

На правах рукописи



Журов Александр Викторович

**СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ
ПРИ АНАЛИЗЕ ВРЕМЕННЫХ РЯДОВ**

05.13.01 – Системный анализ, управление и обработка информации

Автореферат

диссертации на соискание ученой степени
кандидата технических наук

Красноярск – 2007

Работа выполнена в Государственном образовательном учреждении высшего профессионального образования «Красноярский государственный аграрный университет»

Научный руководитель: кандидат физико-математических наук,
доцент Кузнецов Александр Алексеевич

Официальные оппоненты: доктор технических наук,
профессор Семенин Евгений Станиславович

доктор технических наук,
профессор Дулесов Алексей Сергеевич


Ведущая организация: Томский государственный педагогический университет

Защита состоится 15 ноября 2007 года в 14 часов на заседании диссертационного совета Д212.249.02 при Сибирском государственном аэрокосмическом университете им. ак. М.Ф. Решетнева по адресу: 660014, г. Красноярск, пр. им. газ. «Красноярский рабочий», 31.

С диссертацией можно ознакомиться в библиотеке Сибирского государственного аэрокосмического университета.

Автореферат разослан 15 октября 2007 года.

Ученый секретарь
диссертационного совета



И.В. Ковалев

2007 А
237 УФ

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы. Одной из основных составляющих прогнозирования является информационно-математическое моделирование, включающее в себя как частный случай анализ временных рядов.

Все современные теории прогнозирования в той или иной степени используют математический аппарат, связанный с понятием временного ряда. Поэтому задача изучения анализа временных рядов актуальна как отдельная фундаментальная проблема, так и прикладная проблема конкретного частного случая (оценка параметров зависящих от времени технологических процессов; рынок ценных бумаг и т.д.)

Моделями, в основе которых лежат временные ряды занимались такие известные ученые как С.А. Айвзян, И.В. Бестужев-Лада, А.И. Орлов, Е.М. Четыркин, Дж. Бокс, Р. Браун, Г. Джекинсон, С. Хольт, П. Уинтерс и др.

Особую роль в современном менеджменте (принятии решений) играет стратегическое прогнозирование, включающее: выработку главной цели бизнеса; прогнозирование как предвидение результатов развития, происходящего под воздействием существующих факторов; перспективное планирование в качестве системы мер, необходимых для преодоления отклонения прогнозируемых итогов от установленных параметров.

Органической частью планирования является составление прогнозов, показывающих возможные направления будущего развития экономических, финансовых, социальных и промышленных структур, рассматриваемой в тесном взаимодействии с окружающей средой. Вся как плановая, так и практическая работа в организации связана с необходимостью прогнозирования. Каждый менеджер и специалист по планированию должен владеть основными навыками и технологией прикладного информационно-математического прогнозирования для принятия грамотных, обоснованных решений.

Цель диссертационного исследования состоит в создании системы поддержки принятия решений, основанной на анализе временных рядов, для повышения эффективности прогнозов.

Поставленная в диссертации цель достигается путем решения следующих задач:

- анализа известных моделей и методов прогнозирования;
- модификации имеющихся моделей для осуществления более эффективных прогнозов;
- разработки комплексной системы прогнозирования, в основе которой лежит анализ временных рядов;
- содержательной реализации указанной системы на рынке цен вторичного жилья г. Красноярска.



Методы исследования. Основные теоретические и прикладные результаты работы получены на основе методологии системного анализа, а также информационных технологий и методов фундаментальной и прикладной математики.

Научная новизна работы:

1. Разработан метод подбора параметров в моделях экспоненциального сглаживания для осуществления более эффективных прогнозов.
2. Разработана методика построения взвешенного прогноза, основанная на результатах прогноза отдельно взятых моделей.
3. Разработана комплексная система прогнозирования, основанная на анализе временных рядов.
4. Реализована процедура анализа и прогнозирования цен на вторичном рынке жилья г. Красноярска. Показано, что наиболее эффективным методом прогнозирования является взвешенный прогноз.

Значение для теории. Результаты, полученные при выполнении диссертационной работы, являются развитием моделей и методов прогнозирования.

Практическая ценность. Разработанная в диссертации система прогнозирования временных рядов позволяет анализировать и прогнозировать показатели в социальных, экономических, финансовых и промышленных структурах. Архитектура разработанной системы позволяет предотвратить потерю информации, что важно для достоверного прогноза в данных структурах.

Реализация результатов работы. Полученные в диссертационной работе результаты внедрены

- 1) в работу компании «Брокер Кредит Сервис» (г. Красноярск);
- 2) в учебный процесс ФГОУ ВПО «Красноярский государственный аграрный университет».

На защиту выносятся:

1. Метод подбора параметров в моделях экспоненциального сглаживания для осуществления более эффективных прогнозов.
2. Методика построения взвешенного прогноза, основанная на результатах прогноза различных моделей.
3. Комплексная система прогнозирования, основанная на анализе временных рядов.
4. Результаты реализации процедуры анализа и прогнозирования цен на вторичном рынке жилья г. Красноярска.

Апробация работы. Основные положения и результаты работы прошли всестороннюю апробацию на международных и всероссийских научных, научно-практических и научно-технических конференциях. В том числе, на международной конференции «Аграрная наука на рубеже веков» (2006 г., Красноярск); межрегиональной конференции «Актуальные проблемы строительства» (2006 г., Красноярск); всероссийской конференции «Актуальные социально-экономические проблемы развития АПК» (2007 г., Красноярск); международной конференции «Студент и научно-технический прогресс» (2007 г., Новосибирск); международной конференции «Решетневские чтения» (2007 г., Красноярск).

Диссертационная работа неоднократно обсуждалась на научных семинарах Красноярского государственного аграрного университета и Сибирского государственного аэрокосмического университета.

Публикации. По материалам диссертации опубликовано 8 работ (2 из них по списку ВАК). Полный список публикаций представлен в конце автореферата.

Общая характеристика диссертации. Работа состоит из введения, трех глав, заключения и библиографического списка, включающего 57 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении дана общая характеристика проблемы, обоснована актуальность выбранной темы, определены цель и задачи исследования. Сформулированы основные положения, выносимые на защиту, научная новизна, теоретическая и практическая значимость полученных результатов.

В первой главе рассматриваются основные определения, а также цели и задачи анализа временных рядов.

Ряд наблюдений

$$x(t_1), x(t_2), \dots, x(t_N) \quad (1)$$

анализируемой случайной величины $\xi(t)$, произведенных в последовательные моменты времени t_1, t_2, \dots, t_N , называется *временным рядом*.

Совокупность существующих методов и моделей исследования таких рядов зависимых наблюдений называется *анализом временных рядов*.

Главная цель анализа временных рядов состоит в описании по возможности простых и экономично параметризованных моделей, адекватно описывающих имеющиеся ряды наблюдений и составляющих базу для решения, в первую очередь, следующих задач:

- вскрытие *механизма генезиса* наблюдений, составляющих анализируемый временной ряд;

- построение *оптимального* (т. е. наиболее точного) прогноза для будущих значений временного ряда;
- выработка *стратегии управления и оптимизации* анализируемых процессов.

Выделим 4 основных группы факторов, под воздействием которых формируются значения элементов временного ряда.

(А) *Долговременные*, формирующие общую (в длительной перспективе) тенденцию в изменении анализируемого признака $x(t)$. Обычно эта тенденция описывается с помощью той или иной неслучайной функции $f_{mp}(t)$, как правило, монотонной. Эту функцию называют функцией тренда или просто – трендом.

(Б) *Сезонные*, формирующие периодически повторяющиеся в определенное время года колебания анализируемого признака. Обозначим результат действия сезонных факторов в виде неслучайной функции $\varphi(t)$. Поскольку эта функция должна быть периодической (с периодами, кратными "сезонам"), в ее аналитическом выражении участвуют гармоники (тригонометрические функции), периодичность которых, как правило, обусловлена содержательной сущностью задачи.

(В) *Циклические (конъюнктурные)*, формирующие изменения анализируемого признака, обусловленные действием долговременных циклов экономической, демографической или астрофизической природы (волны Кондратьева, демографические "ямы", циклы солнечной активности и т.п.). Результат действия циклических факторов будем обозначать с помощью неслучайной функции $\psi(t)$.

(Г) *Случайные* (нерегулярные), не поддающиеся учету и регистрации. Их воздействие на формирование значений временного ряда как раз и обуславливает стохастическую природу элементов $x(t)$, а, следовательно, и необходимость интерпретации $x(1), x(2), \dots, x(N)$ как наблюдений, произведенных над случайными величинами, соответственно, $\xi(1), \xi(2), \dots, \xi(N)$. Будем обозначать результат воздействия случайных факторов с помощью случайных величин ("остатков", "ошибок") $\varepsilon(t)$.

Конечно, вовсе не обязательно, чтобы в процессе формирования значений всякого временного ряда участвовали одновременно факторы всех четырех типов. *Аддитивная архитектурная схема* влияния факторов А, Б, В, Г на формирование значений $x(t)$, которая означает правомерность представления значений членов временного ряда в виде разложения:

$$x(t) = f_{mp}(t) + \varphi(t) + \psi(t) + \varepsilon(t), \quad t = 1, 2, \dots, N. \quad (2)$$

Выводы о том, участвуют или нет факторы данного типа в формировании значений $x(t)$, могут базироваться как на анализе содержательной сущности задачи (т.е. быть априорно-экспертным по своей природе), так и на специальном статистическом анализе исследуемого временного ряда.

Отправляясь от приведенного выше аддитивного разложения временного ряда $x(t)$, можно дать общую формулировку базисной цели его статистического анализа: по имеющейся траектории анализируемого временного ряда $x(t)$ требуется решить следующие задачи:

- (i) определить, какие из неслучайных функций $f_{np}(t)$, $\varphi(t)$ и $\psi(t)$ присутствуют в разложении (2);
- (ii) построить "хорошие" оценки для тех неслучайных функций, которые присутствуют в разложении (2);
- (iii) подобрать модель, адекватно описывающую поведение "случайных остатков" $\varepsilon(t)$, и статистически оценить параметры этой модели.

Успешное решение задач (i)-(iii), обусловленных базисной целью статистического анализа временного ряда, является основой для достижения конечных прикладных целей исследования, в первую очередь, для решения задачи кратко- и среднесрочного прогноза значений временного ряда.

В главе 2 рассматривается системный подход к анализу временных рядов. Приступая к анализу дискретного ряда наблюдений, расположенных в хронологическом порядке, следует, в первую очередь, убедиться, действительно ли в формировании значений этого ряда участвовали какие-либо факторы, помимо чисто случайных. При этом под «чисто случайными» понимаются лишь те случайные факторы, под воздействием которых генерируются последовательности взаимно не коррелированных и одинаково распределенных случайных величин, обладающих постоянными (не зависящими от времени) средними значениями и дисперсиями. Ответ на поставленный вопрос получают, проводя статистическую проверку соответствующей гипотезы с помощью «критериев серий».

Если в результате проверки такой статистической гипотезы выяснилось, что имеющиеся наблюдения взаимно зависимы и неодинаково распределены (т.е. образуют действительно временной ряд), то приступают к подбору подходящей модели для этого ряда. Множество моделей, в рамках которого ведется этот подбор, ограничивается обычно тремя классами моделей:

- модели авторегрессии и проинтегрированного скользящего среднего (АРСС-модели);
- адаптивные модели одно-, двух-, трехпараметрического экспоненциального сглаживания (метод Брауна, метод Хольта, метод Хольта-Уинтерса);
- модели, основанные на кривых роста.

Далее рассматриваются методы оценок параметров моделей, их адекватности и точности. Для того чтобы модель была адекватна, необходимо выполнение следующих условий для величины ошибки $\varepsilon(t)$:

- случайность колебаний $\varepsilon(t)$;

- соответствие распределения $\varepsilon(t)$ нормальному закону распределения;
- равенство математического ожидания $\varepsilon(t)$ нулю;
- независимость значений уровней $\varepsilon(t)$.

В моделях экспоненциального сглаживания подбор параметров предлагается осуществлять методом минимизации относительной суммарной погрешности.

В случае однопараметрического сглаживания

$$y_{t+1} = \alpha x_t + (1 - \alpha)y_t,$$

где x_t – наблюдаемые значения и y_t – прогнозируемые значения. Здесь, при $t \geq 2$ все получаемые y , будут зависеть от y_1 и α . Пусть $\alpha_j = \frac{j}{n}$, ($j = 0, 1, 2, \dots, n$) и $y_1 = \{y_1^{(1)}, y_1^{(2)}, \dots, y_1^{(p)}, \dots, y_1^{(r)}\}$. Введем функцию суммарной относительной погрешности:

$$S(y_1^{(p)}, \alpha_j) = \sum_{i=1}^N \left| \frac{y_i(y_1^{(p)}, \alpha_j) - x_i}{x_i} \right|.$$

Затем, находим минимум функции S :

$$S(y_1^{(p)}, \alpha_j) = \min_{\substack{j=0,1,2,\dots,n \\ p=1,2,\dots,r}} S(y_1^{(p)}, \alpha_j).$$

В дальнейшем будем использовать $y_1^{(p)}$ и α_j .

В случае двухпараметрического сглаживания

$$\begin{cases} y_{t+1} = A_t + B_t, \\ A_t = \alpha x_t + (1 - \alpha)(A_{t-1} + B_{t-1}), \\ B_t = \beta(A_t - A_{t-1}) + (1 - \beta)B_{t-1}. \end{cases}$$

Пусть $\alpha_j = \frac{j}{n}$, $\beta_k = \frac{k}{n}$ ($j, k = 0, 1, 2, \dots, n$), $A_1 = \{A_1^{(1)}, A_1^{(2)}, \dots, A_1^{(p)}, \dots, A_1^{(r)}\}$ и $B_1 = \{B_1^{(1)}, B_1^{(2)}, \dots, B_1^{(q)}, \dots, B_1^{(r)}\}$. Введем функцию суммарной относительной погрешности:

$$S(A_1^{(p)}, B_1^{(q)}, \alpha_j, \beta_k) = \sum_{i=1}^N \left| \frac{y_i(A_1^{(p)}, B_1^{(q)}, \alpha_j, \beta_k) - x_i}{x_i} \right|.$$

После чего, находим минимум функции S :

$$S(A_1^{(p^*)}, B_1^{(q^*)}, \alpha_{j^*}, \beta_{k^*}) = \min_{\substack{j=0,1,2,\dots,n \\ k=0,1,2,\dots,n \\ p=1,2,\dots,r_A \\ q=1,2,\dots,r_B}} S(A_1^{(p)}, B_1^{(q)}, \alpha_j, \beta_k).$$

В дальнейшем будем использовать $j^*, k^*, A_1(p^*)$ и $B_1(q^*)$.

В случае трехпараметрического сглаживания

$$\begin{cases} y_{t+1} = (A_t + B_t)C_{t+1-T}, \\ A_t = \alpha \frac{x_t}{C_{t-T}} + (1-\alpha)(A_{t-1} + B_{t-1}), \\ B_t = \beta(A_t - A_{t-1}) + (1-\beta)B_{t-1}, \\ C_t = \gamma \frac{x_t}{A_t} + (1-\gamma)C_{t-T}. \end{cases}$$

Пусть $\alpha_j = \frac{j}{n}$, $\beta_k = \frac{k}{n}$, $\gamma_s = \frac{s}{n}$ ($j, k, s = 0, 1, 2, \dots, n$),

$A_1 = \{A_1^{(1)}, A_1^{(2)}, \dots, A_1^{(p)}, \dots, A_1^{(r_A)}\}$, $B_1 = \{B_1^{(1)}, B_1^{(2)}, \dots, B_1^{(q)}, \dots, B_1^{(r_B)}\}$, и $C_h = \{C_h^{(1)}, C_h^{(2)}, \dots, C_h^{(v_h)}, \dots, C_h^{(r_C)}\}$ ($h=1, 2, \dots, T$).

Введем функцию суммарной относительной погрешности:

$$S(A_1^{(p)}, B_1^{(q)}, C_h^{(v_h)}, \alpha_j, \beta_k, \gamma_s) = \sum_{i=1}^N \left| \frac{y_i(A_1^{(p)}, B_1^{(q)}, C_h^{(v_h)}, \alpha_j, \beta_k, \gamma_s) - x_i}{x_i} \right|.$$

После чего, находим минимум функции S :

$$S(A_1^{(p^*)}, B_1^{(q^*)}, C_h^{(v_h^*)}, \alpha_{j^*}, \beta_{k^*}, \gamma_{s^*}) = \min_{\substack{j=0,1,2,\dots,n \\ k=0,1,2,\dots,n \\ s=0,1,2,\dots,n \\ p=1,2,\dots,r_A \\ q=1,2,\dots,r_B \\ v_h=1,2,\dots,r_C}} S(A_1^{(p)}, B_1^{(q)}, C_h^{(v_h)}, \alpha_j, \beta_k, \gamma_s).$$

В дальнейшем будем использовать $j^*, k^*, s^*, A_1(p^*)$ и $B_1(q^*)$ и $C_h(v_h^*)$ ($h=1, 2, \dots, T$).

Для оценки точности прогнозных свойств модели используется ретроспективный прогноз – подход, основанный на выделении участка из ряда последних уровней исходного временного ряда.

Затем строится прогноз на будущее для адекватных моделей. После чего вычисляется взвешенный прогноз, учитывающий прогнозные значения каждой из адекватных моделей.

Рассмотрим два подхода построения взвешенного прогноза, учитывающего прогнозные значения и ошибки каждой из моделей.

Пусть для прогнозирования значений временного ряда (1) используются различные методы и модели M_1, M_2, \dots, M_k . Будем считать, что применительно к ряду (1) данные методы являются адекватными. Пусть, также известно, что в результате проведения n экспериментов в прошлом, модель M_j обеспечивала наилучший прогноз m_j раз. На основе ретроспективного прогноза для каждой модели M_j ($j=1, 2, \dots, k$) рассчитаем относительную ошибку прогноза δ_j для ряда (1). Затем вычислим прогнозные значения $y^{(j)}_{t+l}$ ($l=1, 2, \dots$) по каждому методу M_j .

1. Взвешенный прогноз на основе матрицы парных сравнений. На основе значений m_j и n вычислим матрицу парных сравнений $P_{k \times k}$, значения которой равны $p_{js} = \frac{m_j}{m_s}$:

$$P_{k \times k} = \begin{pmatrix} m_1 / m_1 & \dots & m_1 / m_k \\ m_2 / m_1 & \dots & m_2 / m_k \\ \dots & \dots & \dots \\ m_k / m_1 & \dots & m_k / m_k \end{pmatrix}$$

Далее рассчитаем весовые коэффициенты w_j для каждой модели M_j :

$$w_j = \frac{\sum_{s=1}^k p_{js}}{\sum_{j=1}^k \sum_{s=1}^k p_{js}}$$

Нетрудно проверить, что $\sum_{j=1}^k w_j = 1$.

После чего получаем взвешенный прогноз y_{t+l} :

$$y_{t+l} = \sum_{j=1}^k w_j y^{(j)}_{t+l}. \quad (3)$$

2) Взвешенный прогноз на основе ошибок ретроспективного прогноза. Вычислим суммарную относительную ошибку:

$$\delta = \sum_{j=1}^k \delta_j.$$

Весовые коэффициенты w_j в данном случае будут равны

$$w_j = \frac{\delta - \delta_j}{(k-1)\delta}.$$

Нетрудно убедиться, что $\sum_{j=1}^k w_j = 1$. Взвешенный прогноз осуществляем по формуле (3).

Общая схема комплексного анализа временных рядов представлена на рис. 1.

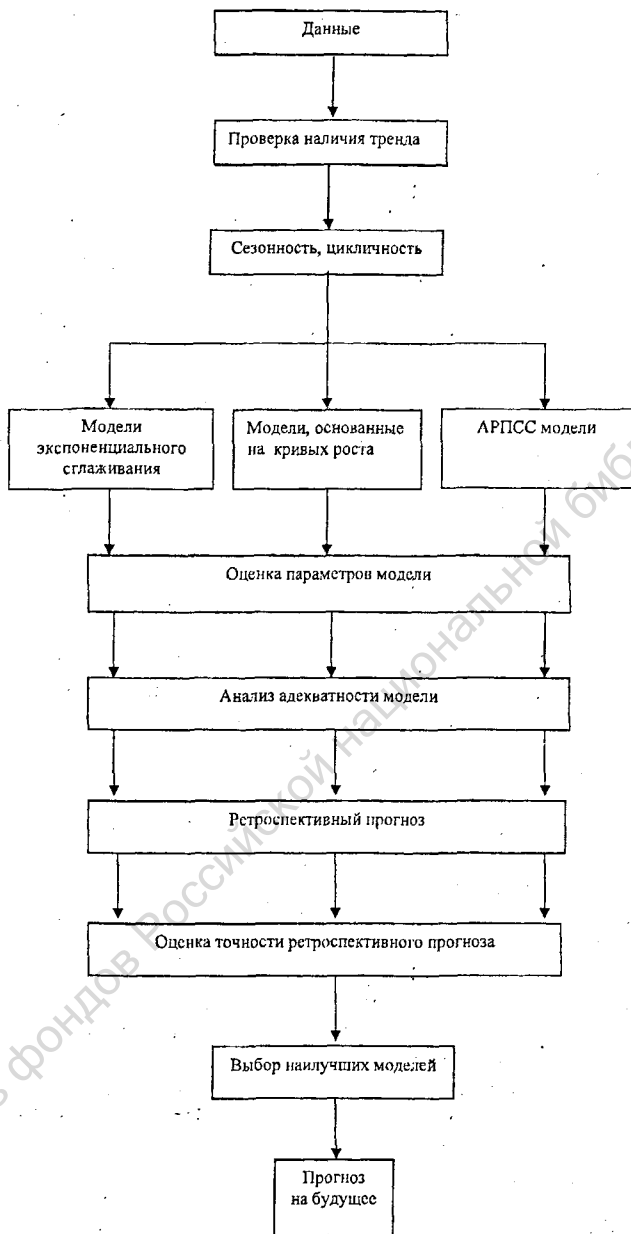


Рисунок 1. Общая схема комплексного анализа временных рядов.

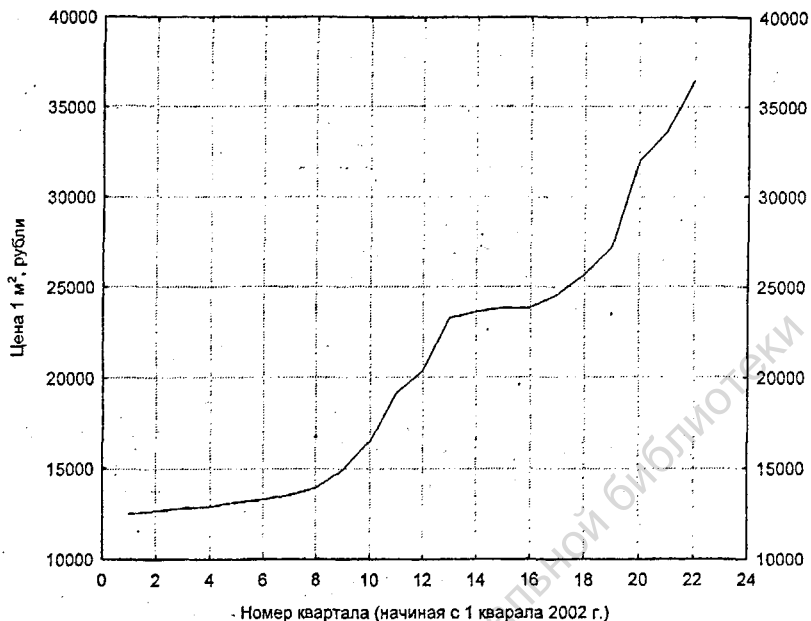


Рисунок 2. График динамики значений цен за 1 м² на вторичном рынке жилья г. Красноярска.

В третьей главе в качестве иллюстрации построенной системы анализа временных рядов рассматривается анализ цен на вторичном рынке жилья г. Красноярска. В качестве данных были взяты значения цен за один квадратный метр за период с 1 квартала 2002 года по 2 квартал 2007 года. Данные были получены в Территориальном органе федеральной службы государственной статистики по Красноярскому краю.

График динамики значений цен за 1 м² на вторичном рынке жилья г. Красноярска приведен на рис. 2.

Используя ретроспективный прогноз, было получено, что наилучшими моделями для прогноза являются АРСС(2,0,1) и метод двухпараметрического экспоненциального сглаживания (метод Хольта), относительные погрешности ретроспективного прогноза которых равны 1,4% и 0,45%, соответственно.

Взвешенный ретроспективный прогноз, позволил повысить точность прогноза. Его ошибка составила 0,1%.

Отметим, что все вычисления были проведены с использованием программ MS Excel, Statistica и MatLab.

В заключении приведены основные результаты и выводы, полученные в ходе выполнения диссертационной работы.

В результате анализа существующих методов и моделей прогнозирования, была создана комплексная система прогнозирования,

составными элементами которой являются как отдельно взятые известные модели (а также их модификации), так и вновь созданные методы прогнозирования.

К научно новым результатам, полученным в работе, мы относим следующее: метод подбора параметров в моделях экспоненциального сглаживания для осуществления более эффективных прогнозов; методику построения взвешенного прогноза, основанную на результатах прогноза отдельно взятых моделей; разработку комплексной системы прогнозирования, основанной на анализе временных рядов.

Работы автора по теме диссертации

1. Ширяева Т.А., Кузнецова Е.М., Шлёпкина А.К., Журов А.В. Применение стохастических моделей для прогнозирования динамики мировых рыночных цен на алюминий и медь. Мат системы, выпуск 3. КрасГАУ, 2005, с.114-120.
2. Ширяева Т.А., Кузнецова, Журов А.В., Студенский А.Г. Прогнозирование в модели скользящего среднего MA(q). Мат системы, выпуск 4. КрасГАУ, 2005, с.115-126.
3. Кузнецов А.А., Журов А.В. Об одном доказательстве основной теоремы теории игр. Актуальные социально-экономические проблемы развития АПК, выпуск 1. КрасГАУ, 2006, с. 129-133.
4. Бородина Е.В., Журов А.В. К вопросу о нахождении константы в методе экспоненциального сглаживания при анализе временных рядов. Материалы XLV Междунар. студ. конф. «Студент и научно-технический прогресс»: Математика. Новосибирский гос. ун-т, 2007, с. 73-74.
5. Кузнецов А.А., Журов А.В. К вопросу о выборе констант в методах экспоненциального сглаживания при анализе временных рядов. Вестник СибГАУ, № 3, СибГАУ, 2007, с. 76-77.
6. А.В. Журов Модификация метода наименьших квадратов при анализе временных рядов. Мат системы, выпуск 6. КрасГАУ, 2007, с.59-60.
7. Кузнецов А.А., Журов А.В. Модель анализа временных рядов с учетом инфляции. Труды междунар. конф. «Решетневские чтения», СибГАУ, 2007, с. 28-29.
8. Кузнецов А.А., Журов А.В. Взвешенный прогноз на основе временных рядов. Вестник СибГАУ, № 4, СибГАУ, 2007, с. 62-63.

Из фондов Российской национальной библиотеки

Санитарно-эпидемиологическое заключение № 24.49.04.953.П. 000381.09.03 от 25.09.2003 г.

Подписано в печать 12.10.2007. Формат 60x84/16. Бумага тип. № 1.

Офсетная печать. Объем 1,0 п.л. Тираж 100 экз. Заказ № 1193

Издательство Красноярского государственного аграрного университета
660017, Красноярск, ул. Ленина, 117

Из фондов Российской национальной библиотеки

2007A
23748

№ 23748

Из фондов Российской национальной библиотеки