

Министерство образования Российской Федерации

Э.Е. Тихонов, В.А. Кузьмищев

Методы и алгоритмы прогнозирования экономических показателей на базе нейронных сетей и модулярной арифметики

igwt@mail.ru

**Невинномысск
2004**

УДК 333.27:681.3

Э.Е. Тихонов, В.А. Кузьмищев.

Методы и алгоритмы прогнозирования экономических показателей на базе нейронных сетей и модулярной арифметики: Монография. - Невинномысск: Издательство НИЭУП, 2004. - 166 с.

ISBN 5-94812-020-1

Монография посвящена вопросам совершенствования методов и алгоритмов прогнозирования с использованием преимуществ теории хаоса, нейронных сетей и модулярной арифметики. В работе представлен подробный обзор на основе критического анализа по методам и алгоритмам прогнозирования.

Монография предназначена для практикующих специалистов, работающих на фондовых и валютных рынках в брокерских компаниях, банках и на крупных предприятиях, а также для преподавателей, аспирантов и студентов экономических специальностей.

Табл. 10. Ил. 42 Библиогр.: 249 назв.

Научный редактор -
канд. тех. наук, доцент Д.В. Болдырев

Рецензенты:
д-р. физ.-мат. наук, проф. В.И. Лебедев;
д-р. физ.-мат. наук, проф. В.И. Дроздова.
(Северо-Кавказский государственный
технический университет)

УДК 333.27:681.3

ISBN 5-94812-020-1

© Тихонов Э.Е. Кузьмищев В.А., 2004
© Невинномысский институт экономики,
управления и права, 2004

Оглавление

Введение.....	6
Глава 1 аналитический обзор моделей и методов прогнозирования.....	10
1.1 Прогнозная экстраполяция.....	13
1.1.1 Метод наименьших квадратов.....	15
1.1.2 Метод экспоненциального сглаживания.....	17
1.1.3 Метод вероятностного моделирования.....	21
1.2 Интуитивные (экспертные) методы прогнозирования.....	26
1.3 Корреляционный и регрессионный анализы.....	34
1.4 Модели стационарных временных рядов и их идентификация.....	41
1.4.1 Модели авторегрессии порядка p (AR(p)-модели).....	43
1.4.2 Модели скользящего среднего порядка q (MA(q)-модели).....	45
1.4.3 Авторегрессионные модели со скользящими средними в остатках (ARMA(p, q)-модели).....	46
1.5 Модели нестационарных временных рядов и их идентификация.....	47
1.5.1 Модель авторегрессии-проинтегрированного скользящего среднего (ARIMA(p, k, q)-модель).....	47
1.5.2 Модели рядов, содержащих сезонную компоненту.....	49
1.5.3 Прогнозирование на базе ARIMA-моделей.....	50
1.6 Адаптивные методы прогнозирования.....	52
1.7 Метод группового учета аргументов.....	56
1.8 Теория распознавания образов.....	63
1.9 Прогнозирование с использованием нейронных сетей, искусственного интеллекта и генетических алгоритмов.....	64
Выводы к главе.....	68
Глава 2 Модели хаоса для задач прогнозирования курсов акций.....	72

2.1	Методика нелинейно-динамического анализа данных.....	72
2.2	Практические исследования стохастических характеристик динамики курса акции «РАО ЕЭС»...	80
2.2.1	Оценка τ по автокорреляционной функции.....	81
2.2.2	Спектральный анализ.....	82
2.2.3	Оценка показателя Херста для курса акций компании «РАО ЕЭС».....	84
2.2.4	Построение фазового портрета. Восстановление аттрактора по временному ряду.....	86
2.2.5	Вычисление корреляционного интеграла.....	88
2.2.6	Вычисление корреляционной энтропии и старшего показателя Ляпунова.....	89
2.3	Исследование основных нелинейных показателей для курсов акций «Ростелеком», «Лукойл», «Сбербанк».....	91
2.3.1	R/S – анализ. Анализ показателя Херста.....	95
2.3.2	Восстановление фазовых портретов и аттракторов по временным рядам.....	96
2.3.3	Вычисление корреляционных интегралов и фрактальной размерности.....	99
2.3.4	Вычисление корреляционной энтропии и показателей Ляпунова.....	102
	Выводы к главе.....	105

Глава 3	Применение нейронных сетей в задачах прогнозирования и проблемы идентификации моделей прогнозирования на нейронных сетях....	107
3.1	Сравнительный анализ нейронных сетей.....	109
3.2	Об отрицательном влиянии процедуры перемешивания данных (Shuffle) при обучении нейронных сетей, применяемых для прогнозирования.....	114
3.3	Исследование нейросетевых структур для курсов акций «Ростелеком», «Лукойл», «Сбербанк».....	117
3.4	Многофакторные нейронные сети.....	124
3.5	Разработка математических моделей непозиционного кодирования для решения задач	

прогнозирования на нейронных сетях.....	127
3.6 Анализ методов повышения качества предсказаний..	136
Выводы к главе.....	141
Глава 4 Сравнительная оценка классических и нейросетевых методов прогнозирования. оценка точности и погрешности вычислений в нейронных сетях.....	142
Литература.....	149

Введение

Современное развитие человека, общества и экономики имеет направленность в будущее, что нашло отражение в возникновении таких понятий, как «предвидение», «предсказание», «прогноз», «план», «программа». Научное предвидение – это та часть, сторона, этап познавательной деятельности субъекта, результатом которого является получение знаний о будущих событиях. Определенное таким образом понятие «предвидение» может служить родовым для понятия «предсказание» и «прогноз».

В настоящее время в литературе используется два понятия, связанных с прогнозированием: прогнозирование (*forecasting*) и предсказание (*prediction*), толкуемые не совсем однозначно. Дословный перевод глагола *to predict*, имеющего латинское происхождение, означает «сказать заранее», а глагола *to forecast* – «бросать вперед».

Развитие прогностики как науки в последние десятилетия привело к созданию множества методов, процедур, приемов прогнозирования, неравноценных по своему значению. Отсутствие их четкой систематизации приводит к дальнейшему расширению инструментария прогностики зачастую малоценными и компилятивными методами. По оценкам зарубежных и отечественных систематиков прогностики, уже насчитывается свыше ста методов прогнозирования. Число базовых методов прогностики, которые в тех или иных вариациях повторяются в других методах, гораздо меньше. Одни из этих «методов» относятся скорее к отдельным приемам или процедурам прогнозирования, другие представляют набор отдельных приемов, отличающихся от базовых или друг от друга количеством частных приемов и последовательностью их применения.

Модели сложных систем, таких как финансовые рынки, не всегда могут давать однозначные рекомендации или прогноз. Эти модели всегда должны указывать, при достижении каких значений параметров, описывающих систему, или какого определенного момента времени может произойти нечто непредвиденное «катастрофа». Порой они должны указывать и область непредсказуемости (т. е. область параметров, в которой поведение системы неконтролируемо или непредсказуемо).

Среди факторов, характеризующих динамику рынка и влияющих на нее, есть изрядное количество данных нечисловой природы, значения которых известны только с определенной долей уверенно-

сти. Можно выделить различные типы неопределенностей, из которых для финансового анализа важны следующие:

- связанные с незнанием или неточным знанием некоторых факторов или процессов, влияющих на развитие ситуации;
- связанные с математической несоизмеримостью численных оценок величин, характеризующих динамику системы;
- связанные с нелинейностью и наличием у системы нескольких состояний равновесия или аттракторов;
- связанные с недостатком или неадекватностью понятийного аппарата и невозможностью отождествления фактов.

Для понимания того, какие же преимущества дают предлагаемые методы анализа данных и прогнозирования, необходимо указать на три принципиальные проблемы, возникающие при создании систем поддержки принятия решений и анализа на финансовых рынках.

Первая – это определение необходимых и достаточных параметров для оценки состояния рынка, а также выбор критериев эффективности действий. Формализация поведения систем, включающих разнородные компоненты, требует единой метрики описания ситуации.

Вторая проблема заключается в так называемом "проклятье размерности". Желание учесть в модели как можно больше показателей и критериев оценки может привести к тому, что требуемая для ее решения компьютерная система вплотную приблизится к "пределу Тьюринга (ограничению на быстродействие и размеры вычислительного комплекса в зависимости от количества информации, обрабатываемого в единицу времени).

Третья проблема – наличие феномена надсистемности. Взаимодействующие системы образуют надсистему – систему более высокого уровня, обладающую собственными свойствами. Феномен заключается в принципиальной недостижимости надсистемного отображения и целевых функций с точки зрения систем, входящих в состав надсистемы.

Для преодоления некоторых из перечисленных проблем, с которыми приходится сталкиваться при анализе финансовой ситуации делаются попытки применения таких разделов современной фундаментальной и вычислительной математики, как нейрокompьютеры, теория стохастического моделирования (теория хаоса) и теория рисков, теория катастроф, синергетика и теория самоорганизующихся систем (включая генетические алгоритмы), теория фракталов, нечет-

кие логики и даже виртуальная реальность. Считается, что эти методы позволят увеличить глубину прогноза на финансовых рынках за счет выявления скрытых закономерностей и взаимосвязей среди плохо формализуемых обычными методами макроэкономических, политических и глобальных финансовых показателей.

Существуют уже разработанные системы и методики (например использующие аппарат нечетких логик), оболочки экспертных систем, поддерживающие работу с нечеткими знаниями, такие, например, как Gold Works, Guru, Nexpert Object with Nextra, Flex, 1stClass HT. Практически все они используют для генерации правил (после заполнения базы знаний) алгоритм Куинлена ID3. Созданы первые в мире электронные таблицы FuziCalc, способные работать с нечеткими данными. Существуют и достаточно мощные средства разработки приложений, использующих аппарат нечетких логик, – пакеты CubiCalc RTS и CubiCalc 2.0 для Windows фирмы HyperLogic. Завоевали признание и нейросетевые технологии. Практика использования нейросетей показала их эффективность в таких областях, как прогнозирование, выявление зависимостей, ситуационное управление. Используя информацию о динамике стоимости ценных бумаг, об изменениях показателей экономической активности и о колебаниях курса, например, государственных облигаций, можно выявить существующие между этими характеристиками взаимозависимости. Это позволяет выявить, как тот или иной показатель либо их комбинация с учетом динамики развития влияют на изменение курса ценных бумаг. На сегодняшний день используются такие системы как:

- Fujitsu (используется в Японии фирмой Nikko Securities);
- "Прогноз макроэкономических индикаторов" фирмы Data Market (нейроплата);
- S&PCBRS для прогнозирования индекса S&P500 и курсов акций, созданная в Chase Manhattan Bank;
- система биржевых прогнозов HNC, работающая в Citibank; а также такие коммерческие продукты для работы на финансовых рынках, как Nestor DLS фирмы Nestor, пакет Nexpert Object фирмы Neuron Data, программы NeuroShell 2 v.3, NeuroWindows v.4.6 и один из наиболее популярных в мире пакетов на основе генетических алгоритмов GeneHunter v.1.0 и пакет Brain Maker Pro.

Еще одним методом, находящим все большее применение при анализе финансовых рынков, и особенно быстротекущих процессов

на них, является теория хаоса, или теория коллективной (хаотической) динамики.

Применительно к области финансов на основе теории хаоса впервые был разработан принципиально новый подход к анализу рынка, отличный от "портфельной теории". Этот подход базируется на положении о том, что рынок представляет собой сложную нелинейную систему с обратной связью, а характер группового взаимодействия участников рынка порождает хаотическую динамику цен вследствие спорадического использования инвесторами информационного потока и, как следствие, возникновение квазистохастических временных интервалов их действия на рынках.

Представленная монография состоит из четырех глав и может быть рекомендована для студентов, аспирантов и преподавателей занимающимся проблемами совершенствования методов и моделей прогнозирования.

ГЛАВА 1 АНАЛИТИЧЕСКИЙ ОБЗОР МОДЕЛЕЙ И МЕТОДОВ ПРОГНОЗИРОВАНИЯ

Бурное развитие прогностики как науки в последние десятилетия привело к созданию множества методов, процедур, приемов прогнозирования, далеко не равноценных по своему значению. Отсутствие их четкой систематизации приводит к дальнейшему расширению инструментария прогностики зачастую малоценными и компилятивными методами.

По оценкам зарубежных и отечественных систематиков прогностики, уже насчитывается свыше 100 методов прогнозирования. Число базовых методов прогностики, которые в тех или иных вариациях повторяются в других методах, гораздо меньше. Многие из этих «методов» относятся скорее к отдельным приемам или процедурам прогнозирования, другие представляют набор отдельных приемов, отличающихся от базовых или друг от друга количеством частных приемов и последовательностью их применения. О проблеме классификации было отмечено во введении.

В литературе имеется большое количество классификационных схем методов прогнозирования. Однако большинство из них или неприемлемы, или обладают недостаточной познавательной ценностью. Основной погрешностью существующих классификационных схем является нарушение принципов классификации. К числу основных таких принципов, на наш взгляд, относятся: достаточная полнота охвата прогностических методов, единство классификационного признака на каждом уровне членения, открытость классификационной. Предлагаемая нами схема классификации методов прогнозирования показана на рисунке 1 [19].

Безусловно, имеют право на существование частные классификационные схемы, предназначенные для определенной цели или задачи.

Каждый уровень детализации (членения) определяется своим классификационным признаком: степенью формализации, общим принципом действия, способом получения прогнозной информации.

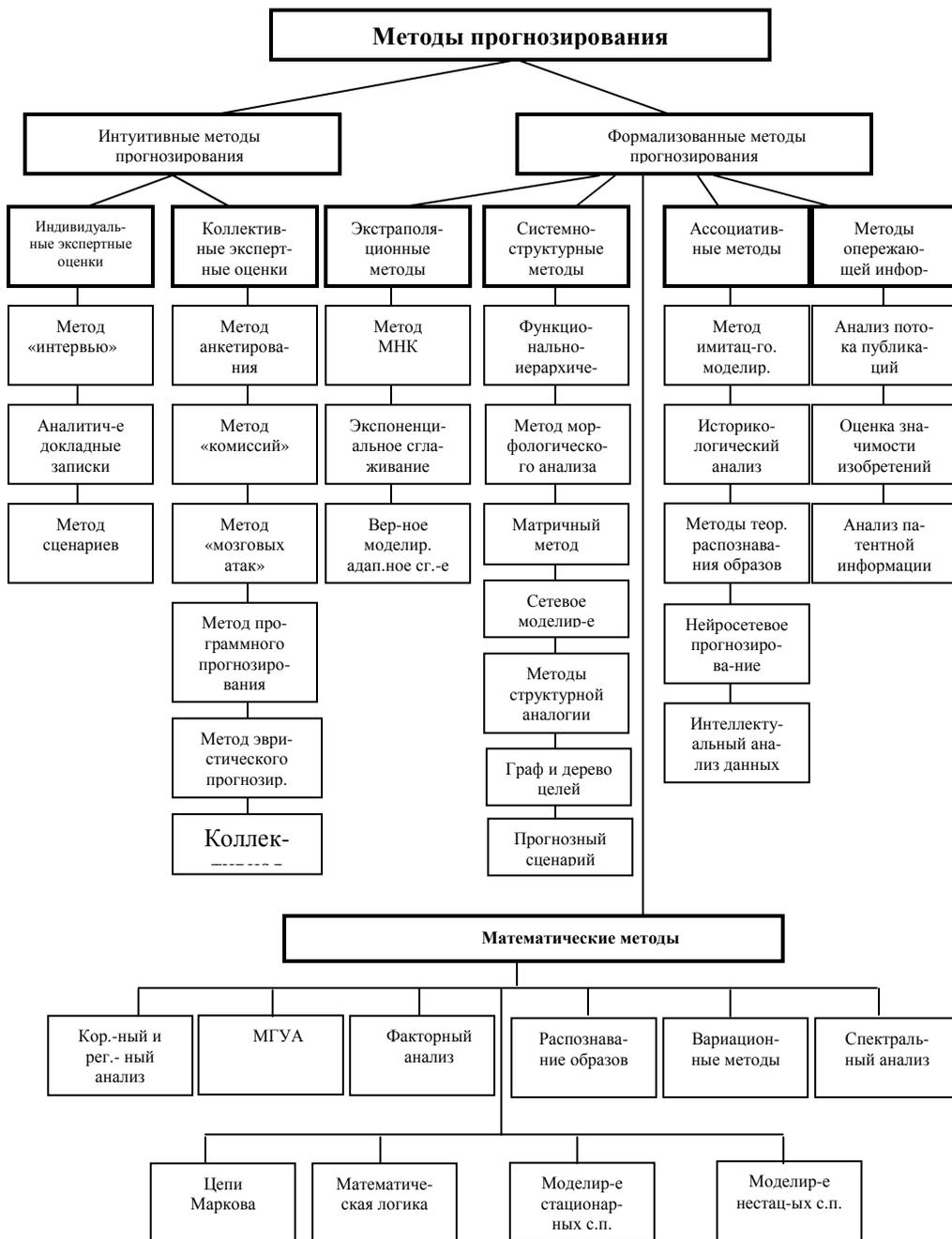


Рис. 1.1 Классификационная схема методов прогнозирования

По степени формализации все методы прогнозирования делятся на интуитивные и формализованные. Интуитивное прогнозирование применяется тогда, когда объект прогнозирования либо слишком прост, либо настолько сложен, что аналитически учесть влияние многих факторов практически невозможно. В этих случаях прибегают к опросу экспертов, Полученные индивидуальные и коллективные экспертные оценки используют как конечные прогнозы или в качестве исходных данных в комплексных системах прогнозирования.

В выборе методов прогнозирования, комплекслируемых в систему, важным показателем является глубина упреждения прогноза. При этом необходимо не только знать абсолютную величину этого показателя, но и отнести его к длительности эволюционного цикла развития объекта прогнозирования. Для этого можно использовать предложенный В. Белоконом безразмерный показатель глубины (дальности) прогнозирования (τ)

$$\tau = \Delta t / t_x ,$$

где Δt – абсолютное время упреждения; t_x – величина эволюционного цикла объекта прогнозирования.

Формализованные методы прогнозирования являются действенными, если величина глубины упреждения укладывается в рамки эволюционного цикла ($\tau \ll 1$). При возникновении в рамках прогнозного периода «скачка» в развитии объекта прогнозирования ($\tau \approx 1$) необходимо использовать интуитивные методы как для определения силы «скачка», так и для оценки времени его осуществления. В этом случае формализованные методы применяются для оценки эволюционных участков развития до и после скачка. Если же в прогножном периоде укладывается несколько эволюционных циклов развития объекта прогнозирования ($\tau \gg 1$), то при комплексировании систем прогнозирования большее значение имеют интуитивные методы.

В зависимости от общих принципов действия интуитивные методы прогнозирования, например, можно разделить на две группы: индивидуальные экспертные оценки и коллективные экспертные оценки.

Методы коллективных экспертных оценок уже можно отнести к комплексным системам прогнозирования (обычно неполным), поскольку в последних сочетаются методы индивидуальных экспер-

ных оценок и статистические методы обработки этих оценок. Но так как статистические методы применяются во вспомогательных процедурах выработки прогнозной информации, на наш взгляд, коллективные экспертные оценки целесообразнее отнести к сингулярным методам прогнозирования.

В группу индивидуальных экспертных оценок можно включить (принцип классификации – способ получения прогнозной информации) следующие методы: метод «интервью», аналитические докладные записки, написание сценария. В группу коллективных экспертных оценок входят анкетирование, методы «комиссий», «мозговых атак» (коллективной генерации идей).

Класс формализованных методов в зависимости от общих принципов действия можно разделить на группы экстраполяционных, системно-структурных, ассоциативных методов и методов опережающей информации.

В группу методов прогнозной экстраполяции можно включить методы наименьших квадратов, экспоненциального сглаживания, вероятностного моделирования и адаптивного сглаживания. К группе системно-структурных методов – отнести методы функционально-иерархического моделирования, морфологического анализа, матричный, сетевого моделирования, структурной аналогии. Ассоциативные методы можно разделить на методы имитационного моделирования и историко-логического анализа. В группу методов опережающей информации – включить методы анализа потоков публикаций, оценки значимости изобретений и анализа патентной информации.

Представленный перечень методов и их групп не является исчерпывающим. Нижние уровни классификации открыты для внесения новых элементов, которые могут появиться в процессе дальнейшего развития инструментария прогностики.

Некоторые не названные здесь методы являются или разновидностью включенных в схему методов, или дальнейшей их конкретизацией.

1.1 Прогнозная экстраполяция

В методическом плане основным инструментом любого прогноза является схема экстраполяции. Различают формальную и прогнозную экстраполяцию. Формальная базируется на предположении

о сохранении в будущем прошлых и настоящих тенденций развития объекта прогноза. При прогнозной экстраполяции фактическое развитие увязывается с гипотезами о динамике исследуемого процесса с учетом в перспективе его физической и логической сущности.

Основу экстраполяционных методов прогнозирования составляет изучение временных рядов, представляющих собой упорядоченные во времени наборы измерений тех или иных характеристик исследуемого объекта, процесса.

Временной ряд y_t может быть представлен в следующем виде

$$y_t = x_t + \varepsilon_t, \quad (1.1)$$

где x_t – детерминированная неслучайная компонента процесса; ε_t – стохастическая компонента процесса.

Если детерминированная компонента (тренд) x_t характеризует существующую динамику развития процесса в целом, то стохастическая компонента ε_t отражает случайные колебания или шумы процесса. Обе составляющие процесса определяются каким-либо функциональным механизмом, характеризующим их поведение во времени. Задача прогноза состоит в определении вида экстраполирующих функций x_t и ε_t на основе исходных эмпирических данных.

Первым этапом экстраполяции тренда является выбор оптимального вида функции, описывающей эмпирический ряд. Для этого проводятся предварительная обработка и преобразование исходных данных с целью облегчения выбора вида тренда путем сглаживания и выравнивания временного ряда, определения функций дифференциального роста, а также формального и логического анализа особенностей процесса. Следующим этапом является расчет параметров выбранной экстраполяционной функции.

Наиболее распространенными методами оценки параметров зависимостей являются метод наименьших квадратов и его модификации, метод экспоненциального сглаживания, метод вероятностного моделирования и метод адаптивного сглаживания.

1.1.1 Метод наименьших квадратов

Сущность метода наименьших квадратов состоит в отыскании параметров модели тренда, минимизирующих ее отклонение от точек исходного временного ряда, т. е.

$$S = \sum_{i=1}^n (\hat{y}_i - y_i)^2 \rightarrow \min \quad (1.2)$$

где \hat{y}_i – расчетные значения исходного ряда; y_i – фактические значения исходного ряда; n – число наблюдений. Если модель тренда представить в виде

$$\hat{y} = f(x_i; a_1, a_2, \dots, a_k, t), \quad (1.3)$$

где a_1, a_2, \dots, a_k – параметры модели; t – время; x_i – независимые переменные, то для того, чтобы найти параметры модели, удовлетворяющие условию (1), необходимо приравнять нулю первые производные величины S по каждому из коэффициентов a_j . Решая полученную систему уравнений с k неизвестными, находим значения коэффициентов a_j .

Использование процедуры оценки, основанной на методе наименьших квадратов, предполагает обязательное удовлетворение целого ряда предпосылок, невыполнение которых может привести к значительным ошибкам.

1. Случайные ошибки имеют нулевую среднюю, конечные дисперсии и ковариации.

2. Каждое измерение случайной ошибки характеризуется нулевым средним, не зависящим от значений наблюдаемых переменных.

3. Дисперсии каждой случайной ошибки одинаковы, их величины независимы от значений наблюдаемых переменных (гомоскедастичность).

4. Отсутствие автокорреляции ошибок, т. е. значения ошибок различных наблюдений независимы друг от друга.

5. Нормальность. Случайные ошибки имеют нормальное распределение.

6. Значения эндогенной переменной x свободны от ошибок измерения и имеют конечные средние значения и дисперсии.

В практических исследованиях в качестве модели тренда в основном используют следующие функции: линейную $y = ax + b$; квадратичную $y = ax^2 + bx + c$; степенную $y = ax^n$; показательную $y = a^x$; экспоненциальную $y = ae^x$; логистическую $y = \frac{a}{1 + be^{-cx}}$.

Особенно широко применяется линейная, или линеаризуемая, т. е. сводимая к линейной, форма как наиболее простая и в достаточной степени удовлетворяющая исходным данным.

Выбор модели в каждом конкретном случае осуществляется по целому ряду статистических критериев, например по дисперсии, корреляционному отношению и др. Следует отметить, что названные критерии являются критериями аппроксимации, а не прогноза. Однако, принимая во внимание принятую гипотезу об устойчивости процесса в будущем, можно предполагать, что в этих условиях модель, наиболее удачная для аппроксимации, будет наилучшей и для прогноза.

Классический метод наименьших квадратов предполагает равноценность исходной информации в модели. В реальной же практике будущее поведение процесса значительно в большей степени определяется поздними наблюдениями, чем ранними. Это обстоятельство породило так называемое дисконтирование, т. е. уменьшение ценности более ранней информации. Дисконтирование можно учесть путем введения в модель (1) некоторых весов $\beta < 1$. Тогда

$$S = \sum_{i=1}^n \beta_i (\hat{y}_i - y_i)^2 \rightarrow \min \quad (1.4)$$

Коэффициенты β , могут задаваться заранее в числовой форме или в виде функциональной зависимости таким образом, чтобы по мере продвижения в прошлое веса убывали, например $\beta_i = a^i$, где $a < 1$. К сожалению, формальных процедур выбора параметра не разработано, и он выбирается исследователем произвольно.

Метод наименьших квадратов широко применяется для получения конкретных прогнозов, что объясняется его простотой и легкостью реализации на ЭВМ. Недостаток метода состоит в том, что модель тренда жестко фиксируется и с помощью МНК можно получить надежный прогноз на небольшой период упреждения. Поэтому МНК относится главным образом к методам краткосрочного прогнозиро-

вания. Кроме того, существенной трудностью МНК является правильный выбор вида модели, а также обоснование и выбор весов во взвешенном методе наименьших квадратов.

1.1.2 Метод экспоненциального сглаживания

Весьма эффективным и надежным методом прогнозирования является экспоненциальное сглаживание. Основные достоинства метода состоят в возможности учета весов исходной информации, в простоте вычислительных операций, в гибкости описания различных динамик процессов. Метод экспоненциального сглаживания дает возможность получить оценку параметров тренда, характеризующих не средний уровень процесса, а тенденцию, сложившуюся к моменту последнего наблюдения. Наибольшее применение метод нашел для реализации среднесрочных прогнозов. Для метода экспоненциального сглаживания основным и наиболее трудным моментом является выбор параметра сглаживания α , начальных условий и степени прогнозирующего полинома [6,64,72,151].

Пусть исходный динамический ряд описывается

$$y_t = a_0 + a_1 t + \frac{a_2}{2} t^2 + \dots + \frac{a_p}{p!} t^p + \varepsilon_t. \quad (1.5)$$

Метод экспоненциального сглаживания, являющийся обобщением метода скользящего среднего, позволяет построить такое описание процесса (1.5), при котором более поздним наблюдениям придается большие веса по сравнению с ранними наблюдениями, причем веса наблюдений убывают по экспоненте. Выражение

$$S_t^{[k]}(y) = \alpha \sum_{i=0}^n (1-\alpha)^i S_{t-i}^{[k]}(y) \quad (1.6)$$

называется экспоненциальной средней k -го порядка для ряда y_t , где α – параметр сглаживания.

В расчетах для определения экспоненциальной средней пользуются рекуррентной формулой [151]

$$S_t^{[k]}(y) = \alpha S_t^{[k-1]}(y) + (1-\alpha) S_{t-1}^{(k)}(y). \quad (1.7)$$

Использование соотношения (1.7) предполагает задание начальных условий $S_0^{[1]}, S_0^{[2]}, \dots, S_0^{[k]}$. Для этого можно воспользоваться формулой Брауна–Мейера, связывающей коэффициенты прогнозирующего полинома с экспоненциальными средними соответствующих порядков

$$S_t^{[k]} = \sum_{p=0}^n (-1)^p \frac{\hat{a}_p}{p!} \frac{\alpha\beta}{(k-1)!} \sum_{j=0}^{\infty} j^p \beta^j \frac{(p-1+j)!}{j!}, \quad (1.8)$$

где $p = 1, 2, \dots, n+1$; \hat{a}_p – оценки коэффициентов; $p = 1 - \alpha$. Можно получить оценки начальных условий, в частности, для линейной модели

$$\begin{aligned} S_0^{[1]} &= a_0 - \frac{\beta}{\alpha} a_1; \\ S_0^{[2]} &= a_0 - \frac{2\beta}{\alpha} a_1; \end{aligned} \quad (1.9)$$

для квадратичной модели –

$$\begin{aligned} S_0^{[1]} &= a_0 - \frac{\beta}{\alpha} a_1 + \frac{\beta(2-\alpha)}{2\alpha^2} a_2; \\ S_0^{[2]} &= a_0 - \frac{2\beta}{\alpha} a_1 + \frac{\beta(3-2\alpha)}{2\alpha^2} a_2; \\ S_0^{[3]} &= a_0 - \frac{3\beta}{\alpha} a_1 + \frac{\beta(4-3\alpha)}{2\alpha^2} a_2. \end{aligned} \quad (1.10)$$

Зная начальные условия $S_0^{[k]}$ и значения параметра α , можно вычислить экспоненциальные средние $S_t^{[k]}$.

Оценки коэффициентов прогнозирующего полинома определяются через экспоненциальные средние по фундаментальной теореме Брауна – Мейера. В этом случае коэффициенты и, находясь решением системы $(p+1)$ уравнений с $k(p+1)$ неизвестными, связывающей параметры прогнозирующего полинома с исходной информацией. Так, для линейной модели получим,

$$\hat{a}_0 = 2S_t^{[1]} - S_t^{[2]}; \quad (1.11)$$

$$\hat{a}_1 = \frac{\alpha}{\beta} (S_i^{[1]} - S_i^{[2]});$$

для квадратичной модели –

$$\begin{aligned} \hat{a}_0 &= 3(S_i^{[1]} - S_i^{[2]}) + S_i^{[3]}, \\ \hat{a}_1 &= \frac{\alpha}{\beta^2} [(6 - 5\alpha)S_i^{[1]} - 2(5 - 4\alpha)S_i^{[2]} + (4 - 3\alpha)S_i^{[3]}], \\ \hat{a}_2 &= \frac{\alpha^2}{\beta^2} [S_i^{[1]} - 2S_i^{[2]} + S_i^{[3]}] \end{aligned} \quad (1.12)$$

Прогноз реализуется по выбранному многочлену соответственно для линейной модели $\hat{y}_{t+\tau} = \hat{a}_0 + \hat{a}_1\tau$; для квадратичной модели

$$\hat{y}_{t+\tau} = \hat{a}_0 + \hat{a}_1\tau + \frac{\hat{a}_2}{2}\tau^2, \text{ где } \tau - \text{период прогноза.}$$

Важную роль в методе экспоненциального сглаживания играет выбор оптимального параметра сглаживания α , так как именно он определяет оценки коэффициентов модели, а, следовательно, и результаты прогноза [72, 103, 215].

В зависимости от величины параметра прогнозные оценки по-разному учитывают влияние исходного ряда наблюдений: чем больше α , тем больше вклад последних наблюдений в формирование тренда, а влияние начальных условий быстро убывает. При малом α прогнозные оценки учитывают все наблюдения, при этом уменьшение влияния более «старой» информации происходит медленно.

Известны два основных соотношения, позволяющие найти приближенную оценку α . Первое соотношение Брауна, выведенное из условия равенства скользящей и экспоненциальной средней

$\alpha = \frac{2}{N+1}$, где N – число точек ряда, для которых динамика ряда считается однородной и устойчивой (период сглаживания). Вторым является соотношение Мейера $\alpha \approx \frac{\sigma_n}{\sigma_\varepsilon}$, где σ_n – среднеквадратическая

ошибка модели; σ_ε – среднеквадратическая ошибка исходного ряда. Однако использование последнего соотношения затруднено тем, что достоверно определить σ_n и σ_ε , из исходной информации очень сложно.

Выбор параметра α целесообразно связывать с точностью прогноза, поэтому для более обоснованного выбора α можно использовать процедуру обобщенного сглаживания, которая позволяет получить следующие соотношения, связывающие дисперсию прогноза и параметр сглаживания [64, 103, 129] для линейной модели –

$$\sigma_{\hat{x}_t}^2 = \frac{\alpha}{(1 + \beta)^2} [1 + 4\beta + 2\beta^2 + 2\alpha(1 + 3\beta)\tau + 2\alpha^2\tau^3] \sigma_\varepsilon^2; \quad (1.13)$$

для квадратичной модели –

$$\sigma_{\hat{x}_t}^2 \approx [2\alpha + 3\alpha^3 + 3\alpha^2\tau] \sigma_\varepsilon^2; \quad (1.14)$$

для обобщенной модели вида

$$y(t) = \sum_{i=1}^n a_i f_i(t) + \varepsilon_t \quad (1.15)$$

дисперсия прогноза имеет следующий вид

$$\sigma_{\hat{x}_t}^2 = \sum_{j=1}^n \sum_{k=1}^n f_j(\tau) \text{cov}(a_j, a_k) f_k(\tau) = \bar{f}^T V f(\tau) \sigma_\varepsilon^2, \quad (1.6)$$

где σ_ε – среднеквадратическая ошибка аппроксимации исходного динамического ряда; $f_i(t)$ – некоторая известная функция; V – матрица ковариации коэффициентов модели.

Отличительная особенность этих формул состоит в том, что при $\alpha = 0$ они обращаются в нуль. Это объясняется тем, что, чем ближе к нулю α , тем больше длина исходного ряда наблюдений $t \rightarrow \infty$ и, следовательно, тем меньше ошибка прогноза. Поэтому для уменьшения ошибки прогноза необходимо выбирать минимальное α .

В то же время параметр α определяет начальные условия, и, чем меньше α , тем ниже точность определения начальных условий, а следовательно, ухудшается и качество прогноза. В работе [103] на примере линейной модели показано, что ошибка прогноза растет по мере уменьшения точности определения начальных условий.

Таким образом, использование формул (1.13)-(1.16) приводит к противоречию при определении параметра сглаживания: с уменьшением α уменьшается среднеквадратическая ошибка, но при этом возрастает ошибка в начальных условиях, что в свою очередь влияет на точность прогноза.

Кроме того, при использовании соотношений (1.13)-(1.16) необходимо принимать во внимание следующие обстоятельства, а именно: эти выражения получены для бесконечно длинных рядов без учета автокорреляции наблюдений. На практике мы имеем дело с конечными рядами, характеризующимися внутренней зависимостью между исходными наблюдениями. Все это снижает целесообразность использования соотношений (1.13)-(1.16).

В ряде случаев параметр α выбирается таким образом, чтобы минимизировать ошибку прогноза, рассчитанного по ретроспективной информации.

Весьма существенным для практического использования является вопрос о выборе порядка прогнозирующего полинома, что во многом определяет качество прогноза. В работах [40, 53] показано, что превышение второго порядка модели не приводит к существенному увеличению точности прогноза, но значительно усложняет процедуру расчета.

Рассмотренный метод является одним из наиболее надежных и широко применяется в практике прогнозирования. Одно из наиболее перспективных направлений развития данного метода представляет собой метод разностного прогнозирования, в котором само экспоненциальное сглаживание рассматривается как частный случай [129, 138].

1.1.3 Метод вероятностного моделирования

Прогнозирование с использованием вероятностных моделей базируется на методе экспоненциального сглаживания. Вероятностные модели по своей сути отличны от экстраполяционных моделей временных рядов, в которых основой является описание изменения во времени процесса.

Во временных рядах модели представляют собой некоторую функцию времени с коэффициентами, значения которых оцениваются по наблюдениям. В вероятностных моделях оцениваются вероятности, а не коэффициенты.

Пусть определено n взаимно независимых и исключających событий. В каждом случае наблюдения измеряются в единой шкале, помещаются в $(n + 1)$ ограниченный класс и обозначаются так: x_1, x_2, \dots, x_n . Событие, связанное с наблюдением $x(t)$, соответствует числу интервалов, в которое это событие попадает, т. е. существует единственное значение k , такое, что $x_{k-1} < x(t) \leq x_k$. И поэтому k -е событие связывается с наблюдением $x(t)$.

Рассмотрим метод оценивания вероятностей $\hat{p}_n(t)$, связанных с различными событиями $x_{k-1} < x(t) \leq x_k$. На первом этапе задаются начальные значения различных вероятностей: $\hat{p}_k(0)$; $k \neq 1, 2, \dots, n$. Наблюдение $x(t)$ связано с k -м событием следующим образом: если $x_{k-1} < x(t) \leq x_k$, то строится единичный вектор \vec{u}_k , $(k - 1)$ компонент которого равен 0 и k -й. компонент равен 1. Это может быть k -м столбцом единичной матрицы ранга k .

Процесс, реализующий оценки вероятностей, описывается вектором сглаживания по формуле

$$\hat{p}(t) = \alpha \vec{u}_k + (1 - \alpha) \vec{p}(t - 1). \quad (1.17)$$

Каждая компонента вектора меняется по закону простого экспоненциального сглаживания между нулем и единицей. Если вектор $\vec{p}(t - 1)$ вероятностный, то все его компоненты должны быть неотрицательными и их сумма должна быть равна 1. Значение оценки $\vec{p}_k(t)$ есть результат экспоненциального сглаживания, и если распределение вероятностей наблюдений $x(t)$ не меняется, то получаемые вероятности и будут действительными вероятностями k -го события. Если существует достаточно длительная реализация процесса, то начальные оценки со временем перестанут оказывать влияние (будут достаточно «взвешены»), и вектор сглаживания будет в среднем описывать вероятности и взаимно исключających, и независимых событий. Значения компонент вектора $\vec{u}(t)$ представляют собой выборку с биномиальным распределением, поэтому дисперсии k -й компоненты будут $\vec{p}_k(1 - \vec{p}_k)$. Дисперсия оценок k -и вероятности определяется соотношением

$$\sigma_k^2 = \frac{\alpha}{2-\alpha} \bar{p}_k (1-\bar{p}_k), \quad (1.18)$$

где α – константа сглаживания ($0 \leq \alpha \leq 1$), используемая для получения оценок вектора вероятностей.

Возможны два варианта. В первом варианте пределы классов заданы так, что p_k может быть или очень большим (около 1), или очень малым (около 0). Тогда дисперсия компонент вектора вероятностей будет небольшой. Если форма распределения меняется со временем, большое значение константы сглаживания может быть использовано, чтобы устранить влияние «старой» информации.

Во втором варианте распределение вероятностей постоянно во времени, нет необходимости «взвешивать» старую информацию. Малое значение константы сглаживания, может быть, позволит уменьшить дисперсию оценок. Тогда можно использовать меньшие интервалы классов с не очень большими вероятностями.

Автоматизированные системы прогнозирования требуют постоянного добавления новых значений информации. Некоторые системы могут просто накапливать информацию, затем использовать ее для прогноза. Если мы имеем дело с поступающей информацией, то система может практически бездействовать в течение значительного промежутка времени. Если информация достаточно важна, следует рассматривать ее как непрерывный во времени поток наблюдений или предсказывать распределение поступлений наблюдений. Очевидно, для таких прогнозов следует использовать модель, изложенную выше. Если в какой-то период нет никаких наблюдений, можно перестроить систему на другой вид информации. Кроме того, оценки коэффициентов (или других параметров) в модели прогноза не изменяются, если наблюдения равны нулю; соответственно и прогноз будет тем же [54, 72].

Вероятностная модель оперирует последовательностью наблюдений с учетом их распределения и игнорирует последовательность этой информации уже непосредственно во времени. Поэтому вектор вероятностей $\vec{p}(t)$, который служит текущей оценкой вероятностей п отдельных событий, является оценкой этих вероятностей в будущем. Последовательность наблюдений может быть представлена как временной ряд $x(t)$, где x измерен по некоторой шкале $x_0 \leq x \leq x_n$ а x_0 и x_n есть минимум и максимум возможных значений наблюдений.

Поэтому p -й процентилю будет такое число x_p , что для p процентов времени реализуется условие $x_0 \leq x(t) \leq x_p$, а $(100 - p)$ процентов - $x_p \leq x(t) \leq x_n$. Кумулятивная вероятность того, что наблюдение попадает левее по шкале, для данного класса записывается так

$$P_r\{x \leq x_k\} = P_k = \sum_{i=1}^k P_i. \quad (1.19)$$

Очевидно, $p_0 = 0$ и $p_k = 1, 0$.

Для связи с вероятностью дается несколько иное представление, нежели процентное. Если $p = p_k$ для класса k , то в этом случае [54]

$$\sum_{i=1}^k p_i(t) = \hat{p}_{k-1}(t) < p < \hat{p}_k(t) = \sum_{i=1}^k p_i(t). \quad (1.20)$$

Кумулятивная вероятность $\hat{p}_{k-1}(t)$ для класса x_{k-1} меньше, чем желаемая вероятность, которая в свою очередь меньше, чем кумулятивная вероятность $\hat{p}_k(t)$ для следующего класса. Простейшая оценка необходимой процентиля может быть получена по линейной интерполяции

$$x_p(t) = \frac{[\hat{p}_k(t) - p]x_{k-1} + (p - \hat{p}_{k-1}(t))x_k}{\hat{p}_k(t) - \hat{p}_{k-1}(t)} \quad (1.21)$$

Если классы очень малы (или p_k близко к p_{k-1}), линейная интерполяция достаточно хороша. Возможно и интерполирование по полиномам более высокой степени. Около хвостов распределения можно ожидать, что кумулятивная вероятность ведет себя как

$$p(x) = 1 - j^x \quad \text{или} \quad p(x) = 1 - \delta^{x^2}, \quad (1.22)$$

где j или δ - числа меньше единицы. Такие функции могут быть оценены на основании имеющейся информации.

Определим дисперсию вероятностей модели следующим образом

$$\sigma_x^2 = \bar{x}^2 - x^2, \quad (1.23)$$

где величина \bar{x}^2 может быть оценена экспоненциальным сглаживанием квадратов наблюдений. Можно считать, что наблюде-

ния почти всегда распределены нормально. Тогда вероятностная модель может быть применена непосредственно к этим наблюдениям.

Пусть x – случайная величина с ожиданием m и конечной дисперсией σ . Тогда сумма n случайных выборок x будет нормально распределена со средним nm и дисперсией $n\sigma^2$, и вероятности как суммы точек наблюдений будут распределены нормально.

Пусть случайная величина x распределена между нулем и единицей. Введем функцию

$$f(x) = \begin{cases} 1, & \text{если } 0 \leq x \leq 1, \\ 0, & \text{если } x < 0 \text{ или } x > 1. \end{cases} \quad (1.24)$$

Пусть y_N – сумма N случайных выборок, тогда функция распределения этих сумм будет

$$f_N(y) = \frac{1}{(N-1)!} \left[y^{N-1} - \binom{N}{1}(y-1)^{N-1} + \binom{N}{2}(y-2)^{N-2} + \dots \right] \quad (1.25)$$

где $0 < y < N$. Находим среднее значение и дисперсию для величины y

$$\bar{y} = \frac{N}{2}; \quad \sigma_y^2 = \frac{N}{12}. \quad (1.26)$$

Тогда можно выразить точку y_p через среднюю и дисперсию распределения

$$y_p = \bar{y} + k_p \sigma_y = \frac{N}{2} + k_p \sqrt{\frac{N}{12}}, \quad (1.27)$$

где k_p – некоторый множитель, учитывающий число степеней свободы распределения.

Данное соотношение может служить основой оценок для вероятностной модели. При достаточном количестве исходной информации вероятностная модель может дать вполне надежный прогноз. Кроме того, эта модель отличается большой простотой и наглядностью. Оценки, получаемые с помощью этой модели, имеют вполне конкретный смысл. Недостатком модели является требование большого количества наблюдений и незнание начального распределения, что может привести к неправильным оценкам. Тем не менее при оп-

ределении процедуры начального распределения или с помощью байесовского метода, корректируя его, можно рассматривать вероятностную модель как эффективный метод прогноза.

1.2 Интуитивные (экспертные) методы прогнозирования

Прогнозные экспертные оценки отражают индивидуальное суждение специалистов относительно перспектив развития объекта и основаны на мобилизации профессионального опыта и интуиции. Методы экспертных оценок используются для анализа объектов и проблем, развитие которых либо полностью, либо частично не поддается математической формализации, т. е. для которых трудно разработать адекватную модель. Применяемые в прогнозировании методы экспертной оценки разделяют на индивидуальные и коллективные.

Индивидуальные экспертные методы основаны на использовании мнений экспертов-специалистов соответствующего профиля независимо друг от друга. Наиболее часто применимыми являются следующие два метода формирования прогноза: интервью и аналитические экспертные оценки.

Метод интервью предполагает беседу прогнозиста с экспертом, в ходе которой прогнозист в соответствии с заранее разработанной программой ставит перед экспертом вопросы относительно перспектив развития прогнозируемого объекта. Успех такой оценки в значительной степени зависит от способности интервьюируемого эксперта экспромтом давать заключения по самым различным фундаментальным вопросам. Аналитические экспертные оценки предполагают длительную и тщательную самостоятельную работу эксперта над анализом тенденций, оценкой состояния и путей развития прогнозируемого объекта. Этот метод дает возможность эксперту использовать всю необходимую ему информацию об объекте прогноза. Свои соображения эксперт оформляет в виде докладной записки.

Основными преимуществами рассматриваемых методов являются возможность максимального использования индивидуальных способностей эксперта и незначительность психологического давления, оказываемого на отдельного работника. Однако эти методы мало пригодны для прогнозирования наиболее общих стратегий из-за ограниченности знаний одного специалиста-эксперта о развитии смежных областей науки.

Методы коллективных экспертных оценок основываются на принципах выявления коллективного мнения экспертов о перспективах развития объекта прогнозирования. В основе применения этих методов лежит гипотеза о наличии у экспертов умения с достаточной степенью достоверности оценить важность и значение исследуемой проблемы, перспективность развития определенного направления исследований, времени свершения того или иного события, целесообразности выбора одного из альтернативных путей развития объекта прогноза и т. д. В настоящее время широкое распространение получили экспертные методы, основанные на работе специальных комиссий, когда группы экспертов за круглым столом обсуждают ту или иную проблему с целью согласования мнений и выработки единого мнения. Этот метод имеет недостаток, заключающийся в том, что группа экспертов в своих суждениях руководствуется в основном логикой компромисса.

В свою очередь в **методе Дельфи** вместо коллективного обсуждения той или иной проблемы проводится индивидуальный опрос экспертов обычно в форме анкет для выяснения относительной важности и сроков свершения гипотетических событий. Затем производится статистическая обработка анкет и формируется коллективное мнение группы, выявляются, обобщаются аргументы в пользу различных суждений. Вся информация сообщается экспертам. Участников экспертизы просят пересмотреть оценки и объяснить причины своего несогласия с коллективным суждением. Эта процедура повторяется 3–4 раза. В результате происходит сужение диапазона оценок. Недостатком этого метода является невозможность учета влияния, оказываемого на экспертов организаторами опросов при составлении анкет.

Как правило, основными задачами при формировании прогноза с помощью коллектива экспертов являются: формирование репрезентативной экспертной группы, подготовка и проведение экспертизы, статистическая обработка полученных документов.

При формировании группы экспертов основными являются вопросы определения ее качественного и количественного состава. Отбор экспертов начинается с определения вопросов, которые охватывают решение данной проблемы; затем составляется список лиц, компетентных в этих областях.

Для получения качественного прогноза к участникам экспертизы предъявляется ряд требований, основными из которых являются:

высокий уровень общей эрудиции; глубокие специальные знания в оцениваемой области; способность к адекватному отображению тенденции развития исследуемого объекта; наличие психологической установки на будущее; наличие академического научного интереса к оцениваемому вопросу при отсутствии практической заинтересованности специалиста в этой области; наличие производственного и (или) исследовательского опыта в рассматриваемой области.

Для определения соответствия потенциального эксперта перечисленным требованиям используется анкетный опрос. Дополнительно к этому часто используют способ самооценки компетентности эксперта. При самооценке эксперт определяет степень своей осведомленности в исследуемом вопросе также на основании анкеты. Обработка данных дает возможность получить количественную оценку компетентности потенциального эксперта по следующей формуле

$$K = 0,5 \left(\frac{\sum_{j=3}^m V_j}{\sum_{j=1}^m V_{j_{\max}}} + \frac{\lambda}{P} \right), \quad (1.28)$$

где V_j – вес градации, перечеркнутой экспертом по j -й характеристике в анкете в баллах; $V_{j_{\max}}$ – максимальный вес (предел шкалы) j -й характеристики в баллах; t – общее количество характеристик компетентности в анкете; λ – вес ячейки, перечеркнутой экспертом в шкале самооценки в баллах; p – предел шкалы самооценки эксперта в баллах.

Установить оптимальную численность группы экспертов довольно трудно. Однако в настоящее время разработан ряд формализованных подходов к этому вопросу. Один из них основан на установлении максимальной и минимальной границ численности группы. При этом исходят из двух условий: высокой средней компетентности групп экспертов и стабилизации средней оценки прогнозируемой характеристики.

Первое условие используется для определения максимальной

численности группы экспертов n_{\max} : $CK_{\max} \leq \frac{\sum_{i=1}^n k_i}{n_{\max}}$, где C - константа;

K_{\max} – максимально возможная компетентность по используемой

шкале компетентности; K_i – компетентность i -го эксперта. Это условие предполагает, что если имеется группа экспертов, компетентность которых максимальна, то среднее значение их оценок можно считать «истинным». Для определения константы используется практика голосования, т. е. группа считается избранной, если за нее подано $\frac{2}{3}$ голосов присутствующих. Исходя из этого, принимается, что $C = \frac{2}{3}$. Таким образом, максимальная численность экспертной группы устанавливается на основании неравенства

$$n_{\max} \leq \frac{3 \sum_{i=1}^n K_i}{2k_{\max}}. \quad (1.29)$$

Далее определяется минимальная численность экспертной группы n_{\min} . Это осуществляется посредством использования условия стабилизации средней оценки прогнозируемой характеристики, которое формулируется следующим образом: включение или исключение эксперта из группы незначительно влияет на среднюю оценку прогнозируемой величины

$$\frac{B - B'}{B_{\max}} < \varepsilon, \quad (1.30)$$

где B – средняя оценка прогнозируемой величины в баллах, данная экспертной группой; B' – средняя оценка, данная экспертной группой, из которой исключен (или в которую включен) один эксперт; B_{\max} – максимально возможная оценка прогнозируемой величины в принятой шкале оценок; ε – заданная величина изменения средней ошибки при включении или исключении эксперта.

Величина средней оценки наиболее чувствительна к оценке эксперта, обладающего наибольшей компетентностью и поставившего наибольший балл при $B \leq B_{\max}$ и минимальный – при $B \geq B_{\max} / 2$. Поэтому для проверки выполнения условия (1.30) предлагается исключить из группы одного эксперта.

В литературе приводится правило расчета минимального числа экспертов в группе в зависимости от заданной (допустимой) величины изменения средней оценки ε

$$n_{\min} = 0,5 \left(\frac{3}{\varepsilon} + 5 \right). \quad (1.31)$$

Таким образом, правила (1.29)-(1.31) дают возможность получить оценочные значения максимального и минимального числа экспертов в группе. Окончательная численность экспертной группы формируется на основании последовательного исключения малокомпетентных экспертов, при этом используется условие $(K_{\max} - K_i) \leq \eta$, где η – задаваемая величина границы допустимого отклонения компетентности 1-го эксперта от максимальной. Одновременно могут включаться в группу новые эксперты. Численность группы устанавливается в пределах $n_{\min} \leq n \leq n_{\max}$.

Кроме описанных выше процедур в методах коллективных экспертных оценок используется подробный статистический анализ экспертных заключений, в результате которого определяются качественные характеристики группы экспертов. В соответствии с этими характеристиками в процессе проведения экспертизы качественный и количественный состав экспертной группы может корректироваться.

Подготовка к проведению экспертного опроса включает разработку анкет, содержащих набор вопросов по объекту прогноза. Структурно-организационный набор вопросов в анкете должен быть логически связан с центральной задачей экспертизы. Хотя форма и содержание вопросов определяются спецификой объекта прогнозирования, можно установить общие требования к ним: вопросы должны быть сформулированы в общепринятых терминах, их формулировка должна исключать всякую смысловую неопределенность, все вопросы должны логически соответствовать структуре объекта прогноза, обеспечивать единственное толкование.

По форме вопросы могут быть открытыми и закрытыми, прямыми и косвенными. Вопрос называют открытым, если ответ на него не регламентирован. Закрытыми считаются вопросы, в формулировке которых содержатся альтернативные варианты ответов, и эксперт должен остановить свой выбор на одном (или нескольких) из них. Косвенные вопросы используют в тех случаях, когда требуется замаскировать цель экспертизы. К подобным вопросам прибегают тогда, когда не уверены, что эксперт, давая информацию, будет вполне искренен или свободен от посторонних влияний, искажающих объективность ответа. Рассмотрим основные группы вопросов, используемых при проведении коллективной экспертной оценки:

- вопросы, предполагающие ответы в виде количественной оценки: о времени свершения событий, о вероятности свершения событий, об оценке относительного влияния факторов. При определении шкалы значений количественных характеристик целесообразно пользоваться неравномерной шкалой. Выбор конкретного масштаба неравномерности определяется характером зависимости ошибки прогноза от периода упреждения;

- вопросы, требующие содержательного ответа в свернутой форме: дизъюнктивные, конъюнктивные, имплицативны;

- вопросы, требующие содержательного ответа в развернутой форме: в виде перечня сведений об объекте; в виде перечня аргументов, подтверждающих или отвергающих тезис, содержащийся в вопросе.

Эти вопросы формируются в два этапа. На первом этапе экспертам предлагается сформулировать наиболее перспективные и наименее разработанные проблемы. На втором – из названных проблем выбираются принципиально разрешимые и имеющие непосредственное отношение к объекту прогноза.

Процедура проведения экспертизы может быть различной, однако здесь также можно выделить три основных этапа. На первом этапе эксперты привлекаются для уточнения формализованной модели объекта прогноза, формулировки вопросов в анкетах, уточнения состава группы. На втором этапе осуществляется непосредственная работа экспертов над вопросами в анкетах. На третьем этапе после предварительной обработки результатов прогноза эксперты привлекаются для консультаций по недостающей информации, необходимой для окончательного формирования прогноза.

При статистической обработке результатов экспертных оценок в виде количественных данных, содержащихся в анкетах, определяются статистические оценки прогнозируемых характеристик и их доверительные границы, статистические оценки согласованности мнений экспертов.

Среднее значение прогнозируемой величины определяется по формуле

$$B = \sum_{i=1}^n B_i / n, \quad (1.32)$$

где B_i - значение прогнозируемой величины, данное i -м экспертом; n – число экспертов в группе.

Кроме того, определяется дисперсия $D = \left[\sum_{i=1}^n (B_i - B)^2 / (n-1) \right]$ и

приближенное значение доверительного интервала $j = t \sqrt{\frac{D}{n-1}}$, где t –

параметр, определяемый по таблицам Стьюдента для заданного уровня доверительной вероятности и числа степеней свободы $k = (n - 2)$.

Доверительные границы для значения прогнозируемой величины вычисляются по формулам: для верхней границы $A_B = B + j$, для нижней границы $A_H = B - j$.

Коэффициент вариации оценок, данных экспертами, определяется по зависимости $v = \frac{\sigma}{B}$, где σ – среднеквадратическое отклонение.

При обработке результатов экспертных оценок по относительной важности направлений среднее значение, дисперсия и коэффициент вариации вычисляются для каждого оцениваемого направления. Кроме того, вычисляется коэффициент конкордации, показывающий степень согласованности мнений экспертов по важности каждого из оцениваемых направлений, и коэффициенты парной ранговой корреляции, определяющие степень согласованности экспертов друг с другом.

Для этого производится ранжирование оценок важности, данных экспертами. Каждая оценка, данная i -м экспертом, выражается числом натурального ряда таким образом, что число 1 присваивается максимальной оценке, а число n – минимальной. Если все оценки различны, то соответствующие числа натурального ряда есть ранги оценок i -го эксперта. Если среди оценок, данных i -м экспертом, есть одинаковые, то этим оценкам назначается одинаковый ранг, равный средней арифметической соответствующих чисел натурального ряда.

Сумма рангов S_j назначенных экспертами направлению j ($j = 1, \dots, t$; x – число исследуемых направлений), определяется по формуле

$$S_j = \sum_{i=1}^n R_{ij}, \quad (1.33)$$

где R_{ij} – ранг оценки, данной i -м экспертом j -му направлению. Среднее значение суммы рангов оценок по всем направлениям равно $\bar{S} = \sum_{j=1}^m S_j / m$. Отклонение суммы рангов, полученных j -м направлением, от среднего значения суммы рангов определяется как $d_j = S_j - \bar{S}$. Тогда коэффициент конкордации, вычисленный по совокупности всех направлений, составляет

$$W = \frac{12 \sum_{j=1}^m d_j^2}{n^2(m^3 - m) - n \sum_{i=1}^n T_i}. \quad (1.34)$$

Величина $T_i = \sum_{i=1}^n t_i^3 - t_e$, рассчитывается при наличии равных рангов (n - количество групп равных рангов, t_e – количество равных рангов в группе).

Коэффициент конкордации принимает значение в пределах от 0 до 1. $W=1$ означает полную согласованность мнений экспертов, при $W=0$ - полную несогласованность. Коэффициент конкордации показывает степень согласованности всей экспертной группы. Низкое значение этого коэффициента может быть получено как при отсутствии общности мнений всех экспертов, так и из-за наличия противоположных мнений между подгруппами экспертов, хотя внутри подгруппы согласованность может быть высокой.

Для выявления степени согласованности мнений экспертов используется коэффициент парной ранговой корреляции

$$\rho_{i,i+1} = \frac{\sum_{j=1}^m \psi_j^2}{\frac{1}{\sigma} (m^3 - m) - \frac{1}{12} (\pi - T_j - 1)}, \quad (1.35)$$

где ψ_j – разность (по модулю) величин рангов оценок j -го направления, назначенных экспертами i и $i + 1$,

$$\psi_j = |R_i - R_{i+1}|. \quad (1.36)$$

Коэффициент парной ранговой корреляции может принимать значения от +1 до -1. Значение $\rho = 1$ соответствует полной согласованности мнений двух экспертов. Значение $\rho = -1$ показывает, что мнение одного эксперта противоположно мнению другого.

Для определения уровня значимости значений коэффициентов W и $\rho_{i,i+1}$ можно использовать критерий χ^2 . Для этого вычисляется величина

$$\chi^2 = \frac{12 \sum_{j=1}^m \alpha_j^2}{m \cdot n(m+1) - \frac{1}{m-1} \sum_{i=1}^n T_i} \quad (1.37)$$

(число степеней свободы $k = t - 1$) и по соответствующим таблицам определяется уровень значимости полученных значений.

1.3 Корреляционный и регрессионный анализы

Одним из наиболее распространенных способов получения многофакторных прогнозов является упоминавшийся ранее классический метод наименьших квадратов и построение на его основе модели множественной регрессии [16]. Для линейного случая модель множественной регрессии записывается в виде

$$y_j = \sum_{i=1}^n \alpha_i x_{ij} + \varepsilon_j \quad (1.38)$$

где α_i - коэффициенты модели; y_j, x_{ij} - соответственно значения j -й функции (зависимой переменной) и i -й независимой переменной; $i = \overline{0, n}$; $j = \overline{1, N}$, ε_j - случайная ошибка; n - число независимых переменных в модели (в ряде случаев полагается, что α_i - свободный член и $x_{0j} = 1$).

В векторном виде эта модель записывается так [131]

$$Y = X\alpha + \varepsilon, \quad (1.39)$$

где вектор $Y^T = (y_1, y_2, \dots, y_n)$ и вектор $\alpha^T = (\alpha_1, \alpha_2, \dots, \alpha_n)$ - соответственно векторы значений зависимой переменной и коэффициентов модели, матрица порядка $(n \times N)$;

$$X = \begin{bmatrix} x_{11}, x_{12}, \dots, x_{1n} \\ \dots\dots\dots \\ x_{N1}, x_{N2}, \dots, x_{Nn} \end{bmatrix} - \text{матрица независимых переменных};$$

$\varepsilon^T = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ – вектор случайных ошибок.

Неизвестные коэффициенты модели находятся из условия минимума функционала рассогласований, который представляет собой сумму квадратов рассогласований реальных значений зависимой переменной и значений. В векторном виде функционал рассогласований записывается как.

$$\Phi = (Y - X\alpha)^T (Y - X\alpha) \rightarrow \min.$$

Условия минимума Φ реализуются при равенстве нулю первых производных функционала по неизвестным коэффициентам, т. е.

$$\frac{\partial \Phi}{\partial \alpha_i} = 0; i = \overline{0, n} \tag{1.41}$$

Данное условие эквивалентно выполнению векторного соотношения $X^T X \alpha = X^T Y$, что дает значения оценок коэффициентов модели

$$\hat{\alpha} = (X^T X)^{-1} X^T Y. \tag{1.42}$$

Надежность получаемой с помощью оценок $\hat{\alpha}$ модели определяется с помощью величины остаточной дисперсии, которая вычисляется по формуле

$$\sigma^2 = \frac{\varepsilon^T \varepsilon}{N - n} = \frac{1}{N - n} [Y Y^T - Y^T X (X^T X)^{-1} X^T Y] \tag{1.43}$$

и коэффициента множественной корреляции, вычисляемого по формуле

$$R = \sqrt{1 - \frac{D}{D_{(n-1)(n-1)}}} \quad (1.44)$$

$D_{(n+1)(n+1)}$ - алгебраическое дополнение определителя корреляционной матрицы $r = r[x_i, x_j] (i, j = \overline{0, n})$ к элементу $r_{x_{n-1}x_{n+1}}$. Величина R^2 – множественный коэффициент детерминации; она показывает, какая доля дисперсии функции объясняется изменениями входящих в уравнение регрессии независимых переменных при полученных значениях коэффициентов модели. Надежность коэффициента множественной корреляции определяется по критерию Фишера

$$F = \frac{R^2(N - n - 1)}{(1 - R^2)(n - 1)} \quad (1.45)$$

при заданном уровне надежности и степени свободы $V_1 = n$, $v_1 = N - n$. Наличие связи между зависимой переменной y_i и независимыми переменными x_{ij} определяется с помощью коэффициентов корреляции

$$r_{yxi} = \frac{S_{yxi}^2}{\sqrt{S_{yy}^2 S_{xixi}^2}} \quad (1.46)$$

где $S_{yxi}^2; S_{yy}^2; S_{xixk}^2 \dots$, - коэффициенты ковариации, определяемые по формуле

$$S_{xixk}^2 = \frac{1}{N - 1} \sum_{j=1}^N (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k), \quad (1.47)$$

где \bar{x}_i, \bar{x}_k - средние значения независимых переменных x_i и x_k . Доверительные интервалы Z_1, Z_2 ; для коэффициента корреляции определяются как

$$Z_{1,2} = \operatorname{arctg} \pm \frac{U \alpha / 2}{\sqrt{N - 3}} - \frac{r}{2(N - 1)}, \quad (1.48)$$

где $U \alpha / 2 - \frac{\alpha}{2} - r$ – процентиль нормального распределения $N(0, 1)$ с нулевым средним и единичной дисперсией, z - преобразование Фишера, определяемое по формуле [16]

$$z_{xixk} = \frac{1}{2} \ln \frac{1 + r_{xixk}}{1 - r_{xixk}}, \quad (1.49)$$

th z - гиперболический тангенс аргумента z , вычисленный по формуле

$$thz = \frac{e^z - e^{-z}}{e^z + e^{-z}}. \quad (1.50)$$

Истинное значение коэффициента корреляции заключено в пределах $th z_1 \leq r_{xixk} \leq th z_2$. Для определения надежности оценок строится доверительный интервал для полученных оценок $\hat{\alpha}$ коэффициентов модели

$$|\hat{\alpha} - \alpha_i| \leq t_{p,v} \sigma \sqrt{c_{ii}}, \quad (1.51)$$

где $t_{p,v}$ - значения критерия Стьюдента с уровнем надежности p и степенями свободы $\nu = (N - n - 1)$; $\sigma = \sqrt{\sigma^2}$ (σ^2 определяется по формуле (6)), c_{ii} - i -й диагональный элемент матрицы $(X^T X)^{-1}$. Поэтому истинное значение коэффициента α модели будет лежать в интервале

$$\hat{\alpha}_i - t_{p,v} \sigma \sqrt{c_{ii}} < \alpha_i < \hat{\alpha}_i + t_{p,v} \sigma \sqrt{c_{ii}} \quad (1.52)$$

Использование вычислительной процедуры по методу наименьших квадратов с целью получения оценок коэффициентов модели $\hat{\alpha}_i (i = \overline{0, n})$, которые удовлетворяли бы условиям несмещенности, состоятельности, эффективности, предполагает выполнение ряда условий. Рассмотрим эти условия:

- независимые переменные представляют собой неслучайный набор чисел, их средние значения и дисперсии конечны;
- случайные ошибки ε_j - имеют нулевую среднюю и конечную дисперсию

$$M(\varepsilon_j) = 0; M(\varepsilon_j \varepsilon_j) = \sigma_\varepsilon^2 < \infty \quad (1.53)$$

- между независимыми переменными отсутствует корреляция и автокорреляция;
- случайная ошибка не коррелирована с независимыми переменными;
- случайная ошибка подчинена нормальному закону распределения.

Кроме того, можно выделить условие отсутствия мультиколлинеарности, когда несколько независимых переменных связаны между собой линейной зависимостью, и условие гомоскедастичности, т. е. одинаковой дисперсии для всех случайных ошибок. Важным является условие линейной формы связи между зависимой и независимой переменными. Зависимость должна быть именно линейной или сводимой к линейной с помощью некоторых преобразований.

Но иногда исследуемый процесс не может быть сведен к линейной зависимости никакими преобразованиями, как, например, в случае логистической зависимости. Тогда используется ряд методов, например метод симплексов. Данный метод отличается сравнительной простотой, легкой реализуемостью на ЭВМ, эффективностью при определении оценок коэффициентов модели.

Важной характеристикой реализованной модели является оценка ошибки прогноза. Так, в [55] предлагается следующая оценка дисперсии прогноза

$$\sigma_{t+\tau}^2 = \sigma^2 \left[1 + X_{t+\tau}^T (X^T X)^{-1} X_{t+\tau} \right], \quad (1.54)$$

где $X_{t+\tau}$ - вектор значений независимых переменных в момент $(t + \tau)$. Поэтому доверительный интервал для значений зависимой переменной определяется в момент t как

$$\left\{ Y_t \pm t_{p,v} \sigma_t \sqrt{I + \frac{I}{t-1} X_t^T (X_t^T X_t)^{-1} X_t} \right\} \quad (1.55)$$

где I - единичный столбец; $t_{p,v}$ - значение критерия Стьюдента. В [14] находится более эффективная оценка доверительного интервала для прогнозных значений

$$\left[Y_{t+\tau} \pm t_{p,v} \sigma_{t+\tau} \sqrt{I + \frac{1-r_1}{t+\tau-1} \frac{X_{t+\tau}^T X_{t+\tau}}{X_t^T X_t}} \right]. \quad (1.56)$$

Важным условием получения надежных оценок для модели по методу наименьших квадратов является отсутствие автокорреляции. Оценка автокорреляции для полученной по МНК модели осуществляется по критерию Дарбина – Уотсона [73, 16]

$$d = \frac{\sum_{t=1}^{T-1} (\varepsilon_{t+1} - \varepsilon_t)}{\sum_{t=1}^T \varepsilon_t^2}, \quad (1.57)$$

где T - длина временного ряда.

Полученное расчетное значение d сравнивается с нижней и верхней границей d_1 и d_2 , критерия. Если $d < d_1$, то гипотеза отсутствия автокорреляции отвергается; если $d > d_2$, то гипотеза отсутствия автокорреляции принимается; если $d_1 \leq d \leq d_2$, то необходимо дальнейшее исследование. Одним из известных способов уменьшения автокорреляции является авторегрессионное преобразование для исходной информации или переход к разностям, т. е. $\Delta Y_t = Y_{t+1} - Y_t$; $\Delta X_t = X_{t+1} - X_t$. Если же автокорреляцию устранить не удастся, то полученные оценки считаются состоятельными, и среднеквадратическое отклонение корректируется на величину Δ_j для j -го коэффициента.

$$\Delta_j = \sqrt{\frac{1 + r_1 R_{1j}}{1 - r_1 R_{1j}}}, \quad (1.58)$$

где r_1 - коэффициент автокорреляции случайных слагаемых первого порядка; R_{1j} - коэффициент автокорреляции для j -й независимой переменной первого порядка.

Другим условием, необходимым для получения состоятельных оценок, является отсутствие мультиколлинеарности. Действительно, при наличии мультиколлинеарности определитель квадратной матрицы $[X^T X]$ равен или близок нулю, следовательно, матрица вырождена, и поэтому решения системы нормальных уравнений не существует.

Эффективный подход к определению мультиколлинеарности предполагает следующую последовательность расчетов. Пусть рассматривается уравнение регрессии $y = f(x_1, \dots, x_n)$. Тогда для выявления существования мультиколлинеарности предлагается критерий

$$\chi^2 = - \left[N - 1 - \frac{1}{6}(2n + 5) \right] \lg \left[\tilde{X}^T \tilde{X} \right] \quad (1.59)$$

где $|\tilde{X}^T X|$ - определитель матрицы $[\tilde{X}^T X]$, имеющий асимптотическое распределение Пирсона с $\frac{1}{2}n(n-1)$ степенями свободы. В формуле (16) N – число наблюдений по каждому переменному, n – число независимых переменных, матрица X включает значения переменных, преобразованных по формуле

$$\tilde{X}_{ik} = \frac{x_{ik} - \bar{x}_i}{S_i \sqrt{N}}, \quad (1.60)$$

где S_i , \bar{x}_i – соответственно оценки среднеквадратического отклонения и среднее значение для i -й независимой переменной. Далее вычисляются величин

$$d_{ii} = \frac{(\tilde{X}^T \tilde{X})_{ii}}{|\tilde{X}^T \tilde{X}|}, \quad (1.61)$$

которые при неколлинеарности переменных близки единице, а при наличии мультиколлинеарности близки к бесконечности, что дает основание оставить или отбросить показатель x_i , что определяется статистикой $w_i = (d_{ii} - 1) \frac{N - n}{n - 1}$, имеет распределение Фишера с

$v_1 = N - n$ и $v_2 = n - 1$ степенями свободы. Существует еще ряд способов определения мультиколлинеарности, В целях устранения или уменьшения ее можно переходить к разностям для исходной информации, использовать метод факторного анализа или метод главных компонент.

Получение прогнозов с помощью многофакторных регрессионных моделей предполагает неизменность значений коэффициентов этих моделей во времени. Тем не менее в процессе исследования объекта возможно появление новой информации, что позволяет с помощью рекуррентного оценивания корректировать значения оценок коэффициентов моделей. В то же время исходная информация может содержать в себе различные динамики изменения независимых переменных, которые возникают в результате различных «режимов» функционирования исследуемого объекта. В этом случае важным является как сам факт установления различия динамик процессов на разных временных интервалах, так и выбор такого интер-

вала для построения на нем модели прогнозирования, который был бы наиболее адекватным будущему поведению объекта. Если оказывается, что для одного интервала времени построена многофакторная модель $y_1 = \sum_{i=1}^m \alpha_{i1} x_i$, а для другого интервала - модель $y_2 = \sum_{i=1}^m \alpha_{i2} x_i$, где $\alpha_{i1} \neq \alpha_{i2}$, то прогноз будет смещен, а следовательно, резко возрастает дисперсия прогноза.

Построение адекватных регрессионных моделей для целей прогнозирования с помощью метода наименьших квадратов предъявляет к исходной информации весьма жесткие требования. В ряде случаев эти требования для реальных наблюдений оказываются невыполненными, поэтому получаемые оценки оказываются неэффективными, а прогноз – недостоверным. Действительно, требование нормальности распределения ошибок, предъявляемое к исходной информации процедурой метода наименьших квадратов, в большом числе случаев оказывается невыполненным. Так, говорится: «Нормальность – это миф. В реальном мире никогда не было и никогда не будет нормального распределения». Поэтому в последнее время интенсивно разрабатывается новое направление в статистике – так называемая робастная статистика, задача которой в том и состоит, чтобы получать эффективные оценки в случаях невыполнения некоторых предпосылок, например нормальности распределения, наличия аномальных наблюдений. Использование робастных методов получения статистических оценок для информации, содержащей аномальные «выбранные» наблюдения, позволяет значительно повысить надежность получаемых оценок по сравнению с обычным методом наименьших квадратов.

1.4 Модели стационарных временных рядов и их идентификация

Здесь рассматривается набор линейных параметрических моделей. Речь здесь идет не о моделировании временных рядов, а о моделировании их случайных остатков ϵ_t , получающихся после элиминирования из исходного временного ряда x_t его неслучайной составляющей. Следовательно, в отличие от прогноза, основанного на регрессионной модели, игнорирующего значения случайных остатков, в

прогнозе временных рядов существенно используется взаимозависимость и прогноз самих случайных остатков.

Введем обозначения. Так как здесь описывается поведение случайных остатков, то моделируемый временной ряд обозначим ε_t , и будем полагать, что при всех t его математическое ожидание равно нулю, т.е. $E\varepsilon_t = 0$. Временные последовательности, образующие «белый шум», обозначим δ_t .

Описание и анализ, рассматриваемых ниже моделей, формулируется в терминах общего линейного процесса, представимого в виде взвешенной суммы настоящего и прошлых значений белого шума, а именно

$$\varepsilon_t = \sum_{k=0}^{\infty} \beta_k \delta_{t-k}, \quad (1.62)$$

где $\beta_0 = 1$ и $\sum_{k=0}^{\infty} \beta_k^2 < \infty$.

Таким образом, белый шум представляет собой серию импульсов, в широком классе реальных ситуаций генерирующих случайные остатки исследуемого временного ряда.

Временной ряд ε_t можно представить в эквивалентном (1.62) виде, при котором он получается в виде классической линейной модели множественной регрессии, в которой в качестве объясняющих переменных выступают его собственные значения во все прошлые моменты времени

$$\varepsilon_t = \sum_{k=1}^{\infty} \pi_k \varepsilon_{t-k} + \delta_t. \quad (1.63)$$

При этом весовые коэффициенты π_1, π_2, \dots связаны определенными условиями, обеспечивающими стационарность ряда ε_t . Переход от (2) к (1) осуществляется с помощью последовательной подстановки в правую часть (2) вместо $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$ их выражений, вычисленных в соответствии с (2) для моментов времени $t-1, t-2$ и т.д.

Рассмотрим также процесс смешанного типа, в котором присутствуют как авторегрессионные члены самого процесса, так и скользящее суммирование элементов белого шума

$$\varepsilon_t = \sum_{k=1}^p \pi_k \varepsilon_{t-k} + \delta_t + \sum_{j=1}^q \beta_j \delta_{t-j}. \quad (1.64)$$

Будем подразумевать, что p и q могут принимать и бесконечные значения, а также то, что в частных случаях некоторые (или даже все) коэффициенты π или β равны нулю.

1.4.1 Модели авторегрессии порядка p (AR(p)-модели)

Рассмотрим сначала простейшие частные случаи.

Модель авторегрессии 1-го порядка – AR(1) (марковский процесс). Эта модель представляет собой простейший вариант авторегрессионного процесса типа (1.63), когда все коэффициенты кроме первого равны нулю. Соответственно, она может быть определена выражением

$$\varepsilon_t = \alpha \varepsilon_{t-1} + \delta_t, \quad (1.65)$$

где α – некоторый числовой коэффициент, не превосходящий по абсолютной величине единицу ($|\alpha| < 1$), а δ_t – последовательность случайных величин, образующая белый шум. При этом ε_t зависит от δ_t и всех предшествующих δ , но не зависит от будущих значений δ . Соответственно, в уравнении (1.65) δ_t не зависит от ε_{t-1} и более ранних значений ε . В связи с этим, δ_t называют инновацией (обновлением).

Последовательности ε , удовлетворяющие соотношению (1.65), часто называют также марковскими процессами.

Модели авторегрессии 2-го порядка – AR(2) (процессы Юла). Эта модель, как и AR(1), представляет собой частный случай авторегрессионного процесса, когда все коэффициенты π_j в правой части (1.63) кроме первых двух, равны нулю. Соответственно, она может быть определена выражением

$$\varepsilon_t = \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \delta_t, \quad (1.66)$$

где последовательность $\delta_1, \delta_2, \dots$ образует белый шум.

Условия стационарности ряда (1.66) (необходимые и достаточные) определяются как

$$\left\{ \begin{array}{l} |\alpha_1| < 2, \\ \alpha_2 < 1 - |\alpha_1|. \end{array} \right. \quad (1.67)$$

В рамках общей теории моделей те же самые условия стационарности получаются из требования, чтобы все корни соответствующего характеристического уравнения лежали бы вне единичного круга. Характеристическое уравнение для модели авторегрессии 2-го порядка имеет вид: $1 - \alpha_1 z - \alpha_2 z^2 = 0$.

Автокорреляционная функция процесса Юла подсчитывается следующим образом. Два первых значения $r(1)$ и $r(2)$ определены соотношениями

$$\begin{aligned} r(1) &= \frac{\alpha_1}{1 - \alpha_2}, \\ r(2) &= \alpha_2 + \frac{\alpha_1^2}{1 - \alpha_2}, \end{aligned} \quad (1.68)$$

а значения для $r(\tau)$, $\tau = 3, 4, \dots$ вычисляются с помощью рекуррентного соотношения

$$r(\tau) = \alpha_1 r(\tau - 1) + \alpha_2 r(\tau - 2). \quad (1.69)$$

Модели авторегрессии p -го порядка – $AR(p)$ ($p \geq 3$). Эти модели, образуя подмножество в классе общих линейных моделей, сами составляют достаточно широкий класс моделей. Если в общей линейной модели (1.63) полагать все параметры π_j , кроме первых p коэффициентов, равными нулю, то мы приходим к определению $AR(p)$ -модели

$$\varepsilon_t = \sum_{j=1}^p \alpha_j \varepsilon_{t-j} + \delta_t, \quad (1.70)$$

где последовательность случайных величин $\delta_1, \delta_2, \dots$ образует белый шум.

Условия стационарности процесса, генерируемого моделью (1.70), также формулируются в терминах корней его характеристического уравнения

$$1 - \alpha_1 z - \alpha_2 z^2 - \dots - \alpha_p z^p = 0. \quad (1.71)$$

Для стационарности процесса необходимо и достаточно, чтобы все корни характеристического уравнения лежали бы вне единичного круга, т.е. превосходили бы по модулю единицу.

1.4.2 Модели скользящего среднего порядка q (МА(q)-модели)

Рассмотрим частный случай общего линейного процесса (1.62), когда только первые q из весовых коэффициентов β_j ненулевые. В это случае процесс имеет вид

$$\varepsilon_t = \delta_t - \theta_1 \delta_{t-1} - \theta_2 \delta_{t-2} - \dots - \theta_q \delta_{t-q}, \quad (1.72)$$

где символы $-\theta_1, \dots, \theta_q$ используются для обозначения конечного набора параметров β , участвующих в (1.62). Процесс (1.72) называется моделью скользящего среднего порядка q (МА(q)).

Двойственность в представлении AR- и МА-моделей и понятие обратимости МА-модели. Из (1.62) и (1.63) видно, что один и тот же общий линейный процесс может быть представлен либо в виде AR-модели бесконечного порядка, либо в виде МА-модели бесконечного порядка.

Соотношение (1.72) может быть переписано в виде

$$\delta_t = \varepsilon_t + \theta_1 \delta_{t-1} + \theta_2 \delta_{t-2} + \dots + \theta_q \delta_{t-q}. \quad (1.73)$$

Откуда

$$\delta_t = \varepsilon_t - \pi_1 \varepsilon_{t-1} - \pi_2 \varepsilon_{t-2} - \dots, \quad (1.74)$$

где коэффициенты π_j ($j = 1, 2, \dots$) определенным образом выражаются через параметры $\theta_1, \dots, \theta_q$. Соотношение (1.74) может быть записано в виде модели авторегрессии бесконечного порядка (т.е. в виде обращенного разложения)

$$\varepsilon_t = \sum_{j=1}^{\infty} \pi_j \varepsilon_{t-j} + \delta_t. \quad (1.75)$$

Известно (см., например, [15]), что условие обратимости МА(q)-модели (т.е. условие сходимости ряда $\sum_{j=1}^{\infty} \pi_j$) формулируется в терминах характеристического уравнения модели (1.75) следующим образом.

Все корни характеристического уравнения $1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_q z^q = 0$ должны лежать вне единичного круга, т.е. $|z_j| > 1$ для всех $j = 1, 2, \dots, q$.

Взаимосвязь процессов AR(q) и MA(q). Сделаем ряд замечаний о взаимосвязях между процессами авторегрессии и скользящего среднего.

Для конечного процесса авторегрессии порядка p δ_t может быть представлено как конечная взвешенная сумма предшествующих ε_t , или ε_t может быть представлено как бесконечная сумма предшествующих δ_t . В то же время, в конечном процессе скользящего среднего порядка q ε_t может быть представлено как конечная взвешенная сумма предшествующих δ_t или δ_t – как бесконечная взвешенная сумма предшествующих ε_t .

Конечный процесс MA имеет автокорреляционную функцию, обращающуюся в нуль после некоторой точки, но так как он эквивалентен бесконечному процессу AR, его частная автокорреляционная функция бесконечно протяженная. Главную роль в ней играют затухающие экспоненты и (или) затухающие синусоиды. И наоборот, процесс AR имеет частную автокорреляционную функцию, обращающуюся в нуль после некоторой точки, но его автокорреляционная функция имеет бесконечную протяженность и состоит из совокупности затухающих экспонент и или затухающих синусоид.

Параметры процесса авторегрессии конечного порядка не должны удовлетворять каким-нибудь условиям для того, чтобы процесс был стационарным. Однако для того чтобы процесс MA был обратимым, корни его характеристического уравнения должны лежать вне единичного круга.

Спектр процесса скользящего среднего является обратным к спектру соответствующего процесса авторегрессии.

1.4.3 Авторегрессионные модели со скользящими средними в остатках (ARMA(p, q)-модели)

Представление процесса типа MA в виде процесса авторегрессии неэкономично с точки зрения его параметризации. Аналогично процесс AR не может быть экономично представлен с помощью модели скользящего среднего. Поэтому для получения экономичной параметризации иногда бывает целесообразно включить в модель как

члены, описывающие авторегрессию, так и члены, моделирующие остаток в виде скользящего среднего. Такие линейные процессы имеют вид

$$\varepsilon_t = \alpha_1 \varepsilon_{t-1} + \dots + \alpha_p \varepsilon_{t-p} + \delta_t - \theta_1 \delta_{t-1} - \dots - \theta_q \delta_{t-q} \quad (1.76)$$

и называются процессами авторегрессии – скользящего среднего порядка (p, q) (ARMA(p, q)).

Стационарность и обратимость ARMA(p, q)-процессов. Записывая процесс (1.76) в виде

$$\varepsilon_t = \sum_{j=1}^p \alpha_j \varepsilon_{t-j} + \bar{\delta}_{qt}, \quad (1.77)$$

где $\bar{\delta}_{qt} = \delta_t - \theta_1 \delta_{t-1} - \dots - \theta_q \delta_{t-q}$, можно провести анализ стационарности (8) по той же схеме, что и для AR(p)-процессов. При этом различие “остатков” $\bar{\delta}_{qt}$ и δ_t никак не повлияет на выводы, определяющие условия стационарности процесса авторегрессии. Поэтому процесс (1.74) является стационарным тогда и только тогда, когда все корни характеристического уравнения AR(p)-процесса лежат вне единичного круга.

1.5 Модели нестационарных временных рядов и их идентификация

1.5.1 Модель авторегрессии-проинтегрированного скользящего среднего (ARIMA(p, k, q)-модель)

Эта модель предложена Дж. Боксом и Г. Дженкинсом [15]. Она предназначена для описания нестационарных временных рядов x_t , обладающих следующими свойствами:

- анализируемый временной ряд аддитивно включает в себя составляющую $f(t)$, имеющую вид алгебраического полинома (от параметра времени t) некоторой степени $k > 1$; при этом коэффициенты этого полинома могут быть как стохастической, так и нестохастической природы;

- ряд $x_t^k, t = 1, \dots, T - k$, получившийся из x_t после применения к нему k -кратной процедуры метода последовательных разностей, может быть описан моделью ARMA(p, q).

Это означает, что ARIMA(p, k, p)-модель анализируемого процесса x_t , может быть записана в виде

$$x_t^k = \alpha_1 x_{t-1}^k + \alpha_2 x_{t-2}^k + \dots + \alpha_p x_{t-p}^k + \delta - \theta_1 \delta_{t-1} - \dots - \theta_q \delta_{t-q}, \quad (1.78)$$

где

$$x_t^k = \Delta^k x_t = x_t - C_k^1 x_{t-1} + C_k^2 x_{t-2} - \dots + (-1)^k x_{t-k}.$$

Заметим, что классу моделей ARIMA принадлежит и простейшая модель стохастического тренда – процесс случайного блуждания (или просто случайное блуждание). Случайное блуждание определяется аналогично процессу авторегрессии первого порядка (1.63), но только у случайного блуждания $\alpha = 1$, так что

$$\varepsilon_t = \varepsilon_{t-1} + \delta_t. \quad (1.79)$$

Ряд первых разностей случайного блуждания δ_t представляет собой белый шум, т.е. процесс ARMA(0, 0). Поэтому само случайное блуждание входит в класс моделей ARIMA как модель ARIMA(0, 1, 0).

Идентификация ARIMA-моделей. В первую очередь, следует подобрать порядок k модели. Первый тип критерия подбора основан на отслеживании поведения величины $\hat{\sigma}^2(k)$ в зависимости от k : в качестве верхней оценки для порядка k определяется то значение k_0 , начиная с которого тенденция к убыванию $\hat{\sigma}^2(k)$ гасится и само значение $\hat{\sigma}^2(k)$ относительно стабилизируется. Второй тип критерия подбора порядка k ARIMA-модели основан на анализе поведения автокорреляционных функций процессов $\Delta x_t, \Delta^2 x_t, \dots$ последовательные преобразования анализируемого процесса x_t с помощью операторов Δ, Δ^2, \dots нацелены на устранение его нестационарности. Поэтому до тех пор, пока $1 < k$ процессы $\Delta^k x_t$ будут оставаться нестационарными, что будет выражаться в отсутствии быстрого спада в поведении их выборочной автокорреляционной функции. Поэтому предполагается, что необходимая для получения стационарности степень k разности Δ достигнута, если автокорреляционная функция ряда $x_t^k = \Delta^k x_t$ быстро затухает.

После подбора порядка k анализируется уже не сам ряд x_t , а его k -е разности. Идентификация этого ряда сводится к идентификации ARMA(p, q)моделей.

1.5.2 Модели рядов, содержащих сезонную компоненту

Под временными рядами, содержащими сезонную компоненту, понимаются процессы, при формировании значений которых обязательно присутствовали сезонные и/или циклические факторы.

Один из распространенных подходов к прогнозированию состоит в следующем: ряд раскладывается на долговременную, сезонную (в том числе, циклическую) и случайную составляющие; затем долговременную составляющую подгоняют полиномом, сезонную – рядом Фурье, после чего прогноз осуществляется экстраполяцией этих подогнанных значений в будущее. Однако этот подход может приводить к серьезным ошибкам. Во-первых, короткие участки стационарного ряда (а в экономических приложениях редко бывают достаточно длинные временные ряды) могут выглядеть похожими на фрагменты полиномиальных или гармонических функций, что приведет к их неправомерной аппроксимации и представлению в качестве неслучайной составляющей. Во-вторых, даже если ряд действительно включает неслучайные полиномиальные и гармонические компоненты, их формальная аппроксимация может потребовать слишком большого числа параметров, т.е. получающаяся параметризация модели оказывается неэкономичной.

Принципиально другой подход основан на модификации ARIMA-моделей с помощью «упрощающих операторов». Схематично процедура построения сезонных моделей, основанных на ARIMA-конструкциях, модифицированных с помощью упрощающих операторов $\nabla_T = 1 - F^T$, может быть описана следующим образом (детальное описание соответствующих процедур см., например, в [15]:

- применяем к наблюдаемому ряду x_t операторы Δ и ∇_T для достижения стационарности;
- по виду автокорреляционной функции преобразованного ряда $x_{k,K}^{(r)}(t)$ подбираем пробную модель в классе ARMA- или модифицированных (в правой части) ARMA-моделей;
- по значениям соответствующих автоковариаций ряда $x_{k,K}^{(r)}(t)$ получаем (методом моментов) оценки параметров пробной модели;

Диагностическая проверка полученной модели (анализ остатков в описании реального ряда x_t с помощью построенной модели) может либо подтвердить правильность модели, либо указать пути ее улучшения, что приводит к новой подгонке и повторению всей процедуры.

1.5.3 Прогнозирование на базе ARIMA-моделей

ARIMA-модели охватывают достаточно широкий спектр временных рядов, а небольшие модификации этих моделей позволяют весьма точно описывать и временные ряды с сезонностью. Начнем обсуждение проблемы прогнозирования временных рядов с методов, основанных на использовании ARIMA-моделей. Мы говорим об ARIMA-моделях, имея в виду, что сюда входят как частные случаи AR-, MA- и ARMA-модели. Кроме того, будем исходить из того, что уже осуществлен подбор подходящей модели для анализируемого временного ряда, включая идентификацию этой модели. Поэтому в дальнейшем предполагается, что все параметры модели уже оценены.

Будем прогнозировать неизвестное значение x_{t+1} , $l \geq 1$ полагая, что x_t – последнее по времени наблюдение анализируемого временного ряда, имеющееся в нашем распоряжении. Обозначим такой прогноз \hat{x}_t^l .

Заметим, что хотя \hat{x}_t^l и \hat{x}_{t-1}^{l+1} обозначают прогноз одного и того же неизвестного значения x_{t+1} , но вычисляются они по-разному, т.к. являются решениями разных задач.

Ряд x_τ , анализируемый в рамках ARIMA(p, k, q)-модели, представим (при любом $\tau > k$) в виде

$$(1 - \alpha_1 L - \dots - \alpha_p L^p) \sum_{j=0}^k (-1)^j C_k^j x_{\tau-j} = \delta_\tau - \theta_1 \delta_{\tau-1} - \dots - \theta_q \delta_{\tau-q}, \quad (1.80)$$

где L – оператор сдвига функции времени на один временной такт назад.

Из соотношения (1.80) можно выразить x_τ для любого $\tau = t - q, \dots, t - 1, t, t + 1, \dots, t + l$. Получаем

$$x_t = \left(\sum_{j=1}^p \alpha_j L^j \right) x_t + \left(1 - \sum_{j=1}^p \alpha_j L^j \right) \left(\sum_{i=0}^k (-1)^i C_k^i x_{t-i} \right) + \left(1 - \sum_{j=1}^q \theta_j L^j \right) \delta_t. \quad (1.81)$$

Правые части этих соотношений представляют собой линейные комбинации $p + k$ предшествующих (по отношению к левой части) значений анализируемого процесса x_t , дополненные линейными комбинациями текущего и q предшествующих значений случайных остатков δ_t . Причем коэффициенты, с помощью которых эти линейные комбинации подсчитываются, известны, т.к. выражаются в терминах уже оцененных параметров модели.

Этот факт и дает возможность использовать соотношения (1.81) для построения прогнозных значений анализируемого временного ряда на l тактов времени вперед. Теоретическую базу такого подхода к прогнозированию обеспечивает известный результат, в соответствии с которым наилучшим (в смысле среднеквадратической ошибки) линейным прогнозом в момент времени t с упреждением l является условное математическое ожидание случайной величины x_{t+l} , вычисленное при условии, что все значения x_t до момента времени t . Этот результат является частным случаем общей теории прогнозирования (см. [237, 198, 235]).

Условное математическое ожидание $E(x_{t+l} | x_1, \dots, x_t)$ получается применением операции усреднения к обеим частям (10) при $\tau = t + l$ с учетом следующих соотношений:

$$E(x_{t-j} | x_1, \dots, x_t) = x_{t-j} \quad \text{при всех } j = 0, 1, 2, \dots, t-1 \quad (1.82)$$

$$E(x_{t+j} | x_1, \dots, x_t) = \hat{x}_t^j \quad \text{при всех } j = 1, 2, \dots; \quad (1.83)$$

$$E(x_{t+j} | x_1, \dots, x_t) = 0 \quad \text{при всех } j = 1, 2, \dots; \quad (1.84)$$

$$E(x_{t-j} | x_1, \dots, x_t) = x_{t-j} - \hat{x}_{t-j-1}^1 \quad \text{при всех } j = 0, 1, 2, \dots, t-1. \quad (1.85)$$

Таким образом, определяется следующая процедура построения прогноза по известной до момента траектории временного ряда: по формулам (1.81) вычисляются ретроспективные прогнозы \hat{x}_{t-q-1}^1 , $\hat{x}_{t-q}^1, \dots, \hat{x}_{t-1}^1$ по предыдущим значениям временного ряда; при этом при вычислении начальных прогнозных значений $\hat{x}_{t-q+m-1}^1$ для x_{t-q+m} ($m = 0, 1, \dots$) по формулам (1.81) вместо условных средних $E(\delta_{t-q+m-j} |$

x_1, \dots, x_{t-q+m}), которые в общем случае следовало бы вычислять по формулам (1.85), подставляются их безусловные значения, равные нулю; используя формулы для $\tau > t$ и правила (1.82)–(1.85) подсчитываются условные математические ожидания для вычисления прогнозных значений.

Описанная процедура выглядит достаточно сложной. Однако при реалистичных значениях параметров p , q и k эта процедура в действительности оказывается весьма простой.

1.6 Адаптивные методы прогнозирования

Считается, что характерной чертой адаптивных методов прогнозирования является их способность непрерывно учитывать эволюцию динамических характеристик изучаемых процессов, «подстраиваться» под эту эволюцию, придавая, в частности, тем больший вес и тем более высокую информационную ценность имеющимся наблюдениям, чем ближе они к текущему моменту прогнозирования. Однако деление методов и моделей на «адаптивные» и «неадаптивные» достаточно условно. В известном смысле любой метод прогнозирования адаптивный, т.к. все они учитывают вновь поступающую информацию, в том числе наблюдения сделанные с момента последнего прогноза. Общее значение термина заключается, по-видимому, в том, что «адаптивное» прогнозирование позволяет обновлять прогнозы с минимальной задержкой и с помощью относительно несложных математических процедур. Однако это не означает, что в любой ситуации адаптивные методы эффективнее тех, которые традиционно не относятся к таковым.

Методы экспоненциального сглаживания [151]. Простейший вариант метода уже рассматривался в связи с задачей выявления неслучайной составляющей анализируемого временного ряда. Постановка задачи прогнозирования с использованием простейшего варианта метода экспоненциального сглаживания формулируется следующим образом.

Пусть анализируемый временной ряд x_t , $t = 1, 2, \dots$, представлен в виде

$$x_t = a_0 + \varepsilon_t, \quad (1.86)$$

где a_0 – неизвестный параметр, не зависящий от времени, а ε_τ – случайный остаток со средним значением, равным нулю, и конечной дисперсией. Как известно, экспоненциально взвешенная скользящая средняя ряда x_τ в точке t $\bar{x}_t(\lambda)$ с параметром сглаживания (параметром адаптации) λ ($0 < \lambda < 1$) определяется формулой

$$\bar{x}_t(\lambda) = \frac{1-\lambda}{1-\lambda^t} \sum_{j=0}^{t-1} \lambda^j x_{t-j}, \quad (1.87)$$

которая дает решение задачи: $\bar{x}_t(\lambda) = \arg \min_a \sum_{j=0}^{t-1} \lambda^j (x_{t-j} - a)^2$.

Коэффициент сглаживания λ можно интерпретировать также как коэффициент дисконтирования, характеризующий меру обесценивания наблюдения за единицу времени.

Для рядов с «бесконечным прошлым» формула (1.87) сводится к виду

$$\bar{x}_t(\lambda) = (1-\lambda) \sum_{j=0}^{\infty} \lambda^j x_{t-j}. \quad (1.88)$$

В соответствии с простейшим вариантом метода экспоненциального сглаживания прогноз \hat{x}_t^1 для неизвестного значения x_{t+1} по известной до момента времени t траектории ряда x_t строится по формуле

$$\hat{x}_t^1 = \bar{x}_t(\lambda), \quad (1.89)$$

где значение $\bar{x}_t(\lambda)$ определено формулой (1.87) или (1.88), соответственно для короткого или длинного временного ряда.

Формула (1.89) удобна, в частности, тем, что при появлении следующего $(t + 1)$ -го наблюдения x_{t+1} пересчет прогнозирующей функции $\hat{x}_{t+1}^1 = \bar{x}_{t+1}(\lambda)$ производится с помощью простого соотношения $\bar{x}_{t+1}(\lambda) = \lambda \bar{x}_t(\lambda) + (1-\lambda)x_{t+1}$.

Метод экспоненциального сглаживания можно обобщить на случай полиномиальной неслучайной составляющей анализируемого временного ряда, т.е. на ситуации, когда вместо (1.86) постулируется

$$x_{t+\tau} = a_0 + a_1\tau + \dots + a_k\tau^k + \varepsilon_\tau, \quad (1.90)$$

где $k \geq 1$. В соотношении (1.90) начальная точка отсчета времени сдвинута в текущий момент времени t , что облегчает дальнейшие вычисления. Соответственно, в схеме простейшего варианта метода прогноз \hat{x}_t^l значения x_{t+1} будет определяться соотношениями (1.90) при $\tau = 1$ и (4)

$$\hat{x}_t^l = \hat{x}_{t+1} = \hat{a}_0^{(k)}(t, \lambda) + \hat{a}_1^{(k)}(t, \lambda) + \dots + \hat{a}_k^{(k)}(t, \lambda), \quad (1.91)$$

где оценки $\hat{a}_j(t, \lambda)$, $j = 0, 1, \dots, k$ получаются как решение оптимизационной задачи

$$\sum_{j=0}^{\infty} \lambda^j (x_{t-j} - a_0 - a_1 j - \dots - a_k j^k)^2 \rightarrow \min_{a_0, a_1, \dots, a_k}. \quad (1.92)$$

Решение задачи (1.92) не представляет принципиальных трудностей.

Рассмотрим еще несколько методов, использующих идеологию экспоненциального сглаживания, которые развивают метод Брауна в различных направлениях.

Метод Хольта. Хольт [195] ослабил ограничения метода Брауна, связанные с его однопараметричностью, введением двух параметров сглаживания λ_1 и λ_2 ($0 < \lambda_1, \lambda_2 < 1$). В его модели прогноз \hat{x}_t^l на l тактов времени в текущий момент t также определяется линейным трендом вида

$$\hat{x}_t^l = \hat{a}_0(t, \lambda_1, \lambda_2) + l\hat{a}_1(t, \lambda_1, \lambda_2), \quad (1.93)$$

где обновление прогнозирующих коэффициентов производится по формулам

$$\begin{aligned} \hat{a}_0(t+1, \lambda_1, \lambda_2) &= \lambda_1 x_t + (1 - \lambda_1)(\hat{a}_0(t, \lambda_1, \lambda_2) + \hat{a}_1(t, \lambda_1, \lambda_2)), \\ \hat{a}_1(t+1, \lambda_1, \lambda_2) &= \lambda_2 (\hat{a}_0(t+1, \lambda_1, \lambda_2) - \hat{a}_1(t, \lambda_1, \lambda_2)) + \\ &+ (1 - \lambda_2)\hat{a}_1(t, \lambda_1, \lambda_2). \end{aligned} \quad (1.94)$$

Таким образом, прогноз по данному методу является функцией прошлых и текущих данных, параметров λ_1 и λ_2 , а также начальных значений $\hat{a}_0(0, \lambda_1, \lambda_2)$ и $\hat{a}_1(0, \lambda_1, \lambda_2)$.

Метод Хольта–Уинтерса. Уинтерс [236] развил метод Хольта так, чтобы он охватывал еще и сезонные эффекты. Прогноз, сделанный в момент t на l тактов времени вперед, равен

$$\hat{x}_t^l = [\hat{a}_0(t) + l\hat{a}_1(t)]\omega_{t+l-N}, \quad (1.95)$$

где ω_τ – коэффициент сезонности, а N – число временных тактов, содержащихся в полном сезонном цикле. Сезонность в этой формуле представлена мультипликативно. Метод использует три параметра сглаживания $\lambda_1, \lambda_2, \lambda_3$ ($0 < \lambda_j < 1, j = 1, 2, 3$), а его формулы обновления имеют вид

$$\begin{aligned} \hat{a}_0(t+1) &= \lambda_1 \frac{x_{t+1}}{\omega_{t+1-N}} + (1 - \lambda_1)[\hat{a}_0(t) + \hat{a}_1(t)], \\ \omega_{t+1} &= \lambda_2 \frac{x_{t+1}}{\hat{a}_0(t+1)} + (1 - \lambda_2)\omega_{t+1-N}, \\ \hat{a}_1(t+1) &= \lambda_3 [\hat{a}_0(t+1) - \hat{a}_0(t)] + (1 - \lambda_3)\hat{a}_1(t). \end{aligned} \quad (1.96)$$

Как и в предыдущем случае, прогноз строится на основании прошлых и текущих значений временного ряда, параметров адаптации λ_1, λ_2 и λ_3 , а также начальных значений $\hat{a}_0(0), \hat{a}_1(0)$ и ω_0 .

Аддитивная модель сезонности Тейла–Вейджа. В экономической практике чаще встречаются экспоненциальные тенденции с мультипликативно наложенной сезонностью. Поэтому перед использованием аддитивной модели члены анализируемого временного ряда обычно заменяют их логарифмами, преобразуя экспоненциальную тенденцию в линейную, а мультипликативную сезонность – в аддитивную. Преимущество аддитивной модели заключается в относительной простоте ее вычислительной реализации. Рассмотрим модель вида (в предположении, что исходные данные прологарифмированы)

$$\begin{aligned} x_\tau &= a_0(\tau) + \omega_\tau + \delta_\tau, \\ a_0(\tau) &= a_0(\tau-1) + a_1(\tau), \end{aligned} \quad (1.97)$$

где $a_0(\tau)$ – уровень процесса после элиминирования сезонных колебаний, $a_1(\tau)$ – аддитивный коэффициент роста, ω_t – аддитивный коэффициент сезонности, δ_t – белый шум.

Прогноз, сделанный в момент t на l временных тактов вперед, подсчитывается по формуле

$$\hat{x}_t^l = \hat{a}_0(t) + l\hat{a}_1(t) + \hat{\omega}_{t-N+l}, \quad (1.98)$$

где коэффициенты \hat{a}_0, \hat{a}_1 и ω вычисляются рекуррентным образом с помощью следующих формул обновления

$$\begin{aligned} \hat{a}_0(\tau) &= \hat{a}_0(\tau-1) + \hat{a}_1(\tau-1) + \lambda_1 [x_\tau - \hat{x}_{\tau-1}^1] \\ \hat{a}_1(\tau) &= \hat{a}_1(\tau-1) + \lambda_1 \lambda_2 [x_\tau - \hat{x}_{\tau-1}^1] \\ \hat{\omega}_t &= \hat{\omega}_{\tau-N} + (1 - \lambda_1) \lambda_3 [x_\tau - \hat{x}_{\tau-1}^1] \end{aligned} \quad (1.99)$$

В этих соотношениях, как и прежде, N – число временных тактов, содержащихся в полном сезонном цикле, а λ_1, λ_2 и λ_3 – параметры адаптации.

1.7 Метод группового учета аргументов

В настоящее время большую популярность для конкретных задач прогнозирования приобретает так называемый метод группового учета аргументов (МГУА), представляющий собой дальнейшее развитие метода регрессионного анализа. Он основан на некоторых принципах теории обучения и самоорганизации, в частности на принципе «селекции», или направленного отбора [52,51].

Метод реализует задачи синтеза оптимальных моделей высокой сложности, адекватной сложности исследуемого объекта (здесь под моделями понимается система регрессионных уравнений). Так, алгоритмы МГУА, построенные по схеме массовой селекции, осуществляют перебор возможных функциональных описаний объекта. При этом полное описание объекта [51]

$$y = f(x_1, x_2, \dots, x_m), \quad (1.100)$$

где f – некоторая функция, например степенной полином, заменяется рядами частных описаний:

$$1\text{-й ряд селекции - } y_1 = f(x_1, x_2), \quad y_2 = f(x_1, x_3), \quad \dots, \quad y_s = f(x_{m-1}, x_m);$$

$$2\text{-й ряд селекции - } z_1 = f(y_1, y_2), \quad z_2 = f(y_1, y_3), \quad \dots, \quad z_k = f(y_{n-1}, y_n).$$

и т. д.

Рассматриваются различные сочетания входных и промежуточных переменных, и для каждого сочетания строится модель, причем при построении рядов селекции используются самые регулярные переменные. Понятие регулярности является одним из основных в методе МГУА. Регулярность определяется минимумом среднеквадратической ошибки переменных на отдельной проверочной последовательности данных (исходный ряд делится на обучающую и проверочную последовательности). В некоторых случаях в качестве показателя регулярности используется коэффициент корреляции. Ряды строятся до тех пор, пока регулярность повышается, т. е. снижается ошибка или увеличивается коэффициент корреляции. Таким образом, из всей совокупности моделей выбирается такая, которая является оптимальной с точки зрения выбранного критерия.

Рассмотрим некоторые алгоритмы МГУА [51].

В алгоритмах с линейными полиномами в качестве частных описаний используются соотношения вида

$$y_k = a_0 + a_1 x_i + a_2 y_i, \quad 0 < i \leq m. \quad (1.101)$$

Алгоритм синтезирует модели с последовательно увеличивающимся числом учитываемых аргументов. Так, модели первого селекционного ряда включают по два аргумента, модели второго ряда – три-четыре и т. д.

Алгоритмы с ковариациями и квадратичными описаниями оперируют с частными описаниями вида

$$y_i = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j; \quad (1.102)$$

$$y_k = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2.$$

В данном случае модели усложняются не только за счет увеличения числа учитываемых аргументов, но и за счет роста степени описания.

В алгоритмах с последовательным выделением трендов в качестве таковых рассматриваются уравнения регрессии по одному аргументу, включая время: $f(x_1)$, $f(x_2)$, ..., $f(x_m)$. Для построения моделей используются частные описания вида

$$y = a_0 + a_1 f(x_1) + a_2 f(x_2) + \dots + a_m f(x_m). \quad (1.103)$$

Алгоритм работает таким образом, что вначале выделяется первый тренд и рассчитывается соответствующее отклонение (первый остаток) истинных значений функции от тренда. После чего это отклонение аппроксимируется вторым трендом и определяется второй остаток и т. д. На практике выделяют до пяти-шести трендов.

Среди основных алгоритмов МГУА наибольший интерес представляет обобщенный алгоритм, обеспечивающий получение наиболее точных моделей благодаря использованию в качестве опорной функции аддитивной и мультипликативной моделей трендов [51].

С целью сокращения числа входных аргументов в обобщенном алгоритме используется рассмотренный выше алгоритм последовательного выделения трендов для выбора оптимальной опорной функции, после чего осуществляется перебор всех возможных комбинаций выделенных трендов, либо в классе сумм, либо в классе произведений. Пусть, например, выбрана зависимость

$$\varphi_{+1} = f(t) + f_1(x_1) + f_2(x_2) + f_3(x_3) + f_4(x_4). \quad (1.104)$$

Обобщенный алгоритм МГУА предусматривает перебор двенадцати комбинаций [51]

$$\begin{aligned}
\varphi_{+1} &= f(t) + f_1(x_1) + f_2(x_2) + f_3(x_3) + f_4(x_4); \\
\varphi_{+1} &= f(t) + f_1(x_1) + f_2(x_2, x_3, x_4); \\
\varphi_{+1} &= f(t) + f_1(x_2) + f_2(x_2, x_3, x_4); \\
\varphi_{+1} &= f(t) + f_1(x_3) + f_2(x_2, x_3, x_4); \\
\varphi_{+1} &= f(t) + f_1(x_4) + f_2(x_2, x_3, x_4); \\
\varphi_{+1} &= f(t) + f_1(x_1) + f_2(x_2) + f_3(x_3, x_4); \\
\varphi_{+1} &= f(t) + f_1(x_1) + f_2(x_3) + f_3(x_2, x_4); \\
\varphi_{+1} &= f(t) + f_1(x_1) + f_2(x_4) + f_3(x_2, x_3); \\
\varphi_{+1} &= f(t) + f_1(x_2) + f_2(x_3) + f_3(x_1, x_4); \\
\varphi_{+1} &= f(t) + f_1(x_2) + f_2(x_4) + f_3(x_1, x_2); \\
\varphi_{+1} &= f(t) + f_1(x_3) + f_2(x_4) + f_3(x_1, x_2); \\
\varphi_{+1} &= f(t) + f_1(x_1, x_2, x_3, x_4).
\end{aligned} \tag{1.05}$$

В результате перебора определяется комбинация, дающая наиболее регулярное решение.

В случаях когда процесс описывается большим числом переменных, использование обобщенного алгоритма затруднено и для сокращения перебора рекомендуется применять алгоритм с многофазной селекцией проекторов (операторов ортогонального проектирования). Понятие проектора введено в МГУА по аналогии с графическим представлением метода наименьших квадратов, согласно которому вектор выходной величины проекта руется на плоскость аргументов. В соответствии с этим все алгоритмы МГУА рассматриваются как варианты последовательного проектирования выходной величины на плоскости переменных на каждом ряду селекции.

Выбирая определенный вид оператора проектирования, можно получить тот или иной алгоритм МГУА. Например, в алгоритме МГУА с последовательным выделением трендов выходная величина у проектируется только на первом ряду, далее проектируется уже остаток $\Delta r = y - f_r$ на оси наиболее эффективной переменной. Аппроксимационная функция имеет вид

$$f_{n+1} = \bar{P}_{j_0}\varphi + \bar{P}_{j_1}\Delta_1 + \dots + \bar{P}_{j_n}\Delta_n, \tag{1.106}$$

где $\overline{P}_{jr} \Delta_r$ – частное описание в виде полинома степени I по одной из наиболее эффективных переменных x_j отбираемых по коэффициенту корреляции: $0 \leq l \leq m$; P_{ij} – оператор ортогонального проектирования на подпространство $L_j = \{x_i^0, x_i^1, \dots, x_i^m\}$, $1 \leq j \leq (m+1)S$, $j = \overline{1, l}$. Оператор P_{ij} можно представить в виде суперпозиции проекторов, соответствующих различным алгоритмам МГУА. Таким образом, выражение для f_{n+1} задает множество алгоритмов МГУА, отличающихся по способу проектирования остатка.

В ряде случаев для упрощения вида частных описаний и простоты определения оценок их коэффициентов используют прием ортогонализации переменных.

Непосредственное использование МГУА для целей прогнозирования основывается на теоремах, изложенных в [52, 51].

1. При любом разделении полного полинома заданной степени на частные полиномы критерий минимума среднеквадратической ошибки, определяемой на обучающей последовательности (первый критерий), позволяет однозначно определить оптимальные оценки всех коэффициентов, если число точек в обучающей последовательности больше числа членов каждого из частных полиномов по крайней мере на единицу.

2. При заданной степени полного полинома имеется много вариантов разбиения его на частные полиномы. Полный перебор всех комбинаций по критерию среднеквадратической ошибки, измеряемой на отдельной проверочной последовательности данных, позволяет найти единственное наилучшее разделение.

3. При постепенном нарастании степени полного полинома до некоторого ограниченного значения ошибка на проверочной последовательности либо непрерывно падает, либо имеет минимум по крайней мере при одном значении степени.

4. Если точки ранжированы по величине дисперсии, то имеется единственное значение отношения числа точек проверочной последовательности к числу точек обучающей последовательности, при котором достигается минимум числа рядов селекции и степени полного полинома.

5. В многорядном процессе алгоритмов МГУА среднеквадратическая ошибка от ряда к ряду не может возрастать независимо от пути, по которому идет селекция.

В качестве критериев получения оптимальной модели по МГУА или критерия регуляризации (точности) используются критерии [51]: $\Delta^2(2) \rightarrow \min$ и $\Delta(1) \rightarrow \min$, где $\Delta(1)$ – среднеквадратическая ошибка на проверочной последовательности данных (первая разность реальных и прогнозных значений); $\Delta^2(2)$ - среднеквадратическая ошибка приращений (вторая разность этих значений). Так, по критерию $\Delta(1) \rightarrow \min$ расчет проводится следующим образом [51]:

1. По заданному уравнению регрессии для периода упреждения прогноза $T_y = 1$ и периода предыстории $T_n = 2$ строится зависимость $x(t)$.

2. Находятся рассогласования $[x^*(t)-x(t)]$ для всех точек проверочной последовательности (N_{np}), $x^*(t)$ - реализации процесса, $x(t)$ – прогнозные значения.

3. Рассчитывается величина ошибки

$$\delta(1) = \frac{1}{N_{np}} \sqrt{\sum_{j=1}^{N_{np}} [x_j^*(t) - x_j(t)]^2}; \quad \Delta(1) = \frac{\delta(1) \cdot 100\%}{\frac{1}{N_{np}} \sqrt{\sum_{j=1}^{N_{np}} [x_j^*(t)]^2}}. \quad (1.107)$$

Расчет по критерию $\Delta^2(2)$ проводится по схеме [51].

1. При $T_y = 1$, $T_n = 2$ при помощи прямого обучения по алгоритму МГУА находится уравнение регрессии для ошибки, например: $\Delta x_0 = f(\Delta x_{-1}, \Delta x_{-2}, x_{-1}, x_{-2})$. Затем строится кривая $\Delta x(t)$ для каждого частного уравнения регрессии.

2. Находятся разности $[\Delta x^*(t) - \Delta x(t)]$ для всех экспериментальных точек.

3. Рассчитывается величина ошибок

$$\delta^2(2) = \frac{1}{N - T_y} \sqrt{\sum_{j=1}^{N - T_y} [\Delta x_j^*(t) - \Delta x_j(t)]^2} \quad (1.108)$$

$$\Delta^2(2) = \frac{\delta^2(2) \cdot 100\%}{\frac{1}{N - T_y} \sum_{j=1}^{N-1} [\Delta x_j^*(t)]^2}. \quad (1.109)$$

Использование критерия $\Delta^2(2)$, как показывает анализ, повышает точность прогноза. Метод МГУА может быть эффективно использован для получения так называемых «системных многократных дифференциальных прогнозов». Под системой в данном случае можно понимать группу определенным образом связанных между собой «входных» и «выходных» показателей с заданным описанием связей, элементов, процессов, структуры.

Для получения прогнозов поведения сложных систем предполагается выполнение определенных условий [52, 51]:

1. Границы системы выбираются таким образом, чтобы можно было исключить лишь наименее важные связи системы с внешней средой.
2. Система включает определенное число переменных M и обратных связей f . Для получения надежного прогноза при построении модели достаточно использовать любые $m \geq M - f$ переменные.
3. Выбранные переменные не должны повторять друг друга.
4. Плохо прогнозируемые переменные следует исключить из модели.

Прогноз называется системным, если одновременно прогнозируются не менее T характеристических переменных системы. Переменные прогнозируются одновременно, шаг за шагом. При этом устраняется один из основных недостатков однократного прогноза – аргументы уравнений прогнозирования «не стареют» (носят последние по времени отсчета индексы).

Многократный прогноз можно вести на основе как алгебраических, так и дифференциальных или интегральных уравнений.

При получении долгосрочных дифференциальных прогнозов важным является установление устойчивости поведения системы. Наиболее распространенным способом установления области устойчивости (для линейных моделей) являются методы Ляпунова, критерии Гурвица – Рауса.

При реализации прогнозов важно установить критерий качества полученных прогнозных результатов. В [53] устанавливается своеобразная иерархия критериев прогноза в зависимости от глуби-

ны прогнозирования. Так, для краткосрочного прогноза в качестве критерия селекции предлагается использовать критерий регулярности – величину среднеквадратической ошибки, определяемой на точках проверочной последовательности, не участвующей в получении оценок коэффициентов. Для среднесрочных прогнозов предлагается использовать критерий несмещенности как более эффективный. При наличии информации об изменении взаимосвязанных переменных появляется возможность использовать критерий, который является одним из наиболее эффективных при долгосрочном прогнозировании, именно критерий баланса переменных, т. е. минимизации суммы квадратов рассогласований самих значений промежуточных переменных и их модельных представлений. Данный критерий определяет «жесткость», неизменность структуры исследуемого объекта.

1.8 Теория распознавания образов

Весьма перспективным в настоящее время является использование для прогнозирования методов теории распознавания образов. Непосредственно с использованием этой теории решается комплекс задач, имеющих важное значение в прогнозировании [4, 24, 25, 67, 75, 105, 126].

Уже использование экстраполяционных методов для прогнозирования временных рядов предполагает однородность динамики. Действительно, исходный временной ряд, являющийся основой прогнозирования какого-либо процесса, может содержать в себе интервалы, внутри которых динамика характеризуется определенными отличными от других интервалов условиями. Естественно, эти интервалы на перспективу искажают полученный прогнозный результат. В этой связи возникает необходимость четкого выделения тех интервалов, для которых характерна однородная динамика. Решение этого вопроса эффективно реализуется с помощью методов теории распознавания образов [126].

Другим важным приложением теории распознавания образов для получения прогнозов является описание и прогнозирование поведения какого-либо объекта по набору признаков, определяющих поведение этого объекта.

Процедура прогнозирования на основе методов распознавания образов состоит в том, что выбираются классы состояний исследуемых объектов, которые могут быть заданы как диапазонами измене-

ния некоторых параметров, так и определенными качественными характеристиками. По совокупности признаков, определяющих состояние объектов, находится соответствие принадлежности каждого нового объекта (или объекта в будущем понятии времени) к определенному классу. Это позволяет дать прогноз состояния объекта или указать диапазон изменения параметров, характеризующих его на прогнозируемый период.

Одной из важнейших проблем, возникающих при получении конкретных прогнозов, является оценка исходной информации. При прогнозировании развития сложной системы возникает ситуация, когда поведение системы может быть описано с помощью многих различных показателей. Реализация прогнозов по всей совокупности этих показателей приводит к необходимости учитывать и взаимосвязи между ними, что подчас бывает весьма затруднительно. Ситуация облегчается, когда для реализации прогнозов используется аппарат распознавания образов и прогнозируются возможные варианты развития сложной системы. В этой связи важной является задача определения качества исходной информации, т. е. рассматриваемых показателей, для возможного описания исследуемой системы.

Интересным является при построении прогнозных моделей использование сочетаний методов, например регрессионного анализа и распознавания образов.

1.9 Прогнозирование с использованием нейронных сетей, искусственного интеллекта и генетических алгоритмов

Жесткие статистические предложения о свойствах временных рядов ограничивают возможности методов математической статистики, теории распознавания образов, теории случайных процессов и т.п. Дело в том, что многие реальные процессы не могут адекватно быть описаны с помощью традиционных статистических моделей, поскольку по сути являются существенно нелинейными и имеют либо хаотическую, либо квазипериодическую, либо смешанную (стохастика + хаос-динамика+детерминизм) основу [14, 20, 21].

В данной ситуации адекватным аппаратом для решения задач диагностики и прогнозирования могут служить специальные искусственные сети [29, 36, 104, 109] реализующие идеи предсказания и классификации при наличии обучающих последовательностей, причем, как весьма перспективные, следует отметить радиально-

базисные структуры, отличающиеся высокой скоростью обучения и универсальными аппроксимирующими возможностями [114].

В его основе нейроинтеллекта лежит нейронная организация искусственных систем, которая имеет биологические предпосылки. Способность биологических систем к обучению, самоорганизации и адаптации обладает большим преимуществом по сравнению с современными вычислительными системами. Первые шаги в области искусственных нейронных сетей сделали в 1943 г. В.Мак-Калох и В.Питс. Они показали, что при помощи пороговых нейронных элементов можно реализовать исчисление любых логических функций [36]. В 1949 г. Хебб предложил правило обучения, которое стало математической основой для обучения ряда нейронных сетей [29]. В 1957-1962 гг. Ф. Розенблатт предложил и исследовал модель нейронной сети, которую он назвал персептроном [104]. В 1959 г. В. Видроу и М. Хофф предложили процедуру обучения для линейного адаптивного элемента – ADALINE. Процедура обучения получила название "дельта правило" [36]. В 1969 г. М. Минский и С. Пайперт опубликовали монографию "Персептроны", в которой был дан математический анализ персептрона, и показаны ограничения, присущие ему. В 80-е годы значительно расширяются исследования в области нейронных сетей. Д. Хопфилд в 1982 г. дал анализ устойчивости нейронных сетей с обратными связями и предложил использовать их для решений задач оптимизации. Т.Кохонен разработал и исследовал самоорганизующиеся нейронные сети. Ряд авторов предложил алгоритм обратного распространения ошибки, который стал мощным средством для обучения многослойных нейронных сетей [29, 36, 104]. В настоящее время разработано большое число нейросистем, применяемых в различных областях: прогнозировании, управлении, диагностике в медицине и технике, распознавании образов и т.д [1, 4, 28, 47, 51, 52].

Нейронная сеть – совокупность нейронных элементов и связей между ними. Основным элементом нейронной сети - это *формальный нейрон*, осуществляющий операцию нелинейного преобразования суммы произведений входных сигналов на весовые коэффициенты

$$y = F\left(\sum_{i=1}^n w_i x_i\right) F(WX), \quad (1.110)$$

где $X = (x_1, x_2, \dots, x_n)^T$ - вектор входного сигнала; $W = (w_1, w_2, \dots, w_n)$ - весовой вектор; F - оператор нелинейного преобразования.

Для обучения сети используются различные алгоритмы обучения и их модификации [9, 11, 22, 42, 70, 139]. Очень трудно определить, какой обучающий алгоритм будет самым быстрым при решении той или иной задачи. Наибольший интерес для нас представляет *алгоритм обратного распространения ошибки*, так как является эффективным средством для обучения многослойных нейронных сетей прямого распространения [85, 127]. Алгоритм минимизирует среднеквадратичную ошибку нейронной сети. Для этого с целью настройки синаптических связей используется метод градиентного спуска в пространстве весовых коэффициентов и порогов нейронной сети. Следует отметить, что настройки синаптических связей сети используется не только метод градиентного спуска, но и методы сопряженных градиентов, Ньютона, квазиньютоновский метод [94]. Для ускорения процедуры обучения вместо постоянного шага обучения предложено использовать адаптивный шаг обучения $\alpha(t)$. Алгоритм с адаптивным шагом обучения работает в 4 раза быстрее. На каждом этапе обучения сети он выбирается таким, чтобы минимизировать среднеквадратическую ошибку сети [29, 36].

В [36] показано, что для прогнозирующих систем на базе НС наилучшие качества показывает гетерогенная сеть, состоящую из скрытых слоев с нелинейной функцией активации нейронных элементов и выходного линейного нейрона. Недостатком большинства рассмотренных нелинейных функций активации является то, что область выходных значений их ограничена отрезком $[0,1]$ или $[-1,1]$. Это приводит к необходимости масштабирования данных, если они не принадлежат указанным выше диапазонам значений. В работе предложено использовать логарифмическую функцию активации для решения задач прогнозирования, которая позволяет получить прогноз значительно точнее, чем при использовании сигмоидной функции.

Анализ различных типов НС показал, что НС может решать задачи сложения, вычитания десятичных чисел, задачи линейного авторегрессионного анализа и прогнозирования временных рядов с использованием метода «скользящего окна» [70].

Проведенный анализ многослойных нейронных сетей и алгоритмов их обучения позволил выявить ряд недостатков и возникающих проблем:

1. неопределенность в выборе числа слоев и количества нейронных элементов в слое;

2. медленная сходимость градиентного метода с постоянным шагом обучения;

3. сложность выбора подходящей скорости обучения α . Так как маленькая скорость обучения приводит к скатыванию НС в локальный минимум, а большая скорость обучения может привести к пропуску глобального минимума и сделать процесс обучения расходящимся;

4. невозможность определения точек локального и глобального минимума, так как градиентный метод их не различает;

5. влияние случайной инициализации весовых коэффициентов НС на поиск минимума функции среднеквадратической ошибки.

Большую роль для эффективности обучения сети играет архитектура НС [114]. Так в [14, 110] показано, что при помощи трехслойной НС можно аппроксимировать любую функцию со сколь угодно заданной точностью. Точность определяется числом нейронов в скрытом слое, но при слишком большой размерности скрытого слоя может наступить явление, называемое перетренировкой сети. Для устранения этого недостатка необходимо, чтобы число нейронов в промежуточном слое было значительно меньше, чем число тренировочных образцов. С другой стороны, при слишком маленькой размерности скрытого слоя можно попасть в нежелательный локальный минимум. Для нейтрализации этого недостатка можно ряд методов описанных в [36, 94, 127].

Прогнозирование с использованием теории генетических алгоритмов. Впервые идея использования генетических алгоритмов для обучения (machine learning) была предложена в 1970-е годы [241, 245, 244, 246]. Во второй половине 1980-х к этой идее вернулись в связи с обучением нейронных сетей. Они позволяют решать задачи прогнозирования (в последнее время наиболее широко генетические алгоритмы обучения используются для банковских прогнозов), классификации, поиска оптимальных вариантов, и совершенно незаменимы в тех случаях, когда в обычных условиях решение задачи основано на интуиции или опыте, а не на строгом (в математическом смысле) ее описании. Использование механизмов генетической эво-

люции для обучения нейронных сетей кажется естественным, поскольку модели нейронных сетей разрабатываются по аналогии с мозгом и реализуют некоторые его особенности, появившиеся в результате биологической эволюции [10, 28, 100].

Основные компоненты генетических алгоритмов – это стратегии репродукций, мутаций и отбор "индивидуальных" нейронных сетей (по аналогии с отбором индивидуальных особей) [28].

Важным недостатком генетических алгоритмов является сложность для понимания и программной реализации. Однако преимуществом является эффективность в поиске глобальных минимумов адаптивных рельефов, так как в них исследуются большие области допустимых значений параметров нейронных сетей. Другая причина того, что генетические алгоритмы не застревают в локальных минимумах – случайные мутации, которые аналогичны температурным флуктуациям метода имитации отжига.

В [28, 88, 100] есть указания на достаточно высокую скорость обучения при использовании генетических алгоритмов. Хотя скорость сходимости градиентных алгоритмов в среднем выше, чем генетических алгоритмов.

Генетические алгоритмы дают возможность оперировать дискретными значениями параметров нейронных сетей. Это упрощает разработку цифровых аппаратных реализаций нейронных сетей. При обучении на компьютере нейронных сетей, не ориентированных на аппаратную реализацию, возможность использования дискретных значений параметров в некоторых случаях может приводить к сокращению общего времени обучения.

В рамках "генетического" подхода в последнее время разработаны многочисленные алгоритмы обучения нейронных сетей, различающиеся способами представления данных нейронной сети в "хромосомах", стратегиями репродукции, мутаций, отбора [241, 244, 245].

Выводы к главе

В главе 1 были рассмотрены и проанализированы некоторые методы и алгоритмы прогнозирования, имеющие четкую математическую формализацию и позволяющие нам работать с временными рядами. Отметим, что на практике, кроме рассмотренных методов, для прогнозирования широко используются методы экспертных оце-

нок, теория межотраслевого баланса, методы основанные на теории игр, вариационного исчисления, спектрального анализа и др. [31, 32, 35, 41, 46, 49, 56, 96]. В последнее время все большее внимание уделяется исследованию и прогнозированию финансовых временных рядов с использованием теории динамических систем, теория хаоса. Это достаточно новая область, которая представляет собой популярный и активно развивающийся раздел математических методов экономики [12, 33, 50, 58, 59, 78, 82, 101, 14039].

Обзор литературы позволил выявить ряд недостатков и задач, которые необходимо решить в данной области.

- Недостатки метода наименьших квадратов (МНК). Использование процедуры оценки, основанной на методе наименьших квадратов, предполагает обязательное удовлетворение целого ряда предпосылок, невыполнение которых может привести к значительным ошибкам: 1. Случайные ошибки имеют нулевую среднюю, конечные дисперсии и ковариации; 2. Каждое измерение случайной ошибки характеризуется нулевым средним, не зависящим от значений наблюдаемых переменных; 3. Дисперсии каждой случайной ошибки одинаковы, их величины независимы от значений наблюдаемых переменных (гомоскедастичность); 4. Отсутствие автокорреляции ошибок, т. е. значения ошибок различных наблюдений независимы друг от друга; 5. Нормальность. Случайные ошибки имеют нормальное распределение; 6. Значения эндогенной переменной x свободны от ошибок измерения и имеют конечные средние значения и дисперсии.
- Проблемы выбора адекватной модели. Выбор модели в каждом конкретном случае осуществляется по целому ряду статистических критериев, например по дисперсии, корреляционному отношению и др.
- Дисконтирование. Классический метод наименьших квадратов предполагает равноценность исходной информации в модели. В реальной же практике будущее поведение процесса значительно в большей степени определяется поздними наблюдениями, чем ранними. Это обстоятельство породило так называемое дисконтирование, т. е. уменьшение ценности более ранней информации.
- Проблема оценки достоверности прогнозов. Важным моментом получения прогноза с помощью МНК является оценка достоверности полученного результата. Для этой цели используется целый ряд статистических характеристик: 1. оценка стандартной

- ошибки; 2. средняя относительная ошибка оценки; 3. среднее линейное отклонение; 4. корреляционное отношение для оценки надежности модели; 5. оценка достоверности выбранной модели через значимость индекса корреляции по Z-критерию Фишера; 6. оценка достоверности модели по F-критерий Фишера; 7. наличие автокорреляций (критерий Дарбина – Уотсона).
- Недостатки обусловленные жесткой фиксацией тренда. Жесткие статистические предположения о свойствах временных рядов ограничивают возможности методов математической статистики, теории распознавания образов, теории случайных процессов и т.п., так как многие реальные процессы не могут адекватно быть описаны с помощью традиционных статистических моделей, поскольку по сути являются существенно нелинейными и имеют либо хаотическую, либо квазипериодическую, либо смешанную основу.
 - Проблемы и недостатки метода экспоненциального сглаживания. Для метода экспоненциального сглаживания основным и наиболее трудным моментом является выбор параметра сглаживания α , начальных условий и степени прогнозирующего полинома. Кроме того, для определения начальных параметров модели остаются актуальными перечисленные недостатки МНК и проблема автокорреляций.
 - Проблемы и недостатки метода вероятностного моделирования. Недостатком модели является требование большого количества наблюдений и незнание начального распределения, что может привести к неправильным оценкам.
 - Проблемы и недостатки метода адаптивного сглаживания. При наличии достаточной информации можно получить надежный прогноз на интервал больший, чем при обычном экспоненциальном сглаживании. Но это лишь при очень длинных рядах. К сожалению, для данного метода нет строгой процедуры оценки необходимой или достаточной длины исходной информации, для конечных рядов нет конкретных условий оценки точности прогноза. Поэтому для конечных рядов существует риск получить весьма приблизительный прогноз, тем более что в большинстве случаев в реальной практике встречаются ряды, содержащие не более 20 – 30 точек.

- Проблемы и недостатки метода Бокса – Дженкинса (модели авторегрессии – скользящего среднего). Проблемы связаны, прежде всего, с неоднородностью временных рядов и практической реализации метода из-за своей сложности.
- Проблемы и недостатки методов, реализованных на базе нейронных сетей. Проблемы неопределенности в выборе числа слоев и количества нейронных элементов в слое, медленная сходимость градиентного метода с постоянным шагом обучения, сложность выбора оптимальной скорости обучения α , влияние случайной инициализации весовых коэффициентов НС на поиск минимума функции среднеквадратической ошибки. Одна из наиболее серьезных трудностей при обучении – это явление переобучения. Кроме того использование немасштабированных данных может привести к «параличу» сети.
- Проблемы и недостатки методов, реализованных на базе генетических алгоритмов. Это проблема кодировки информации, содержащейся в модели нейронной сети, а также сложность понимания и программной реализации.

ГЛАВА 2 МОДЕЛИ ХАОСА ДЛЯ ЗАДАЧ ПРОГНОЗИРОВАНИЯ КУРСОВ АКЦИЙ

2.1 Методика нелинейно-динамического анализа данных

В задачах экономического прогнозирования основной интерес представляет поведение параметров, определяющих характер изучаемых явлений. Ввиду сложности нелинейных взаимодействий этих параметров, далеко не все типы экономико-математических методов и моделей в состоянии сформировать достоверный прогноз.

Основой нелинейно-динамического подхода является учет внутренних особенностей системы, в отличие от статистических методов, в которых все факторы полагаются случайными или неопределенными. Несомненно, что каждый аспект экономической жизни подвержен влиянию неопределенностей. Исследования в области нелинейной динамики показывают, что, даже если возможно найти статистические закономерности управляющих воздействий на ход экономических процессов, предсказать поведение системы с разумной достоверностью не представляется возможным [12, 140]. Вследствие нелинейных взаимодействий между переменными и под влиянием внешней среды для процессов характерно хаотическое поведение. Результаты ряда современных исследований [98, 50, 59] показывают, что поведение экономических систем определяется не только случайностью, но и структурной сложностью. Для многих экономических систем малые изменения внешних параметров или начальных условий могут привести к резким изменениям характера поведения этих систем. В таких системах не исключено существование множества значений, к которым стремятся показатели процессов при различных начальных условиях. Часто структурные изменения, вызванные малыми сдвигами параметров, приводят к регулярному поведению, зависящему от времени – предельным циклам.

Колебания наблюдаемых значений, присущие экономическим данным, весьма различны и по амплитуде, и по области распространения, и по длительности. Для объяснения экономических флуктуаций были выделены две группы факторов [50, 98]. Во-первых, на экономическую систему воздействуют случайные внешние факторы, что приводит к смещению системы от состояния равновесия. Пока система сохраняет близость к равновесию, результирующая траектория динамики экономического процесса может иметь вид осцилля-

торных скачков. Во-вторых, осцилляции могут возникнуть вследствие сложных нелинейных взаимодействий между внутренними переменными (процессами).

Существенным общим положением ряда экономических теорий являлось признание эндогенного характера экономических циклов, т. е. заведомая концентрация внимания на внутренней динамике систем. В целом, эти теории утверждали, что в результате такой динамики современные индустриальные экономики подвержены периодическим флуктуациям с регулярными крупномасштабными закономерностями.

В качестве исследуемого экономического процесса рассматривается динамика курсов акций «РАО ЕЭС», «Ростелеком», «Лукойл», «Сбербанк». Используются данные биржи «ММВБ» в период с 29.05.1997 по 24.06.2003. При этом мы рассматриваем изменения цены на начало торговой сессии.

Предлагается следующий алгоритм исследований для каждого из эмитентов: построение фазового портрета; восстановление аттрактора по временному ряду; вычисление корреляционного интеграла; вычисление корреляционной (фрактальной) размерности аттрактора; вычисление корреляционной энтропии; вычисление характеристических показателей Ляпунова.

При этом должны быть решены промежуточные задачи: разработка вычислительных и программных алгоритмов для определения каждой из перечисленных величин; выбор временной задержки для восстановления аттрактора по временному ряду; выбор метода для определения характеристических показателей Ляпунова.

Для вычисления таких статистических средних, как размерность, энтропия, спектр показателей Ляпунова, и других характеристик аттрактора, необходимо иметь множество точек, определенных в фазовом пространстве размерности n и принадлежащих аттрактору. Число точек M в расчетах конечно, но обязано быть достаточно большим. Согласно формуле

$$M \geq M_{\min} = 10^{2+0.4D}, \quad (2.1)$$

где D – размерность аттрактора. В случае, если требуется вычислить характеристики аттрактора некоторой реальной системы, математическая модель которой неизвестна, используют методику предложенную Такенсом. Доказано, что почти для всех гладких динамиче-

ских систем по имеющейся временной реализации одной наблюдаемой динамической переменной можно сконструировать новый аттрактор, основные свойства которого будут такими же, как у исходного. Порядок расчетов следующий:

1. Построение фазового портрета системы и восстановление аттрактора. Пусть имеется временной ряд экспериментальных данных, представляющий собой отсчеты некоторой физической величины: $\{s_k\}_{k=0}^{M-1}$. Если известен шаг по времени Δt , то время $t = k \Delta t$. Предполагается, что физическая величина s является одной из переменных динамической системы. Система находится в стационарном режиме, т.е. фазовая траектория проходит внутри аттрактора. Для восстановления аттрактора Такенсом предложен метод временной задержки координат [98] n -мерном фазовом пространстве строится последовательность точек вида

$$\begin{aligned} x_k &= (s_k, s_{k+\tau}, \dots, s_{k+(n-1)\tau}), \\ k &= \overline{0, m-1}, m = M - (n-1)\tau, \end{aligned} \quad (2.2)$$

где τ – временная задержка, n – размерность вложения.

Основной результат Такенса состоит в следующем. Если $M \rightarrow \infty$, то множество точек $x_k \in R^n$ задает вложение исходного аттрактора почти при любом выборе наблюдаемой переменной, если n не меньше удвоенной размерности исходного аттрактора. Для оценки характеристик реального исследуемого аттрактора можно вычислять характеристики восстановленного аттрактора. С целью уменьшения ошибки, обусловленной конечностью набора экспериментальных точек $\{s_k\}_{k=0}^{M-1}$, необходимо проводить расчеты при нескольких различных значениях M и n и добиваться независимости получаемых оценок характеристик от M и n в пределах заданной точности.

2. Выбор временной задержки τ . Для малых шагов по времени Δt значения s_k и s_{k+1} будут близкими, поэтому большое значение приобретает правильный выбор временной задержки τ . Необходимо стремиться выбрать τ так, чтобы корреляция между s_k и $s_{k+\tau}$ была по возможности минимальной. Традиционный способ выбора времен-

ной задержки состоит в вычислении автокорреляционной функции временного ряда.

Задержка τ выбирается равной времени первого пересечения нуля автокорреляционной функции. Второй способ требует вычисления спектра мощности временного ряда, т.е. быстрого преобразования Фурье автокорреляционной функции. Если в спектре мощности присутствуют кратные пики, то задержка τ выбирается равной четверти периода самой высокой из доминирующих частот. Третий способ [98] основан на вычислении показателя Херста, который характеризует процесс, как процесс с долговременной памятью примерно в продолжении τ , после которого график начинает следовать случайным блужданиям.

3. Алгоритм вычисления корреляционной размерности аттрактора. В случае модельных данных, когда нам известна размерность n фазового пространства динамической системы и все n координат каждой точки на аттракторе, корреляционную размерность D_2 аттрактора находят следующим образом [82].

Рассмотрим корреляционный интеграл $C(r)$, показывающий относительное число пар точек аттрактора, находящихся на расстоянии, не большем r

$$C(r) = \frac{1}{m(m-1)/2} \sum_{i=0}^{m-2} \sum_{j=i+1}^{m-1} \theta(r - \rho(x_i, x_j)), \quad (2.3)$$

где θ – функция Хевисайда; ρ – расстояние в n -мерном фазовом пространстве; m – число точек x_i на аттракторе.

Если выполняется условие

$$C(r) \sim r^{D_2}, \quad (2.4)$$

то D_2 считают корреляционной размерностью аттрактора.

Справедливость приведенного степенного закона ограничена значениями r , достаточно малыми по сравнению с размером аттрактора. При увеличении r величина $C(r)$ достигает насыщения $C(r)$ (при r , сравнимых с размером аттрактора). С другой стороны, при очень малых значениях r число пар точек x_i, x_j , расстояние между которыми не превышает r , становится малым (из-за конечности числа точек на аттракторе) и статистика становится бедной. Кроме того, приобрета-

ет решающее значение влияние инструментальных ошибок измерения сигнала. Следовательно, на практике степенной закон выполняется только в ограниченном диапазоне значений r (скейлинговом диапазоне), который и может быть использован для определения размерности аттрактора.

Учитывая, что из (2.4) следует

$$\ln C(r) \sim D_2 \ln r \quad (2.5)$$

получаем оценку размерности аттрактора как тангенс угла наклона прямой, аппроксимирующей график корреляционного интеграла $C(r)$ в двойном логарифмическом масштабе.

В случае экспериментальных данных мы обычно не знаем размерность фазового пространства системы и располагаем информацией только об одной координате точек на аттракторе. Поэтому все расчеты проводятся для нескольких размерностей фазового пространства $n = 1, 2, 3, \dots$. Для восстановления аттрактора используется метод Такенса. При этом корреляционная размерность аттрактора $D_2(n)$ сначала возрастает, но затем обычно выходит на постоянный уровень $D_2(n) \approx D_2$. Таким образом, получают искомую корреляционную размерность D_2 аттрактора и оценку размерности фазового пространства системы $n \leq 2D_2 + 1$. Если же выходной сигнал динамической системы сильно зашумлен, то размерность аттрактора постоянно растет.

4. Алгоритм вычисления корреляционной энтропии аттрактора. Корреляционная энтропия K_2 может быть вычислена достаточно просто [82, 98]. Для этого также вычисляют корреляционный интеграл (2.3), но рассматривают не только его зависимость от расстояния r , но и от размерности фазового пространства n . При этом полагают, что

$$C(r, n) \sim r^{D_2} \exp(-nK_2) \quad (2.6)$$

откуда

$$K_2(r, n) = \ln \frac{C(r, n)}{C(r, n+1)}. \quad (2.7)$$

Энтропия K_2 аппроксимируется в приемлемом диапазоне значений r и n .

5. Алгоритмы вычисления характеристических показателей Ляпунова. Если математическая модель динамической системы неизвестна, то требуются специальные алгоритмы нахождения показателей Ляпунова, поскольку в этом случае мы не знаем ни матрицы Якоби, ни размерности фазового пространства системы [80, 97].

В 1985 году Вольф описал алгоритм нахождения наибольшего показателя Ляпунова по временному ряду данных. Он следит за парой точек на аттракторе. Схематическая иллюстрация алгоритма приведена на рисунке 2.1. Он начинается с первой точки данных $y(t_0)$ и ее ближайшего соседа $z_0(t_0)$, которые отделены на расстояние r_0 . Эти две точки эволюционируют во времени с шагом Δt пока расстояние r'_0 между ними не превысит некоторую величину ϵ . Эволюционирующая первая точка данных $y(t_1)$ сохраняется, а новый сосед $z_1(t_1)$ ищется такой, что расстояние

$$r_1 = \|y(t_1) - z_1(t_1)\| \quad (2.8)$$

снова меньше, чем ϵ и такой, что $z_1(t_1)$ лежит так близко, насколько это возможно в одинаковом направлении, что и направление от $y(t_1)$ к $z_0(t_1)$.

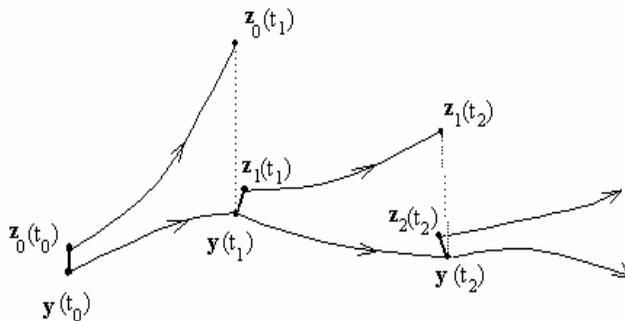


Рис. 2.1 Методика вычисления показателей Ляпунова

Процедура продолжается до тех пор, пока принятая за основу сравнения траектория y не дойдет до конца временного ряда. Наибольший показатель Ляпунова аттрактора оценивается как

$$\lambda_1 = \frac{1}{m \Delta t} \sum_{k=0}^{L-1} \ln \frac{r'_k}{r_k}, \quad (2.9)$$

где L – число шагов замены и m – общее число шагов по времени, в течение которых движется траектория у.

Каждая заменяющая точка должна лежать в том же направлении, что и старая, но из-за ограниченного числа отсчетов данных необходимы компромиссы. Первоначально поиск точки для замены ограничен в конусе угловой ширины и высоты ϵ около $y(t)$. Обычно θ устанавливают равным $\pi/9$. Значение θ увеличивают по мере необходимости пока сосед у не будет найден. Если это не приводит к успеху, то берут ближайшего соседа у безотносительно ϵ . Ориентационные ошибки обычно приводят к малым ошибкам в оценке λ_1 . Таким образом, результаты обычно нечувствительны к выбору θ . Если λ_1 не очень маленькое, то алгоритм Вольфа может быть полезен для определения того, хаотичный ли наблюдаемый временной ряд. Часто, однако, интересуются более чем одним первым показателем Ляпунова. Алгоритм может расширяться для измерения $\lambda_1 + \lambda_2$, но при этом становится заметно сложнее. Он должен следить за треугольником точек и сохранять ориентацию относительно треугольника, которому выбирают замену. Этот подход становится плохим для более чем двух положительных показателей Ляпунова.

Оценка показателя Херста. Причины смещенных случайных блужданий широко изучались Херстом в 40-х гг. и Мандельбротом в 60-70-х гг. Мандельброт назвал их обобщенным броуновским движением. Теперь мы можем назвать их фрактальными временными рядами.

Херст показал, что большинство естественных явлений, включая речные стоки, температуры, осадки, солнечные пятна следуют «смещенному случайному блужданию» – тренду с шумом. Сила тренда и уровень шума могут быть оценены тем, как изменяется нормированный размах со временем, или, другими словами, насколько величина H (показатель Херста) превосходит 0.5 [98].

Мы перенесли метод Херста изучения временных рядов природных явлений на временные ряды в экономике и на рынках капитала, чтобы выяснить, являются ли эти ряды также смещенными слу-

чайными блужданиями. Для этого необходимо определить размах, который был бы сравним с колебаниями уровня

$$X_{t,N} = \sum_{u=1}^t (e_u - M_N), \quad (2.10)$$

где: $X_{t,N}$ – накопленное отклонение за N периодов; e_u – цена; M_N – среднее e_u за N периодов. Тогда размах становится разностью между максимальным и минимальным уровнями, достигнутыми в (2.8)

$$R = \text{Max}(X_{t,N}) - \text{Min}(X_{t,N}), \quad (2.11)$$

где: R – размах отклонения X; $\text{Max}(X_{t,N})$ – максимальное значение для X; $\text{Min}(X_{t,N})$ – минимальное значения для X.

Для сравнения различных типов временных рядов Херст разделил этот размах на стандартное отклонение исходных наблюдений. Этот «нормированный размах» должен увеличиваться со временем. Херст ввел следующее соотношение

$$R/S = (a * N)^H, \quad (2.12)$$

где: R/S – нормированный размах, N – число наблюдений, a – константа, H – показатель Херста.

Когда H отличается от 0.50, то это значит, что наблюдения не являются независимыми. Каждое наблюдение несет память о всех предшествующих событиях. Это не кратковременная память, которую часто называют «марковской». Это другая память – долговременная, теоретически она сохраняется навсегда.

Влияние настоящего на будущее может быть выражено корреляционным соотношением

$$C = 2^{2H-1} - 1, \quad (2.13)$$

где: C – мера корреляции, H – показатель Херста.

Имеются три различных классификации для показателя Херста: а) $H = 0.5$, б) $0 \leq H < 0.5$, в) $0.5 < H < 1.0$. H, равное 0.5, указывает на случайный ряд. События случайны и некоррелированы. Функция плотности вероятности может быть нормальной кривой, однако это не обязательное условие.

В случае если $0 \leq H < 0.5$. Данный диапазон соответствует антиперсистентным, или эргодическим рядам. Такой тип системы часто называют «возврат к среднему». Если система демонстрирует рост в предыдущий период, то скорее всего в следующем периоде начнется спад. И наоборот.

При $0.5 < H < 1.0$ мы имеем персистентные, или трендо-устойчивые ряды. Если ряд возрастает (убывает) в предыдущий период, то вероятно, что он будет сохранять эту тенденцию какое-то время в будущем. Тренды очевидны. Трендо-устойчивость поведения, или сила персистентности, увеличивается при приближении H к единице, или 100% корреляции, в соотношении (2.13). Чем ближе H к 0.5, тем более зашумлен ряд и тем менее выражен его тренд.

Замечание. Результаты R/S-анализа могут быть смещенными под влиянием двух основных обстоятельств: нестационарности данных и наличие процесса с короткой памятью. Поэтому, чтобы исключить процесс с короткой памятью следует, перед применением R/S-анализа, найти AR(1) разности.

Теоретически доказана целесообразность применения теории нелинейных систем. Определены основные алгоритмы и методики. Которые необходимо реализовать на практике.

2.2 Практические исследования стохастических характеристик динамики курса акции «РАО ЕЭС»

Следуя приведенному алгоритму исследований восстановим фазовый портрет для временного ряда динамики курса акций «РАО ЕЭС», данные биржи «ММВБ» за период с 29.05.1997 по 24.06.2003 гг. Массив значений $N = 1511$. На рисунке 2.2 показана динамика курса акций российской компании «РАО ЕЭС».

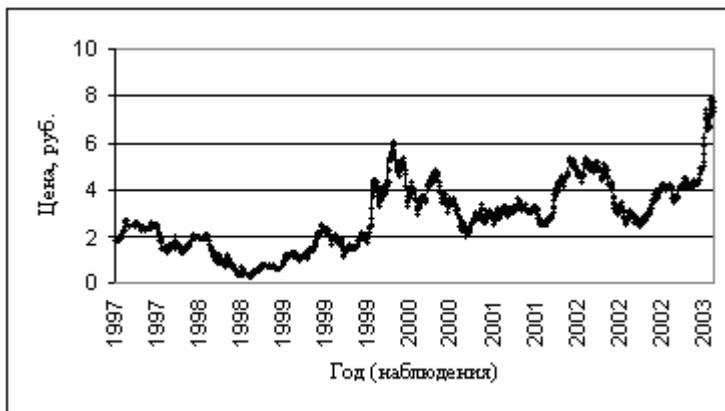


Рис. 2.2 Динамика курса акций российской компании «РАО ЕЭС» в период с 29.05.1997 по 24.06.2003 гг.

Для этого необходимо оценить время задержки τ . Чтобы оценка имела достаточную обоснованность воспользуемся несколькими перечисленными ранее методами.

2.2.1 Оценка τ по автокорреляционной функции

С помощью разработанной программы рассчитывается автокорреляционная функция, показанная на рисунке 2.3.

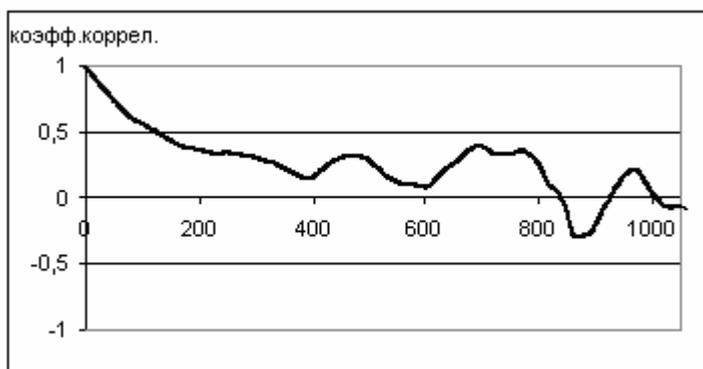


Рис. 2.3 Автокорреляционная функция

Как видно из рисунка 2.3 полученное значение $\tau = 840$ не совсем можно обосновать. На практике может случиться так, что автокорреляционная функция вообще не пересечет ось абсцисс, поэтому необходим уточняющий подход.

2.2.2 Спектральный анализ

Спектральный анализ реализуется для метода при котором в исходный ряд добавляются новые нулевые значения до количества соответствующего ближайшей степени 2. При этом для вычислений применяются следующая процедура: 15% сглаживание на концах ряда (Процедура применяется для устранения ложных пиков (бликов) в периодограмме). На рисунке 2.4 приведены результаты спектрального анализа автокорреляционной функции. Из рисунка 2.4 хорошо видно, что пик самой высокой из доминирующих частот приходится на период равный 400. Поэтому τ следует выбрать равной четверти периода. Примерно 100 дней.

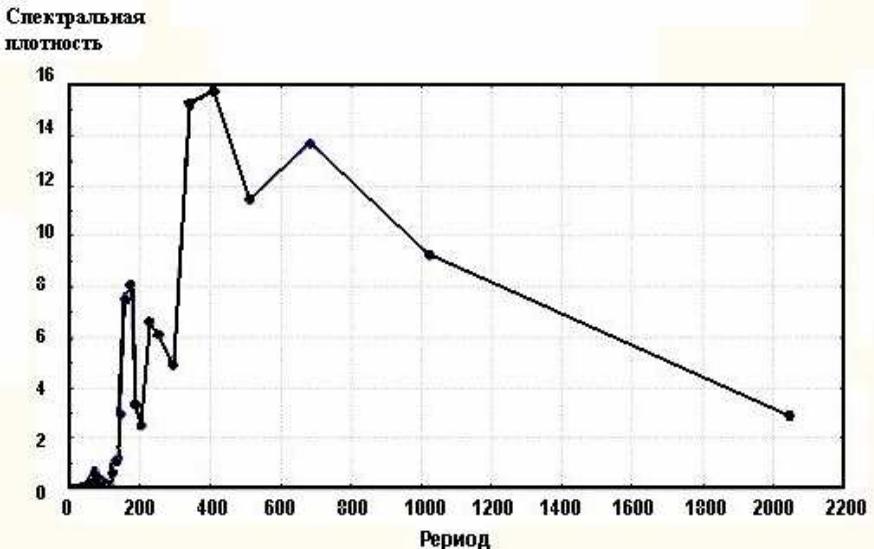


Рис. 2.4 Спектральный анализ автокорреляционной функции

Увеличим масштаб построения рисунок 2.5-2.6.

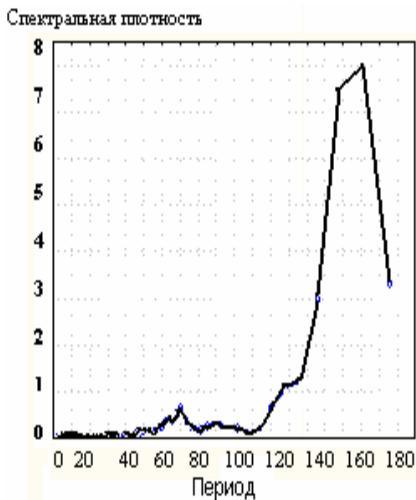


Рис. 2.5 Зависимость спектральной плотности от периода

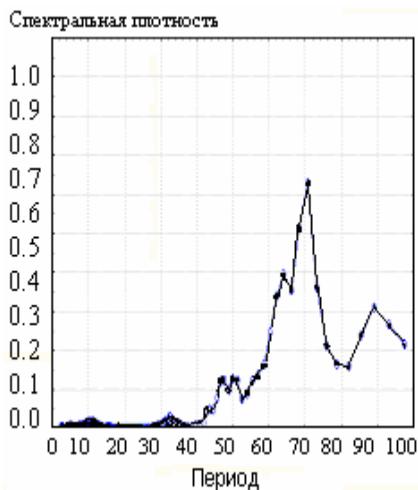


Рис. 2.6 Зависимость спектральной плотности от периода

Легко выделить устойчивые циклы 70 дней, 170 дней, 400 дней. При дальнейшем увеличении масштаба мы можем выделить и другие менее значительные временные периоды.

Построим зависимость частоты от спектральной плотности, где так же можно выделить те же временные циклы рисунок 2.7

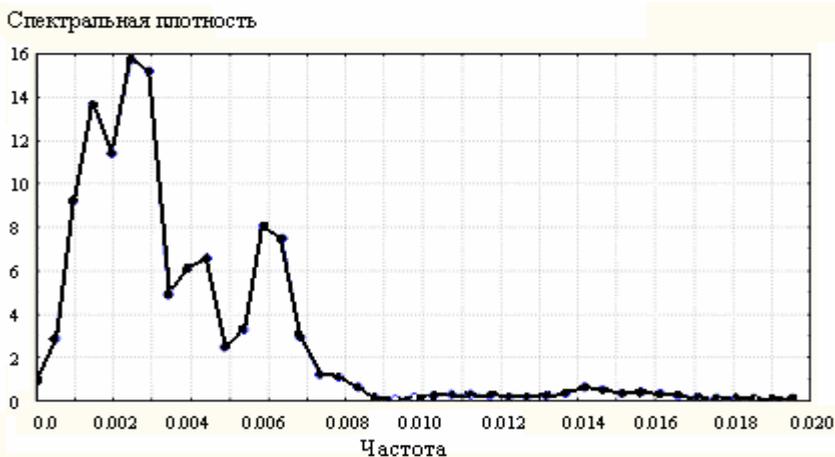


Рис. 2.7 Зависимость спектральной плотности от частоты для динамики курса акций «РАО ЕЭС»

Таким образом для определения времени задержки и построения аттрактора берем четверть периода самого высокого пика. Это примерно 100 дней.

2.2.3 Оценка показателя Херста для курса акций компании «РАО ЕЭС»

Прологарифмируем соотношение (2.12)

$$\log(R/S) = H * (\log(N) + \log(a)) \quad (2.14)$$

Если в двойных логарифмических координатах найти наклон R/S как функцию от N , то тем самым мы получим оценку H . Для очень большого количества наблюдений N можно ожидать сходимости ряда к величине $H = 0.50$, так как эффект памяти уменьшается до того уровня, когда становится незаметным. Регрессия в этом случае должна выполняться до того как H приблизится к 0.5 , так как корреляционная мера не применима ко всем без исключения приращениям.

На рисунке 2.8 в двойных логарифмических координатах представлена кривая зависимости R/S от N для $H = 0.731494$, построенная по данным из рисунка 2.2 (динамика курса акций российской компании «РАО ЕЭС»).



Рис. 2.8 Кривая зависимости R/S от N – количества наблюдений

В сущности, оценка H , которая значительно отличается от 0.50, имеет два возможных объяснения:

1. В изучаемом временном ряду имеется долговременная память. Каждое наблюдение коррелирует до некоторой степени с последующими наблюдениями.

2. Такого рода анализ сам по себе несостоятелен, и аномальная величина H не означает, что имеет место эффект долговременной памяти [82, 98].

Может оказаться, что у нас недостаточно данных для обоснованного теста (при этом не существует четких критериев того, сколько данных необходимо). Тем не менее, в этом случае изучаемый ряд как ряд независимых случайных переменных либо заключает в себе H , отличное от 0.50, либо представляет собой независимый процесс с толстыми хвостами, описанный еще Кутнером.

Мы можем проверить обоснованность наших результатов путем случайного перемешивания данных, в результате чего порядок наблюдений станет полностью отличным от исходного ряда. Ввиду того, что наблюдения остаются теми же, их частотное распределение также останется неизменным. Далее вычислим показатель Херста этих перемешанных данных. Если ряд действительно является независимым, то показатель Херста не изменится, поскольку отсутствовал эффект долговременной памяти, т. е. корреляции между наблюдениями. В этом случае перемешивание данных не оказывает влияния на качественные характеристики данных.

Если имел место эффект долговременной памяти, то порядок данных весьма важен. Перемешивая данные, мы тем самым разрушаем структуру системы. Оценка H при этом окажется значительно ниже, даже если частотное распределение наблюдений не изменится [98].

Нами был выполнен тест на перемешивание значений ряда динамики курса акций «РАО ЕЭС». На рисунке 2.9 в двойных логарифмических координатах представлены перемешанный и не перемешанный ряды.



Рис. 2.9 В двойных логарифмических координатах представлены перемешанный и не перемешанный ряды курса акций «РАО ЕЭС»

Исходный ряд дал результативную оценку $H = 0.516$, перемешанный – $H < 0.5$. Такое падение величины H говорит о том, что при перемешивании была разрушена структура процесса. Перемешанный ряд остался не нормально распределенным, но процесс перемешивания сделал данные независимыми. Это доказывает утверждение Мандельброта о том, что R/S -анализ работоспособен безотносительно к распределению временного ряда.

2.2.4 Построение фазового портрета. Восстановление аттрактора по временному ряду

На рисунке 2.10 показан фазовый портрет для размерности вложения равной трем. На рисунке 2.11 показан восстановленный аттрактор.

Полученный аттрактор является «странным аттрактором». Результаты свидетельствуют о нерегулярном (хаотическом) поведении процесса, что говорит о значительном изменении динамики курса акций при различных начальных условиях (начальной стоимости). Вместе с тем, внутренняя структура системы такова, что курс акций стремится к определенной притягивающей траектории. При этом

можно выделить три области притяжения. В районе 1.5 – 3 рубля, 4.5 – 6 рублей, 7 – 9 рублей.

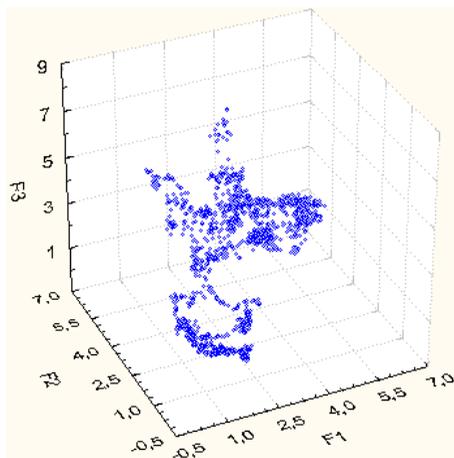


Рис. 2.10 Фазовый портрет

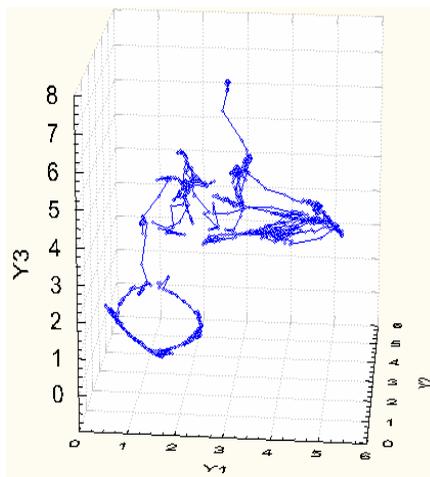


Рис. 2.11 Аттрактор

Если рассмотреть множество начальных условий на фазовой плоскости, занимающую некоторую область размером F , то в следствии диссипации поток приводит к сокращению площадей. Следовательно поверхность F под действием потока вырождается до линейного отрезка на аттракторе. Значит происходит потеря информации относительно взаимного расположения точек, первоначально принадлежавших поверхности F . По достижении аттрактора информация утрачивается необратимо. Этот вывод опирается исключительно на сокращение площадей и на одновременное существование аттрактора [12]. Вывод остается в силе независимо от типа аттрактора.

Как видно из рисунка 2.11, на аттракторе мы имеем три области притяжения, которые связаны между собой фазовым переходом. Можно утверждать, что адекватные прогнозы можно сделать только для цен курсов акций, формирующих области притяжения, а не для всего временного ряда.

2.2.5 Вычисление корреляционного интеграла

Процедура вычислений не так сложна, однако программа реализующая данный метод вычислений, производит расчеты фрактальной размерности аттрактора для вложений $m = 2 \div 9$ порядка 40 минут на PC типа Intel Pentium 500.

Во-первых, в соответствии с (2.3) рассчитываются корреляционные интегралы для последовательно увеличивающихся размерностей вложения. При этом рассматриваются регрессии линейной области двойных логарифмических кривых. Фрактальная размерность должна сходиться к своей истинной величине, по мере увеличения размерности вложения. На рисунке 2.12 показаны диаграммы корреляционных интегралов для вложений $m = 2 \div 9$. Анализ данных рисунка 2.12 позволил выделить область сходимости фрактальной размерности рисунок 2.13.

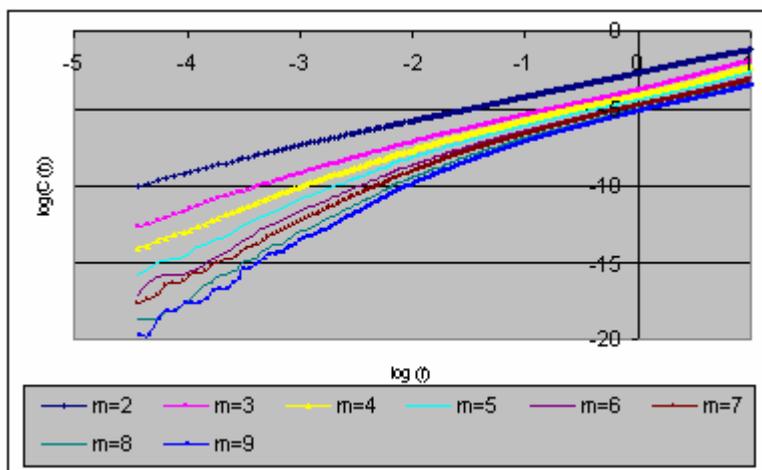


Рис. 2.12 Диаграммы корреляционных интегралов для вложений $m = 2 \div 9$ для курса акций «РАО ЕЭС»

На практике график корреляционного интеграла отклоняется от прямой линии в области больших r и очень малых r , когда количество пар точек становится мало для хорошей статистической оценки. Интервал линейности тем больше, чем больше объем обрабатываемых данных. Чаще всего его выбирают «на глаз», а затем подвер-

гают полученные точки обработке с помощью метода наименьших квадратов для нахождения аппроксимирующей прямой. Угол наклона этой прямой даст значение корреляционной размерности.

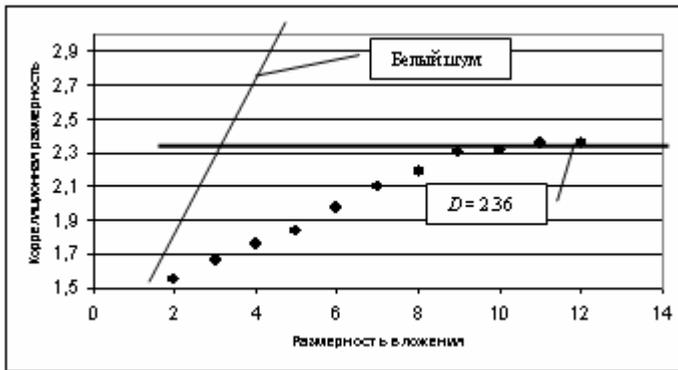


Рис. 2.13 Сходимость фрактальной размерности $D = 2.36$

Реальный анализ показывает, что устойчивая сходимость к фрактальной размерности согласуется с теорией. Система является низко-размерной и для моделирования динамики этого рынка требуется три переменные.

2.2.6 Вычисление корреляционной энтропии и старшего показателя Ляпунова

На рисунке 2.14 показаны в двойных логарифмических координатах зависимости корреляционной энтропии для различных m от r (3).

Характерное время, на которое может быть предсказано поведение системы, обратно пропорционально энтропии Колмогорова [66]. Если энтропия достигает нуля, то система становится полностью предсказуемой. Для истинно случайных процессов энтропия неограниченно велика. Энтропия системы в режиме странного аттрактора положительна, но имеет конечное значение. Числовое значение энтропии является количественной характеристикой степени хаотичности системы.

На рисунке 2.14 хорошо видна сходимость корреляционной энтропии в зависимости от размерности вложения и величины $\log(r)$. Корреляционная энтропия системы примерно равна 0,116. Подтвер-

ждением наших выводов является расчетное значение старшего показателя Ляпунова. Как видно из рисунка 2.15 найдена устойчивая сходимост к величине 0,1162 бит/день.

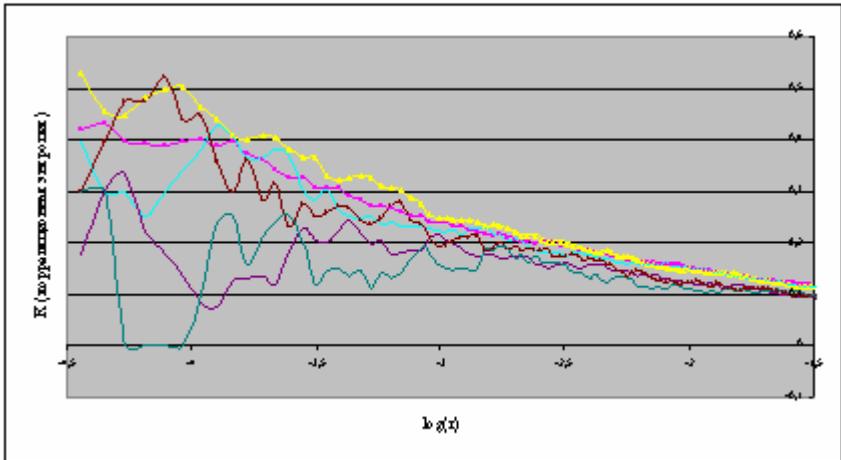


Рис. 2.14 Корреляционная энтропия для курса акций «РАО ЕЭС»

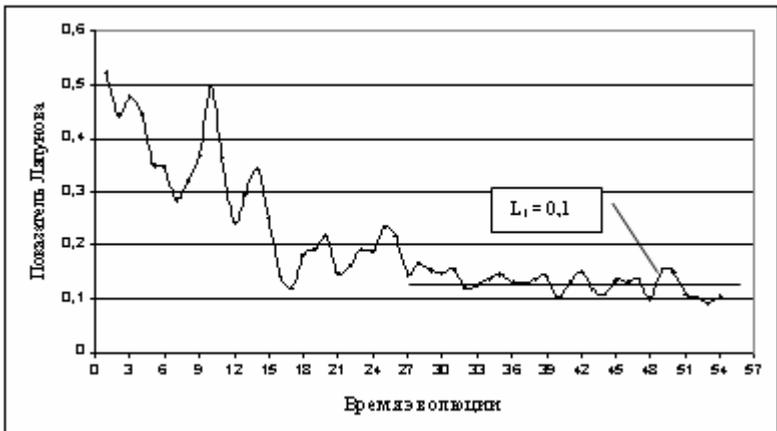


Рис. 2.15 Сходимость наибольшего показателя Ляпунова

Это означает, что мы теряем предсказательную мощност с скоростью 0,02 бит/день. Таким образом, можно утверждать, что можно делать адекватные прогнозы на $1/0,1162 = 8,605$ дней.

2.3 Исследование основных нелинейных показателей для курсов акций «Ростелеком», «Лукойл», «Сбербанк»

Используются данные биржи «ММВБ» в период с 29.05.1997 по 24.06.2003. Используя рассмотренную методику, определим основные показатели временных рядов.

Автокорреляционные функции, рассматриваемых временных рядов показаны на рисунке 2.17. Согласно методу 1, который описан ранее, для определения времени задержки нам необходимо взять то значение периода, где автокорреляционная функция пересекает ноль.

Анализ автокорреляционных функций показал, что для акций «Ростелеком» величина τ должна равняться примерно 200 дней. Для акций нефтяной компании «Лукойл» $\tau = 929$ дней. Для акций «Сбербанк» $\tau = 825$ дней. Такие времена задержки не могут быть использованы в расчетах поэтому необходимо проводить дополнительные исследования.

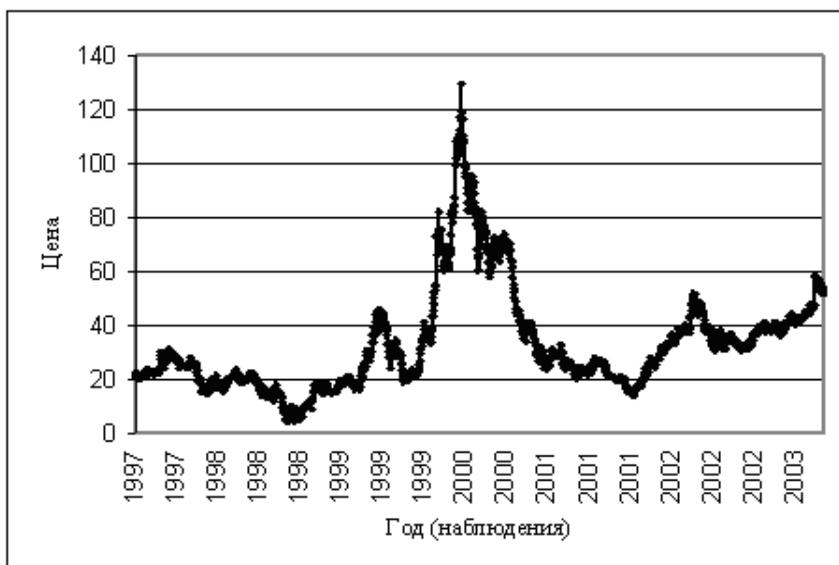


Рис. 2.16 а)



Рис. 2.16 б)

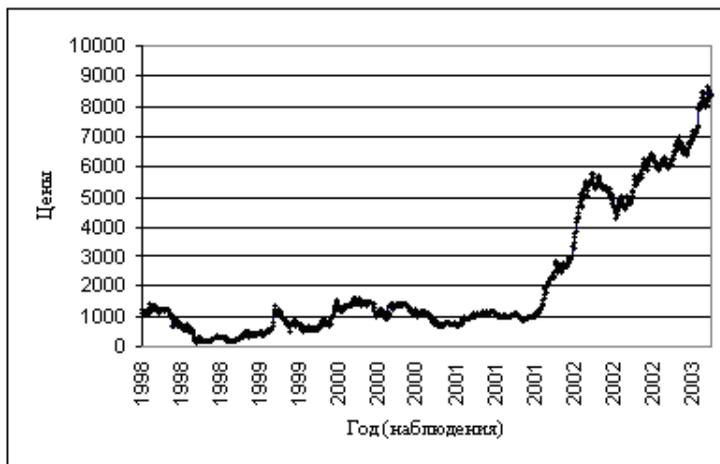


Рис. 2.16 в)

Рис. 2.16 Динамика курсов акций: а) «Ростелеком 1536 наблюдений; б) «Лукойл» – 1511 наблюдений; в) «Сбербанк» – 1304 наблюдений.

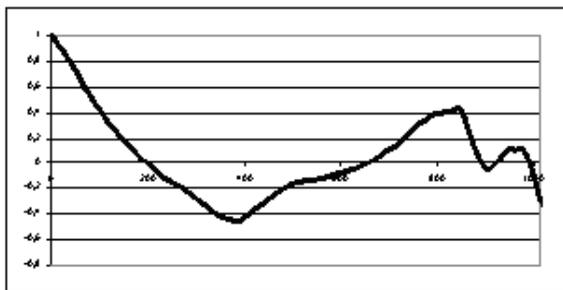


Рис. 2.17 а)

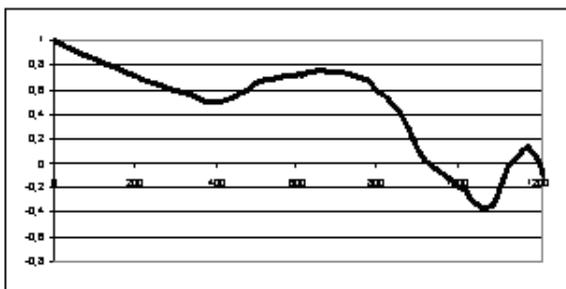


Рис. 2.17 б)

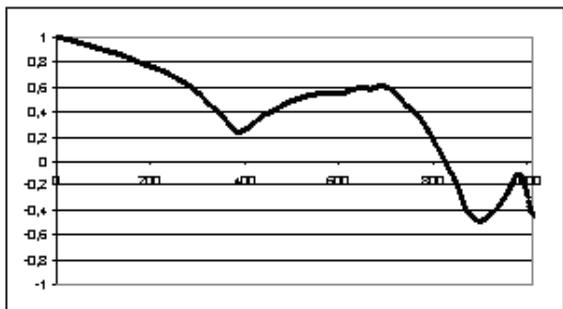
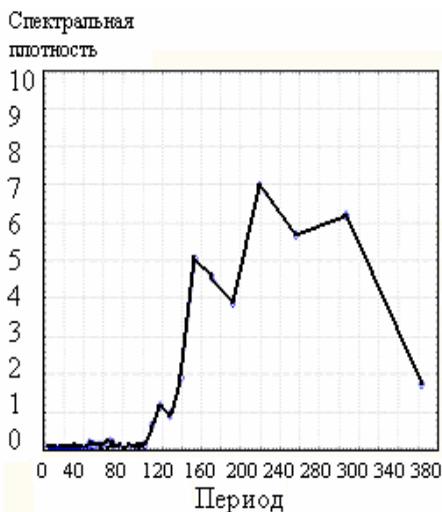


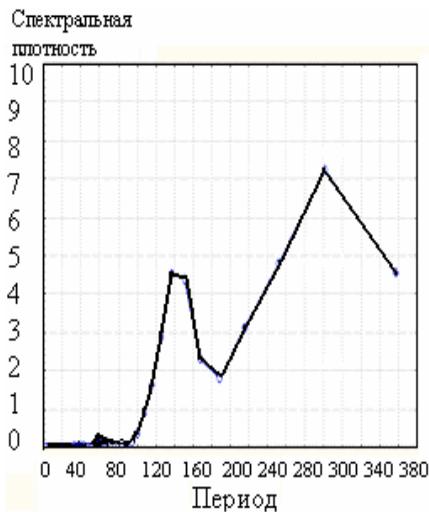
Рис. 2.17 в)

Рис. 2.17 Автокорреляционные функции: а) «Ростелеком» (пересекает 0 на 200 шаге); б) «Лукойл» (пересекает 0 на 929 шаге); в) «Сбербанк» (пересекает 0 на 825 шаге)

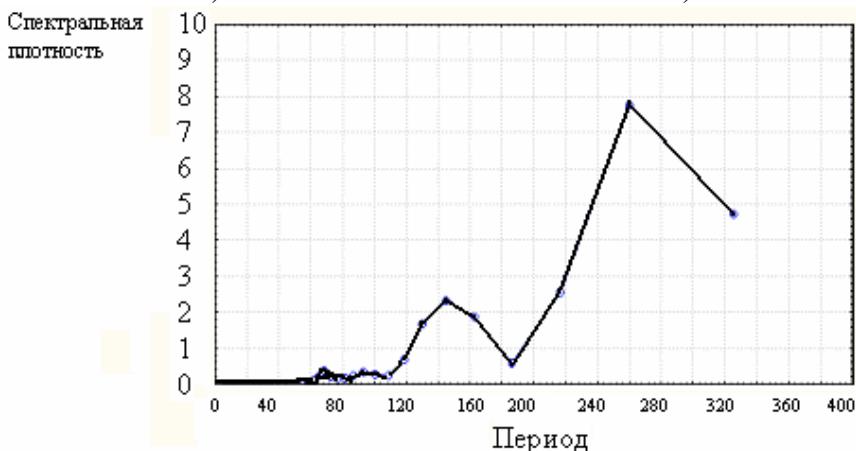
На рисунке 2.18 показаны результаты спектрального анализа. На графиках показана зависимость спектральной плотности от величины периода.



а)



б)



в)

Рис. 2.18 Спектральный анализ автокорреляционных функций: а) «Ростелеком; б) «Лукойл»; в) «Сбербанк».

Из рисунка 2.18 легко видеть, что для акций Ростелекома максимальный пик приходится на 220 значение периода, соответственно τ будет равно четверти этого периода т.е. $\tau = 55$ дней. Для акций Лукойла, период 300 дней, а соответственно $\tau = 75$. Для акций Сбербанка, период составляет 260 дней τ соответственно равно 65.

2.3.1 R/S – анализ. Анализ показателя Херста

На рисунке 2.19 в двойных логарифмических координатах показаны зависимости R/S от N для курсов акций. Для акций компании «Ростелеком» процесс с долговременной памятью наблюдается примерно в продолжении 70 –80 дней, $H = 0,74$. После этой точки график начинает следовать случайным блужданиям $H = 0,44$. Столь высокая оценка показателя Херста свидетельствует о том, что фондовый рынок является фракталом, а не следует случайным блужданиям.

Для акций компании «Лукойл» такой период составляет также 70 – 80 дней $H = 0,77$, а для акций «Сбербанка» 60 – 70 дней $H = 0,91$. Видно, что акции имеют примерно схожие величины H и длины циклов.

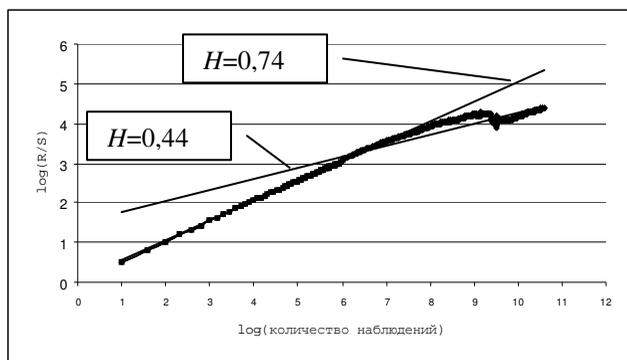


Рис. 2.19 а)

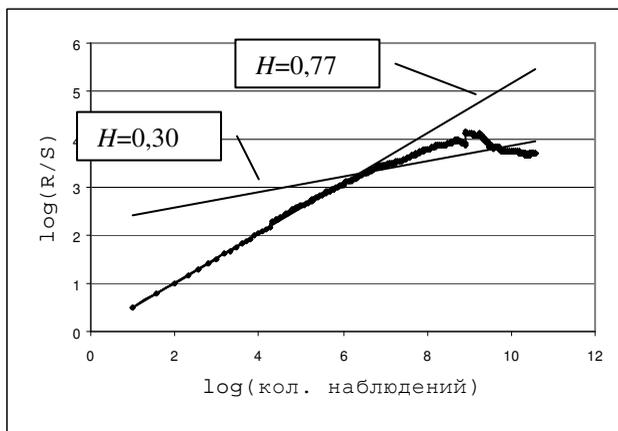


Рис. 2.19 б)

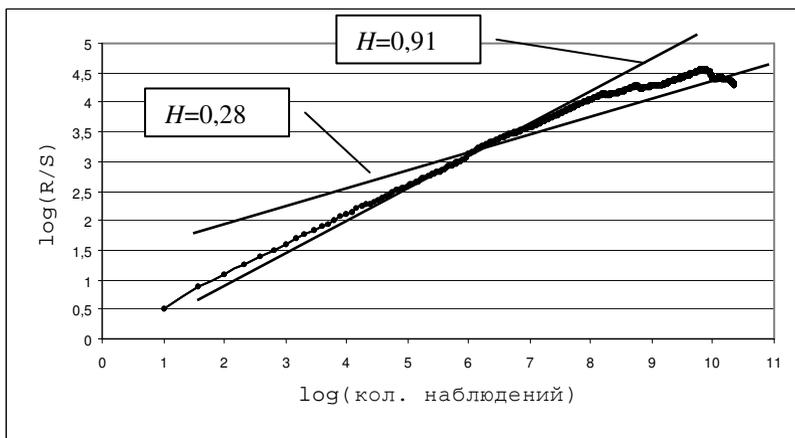


Рис. 2.19 в)

Рис. 2.19 R/S-анализ:

а) «Ростелеком»; б) «Лукойл»; в) «Сбербанк».

2.3.2 Восстановление фазовых портретов и аттракторов по временным рядам

Анализируя результаты предыдущих исследований получена оптимальная величина задержки τ , которую мы используем для построения фазового портрета и восстановления аттрактора системы.

Так для акций «Ростелеком» τ составляет порядка 75 дней, что подтверждено спектральным анализом и R/S – анализом. Для «Лукойла» $\tau = 75$ дней, а для «Сбербанка» $\tau = 65$ дней.

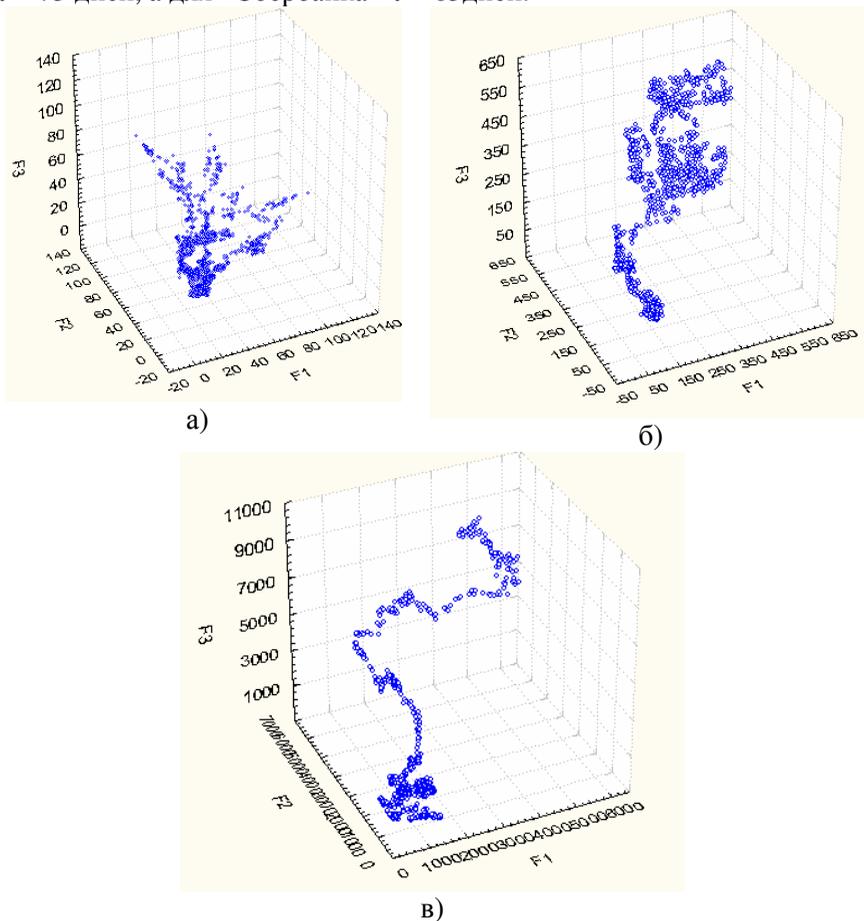
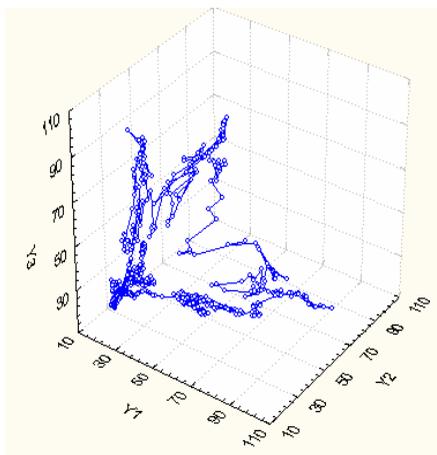
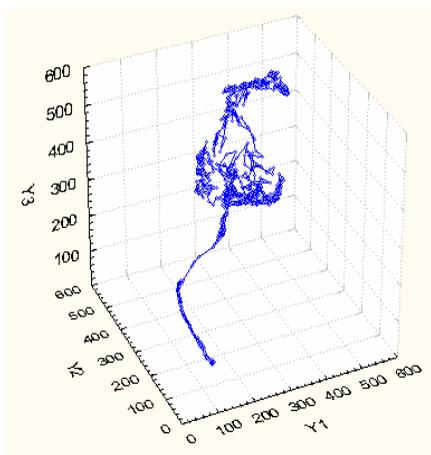


Рис 2.20 Фазовые портреты:
а) «Ростелеком; б) «Лукойл»; в) «Сбербанк».

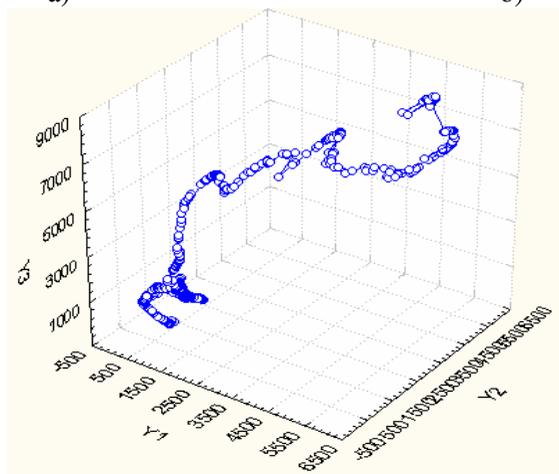
На рисунке 2.20 показаны восстановленные фазовые портреты. Визуальная оценка данных в нелинейных системах очень важна потому, что они как правило не имеют единственного решения. По виду фазового портрета мы восстанавливаем вид аттрактора, который как видно из рисунка 2.21, является «странным аттрактором».



а)



б)



в)

Рис. 2.21 Восстановленные аттракторы: а) «Ростелеком»; б) «Лукойл»; в) «Сбербанк».

На основании полученных результатов можно построить модель системы

$$\begin{aligned} x(t+1) &= F(x(t)), \\ y(t) &= Q(x(t)), \end{aligned} \tag{2.15}$$

где $x(t)$ – n -мерный вектор переменных состояний системы, $y(t)$ – значения курса акций, t – дискретное время, F, Q – нелинейные операторы.

2.3.3 Вычисление корреляционных интегралов и фрактальной размерности

Как было установлено, фрактальная размерность дает информацию относительно изучаемой системы. Целое число, следующее за числом фрактальной размерности, говорит о минимальном количестве переменных, необходимых для моделирования системы. Оно задает нижнюю границу возможного количества степеней свободы [98, 128]. На рисунке 2.22 показаны вычисленные корреляционные интегралы для различных величин вложения m .

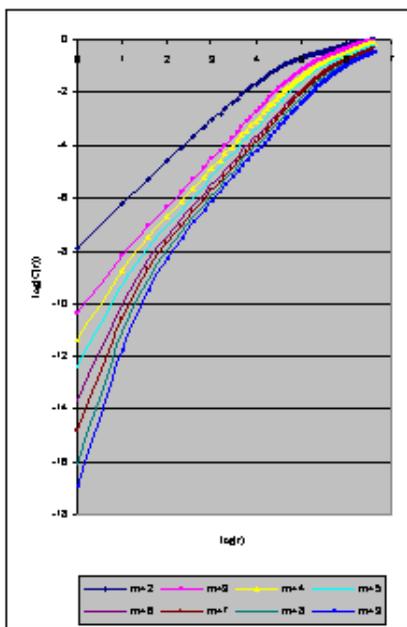


Рис. 2.22 а)

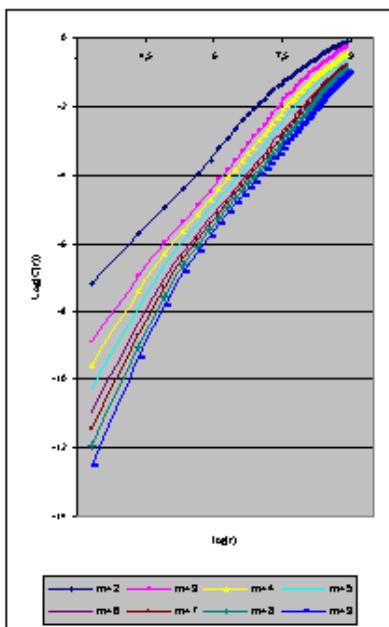


Рис. 2.22 б)

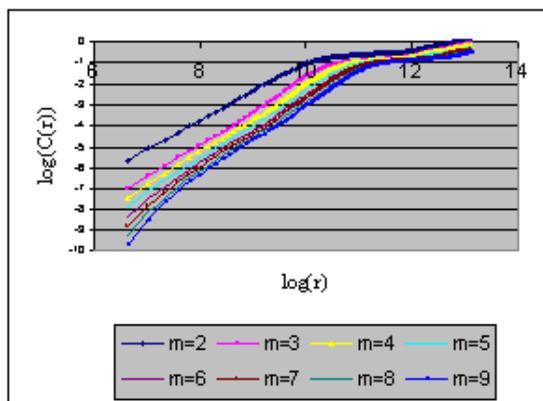


Рис. 2.22 в)

Рис. 2.22 Корреляционные интегралы:
а) «Ростелеком»; б) «Лукойл»; в) «Сбербанк».

Для размерности m вычисляется $C(r)$ (2.3) при увеличении R . Находя наклон прямой на графике (см. рисунок 2.22) линейной регрессии в двойных логарифмических координатах, оценивается корреляционная размерность D для каждой размерности вложения m . При увеличении m размерность D будет сходиться к своей истинной величине. На рисунке 2.23 показана сходимость D для различных m .

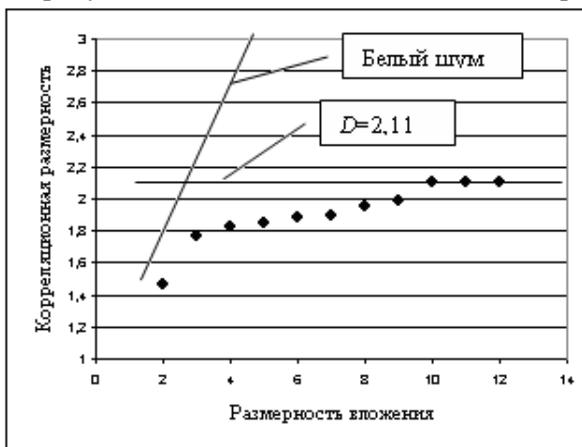


Рис. 2.23 а)

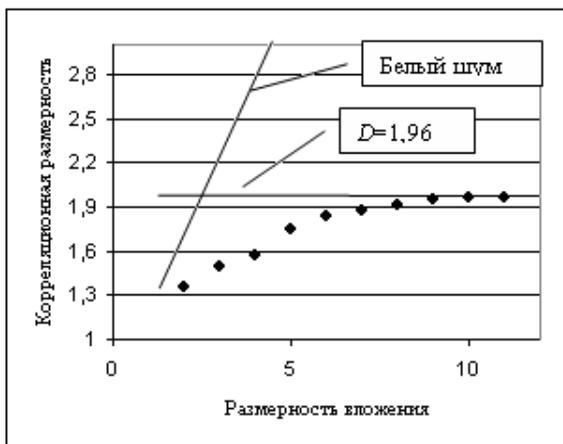


Рис. 2.23 б)

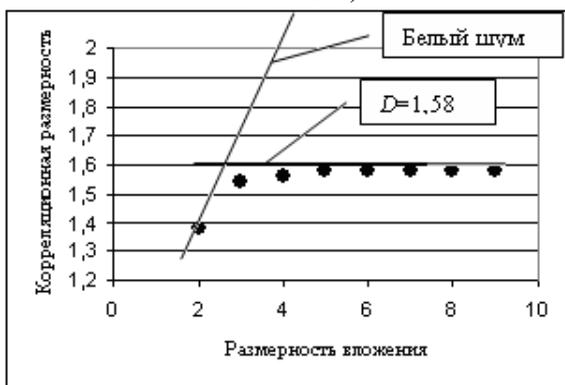


Рис. 2.23 в)

Рис. 2.23 Сходимость фрактальной размерности:
а) «Ростелеком»; б) «Лукойл»; в) «Сбербанк».

Как видно из рисунка 2.23 для акций «Ростлекома» $D = 2,11$, для акций «Лукойла» $D = 1,96$, для акций «Сбербанка» $D = 1,58$. Это значит, что смоделировать динамику этих акций соответственно с помощью трех и двух переменных. Таким образом, решается проблема определения количества переменных для моделирования сложных, динамических систем.

2.3.4 Вычисление корреляционной энтропии и показателей Ляпунова

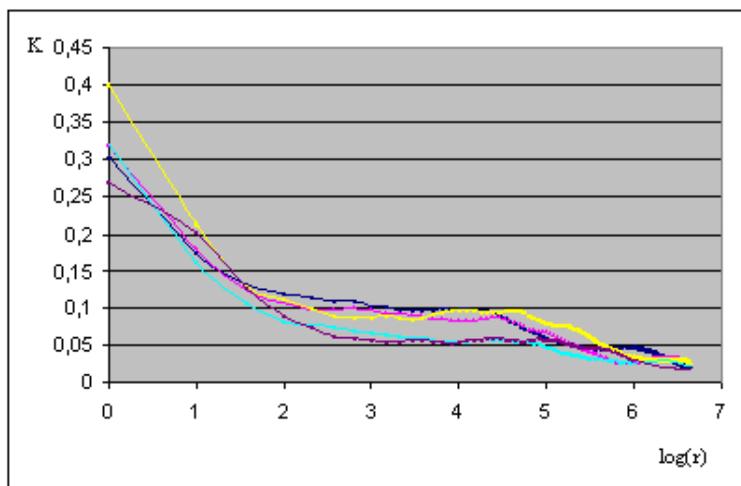


Рис. 2.24 а)

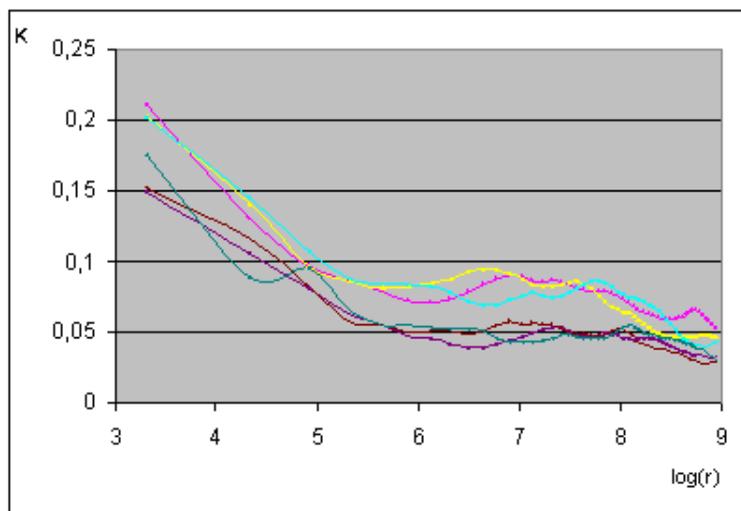


Рис. 2.24 б)

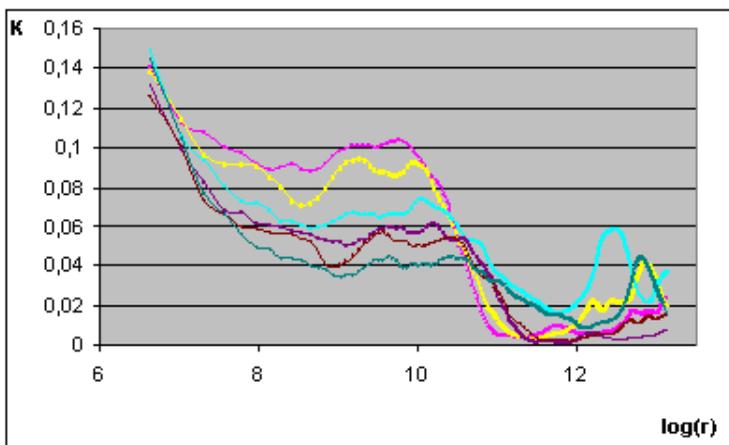


Рис. 2.24 в)

Рис. 2.24 Корреляционная энтропия:

- а) «Ростелеком» – 0,0231; б) «Лукойл» – 0,0326; в) «Сбербанк» – 0,0158.

На графиках хорошо видно, что есть несколько областей сходимости на уровнях примерно $K \approx 0,1$ и $K \approx 0,05$. Из этого следует что есть несколько уровней предсказания. Более точным является уровень краткосрочного предсказания не более 10 дней.



Рис. 2.25 а)



Рис. 2.25 б)



Рис. 2.25 в)

Рис. 2.25 Сходимость наибольшего показателя Ляпунова:
 а) «Ростелеком» – 0,0241; б) «Лукойл» – 0,0331; в) «Сбербанк» – 0,0161.

Наблюдая за сходимостью корреляционной энтропии видно, что у акций «Сбербанка» наблюдается наибольшая неоднородность. Это связано с особенностями ряда. Как можно увидеть из графика начальных данных в ряду имеется два периода. Первый с 1998-2001 гг. когда отчетливо виден параллельный тренд и второй период 2002-2003 гг. – восходящий тренд. Поэтому рассмотрим второй период, так как он является наиболее значимым, и вычислим для него корреляционную энтропию рисунок 2.26. Легко видеть что K сходится к

величине 0,0114. Что означает, что время предсказуемости равно $1/0,0114 = 87,7$ дней.

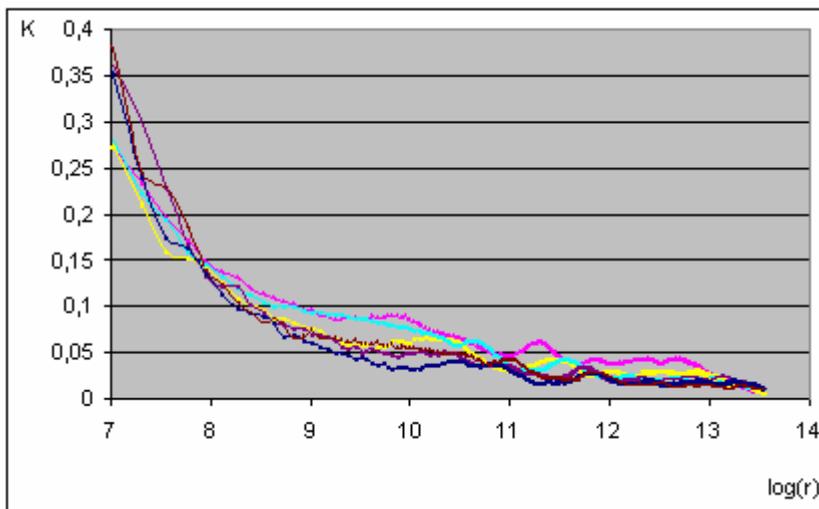


Рис. 2.26. корреляционная энтропия для курса акций «Сбербанка» последние 410 значений показателя

Однако, как было сказано ранее, на графике мы можем выделить несколько областей сходимости на уровне 0,05 и на уровне 0,02. Это позволяет сделать выводы и краткосрочном периоде предсказуемости исследуемого временного ряда.

Выводы к главе

1. Хаотические ряды только выглядят случайными, но, как детерминированный динамический процесс, вполне допускают краткосрочное прогнозирование. Подтверждением частичной предсказуемость финансовых временных рядов, являются проведенные исследования из которых можно сделать следующие выводы. Исследованные временные ряды имеют лептоэксцессное распределение (распределение с толстыми хвостами). В следствие того, что информация поступает на рынок редкими «порциями», а не непрерывно. Рыночная реакция на сгустки информации имеет следствием тяжелые хвосты. Поскольку распределение информации является лептоэксцессным, то распределение ценовых изменений также носит признаки

лептоэуsuccеса. Поскольку установлено, что прибыли акций не следуют нормальному распределению, то стандартное отклонение, как мера сравнения уровней риска неприемлема, а значит, не приемлема оценка риска различных ценных бумаг посредством оценки их волатильности.

2. Для определения временной задержки τ и восстановления аттрактора по временному ряду хорошие результаты удастся получить, используя спектральный (Фурье) анализ. Подтверждением сделанной оценки может служить показатель Херста.

3. Анализ восстановленных «странных аттракторов» курсов акций позволил выделить области притяжения и формирования цен. А рассчитанные фрактальные размерности позволили оценить количество переменных, которое необходимо для моделирования систем. Так для акций «РАО ЕЭС» $D = 2,36$, для акций «Ростелеком» $D = 2,11$, для акций «Лукойла» $D = 1,96$, для акций «Сбербанка» $D = 1,58$. Это означает, что смоделировать динамику этих акций можно с помощью трех и двух переменных. Таким образом, решена проблема определения количества переменных для моделирования сложных, динамических систем.

4. Временные ряды фондовой биржи являются фрактальными. Рассчитанный показатель Ляпунова, больший нуля, свидетельствует о том, что имеется чувствительная зависимость от начальных условий, в системе существует «странный аттрактор». Максимальная предсказуемость для курса акций «РАО ЕЭС» составляет примерно 8,605 дней, для курса акций «Ростелеком» составляет 8,064 дней, для курса акций «Лукойла» составляет 6,535 дней, для курса акций «Сбербанка» составляет 15,384 дней.

5. Для эффективного прогнозирования необходимо привлечение технологии нейронных сетей. Так как: нейросетевой анализ, в отличие от технического, не предполагает никаких ограничений на характер входной информации; в отличие от теханализа, основанного на общих рекомендациях, нейросети способны находить оптимальные индикаторы и строить по ним оптимальную стратегию предсказания; эти стратегии могут быть адаптивны, меняясь вместе с рынком, что особенно важно рынка России.

ГЛАВА 3 ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧАХ ПРОГНОЗИРОВАНИЯ И ПРОБЛЕМЫ ИДЕНТИФИКАЦИИ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ НА НЕЙРОННЫХ СЕТЯХ

Как было отмечено, жесткие статистические предположения о свойствах временных рядов ограничивают возможности классических методов прогнозирования. В данной ситуации адекватным аппаратом для решения задач прогнозирования могут служить специальные искусственные нейронные сети (НС) [22, 29, 36, 116, 121].

В этой связи возникает особо важная задача определения структуры и типа прогнозирующей нейронной сети. На сегодняшний день нет алгоритма или метода, позволяющего дать однозначный ответ на этот вопрос. Однако предложены способы настройки числа нейронов в процессе обучения, которые обеспечивают построение нейронной сети для решения задачи и дают возможность избежать избыточности [37]. Эти способы настройки можно разделить на две группы: конструктивные алгоритмы и алгоритмы сокращения.

В основе алгоритмов сокращения лежит принцип постепенного удаления из нейронной сети синапсов и нейронов. Вначале работы алгоритма обучения с сокращением число нейронов скрытых слоев сети заведомо избыточно.

Алгоритмы сокращения можно рассматривать как частный случай алгоритмов контрастирования [37]. Существуют два подхода к реализации алгоритмов сокращения метод штрафных функций и метод проекций. Для реализации первого в целевую функцию алгоритма обучения вводится штраф за то, что значения синаптических весов отличны от нуля; пример штрафа

$$C = \sum_{i=1}^N \sum_{j=1}^M w_{ij}^2 \quad (3.1)$$

где w_{ij}^2 – синаптический вес, i – номер нейрона, j – номер входа, N – число нейронов, M – размерность входного сигнала нейронов;

Метод проекций реализуется следующим образом. Синаптический вес обнуляется, если его значение попало в заданный диапазон

$$w_{ij} = \begin{cases} 0, & -\varepsilon \leq w_{ij} \leq \varepsilon \\ w_{ij}, & w_{ij} < -\varepsilon, w_{ij} > \varepsilon \end{cases} \quad (3.2)$$

где ε – некоторая константа.

Алгоритмы сокращения имеют по крайней мере два недостатка:

1. Нет методики определения числа нейронов скрытых слоев, которое является избыточным, поэтому перед началом работы алгоритма нужно угадать это число.

2. В процессе работы алгоритма сеть содержит избыточное число нейронов, поэтому обучение идет медленно.

Предшественником конструктивных алгоритмов можно считать методику обучения многослойных сетей, включающую в себя следующие шаги:

1. Выбор начального числа нейронов в скрытых слоях.
2. Инициализация сети, то есть присваивание синаптическим весам смещениям сети случайных значений из заданного диапазона.
3. Обучение сети по заданной выборке.
4. Завершение в случае успешного обучения; если сеть обучить не удалось, то число нейронов увеличивается и повторяются шаги со второго по четвертый.

В конструктивных алгоритмах число нейронов в скрытых слоях также изначально мало и постепенно увеличивается. В отличие от описанной методики, в конструктивных алгоритмах сохраняются навыки, приобретенные сетью до увеличения числа нейронов.

Конструктивные алгоритмы различаются правилами задания значений параметров в новых – добавленных в сеть – нейронах:

1. значения параметров – случайные числа из заданного диапазона;
2. значения синаптических весов нового нейрона определяются путем расщепления (splitting) одного из старых нейронов.

Первое правило не требует значительных вычислений, однако его использование приводит к некоторому увеличению значения функции ошибки после каждого добавления нового нейрона. В результате случайного задания значений параметров новых нейронов может появиться избыточность в числе нейронов скрытого слоя. Расщепление нейронов лишено двух указанных недостатков.

Самым большим недостатком алгоритма является экспоненциальный рост времени вычислений при увеличении размерности сети. Для преодоления указанного недостатка предлагается упрощенный алгоритм расщепления, который не требует значительных вычислений.

Кроме описанных способов выбора нейронов для расщепления, может быть использован анализ чувствительности [37], в процессе которого строятся матрицы Гессе - матрицы вторых производных функции ошибки по параметрам сети. По величине модуля второй производной судят о важности значения данного параметра для решения задачи. Параметры с малыми значениями вторых производных обнуляют. Анализ чувствительности имеет большую вычислительную сложность и требует много дополнительной памяти.

В исследовании рассмотрены два типа нейронных сетей: многослойный персептрон и сеть типа радиально-базисная функция.

3.1 Сравнительный анализ нейронных сетей

Актуальность данной тематики продиктована поиском адекватных моделей нейронных сетей (НС), определяемые типом и структурой НС, для задач прогнозирования. В ходе исследования установлено, что радиальные базисные сети (RBF) обладают рядом преимуществ перед сетями типа многослойных персептрон (MLP) [121, 123]. Во-первых, они моделируют произвольную нелинейную функцию с помощью одного промежуточного слоя. Тем самым отпадает вопрос о числе слоев. Во-вторых, параметры линейной комбинации в выходном слое можно полностью оптимизировать с помощью известных методов моделирования, которые не испытывают трудностей с локальными минимумами, мешающими при обучении MLP. Поэтому сеть RBF обучается очень быстро (на порядок быстрее MLP) [85,107].

С другой стороны, до того как применять линейную оптимизацию в выходном слое сети RBF, необходимо определить число радиальных элементов, положение их центров и величины отклонений. Для устранения этой проблемы предлагается использовать автоматизированный конструктор сети, который выполняет за пользователя основные эксперименты с сетью.

Другие отличия работы RBF от MLP связаны с различным представлением пространства модели: «групповым» в RBF и «пло-

скостным» в MLP. Опыт показывает, что для правильного моделирования типичной функции сеть RBF, с ее более эксцентричной поверхностью отклика, требует несколько большего числа элементов. Следовательно, модель, основанная на RBF, будет работать медленнее и потребует больше памяти, чем соответствующий MLP (однако она гораздо быстрее обучается, а в некоторых случаях это важнее) [92].

С «групповым» подходом связано и неумение сетей RBF экстраполировать свои выводы за область известных данных. При удалении от обучающего множества значение функции отклика быстро падает до нуля. Напротив, сеть MLP выдает более определенные решения при обработке сильно отклоняющихся данных, однако, в целом, склонность MLP к некритическому экстраполированию результата считается его слабостью. Сети RBF более чувствительны к «проклятию размерности» и испытывают значительные трудности, когда число входов велико.

Для оценки точности и адекватности результатов прогнозирования, а также структуры нейронной сети использовались следующие статистические показатели [122]:

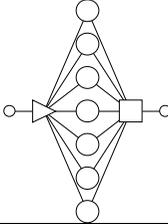
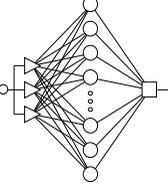
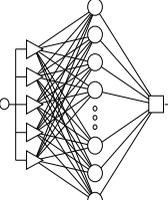
1. Data Mean. – среднее значение целевой выходной переменной;
2. Data S.D. – среднеквадратическое отклонение целевой выходной переменной;
3. Error Mean – средняя ошибка выходной переменной (остаток между целевой и реальной переменной);
4. Abs. E. Mean – средняя абсолютная ошибка (разница между целевой и реальной выходной переменной);
5. Error S.D. – стандартное отклонение ошибки выходной переменной;
6. S.D. Ratio – среднеквадратическое отклонение ошибок выходной переменной;
7. Correlation – коэффициент корреляции Спирмена вычисленный между целевым вектором и реальным выходным вектором.

Исследования проводились в пакете STATISTICA Neural Networks 4.0.В для курса акций «РАО ЕЭС» [92]. Каждая таблица показывает найденные типы нейронных структур для исследуемого временного ряда. В первом столбце таблиц стоит значение лага, с которым данные подаются на вход НС. Во втором столбце указано коли-

чество проведенных испытаний, следствием которых стал выбор наилучшей, по всем характеристикам, НС.

Таблица 3.1

Нейронные сети типа RBF

Лег	Количество испытаний	Параметры отобранной НС	Вид лучшей НС	Характеристики
		Тип НС Inputs: Hidden:Outputs		
1	16	RBF 1:7:1 TPerf, VPerf, TePerf 0,1198; 0,1174; 0,1321		Найден адекватный тип сети (регрессионное отношение 0,08168, корреляция 0,995359, среднеквадратическая ошибка предсказания 0,1172432)
3	24	RBF 3:14:1 TPerf, VPerf, TePerf 0,1234; 0,1215; 0,1315		Найден адекватный тип сети (регрессионное отношение 0,089525, корреляция 0,995992, среднеквадратическая ошибка предсказания 0,1199467)
5	21	RBF 5:14:1 TPerf, VPerf, TePerf 0,1241; 0,195; 0,1354		Найден адекватный тип сети (регрессионное отношение 0,095182, корреляция 0,995465, среднеквадратическая ошибка предсказания 0,1403166)

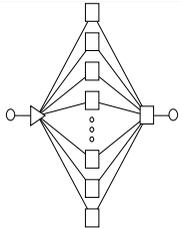
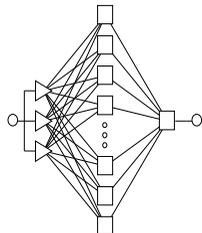
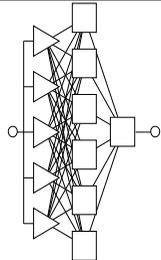
В третьем столбце указаны основные характеристики НС: Тип сети RBF или MPL; внутренняя структура (например 1:7:1 означает, что сеть имеет один входной нейрон, 7 нейронов скрытого слоя и один выходной нейрон), среднеквадратическая ошибка предсказания (TPerf, VPerf, TePerf) для обучающей, валидационной и тестовой выборок.

В четвертом столбце таблиц показан внешний вид НС. В пятом столбце указаны дополнительные характеристики НС, такие как рег-

рессионное отношение, корреляция, среднеквадратическая ошибка предсказания предсказания НС.

Таблица 3.2

Нейронные сети типа MLP (трехслойная сеть)

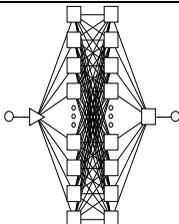
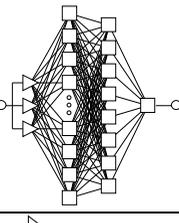
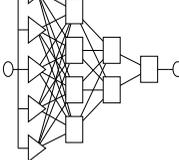
Лег	Количество испытаний	Параметры отобранной НС	Вид лучшей НС	Характеристики
		Тип НС Inputs: Hidden:Outputs		
1	16	MLP 1:10:1 TPerf, VPerf, TePerf 0,1221; 0,1193; 0,1328		Найден адекватный тип сети (регрессионное отношение 0.083055, корреляция 0.996550, среднеквадратическая ошибка предсказания 0.1191367)
3	20	MLP 3:12:1 TPerf, VPerf, TePerf 0,127; 0,125; 0,134		Найден адекватный тип сети (регрессионное отношение 0.085872, корреляция 0.996326, среднеквадратическая ошибка предсказания 0.1253478)
5	25	MLP 5:7:1 TPerf, VPerf, TePerf 0,1227; 0,1203; 0,1313		Найден адекватный тип сети (регрессионное отношение 0.095182, корреляция 0.995465, среднеквадратическая ошибка предсказания 0.1403166)

Можно сделать следующие выводы. Каждый из двух описанных подходов имеет свои достоинства и недостатки. Действие радиальных функций очень локально, в то время как при линейном подходе охватывается все пространство входов. Поэтому, как правило, RBF-сети имеют больше элементов, чем MLP-сети, однако MLP мо-

жет делать необоснованные обобщения в ситуациях, когда ему попадает набор данных, непохожий ни на какие наборы из обучающего множества, в то время как RBF в таком случае всегда будет выдавать почти нулевой отклик [91, 92].

Таблица 3.3

Нейронные сети типа MLP (четырёхслойная сеть)

Лег	Количество испытаний	Параметры отобранной НС	Вид лучшей НС	Характеристики
		Тип НС Inputs: Hidden:Outputs		
1	16	MLP 1:13:13:1 TPerf, VPerf, TePerf 0,1227; 0,1197; 0,133		Найден адекватный тип сети (регрессионное отношение 0.083461, корреляция 0.996511, среднеквадратическая ошибка предсказания 0.1180339)
3	16	MLP 3:13:8:1 TPerf, VPerf, TePerf 0,1233; 0,1207; 0,1338		Найден адекватный тип сети (регрессионное отношение 0.084064, корреляция 0.996471, среднеквадратическая ошибка предсказания 0.1207132)
5	16	MLP 5:4:2:1 TPerf, VPerf, TePerf 0,1227; 0,1194; 0,1325		Найден адекватный тип сети (регрессионное отношение 0.083177, корреляция 0.996543, среднеквадратическая ошибка предсказания 0,1194306)

Анализируя результаты предварительных исследований таблицы 3.1 – 3.3 можно заключить, что сеть типа RBF (1:7:1) таблица 3.1 имеет большее предпочтение. Об этом можно судить по регрессионному отношению и среднеквадратической ошибке предсказания. TPerf, VPerf, TePerf показывает ошибку, которая вычисляется как корень квадратный из среднего значения ошибки на каждом шаге обучения для обучающей, верификационной и тестовой выборок.

3.2 Об отрицательном влиянии процедуры перемешивания данных (Shuffle) при обучении нейронных сетей, применяемых для прогнозирования

Как было показано, процедура перемешивания данных приводит к потере долговременной памяти внутри временного ряда. Проведенные эксперименты наглядно это доказывают [117]. Производились вычислительные эксперименты с последними 185 значениями курса акций «РАО ЕЭС» в период с 23.09.2002 по 24.06.2003 гг. Они показали, что для не перемешанного ряда возможно определить оптимальную структуру прогнозирующей нейронной сети. Основные экспериментальные параметры и показатели приведены в таблицах 4 – 5 и показаны на рисунках 3.1 – 3.2.

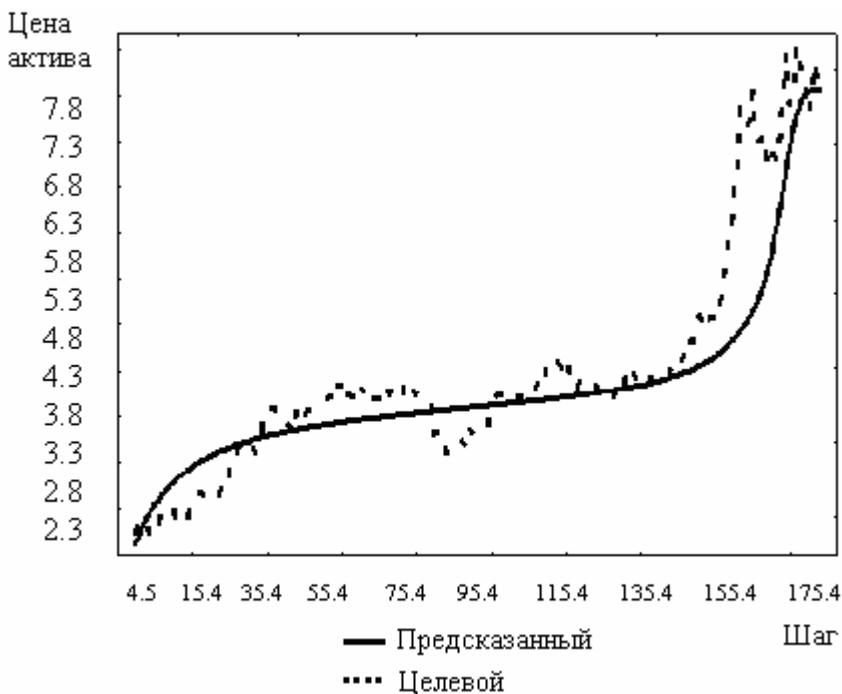
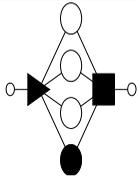


Рис. 3.1 Прогноз курса акций «РАО ЕЭС», полученного нейронной сетью для последних 185 не перемешанных значений (Predicted – предсказанный и Target – целевой временные ряды)

Исходный временной ряд (23.09.2002 по 24.06.2003 гг. 185 значений) для RBF сети

Наименование временного ряда	Количество испытаний	Параметры отобранных НС			Вид лучшей НС	Характеристики
		Тип НС	Inputs: Hidden:Outputs	Error		
1. Динамика курса акций российской компании «РАО ЕЭС» в период с 23.09.2002 по 24.06.2003 гг.	287 Отобрано 2 различных сети с лучшими показателями	RB	1:1-5-1:1(3 layer)	0,174		Найден адекватный тип сети для данного временного ряда (регрессионное отношение 0.121674, корреляция 0.992602)
		F		6		
		RB	1:1-4-1:1(3 layer)	0,173		

RBF-радиально базисная функция

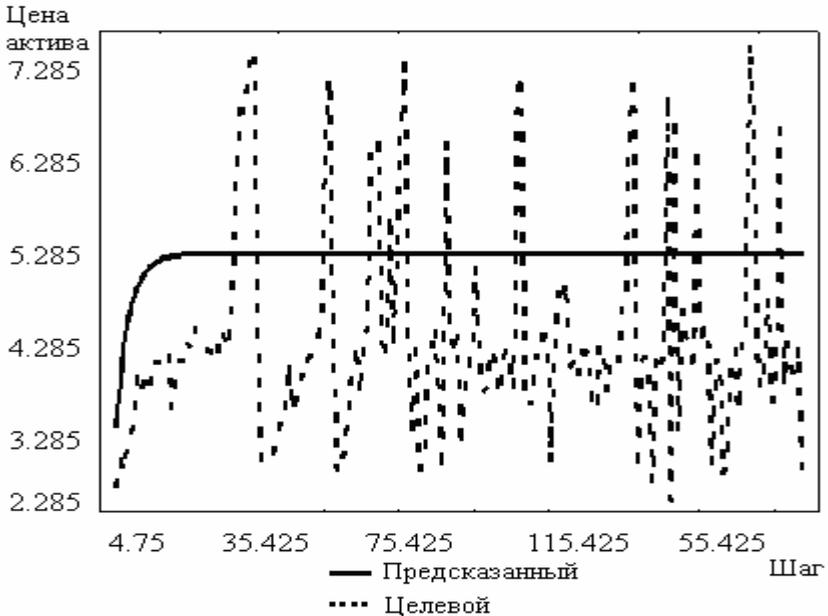
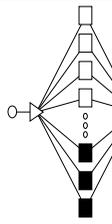


Рис. 3.2 Прогноз курса акций «РАО ЕЭС», полученного нейронной сетью для последних 185 перемешанных значений (Predicted – предсказанный и Target – целевой временные ряды)

Перемешанный ряд (23.09.2002 по 24.06.2003 гг. 185 значений)
для MLP сети

Наименование временного ряда	Количество испытаний	Параметры отобранных НС			Вид лучшей НС	Характеристики
		Тип НС	Inputs: Hidden:Outputs	Error		
1. Динамика курса акций российской компании «РАО ЕЭС» в период с 23.09.2002 по 24.06.2003 гг. Перемешанный ряд	285 Отобрано 2 различных сети с лучшими показателями	ML	1:38:1	1,27		Адекватный тип сети для данного временного ряда не найден
		P	1:58:1	1		
		ML P		1,28 1		

MLP-многослойный перцептрон

Легко видеть (см. таблицу 3.5 и рисунок 3.2), что перемешивание данных при обучении нейронных сетей не позволяет определить оптимальной структуры сети и получить адекватных прогнозов, что также подтверждено расчетным показателем Херста [98, 117].

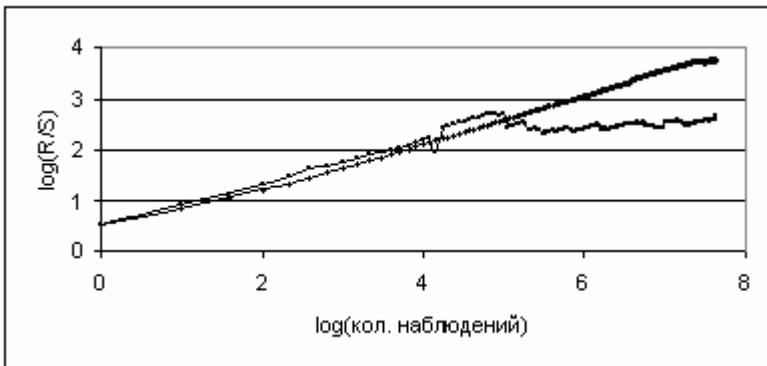


Рис. 3.3 В двойных логарифмических координатах представлены перемешанный и не перемешанный ряды курса акций «РАО ЕЭС» для последних 185 значений

На рисунке 3.3 в двойных логарифмических координатах представлены неперемешанный ряд и тот же ряд – перемешанный для последних 185 значений исходного ряда. Исходный ряд дал результативную оценку показателя Херста $H = 0.87$, перемешанный – $H = 0.41$. Такое падение величины H говорит о том, что при перемешивании была разрушена структура процесса (внутренняя память). Перемешанный ряд остался не нормально распределенным, но процесс перемешивания сделал данные независимыми.

3.3 Исследование нейросетевых структур для курсов акций «Ростелеком», «Лукойл», «Сбербанк»

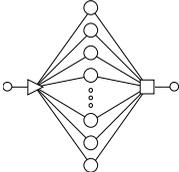
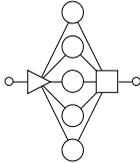
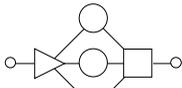
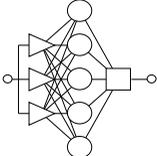
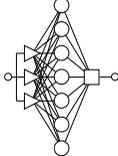
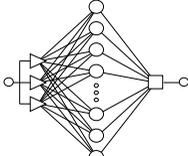
Исследования проводились для рассмотренных временных рядов. Все результаты исследований также сведены в таблицы. В таблице 6 показаны результаты поиска оптимальных нейросетевых структур для курсов акций «Ростелеком», «Лукойл», «Сбербанк». Исследования проводились для сетей типа RBF.

В таблицах 3.7 – 3.8 показаны результаты поиска оптимальных нейросетевых структур для тех же временных рядов. Исследования проводились для сетей типа MLP (трех и четырехслойная сеть). Известно, что поиск типа нейронной сети и структуры достаточно трудоемкая процедура, поэтому решалась промежуточная задача определения начального «прототипа», после определения которого, велось экспериментальное уточнение дальнейшей структуры НС.

Выводы по результатам вычислительных экспериментов: Анализируя таблицы 3.6 – 3.8 для курса акций «Ростелеком» выбрана оптимальная структура нейронной сети со следующими показателями: тип – MLP; трехслойная структура: 3 входных нейронов – 6 нейронов скрытого слоя – 1 выходной нейрон (таблица 3.7). Однако возможно использование сети типа – MLP; четырехслойная структура: 1 входных нейрона – 10 нейронов скрытого слоя – 1 выходной нейрон (таблица 3.7).

Таблица 3.6

Нейронные сети типа RBF

Лаг	Параметры отобранной НС Для акций «Ростелеком»	Параметры отобранной НС для акций «Лукойл»	Параметры отобранной НС для акций «Сбербанк» (для последних 405 знач.)
1	<p>8 тестов</p>  <p>RBF 1:11:1 TPerf, VPerf, TePerf 2,054; 1,074; 1,266 (регрессионное отношение 0.114370, корреляция 0.993442, среднеквадрати- ческая ошибка предска- зания 1.073946)</p>	<p>15 тестов</p>  <p>RBF 1:5:1 TPerf, VPerf, TePerf 9.223; 13.890; 12.320 (регрессионное отноше- ние 0.181738, корреляция 0.983840, среднеквадра- тическая ошибка пред- сказания 13.88914) Сеть претерпела пере- обучение. Нужно увели- чить обучающую выбор- ку.</p>	<p>13 тестов</p>  <p>RBF 1:3:1 TPerf, VPerf, TePerf 110.8; 103.7; 152.0 (регрессионное отношение 0.190076, корреляция 0.981947, среднеквадрати- ческая ошибка предсказания 103.6565)</p>
3	<p>8 тестов</p>  <p>RBF 3:5:1 TPerf, VPerf, TePerf 2,483; 1,263; 1,35 (регрессионное отношение 0.134435, корреляция 0.990930, среднеквадрати- ческая ошибка предсказа- ния 1.262944)</p>	<p>8 тестов</p>  <p>RBF 3:7:1 TPerf, VPerf, TePerf 11,06; 12,95; 13,49 (регрессионное отноше- ние 0.157377, корреляция 0.987700, среднеквадра- тическая ошибка пред- сказания 14.08784) Сеть претерпела пере- обучение. Нужно увели- чить обучающую выбор- ку.</p>	<p>24 тестов</p>  <p>RBF 3:22:1 TPerf, VPerf, TePerf 120.7; 124.3; 2076.0 (регрессионное отношение 0.223885v, корреляция 0.974682, среднеквадрати- ческая ошибка предсказания 124.2555)</p>

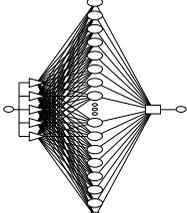
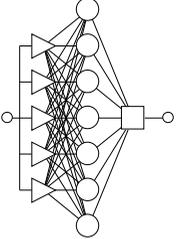
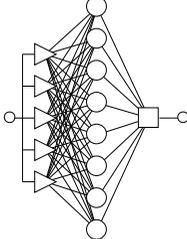
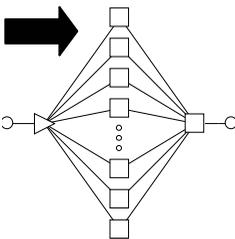
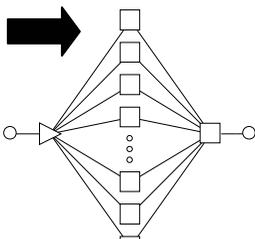
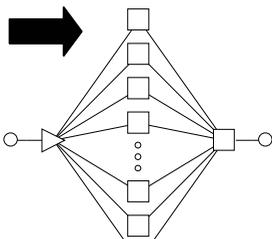
<p>5</p>	<p>27 тестов</p>  <p>RBF 5:44:1 TPerf, VPerf, TePerf 2,484; 1,858; 2,916 (регрессионное отношение 0.098558, корреляция 0.995316, среднеквадрати- ческая ошибка предска- зания 1,857529)</p>	<p>16 тестов</p>  <p>RBF 5:7:1 TPerf, VPerf, TePerf 12.10; 14.57; 303.80 (регрессионное отноше- ние 0.265997, корреляция 0.964260, среднеквадра- тическая ошибка пред- сказания 14.57255)</p>	<p>15 тестов</p>  <p>RBF 5:8:1 TPerf, VPerf, TePerf 120.6; 135.6; 5162.0 (регрессионное отношение 0.251100, корреляция 0.968655, среднеквадрати- ческая ошибка предсказания 135.5618)</p>
----------	---	---	---

Таблица 3.7

Нейронные сети типа MLP (трехслойная сеть)

Ла г	Параметры отобранной НС для акций «Ростелеком»	Параметры отобранной НС для акций «Лукойл»	Параметры отобранной НС для акций «Сбербанк» (для последних 405 знач.)
<p>1</p>	<p>16 тестов</p>  <p>MLP 1:10:1 TPerf, VPerf, TePerf 2,128; 1,07; 1,201 (регрессионное отноше- ние 0.113987, корреля- ция 0.993490, средне- квадратическая ошибка предсказания 1.070176)</p>	<p>16 тестов</p>  <p>MLP 1:11:1 TPerf, VPerf, TePerf 9.217; 10.580; 12.610 (регрессионное отношение 0.195551, корреляция 0.980696, среднеквадрати- ческая ошибка предсказа- ния 10.58064)</p>	<p>69 тестов</p>  <p>MLP 1:47:1 TPerf, VPerf, TePerf 106,7; 97.19; 188,5 (регрессионное отношение 0.179922, корреляция 0.983726, среднеквадрати- ческая ошибка предсказания 97.18809)</p>

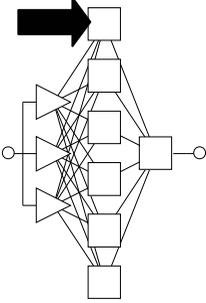
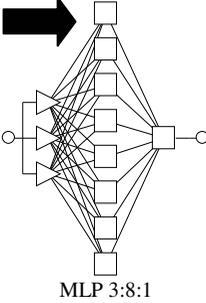
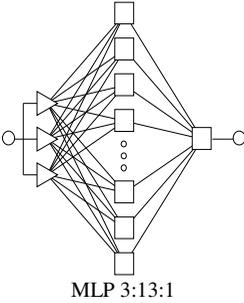
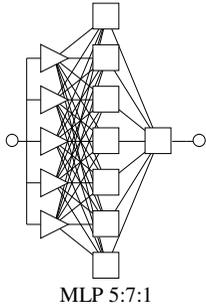
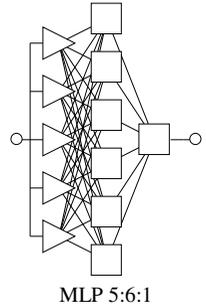
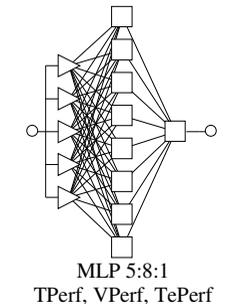
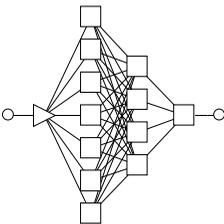
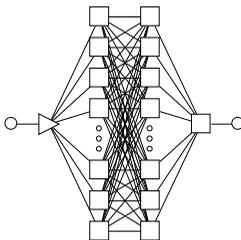
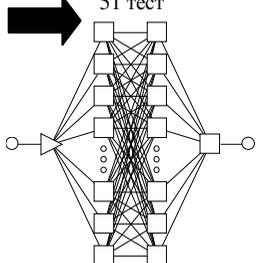
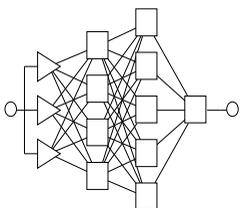
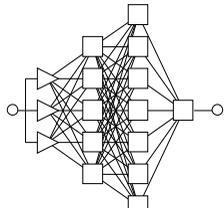
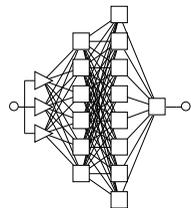
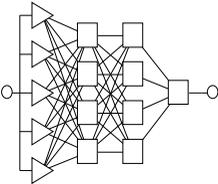
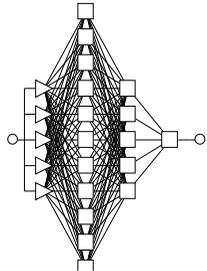
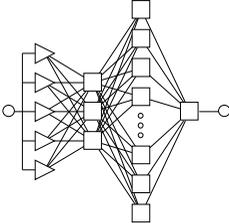
<p>3</p>	<p>16 тестов</p>  <p>MLP 3:6:1 TPerf, VPerf, TePerf 2.104; 1.7; 1.529 (регрессионное отношение 0.085128, корреляция 0.996371, среднеквадратическая ошибка предсказания 1.699702)</p>	<p>16 тестов</p>  <p>MLP 3:8:1 TPerf, VPerf, TePerf 9,259; 10,570; 11,840 (регрессионное отношение 0.195361, корреляция 0.980733, среднеквадратическая ошибка предсказания 10.5714)</p>	<p>16 тестов</p>  <p>MLP 3:13:1 TPerf, VPerf, TePerf 108.9; 98.74; 248.4 (регрессионное отношение 0.179700, корреляция 0.983764, среднеквадратическая ошибка предсказания 98.74388)</p>
<p>5</p>	<p>16 тестов</p>  <p>MLP 5:7:1 TPerf, VPerf, TePerf 2.022; 1.233; 1.200 (регрессионное отношение 0.187880, корреляция 0.982552, среднеквадратическая ошибка предсказания 1.232779)</p>	<p>16 тестов</p>  <p>MLP 5:6:1 TPerf, VPerf, TePerf 9,248; 10,510; 13,850 (регрессионное отношение 0.194242, корреляция 0.981016, среднеквадратическая ошибка предсказания 10.50778)</p>	<p>16 тестов</p>  <p>MLP 5:8:1 TPerf, VPerf, TePerf 117.1; 103.3; 295.6 (регрессионное отношение 0.187530, корреляция 0.982329, среднеквадратическая ошибка предсказания 103.2507)</p>

Таблица 3.8

Нейронные сети типа MLP (четырёхслойная сеть)

Лег	Параметры отобранной НС для акций «Ростелеком»	Параметры отобранной НС для акций «Лукойл»	Параметры отобранной НС для акций «Сбербанк» (для последних 405 знач.)
1	<p>16 тестов</p>  <p>MLP 1:7:4:1 TPerf, VPerf, TePerf 2.051; 1.248; 1.202 (регрессионное отношение 0.190008, корреляция 0.982021, среднеквадратическая ошибка предсказания 1.247681)</p>	<p>16 тестов</p>  <p>MLP 1:13:13:1 TPerf, VPerf, TePerf 9,337; 10,600; 16,120 (регрессионное отношение 0.195906, корреляция 0.980633, среднеквадратическая ошибка предсказания 10.60465)</p>	<p>51 тест</p>  <p>MLP 1:15:13:1 TPerf, VPerf, TePerf 106,6; 98,1; 144,2 (регрессионное отношение 0.181767, корреляция 0.983668, среднеквадратическая ошибка предсказания 98.09992)</p>
3	<p>16 тестов</p>  <p>MLP 3:4:5:1 TPerf, VPerf, TePerf 2.029; 1.255; 1.200 (регрессионное отношение 0.190985, корреляция 0.982249, среднеквадратическая ошибка предсказания 1.254811)</p>	<p>16 тестов</p>  <p>MLP 3:5:7:1 TPerf, VPerf, TePerf 9,382; 10,570; 14,370 (регрессионное отношение 0.195302, корреляция 0.980751, среднеквадратическая ошибка предсказания 10.56612)</p>	<p>32 тестов</p>  <p>MLP 3:6:8:1 TPerf, VPerf, TePerf 112.3; 99.66; 240.7 (регрессионное отношение 0.181885, корреляция 0.983326, среднеквадратическая ошибка предсказания 99.6591)</p>

<p>5</p>	<p>16 тестов</p>  <p>MLP 5:4:4:1 TPerf, VPerf, TePerf 2.105; 1.231; 1.243 (регрессионное отношение 0.187464, корреляция 0.982542, среднеквадратическая ошибка предсказания 1.230627)</p>	<p>16 тестов</p>  <p>MLP 5:11:5:1 TPerf, VPerf, TePerf 9,256; 10,580; 17,210 (регрессионное отношение 0.195390, корреляция 0.980889, среднеквадратическая ошибка предсказания 10.57533)</p>	<p>16 тестов</p>  <p>MLP 5:3:9:1 TPerf, VPerf, TePerf 126.6; 102.7; 263.4 (регрессионное отношение 0.189757, корреляция 0.982160, среднеквадратическая ошибка предсказания 102.6798)</p>
----------	--	---	---

На рисунке 3.4 показано изменение ошибки прогнозирования в процессе обучения нейронной сети для курса акций «Ростелеком», которая составляет в среднем 1,5-2%. Выброс ошибок на интервале 600-800 вызван значительным ростом цен в период 2000 года.

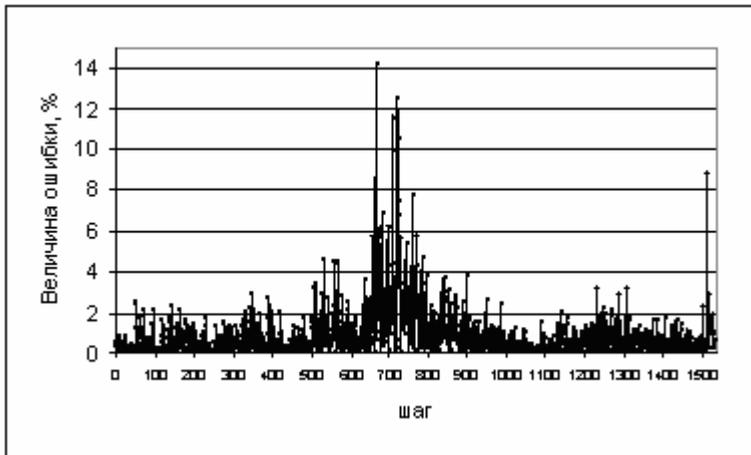


Рис. 3.4 Изменение ошибки прогнозирования в процессе обучения нейронной сети для курса акций «Ростелеком»

Для курса акций «Лукойл» выбрана оптимальная структура нейронной сети со следующими показателями: тип – MLP; трехслойная структура: 3 входных нейрона – 8 нейронов скрытого слоя – 1 выходной нейрон (таблица 3.7). Однако возможно использование сети типа – MLP; трехслойная структура: 1 входных нейрона – 11 нейронов скрытого слоя – 1 выходной нейрон (таблица 3.7). На рисунке 3.5 показано изменение ошибки прогнозирования, которая составляет в среднем 3%.

Для эмитента – «Сбербанк» проанализировав общую тенденцию развития можно выделить два временных интервала 1998-2001 и конец 2001 – 2003. Поэтому в целях получения адекватных прогнозов рассмотрен второй период (примерно 405 значений временного ряда).

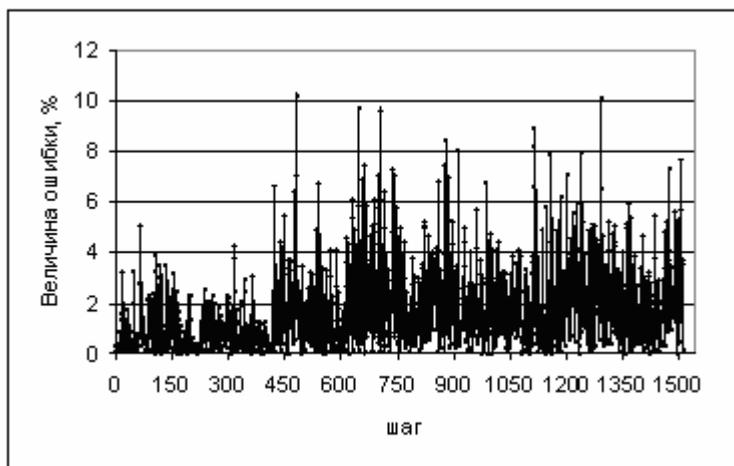


Рис. 3.5 Изменение ошибки прогнозирования в процессе обучения нейронной сети для курса акций «Лукойл»

Выбрана оптимальная структура нейронной сети со следующими показателями: тип – MLP; трехслойная структура: 1 входных нейрона – 13 нейронов скрытого слоя – 1 выходной нейрон (таблица 3.7).

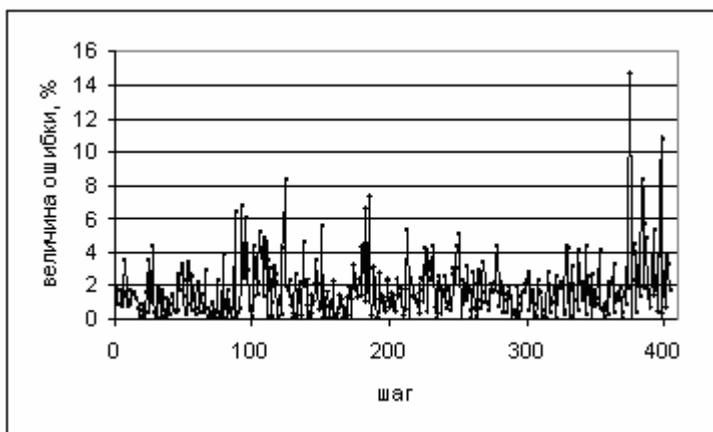


Рис. 3.6. Изменение ошибки прогнозирования в процессе обучения нейронной сети для курса акций «Сбербанк»

Однако возможно использование сети типа – MLP; четырех-слойная структура: 1 входных нейрона – 15 нейронов 1-го скрытого слоя – 13 нейронов 2-го скрытого слоя – 1 выходной нейрон (таблица 3.8). На рисунке 3.6 показано изменение ошибки прогнозирования, которая составляет в среднем 2%.

Можно видеть, что наблюдается рост величины ошибки для последних тестовых значений временного ряда. Это свидетельствует о том, что величина обучающей выборки имеет не достаточное количество данных, однако такой рост ошибки не приведет к росту ошибок прогнозирования НС и составит в среднем 3-7%.

3.4 Многофакторные нейронные сети

Анализируя динамику рынка акций среди них можно выделить так называемые «голубые фишки», которые названы так потому, что определяют динамику поведения всего рынка и наиболее прибыльны. Многофакторные нейронные сети позволяют не только прогнозировать курс акций, но и учитывать влияние на цену других ценных бумаг, факторов [131]. Кроме того, возможно в качестве входной информации (обучающего множества) подавать на вход сети различные типы биржевых индексов, а также индексов применяемых в техническом анализе, применяемых большинством биржевых трейдеров.

Хорошим примером могут служить разно-периодные скользящие средние. Совокупное использование которых, служит индексом разворота рынка.

В таблице 3.9 показаны полученные структуры трехслойных и четырехслойных нейронных сетей. На вход такой сети подавались с лагом равным 3 значения курсов акций «РАО ЕЭС», «Ростелеком» и «Лукойл», на выходе сети т.е. обучающим множеством были акции «РАО ЕЭС». Таким образом прогнозируя курс акций «РАО ЕЭС» мы сможем учитывать изменения цен други ценных бумаг. На рисунке 3.7 показано изменение относительной ошибки предсказания на обучающей, валидационной, тестовой выборках многофакторной нейронной сети для курса акций «РАО ЕЭС», которая составляет в среднем 3%.

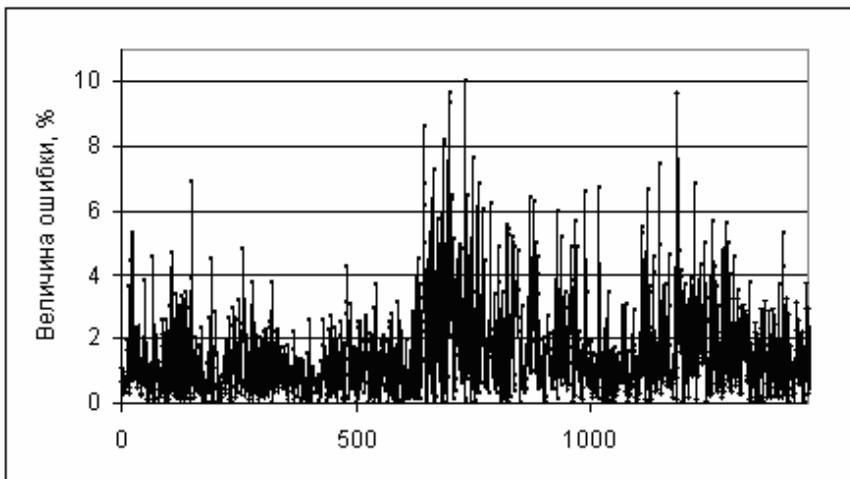
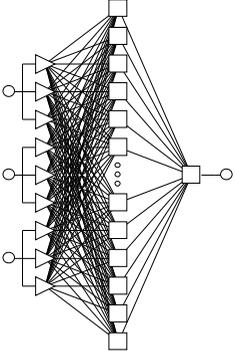
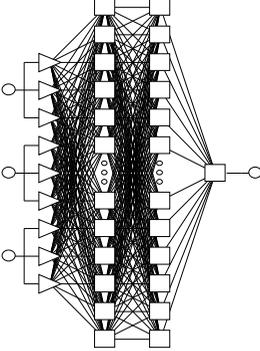
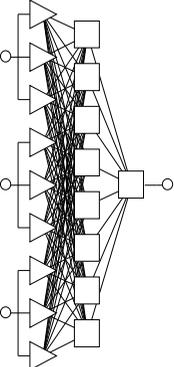
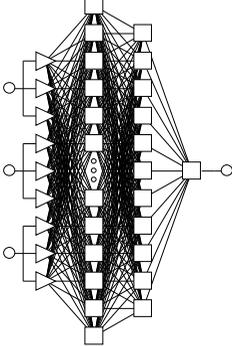


Рис. 3.7 Изменение ошибки прогнозирования в процессе обучения многофакторной нейронной сети для курса акций «РАО ЕЭС»

Результаты исследования многофакторных НС

Трехслойная нейронная сеть	Четырехслойная нейронная сеть
 <p data-bbox="244 655 530 770"> MLP 9:16:1 TPerf, VPerf, TePerf 0,1183; 0,1313; 0,1928 среднеквадратическая ошибка предсказания 0.1312699 </p>	 <p data-bbox="666 655 997 770"> MLP 9:16:16:1 TPerf, VPerf, TePerf 0,1217; 0,1345; 0,2615 среднеквадратическая ошибка пред- сказания 0.1345437 </p>
 <p data-bbox="247 1158 530 1273"> MLP 9:8:1 TPerf, VPerf, TePerf 0,1204; 0,1332; 0,2448 среднеквадратическая ошибка предсказания 0.1332476 </p>	 <p data-bbox="666 1134 997 1249"> MLP 9:16:11:1 TPerf, VPerf, TePerf 0,1244; 0,1285; 0,1817 среднеквадратическая ошибка пред- сказания 0.1285486 </p>

3.5 Разработка математических моделей непозиционного кодирования для решения задач прогнозирования на нейронных сетях

Как было отмечено адекватным аппаратом для решения задач прогнозирования и создания устойчивых необслуживаемых систем управления, диагностики, контроля и т.п., могут служить специальные искусственные нейронные сети (НС). Однако одним из важных критериев работы прогнозирующей системы на нейронных сетях является ее надежность и устойчивость к отказам элементов нейронной сети. Исследование отказоустойчивости НС прямого распространения показало, что параллельная организация является одним из механизмов, обеспечивающим надежность нейронных структур, но запас устойчивости очень мал и при реализации возникает ряд проблем:

Аппаратная реализация НС затруднена при росте количества скрытых слоев или нейронов в отдельном слое.

Отсутствует возможность оперативного контроля количества неисправных нейронов: при неисправности нейрона ошибка не проявляется на выходе.

Отсутствует возможность оперативной замены отказавших элементов сети.

При использовании алгоритмов самообучения НС в условиях возникновения неисправностей увеличивается вероятность потери устойчивости в петле обучения.

При обучении НС становится актуальной проблема связанная с явлением переобучения. Кроме того, использование не масштабированных данных может привести к «параличу» сети.

Поиски новых путей повышения эффективности выполнения арифметических операций привели к заключению, что в рамках обычной позиционной системы значительного ускорения выполнения операций и надежности добиться почти невозможно [2]. Поэтому с развитием теории нейронных сетей, в частности решения задачи отображения модулярных вычислений на нейронные структуры [137, 136], появилась возможность использовать преимущества модулярного кодирования для решения различных задач, в том числе и задач прогнозирования. Система остаточных классов обладает рядом существенных преимуществ, так как является непозиционной системой счисления, которая позволяет сократить разрядность обрабатываемых данных, представляя их в модулярном коде в системе взаимно

простых модулей, обладает способностью к обнаружению ошибок и самокоррекции.

Основной Системы остаточных классов (СОК) является теория сравнений. СОК дает нестандартное представление чисел и используется для повышения эффективности операций над кодами в остатках [3, 5].

Число A в модулярном коде представляется в виде набора $(\alpha_1, \alpha_2, \dots, \alpha_k)$ k остатков (вычетов), вычисленных по модулю каждого из оснований непозиционной системы счисления p_i

$$\alpha_i = |A|_{p_i} = \left[\frac{A}{p_i} \right] \cdot p_i, \quad \text{для } i = 1, 2, \dots, k. \quad (3.3)$$

где: $[]$ – целая часть числа, k - количество оснований. В отличие от позиционной системы счисления (ПСС), в которой число A представляется в виде

$$A = A_n N^n + A_{n-1} N^{n-1} + \dots + A_1 N^1 + A_0 N^0 = \sum_{i=0}^n A_i N^i, \quad (3.4)$$

где N – основание ПСС, значение числа в модулярном коде не зависит от местоположения каждого разряда в его представлении, а зависит от значения основания соответствующего разряда. Поэтому модулярный код является непозиционным.

Если основания в представлении p_1, p_2, \dots, p_k выбрать попарно взаимно простыми, то, согласно Китайской теореме об остатках (КТО), это обеспечивает однозначность представления любых целых чисел из диапазона $[0, P)$

$$P = \prod_{i=1}^k p_i. \quad (3.5)$$

В литературе [3, 5] такую непозиционную систему счисления называют системой остаточных классов (СОК). СОК обеспечивает возможность выполнения операций сложения и умножения с высокой скоростью. Образование остатков от числа в СОК производится независимо друг от друга, вследствие чего, каждый разряд содержит

в себе информацию обо всем числе. При выполнении сложения, вычитания и умножения каждая цифра результата зависит от соответствующих цифр операндов. Так, если операнды A и B представлены соответственно остатками α и β по основаниям p_i при: $i = 1, 2, \dots, k$, то

$$A = (\alpha_1, \alpha_2, \dots, \alpha_k) \text{ и } B = (\beta_1, \beta_2, \dots, \beta_k) \quad (3.6)$$

$$|A + B|_p = (|\alpha_1 + \beta_1|_{p_1}, |\alpha_2 + \beta_2|_{p_2}, \dots, |\alpha_n + \beta_n|_{p_k}), \quad (3.7)$$

$$|A - B|_p = (|\alpha_1 - \beta_1|_{p_1}, |\alpha_2 - \beta_2|_{p_2}, \dots, |\alpha_n - \beta_n|_{p_k}), \quad (3.8)$$

$$|AB|_p = (|\alpha_1 \beta_1|_{p_1}, |\alpha_2 \beta_2|_{p_2}, \dots, |\alpha_n \beta_n|_{p_k}). \quad (3.9)$$

Таким образом, выполнение арифметических операций в модулярном коде производится независимо по каждому из модулей, что и указывает на параллелизм данной системы. Это обстоятельство определяет возможность независимой их обработки, то есть поразрядного выполнения операций, что избавляет от необходимости "занимать" или "переносить" единицу старшего разряда, и приводит к появлению кодов с параллельной структурой, которые позволяют распараллелить основные арифметические операции, выполняемые в нейронной сети (сложение и умножение).

Достаточно просто в СОК реализуются операции основного кольца. Пусть операнды A и B представлены соответственно остатками α_i и β_i по основаниям p_i при $i = 1, 2, \dots, n$. Тогда результаты операции сложения и умножения $A+B$ и AB представлены соответственно остатками γ_i и σ_i по тем же основаниям

$$A = (\alpha_1, \alpha_2, \dots, \alpha_n), B = (\beta_1, \beta_2, \dots, \beta_n), \\ A+B = (\gamma_1, \gamma_2, \dots, \gamma_n), AB = (\sigma_1, \sigma_2, \dots, \sigma_n),$$

и при этом имеют место соотношения

$$\gamma_i = \alpha_i + \beta_i - \left[\frac{\alpha_i + \beta_i}{p_i} \right] p_i, \quad (3.10)$$

$$\sigma_i = \alpha_i \beta_i - \left[\frac{\alpha_i + \beta_i}{p_i} \right] p_i. \quad (3.11)$$

Одной из проблем использования СОК является ограниченность действия системы сферой целых положительных чисел. Эту проблему предлагается решить, используя алгоритм представления и работы с дробными числами.

Покажем на примере реализации в СОК метода наименьших квадратов (МНК) для выделения тренда и определения коэффициентов модели регрессии [115, 123].

Пусть дано: $p_1 = 3$ $p_2 = 5$ $p_3 = 7$ $p_4 = 11$ тогда $P = 1155$

Таблица 3.10

Исходные данные

X	Y	x*y	x ²
1 (1,1,1,1)	1 (1,1,1,1)	1 (1,1,1,1)	1 (1,1,1,1)
2 (2,2,2,2)	4 (1,4,4,4)	8 (2,3,1,8)	4 (1,4,4,4)
3 (0,3,3,3)	3 (0,3,3,3)	9 (0,4,2,9)	9 (0,4,2,9)
4 (1,4,4,4)	5 (2,0,5,5)	20 (2,0,6,9)	16 (1,1,2,5)
5 (2,0,5,5)	4 (1,4,4,4)	20 (2,0,6,9)	25 (1,0,4,3)
∑ 15 (0,0,1,4)	∑ 17 (2,2,3,6)	∑ 58 (1,3,2,3)	∑ 55 (1,0,6,0)

Необходимо: построить линию регрессии $y = a_0 + a_1 * x$

Система уравнений для определения коэффициентов модели по МНК

$$\left\{ \begin{array}{l} na_0 + a_1 \sum_{i=1}^n x_i = \sum_{j=1}^n y_j \\ a_0 \sum_{i=1}^n x_i + a_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{array} \right.$$

Решая систему уравнений (в десятичной системе счисления) получаем модель

$$y = 1.3 + 0.7x \quad (3.12)$$

Решим систему в СОК

$$(2,0,5,5)_{(5)} * a_0 + a_1 * (0,0,1,4)_{(15)} = (2,2,3,6)_{(17)}$$

$$a_1 * (0,0,1,4)_{(15)} + a_1 * (1,0,6,0)_{(55)} = (1,3,2,3)_{(58)}$$

В нижнем индексе, в скобках, указаны значения чисел СОК в десятичной системе счисления. Выразим a_0

$$a_0 = \frac{(2,2,3,6)_{(17)} - a_1(0,0,1,4)_{(15)}}{(2,0,5,5)_{(5)}}$$

Подставим во второе уравнение

$$(0,0,1,4)_{(15)} * \frac{(2,2,3,6)_{(17)}}{(2,0,5,5)_{(5)}} - a_1 * (0,0,1,4)_{(15)} * \frac{(0,0,1,4)_{(15)}}{(2,0,5,5)_{(5)}} + a_1 * (1,0,6,0)_{(55)} = (1,3,2,3)_{(58)}$$

$$\begin{aligned} (0,1,2,7)_{(51)} - a_1 * (0,0,3,1)_{(45)} + a_1 * (1,0,6,0)_{(55)} &= (1,3,2,3)_{(58)} \\ a_1 * ((1,0,6,0)_{(55)} - (0,0,3,1)_{(45)}) &= (1,3,2,3)_{(58)} - (0,1,2,7)_{(51)} \\ a_1 * (1,0,3,10)_{(10)} &= (1,2,0,7)_{(7)} \end{aligned}$$

Тогда

$$a_1 = \frac{(1,2,0,7)_{(7)}}{(1,0,3,10)_{(10)}} = 0,7$$

Таким образом, далее числа СОК можно перевести либо в десятичную систему счисления либо в двоичную и там осуществить деление, либо применить алгоритм приближенного деления реализуемого в СОК, либо использовать при делении теорию индексов, либо воспользоваться алгоритмом вычитания, реализуемого в СОК. Вычислим a_0

$$\begin{aligned} a_0 &= \frac{(2,2,3,6)_{(17)} - \frac{(1,2,0,7)_{(7)}}{(1,0,3,10)_{(10)}} * (0,0,1,4)_{(15)}}{(2,0,5,5)_{(5)}} = \frac{(2,2,3,6)_{(17)} - \frac{(1,2,0,7)_{(7)}}{(2,2,2,2)_{(2)}} * (0,3,3,3)_{(3)}}{(2,0,5,5)_{(5)}} = \\ &= \frac{\frac{(1,4,6,1)_{(34)}}{(2,2,2,2)_{(2)}} - \frac{(0,1,0,10)_{(10)}}{(2,2,2,2)_{(2)}}}{(2,0,5,5)_{(5)}} = \frac{(1,3,6,2)_{(13)}}{(1,0,3,10)_{(10)}} = 1,3 \end{aligned}$$

Используя указанный подход появляется возможность реализовывать любые математические модели в СОК без представления в СОК числа с плавающей точкой.

Покажем возможность реализации в СОК алгоритма обратного распространения для обучения нейронных сетей. Рассмотрим пример, демонстрирующий процесс обучения сети методом обратного распространения реализованный в СОК. Допустим, обучающее множество состоит из одного набора $((1,1,1,1)_{(1)}; (2,2,2,2)_{(2)})$. Сеть должна научиться давать на выходе число $(0,3,3,3)_{(3)}$.

Пусть сеть состоит из двух слоев нейронов: два нейрона в скрытом слое и один нейрон в выходном слое рисунок 3.8.

Где в СОК $0,1 = \frac{(1,1,1,1)_{(1)}}{(1,0,3,10)_{(10)}}$, $0,2 = \frac{(2,2,2,2)_{(2)}}{(1,0,3,10)_{(10)}}$, $3 = (0,3,3,3)_{(3)}$

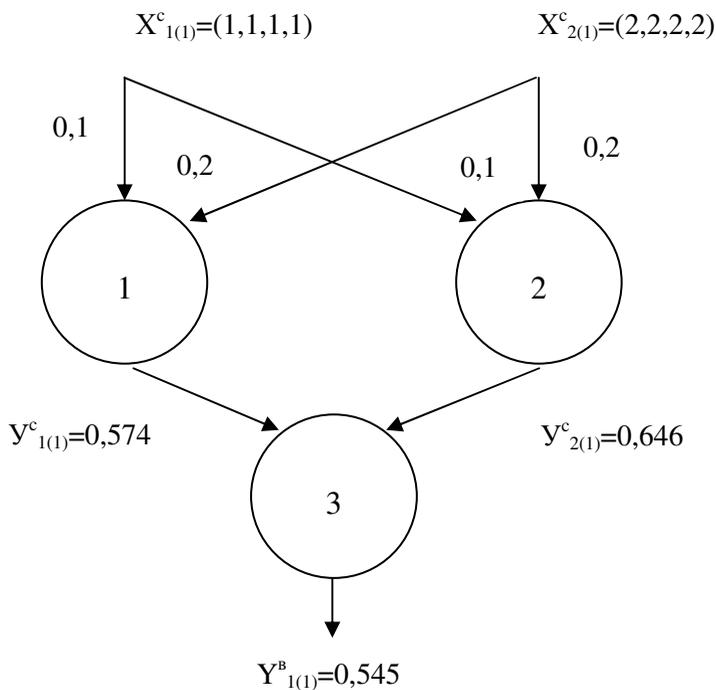


Рис. 3.8 Нейронная сеть, состоящая из одного скрытого слоя и одного выходного нейрона.

Далее будут использоваться следующие обозначения:

Буквой X обозначается вход какого-либо нейрона, идентификатор сверху указывает на слой, в котором находится нейрон: X^c – вход скрытого слоя, X^B – вход выходного слоя. Кроме того, с помощью индекса в скобках указывается номер итерации обучения: $X^B_{(1)}$ – вход скрытого слоя в процессе первой итерации, $X^B_{(2)}$ – вход выходного слоя в процессе второй итерации и т.д.

Буквой Y обозначается выход какого-либо нейрона. Идентификаторы и индексы те же, что и у буквы X .

Буквой W_i обозначается синаптическая матрица, используемая на i -и итерации обучения. Идентификаторы «с» и «в» указывают на синаптические веса соответствующего слоя (скрытого или выходного).

Ошибка сети или ошибка, полученная на выходе j -го нейрона, рассчитывается так

$$E_j = \frac{1}{2}(Y_j - D_j)^2, \quad (3.13)$$

где D_j – желаемый выход j -го нейрона.

Все нейроны будут возбуждаться с помощью одной функции – сигмоидной

$$F(X) = \frac{1}{1 + e^{-x}} \quad (3.14)$$

Для реализации функции в СОК представим e^{-x} разложением в ряд Тейлора

$$e^{-x} = 1 - x + \frac{x^2}{2} - \frac{x^3}{6} + \dots \quad (3.15)$$

Продемонстрируем две итерации обучения и один обратный проход.

Первая итерация. Первый прямой проход

На вход скрытого слоя сети подается набор обучающих данных

$$X_{(1)}^c = [X_{1(1)}^c; X_{2(1)}^c] = [1, 2]$$

$$X_{(1)}^c = [X_{1(1)}^c; X_{2(1)}^c] = [(1,1,1,1); (2,2,2,2)]$$

Здесь $X_{j(1)}^c$ – обозначает значение j -го входа скрытого слоя во время первой итерации.

Синаптическая матрица для скрытого слоя имеет вид

$$W_{(1)}^c = \begin{bmatrix} 0,1 & 0,2 \\ 0,1 & 0,2 \end{bmatrix}. \text{ в СОК}$$

$$W_{(1)}^c = \begin{bmatrix} \frac{(1,1,1,1)_{(1)}}{(1,0,3,10)_{(10)}} & \frac{(2,2,2,2)_{(2)}}{(1,0,3,10)_{(10)}} \\ \frac{(1,1,1,1)_{(1)}}{(1,0,3,10)_{(10)}} & \frac{(2,2,2,2)_{(2)}}{(1,0,3,10)_{(10)}} \end{bmatrix}.$$

Для выходного слоя такая матрица соответственно равна

$$W_{(1)}^e = \begin{bmatrix} 0,2 \\ 0,1 \end{bmatrix} \dots \quad \text{в СОК } W_{(1)}^e = \begin{bmatrix} \frac{(2,2,2,2)_{(2)}}{(1,0,3,10)_{(10)}} \\ \frac{(1,1,1,1)_{(1)}}{(1,0,3,10)_{(10)}} \end{bmatrix} \dots$$

Индекс (1) указывает на значения матрицы во время первой итерации.

Рассчитаем, чему равен взвешенный вход для каждого из нейронов скрытого слоя. Выход скрытого слоя сети, преобразованный функцией возбуждения, рассчитывается так

$$\bar{X}_{(1)}^c = \left[\bar{X}_{(1)}^c \bar{X}_{(2)}^c \right] = X_{(1)}^c W_{(1)}^c = [1; 2] \cdot \begin{bmatrix} 0,1 & 0,2 \\ 0,1 & 0,2 \end{bmatrix} = [0,3; 0,6]$$

в СОК

$$\bar{X}_{(1)}^c = [(1,1,1,1); (2,2,2,2)] \cdot \begin{bmatrix} \frac{(1,1,1,1)_{(1)}}{(1,0,3,10)_{(10)}} & \frac{(2,2,2,2)_{(2)}}{(1,0,3,10)_{(10)}} \\ \frac{(1,1,1,1)_{(1)}}{(1,0,3,10)_{(10)}} & \frac{(2,2,2,2)_{(2)}}{(1,0,3,10)_{(10)}} \end{bmatrix} = \left[\frac{(0,3,3,3)_{(3)}}{(1,0,3,10)_{(10)}}; \frac{(0,1,6,6)_{(6)}}{(1,0,3,10)_{(10)}} \right]$$

$$Y_{(1)}^c = [Y_{(1)}^c; Y_{(2)}^c] = F(\bar{X}_{(1)}^c) = \left[F(\bar{X}_{(1)}^c); F(X_{(2)}^c) \right] = [F(0,3); F(0,6)] =$$

$$= \left[\frac{1}{1+e^{-0,3}}; \frac{1}{1+e^{-0,6}} \right] = [0,574; 0,646]$$

в СОК

$$Y_{(1)}^c = \left[\frac{1}{1+e^{-\frac{(0,3,3,3)_{(3)}}{(1,0,3,10)_{(10)}}}}; \frac{1}{1+e^{-\frac{(0,1,6,6)_{(6)}}{(1,0,3,10)_{(10)}}}} \right] = \left[\frac{(0,2,1,2)_{(57)}}{(1,0,2,1)_{(100)}}; \frac{(2,0,2,10)_{(65)}}{(1,0,2,1)_{(100)}} \right]$$

где $e^{-x} = 1 - x + \frac{x^2}{2} - \frac{x^3}{6} + \dots$

$$\text{в СОК } e \frac{\frac{(0,3,3,3)_{(3)}}{(1,0,3,10)_{(10)}}}{(1,1,1,1)} = \frac{(0,3,3,3)_{(3)}}{(1,0,3,10)_{(10)}} +$$

$$\frac{\left(\frac{(0,3,3,3)_{(3)}}{(1,0,3,10)_{(10)}}\right)^2}{(2,2,2,2)_{(2)}} - \frac{\left(\frac{(0,3,3,3)_{(3)}}{(1,0,3,10)_{(10)}}\right)^3}{(0,1,6,6)_{(6)}}$$

для второго слагаемого применим ту же методику.

Выход скрытого слоя является входом для выходного слоя, т.е.

$$X_{(1)}^e = Y_{(1)}^c.$$

Как и для скрытого слоя, рассчитаем взвешенный вход для выходного слоя

$$X_{(1)}^e = \left[\bar{X}_{(1)}^e; \bar{X}_{(2)}^e \right] = Y_{(1)}^c W_{(1)}^e = [0,574; 0,646] \cdot \begin{bmatrix} 0,2 \\ 0,1 \end{bmatrix} = [0,179],$$

в СОК

$$X_{(1)}^e = \left[\bar{X}_{(1)}^e; \bar{X}_{(2)}^e \right] = Y_{(1)}^c W_{(1)}^e = \left[\frac{(0,2,1,2)_{(57)}}{(1,0,2,1)_{(100)}}; \frac{(2,0,2,1)_{(65)}}{(1,0,2,1)_{(100)}} \right] * \frac{\begin{bmatrix} (2,2,2,2)_{(2)} \\ (1,0,3,10)_{(10)} \\ (1,1,1,1)_{(1)} \end{bmatrix}}{(1,0,3,10)_{(10)}} = \left[\frac{(0,3,4,7)_{(18)}}{(1,0,2,1)_{(100)}} \right].$$

где $W_{(1)}^e$ – синоптическая матрица выходного слоя сети, значения которой соответствуют первой итерации.

Функция возбуждения для 3-го нейрона та же. Поэтому получим для него следующий выход

$$Y_{(1)}^e = F(\bar{X}_{(1)}^e) = F(0,179) = \frac{1}{1 + e^{-0,179}} = 0,545.$$

в СОК

$$Y_{(1)}^e = F(\bar{X}_{(1)}^e) = F\left(\frac{(0,3,4,7)_{(18)}}{(1,0,2,1)_{(100)}}\right) = \frac{1}{1 + e^{-\frac{(0,3,4,7)_{(18)}}{(1,0,2,1)_{(100)}}}} = \frac{(1,0,6,0)_{(55)}}{(1,0,2,1)_{(100)}}.$$

В то же время желаемое значение выхода сети, преобразованное функцией возбуждения, равно $D=F(3)=0,952$.

$$\text{в СОК } D=F(0,3,3,3) = \frac{(2,0,4,7)_{(95)}}{(1,0,2,1)_{(100)}}.$$

При этом среднеквадратичная ошибка равна

$$E = \frac{1}{2}(0,545 - 0,952)^2 = 0,0828 .$$

$$\text{Ошибку легко вычислить в СОК } E = \frac{(2,3,1,8)_{(8)}}{(1,0,2,1)_{(100)}} .$$

Значения фактического выхода и желаемого не совпадают. Поэтому синоптические веса сети следует изменить. Для этого необходимо выяснить, каким образом повлияют эти изменения на общий результат, т.е. на величину ошибки. Анализ согласно алгоритму обратного распространения выполняют, начиная с выходного слоя сети и продвигаясь к ее входу и т.д.

Рассмотренный подход реализации модульных вычислений для задач прогнозирования на нейронных сетях, позволяет нам получить ряд существенных преимуществ. Повышенная надежность работы системы. Так как параллельная структура нейронных сетей является источником повышения надежности и кроме этого система остаточных классов обладает способностью самокоррекции, обнаружения и исправления ошибок. Система остаточных классов, являясь непозиционной системой счисления, позволяет сократить разрядность обрабатываемых данных, представляя их в модулярном коде в системе взаимно простых модулей. В связи с этим, становится возможным упростить структуру нейронной сети и избежать проблем связанных с масштабированием входных данных.

3.6 Анализ методов повышения качества предсказаний

Хотя предсказание финансовых рядов и сводится к задаче аппроксимации многомерной функции, оно имеет свои особенности как при формировании входов, так и при выборе выходов нейросети.

Связь предсказуемости с нормой прибыли. Особенностью предсказания финансовых временных рядов является стремление к получению максимальной прибыли, а не минимизации среднеквадратичного отклонения, как это принято в случае аппроксимации функций. В простейшем случае ежедневной торговли прибыль зависит от верно угаданного знака изменения котировки. Поэтому НС нужно ориентировать именно на точность угадывания знака, а не самого значения. Найдем как связана норма прибыли с точностью определения знака в простейшей постановке ежедневного вхождения в рынок.

Обозначим на момент t : K_t – полный капитал игрока, $x_t = \Delta C_t / C_t$ относительное изменение котировки, а в качестве выхода сети возьмем степень ее уверенности в знаке этого изменения $y_t \in [-1, 1]$. Такая сеть с выходной нелинейностью вида $y = \tanh(a)$ обучается предсказывать знак изменения и выдает прогноз знака с амплитудой пропорциональной его вероятности. Тогда возрастание капитала на шаге t запишется в виде

$$K_t = K_{t-1} \left[1 + |x_t| \delta \operatorname{sgn}(x_t y_t) \right], \quad (3.16)$$

где δ – доля капитала, "в игре". Выигрыш за все время игры

$$K_t = K_0 \exp \left(\sum_{k=1}^t \ln \left[1 + x_k \delta \operatorname{sgn}(y_k) \right] \right), \quad (3.17)$$

нам и предстоит максимизировать, выбрав оптимальный размер ставок δ . Пусть в среднем игрок угадывает долю $p = \frac{1}{2} + \varepsilon$ знаков и, соответственно, ошибается с вероятностью $q = \frac{1}{2} - \varepsilon$. Тогда логарифм нормы прибыли

$$\langle \ln(K_t / K_0) \rangle = t \langle p \ln(1 + |x| \delta) + q \ln(1 - |x| \delta) \rangle, \quad (3.18)$$

а следовательно и сама прибыль, будет максимальным при значении $\delta = (p - q) \langle |x| \rangle / \langle x^2 \rangle$ и составит в среднем

$$\langle \ln(K_t / K_0) \rangle \approx t (p - q)^2 \frac{\langle |x| \rangle^2}{2 \langle x^2 \rangle} = 2at\varepsilon^2. \quad (3.19)$$

Здесь мы ввели коэффициент $a = \langle |x| \rangle^2 / \langle x^2 \rangle \leq 1$. Например, для Гауссова распределения $a \approx 0.8$. Степень предсказуемости знака напрямую связана с кросс-энтропией, которую можно оценить а priori методом box-counting. Для бинарного выхода (см. рисунок 3.9)

$$I = H_{\max} - H(p), \quad (3.20)$$

В итоге получаем следующую оценку нормы прибыли при заданной величине предсказуемости знака I , выраженной в битах

$$K_t = K_0 2^{aIt}. \quad (3.21)$$

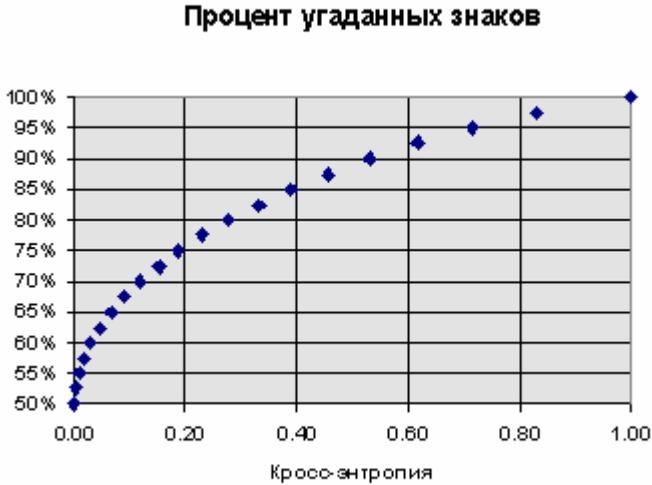


Рис. 3.9 Доля правильно угаданных направлений изменений ряда, как функция кросс-энтропии знака выхода при известных входах

То есть, для ряда с предсказуемостью I в принципе возможно удвоить капитал за $t = 1/(aI)$ вхождений в рынок. Так, например, если предсказуемость временного ряда равная $I = 0.17$ предполагает удвоение капитала в среднем за $t = 1/(0.8 \cdot 0.17) \approx 8$ вхождений в рынок. Таким образом, даже небольшая предсказуемость знака изменения котировок способна обеспечить весьма заметную норму прибыли.

Подчеркнем, что оптимальная норма прибыли требует достаточно аккуратной игры, когда при каждом вхождении в рынок игрок рискует строго определенной долей капитала

$$\frac{\langle \Delta K \rangle}{K} = \delta \langle |x| \rangle = (p - q) \langle |x| \rangle^2 / \langle x^2 \rangle = 2a\varepsilon \approx 1.6\varepsilon, \quad (3.22)$$

где ΔK - типичная при данной волатильности рынка $\langle |x| \rangle$ величина выигрыша или проигрыша. Как меньшие, так и большие значения ставок уменьшают прибыль. Причем, чересчур рискованная игра может привести к проигрышу при любой предсказательной способности. Этот факт иллюстрирует рисунок 3.10.

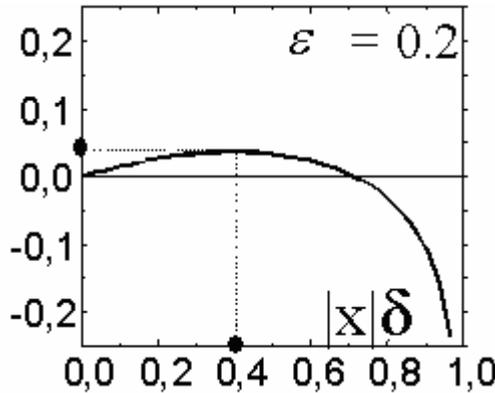


Рис. 3.10 Зависимость средней нормы прибыли от выбора доли капитала «на кону»

Поэтому приведенные выше оценки дают представление лишь о верхнем пределе нормы прибыли.

Выбор функционала ошибки. Если принять, что целью предсказаний финансовых временных рядов является максимизация прибыли, логично настраивать НС именно на этот конечный результат. Например, при игре по описанной выше схеме для обучения НС можно выбрать следующую функцию ошибки обучения, усредненную по всем примерам из обучающей выборки

$$E = -\left\langle \ln \left[1 + x_i \delta_i \operatorname{sgn}(y_i) \right] \right\rangle. \quad (3.23)$$

Здесь доля капитала в игре введена в качестве дополнительного выхода сети, настраиваемого в процессе обучения. При таком подходе, первый нейрон, y_t , с функцией активации $f = \tanh(\cdot)$ даст вероятность возрастания или убывания курса, в то время как второй выход сети δ_t даст рекомендованную долю капитала в игре на данном шаге.

Поскольку, однако, в соответствии с предыдущим анализом, эта доля должна быть пропорциональна степени уверенности предсказания, можно заменить два выхода сети – одним, положив $\delta_t = \delta |y_t|$, и ограничиться оптимизацией всего одного глобального параметра δ минимизирующего ошибку

$$E = -\langle \ln[1 + \delta x_t y_t] \rangle \quad (3.24)$$

Тем самым, появляется возможность регулировать ставку в соответствии с уровнем риска, предсказываемым сетью. Игра с переменными ставками приносит большую прибыль, чем игра с фиксированными ставками.

Использование комитетов сетей. Из-за случайности в выборе начальных значений синаптических весов, предсказания сетей, обученных на одной и той же выборке, будут различаться. Этот недостаток (элемент неопределенности) можно превратить в достоинство, организовав комитет нейро-экспертов, состоящий из различных нейросетей. Разброс в предсказаниях экспертов даст представление о степени уверенности этих предсказаний, что можно использовать для правильного выбора стратегии игры.

Легко показать, что среднее значений комитета должно давать лучшие предсказания, чем средний эксперт из этого же комитета. Пусть ошибка i -ого эксперта для значения входа \mathbf{x} равна $\varepsilon_i(\mathbf{x})$. Средняя ошибка комитета всегда меньше среднеквадратичной ошибки отдельных экспертов в силу неравенства Коши – Буняковского

$$\left(\frac{1}{L} \sum_{i=1}^L \varepsilon_i \right)^2 \leq \frac{1}{L} \sum_{i=1}^L \varepsilon_i^2 . \quad (3.25)$$

Причем, снижение ошибки может быть довольно заметным. Так, если ошибки отдельных экспертов не коррелируют друг с другом, т.е. $\langle \varepsilon_i \varepsilon_j \rangle = \langle \varepsilon_i^2 \rangle \delta_{ij}$, то среднеквадратичная ошибка комитета из L экспертов в \sqrt{L} раз меньше, чем средняя индивидуальная ошибка одного эксперта

$$\begin{aligned}
 E_L^2 &= \left\langle \left(\frac{1}{L} \sum_{i=1}^L \varepsilon_i \right)^2 \right\rangle = \frac{1}{L^2} \left\langle \sum_{i=1}^L \varepsilon_i^2 + \sum_{i \neq j} \varepsilon_i \varepsilon_j \right\rangle = \\
 &= \frac{1}{L^2} \sum_{i=1}^L \langle \varepsilon_i^2 \rangle = \frac{1}{L} E_1^2
 \end{aligned}
 \tag{3.26}$$

Поэтому, в предсказаниях всегда лучше опираться на средние значения всего комитета.

Выводы к главе

Приведенные выше результаты свидетельствуют о перспективности нейросетевого трейдинга. Более того, в силу самоподобия (фрактальности) финансовых временных рядов (Peters, 1994), норма прибыли за единицу времени будет тем выше, чем меньше характерное время трейдинга. Таким образом, автоматические нейросетевые трейдеры оказываются наиболее эффективны при торговле в реальном времени. Обученная нейросеть, подсоединенная к электронной системе торгов, может принимать решения еще до того, как брокер-человек успеет распознать изменения графика котировок на своем терминале.

ГЛАВА 4 СРАВНИТЕЛЬНАЯ ОЦЕНКА КЛАССИЧЕСКИХ И НЕЙРОСЕТЕВЫХ МЕТОДОВ ПРОГНОЗИРОВАНИЯ. ОЦЕНКА ТОЧНОСТИ И ПОГРЕШНОСТИ ВЫЧИСЛЕНИЙ В НЕЙРОННЫХ СЕТЯХ

Сравнительная оценка прогнозов проведена для исследуемых временных рядов динамики курсов акций для последних 200 значений, так как именно эта часть выборок считается тестовой для всех исследуемых моделей прогнозирования. Исследование проведено для моделей экспоненциальной средней по следующей схеме.

Адаптивная полиномиальная модель первого порядка ($p = 1$).

Экспоненциальные средние 1-го и 2-го порядков определяются как

$$\begin{aligned} S_t &= \alpha x_t + \beta S_{t-1} \\ S_t^{[2]} &= \alpha S_t + \beta S_{t-1}^{[2]}, \end{aligned} \quad (4.1)$$

где $\beta = 1 - \alpha$.

Отсюда начальные условия

$$\begin{aligned} S_0 &= \hat{a}_{1,0} - \frac{\beta}{\alpha} \hat{a}_{2,0} \\ S_0^{[2]} &= \hat{a}_{1,0} - \frac{2\beta}{\alpha} \hat{a}_{2,0} \end{aligned} \quad (4.2)$$

Оценка модельного значения ряда с периодом упреждения τ равна

$$\hat{x}_t^* = \left(2 + \frac{\alpha}{\beta} \tau \right) S_{t-\tau} - \left(1 + \frac{\alpha}{\beta} \tau \right) S_{t-\tau}^{[2]} \quad (4.3)$$

Адаптивная полиномиальная модель второго порядка ($p=2$). Экспоненциальные средние 1-го, 2-го и 3-го порядков

$$\begin{aligned} S_t &= \alpha x_t + \beta S_{t-1}; \\ S_t^{[2]} &= \alpha S_t + \beta S_{t-1}^{[2]}; \\ S_t^{[3]} &= \alpha S_t^{[2]} + \beta S_{t-1}^{[3]}. \end{aligned} \quad (4.4)$$

Начальные условия определим как

$$\begin{aligned}
 S_0 &= \hat{a}_{1,0} - \frac{\beta}{\alpha} \hat{a}_{2,0} + \frac{\beta(2-\alpha)}{2\alpha^2} \hat{a}_{3,0}; \\
 S_0^{[2]} &= \hat{a}_{1,0} - \frac{2\beta}{\alpha} \hat{a}_{2,0} + \frac{\beta(3-2\alpha)}{\alpha^2} \hat{a}_{3,0}; \\
 S_0^{[3]} &= \hat{a}_{1,0} - \frac{3\beta}{\alpha} \hat{a}_{2,0} + \frac{3\beta(4-3\alpha)}{2\alpha^2} \hat{a}_{3,0}.
 \end{aligned} \tag{4.5}$$

где $\hat{a}_{1,0} = \hat{a}_1$; $\hat{a}_{2,0} = \hat{a}_2$; и $\hat{a}_{3,0} = \hat{a}_3$.

Оценка модельного значения с периодом упреждения τ находим из выражения

$$\begin{aligned}
 \hat{x}_t^* &= [6\beta^2 + (6-5\alpha)\alpha \cdot \tau + \alpha^2 \tau^2] \frac{S_{t-\tau}}{2\beta^2} - [6\beta^2 + 2(5-4\alpha)\alpha\tau + 2\alpha^2 \tau^2] \frac{S_{t-\tau}^{[2]}}{2\beta^2} + \\
 &+ [2\beta^2 + (4-3\alpha)\alpha\tau + \alpha^2 \tau^2] \frac{S_{t-\tau}^{[3]}}{2\beta^2}.
 \end{aligned} \tag{4.6}$$

Весьма эффективным и надежным методом прогнозирования является экспоненциальное сглаживание. Основные достоинства метода состоят в возможности учета весов исходной информации, в простоте вычислительных операций, в гибкости описания различных динамик процессов. Метод экспоненциального сглаживания дает возможность получить оценку параметров тренда, характеризующих не средний уровень процесса, а тенденцию, сложившуюся к моменту последнего наблюдения. Наибольшее применение метод нашел для реализации среднесрочных прогнозов. Для метода экспоненциального сглаживания основным и наиболее трудным моментом является выбор параметра сглаживания α , начальных условий и степени прогнозирующего полинома. На рисунке 4.1 показана зависимость качества прогнозов от изменения параметра сглаживания α для адаптивной модели первого порядка, построенной для курса акций «РАО ЕЭС». Из рисунка хорошо видно как влияет на качество прогнозов изменение параметра сглаживания α .

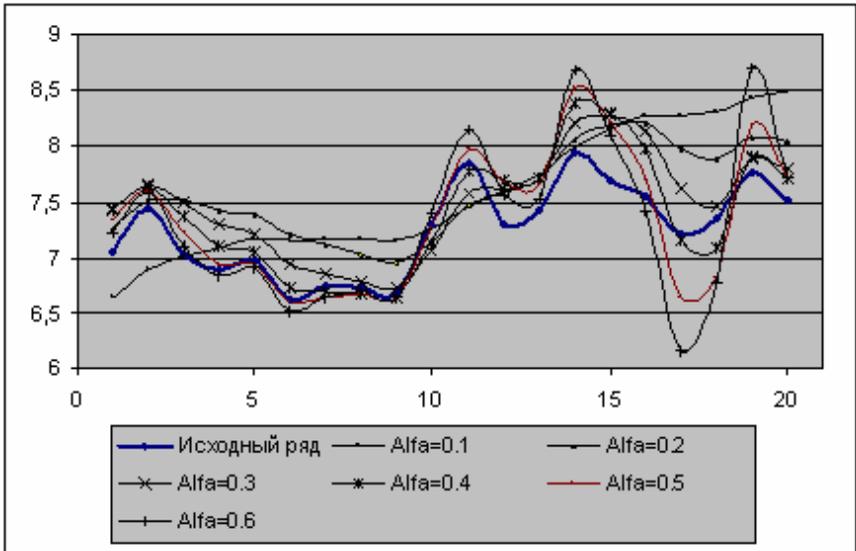


Рис. 4.1 Зависимость качества прогнозов от изменения параметра сглаживания α на примере курса акций «РАО ЕЭС» для последних 20 значений ряда

В зависимости от величины параметра прогнозные оценки по-разному учитывают влияние исходного ряда наблюдений: чем больше α , тем больше вклад последних наблюдений в формирование тренда, а влияние начальных условий быстро убывает. При малом α прогнозные оценки учитывают все наблюдения, при этом уменьшение влияния более «старой» информации происходит медленно. На рисунке 4.2 показана зависимость изменения ошибки прогноза в зависимости от изменения α для модели первого порядка.

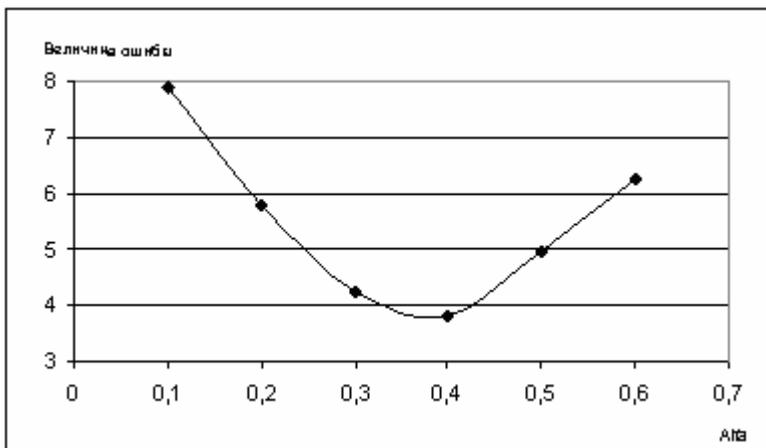


Рис. 4.2 Зависимость изменения ошибки прогноза в зависимости от изменения α

В результате численных экспериментов удалось найти оптимальные значения параметра α для моделей первого и второго порядка, которые составили 0,38 и 0,2 соответственно. Однако полученные оптимальные модели не показали лучших прогнозов, чем нейронные сети. Так на рисунках 4.3 – 4.6 показаны результаты прогнозов исследованных моделей и полученных нейронных сетей для исследуемых временных рядов.

Так для акций «РАО ЕЭС» средняя относительная ошибка прогноза для модели $p = 1$ составила $E = 3,81\%$, для модели $p = 2$, $E = 4,02\%$, для НС (структуры 1:10:1) $E = 3,90\%$; для НС (структуры 1:13:13:1) $E = 3,88\%$.

Для акций «Ростелеком» средняя относительная ошибка прогноза для модели $p = 1$, составила $E = 2,56\%$, для модели $p = 2$, $E = 4,26\%$, для подобранных НС (структуры 1:10:1) $E = 2,41\%$, для НС (структуры 3:6:1) $E = 2,46\%$.

Для акций «Лукойла» средняя относительная ошибка прогноза для модели $p = 1$, составила $E = 2,60\%$, для модели $p = 2$, $E = 3,11\%$, для подобранных НС (структуры) $E = 2,67\%$, для НС (структуры) $E = 2,46\%$.

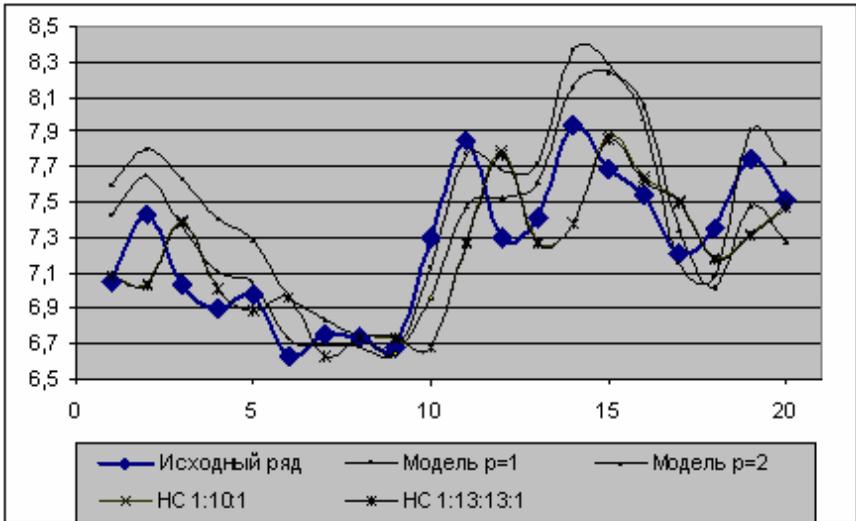


Рис. 4.3 Результаты прогнозов адаптивных моделей и полученных нейронных сетей для динамики курса акций «РАО ЕЭС» для последних 20 значений ряда

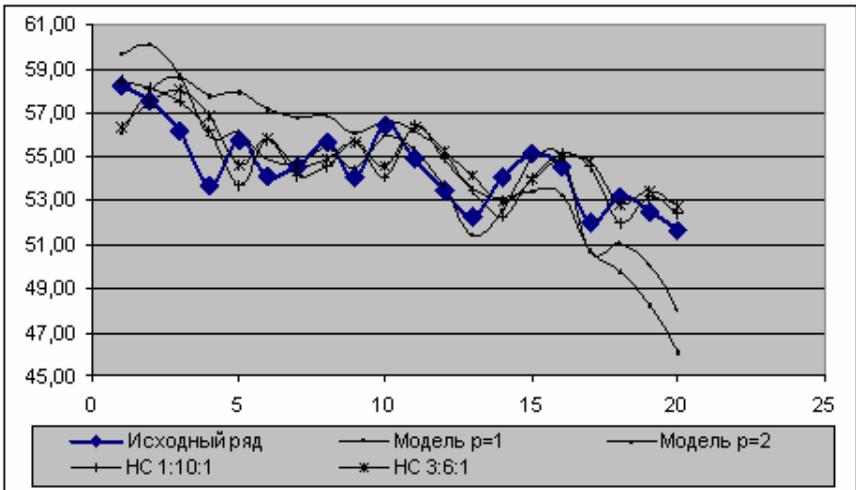


Рис. 4.4 Результаты прогнозов адаптивных моделей и полученных нейронных сетей для динамики курса акций «Ростелеком» для последних 20 значений ряда

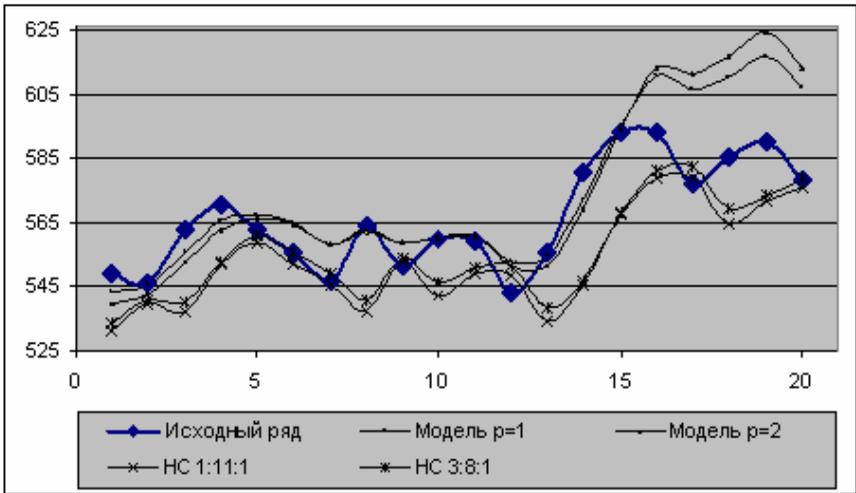


Рис. 4.5 Результаты прогнозов адаптивных моделей и полученных нейронных сетей для динамики курса акций «Лукойл» для последних 20 значений ряда

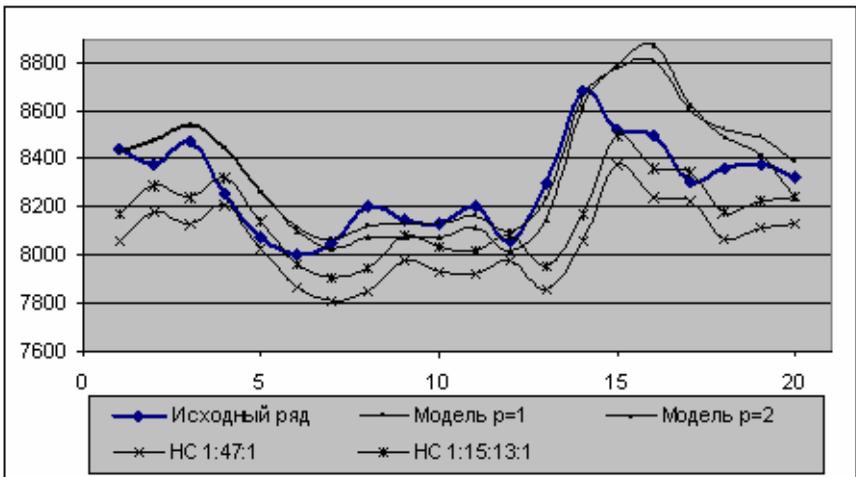


Рис. 4.6 Результаты прогнозов адаптивных моделей и полученных нейронных сетей для динамики курса акций «Сбербанк» для последних 20 значений ряда

Для акций «Сбербанка» средняя относительная ошибка прогноза для модели $p = 1$, составила $E = 1,61\%$, для модели $p = 2$, $E =$

1,87%, для подобранных НС (структуры 1:47:1) $E = 3,18\%$, для НС (структуры 1:15:13:1) $E = 1,98\%$. Высокая ошибка прогнозирования для НС вызвана тем, что рассматриваемый ряд имеет не достаточное количество данных, что привело к недообученности НС. Но в целом нейронные сети показали наилучшие результаты по сравнению с классическими моделями прогнозирования.

Литература

1. Абовский Н.П. и др. Разработка практического метода нейросетевого прогнозирования. //Труды VIII Всероссийской конференции «Нейрокомпьютеры и их применение» Сб.докл., 2002. – С. 1089 – 1097.
2. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрии. – М.: ЮНИТИ, 1998. – 465 с.
3. Акушский И.Я. Машинная арифметика в остаточных классах. – М: Советское радио, 1968. – 440 с.
4. Алексеев В.И., Максимов А.В. Использование нейронных сетей с двухмерными слоями для распознавания образов//Труды VIII Всероссийской конференции «Нейрокомпьютеры и их применение»: Сб. докл., 2002. – С. 69–72.
5. Амербаев В.М. Теоретические основы машинной арифметики. – Алма-Ата: Наука КазССР, 1976. – 324 с.
6. Андерсон Т. Статистический анализ временных рядов. – М.: Наука, 1976. – 343 с.
7. Аркин В. И, Евстигнеев И.В. Вероятностные модели управления экономической динамики. – М.: Наука, 1979. – 176с.
8. Ашманов С.А. Математические модели и методы в экономике. М.: Изд. МГУ, 1981. – 158 с.
9. Барский А.Б. Обучение нейросети методом трассировки //Труды VIII Всероссийской конференции «Нейрокомпьютеры и их применение»: Сб. докл., 2002. – С. 862 – 898.
10. Батищев Д.И. Генетические алгоритмы решения экстремальных задач. – Воронеж: ВГУ, 1994. – 135 с.
11. Белим С.В. Математическое моделирование квантового нейрона//Труды VIII Всероссийской конференции «Нейрокомпьютеры и их применение»: Сб.докл., 2002. – С. 899 – 900.
12. Берже П., Помо И., Видаль К. Порядок в хаосе. О детерминированном подходе к турбулентности: Пер. с франц. – М.: Мир, 1991. – 368 с.
13. Бирман Э.Г. Сравнительный анализ методов прогнозирования //НТИ. Сер.2 – 1986. – №1. – С. 11–16.
14. Бодянский Е.В., Кучеренко Е.И. Диагностика и прогнозирование временных рядов многослойной радиально-базисной нейронной сети //Труды VIII Всероссийской конференции «Нейрокомпьютеры и их применение»: Сб. докл., 2002. – С. 69–72.

15. Бокс Дж., Дженкинс Г. (1974) Анализ временных рядов. Прогноз и управление. – М.: Мир, 1974. – Вып. 1, 2.
16. Болн Б., Хуань К. Дж. Многомерные статистические методы для экономики. М.: Наука, 1979. – 348 с.
17. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1965. – 35 с.
18. Боярский А.Я., и др. Математическая статистика для экономистов. – М.: Статистика, 1979. – 253 с.
19. Бурдо А.И., Тихонов Э.Е. К вопросу систематизации методов и алгоритмов прогнозирования//Материалы межрегиональной конференции "Студенческая наука – экономике научно-технического прогресса". Ставрополь: СевКав ГТУ, 2001. – С. 33 – 34.
20. Бурдо А.И., Тихонов Э.Е. К вопросу совершенствования систем прогнозирования//Материалы XXXVIII юбилейной отчетной научной конференции за 1999 год: В 3 ч./Воронеж. гос. технол. акад. Воронеж, 2000.– Ч.2. – С. 211–215.
21. Бурдо А.И., Тихонов Э.Е. Об одном подходе к проблеме прогнозирования количественных характеристик производственных систем//Материалы XXX НТК профессорско-преподавательского состава. Ставрополь: СевКав ГТУ, 2000.– С.225–226.
22. Бутенко А.А. и др. Обучение нейронной сети при помощи алгоритма фильтра Калмана. //Труды VIII Всероссийской конференции «Нейрокомпьютеры и их применение»: Сб. докл., 2002. – С. 1120 – 1125.
23. Бухштаб А.А. Теория чисел. – М.: Государственное учебно-методическое издательство мин. Просвещения РСФСР, 1960. – 375 с.
24. Бытачевский Е.А., Козуб В.В. Использование нейронных сетей для распознавания визуальных образов//Материалы IV РНТ конференции «Вузовская наука – Северо-Кавказскому региону» Ставрополь, 2000. – С. 52–54.
25. Васильев В.И. Распознающие системы. – Киев: Наукова думка, – 1969. – 354 с.
26. Виноградов И.М. Основы теории чисел. – М.: Издательство «Наука», Гл. ред. физ.-мат.лит., 1965.– 173 с.

27. Владимирова Л.П. Прогнозирование и планирование в условиях рынка: Учебное пособие. – М.: Издательский дом «Дашков и К», 2000. – 308 с.
28. Вороновский Г.К., и др. Генетические алгоритмы, нейронные сети и проблемы виртуальной реальности. – Х.: ОСНОВА, 1997. – 112 с.
29. Галушкин А.И. Теория нейронных сетей. Кн. 1: Учеб. Пособие для вузов. – М.: ИПРЖР, 2001. – 385 с.:ил.
30. Гвишиани Д.М., Лисичкин В.А. Прогностика. М.: «Знание», 1968. – 421 с.
31. Гельфан И.М., Фомин С.В. Вариационное исчисление. – М.: Мир, 1961. – 321 с.
32. Гладышевский А.И. Методы и модели отраслевого экономического прогнозирования. – М.: «Экономика», 1997. – 143 с.
33. Гласс Л., Мэки М. От часов к хаосу: ритмы жизни. – М.: Мир, 1991. – 153с.
34. Глущенко В.В. Прогнозирование. 3-е издание. – М.: Вузовская книга, 2000. – 208 с.
35. Голованова Н.Б., Кривов Ю.Г. Методические вопросы использования межотраслевого баланса в прогнозных расчетах//Взаимосвязи НТП и экономического развития: Сб.науч.тр./АН СССР. СО, ИЭиОПП. – Новосибирск, 1987. – С. 62–77.
36. Головкин В.А. Нейронные сети: обучение, организация и применение. Кн.4:Учеб.пособие для вузов/Общая ред. А.И. Галушкина. – М.: ИПРЖР, 2001. – 256 с.
37. Горбань А.Н.Обучение нейронных сетей.–М.:СП“ПараГраф”, 1990. – 159с.
38. Горелик Е.С. и др. Об одном подходе к задаче формализации процесса прогнозирования //Автоматика и телемеханика. – 1987. – №2. – С.129-136.
39. Гренандер У. Случайные процессы и статистические выводы. (пер. с нем.) ИЛ. 1961. – 167с.
40. Гренджер К., Хатанака М. Спектральный анализ временных рядов в экономике. Пер.с англ. – М.: Статистика., 1972. – 312 с.
41. Грень Е. Статистические игры и их применение. – М.: Наука, 1975. – 243 с.
42. Гусак А.Н. и др. Подход к послойному обучению нейронной сети прямого распространения//Труды VIII Всероссийской

- конференции «Нейрокомпьютеры и их применение » Сб.докл., 2002. – С. 931 – 933.
43. Давидович Б.Я. и др. Методы прогнозирования спроса. М., 1972. –193с.
 44. Дженкинс Г., Ватс Д. (1971, 1972) Спектральный анализ и его применения. – М.: Мир, 1971, 1972. – Вып. 1,2.
 45. Джонстон Дж. (1980) Эконометрические методы. – М.: Статистика, 1980. – 431 с.
 46. Добров Г.М., Ершов Ю.В. и др. Экспертные оценки в научно-техническом прогнозировании – Киев: Наукова Думка, 1974. – 159 с.
 47. Еремин Д.М. Система управления с применением нейронных сетей//Приборы и системы. Управление, контроль, диагностика. – 2001. –№9 – С. 8–11.
 48. Заенцев И.В. Нейронные сети: основные модели/Учебное пособие к курсу «Нейронные сети» – Воронеж: ВГУ, 1999. – 76 с.
 49. Зайкин В.С. Применение простых цепей Маркова для прогнозирования расходов населения//Проблемы моделирования народного хозяйства, 4IV. Новосибирск, 1973. – С. 45 –47.
 50. Занг В.Б. Синергитическая экономика. Время и перемены в нелинейной экономической теории. – М.: Мир, 1999. – 216с.
 51. Ибираимова Т.Б. Прогнозирование тенденций финансовых рынков с помощью нейронных сетей//Труды VIII Всероссийской конференции «Нейрокомпьютеры и их применение » Сб.докл., 2002 г. – С. 745 – 755.
 52. Иванов М.Н. Анализ роста курса акций с применением нейронных сетей. //Труды VIII Всероссийской конференции «Нейрокомпьютеры и их применение» Сб.докл., 2002 г. – С. 756 – 772.
 53. Ивахненко А.Г. Долгосрочное прогнозирование и управление сложными системами. – Киев: Наукова думка, 1975. – 340 с.
 54. Ивахненко А.Г., Лапа Р.Г., Предсказание случайных процессов. – Киев: Наукова думка, 1971. – 416с.
 55. Ивахненко А.Г., Степаненко В.С. Особенности применения метода группового учета аргументов в задачах прогнозирования случайных процессов//Автоматика. –1986. – №5. – С. 3-14.
 56. Ивахненко А.Г., Юрачков Ю.П. Моделирование сложных систем по экспертным данным. – М.: Радио и связь, 1987. – 119 с.
 57. Калан, Роберт. Основные концепции нейронных сетей.: Пер. с англ. – М.: Издательский дом «Вильямс», 2001.– 288 с.

58. Капица С. П., Курдюмов С. П., Малинецкий Г. Г. Синергетика и прогнозы будущего. М.: Наука, 1997. – 236с.
59. Касти Дж. Связность, сложность и катастрофы: Пер. с англ. – М.: Мир, 1982, – 216 с.
60. Кейн Э. Экономическая статистика и эконометрия. М.: Наука, 1977, Вып. 1, 2.
61. Кемени Дж., Снелл Дж. Конечные цепи Маркова. М.: Наука, 1970. – 136 с.
62. Кендэл М. Временные ряды. Пер. с англ. Ю.П. Лукашина. – М.: «Финансы и статистика», 1979. – 198 с.
63. Кильдинов Г.С., Френкель А.А. Анализ временных рядов и прогнозирование. М. Статистика. 1973. – 432 с.
64. Клеопатров Д.И., Френкель А.А. Прогнозирование экономических показателей с помощью метода простого экспоненциального сглаживания. – Статистический анализ экономических временных рядов и прогнозирование. – М.: Наука, 1973. – 298с.
65. Кобринский Н.Е., и др. Экономическая кибернетика: Учебник для студентов вузов и фак., обучающихся по спец. «Экономическая кибернетика». – М.: Экономика, 1982. – 408 с.
66. Колмогоров А.Н. Об энтропии на единицу времени как метрическом инварианте автоморфизмов. ДАН СССР, 1959. – Т.124 – С.754-755
67. Кондратьев А.И. Теоретико-игровые распознающие алгоритмы. – М.: Наука, 1990. – 272 с.
68. Кульбак С. Теория информации и статистика. – М.: Наука, 1967. – 408с.
69. Кучин Б.Л., Якушева Е.В. Управление развитием экономических систем: технический прогресс, устойчивость. – М.: Экономика, 1990. – 156с.
70. Лашев А.Я., Глушич Д.В. Синтез алгоритмов обучения нейронных сетей. //Труды VIII Всероссийской конференции «Нейрокомпьютеры и их применение» Сб.докл., 2002 г. – С. 997 – 999.
71. Левин В.Л. Выпуклый анализ в пространстве измеримых функций и его применение в математике и экономике. М.: Наука, 1985. – 352с.
72. Легостаева И.Л., Ширяев А.Н. Минимальные веса в задаче выделения тренда случайного процесса. – «Теория вероятностей и ее применение», 1971, – Т. XVI, – №2.

73. Лизер С. Эконометрические методы и задачи. – М.: Статистика, 1971. – 141с.
74. Лисичкин В.А. Теория и практика прогностики. – М.: Наука, 1972. – 223с.
75. Литовченко Ц.Г. Нейрокомпьютер для обнаружения и распознавания сложных динамических образов//Труды VIII Всероссийской конференции «Нейрокомпьютеры и их применение » Сб.докл., 2002 г. – С. 69–72.
76. Ллойд Э., Ледерман У. (1990) (ред.) Справочник по прикладной статистике. – М.: Финансы и статистика, 1990. – Том 2.
77. Лопатников Л.И. Экономико-математический словарь /АНССР. ЦЭМИ, – М.: Наука, 1987. – 506 с.
78. Лоскутов А.Ю., Михайлов А.С. Введение в синергетику: Учеб. Руководство. –М.: Наука. гл. ред. физ.-мат.лит., 1990.– 272 с.
79. Льюис К.Д. Методы прогнозирования экономических показателей./Пер. с англ. Демиденко Е.З. – М.: Финансы и статистика, 1986 г. – 132 с.
80. Ляпунов А.М. Собр. соч. Т.1,2. –М.:Изд-во АН СССР, 1954–1956.
81. Максимов В.А. Прогнозирование доходности инвестиций на фондовом рынке//Экономика и математические методы, 2001. Т. 37–№1. – С. 37 – 46.
82. Малинецкий Г.Г., Потапов А.Б. Современные проблемы нелинейной динамики – М.: Эдиториал УРСС, 2000.– 336с.
83. Математическая энциклопедия: Гл.ред. И.М. Виноградов, т.3 Коо-Од – М.: Советская энциклопедия, 1982. – 1184 с.
84. Махортых С.А., Сычев В.В. Алгоритм вычисления размерности стохастического аттрактора и его применение к анализу электрофизиологических данных.– Пущино, 1998. – 34с.
85. Медведев В.С., Потемкин В.Г. Нейронные сети. MATLAB 6 /Под общ.ред. к.т.н. В.Г. Потемкина. – М.: ДИАЛОГ-МИФИ, 2002. – 496 с.
86. Минский М., Пайперт С. Перцептроны. – М.: Мир, 1971. – 365с.
87. Михайлов Ю.Б. Алгоритм выбора прогнозирующей зависимости, обеспечивающей наилучшую точность прогноза//Приборы и системы. Управление, контроль, диагностика., 2000. – №12. – С. 11 – 19.

88. Моделирование функционирования развивающихся систем с изменяющейся структурой. Сб. науч. тр./АН УССР. Ин-т кибернетики им. В.М. Глушакова. – Киев: 1989. – 140 с.
89. Моришма М. Равновесие, устойчивость, рост. – М.: Наука, 1972. 314 с.
90. Морозова Т.Г., Пикулькин А.В., Тихонов В.Ф., и др. Прогнозирование и планирование в условиях рынка. Учеб. Пособие для вузов. Под ред. Т.Г. Морозовой, А.В. Пикулькина. – М.: ЮНИТИ-ДАНА, 1999. – 318 с.
91. Нейроинформатика и ее приложения //Материалы 3 Всероссийского семинара, 6-8 октября 1995 г. Ч. 1 /Под редакцией А.Н.Горбаня. – Красноярск: изд. КГТУ, 1995. – 229 с.
92. Нейронные сети. STSTISTICA Neural Networks: Пер. с англ. – М.: Горячая линия – Телеком. 2001. – 182 с.
93. Никайдо Х. Выпуклые структуры и математическая экономика. – М.: Мир. 1972. – 127с.
94. Новиков А.В., и др. Метод поиска экстремума функционала оптимизации для нейронной сети с полными последовательными связями //Труды VIII Всероссийской конференции «Нейрокомпьютеры и их применение» Сб.докл., 2002 г. – С. 1000 – 1006.
95. Осуга М. Обработка знаний. – М.: Мир, 1989. – 239 с.
96. Оуэн Г. Теория игр. – М.: Наука, 1971. – 359 с.
97. Песин Я.Б. Характеристические показатели Ляпунова и гладкая эргодическая теория. УМН, 1977.– Т.32. – С.55–112.
98. Петерс Э. Хаос и порядок на рынке капитала. – М.: Мир, 2000. – 333с.
99. Половников В.А., и др. Оценивание точности и адекватности моделей экономического прогнозирования // Математическое моделирование экономических процессов: Сб. науч. тр./МЭСИ – М., 1986. – С. 37–47.
100. Пятецкий В.Е., Бурдо А.И. Имитационное моделирование процесса создания обучающихся систем. – В сб.: Имитационное моделирование производственных процессов. Под. ред. Мироносецкого Н.Б., – Новосибирск. 1979. – 68 с.
101. Пятецкий В.Е., Бурдо А.И., Литвяк В.С. Построение стохастических моделей прогнозирования параметров производственных систем//Рук. деп. в Черметинформации, 1987.– № 4161.

102. Развитие российского финансового рынка и новые инструменты привлечения инвестиций. – М., 1998. – 233 с.
103. Рожков Л.Н., Френкель А.А. Выбор оптимального параметра сглаживания в методе экспоненциального сглаживания. – Основные проблемы и задачи научного прогнозирования. – М.: Наука, 1972.- 154 с.
104. Розенблат Ф. Принципы нейродинамики: Персептрон и теория механизмов мозга. Пер. с англ. – М.: Мир, 1965. - 175 с.
105. Розин Б.Б. Распознавание образов в экономических исследованиях. – М.: Статистика, 1973. – 198 с.
106. Романов А.Н., Одинцов Б.Е. Советующие системы в экономике: Учебное пособие для вузов.– М.: ЮНИТИ-ДАНА, 2000. – 487 с.
107. Сергеев А.В. Прогнозирование временных рядов с помощью нейронных сетей с радиальными базисными функциями//Труды VIII Всероссийской конференции «Нейрокомпьютеры и их применение» Сб.докл, 2002 г. – С. 1187 – 1191.
108. Серебрянников М.Г., Первозванский А.А. Выявление скрытых периодичностей. М.: Наука, 1965. – 244с.
109. Сигеру О., и др. Нейроуправление и его приложения. Пер. с англ. под ред. А.И. Галушкина. – М.: ИПРЖР, 2001. – 321 с.
110. Статевич В.П., Шумков Е.А. Новый принцип построения самообучаемых нейросетей//Труды VIII Всероссийской конференции «Нейрокомпьютеры и их применение» Сб.докл., 2002. – С. 1037 – 1040.
111. Стратонович Р.Л. Теория информации. –М.: Сов. Радио. – 1975. – 424с.
112. Теория прогнозирования и принятия решений. Учеб. пособие. Под. ред. С.А. Саркисяна – М.: Высш. Школа, 1977. – 351 с.
113. Тихонов Э.Е., Бурдо А.И. К вопросу совершенствования автоматизированных систем прогноза//Материалы межрегиональной конференции "Студенческая наука – экономике научно-технического прогресса". Ставрополь: СевКав ГТУ, 2000. – С.30–31.
114. Тихонов Э.Е. Об одном подходе к прогнозированию с помощью нейронных сетей//Материалы третьей МНК "Студенческая наука – экономике России". Ставрополь: СевКав ГТУ, 2002.– С. 69 – 70.

115. Тихонов Э.Е. Об одном подходе к вопросу повышения надежности работы нейронных сетей с использованием непозиционной системы остаточных классов//Материалы XXXII НТК профессорско-преподавательского состава, аспирантов и студентов СевКав ГТУ за 2002 год. Т 2: Технические и прикладные науки г. Ставрополь: СевКав ГТУ, 2003. – С. 29 – 30.
116. Тихонов Э.Е. Об одном подходе к разработке системы прогнозирования с использованием модулярных вычислений на нейронных сетях//Материалы РНК «Теоретические и прикладные проблемы современной физики», Ставрополь: СГУ, 2002. – С. 449 – 453. Тихонов Э.Е. Проблемы идентификации моделей прогнозирования на нейронных сетях//Компьютерная техника и технологии: Сб. трудов регион. НТК. Ставрополь: СевКав ГТУ, 2003. – С.187–191.
117. Тихонов Э.Е. Сравнительный анализ радиально базисной нейронной сети и сети типа – многослойный персептрон на примере прогнозирования объема экспорта//Компьютерная техника и технологии: Сб. трудов регион. НТК. Ставрополь: СевКав ГТУ, 2003. – С.184–186.
118. Тихонов Э.Е., Зайцева И.В. Прогнозирование на основе рядов Фурье//Компьютерная техника и технологии: Сб. трудов регион. НТК. Ставрополь: СевКав ГТУ, 2003. – С.173–175.
119. Тихонов Э.Е., Кучаев А.Г. Об одном подходе к вопросу разработки информационной системы прогнозирования для задач поддержки принятия решений// Компьютерная техника и технологии: Сб. трудов регион. НТК. Ставрополь: СевКав ГТУ, 2003.– С.191–193.
120. Тихонов Э.Е. Об одном подходе к разработке алгоритмов модулярных вычислений на нейронных сетях//Материалы РНК «Теоретические и прикладные проблемы современной физики», Ставрополь: СГУ, 2002. – С.443 – 448.
121. Тихонов Э.Е. Сравнительный анализ традиционных методов прогнозирования с методами прогнозирования на нейронных сетях//Компьютерная техника и технологии: Сб. трудов регион. НТК. Ставрополь: СевКав ГТУ, 2003. – С.179–183.
122. Тихонов Э.Е. Проблемы идентификации моделей прогнозирования на нейронных сетях//Компьютерная техника и технологии: Сб. трудов регион. НТК. Ставрополь: СевКав ГТУ, 2003. – С.187–191.

123. Тихонов Э.Е. Методы и алгоритмы прогнозирования экономических показателей на базе нейронных сетей и модулярной арифметики. Дисс. ...канд. тех. наук. – Ставрополь, 2003.–139с.
124. Томпсон Дж.М.Т. Неустойчивости и катастрофы в науке и технике: Пер. с англ. – М.: Мир, 1985. – 254 с.
125. Трисеев Ю.П. Долгосрочное прогнозирование экономических процессов (системные методы). – Киев: Наукова Думка, 1987. – 132с.
126. Ту Дж., Гонсалес Р. Принципы распознавания образов. – М.: Наука, 1978. – 139 с.
127. Уоссермен Ф. Нейрокомпьютерная техника: теория и практика. – М.: ЮНИТИ, 1992. – 240 с.
128. Федер Е. Фракталы. - М.: Мир, 1991. – 143с.
129. Френкель А.А. Математические методы анализа динамики и прогнозирования производительности труда. – М.: Наука, 1972.
130. Хенан Э. Анализ временных рядов. – М.: Статистика, 1964. – 215с.
131. Хенан Э.Дж. Многомерные временные ряды. – М.: Мир, 1986. – 346с.
132. Ховард Р.А. Динамическое прогнозирование и марковские процессы. – М.: сов. Радио, 1964. – 365с.
133. Червяков Н.И., Тихонов Э.Е. Применение нейронных сетей для задач прогнозирования и проблемы идентификации моделей прогнозирования//Нейрокомпьютеры: разработка, применение. – М.: Радиотехника, – 2003. – № 10. – С.25-31.
134. Червяков Н.И., Тихонов Э.Е. Предсказание фрактальных временных рядов с помощью нейронных сетей // Нейрокомпьютеры: разработка, применение. – М.: Радиотехника, – 2003. – № 10. – С.19-24.
135. Червяков Н.И., Тихонов Э.Е. Совершенствование методов прогнозирования на базе нейронных сетей с использованием позиционной системы остаточных классов//Нейрокомпьютеры: разработка, применение. – М.: Радиотехника, – 2003. – № 10. – С.13-18.
136. Червяков Н.И., Сахнюк П.А. Применение нейроматематики для реализации вычислений в конечных кольцах //Нейрокомпьютеры: разработка, применение, 1999. – № 1.– С. 75–84

137. Червяков Н.И., Шапошников А.В., Сахнюк П.А., Калмыков И.А. Применение модулярных вычислений для нейрообработки//Труды VIII Всероссийской конференции «Нейрокомпьютеры и их применение» НКП, 2002. Под ред. проф. А. Галушкина – М., 2002. – С. 1053–1056.
138. Чуев Ю.В., Михайлов Ю.Б., Кузьмин В.И. Прогнозирование количественных характеристик процессов. М., «Сов. радио», 1975. – 400 с.
139. Шибхузov З.М. Конструктивный TOWER алгоритм для обучения нейронных сетей из $\Sigma\Pi$ – нейронов//Труды VIII Всероссийской конференции «Нейрокомпьютеры и их применение » Сб.докл., 2002. – С. 1066 – 1072.
140. Шустер Г. Детерминированный хаос: Введение: Пер. с англ. – М.: Мир, 1988. – 240 с.
141. Экономика переходного периода. Очерки экономической политики посткоммунистической России 1991 – 1997. – М., 1998.
142. Экономические межотраслевые модели целевого прогнозирования экономики/Б.В. Седелев, и др.; ВНИИСИ. – Препр. – М., 1987. – 59с.
143. Ямпольский С.М., Хилюк Ф.М., Лисичкин В.А. Проблемы научно-технического прогнозирования. М.: Экономика, 1969. – 189 с.
144. Almon S. (1960) “The Distributed Lag between Capital Appropriations and Expenditures”, *Econometrica*, 30, 178-196.
145. Andrews D.W.K. (1991) “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59, 817–858.
146. Ardeni P.G., D. Lubian (1991) “Is There Trend Reversion in Purchasing Power Parity”, *European Economic Review*, 35, 1035-1055.
147. Banerjee A., R.L. Lumsdaine, J.H. Stock (1992) “Recursive and Sequential Tests of the Unit Root and Trend Break Hypotheses: Theory and International Evidence”, *Journal of Business and Economic Statistics*, 10, 271-287.
148. Bierens H.J. (1997) “Testing the Unit Root with Drift Hypothesis Against Nonlinear Trend Stationarity, with an Application to the US Price Level and Interest Rate”, *Journal of Econometrics*, 81, 29-64.
149. Bollerslev, Tim (1986) “Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, 31, 307–327.

150. Brown R.G. (1962) "Smoothing, Forecasting and Prediction of Discrete Time-Series". Prentice-Hall, New Jersey.
151. Brown R.G. (1963) "Smoothing, Forecasting and Prediction". Prentice-Hall, Englewood Cliffs, N.Y.
152. Cagan P. (1956) "The Monetary Dynamics of Hyperinflation", in: "Studies in the Quantity Theory of Maney". Chicago, University of Chicago Press.
153. Chan K.H, J.C.Hayya, J.K.Ord (1977) "A Note on Trend Removal Methods: The Case of polynomial versus variate differencing", *Econometrica*, 45, 737-744.
154. Cheung Y.-W., K.S.Lay (1995) "Lag Order and Critical Values of a Modified Dickey-Fuller Test", *Oxford Bulletin of Economics and Statistics*, 57, №3, 411-419.
155. Cheung Y.-W., M.D. Chinn (1996) "Deterministic, Stochastic, and Segmented Trends in Aggregate Output: a Cross-country Analysis", *Oxford Economic Papers*, 48, №1, 134-162.
156. Cheung Y.-W., M.D. Chinn (1997) "Further Investigation of the Uncertain Unit Root in GDP", *Journal of Business and Economic Statistics*, 15, 68-73.
157. Christiano L.J., M. Eichenbaum (1990) "Unit Roots in Real GDP: Do We Know, and Do We Care?", *Carnegie-Rochester Conference Series on Public Policy*, 32, 7-62.
158. Clark P.K. (1989) "Trend Reversion in Real Output and Unemployment", *Journal of Econometrics*, 40,15-32.
159. Cochrane J.H. (1998) "How Big is the Random Walk in GNP?", *Journal of Political Economy*, 96, 893-920.
160. Cogley T. (1990) "International Evidence on the Size of the Random Walk in Output", *Journal of Political Economy*, 98, 501-518.
161. Copeland L.S. (1991) "Cointegration Tests with Daily Exchange Rate Data", *Oxford Bulletin of Economics and Statistics*, 53, 185-198.
162. Davidson R., J.G. MacKinnon (1993) *Estimation and Inference in Econometrics*, Oxford University Press
163. den Haan W.J. (2000) "The Comovement Between Output and Prices", *Journal of Monetary Economics*, 46, №1, 3-30.
164. Dickey D.A. (1976) "Estimation and Hypothesis Testing for Non-stationary Time Series", Ph.D. dissertation, Iowa State University.

165. Dickey D.A., S. Pantula (1987) "Determining the Order of Differencing in Autoregressive Processes", *Journal of Business and Economic Statistics*, 15, 455-461.
166. Dickey D.A., W.A. Fuller (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, 74, 427-431.
167. Dickey D.A., W.R. Bell, R.B. Miller (1986) "Unit Roots in Time Series Models: Tests and Implications", *American Statistician*, 40, 12-26.
168. Dickey, D.A., W.A. Fuller (1981) "Likelihood Ratio Statistics for Autoregressive Time Series With a Unit Root", *Econometrica*, 49, 1057-1072.
169. Dolado H., T. Jenkinson, S. Sosvilla-Rivero (1990) "Cointegration and Unit Roots", *Journal of Economic Surveys*, 4, 243-273.
170. Dutt S.D. (1998) "Purchasing Power Parity Revisited: Null of Cointegration Approach", *Applied Economic Letters*, 5, 573-576.
171. Dutt S.D., D. Ghosh (1999) "An Empirical Examination of Exchange Market Efficiency", *Applied Economic Letters*, 6, №2, 89-91.
172. Dwyer G.P., Wallace M.S. (1992) "Cointegration and Market Efficiency", *Journal of International Money and Finance*, 11 318-327.
173. Elliott G., T.J. Rothenberg, J.H. Stock (1996) "Efficient Tests for an Autoregressive Unit Root", *Econometrica*, 64, 813-836.
174. Enders W. (1995) "Applied Econometric Time Series", Wiley, New York
175. Engle R., Kraft D. (1983) *B Applied Time Series Analysis of Economic Data*, Washington D.C.: Bureau of the Census.
176. Engle R.F., C.W.J. Granger (1991) "Cointegrated Economic Time Series: An Overview with New Results", in R.F. Engle and C.W.J. (eds.), *Long-Run Economic Relationships, Readings in Cointegration*, Oxford University Press, 237-266.
177. Engle, R. (1983) "Estimates of the Variance of U.S. Inflation Based on the ARCH Model," *Journal of Money, Credit and Banking*, 15, 286-301.
178. Engle, R. F. (1982) "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1008.

179. Fama E.F., French K.R. (1988) "Permanent and Temporary Components of Stock Prices", *Journal of Political Economy*, 96, 246-273.
180. Friedman M. (1957) "Theory of the Consumption Function". Princeton, N.J.: Princeton University Press.
181. Fuller W.A. (1976) *Introduction to Statistical Time Series*, Wiley, New York.
182. Fuller W.A. (1996) *Introduction to Statistical Time Series*, 2nd Ed, Wiley, New York
183. Funke N., J. Thornton (1999) "The demand for money in Italy, 1861-1988", *Applied Economic Letters*, 6, №5, 299-301.
184. Ghysels E., H.S. Lee, J. Noh (1994) "Testing for Unit Roots in Seasonal Time Series: Some Theoretical Extensions and a Monte Carlo Investigation", *Journal of Econometrics*, 62, 415-442.
185. Ghysels E., Perron P. (1992) "The Effect of Seasonal Adjustment Filters on Tests for a Unit Root", *Journal of Econometrics*, 55, 57-98.
186. Gragg J. (1983) "More Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form", *Econometrica*, 51, 751-763.
187. Granger C.W.J. (1963) "The Effect of Varying Month-Length in the Analysis of Economic Time Series", *L'Industria*, 1, 3, Milano.
188. Green W.H. (1997) "Econometric Analysis". 3rd edition, Prentice-Hall.
189. Hafer R.W., D.W. Jansen (1991) "The Demand for Money in the United States: Evidence from Cointegration Tests", *Journal of Money, Credit, and Banking*, 23 (1991), 155-168.
190. Hall A. (1994) "Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection", *Journal of Business and Economic Statistics*, 12, 451-470.
191. Hamilton, James D. (1994) *Time Series Analysis*, Princeton University Press, Princeton.
192. Hasan M.S. (1998) "The Choice of Appropriate Monetary Aggregate in the United Kingdom", *Applied Economic Letters*, 5, №9, 563-568.
193. Hatanaka M. (1996) *Time Series-Based Econometrics: Unit Roots and Cointegration*, Oxford University Press.
194. Holden D., Perman R. (1994) "Unit Roots and Cointegration for Economist", в сборнике *Cointegration for the Applied Economists* (редактор Rao B.B.), Macmillan.

195. Holt C.C. (1957) "Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages", Carnegie Inst. Tech. Res. Mem., 52.
196. Johansen S., K. Juselius (1990) "Maximum Likelihood Estimation and Inferences on Cointegration—with applications to the demand for money," Oxford Bulletin of Economics and Statistics, 52, 169–210.
197. Kim B. J.C., Mo Soowon (1995) "Cointegration and the long-run forecast of exchange rates", Economics Letters, 48, №№ 3-4, 353-359.
198. Kolmogoroff A. (1939) "Sur L'interpolation et L'extrapolation des Suites Stationnaires", Compt. Rend., 208, 2043.
199. Koyck L.M. (1954) "Distributed Lags and Investment Analysis". North-Holland Publishing Company, Amsterdam.
200. Kwiatkowski D., P.C.B. Phillips, P. Schmidt, Y. Shin (1992) "Testing of the Null Hypothesis of Stationary against the Alternative of a Unit Root", Journal of Econometrics, 54, 159-178.
201. Leybourne S., T. Mills, P. Newbold (1998) "Spurious Rejections by Dickey-Fuller Tests in the Presence of a Break Under Null", Journal of Econometrics, 87, 191-203.
202. Leybourne S.J. (1995) "Testing for Unit Roots Using Forward and Reverse Dickey-Fuller Regressions", Oxford Bulletin of Economics and Statistics, 57, 559-571.
203. Lumsdaine R.L., Kim I.M. (1997) "Structural Change and Unit Roots", The Review of Economics and Statistics, 79, 212-218.
204. MacKinnon, J.G. (1991) "Critical Values for Cointegration Tests," Глава 13 в Long-run Economic Relationships: Readings in Cointegration, edited by R.F.Engle and C.W.J. Granger, Oxford University Press.
205. Maddala G.S., In-Moo Kim (1998) Unit Roots, Cointegration, and Structural Change. Cambridge University Press, Cambridge.
206. Metin K. (1995) "An Integrated Analysis of Turkish Inflation", Oxford Bulletin of Economics and Statistics, 57, №4, 513-532.
207. Milas C. (1998) "Demand for Greek Imports Using Multivariate Cointegration Technique", Applied Economics, 30, №11, 1483-1492.
208. Mills T.C. (1993) The Econometric Modeling of Financial Time Series. Cambridge University Press, Cambridge.

209. Molana H. (1994) "Consumption and Fiscal Theory. UK Evidence from a Cointegration Approach", Dundee Discussion Papers, University of Dundee, Dundee, Scotland.
210. Murray C.J., C.R. Nelson (2000) "The Uncertain Trend in U.S. GDP", *Journal of Monetary Economics*, 46, 79-95.
211. Nadal-De Simone F., W.A. Razzak (1999) "Nominal Exchange Rates and Nominal Interest Rate Differentials", IMF Working Paper WP/99/141.
212. Nelson C.R., C.I. Plosser (1982) "Trends and Random Walks in Macroeconomic Time Series", *Journal of Monetary Economics*, 10, 139-162.
213. Nelson C.R., H. Kang (1981) "Spurious Periodicity in Inappropriately Detrended Time Series", *Journal of Monetary Economics*, 10, 139-162.
214. Nerlove M. (1956) "Estimates of the Elasticities of Supply of Selected Agricultural Commodities", *Jorn. Farm Econ.*, 38, 496-509.
215. Nerlove M. (1958) "The Dynamics of Supply: Estimation of Farmers Response to Price". The Johns Hopkins Press. Baltimore.
216. Newey W., K. West (1987) "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708
217. Newey W., K. West (1994) "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631-653.
218. Ng S., P. Perron (1995) "Unit Root Tests in ARMA models With Data-Dependent Methods for the Selection of the Truncation Lag", *Journal of American Statistical Association*, 90, 268-281.
219. Nunes L.S., Newbold P., C.-M. Kuan (1997) "Testing for Unit Roots With Breaks. Evidence on the Great Crash and the Unit Root Hypothesis Reconsidered", *Oxford Bulletin of Economics and Statistics*, 59, №4, 435-448.
220. Perron P. (1988) " Trends and Random Walks in Macroeconomic Time Series: Furter Evidence from a New Approach", *Journal of Economic Dynamic and Control*, 12, 297-332.
221. Perron P. (1989a) "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis, *Econometrica*, 577, 1361-1401.
222. Perron P. (1989b) "Testing for a Random Walk: A Simulation Experiment When the Sampling Interval Is Varied" – в сборнике Ad-

- vances in Econometrics and Modeling (редактор B.Ray), Kluwer Academic Publishers, Dordrecht and Boston.
223. Perron P. (1997) "Further evidence on breaking trend functions in macroeconomic variables, *Journal of Econometrics*, 80, №2, 355-385.
 224. Perron P., Vogelsang T.J. (1993) "Erratum", *Econometrica*, 61, №1, 248-249.
 225. Phillips P.C.B. (1987) "Time Series Regression with a Unit Root", *Econometrica*, 55, 277-301.
 226. Phillips P.C.B., P. Perron (1988) "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335-346.
 227. Said E., D.A. Dickey (1984) "Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order," *Biometrika*, 71, 599-607.
 228. Schmidt P., Phillips P.C.B. (1992) "LM Tests for a Unit Root in the Presence of Deterministic Trends", *Oxford Bulletin of Economics and Statistics*, 54, 257-287.
 229. Schwert G.W. (1989) "Tests for Unit Roots: A Monte Carlo Investigation", *Journal of Business and Economic Statistics*, 7, 147-159.
 230. Shiller R.J., Perron P. (1985) "Testing the Random Walk Hypothesis: Power versus Frequency of Observation", *Economic Letters*, 18, 381-386.
 231. Solow R.M. (1960) "On a Family of Lag Distributions", *Econometrica*, 28, 393-406.
 232. Taylor A.M.R. (2000) "The Finite Sample Effects of Deterministic Variables on Conventional Methods of Lag-Selection in Unit-Root Tests", *Oxford Bulletin of Economics and Statistics*, 62, 293-304.
 233. Walker G. (1931) "On Periodicity in Series of Related Terms", *Proc. Royal Soc.*, A131, 518.
 234. White H., I. Domovitz (1984) "Nonlinear Regression with Dependent Observations", *Econometrica*, 52, 143-162.
 235. Wiener N. (1949) "Extrapolation, Interpolation and Smoothing of Stationary Time Series". John Wiley, New York.
 236. Winters P.R. (1960) "Forecasting Sales by Exponentially Weighted Moving Averages", *Mgmt. Sci.*, 6, 324.
 237. Wold H.O. (1932) "A Study in the Analysis of Stationary Time Series". Almquist and Wiksell, Uppsala.
 238. Woodward G., R. Pillarisetti (1999) "Empirical Evidence on Alternative Theories of Inflation and Unemployment: a Re-Evaluation

- for the Scandinavian Countries”, Applied Economic Letters, 6, №1, 55-58.
239. Yule G.U. (1927) “On a Method of Investigating Periodicities in Disturbed Series”, Phil. Trans., A226, 227.
 240. Zivot E., Andrews D.W.K. (1992) “Further Evidence on the Great Crash, the Oil Price Shock and the Unit Root Hypothesis”, Journal of Business and Economic Statistics, 10, 251-270.
 241. (EHIPS) Генетические алгоритмы. Режим доступа [<http://www.iki.rssi.ru/ehips/genetic.htm> 29.08.2002]
 242. Аргуткина Н.Л. О совершенствовании методов прогнозирования, основанных на экспоненциальном сглаживании. Конф. Маркетологов ВНИК «Прогнозирование» Режим доступа [<http://www.marketing.spb.ru/conf> 10.04.2001]
 243. Билл Вильямс Торговый Хаос – М.: ИК Аналитика, 2000. – 328 с. Режим доступа [<http://yamdex.narod.ru/books/Williams/Tradxaos/Ссылки.htm>]
 244. Генетические алгоритмы и машинное обучение. Режим доступа [http://www.math.tsu.ru/russian/center/ai_group/ai_collection/docs/faqs/ai/part5/faq3.html 29.08.2002]
 245. Генетические алгоритмы обучения. Режим доступа [<http://www.hamovniki.net/~alchemist/NN/DATA/Gonchar/Main.htm> 29.08.2002]
 246. Лекции по нейронным сетям и генетическим алгоритмам. Режим доступа [<http://infoart.baku.az/inews/30000007.htm> 29.08.2002]
 247. Прогностика. Терминология, вып. 92. – М.: «Наука», 1978. Режим доступа [<http://www.icc.jamal.ru/library/koi/POLITOLOG/bunchuk.txt> 10.04.2001]
 248. Яковлев В.Л., Яковлева Г.Л., Лисицкий Л.А. Применение нейросетевых алгоритмов к анализу финансовых рынков. Режим доступа [<http://neurnews.iu4.bmstu.ru/neurnews.html>. 29.08.2002]
 249. Яковлев В.Л., Яковлева Г.Л., Лисицкий Л.А. Создание математических моделей прогнозирования тенденций финансовых рынков, реализуемых при помощи нейросетевых алгоритмов. Режим доступа [<http://neurnews.iu4.bmstu.ru/neurnews.html>. 29.08.2002]

Научное издание

Тихонов Эдуард Евгеньевич
Кузьмищев Валерий Анатольевич

Методы и алгоритмы прогнозирования экономических показателей на базе нейронных сетей и модулярной арифметики

Монография

Редактор Э.Е. Тихонов
Корректурa авторов

Лицензия ИД № 03184 от 10.11.2000

Подписано в печать
Бумага-офсетная. Гарнитура-«Таймс».
Усл. печ. л. 9,5. Уч.-изд. л. 8,1. Формат 60×84/16
Тираж 500 экз.

Издатель- Невинномысский институт экономики,
Управления и права (НИЭУП)
357101, г. Невинномысск,
ул. З.Космодемьянской, 1
Отпечатано-НИЭУП