

УДК 519.242.33

На правах рукописи

Осминин Константин Павлович

**АЛГОРИТМЫ ПРОГНОЗИРОВАНИЯ  
НЕСТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ**

Автореферат диссертации на соискание ученой степени  
кандидата физико-математических наук

Специальность

05.13.18 – математическое моделирование, численные методы  
и комплексы программ

Москва – 2008

Работа выполнена на кафедре высшей геометрии и топологии механико-математического факультета Московского государственного университета им. М.В. Ломоносова

Научный руководитель: доктор физико-математических наук  
доцент Юрий Николаевич Орлов

Официальные оппоненты: доктор физико-математических наук  
Ирина Федоровна Потапенко

кандидат физико-математических наук  
Михаил Леонидович Нечаев

Ведущая организация: Международный институт теории прогноза землетрясений и математической геофизики РАН

Защита состоится " 23 " октября 2008 г. в 15 час. на заседании Диссертационного совета Д 002.058.01 при Институте математического моделирования РАН по адресу: 125047, Москва, Миусская пл., д4.

С диссертацией можно ознакомиться в библиотеке Института математического моделирования РАН.

Автореферат разослан "      "                                      2008 г.

Ученый секретарь диссертационного совета

Н.В. Змитренко

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

### Актуальность темы

Роль статистических методов анализа в настоящее время неуклонно растет. Область применения и разнообразие методов, используемых в задачах прогнозирования временного ряда, велико. Но при этом большинство применяемых методов относятся к анализу стационарных временных рядов и используют следующие основополагающие утверждения (см., напр., Уилкс С. Математическая статистика. – М.: Наука, 1967. – 632 с.): критерий согласия Колмогорова, теорема Вальда, теорема Гофдинга.

Критерий согласия Колмогорова (1933) является одним из центральных результатов математической статистики. Им определяется близость интегральной выборочной функции распределения случайной величины  $\xi$  к стационарному распределению, если оно есть. Именно, супремум модуля разности  $D_n = \sup_x |F_n(x) - F(x)|$  выборочной и точной интегральных функций

распределения стационарной случайной величины  $\xi$  по вероятности стремится к нулю с ростом объема выборки  $n$  так, что случайная величина  $\sqrt{n}D_n$  имеет асимптотическое распределение в виде  $K$ -функции Колмогорова. В дальнейшем на основе этого утверждения были получены различные широко применяемые асимптотические критерии о принадлежности двух выборочных распределений одной генеральной совокупности: Колмогорова-Смирнова (1939), Вальда-Волфовица (1940), Вилкоксона (1945), Манна-Уитни (1947), Гнеденко-Королюка (1951) и другие, более узкие по применимости критерии (Стьюдента, Фишера, Крамера-Уэлча, «омега-квадрат» и др.).

Другим фундаментальным утверждением является теорема Вальда (1938) о разложении, согласно которой любой стационарный случайный процесс представляется в виде суперпозиции детерминированного процесса и белого шума.

Еще одним методологически важным результатом является теорема Гофдинга (1948) о том, что умноженные на  $\sqrt{n}$  отклонения моментов эмпирического распределения, построенного по выборке объема  $n$ , от моментов генеральной совокупности для стационарной случайной величины распределены асимптотически нормально. Эта теорема позволяет определить скорость сходимости по вероятности выборочных моментов и вероятность отклонения их значений от теоретических, если таковые известны.

Перечисленные утверждения математической статистики определяют основные принципы моделирования стационарных временных рядов. Обычно ряд представляется в виде суммы некоторой детерминированной составляющей и остатка, причем желательно, чтобы автокорреляционная функция остатка с достаточной точностью была близка к нулю, что свидетельствует о близости остатка к белому шуму. После этого параметрическими или иными методами находят наиболее близкую статистику, моделирующую поведение остатка. Существуют различные

модификации такого подхода. В настоящее время основными статистическими методами исследования временных рядов являются: метод выделения тренда (временного сглаживания), регрессионный, автокорреляционный, адаптивный (скользящих средних), метод гармонического анализа, сингулярного спектрального анализа, бутстрепа (численного размножения выборок) и нейросетевой.

Подчеркнем, что вышеперечисленные методы корректно обоснованы только для стационарных рядов. Однако на практике они применяются ко всем рядам, которые возникают в случайном эксперименте. Тогда перед исследователем встает проблема оценки точности получаемых им результатов. Если в стационарном случае есть доказательная уверенность в асимптотической состоятельности оценок той или иной статистики, то в нестационарном случае отсутствует само понятие генеральной совокупности, что делает неприменимым весь развитый аппарат современной математической статистики, кроме тех случаев, когда априори задана функциональная принадлежность модели процесса. Однако на практике часто бывает не известно, к какому классу принадлежит распределение и является ли оно стационарным, причем оба этих фактора могут быть определены лишь с некоторой доверительной вероятностью – корректно определенной, однако, только для стационарных процессов.

Если ряд нестационарный, то статистики, используемые, например, в моделях скользящих средних, не являются состоятельными оценками моментов распределения, т.к. сходимости по вероятности в общем случае нет.

Кроме того, в адаптивных методах исследования рядов, про которые априори не известно, являются ли они (ряды) стационарными или нет, не решен вопрос, по выборке какого объема следует проводить скользящее усреднение, чтобы получить наименьшую ошибку прогноза. Решение этой проблемы в существующих критериях оставляется на усмотрение пользователя в соответствии с его жизненным опытом.

Одной из основных составляющих методик прогнозирования временного ряда (в том числе и стационарного) является оценка сверху среднеквадратичной ошибки, которую можно допустить, если применить к нему тот или иной метод анализа. Имеющиеся методы, предполагающие стационарность ряда, не учитывают вклад изменения характеристик ряда в ошибку прогноза. Таким образом, существующие оценки методов для стационарных рядов следует дополнить оценками временных границ их применимости. Поскольку же временные ряды, возникающие в реальных ситуациях, как правило, нестационарны, и анализ их имеет практическую важность, как в примерах с ценовыми рядами, то разработка методики и численного алгоритма, увеличивающего точность прогнозирования таких рядов, является актуальной задачей.

## **Цель и задачи исследования**

Диссертация посвящена разработке моделей эволюции нестационарных временных рядов, анализ которых требуется в практической деятельности. К ним относятся, например, ряд почасовых цен на российском оптовом рынке электроэнергии и мощности, биржевые котировки акций компаний, таких, как General Motors, General Electric, Microsoft, Ford и др.

Временные ряды, встречающиеся в экономических, геологических, медицинских и т.п. задачах, имеют, как правило, единственную реализацию, поэтому статистический ансамбль отсутствует. Следовательно, невозможно оценить эмпирическую вероятность значений случайной величины с любой наперед заданной точностью. В качестве оценки точности такой вероятности в данной работе использовался интегральный критерий близости выборочных плотностей функций распределения (далее просто функций распределения).

Основной задачей, решаемой в диссертации, является задача снижения ошибки прогноза временного ряда на заданном временном горизонте при условии, что ошибка прогноза выборочной функции распределения этого ряда не превосходит заданной величины. Ошибка прогноза ряда понимается в смысле среднего квадратичного. Параметром, по которому минимизируется эта ошибка, является текущий объем выборки элементов ряда.

Оптимизация объема выборки возможна в силу того, что ошибка прогноза складывается из ошибок двух типов. Ошибка первого типа – это погрешность оценки статистических свойств временного ряда за счет конечности объема выборки, т.е. за счет недостаточной репрезентативности. Ошибка второго типа – это погрешность за счет нестационарности статистики. Ошибка первого типа с ростом объема выборки уменьшается, а ошибка второго типа возрастает.

Для определения оптимального объема выборки при условии фиксированного максимально допустимого отклонения между прогнозной и фактической выборочными функциями распределения, а также для создания прогнозной модели нестационарного временного ряда в диссертации решаются следующие задачи.

Первой задачей является создание системы необходимых понятий для корректного прогнозирования нестационарных временных рядов, когда отсутствует само понятие генеральной совокупности.

Второй задачей является разработка подходящего теоретического инструмента (аналога статистики Колмогорова-Смирнова) для определения выборки оптимального объема с точки зрения общей цели исследования, сформулированной выше. Таким инструментом стала статистика «горизонтного» ряда (терминология автора), степень нестационарности которого существенно меньше, чем исходного анализируемого ряда.

Третья задача связана с созданием математической прогнозной модели для выборочной функции распределения, а также для самого ряда. Рассматривая задачу прогнозирования как задачу о построении оператора эволюции функции распределения, можно сконструировать соответствующее

«эмпирическое» эволюционное уравнение. Это может быть либо уравнение Лиувилля, для которого строится некоторый кинетический аналог скорости изменения плотности вероятности, либо уравнение Фоккера-Планка, корректирующее уравнение Лиувилля с помощью модели случайного блуждания.

Четвертая задача – создание численного алгоритма, реализующего предложенную методику, и проведение тестовых расчетов.

Таким образом, создание моделей включает в себя следующие шаги: разработка терминов и понятий для работы с нестационарными временными рядами, методик прогнозирования нестационарных временных рядов и их выборочных функций распределения, математическое обоснование этих методик, описание прогнозной модели, построение соответствующих численных алгоритмов, их реализация в программном коде и сравнение точности предложенных моделей с точностью существующих.

### **Научная новизна**

По теории случайных процессов существует огромное количество литературы, как учебно-методической, так и узко-специальной. В основном они посвящены исследованию стационарных процессов. Нестационарные процессы, изучаемые в относительно малом числе публикаций, чаще всего относятся к определенным функциональным классам, проверка принадлежности к которым реальных процессов является гораздо более трудной задачей, чем проверка их на стационарность.

Таким образом, изучение нестационарных процессов как таковых является сравнительно новой областью исследований. Диссертантом была построена система определений и доказаны свойства новых введенных им статистик, позволяющих корректно использовать оценки, получаемые на нестационарных выборках, в практической деятельности.

Новой является также прогнозная модель выборочной функции распределения и собственно временного ряда, использующая эмпирические уравнения Лиувилля и Фоккера-Планка. Эти модели позволяют локально сопоставить текущей выборке некоторую квазидинамическую систему, эволюционирующую в условиях воздействия случайной силы.

Впервые получена конструктивная связь между объемом выборки, при котором ошибка прогноза не превышает наперед заданную величину, горизонтом прогноза и точностью прогнозирования функции распределения, понимаемой как расстояние в пространстве суммируемых функций.

Новизна результатов диссертации состоит не только в получении новых теоретических утверждений о свойствах некоторых статистик, но и в разработке численного алгоритма для их построения, поскольку такой алгоритм отсутствует во всех широко используемых программных статистических пакетах. В частности, такие универсальные пакеты как STATISTICA, SPSS, Minitab, Eviews, SAS, SYSTAT, Statgraphics и др. (см. Айвазян С.А., Степанов В.С. Инструменты статистического анализа данных. //

Мир ПК, 1997, № 8, с. 32-41.; Статистические и математические системы – «Тысячи программных продуктов»: Каталог. М., 1995, №2, с.88–92) не содержат инструментов для решения задачи о получении т.н. «горизонтной» статистики ряда. Также в существующем программном обеспечении нет алгоритма для прогноза функции распределения на основе эмпирического кинетического уравнения.

### **Научная и практическая значимость**

Результаты настоящей диссертации имеют научную, методическую и практическую значимость.

Научная значимость полученных результатов состоит в том, что автором была построена новая производная статистика от исходного временного ряда, являющаяся эффективным индикатором его нестационарности. Эта «горизонтная» статистика построена на базе нормы в пространстве суммируемых функций. Было показано, что для стационарных рядов эта статистика имеет существенно меньшую дисперсию нежели дисперсия исходного распределения, что позволяет предложить ее в качестве квазистационарного критерия близости двух выборочных распределений.

Был получен также новый теоретический результат, относящийся к математической статистике стационарных процессов. Именно, с использованием методов теории функций комплексного переменного была найдена функция распределения «горизонтной» статистики для произвольных стационарных временных рядов в практически важных случаях: при сдвиге на один шаг по времени, а также распределение значений «горизонтного» ряда вблизи своего максимума, определяемого заданной величиной близости двух выборочных распределений, при сдвиге на фиксированное число шагов по времени.

Создание прогнозной модели для выборочной функции распределения нестационарного временного ряда имеет практическую значимость. В условиях, когда нет классических доверительных интервалов для оценок точности моментов распределения, построение прогнозного распределения позволяет использовать его квантили для получения эмпирических оценок доверительных интервалов. Прогнозная модель построена на основе подхода с использованием кинетических уравнений, развитого в статистической механике для описания эволюции сложных динамических систем. Таким образом, предложенная модель обобщает кинетический подход на случай временных рядов, которые могут и не иметь под собой динамического основания.

Эти результаты составляют теоретическую основу для разработанной в диссертации методики прогнозирования выборочных нестационарных распределений путем выделения временных промежутков, на которых ряд является квазистационарным.

Методическая значимость исследования определяется тем, что была разработана система понятий, позволяющих корректно использовать

выборочную статистику, если временной ряд не является стационарным. В диссертации особое внимание обращено на взаимосвязь между объемом выборки и горизонтом прогноза. В существующих прогнозных моделях стационарных временных рядов эти понятия разобщены, причем горизонт и точность прогноза зачастую вообще отсутствуют. Последнее связано с тем, что если ряд приведен к форме разложения на детерминированный и полностью случайный процессы по теореме Вальда, то прогноз белого шума, по предположению, например, гауссовского, делается с известной точностью, определяемой дисперсией шума, на произвольный промежуток времени. В то же время ясно, что такая идеализация в реальных процессах не наблюдается. Исследователь же все возникающие погрешности часто перекладывает «по привычке», выработанной применением различных стандартных критериев, на доверительную вероятность оценок нормальности и стационарности. Тем самым он лишает себя возможности понизить ошибку прогноза по некоторой используемой им стационарной модели, поскольку не рассматривает оптимальный объем выборки, определяемой величиной «приближенной стационарности» процесса.

Введенная «горизонтная» статистика оказалась практически полезным и эффективным инструментом для конструирования более точных прогнозных моделей. Например, анализ этой статистики показал, что для многих рядов, генерируемых процессами на финансовых рынках, характерно представление в виде процесса с переменной долей хаоса. «Горизонтный» ряд, являющийся оценкой промежутка времени, на котором ряд достаточно репрезентативен, таким образом, может выступать в качестве квазистационарного индикатора состояния рынка, что является полезной информацией при принятии решений.

Практическая важность полученных результатов состоит также и в том, что на основе корректной математической модели процесса создан прогнозный алгоритм, гарантирующий заданную точность прогноза выборочной функции распределения и нестационарного временного ряда.

Результаты работы могут быть применены для повышения точности статистического анализа временных рядов, возникающих в различных областях практической деятельности, а также для усовершенствования методов математической статистики. Они могут использоваться в МГУ им. М.В. Ломоносова, в ВЦ РАН, в ИПМ им. М.В. Келдыша РАН, в ИММ РАН, в Международном институте теории прогноза землетрясений и математической геофизики, в МФТИ и других научных и учебных организациях, в планах работ которых существуют задачи, связанные с анализом и прогнозированием случайных процессов.

### **Апробация работы**

Результаты исследований докладывались на конференциях:

1. Международная конференция «Синергетика в естественных науках», Тверь, Россия, 10-13 апреля, 2008.

2. 31-ая конференция молодых ученых и специалистов ИППИ РАН «Информационные технологии и системы», Геленджик, Россия, 29 сентября – 03 октября, 2008.

Результаты исследований докладывались на научных семинарах:

1. Семинар по математической физике ИПМ РАН (рук. В.В. Веденяпин, М.В. Масленников, Ю.Н. Орлов);
2. Семинар «Будущее прикладной математики» ИПМ РАН (рук. Г.Г. Малинецкий);
3. Семинар кафедры высшей математики МФТИ (зав. Е.С. Половинкин);
4. Семинар Международного института теории прогноза землетрясений и математической геофизики РАН (рук. И.В. Кузнецов);
5. Расширенный семинар кафедры высшей геометрии и топологии мехмата МГУ (зав. С.П. Новиков)
6. Семинар Института математического моделирования РАН (рук. Е.И. Леванов)

### **Публикации**

Все представленные в диссертации результаты являются новыми. Они опубликованы в 6 работах диссертанта в 2007 – 2008 годах. В их число входят 2 препринта, статья в трудах конференции, статья в сборнике и 2 публикации в журналах из перечня ВАК, рекомендуемых по данной специальности. Из перечисленных работ одна опубликована диссертантом самостоятельно.

### **Вклад автора в совместных работах**

В работе [1] автором предложен метод исследования временного ряда и проведены численные расчеты. В работах [2-3] автором разработана методика прогнозирования нестационарного временного ряда, предложена модель эволюции соответствующей выборочной функции распределения и создан расчетный алгоритм. В работе [4] автором предложен класс моделей эволюционного типа для функций распределения временных рядов. В работе [5] автором найдены явные выражения для функции распределения горизонтного ряда для стационарных распределений.

### **Структура диссертации**

Диссертация «Алгоритмы прогнозирования нестационарных временных рядов» состоит из введения, четырех глав, заключения, приложения и списка литературы из 96 наименований, расположенных в алфавитном порядке. Каждая глава разбита на параграфы, имеющие двойную нумерацию с указанием на соответствующую главу. Формулы внутри каждого параграфа имеют двойную нумерацию, с указанием на параграф; при ссылке на формулы из другой главы используется тройная нумерация, где первым идет номер главы. Рисунки, определения и утверждения имеют сквозную нумерацию.

## СОДЕРЖАНИЕ РАБОТЫ

Во **введении** описывается та область знаний, которая исследуется автором методами математического моделирования, и формулируется проблема, решаемая в диссертации.

Областью исследований данной работы является математическая статистика нестационарных процессов. Центральная задача состоит в разработке математической модели эволюции выборочной функции распределения (далее ВФР) для некоторых нестационарных временных рядов, таких, как ряд почасовых цен на российском оптовом рынке электроэнергии и мощности ОРЭМ, биржевые котировки акций компаний на примерах General Motors, General Electric, Microsoft, Ford. В качестве базовой модели используется эмпирическое уравнение Лиувилля. Более точная модель использует уравнение Фоккера-Планка.

В статистической механике эволюция функции распределения происходит в силу уравнений динамики, например, в гамильтоновой форме. Для задач статистики динамическая система чаще всего отсутствует, если не иметь в виду так называемых моделей динамического хаоса. В то же время выборочная функция распределения временного ряда некоторым образом изменяется, т.е. эволюционирует. Оператор этой эволюции оказывается возможным записать в форме, аналогичной традиционному уравнению Лиувилля. Отличие в том, что в качестве скорости в нем используется рассчитываемая по эмпирическим данным скорость изменения плотности вероятности. Прогнозом ВФР в работе называется результат численного решения эмпирического уравнения Лиувилля, построенного по данным за некоторый предшествующий период времени. По прогнозу ВФР строится прогноз значения собственно временного ряда.

При прогнозировании ВФР по уравнению Лиувилля производная по времени первого момента ВФР (т.е. среднего выборочного значения ряда приростов случайной величины) определяется скоростью, фигурирующей в эмпирическом уравнении Лиувилля. Однако эволюция в силу уравнения Лиувилля дисперсии исходного ряда будет отличаться от фактического значения производной по времени второго центрального момента ВФР на величину удвоенной выборочной ковариации значений ряда и их приращений. Оказалось, что если взять образующуюся невязку в качестве коэффициента диффузии в уравнении Фоккера-Планка для ВФР, то изменение эмпирической дисперсии будет совпадать с эволюцией второго момента ВФР в силу этого уравнения.

Второй, но не менее важной задачей является разработка системы понятий, в которых можно анализировать и прогнозировать нестационарные случайные процессы. Необходимость такой разработки вызвана тем, что для нестационарных временных рядов в общем случае отсутствуют такие понятия классической математической статистики как генеральная совокупность, состоятельность и несмещенность оценки выборочной статистики,

доверительная вероятность. В качестве базового инструмента для формирования новой системы понятий выступает норма в пространстве суммируемых функций. Порождаемый ей функционал, построенный на паре выборочных функций распределения, определяет новую статистику, названную в работе «горизонтным рядом». Эта статистика позволяет найти оптимальный объем выборки для прогнозирования ВФР на заданный промежуток времени.

Построенная в работе модель эволюции ВФР с оптимальным объемом выборки реализована в виде численного алгоритма на языке Java.

**Первая глава** носит вводный характер.

В параграфе 1.1 представлен обзор основных методов аналитического и численного анализа и прогнозирования временных рядов. Рассмотрены методы анализа, развитые в теории стационарных (регрессионные, автокорреляционные, спектральные) и нестационарных (адаптивные модели) временных рядов, с которыми в дальнейшем сравниваются результаты прогнозной модели, построенной в диссертации.

В параграфе 1.2 анализируются существующие модификации стационарных методов для исследования нестационарных процессов и ограничения их применимости. При определении ошибки прогноза нестационарного временного ряда надо учесть два фактора: конечность выборки и различие распределений для разных выборок вследствие нестационарности процесса. Существующие стационарные методы имеют неодинаковую чувствительность точности аппроксимации данных к действию указанных факторов.

В частности, в моделях регрессионного анализа средние величины (математическое ожидание, дисперсия, ковариация) постоянны. Уточнение этой модели в случае зависимости указанных величин от времени, т.е. от текущего значения  $t$ , может быть сделано посредством аналитического моделирования такой зависимости, либо переходом к первым, вторым и т.д. разностям в нестационарных временных рядах, выражающих зависимость средних величин от времени. Окно усреднения становится при этом скользящим, однако остается невыясненным, какой ширины должно быть это окно. Адаптивные модели, использующие весовые коэффициенты в обобщениях авторегрессионных моделей, требуют весьма тонкой настройки сглаживающих функций в нестационарном случае, поскольку даже для стационарных процессов оптимальный выбор этих функций является отдельной достаточно сложной задачей.

Метод сингулярного спектрального анализа представляется в этом контексте наиболее устойчивым к временному тренду, поскольку его задачей и является выделение соответствующих главных компонент ряда. Изменение с течением времени размерности пространства базисных векторов матрицы задержек представляется маловероятным событием: размерность является своеобразным индикатором данного процесса, обусловленного

определенными физическими явлениями, и ее изменение будет свидетельствовать о том, что процесс изменился по своему качеству. Тем не менее, вопрос о размерности самой матрицы и количественной зависимости от этой размерности числа базисных векторов остается в этом методе открытым.

Таким образом, естественно ставится задача о построении статистик, индикативно описывающих изменчивость ряда. Такими статистиками, о которых речь будет идти ниже, являются: ряд оптимальных объемов выборки для прогнозирования ВФР на заданном интервале времени с заданной точностью, называемый горизонтным рядом, а также ряд максимальных горизонтов прогнозирования, использующих данный объем выборки.

В параграфе 1.3 дается обзор основных программных продуктов в области математической статистики. Делается вывод о том, что существующих статистических пакетов недостаточно для повышения точности прогнозирования нестационарного временного ряда, если ряд не сводится простыми операциями типа временного сглаживания или разностного дифференцирования к стационарному.

В параграфе 1.4 формулируется новый критерий близости двух ВФР. Для объяснения идеи такого критерия рассмотрим операцию усреднения по некоторому промежутку времени. Пусть  $x(t)$  есть значение случайной величины  $\xi$  в момент времени  $t$ . Текущие оценки статистических характеристик процесса  $x(t)$  используют усреднение некоторых операторов  $G[x(t)]$  по скользящему промежутку  $\Delta = [t - T, t]$ :

$$\langle G[x(t)] \rangle_{\Delta} = \frac{1}{T} \int_{t-T}^t G[x(t)] dt. \quad (1.4.1)$$

В дискретном случае под интегралом следует понимать конечную сумму по некоторому разбиению промежутка  $\Delta$ . Обозначения операций в виде интеграла или производной используются для сокращения записи и облегчения восприятия.

Применение преобразований вида (1.4.1) для вычисления текущих оценок статистических характеристик нестационарного процесса  $x(t)$  приводит к появлению двух типов принципиально неустранимых погрешностей: погрешности, возникающей за счет конечности времени усреднения, т.е. вследствие недостаточной репрезентативности объема выборки, а также погрешности, возникающей за счет изменения статистики на интервале усреднения. Чтобы уменьшить ошибки первого типа, следует увеличивать объем выборки, а для уменьшения ошибок второго типа объем выборки следует уменьшать. Трудность задачи минимизации ошибки в оценке статистики (1.4.1) по величине промежутка  $\Delta$  состоит в том, что указанные типы ошибок не могут быть разделены только на основе наблюдений за значениями  $x(t)$ , поэтому функционал полной ошибки неизвестен. Однако, если наряду со статистикой (1.4.1) рассмотреть также и выборочную функцию распределения  $f_T(x, t)$  случайной величины  $\xi$  в момент времени  $t$  по выборке

объема  $T$ , то для величины полной ошибки можно получить оценочные неравенства, ограничивающие ее сверху. Это позволяет проанализировать величину верхней грани полной ошибки как функцию объема выборки, предложить метод уменьшения этой верхней грани и построить соответствующий алгоритм.

Для получения оценки верхней грани ошибки выборочной статистики рассмотрим ошибку прогноза  $\delta$  некоторого временного ряда  $x_i$ , которую определим как среднеквадратичное отклонение прогнозных значений  $\tilde{x}_i$  от фактических  $x_i$  на промежутке горизонта прогноза  $\tau$ :

$$\delta = \sqrt{\frac{1}{\tau} \sum_{i=t+1}^{t+\tau} \delta_i^2}, \quad \delta_i = \tilde{x}_i - x_i. \quad (1.4.3)$$

Определим также ошибку прогноза ВФР в момент времени  $t$ . Этой ошибкой будем называть интегральное абсолютное отличие прогнозной ВФР  $\tilde{f}(x, t)$  от фактической  $f(x, t)$ , построенных по выборкам равных объемов. Обозначим соответствующую величину  $\varepsilon(t)$ . Пусть значения ряда принадлежат некоторому конечному промежутку  $[a; b]$ , который без ограничения общности можно принять за отрезок  $[-1; 1]$ . Тогда

$$\varepsilon(t) = \int_{-1}^1 |\tilde{f}(y, t) - f(y, t)| dy. \quad (1.4.4)$$

Важно подчеркнуть, что ошибка прогноза ВФР и ошибка прогноза временного ряда, по выборке из которого построена данная ВФР, – это разные ошибки, определенные для разных математических объектов и в различных функциональных пространствах. Следовательно, чтобы использовать величины (1.4.3) и (1.4.4) совместно, надо придать корректный теоретико-вероятностный смысл функционалу совокупной ошибки. Такая формализация может быть проведена на основе нижеследующих оценок, получающихся из (1.4.3), если записать выборочную дисперсию в терминах ВФР.

Для простоты рассмотрим прогноз на один шаг вперед по времени. Предположим, что построена некоторая прогнозная ВФР  $\tilde{f}$ . Тогда оценка среднеквадратичной ошибки прогноза ряда определяется как корень из дисперсии прогнозной ВФР  $\tilde{f}$ . Эту дисперсию обозначим через  $\tilde{\sigma}^2$ :

$$\tilde{\sigma}^2(t) = \int_{-1}^1 (x - \tilde{x}(t))^2 \tilde{f}(x, t) dx, \quad \tilde{x}(t) = \int_{-1}^1 x \tilde{f}(x, t) dx. \quad (1.4.5)$$

Различие между прогнозными и фактическими средними величинами можно оценить из неравенства

$$|\tilde{\bar{x}} - \bar{x}| = \int_{-1}^1 |x \cdot (\tilde{f} - f)| dx = \int_{-1}^1 |x| |\tilde{f} - f| dx \leq \varepsilon. \quad (1.4.6)$$

Рассмотрим ошибку прогнозирования значения ряда в некоторый момент времени по среднему значению  $\tilde{\bar{x}}(t)$ . Формальной ошибкой такого прогноза является  $\tilde{\sigma}(t)$ , если ряд стационарный. Оценка квадрата фактической ошибки прогноза ряда в силу (1.4.6) составит

$$\begin{aligned} M\delta^2 &= \int_{-1}^1 (x - \tilde{\bar{x}})^2 f(x, t) dx = \int_{-1}^1 (x - \tilde{\bar{x}} + \bar{x} - \bar{x})^2 f(x, t) dx = \\ &= \sigma^2 + (\bar{x} - \tilde{\bar{x}})^2 \leq \sigma^2 + \varepsilon^2. \end{aligned} \quad (1.4.7)$$

Оценка (1.4.7) позволяет в качестве вышеупомянутой ошибки первого типа  $\Sigma_1$  взять дисперсию ряда на рассматриваемом промежутке, а в качестве ошибки второго типа  $\Sigma_2$  максимум функционала (1.4.4) на том же промежутке. При этом в качестве функционала полной ошибки, т.е. меры неточности прогноза нестационарного временного ряда, можно взять величину  $\Sigma = \sqrt{\Sigma_1^2 + \Sigma_2^2}$ .

Таким образом, для более точного прогнозирования требуется определить такой объем выборки, при котором верхняя оценка величины в левой части (1.4.7) минимальна. В настоящей диссертации формулируется критерий, связывающий между собой точность прогноза, горизонт прогноза и объем требуемой для этого выборки, после чего решается задача минимизации величины (1.4.7) при фиксированной верхней грани ошибки прогноза ВФР на основе анализа соответствующей статистики.

Во **второй главе** приведены примеры часто встречающихся нестационарных временных рядов, прогнозирование которых требуется для конкретных практических нужд. Эти примеры служат иллюстрацией метода построения горизонтной статистики и показывают, что в достаточно большом числе случаев эта статистика представляет практический интерес.

В параграфе 2.1 изучены ценовые ряды на рынке электроэнергии, а в параграфе 2.2 – ряды, возникающие на рынке ценных бумаг. В качестве примеров биржевых рядов рассмотрены ряды цен на акции компаний «Майкрософт» (обозначение на рисунках MS), «Дженерал Моторз» (GM) и «Дженерал Электрик» (GE). Поведение этих временных рядов сравнивается в параграфе 2.3 с рядами, генерируемыми некоторыми классическими динамическими системами с дискретным временем, обладающими хаотической динамикой: логистической, системой Эно и системой Лоренца.

В параграфе 2.4 проводится статистический анализ этих рядов и выделяются некоторые общие черты поведения выборочных статистик. Из оценки (1.4.7) следует, что важным является анализ поведения выборочной дисперсии ряда как функции объема выборки. Для стационарного ряда дисперсия стабилизируется с ростом объема выборки, поскольку дисперсия

выборочной дисперсии стремится к нулю как  $1/n$  согласно теореме Гофдинга. Для нестационарного ряда характерен рост выборочной дисперсии, а стабилизация наступает при значительно большем объеме выборки в силу естественной ограниченности значений ценовых временных рядов.

В параграфе 2.5 строятся горизонтные статистики для рассмотренных рядов и на основе проведенного качественного анализа формулируется модель прогнозирования квазистационарных временных рядов, использующая статистику оптимального объема выборки, согласованного с горизонтом и точностью прогноза.

Обозначим через  $f_T(x, t)$  ВФР, построенную по выборке объема  $T$  в момент времени  $t$ , т.е. на основе данных из окна  $[t - T + 1; t]$ . Далее предполагается, что значения временного ряда равномерно ограничены и принадлежат отрезку  $[-1; 1]$ .

Пусть  $\tau$  есть величина сдвига по времени между выборками. Тогда, рассматривая две ВФР, можно говорить об их интегральной  $\varepsilon$ -близости в том смысле, что

$$\int_{-1}^1 |f_T(y, t + \tau) - f_T(y, t)| dy \leq \varepsilon. \quad (2.5.2)$$

Основной задачей, решаемой на этапе формирования данных для статистического анализа, является изучение статистики объемов выборок как функций времени  $t$ , при которых удовлетворяется неравенство (2.5.2) для заданной неточности  $\varepsilon$  и заданном сдвиге на  $\tau$  дискретных шагов по времени. Рассмотрение этой статистики для прогнозирования временных рядов помогает исследователю определить некоторый оптимальный объем выборки или допустимый горизонт прогноза в зависимости от заданной максимальной величины различия между двумя ВФР. Статистику минимальных объемов выборок  $T_0$  таких, что при всех  $T$ , больших либо равных  $T_0$ , в данный момент времени выполняется условие (2.5.2), будем называть горизонтным рядом и обозначать  $h(t, \tau; \varepsilon)$ .

Можно ожидать, что статистика горизонтного ряда позволит оценить вероятность того, что различие между прогнозной и фактической ВФР не будет превосходить заданной величины в смысле определения (2.5.2). Желательно, чтобы при этом и сам временной ряд прогнозировался бы с небольшой ошибкой, оценкой которой служит выборочная дисперсия. Рассмотренные в работе примеры показывают, что в ряде случаев можно определить такой оптимальный объем выборки, для которого и функции распределения отличаются незначительно, и дисперсия относительно мала.

**В третьей главе** выводятся основные свойства статистики оптимального объема выборки: строится функционал соответствующего критерия, находятся явные формулы для важных случаев и дается алгоритм нахождения для остальных случаев для стационарного временного ряда и

дается их обобщение на нестационарные процессы. Эти результаты представляют теоретическое обоснование прогнозной модели, построение которой является одним из центральных результатов диссертации.

В параграфах 3.1 и 3.2 формулируются основные понятия, которые используются далее для построения прогнозной модели для ВФР.

*Определение 3.* ВФР  $f_T(x, t)$  временного ряда  $x(t)$  будем называть  $\theta$ - $\varepsilon$ -стационарной на временном промежутке  $\theta$ , если

$$\forall \tau : 1 \leq \tau \leq \theta, \quad \forall t \quad V(T, \tau; t) = \int_{-1}^1 |f_T(y, t + \tau) - f_T(y, t)| dy \leq \varepsilon \quad (3.1.7)$$

Если неравенство (3.1.7) при данном  $T$  выполнено для одного конкретного значения  $\tau$ , то ВФР называем ограниченно  $\tau$ - $\varepsilon$ -стационарным. Если неравенство  $V(T, \tau; x, t) \leq \varepsilon$  при данном  $T$  выполняется для всех  $\tau \geq 1$ , то ВФР будем называть просто  $\varepsilon$ -стационарной.

Некоторые важные свойства функционала (3.1.7) составляют содержание следующих доказанных в работе утверждений:

*Теорема 1.* Если временной ряд  $x(t)$  стационарный, то  $\forall \varepsilon > 0 \exists T \geq 1$  такое, что ВФР  $f_T(x, t)$  является  $\varepsilon$ -стационарной.

*Следствие 1.* Если две ВФР  $f_T(x)$  и  $\tilde{f}_T(x)$  удовлетворяют условию (3.1.7), то соответствующие интегральные ВФР  $F_T(x)$  и  $\tilde{F}_T(x)$  удовлетворяют критерию Колмогорова-Смирнова с тем же значением  $\varepsilon$ .

*Теорема 2.* Для функционала (3.1.7) имеет место равномерная оценка

$$0 \leq V(T, \tau; x, t) \leq \min(2\tau / T; 2). \quad (3.2.1)$$

*Определение 4.* Горизонтным рядом для ряда  $x(t)$  при сдвиге на промежуток  $\tau$  называется такой минимальный объем выборки  $h(t, \tau; \varepsilon)$ , что при всех  $T \geq h(t, \tau; \varepsilon)$  выполнено условие  $V(T, \tau; t) \leq \varepsilon$ .

Рассмотрим величину

$$H(\tau; \varepsilon) = \max_t h(t, \tau; \varepsilon). \quad (3.2.2)$$

Из (3.2.1) следует, что верхней оценкой минимально допустимого объема выборки  $H(\tau; \varepsilon)$  является  $2\tau / \varepsilon$ . Из теоремы 2 также следует, что если взять  $H(\tau; \varepsilon) = 2\tau / \varepsilon$ , то для всех  $T \geq H(\tau, \varepsilon)$  и для всех  $\tau' \leq \tau$  выполняется  $V(T, \tau; t) \leq \varepsilon$ , т.е. такая выборка гарантированно будет  $\tau$ - $\varepsilon$ -стационарной.

Введем плотность распределения  $\psi_{\tau, \varepsilon}(T)$  значений горизонтного ряда  $h(t, \tau; \varepsilon)$ , т.е. определим вероятность того, что расстояние между двумя ВФР, построенных по выборке объема  $T$  и сдвинутых по времени на фиксированный промежуток  $\tau$ , не превосходит  $\varepsilon$  для всех  $T' \geq T$ . Для краткости зависимость от момента времени  $t$  в аргументах  $\psi_{\tau, \varepsilon}(T)$  опущена.

Определим также интегральную функцию распределения

$$\Psi_{\tau,\varepsilon}(T) = \sum_{k=1}^T \psi_{\tau,\varepsilon}(k), \quad \Psi_{\tau,\varepsilon}(2\tau/\varepsilon) = 1. \quad (3.2.3)$$

Смысл функции (3.2.3) в том, что если вместо величины  $H(\tau; \varepsilon)$  из (3.2.2), во многих случаях оказывающейся равной своему теоретическому максимуму  $2\tau/\varepsilon$ , взять некоторое меньшее значение  $T^*$ , то с вероятностью  $\Psi_{\tau,\varepsilon}(T^*)$  будут выполнены условия  $\tau$ - $\varepsilon$ -стационарности.

*Определение 5.* Оптимальным объемом  $T_{opt}$  для прогнозирования временного ряда  $x(t)$  на горизонт  $\tau$  будем называть такой объем выборки, при котором ВФР этого ряда  $\tau$ - $\varepsilon$ -стационарна, причем относительная выборочная дисперсия ВФР минимальна среди всех дисперсий по выборкам, объем которых удовлетворяет условию  $\tau$ - $\varepsilon$ -стационарности и не превосходит величины  $2\tau/\varepsilon$ .

*Теорема 3.* Пусть ВФР  $f_T(x, t)$   $\tau$ - $\varepsilon$ -стационарна. Тогда функционал  $V(T^*, \tau; t)$ , где  $T^*$  есть  $1-\varepsilon$ -квантиль распределения  $\psi_{\tau,\varepsilon}(T)$  ее горизонтного ряда, будет в среднем отличаться от функционала  $V(T, \tau; t)$  на величину  $o(\varepsilon)$ .

На основе теоремы 3 в диссертации предложена следующая модель временного ряда, представляющего оптимальный объем выборки:

$$T_{2opt} = \max \left\{ T^*, \arg \min \sigma(T) \right\}, \quad (3.2.9)$$

где  $T^*$  определяется как  $1-\varepsilon$ -квантиль распределения  $\psi_{\tau,\varepsilon}(T)$ , т.е. такая величина, для которой

$$\Psi_{\tau,\varepsilon}(T^*) = 1 - \varepsilon.$$

В параграфах 3.3-3.5 строится распределение  $\psi_{\tau,\varepsilon}(T)$  для произвольного стационарного временного ряда  $x(t)$  и исследуются свойства введенной горизонтной статистики  $h(t, \tau; \varepsilon)$ .

Для стационарного ряда  $x(t)$  отрезок принимаемых им значений  $[-1; 1]$  разбивается на  $N$  отрезков длины  $2/N$ . Пусть  $p_i$  есть вероятность того, что значение ряда попадает в промежуток  $\Delta_i$ :

$$x(t) \in \Delta_i \subset [-1; 1], \quad \Delta_i = [-1 + 2(i-1)/N; -1 + 2i/N], \quad i = 1, 2, \dots, N.$$

В случае совпадения значения  $x(t)$  с граничной точкой двух отрезков  $\Delta_i, \Delta_{i+1}$ , будем для определенности считать его принадлежащим левому из них. Тогда для распределения  $\psi_{\tau,\varepsilon}(T)$  при сдвиге ВФР на  $\tau = 1$  шаг в работе получена формула:

$$\psi_{1,\varepsilon}(m) \equiv P\{h(t, 1; \varepsilon) = m\} = \begin{cases} \sum_{i=1}^N p_i^M, & m=1 \\ \sum_{i=1}^N p_i^{M-m+1}(1-p_i), & 1 < m \leq M = [2/\varepsilon]. \\ 0, & m > M \end{cases} \quad (3.3.1)$$

Распределение (3.3.1) можно назвать, по аналогии с известным геометрическим распределением  $p_k = a^k(1-a)$ ,  $0 < a < 1$ ,  $k = 1, 2, \dots$ , конечно-аддитивным геометрическим распределением. Отсюда следует

*Теорема 4. Каково бы ни было распределение вероятностей  $\{p_i\}_1^N$ , отделенное от нуля и единицы, относительное среднее квадратичное отклонение распределения горизонтного ряда при сдвиге на 1 шаг в смысле (1.4.4) имеет порядок малости не больше расстояния между двумя ВФР исходного ряда, т.е. может быть сделано сколь угодно малым.*

Это свойство квазистационарности горизонтной статистики сохраняется и при сдвиге на произвольное число шагов. Выражение для распределения горизонтного ряда в этом общем случае также имеет полиномиальную структуру по вероятностям  $p_i$ , однако в явном виде получить точные аналитические формулы затруднительно. Явный вид распределения найден в практически важном случае, когда горизонтный ряд рассматривается вблизи максимальных значений, определяемых заданной величиной близости двух выборочных распределений при сдвиге на  $\tau$  шагов по времени.

Обозначим вероятность того, что  $\tau$  значений временного ряда попали соответственно в промежутки  $\Delta_{i_1}, \Delta_{i_2}, \dots, \Delta_{i_\tau}$ , через  $a \equiv a(i_1, \dots, i_\tau)$ :

$$a(i_1, \dots, i_\tau) = \sum_{k=1}^{\tau} p_{i_k} - \sum_{\substack{k,l=1 \\ k \neq l}}^{\tau} p_{i_k} p_{i_l} + \dots + (-1)^{\tau-1} \prod_{k=1}^{\tau} p_{i_k}. \quad (3.3.13)$$

Обозначим также условную вероятность того, что значения горизонтного ряда не превосходят заданного числа  $m = [2\tau/\varepsilon] + \tau - k$  при условии попадания значений исходного ряда в промежутки  $\Delta_{i_1}, \Delta_{i_2}, \dots, \Delta_{i_\tau}$ , через  $A_k \equiv A_k(i_1, \dots, i_\tau)$ :

$$A_k = P\{h(t, \tau; \varepsilon) \leq m = [2\tau/\varepsilon] + \tau - k | a(i_1, \dots, i_\tau)\}. \quad (3.3.14)$$

Для вероятностей  $A_k$  в работе получено соотношение

$$\begin{aligned} A_k &= a(t-m-1) \cdot A_{k-1} + a(t-m-2) \cdot (1-a(t-m-1)) \cdot A_{k-2} + \\ &+ a(t-m-3) \cdot (1-a(t-m-1)) \cdot (1-a(t-m-2)) \cdot A_{k-3} + \dots \\ &+ a(t-m-n) \cdot \prod_{i=1}^{n-1} (1-a(t-m-i)) A_{k-n}. \end{aligned}$$

Если ряд  $\{x(t)\}$  стационарен, то величины  $a(t)$  не зависят от времени и  $A_k$  образуют линейную рекуррентную последовательность, характеристическое уравнение которой имеет вид:

$$Q(\lambda) = \frac{\lambda^{n+1} - \lambda^n + a(1-a)^n}{\lambda - (1-a)} = 0. \quad (3.3.15)$$

Обозначим  $\lambda_1, \dots, \lambda_n$  корни уравнения  $Q(\lambda) = 0$ . Из (3.3.15) и из того, что  $0 < a < 1$  следует, что все  $\lambda_i$  заключены внутри объединения двух кругов

единичного радиуса с центрами в 0 и 1. Тогда  $A_k$  имеют вид  $A_k = \sum_{i=1}^n C_i \lambda_i^k$ ,

где константы  $C_i$  находятся с помощью определителей Вандермонда:

$$C_i = \frac{W(\lambda_1, \dots, \overset{i}{1}, \dots, \lambda_n)}{W(\lambda_1, \dots, \lambda_i, \dots, \lambda_n)} = \frac{(\lambda_1 - 1) \cdot \dots \cdot (\lambda_n - 1)}{(\lambda_1 - \lambda_i) \cdot \dots \cdot (\lambda_n - \lambda_i)} = - \frac{Q(1) / (\lambda_i - 1)}{Q(z) / (z - \lambda_i) \Big|_{z=\lambda_i}}. \quad (3.3.16)$$

Из (3.3.16) следует, что величины  $A_k$  могут быть представлены с помощью вычетов в виде интеграла Коши:

$$\begin{aligned} A_k &= - \sum_{i=1}^n \frac{Q(1) / (\lambda_i - 1)}{Q(z) / (z - \lambda_i) \Big|_{z=\lambda_i}} \lambda_i^k = -Q(1) \sum_{i=1}^n \frac{z^k (z - \lambda_i)}{Q(z)(z - 1)} \Big|_{z=\lambda_i} = \\ &= -Q(1) \sum_{i=1}^n \operatorname{res}_{\lambda_i} \frac{z^k}{Q(z)(z - 1)} = 1 - \frac{Q(1)}{2\pi i} \int_{\Pi} \frac{z^k dz}{Q(z) \cdot (z - 1)}. \end{aligned} \quad (3.3.17)$$

Здесь  $\Pi$  – контур, содержащий все корни характеристического уравнения и образованный отрезками прямых  $\operatorname{Re} z = -2$ ,  $\operatorname{Re} z = 3$ ,  $\operatorname{Im} z = \pm 2$ . Тогда распределение ряда прогнозных объемов имеет вид

$$P\{h(t, \tau; \varepsilon) = m = [2\tau / \varepsilon] + \tau - k\} = \sum_{i_1, \dots, i_\tau=1}^N p_{i_1} \dots p_{i_\tau} (A_k(i_1, \dots, i_\tau) - A_{k+1}(i_1, \dots, i_\tau)). \quad (3.3.18)$$

*Следствие 2.* В интервале  $M_{\tau-1} < h(t, \tau; \varepsilon) \leq M_\tau$  распределение  $\psi_{\tau, \varepsilon}([2\tau / \varepsilon] - m)$  при  $1 \leq m \leq \tau$  равномерно.

*Следствие 3.* Справедлива формула

$$\psi_{\tau, \varepsilon}(M_\tau) = \left( \sum_{i=1}^N (1 - f_i)^\tau f_i \right)^\tau.$$

Дальнейшее развитие методики использования горизонтной статистики для построения прогнозной модели нестационарного временного ряда основано на следующем доказываемом в работе утверждении:

*Теорема 5. Пусть распределение ряда  $x(t)$  является  $\varepsilon$ -стационарным. Тогда с точностью  $o(\varepsilon)$  распределение горизонтного ряда для  $\tau = 1$  также  $\varepsilon$ -стационарно.*

В четвертой главе на основе доказанных свойств статистики оптимального объема выборки формулируется методика прогнозирования выборочной функции распределения нестационарного временного ряда и описывается ее численная реализация. Проводится сравнение результатов прогнозирования конкретных временных рядов разными методами: развитым в диссертации методом выделения горизонта прогноза и использованием уравнения Лиувилля для выборочной функции распределения с оптимальной прогнозной выборкой, и методом скользящих средних для регрессионного и автокорреляционного анализа.

В параграфе 4.1 описывается численный алгоритм построения горизонтного ряда. Затем на основе равенства (3.2.9) определяется текущий оптимальный объем выборки и строятся выборочные функции распределения исходного ряда и его разностной производной.

В параграфах 4.2-4.3 описывается математическая модель прогноза нестационарной ВФР с помощью уравнений Лиувилля и Фоккера-Планка. Вводится двумерная ВФР  $F(x, \dot{x}, t)$  совместного распределения случайных величин  $x$  и  $\dot{x}$  исходного ряда и ряда его производной, полученного взятием разностей  $x(t) - x(t-1)$  в соседние моменты времени. Одномерная ВФР определяется как  $f(x, t) = \int F(x, \dot{x}, t) d\dot{x}$ . Формально из сохранения во времени нормировки функции распределения следует уравнение Лиувилля

$$\frac{\partial F(\xi, t)}{\partial t} + \text{div}_{\xi}(\dot{\xi}F(\xi, t)) = F \text{div}_{\xi} \dot{\xi}, \quad \xi = (x, \dot{x}, \dots), \quad (4.2.1)$$

где через  $\dot{\xi}$  обозначена фазовая скорость, возможно, для большего, чем 2, числа фазовых переменных.

Если фазовая мера динамически-инвариантна, то  $\text{div}_{\xi} \dot{\xi} = 0$ . Однако в нашем случае временной ряд не является динамической системой, так что условие  $\text{div}_{\xi} \dot{\xi} = 0$  на эмпирических данных может и не выполняться. Обозначим через  $u(x, t)$  функцию средней эмпирической скорости по распределению  $F(x, \dot{x}, t)$ , а через  $g(x, t)$  среднее значение дивергенции фазовой скорости:

$$u(x, t)f(x, t) = \int \dot{x}F(x, \dot{x}, t) d\dot{x}, \quad g(x, t)f(x, t) = \int F(x, \dot{x}, t) \left( \frac{\partial \dot{x}}{\partial x} + \frac{\partial \ddot{x}}{\partial \dot{x}} \right) d\dot{x}. \quad (4.2.2)$$

Из (4.2.1) следует, что уравнение эволюции для  $f(x, t)$  имеет вид

$$\frac{\partial f(x, t)}{\partial t} + \frac{\partial}{\partial x} u(x, t)f(x, t) = g(x, t)f(x, t). \quad (4.2.4)$$

Поскольку уравнение (4.2.4) записано относительно эмпирической ВФР, то будем его называть «эмпирическим» уравнением Лиувилля. Это означает,

что скорость изменения ВФР  $u(x, t)$  является параметром, определяемым по данной конкретной выборке в момент  $t$ , а не выводится из каких-либо уравнений, как в традиционной статистической механике.

Введем следующие средние по  $x$  величины в момент времени  $t$ :

$$\mu(t) = \langle x \rangle = \int x f(x, t) dx, \quad U(t) = \langle u(x, t) \rangle = \int u(x, t) f(x, t) dx. \quad (4.2.8)$$

Тогда скорость изменения скользящего среднего  $\mu(t)$  в силу уравнения (4.2.4) имеет вид

$$\frac{d\mu}{dt} = \int x \frac{\partial f}{\partial t} dx = U + \langle xg(x, t) \rangle. \quad (4.2.9)$$

Это уравнение означает, что изменение со временем среднего значения ряда равно средней скорости изменения значений ряда по скользящей выборке плюс среднее значение произведения  $x$  на дивергенцию фазовой скорости, что совпадает с представлениями математической статистики в дискретном случае. Таким образом, уравнение (4.2.4) является адекватной моделью для описания эволюции первого момента ВФР.

*Определение 7.* Прогнозом ВФР на шаг  $\tau$  будем называть  $\tau$ - $\varepsilon$ -стационарную ВФР, построенную к моменту времени  $t + \tau$  на основе эмпирического уравнения эволюции по данным выборки  $\Delta(t - T, t) = [t - T, t] \equiv \Delta(t)$ , известных к моменту времени  $t$ .

В работе принята вычислительная модель для определения прогнозной ВФР с центральной разностной аппроксимацией производных по  $x$ :

$$\begin{aligned} \tilde{f}(x, t + 1) = & f(x, t) \cdot (1 + g(x, t)) + \\ & + \frac{1}{2} u(x - 1, t) f(x - 1, t) - \frac{1}{2} u(x + 1, t) f(x + 1, t). \end{aligned} \quad (4.2.12)$$

Уточнение прогнозной модели, использующей уравнение Лиувилля, может быть проведено на основе анализа эволюции центральных моментов высших порядков. Введем

$$m_k(t) = \int (x - \mu(t))^k f(x, t) dx, \quad k \geq 2. \quad (4.3.1)$$

Эволюция моментов (4.3.1) эмпирического распределения в силу (4.2.4) определяется уравнением

$$\frac{dm_k}{dt} = k \sum_{j=0}^{k-2} (-1)^j \binom{k-1}{j} \mu^j \text{cov}(x^{k-1-j}, u) + s_k, \quad (4.3.2)$$

где  $s_k = \langle (x - \mu(t))^k g(x, t) \rangle$ . С другой стороны, можно непосредственно вычислить значение производной выборочных моментов по времени:  $\dot{m}_k^{(e)}(t) = m_k^{(e)}(t) - m_k^{(e)}(t - 1)$ . Если потребовать, чтобы эволюция выборочных моментов (4.3.2) численно совпадала бы с  $\dot{m}_k^{(e)}$ , то это приведет к появлению дополнительного члена в модельном уравнении эволюции для ВФР.

Например, пусть дисперсия данной ВФР в текущий момент времени равна  $\dot{m}_2^{(e)}$ . Обозначим невязку этого значения с уравнением (4.3.2) через  $\lambda$ :

$$\lambda(t) = \dot{m}_2^{(e)} - 2\text{cov}(x, u) - s_2. \quad (4.3.6)$$

Рассмотрим для ВФР вместо уравнения Лиувилля (4.2.4) уравнение Фоккера-Планка:

$$\frac{\partial f}{\partial t} + \frac{\partial(uf)}{\partial x} - \frac{\lambda(t)}{2} \frac{\partial^2 f}{\partial x^2} = gf. \quad (4.3.7)$$

Для этого уравнения в силу его дивергентной формы уравнение эволюции первого момента остается тем же, что и для уравнения Лиувилля, т.е. имеет вид (4.2.9), а уравнение эволюции выборочной дисперсии, как легко проверить, совпадает с соответствующей эмпирической производной  $\dot{m}_2^{(e)}$ .

В параграфе 4.4 формулируется методика прогнозирования функции распределения и временного ряда, объединяющая теорию глав 3-4 и отдельные расчетные алгоритмы в единую прогнозную систему. Эта методика включает: исходные требования к постановке задачи, формализацию задаваемых параметров, правила формирования массивов данных, порядок работы с массивами большого объема, расчет и создание массивов выборочных средних и дисперсий, определение на их основе оптимального объема выборки, алгоритм формирования массива ВФР в различные моменты времени, алгоритм прогноза ВФР на  $\tau$  шагов вперед, модель прогноза исходного ряда на  $\tau$  шагов вперед. В частности, прогнозным значением  $\tilde{x}(t+1)$  временного ряда считается среднее значение случайной величины по найденному прогнозному распределению вероятностей:

$$\tilde{x}(t+1) = \int x\tilde{f}(x, t+1)dx. \quad (4.4.12)$$

В параграфе 4.5 приводятся результаты численных расчетов по прогнозной модели (3.2.9), (4.2.12) и (4.3.7), и сравнивается точность прогнозов, полученных другими методами, не использующими алгоритм оптимального объема выборки. Для рассмотренных в диссертации примеров точность расчета кинетическим методом, предложенным соискателем, была выше, чем точность стандартных регрессионных или автокорреляционных методов и адаптивных методов скользящего усреднения.

**В заключении** кратко подытоживаются основные результаты диссертации и обсуждаются возможные области их применения, указываются ограничения построенной прогнозной модели и возможности ее совершенствования.

**В приложении** представлены блок-схемы разработанных численных алгоритмов прогнозирования нестационарных выборочных функций распределения.

## ОСНОВНЫЕ РЕЗУЛЬТАТЫ, ВЫНОСИМЫЕ НА ЗАЩИТУ

1. На основе равномерных оценок нормы в пространстве суммируемых функций построена «горизонтная» статистика минимальных объемов выборки для прогнозирования ВФР, согласованная с горизонтом и точностью прогноза. Показано, что для широкого класса нестационарных рядов существует интервал времени, называемый горизонтом прогноза, на котором ВФР является квазистационарной. Для стационарных временных рядов получена аналитическая формула для функции распределения «горизонтного» ряда в практически важных случаях.
2. На основе кинетических уравнений Лиувилля и Фоккера-Планка построена математическая модель прогнозирования ВФР нестационарных временных рядов, таких, как почасовые цены на российском оптовом рынке электроэнергии и мощности, биржевые котировки акций компаний General Motors, General Electric, Microsoft, Ford и др.
3. Создан и протестирован (на примерах динамических систем с хаосом) численный алгоритм, реализующий математическую модель и осуществляющий прогноз выборочной функции распределения с ошибкой, не превышающей заданную величину, на заданный горизонт. Этот алгоритм включает в себя: генерацию «горизонтной» статистики по данному временному ряду, определение максимального объема выборки для прогнозирования с заданной точностью, сопоставление временному ряду некоторой квазидинамической системы, решение кинетического уравнения, прогнозирование ВФР. По найденной прогнозной ВФР построен прогноз значений временного ряда.

### Список работ по теме диссертации

1. Орлов Ю.Н., Осминин К.П. Анализ нестационарных временных рядов / Препринт ИПМ им. М.В. Келдыша РАН, № 36, 2007. – 24 с.
2. Орлов Ю.Н., Осминин К.П. Построение выборочной функции распределения для прогнозирования нестационарного временного ряда // Математическое моделирование, 2008, № 9, с 23-33.
3. Орлов Ю.Н., Осминин К.П. Методика определения оптимального объема выборки для прогнозирования нестационарного временного ряда // Информационные технологии и вычислительные системы, 2008, № 3, с. 3-13.
4. Орлов Ю.Н., Осминин К.П. Кинетические уравнения для прогнозирования нестационарных временных рядов / Препринт ИПМ им. М.В. Келдыша РАН, № 47, 2008. – 28 с.
5. Орлов Ю.Н., Осминин К.П. О квазистационарности статистики «горизонтного» ряда для нестационарных временных рядов / В сб. Современные проблемы фундаментальной и прикладной математики. М.: МФТИ, 2008, с. 113-130.
6. Осминин К.П. Горизонтная статистика в прогнозировании нестационарных временных рядов / Труды Международной конференции «Синергетика в естественных науках», Тверь, Россия, 10-13 апреля, 2008, с. 47-49.