов [3], однако они не подходят для моделирования нестационарных рядов, каковыми являются на практике большинство случайных процессов. Концепция построения моделей таких рядов была предложена в [4]. Она состоит в том, чтобы по выборочным распределениям временного ряда, построенным в различные моменты времени, определить эмпирическое уравнение эволюции этих распределений, построить зацепляющуюся систему уравнений для распределений, выписать уравнения эволюции их моментов, а затем замкнуть систему и решить ее относительно самих нестационарных распределений. Прогноз временного ряда будет определяться первым моментом эмпирического распределения, а модель этого ряда будет представлять собой динамическую систему, т.е. систему уравнений для моментов

распределения, наподобие того, как уравнения гидродинамики в статистической механике моделируют движение большого числа частиц посредством сведения его к движению одной частицы со средними характеристиками.

1.2. Актуальность темы

В настоящее время нестационарный анализ данных представляет большую практическую важность и востребованность в различных сферах человеческой деятельности: прогнозирование погоды в метеорологии, анализ функционирования организма человека в биомедицине, прогноз ценовых рядов на финансовых и сырьевых рынках, корректная обработка результатов социологических опросов. Модели стационарных процессов не отвечают потребностям специалистов в этих областях, поскольку они дают слишком большую ошибку. Поэтому построение статистически корректной модели случайного процесса, учитывающей свойство его нестационарности, является актуальной задачей. Трудность представляет математически строгое обоснование достоверности статистических оценок, а также вывод самой модели динамической системы из уравнений эволюции эмпирических статистик.

1.3. Новизна разработки

Основными статистическими методами исследования временных рядов являются: метод выделения тренда (временного сглаживания), регрессионный, автокорреляционный, скользящих средних, адаптивный, метод гармонического анализа, сингулярного спектрального анализа, бутстреп (численное размножение выборок) и нейросетевой метод.

Перечисленные подходы корректно применимы только к стационарным процессам, в которых средние величины (математическое ожидание, дисперсия, ковариация и т.п.) постоянны во времени. Уточнение этих моделей в случае зависимости указанных величин от времени может быть сделано посредством аналитического моделирования такой зависимости, либо переходом к первым, вторым и т.д. разностям в нестационарных временных рядах с полиномиальным трендом, либо регрессией изучаемых величин на время. Окно усреднения становится при этом скользящим, однако остается невыясненным, какой ширины должно быть это окно.

Неоптимальность объема выборки и неизвестный горизонт прогноза, на котором этот прогноз имеет гарантированную точность, например, в среднем квадратичном, являются основными препятствиями на пути применения стационарных моделей к нестационарным временным рядам. Также следует сказать, что если временной ряд нестационарный, то классические теоремы об эффективности, состоятельности и

асимптотической нормальности выборочных оценок и их дисперсий в общем случае не выполняются.

Первая модель нестационарного временного ряда, полученная с использованием метода эволюции выборочных распределений, была построена в работе [4]. В монографии [5] был описан класс моделей, получающихся из эмпирического уравнения Лиувилля для выборочной функции распределения нестационарного временного ряда.

В настоящей работе впервые строится гидродинамическая модель временного ряда, получаемая как обрыв моментной системы на уравнении эволюции момента второго порядка совместной функции распределения случайной величины и ее приращений. Эта модель является новой, она представляет собой реализацию подхода, намеченного в [5].

2. ЭМПИРИЧЕСКОЕ УРАВНЕНИЕ ЛИУВИЛЛЯ

2.1. Нестационарные временные ряды

Временной ряд $x(t)$, заданный в дискретные моменты времени, для удобства считающиеся равноотстоящими, называется стационарным в узком смысле, если его плотность распределения $f(x)$ стационарна. Ряд называется стационарным в широком смысле, если его среднее значение постоянно, а автокорреляционная функция зависит только от разности моментов времени. В противном случае при невыполнении любого из вышеприведенных условий ряд называется нестационарным.

Выборочной плотностью функции распределения (далее ВПФР) $f_T(x, t)$ будем называть ВПФР, построенную по выборке данных $x(t), x(t-1), \dots, x(t-T+1)$. Расстоянием между двумя ВПФР $f_1 = f_{T_1}(x, t_1)$ и $f_2 = f_{T_2}(x, t_2)$ называется величина [4, 5]

$$\rho(f_1, f_2) = \|f_{T_1}(x, t_1) - f_{T_2}(x, t_2)\| = \int |f_{T_1}(x, t_1) - f_{T_2}(x, t_2)| dx. \quad (2.1)$$

Используя (2.1), можно определить вероятность того, что две ВПФР, сдвинутые одна относительно другой на τ шагов, интегрально отличаются не более чем на заданную величину ε . Соответствующий интегральный функционал обозначим через $V(T, \tau; t)$:

$$V(T, \tau; x, t) \equiv \rho(f_T(x, t + \tau), f_T(x, t)) = \int_0^1 |f_T(x, t + \tau) - f_T(x, t)| dx \leq \varepsilon. \quad (2.2)$$

ВПФР $f_T(x, t)$ временного ряда $x(t)$ будем называть ε -стационарной, если выполнено условие (3.2), т.е. если $V(T, \tau; t) \leq \varepsilon$ для всех моментов времени t .

2.2. Иерархия зацепляющихся уравнений эволюции

Введем бесконечномерное фазовое пространство с мерой

$$d\Gamma = F_\infty \left(x, \dot{x}, \ddot{x}, \dots, x^{(k)}, \dots; t \right) \prod_{k=0}^{\infty} dx^{(k)}, \quad (2.3)$$

представляющей число кинематических состояний частиц в границах указанных промежутков. Определим n -частичные ПФР $f_n(x, \dot{x}, \dots, x^{(n)}; t)$ согласно формулам теории вероятностей для плотностей совместных многомерных распределений:

$$\begin{aligned} f_n(x, \dot{x}, \dots, x^{(n)}; t) &= \int F_\infty \left(x, \dot{x}, \ddot{x}, \dots, x^{(k)}, \dots; t \right) \prod_{k=n+1}^{\infty} dx^{(k)} = \\ &= \int f_{n+1} \left(x, \dot{x}, \dots, x^{(n)}, x^{(n+1)}; t \right) dx^{(n+1)}. \end{aligned} \quad (2.4)$$

После этого, считая, что при стремлении скоростей и ускорений всех порядков к бесконечности ПФР стремится к нулю, из уравнения Лиувилля для F_∞ , выражающего закон сохранения нормировки меры

$$\frac{\partial F_\infty}{\partial t} + \sum_{k=0}^{\infty} \operatorname{div}_{x^{(k)}} \left(x^{(k+1)} F_\infty \right) = 0, \quad (2.5)$$

и из соотношений (2.4) получаем уравнения эволюции для n -частичных ПФР:

$$\begin{aligned} \frac{\partial f_0(x, t)}{\partial t} + \operatorname{div}_x \int \dot{x} f_1(x, \dot{x}; t) d\dot{x} &= 0; \\ \frac{\partial f_1(x, \dot{x}; t)}{\partial t} + \operatorname{div}_x \dot{x} f_1(x, \dot{x}; t) + \operatorname{div}_{\dot{x}} \int \ddot{x} f_2(x, \dot{x}, \ddot{x}; t) d\ddot{x} &= 0; \quad \dots; \\ \frac{\partial f_n(x, \dot{x}, \dots, x^{(n)}; t)}{\partial t} + \sum_{k=0}^{n-1} \operatorname{div}_{x^{(k)}} x^{(k+1)} f_n(x, \dot{x}, \dots, x^{(n)}; t) + \\ + \operatorname{div}_{x^{(n)}} \int x^{(n+1)} f_{n+1}(x, \dot{x}, \dots, x^{(n+1)}; t) dx^{(n+1)} &= 0; \quad \dots \end{aligned} \quad (2.6)$$

Таким образом, уравнение эволюции для каждой частичной ПФР не является замкнутым. Оно связывает между собой концентрации и потоки, причем последние выражаются через ПФР от большего числа переменных. Если на каком-либо этапе оборвать зацепление, т.е. задать интеграл в (2.6) независимо, то в пределе нулевых центральных моментов всех порядков относительно координат, скоростей и ускорений всех порядков будет получаться та или иная динамическая система. Уравнения (2.6) служат основой для конструирования эмпирической модели временного ряда.

Зацепляющиеся кинетические уравнения получаются также и для многомерных динамических систем. Важным примером таких систем, для которых, собственно, и был развит метод кинетических уравнений, являются гамильтоновы системы с большим (в

пределе – бесконечным) числом переменных. Метод вывода эволюционных уравнений для частичных ПФР, отвечающих многомерной гамильтоновой системе, был разработан Боголюбовым [1, 2]. В дальнейшем, выбирая способ обрыва цепочки на каком-либо уровне, можно получить замкнутые уравнения относительно соответствующих ПФР. Методологическая важность уравнений цепочки Боголюбова в том, что они лежат в основе уравнений классической гидродинамики. Эти уравнения представляют собой законы эволюции моментов функций распределения.

Как и уравнения цепочки (2.6), уравнения для моментов не образуют замкнутой системы. Например, пусть $f(x, \dot{x}; t)$ представляет собой совместную ПФР значений случайной величины и ее приращений (скоростей). Введем координатную плотность

$$\rho(x, t) = \int f(x, \dot{x}; t) d\dot{x}. \quad (2.7)$$

Из (2.6) следует, что уравнение эволюции для плотности (2.7), т.е. для нулевого момента ПФР $f(x, \dot{x}; t)$ по второму аргументу, выражается через первый момент:

$$\frac{\partial \rho(x, t)}{\partial t} = -\frac{\partial}{\partial x}(\rho(x, t)u(x, t)), \quad u(x, t) = \frac{1}{\rho(x, t)} \int \dot{x} f(x, \dot{x}; t) d\dot{x}. \quad (2.8)$$

В свою очередь, уравнение для первого момента (т.е. для средней скорости $u(x, t)$) выражается через моменты второго порядка и т.д.:

$$\begin{aligned} \frac{\partial}{\partial t}(\rho(x, t)u(x, t)) &= -\frac{\partial}{\partial x}(\rho(x, t)e(x, t)) + \rho(x, t)w(x, t), \\ e(x, t) &= \frac{1}{\rho(x, t)} \int \dot{x}^2 f(x, \dot{x}; t) d\dot{x}, \quad w(x, t) = \frac{1}{\rho(x, t)} \int \ddot{x} f(x, \dot{x}, \ddot{x}; t) d\dot{x} d\ddot{x}. \end{aligned} \quad (2.9)$$

При моделировании временного ряда можно привлечь похожую аналогию, используя понятие квазистационарной выборки для построения уравнения эволюции ВПФР.

2.3. Уравнение Лиувилля для ВПФР

Предположительно, модель эволюции ВПФР должна основываться на некотором аналоге уравнения Лиувилля (2.6) классической статистической механики, выражающем закон сохранения вероятности со временем, и содержать алгоритм его решения с заданной точностью. Тем самым выборочной функции распределения будет локально по времени сопоставлена некоторая динамическая система.

Предположим, что имеется достаточно большое количество данных $\{x(t)\}$, анализ которых позволяет определить достаточный объем T выборки для построения ВПФР $f_T(x, t)$ и ее прогнозирования, скажем, на один шаг вперед с точностью не хуже ε .

Рассматривая набор таких ВПФР $f_T(x, t)$, $f_T(x, t-1)$, ... в предшествующие моменты времени, можно определить соответствующие эмпирические изменения этих плотностей за один шаг:

$$\frac{\partial f_T(x, t-k)}{\partial t} = f_T(x, t-k+1) - f_T(x, t-k), \quad k = 1, 2, \dots \quad (2.10)$$

Частная производная по времени в (2.10) написана для удобства восприятия. Фактически это есть обозначение для выражения, стоящего в правой части (2.10). Определим теперь «скорость» изменения ВПФР $u_T(x, t)$ так, чтобы (2.10) можно было записать в виде

$$\frac{\partial f_T(x, t)}{\partial t} + \frac{\partial}{\partial x} u_T(x, t) f_T(x, t) = 0. \quad (2.11)$$

Скорость $u_T(x, t)$ была введена в [5]. Она называется эмпирической лиувиллевой скоростью (ЭЛС). Далее считаем объем выборки T фиксированным, равным оптимальному объему, при котором выполняется (2.2), и соответствующий индекс для краткости опускаем.

ЭЛС можно определить для каждой дискретизации области значений $x(t)$. При этом удобно считать, что на границах этой области ВПФР равна нулю. Для равномерно ограниченного ряда $x(t)$ этому условию всегда можно удовлетворить, введя нужное количество дополнительных граничных ячеек разбиения, в которых плотность распределения равна нулю во все моменты времени.

В результате ЭЛС в каждой ячейке разбиения, где существуют значения временного ряда, определяется из соотношений (2.10) и (2.11) по формулам

$$u(i+1, t) = \frac{u(i, t) f(i, t) - f(i, t) + f(i, t+1)}{f(i+1, t)}. \quad (2.12)$$

Если же в i -ой ячейке плотность распределения $f(i, t)$ оказалась равной нулю, то формально скорость в такой ячейке не определена. В фиктивных граничных ячейках скорость естественно положить равной нулю. Во внутренних ячейках с нулевой плотностью скорость можно определить различными способами. Если рассматривается только дискретная схема, то для единообразия определения скорость во всех таких ячейках можно положить равной нулю. Это не приведет к противоречиям, поскольку уравнение (2.11) содержит скорость в комбинации с произведением на соответствующую плотность. Если же предполагается, что дискретная схема аппроксимирует некоторое непрерывное распределение, то желательно также считать, что и скорость является непрерывной по x величиной в пределе уменьшения мелкости разбиения до нуля. Тогда ее можно

доопределить в указанных ячейках, например, как среднее арифметическое скоростей в ближайших к ней левой и правой ячейках с ненулевыми скоростями.

Статистический смысл эмпирической скорости эволюции ВПФР выясняется с помощью следующего утверждения.

Теорема 1. Пусть $f_T(x, t)$ - ВПФР, построенная по выборке объема T в момент времени t , и $u_T(x, t)$ - эмпирическая скорость, так что $f_T(x, t)$ удовлетворяет уравнению

$\frac{\partial f_T(x, t)}{\partial t} + \frac{\partial}{\partial x} u_T(x, t) f_T(x, t) = 0$. Тогда скорость изменения среднего выборочного значения

временного ряда совпадает со средней скоростью, определяемой уравнением

$$U_T(t) = \int u_T(x, t) f_T(x, t) dx.$$

Доказательство. Рассмотрим уравнение эволюции первого выборочного момента, определяемого как

$$m(t) = \langle x \rangle_t = \int x f(x, t) dx. \quad (2.13)$$

Из уравнения Лиувилля следует, что эволюция первого момента записывается в виде

$$\frac{dm(t)}{dt} = \int x \frac{\partial f(x, t)}{\partial t} dx = - \int x \frac{\partial}{\partial x} u(x, t) f(x, t) dx.$$

После интегрирования по частям с учетом того, что в граничных ячейках плотность распределения равна нулю, получаем

$$\frac{dm(t)}{dt} = \int u(x, t) f(x, t) dx = \langle u \rangle_t = U(t), \quad (2.14)$$

где $U(t)$ есть среднее значение эмпирической скорости $u(x, t)$ по распределению, которое задано в момент времени t .

С другой стороны, то же самое изменение первого выборочного момента есть

$$\frac{dm(t)}{dt} = m(t+1) - m(t) = \frac{1}{T} \sum_{k=1}^T x(t-T+k+1) - \frac{1}{T} \sum_{k=1}^T x(t-T+k) = \frac{1}{T} \sum_{k=1}^T \dot{x}(t-T+k), \quad (2.15)$$

т.е. $U(t)$ есть среднее значение выборочной скорости, определяемой по приращениям значений исходного временного ряда.

Выражение (2.15), как и (2.14), представляет среднее значение одной и той же скорости, с точностью до ошибки кластеризации данных, связанной с конечным числом разбиений промежутка изменения случайной величины при построении ВПФР. Но, в отличие от (2.14), среднее в (2.15), будучи скоростью \dot{x} , должно вычисляться не через распределение $f(x, t)$ значений ряда, а через распределение его приращений. Введем тогда двумерную ВПФР $F(x, \dot{x}, t)$ совместного распределения случайных величин x и \dot{x}

исходного ряда и ряда его производной, полученного взятием разностей $x(t+1) - x(t)$ в соседние моменты времени. Эта разность и есть аналог микроскопической скорости системы из T «частиц» (значений ряда).

Одномерная ВПФР $f(x, t)$ определяется затем по $F(x, \dot{x}, t)$ формулой

$$f(x, t) = \int F(x, \dot{x}, t) d\dot{x}. \quad (2.16)$$

Тогда среднее значение эмпирической скорости из правой части (2.15) равно

$$U(t) = \int \dot{x} F(x, \dot{x}, t) dx d\dot{x},$$

и эта же величина равна выражению $\int u(x, t) f(x, t) dx$ в (2.14).

Следовательно, средняя локальная скорость $u(x, t)$ для выборочных распределений определяется формально так же, как и для генеральных совокупностей:

$$u(x, t) f(x, t) = \int \dot{x} F(x, \dot{x}, t) d\dot{x}. \quad (2.17)$$

Теорема 1 доказана.

Заметим теперь, что одного лишь уравнения Лиувилля методически недостаточно для практического моделирования эволюции ВПФР. Дело в том, что для построения двумерной ВПФР $F(x, \dot{x}, t)$ в момент времени t необходимо иметь значение случайной величины в момент $t+1$, чтобы можно было построить микроскопическую скорость в момент времени t . Из соотношений (2.12) следует, что скорость изменения ВПФР в данный момент времени определяется значениями распределения в этот и последующий моменты. Это означает, что в текущий момент времени t ЭЛС известна для предыдущего момента $t-1$, что следует учитывать при составлении прогнозной модели. Именно, чтобы спрогнозировать ВПФР в момент времени $t+1$, надо спрогнозировать ЭЛС в момент времени t . Следовательно, необходимо иметь уравнение эволюции для самой ЭЛС. Такое уравнение может быть выведено с помощью уравнения Лиувилля тем же способом, как это делается в классической гидродинамике.

3. МОДЕЛИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ С ПОМОЩЬЮ КИНЕТИЧЕСКИХ УРАВНЕНИЙ

3.1. Вывод уравнений эволюции многомерных ВПФР

Как было установлено выше, уравнение Лиувилля (2.11) не является замкнутым: кроме неизвестной плотности вероятности $f(x, t)$, относительно которой оно записано, в нем фигурирует и неизвестная скорость $u(x, t)$. Если исходить только из требования сохранения нормировки ВПФР, то скорость $u(x, t)$ может быть, вообще говоря, любой

функцией, поскольку на нормировку распределения ее конкретный вид не влияет. Однако важно, чтобы вычисленная по уравнению (2.11) плотность $f(x, t)$ была бы неотрицательна, что накладывает на функцию $u(x, t)$ определенные ограничения.

Дополним уравнение (2.11) уравнением (2.17), определяя тем самым функцию эмпирической скорости. Полученная система все равно не является замкнутой в смысле кинетического подхода, поскольку возникающая в (2.17) двумерная ВПФР $F(x, \dot{x}, t)$ пока еще не определена как функция времени. Для нее также следует записать уравнение Лиувилля, в котором x и \dot{x} формально считаются независимыми переменными.

По аналогии с одномерным случаем, производную скорости по времени следует рассматривать в некотором усредненном смысле, для чего надо расширить пространство фазовых переменных. Тогда уравнение Лиувилля для $F(x, \dot{x}, t)$ имеет вид

$$\frac{\partial F}{\partial t} + \frac{\partial}{\partial x}(F\dot{x}) + \frac{\partial}{\partial \dot{x}}\left(F\left\langle\frac{\delta \ddot{x}}{\delta t}\right\rangle\right) = 0. \quad (3.1)$$

Смысл этого уравнения определяется тем, как трактуется величина $\langle \delta \ddot{x} / \delta t \rangle$. Если рассматривается динамическая система, для которой известна зависимость ускорения $\ddot{x} = w(x, \dot{x}, t)$ от координат, скоростей и времени, то следует считать, что $\langle \delta \ddot{x} / \delta t \rangle$ и есть это ускорение. Тогда из (3.1) получаем

$$\frac{\partial F}{\partial t} + \frac{\partial}{\partial x}(F\dot{x}) + \frac{\partial}{\partial \dot{x}}(Fw) = 0. \quad (3.2)$$

Если же считать, что $\delta \ddot{x} / \delta t$ является случайной величиной, то в уравнении Лиувилля аналог динамического ускорения если и можно ввести, то в некотором усредненном смысле, т.е. в виде $\langle \delta \ddot{x} / \delta t \rangle$. Разумеется, это не «вывод» уравнения эволюции, а всего лишь эвристические соображения относительно того, каким могло бы быть кинетическое уравнение выборочной плотности функции распределения для нестационарного временного ряда.

Для определения величины $\langle \delta \ddot{x} / \delta t \rangle$ потребуется дальнейшее расширение фазового пространства. Положим

$$w(x, \dot{x}, t)F(x, \dot{x}, t) = \int \ddot{x}F_3(x, \dot{x}, \ddot{x}, t)d\ddot{x}. \quad (3.3)$$

Видно, что и для выборочных статистик возникает зацепление многочастичных распределений. Замыкание цепочки можно провести, задав априори выражение для интеграла от неизвестной величины $\int \ddot{x}F_3(x, \dot{x}, \ddot{x}, t)d\ddot{x}$ в правой части уравнения (3.3).

Характерно, что такое замыкание совершенно не повлияет на вид уравнения для одномерной ВПФР в силу дивергентного характера слагаемых.

Если же нет оснований считать интеграл $\int \ddot{x}F_3(x, \dot{x}, \ddot{x}, t)d\ddot{x}$ известным, то цепочка уравнений Лиувилля выписывается для ВПФР следующего порядка, и т.д. Однако, чем выше порядок ВПФР, тем труднее найти разумное основание для того или иного способа обрыва цепочки кинетических уравнений. Следовательно, желательно построить систему уравнений, замкнутую относительно одномерной ВПФР. Чтобы замкнуть уравнение Лиувилля (2.6), которое используется для прогнозирования ВПФР $f(x, t)$, следует независимо определить величину $u(x, t)$ как значение средней локальной скорости.

Для получения замкнутой теории будем исходить из определения (2.17) и уравнения эволюции двумерной ВПФР (3.2). Дифференцируя (2.17) по времени, получаем с учетом (3.2) и при условии независимости \dot{x} от x эволюционное уравнение для средней локальной скорости, аналогичное (2.9):

$$\frac{\partial}{\partial t}(uf) = -\frac{\partial}{\partial x}(ef) + wf, \quad (3.4)$$

где $w(x, t)$ определяется через ВПФР третьего порядка. Далее, из уравнения эволюции для $f(x, t)$ вида $\frac{\partial f}{\partial t} = -\frac{\partial}{\partial x}(fu)$ можно выразить производную $\frac{\partial f}{\partial t}$ и подставить его в уравнение (3.4). В результате получаем

$$\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x}\right)f = -\frac{\partial}{\partial x}(e - u^2)f + wf. \quad (3.5)$$

Таким образом, оказалось, что для определения эволюции первого момента двумерной ВПФР по скорости надо знать эволюцию ее второго момента по той же переменной и эволюцию среднего ускорения. Для этих величин аналогичными вычислениями получаем:

$$\left(\frac{\partial e}{\partial t} - e\frac{\partial u}{\partial x}\right)f - ue\frac{\partial f}{\partial x} = -\frac{\partial}{\partial x}(\langle \dot{x}^3 \rangle f) + \langle \dot{x}w \rangle f, \quad (3.6)$$

$$\left(\frac{\partial w}{\partial t} - w\frac{\partial u}{\partial x}\right)f - uw\frac{\partial f}{\partial x} = -\frac{\partial}{\partial x}(\langle \dot{x}\ddot{x} \rangle f) + Af, \quad (3.7)$$

где $A(x, t)$ есть среднее ускорение следующего порядка, определяемое через F_3 и F_4 :

$$\begin{aligned} A(x, t)f(x, t) &= \int a(x, \dot{x}, \ddot{x}, t)F_3(x, \dot{x}, \ddot{x}, t)d\dot{x}d\ddot{x}, \\ a(x, \dot{x}, \ddot{x}, t)F_3(x, \dot{x}, \ddot{x}, t) &= \int \ddot{x}F_4(x, \dot{x}, \ddot{x}, \ddot{x}, t)d\ddot{x}. \end{aligned} \quad (3.8)$$

Видно, что точная система уравнений не может быть получена в замкнутой форме, поскольку в ней возникают зацепления более высокого порядка. Кроме того, появились и моменты более высокого порядка (третьего по скорости в (3.6)).

Возникшая ситуация полностью аналогична классическому выводу уравнений гидродинамики из кинетических уравнений. Для ее разрешения следует замкнуть моментную систему, т.е. определить независимо правые части эволюционных уравнений (3.5), или (3.6), или (3.7). Это можно сделать различными способами, которые, однако, не могут быть основаны на точных математических следствиях, а должны исходить из более или менее правдоподобных практических предпосылок. Каждый из способов замыкания будет приводить к новой модели эволюции. В настоящей работе предложен один из способов, который можно использовать при прогнозировании ВПФР внутри промежутка ее ε -стационарности.

3.2. Модели временных рядов

Метод кинетических уравнений является основным инструментом для вывода моделей макроскопических процессов «из первых принципов», т.е. из микроскопического описания динамики отдельных частиц. Этот метод может быть распространен и на модели временных рядов.

Модели стационарных рядов представляют собой соотношения между некоторыми статистиками, которые (соотношения) перенесены на сами значения временного ряда. Например, регрессионные модели предполагают, что связь между переменными такая же, как и между их средними величинами, и т.п. В этой связи следует рассмотреть статистические модели рядов как следствия вероятностных гипотез относительно их выборочных распределений.

Обратимся снова к уравнению Лиувилля (2.11) для одномерной ВПФР. Из него прямым следствием получается уравнение эволюции среднего значения первого выборочного момента: $\dot{m} = U(t)$. Если прогнозом значения ряда считать прогноз среднего значения соответствующего распределения, то уравнение $\dot{m} = U(t)$, понимаемое в конечно-разностном смысле, при заданной априори средней скорости $U(t)$ представляет модель выделения заданного временного тренда:

$$x(t+1) = x(t) + U(t). \quad (3.9)$$

В другом случае, если ковариация $\text{cov}(x, \dot{x})$ достаточно мала, допустимо рассмотреть модель, в которой ряд и его приращения независимы. Тогда можно считать, что двумерная ВПФР факторизуется: $F(x, \dot{x}, t) = f(x, t)\varphi(\dot{x}, t)$. В этом случае ЭЛС не

зависит от x и совпадает со средней скоростью U : $u(x, t) = U(t)$. Следовательно, трендовый прогноз (3.9) получается, в частности, при принятии гипотезы статистической независимости координат и скоростей. В случае, если $U(t) = U_0 \neq 0$, из (3.9) следует линейная по времени регрессионная модель.

Если априорные соображения относятся к локальной скорости $u(x, t)$, получаются более сложные модели динамических систем. Эти модели возникают из статистического описания в тех случаях, если прогнозируется не сам ряд, а скорость его изменения. Тогда вместо (3.9) естественно предложить модель временного ряда, в которой приращения определяются средней локальной скоростью:

$$\dot{x} = u(x, t) \Rightarrow x_{n+1} - x_n = u(x_n; n). \quad (3.10)$$

Например, считая, что $u(x, t)$ не зависит от времени, получаем одномерную динамическую систему, порождающую стационарный временной ряд:

$$x_{n+1} = x_n + u(x_n) \equiv g(x_n). \quad (3.11)$$

В частном случае, когда $u(x)$ линейна по x , причем $U = 0$, из (3.11) получается авторегрессионная модель первого порядка.

Аналогично можно рассмотреть ВПФР, зависящую от большего числа производных, т.е. функцию вида $F(x, \dot{x}, \ddot{x}, \dots, t)$, но надо помнить об ограниченности объема выборки определенной величиной T . В силу конечности объема выборки невозможно эмпирически определить производную более высокого порядка, чем $T-1$. Более того, чтобы оперировать одинаковыми массивами данных по каждой из r производных, необходимо иметь исходный ряд в количестве по крайней мере $T+r$ элементов. Если это условие выполнено, то теми же соображениями можно получить многомерные аналоги вышеперечисленных прогнозных моделей.

Например, уравнение Лиувилля для многомерной ВПФР при обрыве цепочки на производной по времени r -го порядка имеет вид

$$\frac{\partial F_r(x, \dot{x}, \dots, x^{(r-1)}; t)}{\partial t} + \frac{\partial}{\partial x} \dot{x} F_r + \frac{\partial}{\partial \dot{x}} \ddot{x} F_r + \dots + \frac{\partial}{\partial x^{(r-1)}} w(x, \dot{x}, \dots, x^{(r-1)}; t) F_r = 0, \quad (3.12)$$

$$w(x, \dot{x}, \dots, x^{(r-1)}; t) F_r(x, \dot{x}, \dots, x^{(r-1)}; t) = \int x^{(r)} F_{r+1}(x, \dot{x}, \dots, x^r; t) dx^{(r)}.$$

Замыкание цепочки эволюционных уравнений на функции распределения r -ой размерности состоит в том, что среднее ускорение следующего порядка полагается равным нулю: $w(x, \dot{x}, \dots, x^{(r-1)}, x^{(r)}; t) = 0$. В этом случае уравнение эволюции $w(x, \dot{x}, \dots, x^{(r-1)}; t)$ будет содержать только моменты плотности распределения F_{r+1} :

$$\frac{\partial}{\partial t} w(x, \dot{x}, \dots, x^{(r-1)}; t) F_r(x, \dot{x}, \dots, x^{(r-1)}; t) = F_r \frac{\partial w}{\partial t} + w \frac{\partial F_r}{\partial t} = \int x^{(r)} \frac{\partial F_{r+1}(x, \dot{x}, \dots, x^r; t)}{\partial t} dx^{(r)} .$$

Используя (3.12) и записывая аналогичное ему уравнение эволюции для F_{r+1} с той лишь разницей, что в последнем следует положить $w(x, \dot{x}, \dots, x^{(r-1)}, x^{(r)}; t) = 0$, получаем

$$\begin{aligned} \frac{\partial}{\partial t} w(x, \dot{x}, \dots, x^{(r-1)}; t) F_r(x, \dot{x}, \dots, x^{(r-1)}; t) &= F_r \frac{\partial w}{\partial t} + w \frac{\partial F_r}{\partial t} = \int x^{(r)} \frac{\partial F_{r+1}(x, \dot{x}, \dots, x^r; t)}{\partial t} dx^{(r)} = \\ &= - \int x^{(r)} \left(\frac{\partial}{\partial x} \dot{x} F_{r+1} + \frac{\partial}{\partial \dot{x}} \ddot{x} F_{r+1} + \dots + \frac{\partial}{\partial x^{(r-1)}} x^{(r)} F_{r+1} \right) dx^{(r)}, \end{aligned}$$

откуда следует

$$\begin{aligned} &\left(\frac{\partial w}{\partial t} + \frac{\partial \langle \dot{x} x^{(r)} \rangle}{\partial x} + \frac{\partial \langle \ddot{x} x^{(r)} \rangle}{\partial \dot{x}} + \dots + w \frac{\partial w}{\partial x^{(r-1)}} \right) F_r = \\ &= \left(w \dot{x} - \langle \dot{x} x^{(r)} \rangle \right) \frac{\partial F_r}{\partial x} + \left(w \ddot{x} - \langle \ddot{x} x^{(r)} \rangle \right) \frac{\partial F_r}{\partial \dot{x}} + \dots + \\ &+ \left(w x^{(r-1)} - \langle x^{(r-1)} x^{(r)} \rangle \right) \frac{\partial F_r}{\partial x^{(r-2)}} + \frac{\partial}{\partial x^{(r-1)}} \left(w^2 - \langle (x^{(r)})^2 \rangle \right) F_r . \end{aligned} \quad (3.13)$$

Что касается модели временного ряда, основанной на уравнении (3.12), то соответствующей модельной функцией здесь является $w(x, \dot{x}, \dots, x^{(r-1)}; t)$. Именно, r -мерная прогнозная модель имеет вид

$$x^{(r)}(t) = w(x, \dot{x}, \dots, x^{(r-1)}; t). \quad (3.14)$$

Можно проинтегрировать (3.12) по выборочным аргументам, идущим не обязательно подряд, после чего полученное уравнение позволит связать между собой переменные, разделенные заданными лагами, что дает в итоге модели с распределенным запаздыванием. Если затем сделать предположения о статистической зависимости распределений – например, считать, что многомерные распределения переменных с лагами функционально зависят только от одной одномерной плотности распределения (по координатам или по скоростям и т.п.), то в результате получится некоторая самосогласованная модель. В класс самосогласованных моделей попадают, в частности, адаптивные модели, в которых прогноз значений ряда использует данные о точности прогноза в предыдущие моменты времени.

Перечисленные возможности применения кинетических уравнений показывают, что эти уравнения могут рассматриваться как инструмент для создания статистических моделей

временных рядов с заданными свойствами, а также для отбора моделей, наиболее близко подходящих для описания тех или иных реальных ситуаций.

Таким образом, метод кинетических уравнений позволяет получать не только существующие общеизвестные модели, но и некоторые другие, которые аналогичны гидродинамическим моделям сплошной среды в представлении о существовании локального равновесия.

Далее строится модель временного ряда на основе системы уравнений (2.11), (3.4) и (3.5). По аналогии с уравнениями гидродинамики в статистической механике будем называть эти модели «моделями гидродинамического типа».

4. МОДЕЛИ ГИДРОДИНАМИЧЕСКОГО ТИПА

4.1. Концепция моделирования временного ряда

Идея моделирования нестационарного временного ряда состоит в следующем. Сначала определяется оптимальный объем выборки для прогнозирования ВПФР на 1 шаг вперед с заданной точностью ε в соответствии с определением (2.2). Затем по имеющимся данным вычисляются эмпирические статистики: лиувиллева скорость, второй и третий моменты, фигурирующие в уравнениях (3.5)-(3.7). Если какие-либо из этих статистик оказалось возможным функционально смоделировать с требуемой точностью ε , то выбирается соответствующий способ замыкания моментных уравнений, после чего решается система этих уравнений, включая уравнение Лиувилля. Прогнозом временного ряда на следующем шаге по времени называется среднее значение построенного прогнозного распределения.

Продемонстрируем эту схему решения на примере эмпирического уравнения Лиувилля, записанного в конечных разностях. Из формулы (2.12) следует, что скорость $u(i, t)$ изменения ВПФР в i -ой ячейке в момент t может быть представлена в виде

$$u(i+1, t)f(i+1, t) = \sum_{k=1}^i (f(k, t+1) - f(k, t)), \quad (4.1)$$

т.е. выражается через изменение ВФР за один шаг по времени. Предположим, что нам удалось замкнуть моментную систему, описанную в разделе 3. Тогда эта система оказывается рекуррентно разрешима, так что после численного нахождения $u(i, t)$ распределение $f(i, t+1)$ определяется из (4.1):

$$\begin{aligned} f(1, t+1) &= f(1, t) + u(2, t)f(2, t); \\ f(2, t+1) &= f(2, t) + u(3, t)f(3, t) - u(2, t)f(2, t); \dots \end{aligned} \quad (4.2)$$

Прогнозным значением $x(t+1)$ будет считаться среднее значение по распределению $f(x, t+1)$:

$$\tilde{x}(t+1) = \int xf(x, t+1)dx = \sum_i if(i, t+1). \quad (4.3)$$

Динамическая система, порождающая этот временной ряд, приближенно находится из (4.1). Обозначая правую часть (4.1), деленную на $f(i+1, t)$, через $g(i, t)$, получаем

$$u(i, t) = g(i-1, t). \quad (4.4)$$

Если связать с формулой (4.4) динамическую систему, порождающую данный временной ряд, то скорость следует отождествить с изменением значения самого временного ряда $u(i, t) = x(t+1) - x(t)$. Если окажется, что правая часть в некотором приближении может считаться независимой явно от времени t , то получающаяся ДС будет автономна, т.е. приращение значения случайной величины $x(t)$ (номера ячейки i) будет определяться только значением самой этой величины. В нестационарном случае правая часть (4.4) помимо $x(t)$ зависит также явно от времени:

$$x(t+1) = x(t) + g(x(t), t). \quad (4.5)$$

Очевидно, для стационарного распределения лиувиллева скорость равна нулю, и прогноз дает одну и ту же величину, равную среднему значению по этому распределению.

Динамическая система первого порядка (4.5), полученная из уравнения Лиувилля для ВПФР, показывает, как могут быть построены более сложные многомерные системы, если точность одномерных моделей недостаточна. С одной стороны, можно использовать совместную плотность распределения значений ряда и его приращений, записать уравнение Лиувилля в виде (2.6) и попробовать аппроксимировать приращение совместной плотности по времени некоторой подходящей функцией. В результате получится формула для ускорений, аналогичная (4.5), связывающая между собой величины $x(t+2)$, $x(t+1)$ и некоторую функцию $g(x(t), t)$. Однако исследовать двумерную плотность технически значительно более сложная задача, чем одномерный анализ. Поэтому к построению динамической системы второго (и следующих) порядков удобнее подойти, используя уравнения эволюции для моментов исходной одномерной функции распределения.

Из уравнения (3.5) для изменения средней локальной скорости (т.е. для ускорения) можно получить замкнутую систему, полагая среднее ускорение $w(x, t)$ равным нулю. Тогда, исходя из разностного аналога уравнения (3.5) как правила рекуррентного определения эмпирической величины $e(x, t)$

$$\left(\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x}\right) f = \frac{\partial}{\partial x} (u^2 - e) f,$$

получаем следующую формулу:

$$e(i, t) = u^2(i, t) + \sum_{k=1}^{i-1} \left(\frac{f(k, t) u^2(k, t)}{f(k+1, t)} + \frac{f(k, t)(u(k, t) - u(k, t+1) - u(k, t)u(k+1, t))}{f(i, t)} \right). \quad (4.6)$$

После численного решения замкнутой системы моментов правая часть (4.6) становится известной. Тогда динамическая система, приближенно порождающая рассматриваемый временной ряд, строится следующим образом.

Считая $u(i, t)$ известной функцией, уравнение (3.5) можно рассматривать как уравнение относительно ускорения $\partial u / \partial t$ с известной правой частью. Обозначим эту правую часть через $W(i, t)$. Следуя правилу сопоставления динамической модели временного ряда и кинетического уравнения, получаем, что $\partial u / \partial t$ надо заменить на разностное ускорение, а аргумент i номера ячейки разбиения – на саму случайную величину $x(t)$ в этот момент времени. Тогда вместо (5.7) имеем

$$x(t+2) - 2x(t+1) + x(t) = \frac{W(x(t), t)}{f(x(t), t)}. \quad (4.7)$$

Аналогично можно получить и модель следующего порядка. Для этого надо еще раз продифференцировать по времени уравнение (3.5), а для закона эволюции второго момента $e(x, t)$ использовать уравнение (3.6), следующее формально из уравнения Лиувилля для одномерной ВПФР.

5.2. Методы замыкания моментной системы

Из предыдущего анализа уравнений (3.6) и (3.7) следует, что для замыкания уравнений относительно одномерной ВПФР надо, во-первых, задать среднее ускорение, т.е. интеграл от трехмерной ВПФР, и, во-вторых, расцепить систему уравнений для моментов двумерной ВПФР.

Если рассматривать промежутки времени, на которых ВПФР квазистационарна, то естественным условием замыкания является нулевое ускорение: $w(x, \dot{x}, t) = 0$. Напомним, что на таких промежутках времени можно использовать более сильное приближение нулевой скорости $u(x, t) = 0$, поскольку даже в этом случае гарантирована квазистационарность одномерного распределения. Уравнение Лиувилля с ненулевой скоростью является следующим приближением, уточняющим эволюцию ВПФР. Таким образом, приближение нулевого ускорения вполне согласовано с квазистационарностью распределения. Тогда для $e(x, t)$ из (3.6) получаем

$$f \frac{\partial e}{\partial t} - e \frac{\partial(uf)}{\partial x} = - \frac{\partial(e_3 f)}{\partial x}, \quad e_3 \equiv \langle \dot{x}^3 \rangle.$$

Переносим слагаемое $e \frac{\partial(uf)}{\partial x}$ в правую часть, преобразуем это уравнение к виду

$$\left(\frac{\partial e}{\partial t} + u \frac{\partial e}{\partial x} \right) f = \frac{\partial}{\partial x} (ue - e_3) f.$$

Объединяя полученное уравнение с (2.11) и (3.5), получаем следующую систему уравнений:

$$\begin{aligned} \frac{\partial f(x,t)}{\partial t} + \frac{\partial}{\partial x} f(x,t)u(x,t) &= 0; \\ \left(\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} \right) f &= \frac{\partial}{\partial x} (u^2 - e) f; \\ \left(\frac{\partial e}{\partial t} + u \frac{\partial e}{\partial x} \right) f &= \frac{\partial}{\partial x} (ue - e_3) f. \end{aligned} \quad (4.8)$$

В правых частях уравнений для моментов в (4.8) стоят производные от ковариаций степеней скорости, так что общее уравнение такого типа имеет вид

$$\left(\frac{\partial e_k}{\partial t} + e_1 \frac{\partial e_k}{\partial x} \right) f = \frac{\partial}{\partial x} (e_1 e_k - e_{k+1}) f, \quad e_k \equiv \langle \dot{x}^k \rangle. \quad (4.9)$$

Замыкание моментной цепочки (4.9) может быть проведено разными способами, позволяющими на определенном k -ом шаге выразить старший момент через моменты низших порядков. Ниже приводятся некоторые варианты такого замыкания как примеры прогнозных моделей «гидродинамического» типа, т.е. основанных на анализе эволюции моментов выборочного распределения большего порядка, чем то, которое непосредственно прогнозируется.

1). Положим $e_{k+1} = 0$, т.е. будем считать эту величину пренебрежимо малой. Тогда система замкнется, но, решив ее, следует убедиться в том, что найденное решение совместимо с этим условием малости. Замыкающее уравнение при этом имеет вид

$$f \frac{\partial e_k}{\partial t} = e_k \frac{\partial(uf)}{\partial x}. \quad (4.10)$$

Такой подход справедлив в том случае, если последовательность $e_k(x,t)$ убывает и стремится к нулю с увеличением k , так что для любой заданной точности ε найдется такой номер k , начиная с которого всеми $e_k(x,t)$ можно пренебречь. Это – случай слабо меняющегося ряда $x(t)$.

2). Будем считать пренебрежимо малой величину ковариации, т.е. положим $e_1 e_k - e_{k+1} = 0$. Эта ситуация возникает, например, тогда, когда скорость $u(x,t)$ как

функция от x приблизительно постоянна, за исключением, возможно, точек вблизи границ промежутка изменения величины x . В таком случае можно считать, что $\partial u / \partial x$ с принятой точностью является пренебрежимо малой величиной, так что замыкающее уравнение имеет вид первого уравнения системы – уравнения Лиувилля:

$$\frac{\partial e_k}{\partial t} + \frac{\partial(ue_k)}{\partial x} = 0. \quad (4.11)$$

В частности, уравнение (4.11) выполняется в этом случае и для $k = 2$.

3). Заметим, что, как и скорость $u(x, t)$, величины $e_k(x, t)$ известны по исходной выборке не в текущий, а в предыдущий момент времени. Подход к замыканию в духе наивного прогноза состоит в том, чтобы оценить величины $e_k(x, t)$ по их значениям в момент $t - 1$, которые известны по имеющейся статистике к этому времени. Положим тогда $e_k(x, t) = e_k(x, t - 1) \equiv g_k(x, t)$, где $g_k(x, t)$ – в данном случае известная функция, определяемая непосредственно по значениям временного ряда. Тогда замыкающее уравнение представляет собой уравнение (4.11) с неоднородностью:

$$f \frac{\partial e_k}{\partial t} = e_k \frac{\partial(uf)}{\partial x} - \frac{\partial(g_k f)}{\partial x}. \quad (4.12)$$

Итак, каждый из этих трех способов, будучи применен к системе (4.9), либо к первым двум ее уравнениям или только к первому, приведет к замкнутой модели эволюции одномерной ВПФР.

Например, если применить второй метод (нулевой ковариации) ко второму уравнению (4.8), то он в данном случае будет означать, что средний квадрат скорости и квадрат средней скорости совпадают. Следовательно, средняя скорость $u(x, t)$ как функция от x равна постоянной. Тогда достаточно использовать только первое уравнение из (4.8), решение которого будет иметь вид $f(x, t) = f(x - u, t - 1)$. Поскольку в граничных промежутках, по условию, плотность всегда нулевая, то решение в виде сдвига вдоль характеристик возможно, только если скорость нулевая. Тогда получается, что $f(x, t)$ не меняется во времени, т.е. мы получили наивный прогноз относительно плотности распределения.

Если же применить ко второму уравнению в (4.8) метод наивного прогноза, касающийся величины $e(x, t)$, то система замыкается следующим образом: правая часть этого уравнения считается известной по данным в предыдущий момент времени, после чего определяется $u(x, t)$ в текущий момент, что позволяет определить $f(x, t)$ из первого уравнения.

Отметим, что статистические гипотезы можно выдвигать не только относительно моментов распределений, содержащихся в гидродинамической модели (4.8), но и относительно распределений более высокого порядка.

4.3. Уточнение способов замыкания

В прогнозных моделях гидродинамического типа подразумевается, что эволюция ВПФР данной размерности определяется эволюцией нескольких первых моментов ВПФР расширенной размерности. Примеры таких моделей были приведены в 5.2 для одномерной ВПФР, эволюция которой определялась замкнутым образом. Исследуем более общую ситуацию. С этой целью вернемся к цепочке (3.12), в которой положим $r = 3$. Тогда получаем следующую систему кинетических уравнений:

$$\begin{aligned} \frac{\partial F(x, \dot{x}; t)}{\partial t} + \frac{\partial}{\partial x} \dot{x} F(x, \dot{x}; t) + \frac{\partial}{\partial \dot{x}} w(x, \dot{x}; t) F(x, \dot{x}; t) &= 0, \quad F \equiv F_2, \\ \frac{\partial f(x, t)}{\partial t} + \frac{\partial}{\partial x} u(x, t) f(x, t) &= 0. \end{aligned} \quad (4.13)$$

В цепочке (4.13) функция $u(x, t)$ определяется, как и раньше, через $F(x, \dot{x}; t)$ согласно (2.14), а ускорение $w(x, \dot{x}; t)$ определяется через выборочную плотность более высокого порядка согласно (3.3):

$$w(x, \dot{x}; t) F(x, \dot{x}; t) = \int \ddot{x} F_3(x, \dot{x}, \ddot{x}; t) d\ddot{x}.$$

Если замыкание цепочки проводится на функции F_3 , то, согласно этому предположению, эта функция должна удовлетворять уравнению Лиувилля без зацепления с F_4 , т.е.

$$\frac{\partial F_3(x, \dot{x}, \ddot{x}; t)}{\partial t} + \frac{\partial}{\partial x} \dot{x} F_3 + \frac{\partial}{\partial \dot{x}} \ddot{x} F_3 = 0. \quad (4.14)$$

В таком приближении замыкание уравнения (4.13) состоит в выборе модели для эволюции ускорения $w(x, \dot{x}; t)$. Дифференцируя (3.3) и используя эволюционные уравнения Лиувилля (4.13) и (4.14) для F и F_3 , получаем для $w(x, \dot{x}; t)$ следующее уравнение:

$$F \frac{\partial w}{\partial t} = \left(\frac{\partial(\dot{x}F)}{\partial x} + \frac{\partial(wF)}{\partial \dot{x}} \right) w - \int \ddot{x} \left(\frac{\partial(\dot{x}F_3)}{\partial x} + \frac{\partial(\ddot{x}F_3)}{\partial \dot{x}} \right) d\ddot{x}.$$

В дополнение к приближению (4.14) будем также считать результат усреднения величин $\dot{x} \frac{\partial \ddot{x}}{\partial x}$ и $\ddot{x} \frac{\partial \ddot{x}}{\partial \dot{x}}$ пренебрежимо малым. Тогда получаем уравнение

$$\left(\frac{\partial w}{\partial t} + \dot{x} \frac{\partial w}{\partial x} + w \frac{\partial w}{\partial \dot{x}} \right) F = \frac{\partial}{\partial x} (\dot{x} w - \pi_{11}) F + \frac{\partial}{\partial \dot{x}} (w^2 - \pi_{02}) F, \quad (4.15)$$

где введены обозначения $\pi_{kl}(x, \dot{x}, t)$ для средних величин, определяемых формулой

$$\pi_{kl}(x, \dot{x}, t)F(x, \dot{x}, t) = \int \dot{x}^k \ddot{x}^l F_3(x, \dot{x}, \ddot{x}; t) d\ddot{x}. \quad (4.16)$$

Введем далее величины Π_{kl} так, что

$$\Pi_{kl}(x, t)f(x, t) = \iint \dot{x}^k \ddot{x}^l F_3(x, \dot{x}, \ddot{x}) d\dot{x} d\ddot{x} = \int \pi_{kl}(x, \dot{x}, t)F(x, \dot{x}, t) d\dot{x}. \quad (4.17)$$

Тогда для эволюции одномерной ВПФР $f(x, t)$ получаем цепочку уравнений

$$\begin{aligned} \frac{\partial f}{\partial t} &= -\frac{\partial(uf)}{\partial x}; \\ \left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x}\right)f &= \frac{\partial}{\partial x}(u^2 - e)f + \Pi_{01}f; \\ \left(\frac{\partial e}{\partial t} + u\frac{\partial e}{\partial x}\right)f &= \frac{\partial}{\partial x}(ue - e_3)f + 2\Pi_{11}f; \dots \end{aligned} \quad (4.18)$$

Система (4.18) уточняет систему (4.8), полученную в предположении обрыва цепочки на двумерной ВПФР. Замыкание моментной системы и будет являться прогнозной моделью для эволюции одномерного выборочного распределения.

Прямым следствием замкнутой системы эволюционных уравнений относительно одномерной ВПФР и моментов двумерной ВПФР является то, что в такой модели зависимость выборочного распределения от времени входит только через несколько первых моментов. Тогда эти моменты с принятой точностью определяют и само распределение, что составляет содержание одного из положений функциональной гипотезы Боголюбова [1, 2] относительно локально-равновесных распределений в классической статистической механике.

Например, если оборвать (4.18) на третьем уравнении, то такой обрыв будет означать, что принята следующая функциональная гипотеза о временной зависимости высших ВПФР: зависимость от времени в $F(x, \dot{x}; t)$, $F_3(x, \dot{x}, \ddot{x}; t)$ и т.д. входит только через $f(x, t)$, $u(x, t)$ и $e(x, t)$. Тогда задача моделирования будет состоять в том, чтобы выдвинуть дополнительные предположения, с помощью которых можно будет выразить e_3 , Π_{01} , Π_{11} через f , u , e , что позволит замкнуть систему (4.18). Продемонстрируем, как может выглядеть один из вариантов такого обрыва.

Пусть в выражении для e_3 можно пренебречь величиной $\langle (\dot{x} - u(x, t))^3 \rangle$. Тогда считаем, что $\int (\dot{x} - u)^3 F d\dot{x} \approx 0$, после чего получаем

$$e_3 f = \int ((\dot{x} - u) + u)^3 F d\dot{x} = 3uef - 2u^3 f, \text{ откуда следует } e_3 = 3ue - 2u^3.$$

Предположим далее, что $\langle w \rangle \equiv \pi_{01} \approx \partial u / \partial t$, и, кроме того, приближенно выполняется условие факторизации $\Pi_{11} \approx u \Pi_{01}$. Тогда сделанные три допущения замыкают систему (4.18), которая в этом случае переписывается в виде:

$$\begin{aligned} \frac{\partial f}{\partial t} &= -\frac{\partial(uf)}{\partial x}; & uf \frac{\partial u}{\partial x} &= \frac{\partial}{\partial x}(u^2 - e)f; \\ f \frac{\partial(e - u^2)}{\partial t} &= e \frac{\partial(uf)}{\partial x} - \frac{\partial}{\partial x}(3ue - 2u^3)f. \end{aligned} \quad (4.19)$$

Аналогично можно строить модели в пространствах большей размерности.

Модель прогнозирования ВПФР может использовать гипотезы не только на уровне макроскопического описания через средние величины, но и на уровне предполагаемых связей между значениями временного ряда. В таком подходе уравнение эволюции ВПФР применяется для прогнозирования изменения выборочных моментов.

Например, пусть ускорение предполагается зависящим от самой координаты, так что $\ddot{x} = \varphi(x, t)$. В частности, пусть $\varphi(x, t)$ аппроксимируется линейной регрессией по x , т.е. $\varphi(x, t) = ax(t) + b$. Согласно формулам регрессионного анализа [3], коэффициент a представляет собой скользящее отношение ковариации между координатой и ускорением к дисперсии координаты: $a = \text{cov}(x, \ddot{x}) / \sigma^2(x)$. Тогда из уравнения Лиувилля (2.11) следует уравнение эволюции для средней скорости: $\partial U / \partial t = a\mu(t) + b$, где $\mu(t)$ есть первый момент ВПФР $f(x, t)$. Зададим начальные значения первого момента и скорости μ_0, U_0 в момент времени t_0 . Тогда получаем прогнозную формулу для первого момента:

$$\mu(t_0 + t) = -\frac{b}{a} + \left(\mu_0 + \frac{b}{a} \right) \cos(\sqrt{a}t) + \frac{U_0}{a} \sin(\sqrt{a}t), \quad t = 1, \dots, \tau. \quad (4.20)$$

Эта формула затем может быть использована и для прогнозирования самого временного ряда, если считать прогнозным значением величину первого момента.

5. ПРИМЕР МОДЕЛИРОВАНИЯ ВРЕМЕННОГО РЯДА

5.1. Временной ряд курса доллара к рублю

Рассмотрим применение замыкания модели (4.8) на первых трех уравнениях на примере временного ряда, образованного суточными данными $c(t)$ курса ЦБ рубль/доллар за период с 12.01.2005 по 26.08.2010 [6].

Коэффициент корреляции между элементами ряда $c(t)$ с лагом 1 равен 0,997. Это означает, что переход к первым разностям приведет к заметному очищению ряда от взаимозависимости между членами. Обозначим

$$M = \max_t (c(t+1) - c(t)), \quad m = \min_t (c(t+1) - c(t)) \quad (5.1)$$

и рассмотрим нормированные приросты курса

$$x(t) = \frac{c(t+1) - c(t) - m}{M - m}. \quad (5.2)$$

Временной ряд $x(t)$ представлен на рис. 1. Автокорреляция с лагом 1 для этого ряда равна -0,003 (близка к нулю), так что соответствующие величины можно считать статистически независимыми.

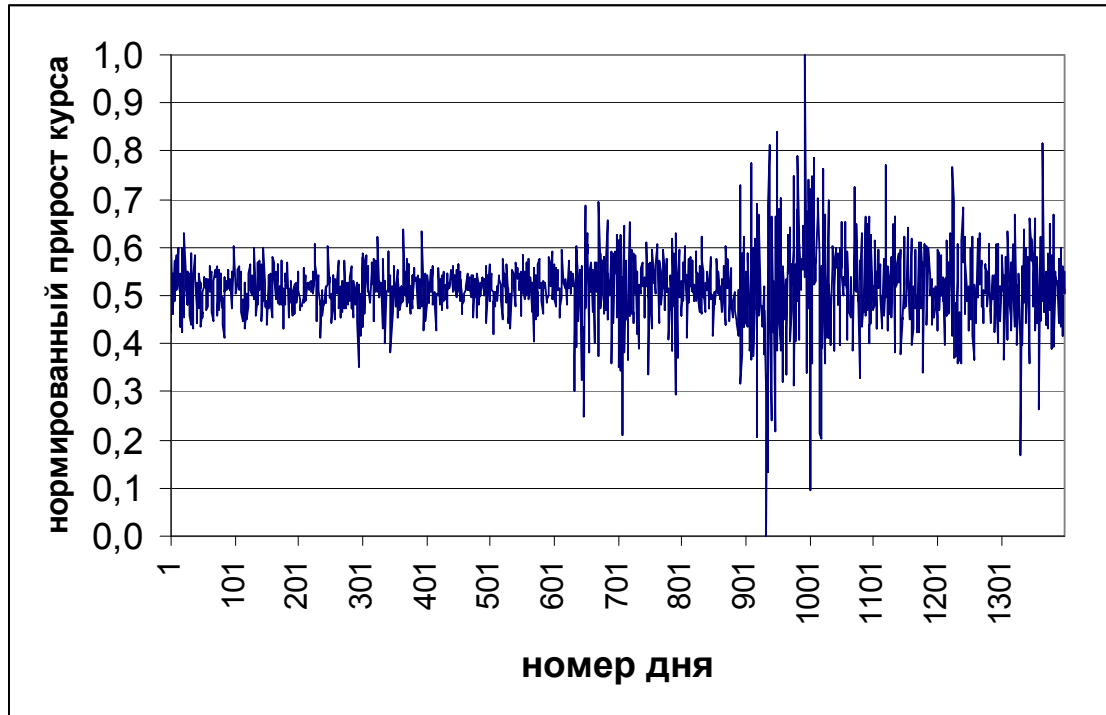


Рис. 1. Временной ряд $x(t)$ нормированных приростов курса ЦБ рубль/доллар

Тем не менее, ряд $x(t)$ не является стационарным, что видно из типичного поведения его выборочной дисперсии (в некоторый момент времени) в зависимости от объема выборки (рис. 2). Дисперсия с увеличением объема выборки не стабилизируется, как это должно быть согласно центральной предельной теореме, а колеблется на возрастающем тренде.

Среднеквадратичное относительное отклонение суточных приростов курса составило 0,008, а среднеквадратичное отклонение нормированных величин $x(t)$ равно примерно 0,05. Чтобы с такой же достаточной точностью $\varepsilon = 0,05$ спрогнозировать изменение ВПФР ряда $x(t)$ на 1 шаг вперед, минимально достаточный объем выборки, как показано в [5], должен быть равен $N_{\min} = 2/\varepsilon = 40$. Однако минимум выборочной дисперсии (рис. 2) находится значительно правее этого значения (в точке 290), поэтому

объем выборки $N_{opt} = 290$ является оптимальным для прогнозирования ВПФР с точностью $\varepsilon = 0,05$. Строго говоря, в каждый момент времени реализуется свое значение N_{opt} , но его относительная дисперсия в данном примере на порядок меньше ε^2 , так что N_{opt} можно считать постоянным на достаточно большом промежутке прогнозирования.

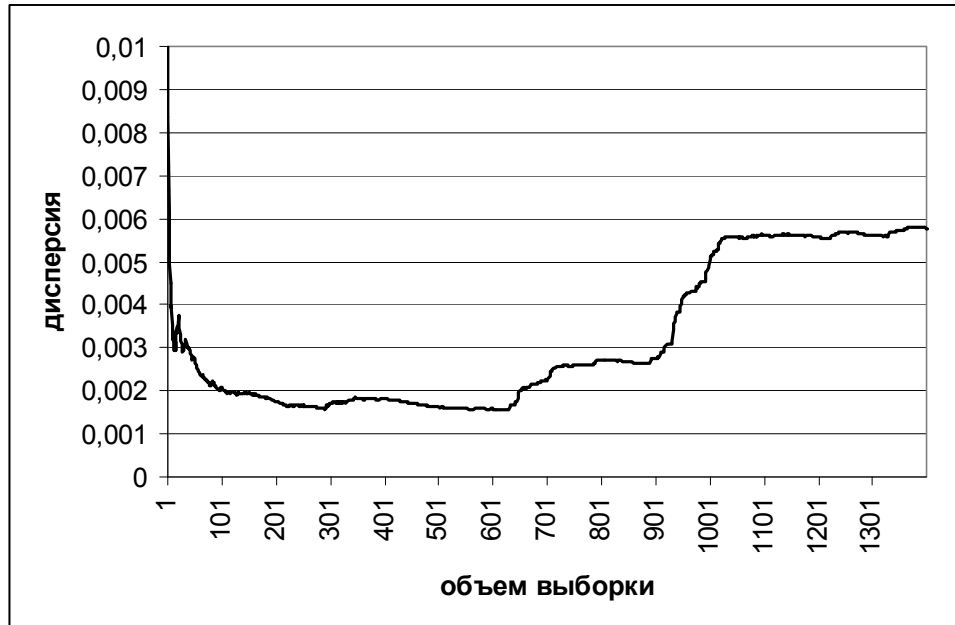


Рис. 2. Дисперсия ряда $x(t)$ как функция объема выборки

5.2. Гидродинамическая модель эволюции ВПФР

В каждый момент времени t построим по выборке объема N_{opt} выборочную плотность функции распределения $f(x, t)$ и запишем для нее систему уравнений (4.8):

$$\begin{aligned} \frac{\partial f(x, t)}{\partial t} + \frac{\partial}{\partial x} f(x, t) u(x, t) &= 0; \\ \left(\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} \right) f &= \frac{\partial}{\partial x} (u^2 - e) f; \\ f \frac{\partial e}{\partial t} &= e \frac{\partial (uf)}{\partial x} - \frac{\partial (g_3 f)}{\partial x}. \end{aligned} \quad (5.3)$$

Замкнем систему (5.3) с помощью гипотезы наивного прогноза относительно момента третьего порядка $e_3(x, t)$. Тогда при дискретизации модели по времени следует положить $g_3(x, t) = e_3(x, t-1)$. Тогда получаем дискретную прогнозную модель в виде (тильдой обозначены прогнозируемые величины):

$$\begin{aligned}
\mathbf{f}: \quad & \tilde{f}(x, t+1) = f(x, t) + f(x, t)\tilde{u}(x, t) - f(x+1, t)\tilde{u}(x+1, t); \\
\mathbf{u}: \quad & (\tilde{u}(x, t) - u(x, t-1))f(x, t-1) = \\
& = u(x, t-1)(u(x+1, t-1)f(x+1, t-1) - u(x, t-1)f(x, t-1)) + \\
& + \tilde{e}(x, t)f(x, t-1) - \tilde{e}(x+1, t)f(x+1, t-1); \\
\mathbf{e}: \quad & (\tilde{e}(x, t) - e(x, t-1))f(x, t-1) = \\
& = e(x, t-1)(u(x+1, t-1)f(x+1, t-1) - u(x, t-1)f(x, t-1)) + \\
& + e_3(x, t-1)f(x, t-1) - e_3(x+1, t-1)f(x+1, t-1).
\end{aligned} \tag{5.4}$$

Видно, что система (5.4) рекуррентно разрешима, если известны $f(x, t)$ и $f(x, t-1)$. Фактически прогноз строится относительно скорости $\tilde{u}(x, t)$ и «энергии» $\tilde{e}(x, t)$, после чего вычисляется $\tilde{f}(x, t+1)$.

Точность прогнозирования ВПФР оказалась равной 0,033, т.е. почти в 2 раза лучше, чем среднее квадратичное отклонение нормированных величин.

Сравнение прогноза и факта курса рубль/доллар приведено на рис. 3.

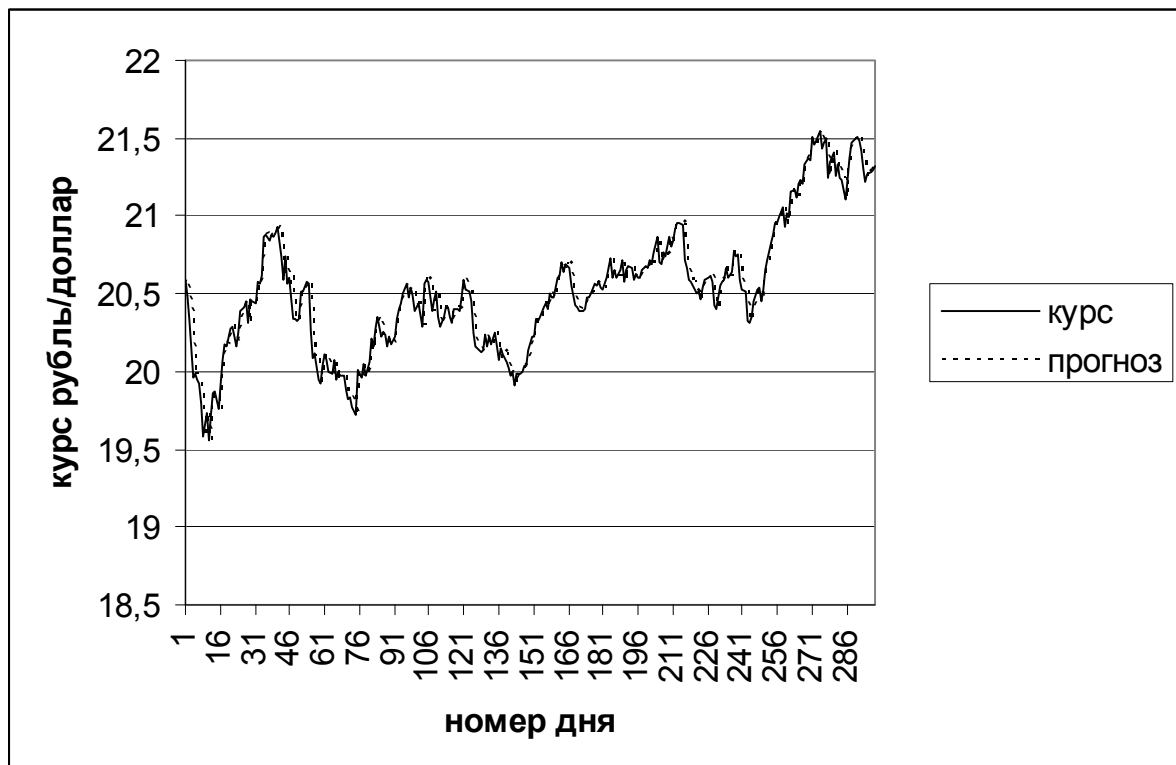


Рис. 3. Сравнение прогноза и факта курса рубль/доллар

Относительное среднее квадратичное отклонение прогноза от факта составило 0,0045, что, как и для прогноза нормированных приростов, примерно в 2 раза меньше, чем фактическое. Это означает, что прогноз по построенной модели в 2 раза точнее наивного прогноза по самому курсу, который состоит в том, что «завтра курс будет такой же, как и

сегодня». Здесь следует сказать, что стандартные прогнозы приростов курса по стационарным регрессионным и авторегрессионным моделям любого порядка на оптимальных объемах выборки дают ошибку, приблизительно совпадающую (до третьего знака) с ошибкой наивного прогноза, т.е. они не являются эффективными.

Прогноз курса по модели (5.4) можно представить как динамическую систему вида (4.7): этот прогноз будет представлять собой среднее значение нестационарной функции распределения, численно определяемой из (5.4). Соответствующая динамическая система показана пунктиром на рис. 3. Отметим, что прогноз по модели ДС второго порядка существенно лучше, чем по модели первого порядка, если ограничиться только моделированием самой эмпирической скорости, поскольку прогноз по ДС первого порядка имеет ошибку на уровне наивного прогноза. При этом следует указать, что стандартные адаптивные регрессионные и автокорреляционные методы прогнозирования, описанные, например, в [3], дают ошибку, в 1,5-2 раза превосходящую ошибку наивного прогноза.

Описанные в работе модели не исчерпывают все способы замыкания системы прогнозных уравнений, получаемых в рамках кинетического подхода. Эти методы носят эвристический характер и не являются универсальными для любых временных рядов. Тем не менее, они показывают, что выбор адекватной модели может быть сделан и для нестационарных процессов без предположений об их функциональной принадлежности.

В дальнейшем предполагается классифицировать временные ряды по типам нестационарности и выбрать наиболее подходящую модель для их прогнозирования.

ИСТОЧНИКИ

1. Боголюбов Н.Н. Проблемы динамической теории в статистической физике. – М.-Л.: ГИТТЛ, 1946.
2. Боголюбов Н.Н. Уравнения гидродинамики в статистической механике. – Сб. трудов Института математики АН УССР, 1948. Т.10. № 41.
3. Кремер Н.Ш., Путко Б.А. Эконометрика. – М.: ЮНИТИ-ДАНА, 2002. – 311 с.
4. Орлов Ю.Н., Осминин К.П. Построение выборочной функции распределения для прогнозирования нестационарного временного ряда. // Мат. Мод., 2008. № 9. С. 23-33.
5. Орлов Ю.Н., Осминин К.П. Нестационарные временные ряды: методы прогнозирования с примерами анализа финансовых и сырьевых рынков. – М.: Книжный дом «ЛИБРОКОМ», 2011. – 384 с.
6. Суточные курсы валют. / www.finam.ru