Оглавление

1.Введение…… 1

2.Арпсс модели…… 5

1)Модели без сезонной составляющей……12

2)Модели с периодом сезонности 40 …… 21

3)Модели АРПСС с интервенцией…… 26

3.Сезонная декомпозиция …… 32

4.Экспоненциальное сглаживание …… 39

1) Линейный тренд…… 41

2)Экспоненциальный тренд…… 45

5. Применение нейронных сетей…… 50

6. Применение метода многомерного прогнозирования…… 62

7.Заключение…….... 79

8.Приложение…… 83

1)Построение линейных графиков…… 83

2)Корреляции переменных…….. 84

3)Построение графика временного ряда в модуле «Анализ временных рядов»…… 85

4)Построение модели АРПСС с заданными параметрами в модуле «Анализ временных рядов» «АРПСС и автокорреляционные функции»…… 86

1.Введение

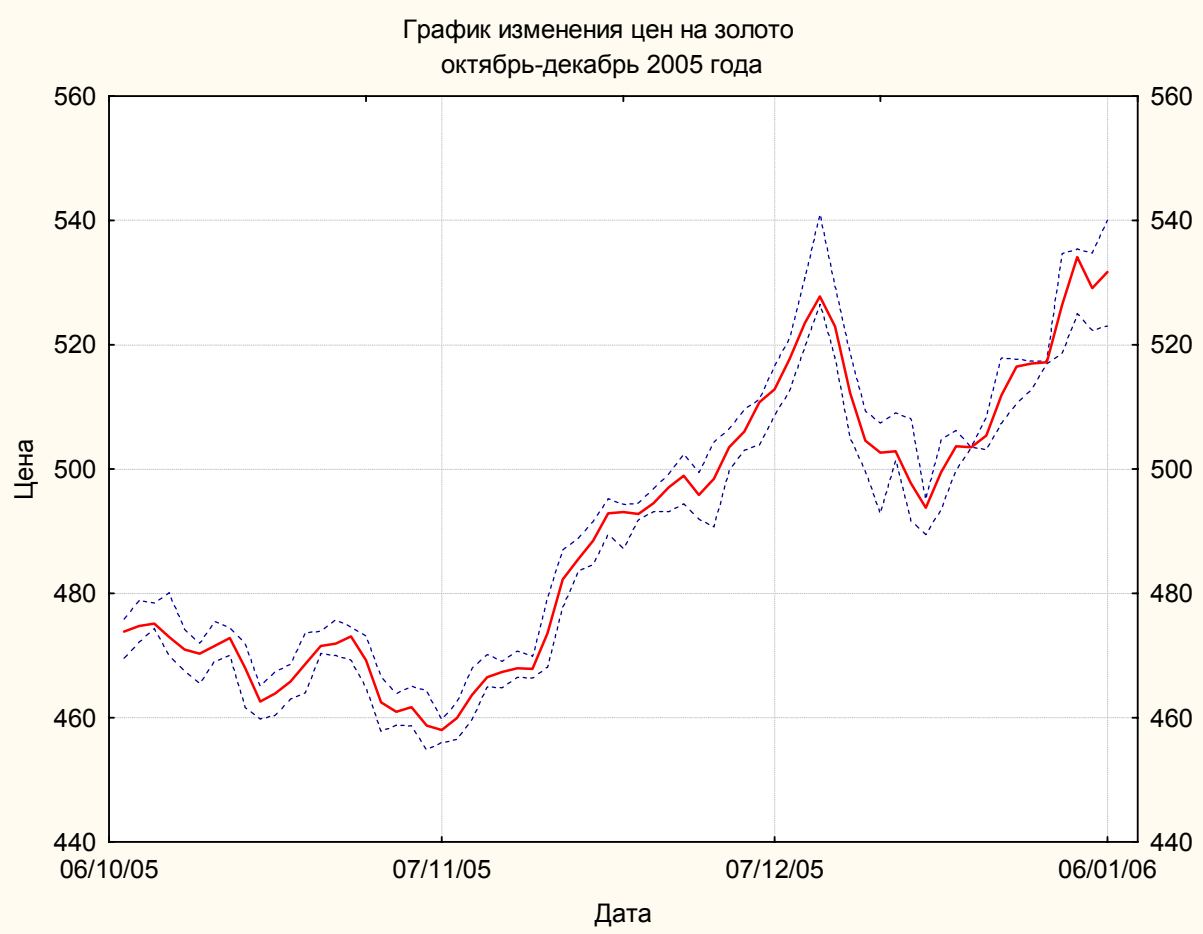


Рис 1.1

График изменения цен на золото в период октябрь-декабрь 2005 года

Красная линия – среднее между ценами закрытия (CLOSE) и открытия(OPEN)

Пунктирные линии – наименьшая и наибольшая цены за день

Мы имеем ежедневную (по рабочим дням) сводку цен на золото в период с 03.01.2005 по 29.12.2006. Задача состоит в том, что на основе представленной статистической информации необходимо сделать прогноз на 15 дней. Для получения более точного прогноза воспользуемся различными методами, а именно: АРПСС, экспоненциальное сглаживание, сезонная декомпозиция, прогнозирование с использованием нейронных сетей, а также построим регрессионную модель, которая также поможет сделать некоторые выводы.

Сначала построим соответствующие ряды и проанализируем их визуально.

1. создадим новую переменную, несущую информацию о днях недели.

2. построим в одной координатной плоскости 4 ряда, отвечающих за различные параметры (см. Приложение).

3. визуально заметно, что в течение дня видимых различий между величинами параметров нет. Это и понятно.

Рис.1.2

График изменения цен на золото (HIGH, LOW, OPEN и CLOSE) за день в период

январь 2005 - декабрь 2006

Можем заметить, что в период с января 2005 по апрель 2006 происходил относительно стабильный и плавный рост цен. Затем произошел сильный взлет, а затем падение (май – середина июля). Приблизительно с июля ситуация вновь стала стабильной, но четко выраженный рост цен практически прекратился. Эта информация пригодится нам в построении модели АРПСС с интервенцией.

4. В данном примере будем строить прогноз для одной переменной (остальные аналогично), но для того, чтобы выбрать ту из четырех переменных, которая наиболее сильно поможет спрогнозировать остальные, построим корреляционную матрицу в анализе «Основные статистики и таблицы».

   Итак, построив матрицу парных корреляций, делаем вывод о том, что переменная LOW наиболее сильно коррелирует с остальными.

   Её прогнозом мы и займемся.

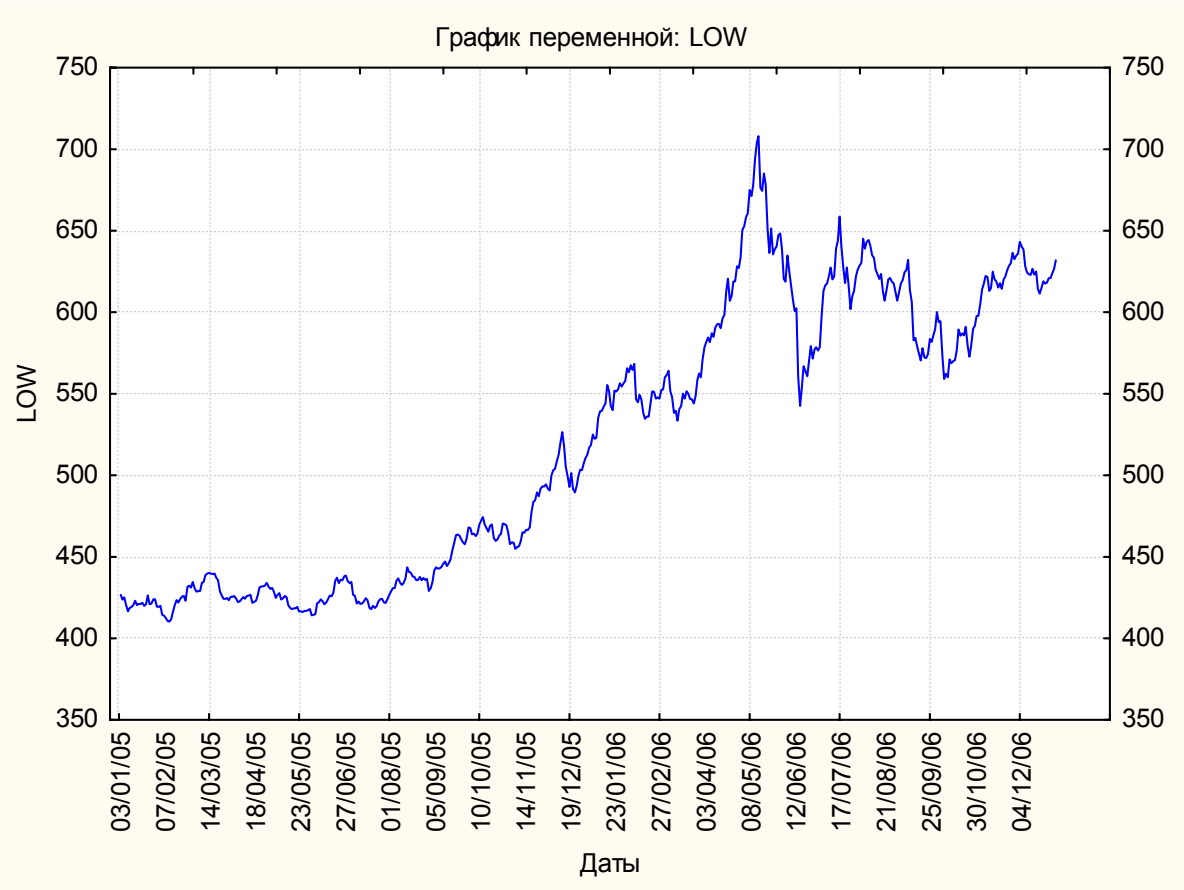


Рис.1.3

График изменения минимальной цены (LOW) за день в период январь 2005 - декабрь 2006

2. АРПСС модели

(Модели Авторегрессии и Проинтегрированного Скользящего среднего)

В меню анализ выберем «Углубленные методы анализа» «Временные ряды и прогнозирование».

В появившемся диалоге можно найти доступ к многочисленным методам анализа. Несмотря на то, что они все являются независимыми, их интеграция позволяет использовать результаты, полученные при работе одного метода как входные данные для работы другого.

Отобразив график ряда LOW, где по оси абсцисс мы будем откладывать номера наблюдений, отметим, что имеет место интервенция типа «скачкообразное, временное» с моментом воздействия №354.

Интервенция началась на наблюдении № 320 и закончилась на № 378 .

Предположим верным, что имеет место вторая и третья интервенции типа «скачкообразное устойчивое воздействие», а именно с номерами воздействий №№ 395 и 445. Время воздействия интервенций: с номера 388 по номер 448.

Итак, мы сделали некоторые предположения об интервенциях.

Рассмотрим начальный отрезок ряда, до момента начала воздействия. Будем рассматривать первоначальный кусок ряда. Итак, нас интересует отрезок ряда с номерами наблюдений 1- 320. Для того чтобы анализу была доступна только часть ряда в главном диалоге «Временные ряды» необходимо нажать кнопку «Выбрать наблюдения» и задать соответствующие условия, задающие требуемые наблюдения.

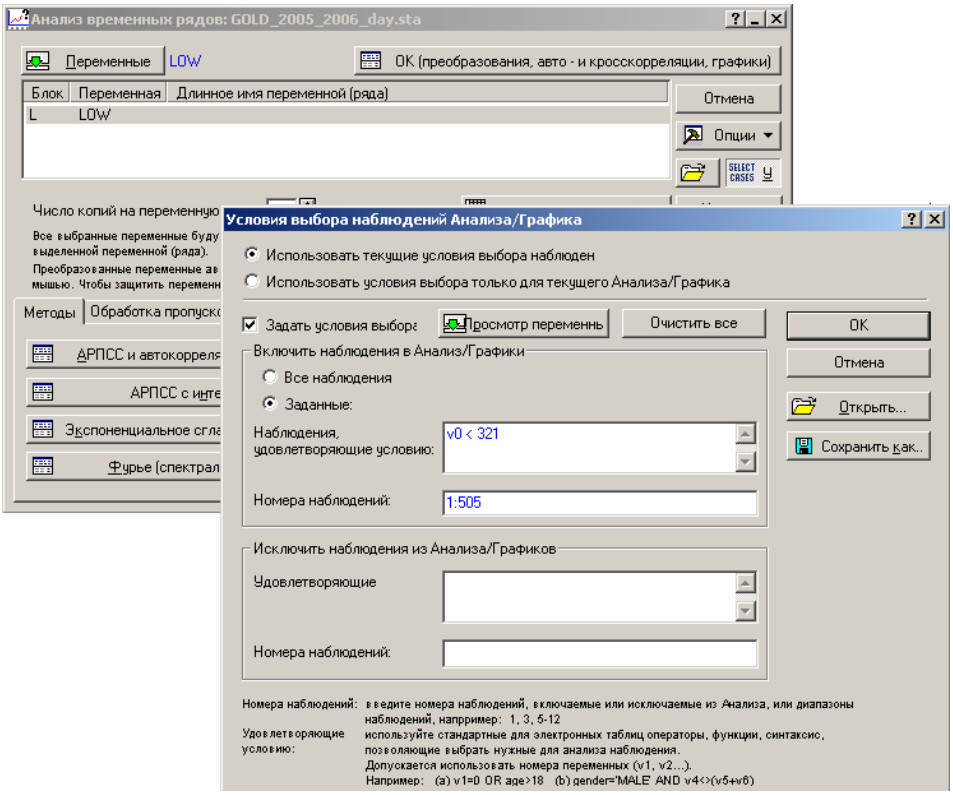


Рис 2.1

Выбор требуемых наблюдений для анализа.

Замечание: *Когда необходимость в рассмотрении именно этого отрезка ряда отпадет необходимо обеспечить и проконтролировать отключение этих условий тем же способом.*

После того, как мы выбрали переменные для анализа, построим график урезанного ряда.

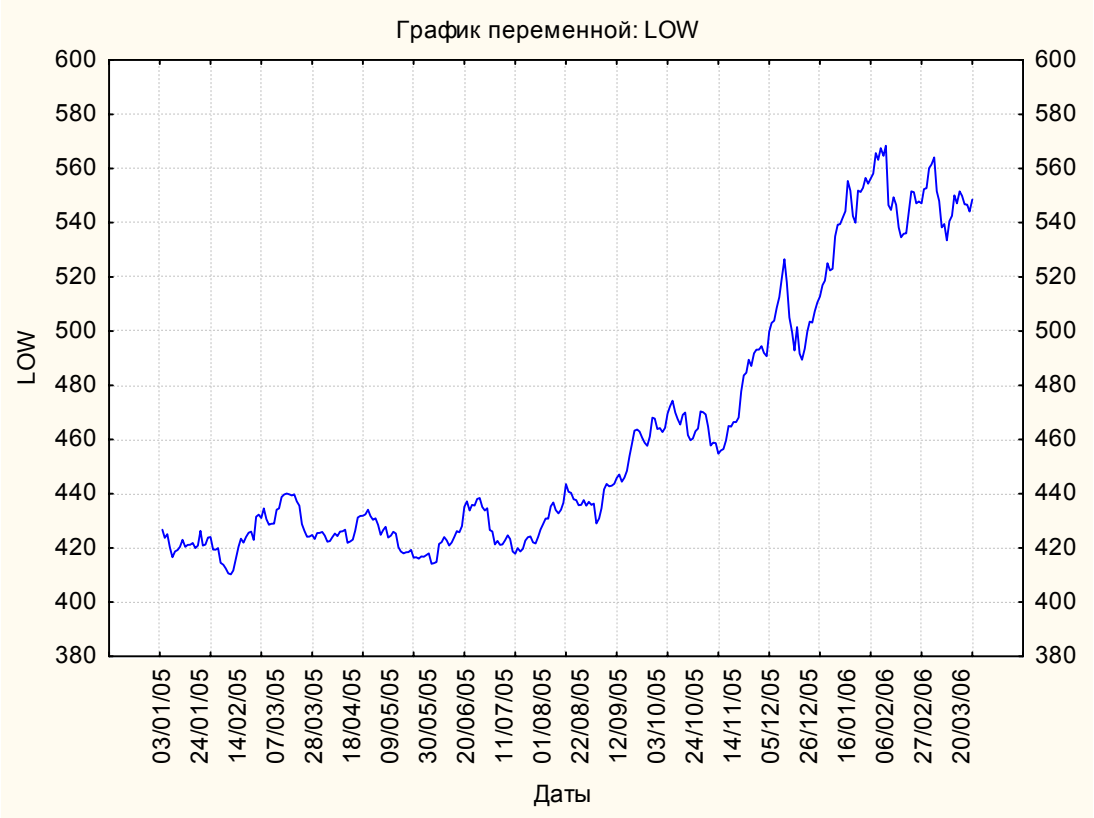


Рис. 2.2

График изменения минимальной цены (LOW) за день в период январь 2005 - апрель 2006(до предполагаемой интервенции)

Теоретически известно, что для любого стационарного временного ряда существует модель авторегрессии и скользящего среднего, которая описывает его поведение.

Основная трудность состоит в том, что подавляющее большинство реальных рядов таковыми не являются. Причинами этому служат либо зависимость дисперсии или математического ожидания от номера наблюдения.

Известно также, что посредством применения некоторых преобразований можно добиться, того, чтобы приведенный ряд был стационарен.

Такими преобразованиями являются:

1. Взятие разностей с различными лагами – добиваемся постоянства математического ожидания.
2. Применение различных арифметических операций (логарифмирование, возведение в степень ...) ко всем членам ряда – добиваемся постоянства дисперсии.

В данном модуле реализованы различные способы преобразования рядов. Для того чтобы начать преобразования, необходимо в любом методе анализе нажать кнопку «Другие преобразования и графики».

Помимо преобразования переменных данный диалог позволяет строить коррелограммы, графики рядов, вычислять значения описательных статистик, а также осуществлять различного рода сглаживания. Сглаживание применяется в тех случаях, когда ряд слишком «игольчатый». Этот факт зачастую затрудняет процесс прогнозирования, обеспечивая большое количество существенных ошибок.

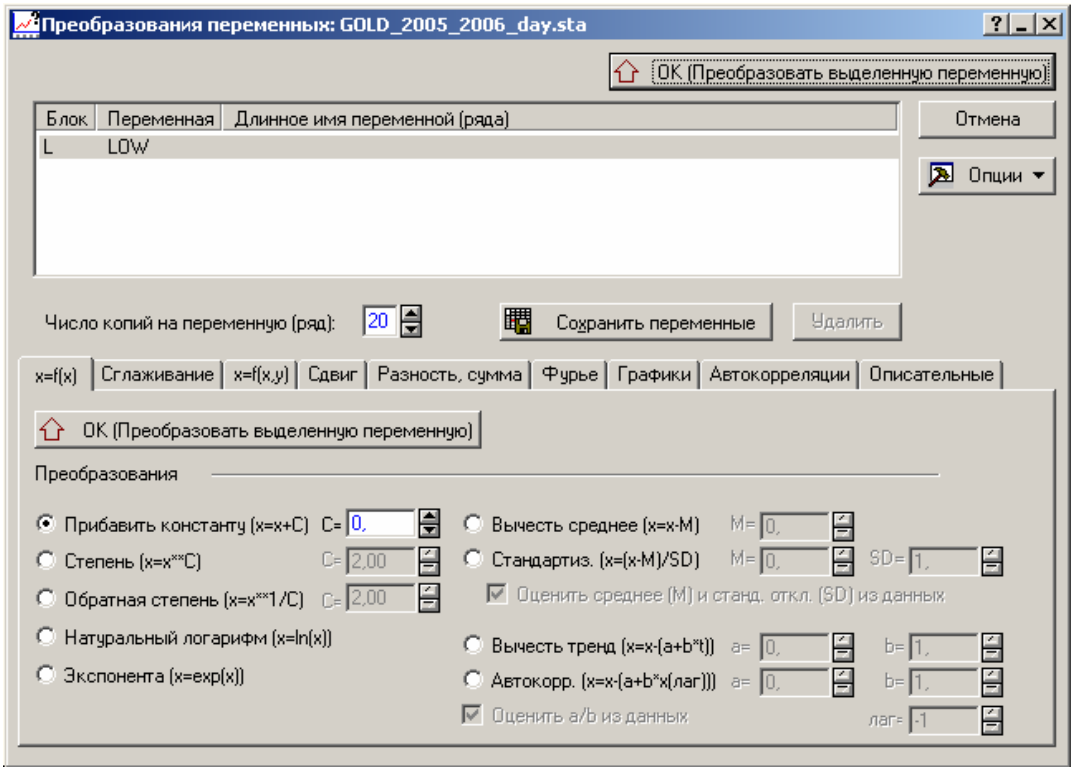


Рис.2.3

Преобразование переменных

Показателем стационарности ряда является его автокорреляционная функция (АКФ) и частная автокорреляционная функция (ЧАКФ).

У стационарного ряда их значения данных функций не равны нулю только лишь на конечном количестве начальных лагов. В статистических исследованиях нас интересует не точное равенство нулю значений автокоррелограмм на больших лагах, а попадание их в определенный доверительный интервал.

Итак, после применения к данному отрезку ряда стандартных процедур получаем:

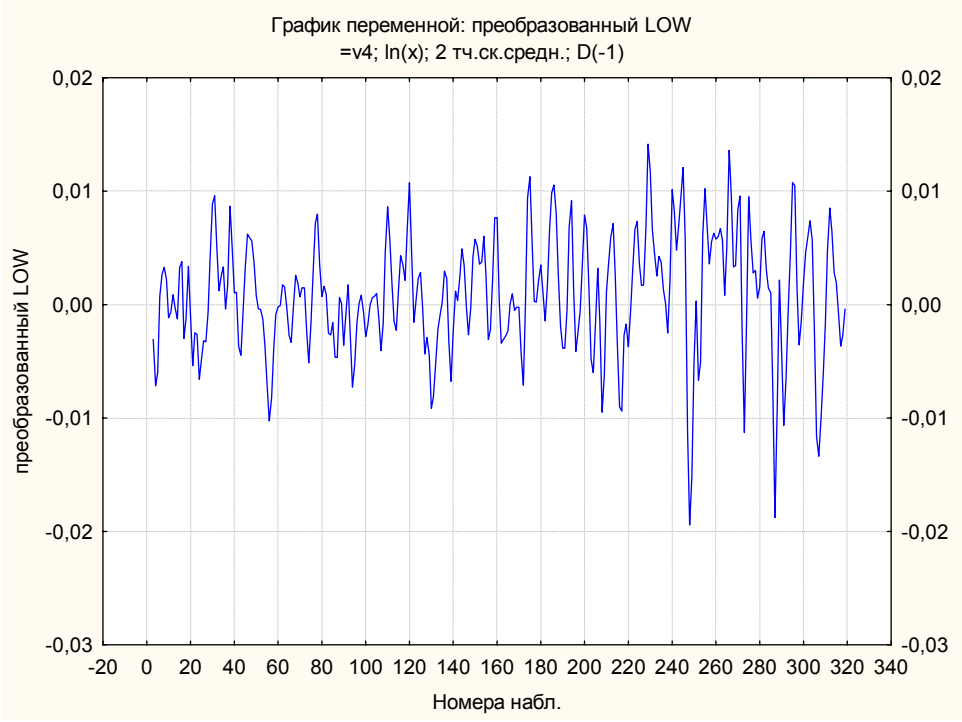


Рис 2.4

График изменения цены LOW после

А) взятия натурального логарифма членов ряда

Б) сглаживания скользящим средним двумя соседними точками

В) взятие одной разности с лагом 1

АКГ и ЧАКГ следующие.

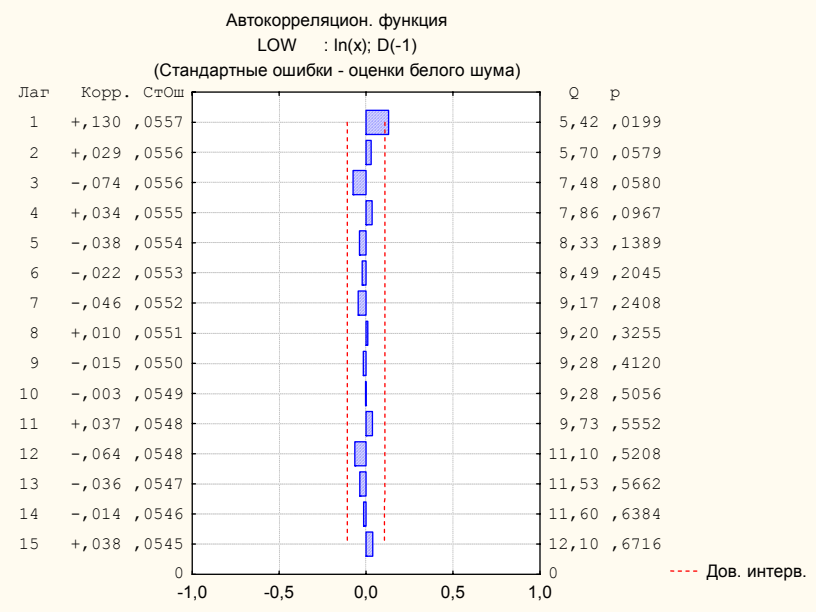


Рис.2.5

Автокоррелограмма (АКГ) преобразованного ряда LOW

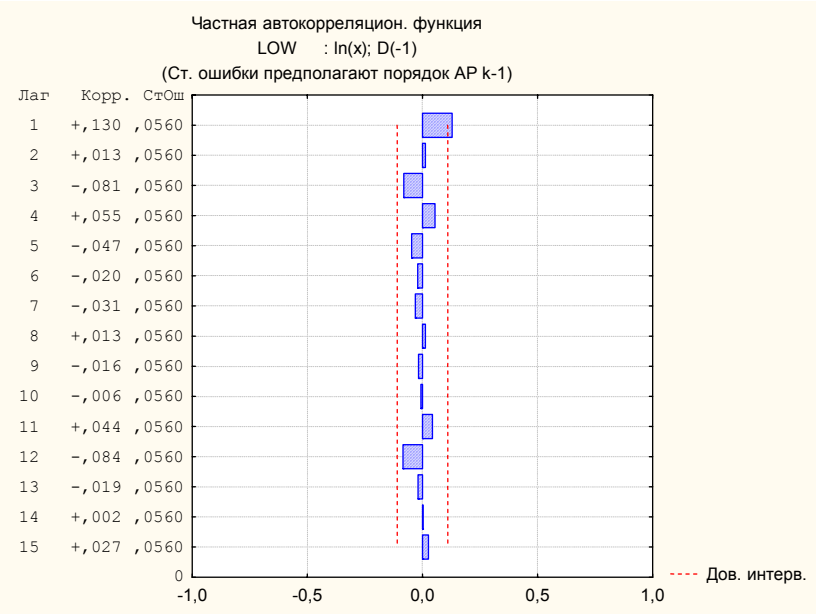


Рис.2.6

Частная автокоррелограмма (ЧАКГ) преобразованного ряда LOW

Визуально проанализировав коррелограммы, сделаем следующие предположения о виде модели. Поскольку значения обеих коррелограмм очень небольшие в совокупности, то можно предположить, что справедлива одна из следующих ситуаций:

1)есть выделяющееся значение АКГ только на единичном лаге и убывание ЧАКГ;

2)есть выделяющееся значение ЧАКГ только на единичном лаге и убывание АКГ;

3)есть одновременно выделяющееся значение АКГ и ЧАКГ только на единичном лаге.

Согласно методологии Бокса-Дженкинса необходимо рассмотреть модели

1) АРПСС(0,1,1),

2)АРПСС(1,1,0),

3)АРПСС(1,1,1)

Для того чтобы определить период сезонности, применим процедуру анализа Фурье. Для этого в диалоге «Анализ временных рядов» выберем вкладку «Фурье (спектральный) анализ».

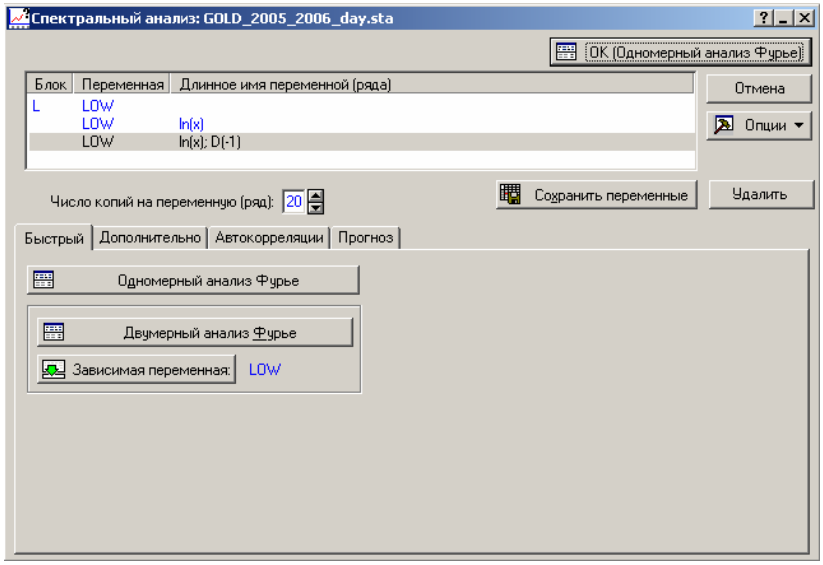


Рис. 2.7

Диалог «Спектральный анализ»

Затем остается лишь нажать кнопку «Одномерный анализ Фурье» и прейти к диалогу «Результаты спектрального анализа»

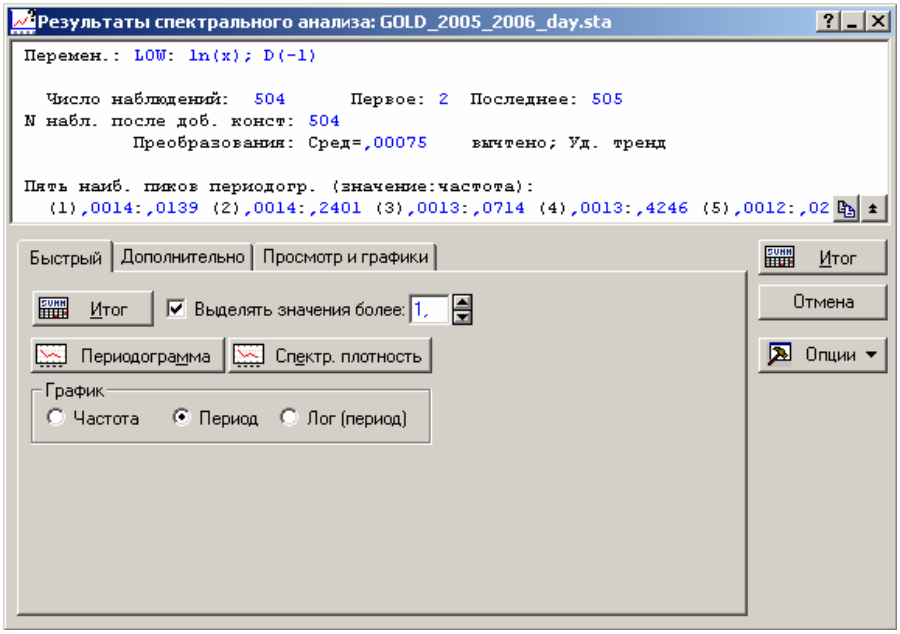


Рис 2.8

Диалог «Результаты спектрального анализа»

В данном диалоге мы можем построить периодограмму, которая и покажет наиболее вероятные значения частоты (или периода при выборе надлежащего флажка) сезонной компоненты. Также модно построить таблицу зависимости значений периодограммы от частоты (либо периода). Значение периодограммы представляет собой сумму квадратов коэффициентов при синусе и косинусе соответствующего аргумента в разложении данного ряда в ряд Фурье.

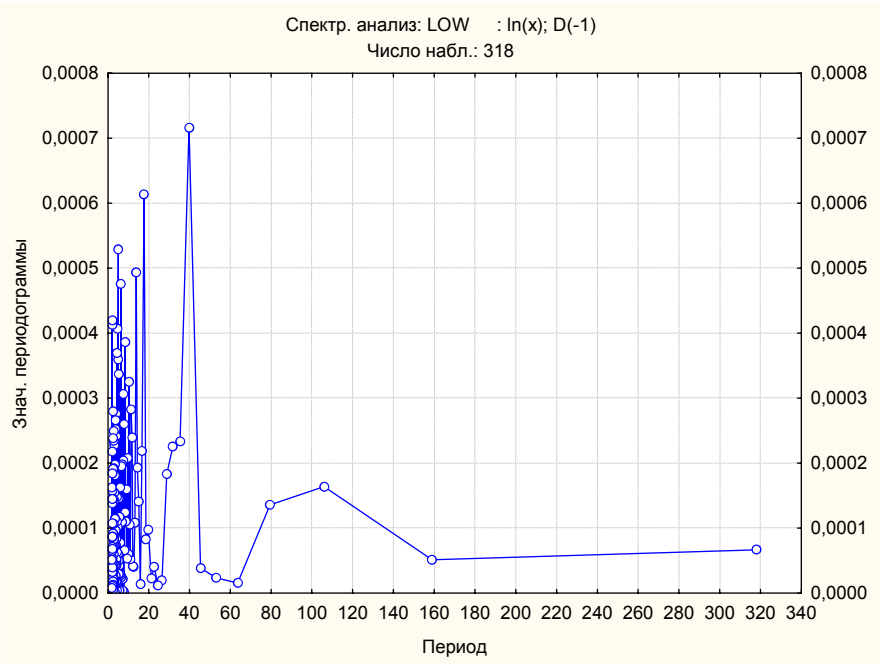


Рис. 2.9

Периодограмма разложения преобразованного ряда по тригонометрическому базису Фурье

Наибольший всплеск (не считая всплеска на периоде равном длине ряда) на значении 40.

Если мы будем строить периодограммы для немодифицированного ряда, то мы опять же заметим наибольший всплеск на значении 40.

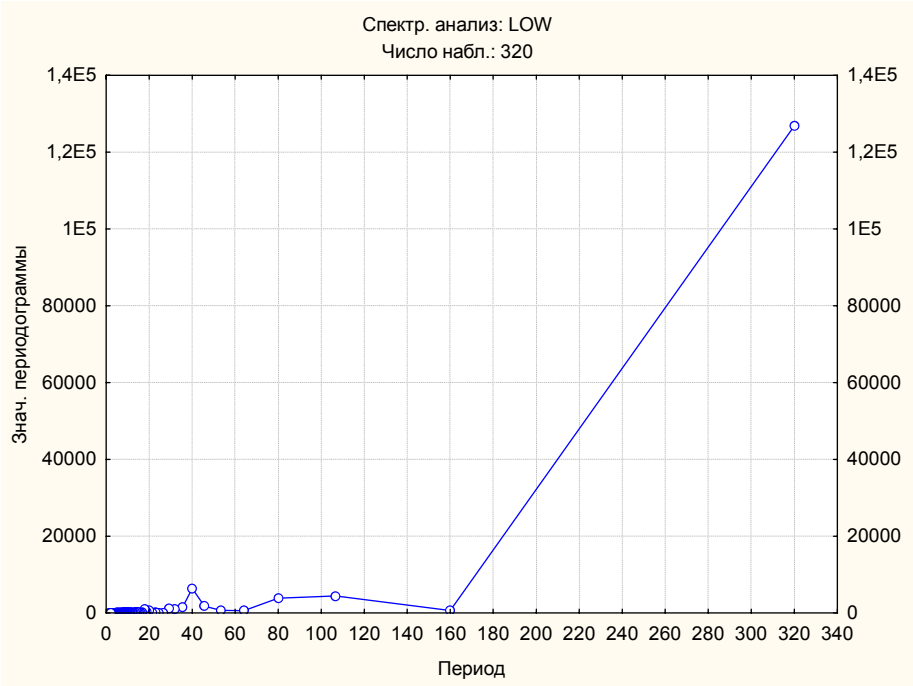


Рис. 2.11

Периодограмма разложения исходного ряда по тригонометрическому базису Фурье

Поскольку это значение весьма необычно и трудно объяснимо с точки зрения житейской логики, то мы проверим еще и случай, когда сезонного лага нет, то есть период сезонности равен нулю (тем самым мы проверяем гипотезу о том, что сезонная составляющая очень мала, или другими словами её значение статистически незначимо)

Случай 1.Сезонный лаг = 0

Для того, чтобы облегчить процедуру поиска оптимальной модели АРПСС был написан макрос автоматического переборного поиска этих параметров.

После запуска отобразиться диалог «АвтоАРПСС».

Указав значения параметров работы параметров (Сезонный лаг установили равным нулю) необходимо нажать «ОК» для начала работы алгоритма программы.

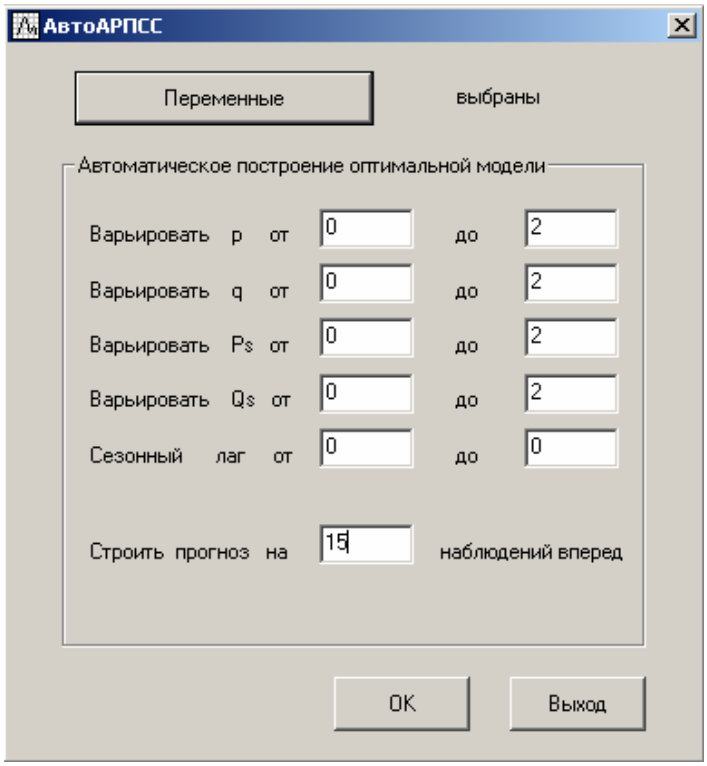


Рис. 2.11

Диалог «АвтоАРПСС»

Результатом работы макроса является таблица, в которой представлены параметры моделей и различные типы ошибок, соответствующие этим моделям.

	Original	Var1	Var2	Var3	Var4	Var5	Var6
p		1	1	0	0	1	1
q		0	0	1	1	1	1
Ps		0	0	0	0	0	0
Qs		0	0	0	0	0	0
SL		0	0	0	0	0	0
AbsMeanDev		4,430015	4,43001589	4,45141709	4,45141709	4,4529463	4,4529463
RelMeanDev		0,007155	0,007155777	0,007190921	0,007190921	0,007193440	0,007193440
bLag		1	1	1	1	1	1
Lag		1	1	1	1	1	1
LagS		0	1	0	1	0	1

Рис. 2.12

Результаты работы макроса «АвтоАРПСС»

Опишем, что содержится в строках таблицы:

- *p* – количество членов авторегрессии;
- *q* – количество членов скользящего среднего;
- *Ps* – количество сезонных членов авторегрессии;
- *Qs* – количество сезонных членов скользящего среднего;
- *SL* – величина сезонного лага;
- *AbsMeanDev* – средняя абсолютная ошибка;
- *RelMeanDev* – средняя относительная ошибка;
- *bLag* – флаговый параметр, указывающий на то, брались разности (1) или нет (0);
- *Lag* - флаговый параметр, указывающий на то, брались разности с лагом 1 (1) или нет (0);
- *LagS* - флаговый параметр, указывающий на то, брались разности с лагом *SL* (1) или нет (0);

Программа автоматической подгонки АРПСС моделей советует остановиться на моделях АРПСС(0,1,1), АРПСС(1,1,0), АРПСС(1,1,1). Во всех моделях относительное абсолютное отклонение порядка 0,07. Строим прогнозы.

Итак, займемся построением прогнозов для выбранных нами моделей.

Модель АРПСС(1,1,0)

Оценки параметров

Параметр является статистически значимым, но доверительный интервал слишком велик.

Исход.:LOW (GOLD_2005_2006_day.sta) Преобразования: ln(x),D(1) Модель(1,1,0) MS Остаток= ,00006						
	Парам.	**Асимпт.**	**Асимпт.**	**p**	**Нижняя**	**Верхняя**
p(1)	0,138651	0,055693	2,489565	0,013301	0,029078	0,248225

Также проверим остатки на предмет соответствия стандартным предположениям (независимость, нормальный тип распределения). Для того чтобы проверить гипотезу о нормальной распределенности остатков, построим гистограмму частот и вероятностный график.

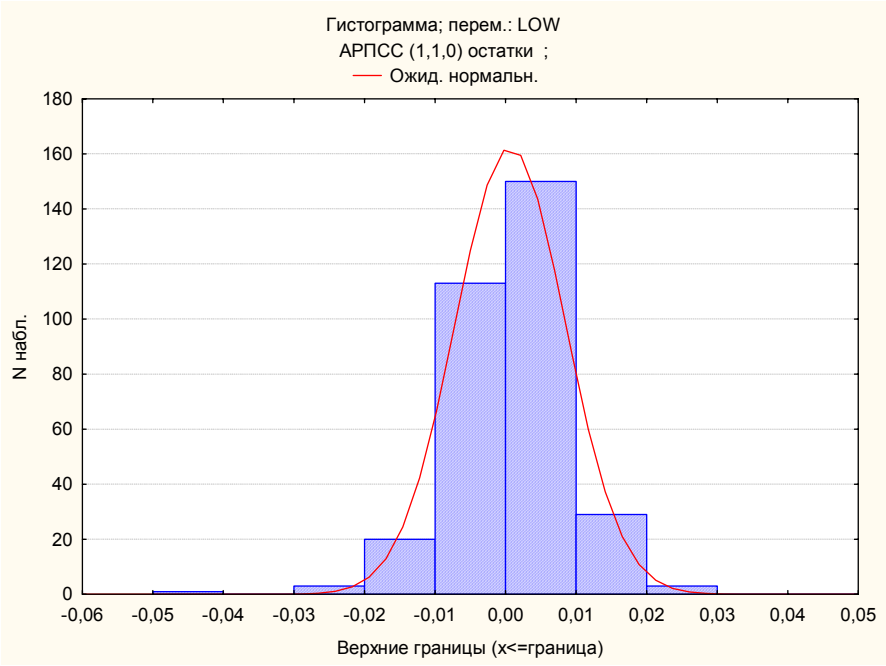


Рис. 2.13

Гистограмма распределения частот остатков в модели АРПСС(1,1,0)

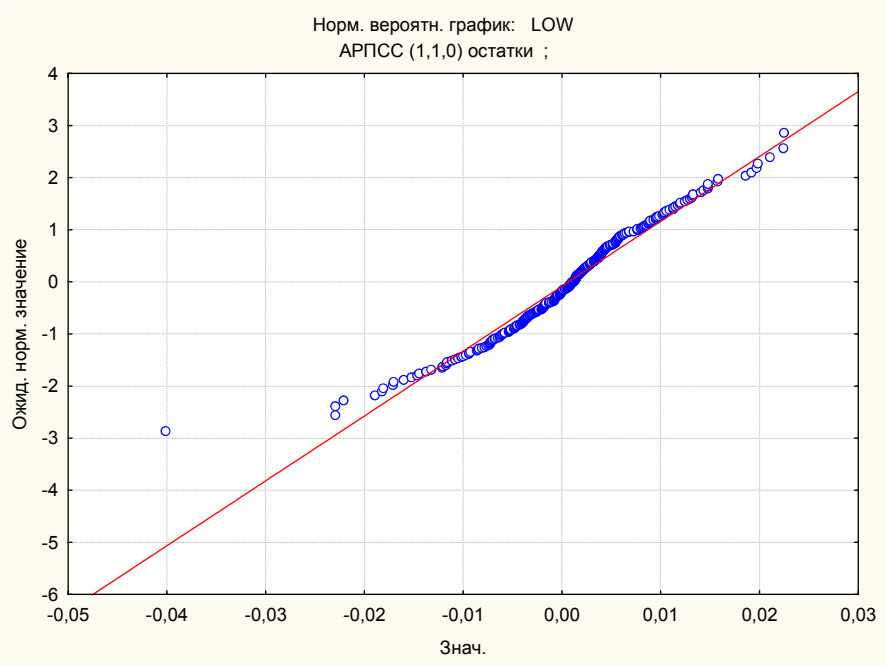


Рис 2.14

Нормальный вероятностный график остатков модели АРПСС(1,1,0)

Для проверки независимости построим автокоррелограммы. В случае справедливости предположения о независимости остатков, значения обеих коррелограммы не будут статистически значимы для любого лага.

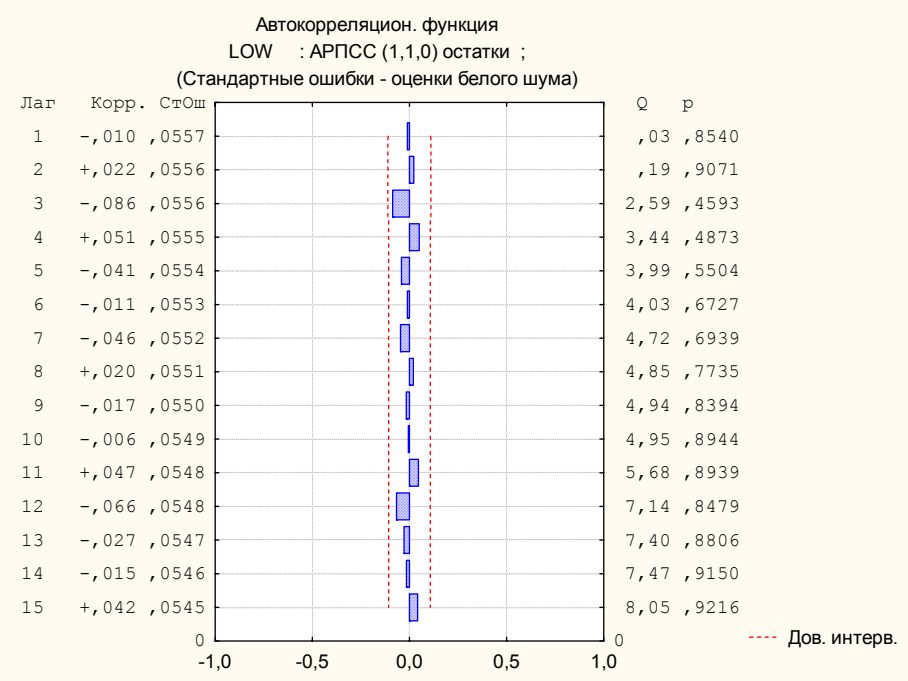


Рис 2.15

Автокоррелограмма остатков модели АРПСС(1,1,0)

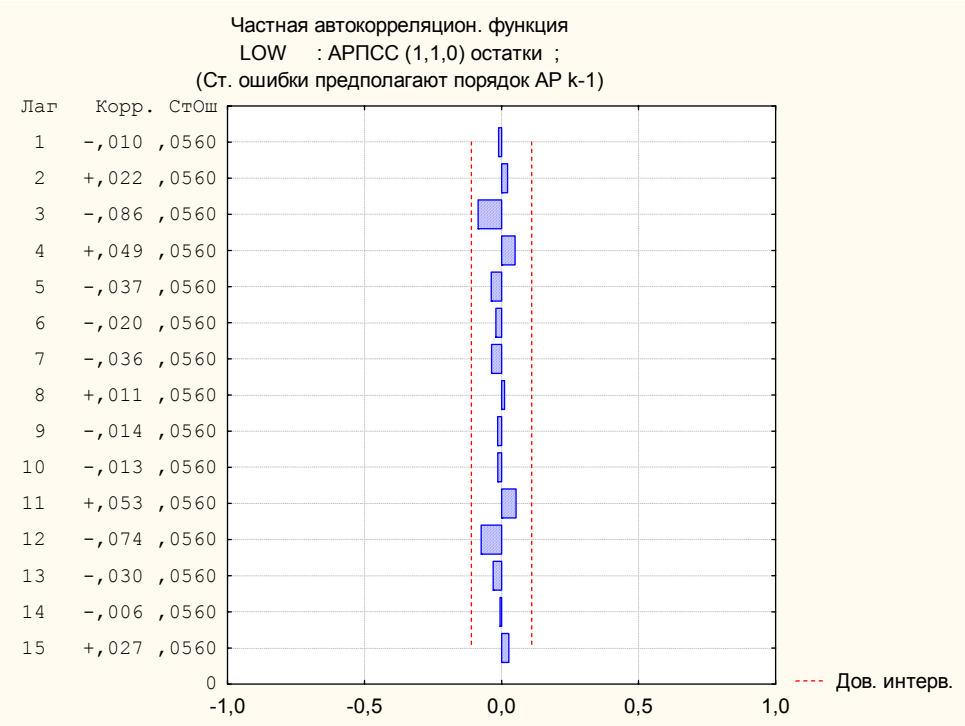


Рис 2.16

Частная автокоррелограмма остатков модели АРПСС(1,1,0)

Итак, с большой степенью точности можно сказать, что остатки действительно удовлетворяют стандартным предположениям. Наконец, построим сам прогноз на 15 наблюдений (3 рабочие недели) вперед для данной модели.

Рис. 2.17

Прогноз модифицированного ряда при помощи модели АРПСС(1,1,0)

	Прогноз	Нижний	Верхний
24 ямарта 2006 г.	543,7592	536,7462	550,8639
27 марта 2006 г.	543,7592	533,2143	554,5127
28 марта 2006 г.	543,7592	530,6181	557,2258
29 марта 2006 г.	543,7592	528,4695	559,4914
30 марта 2006 г.	543,7592	526,5977	561,4801
31 марта 2006 г.	543,7592	524,9190	563,2757
3 апреля 2006 г.	543,7592	523,3848	564,9268
4 апреля 2006 г.	543,7592	521,9642	566,4644
5 апреля 2006 г.	543,7592	520,6357	567,9098
6 апреля 2006 г.	543,7592	519,3840	569,2784
7 апреля 2006 г.	543,7592	518,1977	570,5817
10 апреля 2006 г.	543,7592	517,0677	571,8286
11 апреля 2006 г.	543,7592	515,9870	573,0263
12 апреля 2006 г.	543,7592	514,9500	574,18,03
13 апреля 2006 г.	543,7592	513,9519	575,2953

Рис. 2.17

Прогноз модифицированного ряда при помощи модели АРПСС(1,1,0)

Модель АРПСС(1,1,1)

Оценим параметры.

Исход.:LOW (GOLD_2005_2006_day.sta) Преобразования: ln(x),D(1) Модель(1,1,1) MS Остаток= ,00006

	Парам.	Асимпт.	Асимпт.	p	Нижняя	Верхняя
p(1)	0,211409	0,291445	0,725383	0,468752	-0,362002	0,784821
q(1)	0,073778	0,294989	0,250105	0,802669	-0,506606	0,654163

Оцененные параметры не являются статистически значимыми, на что указывает чрезмерно большое p – значение.

Модель АРПСС(0,1,1)

Оценки параметров

Исход.:LOW (GOLD_2005_2006_day.sta) Преобразования: ln(x),D(1) Модель(0,1,1) MS Остаток= ,00006						
	Парам.	**Асимпт.**	**Асимпт.**	**p**	**Нижняя**	**Верхняя**
q(1)	-0,127628	0,052344	-2,43824	0,015306	-0,230612	-0,024643

Параметр является статистически значимым, а поэтому проделаем анализ остатков, как для предыдущей модели.

Проверка остатков на нормальность

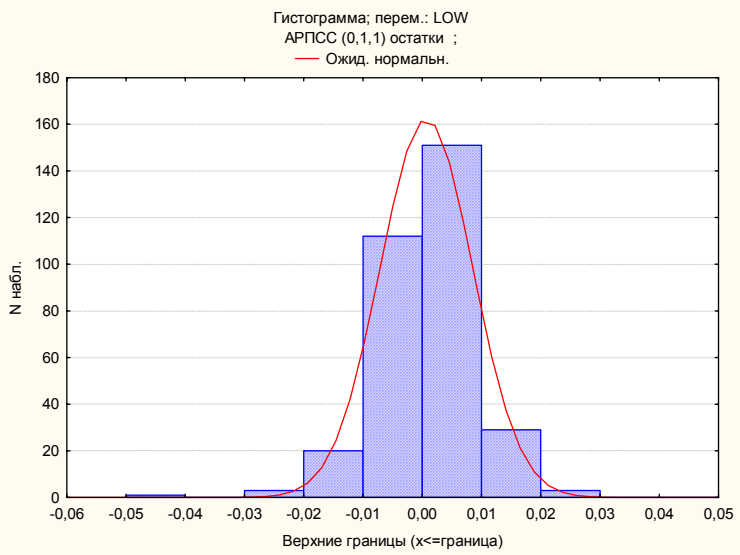


Рис. 2.18

Гистограмма распределения частот остатков в модели АРПСС(0,1,1)

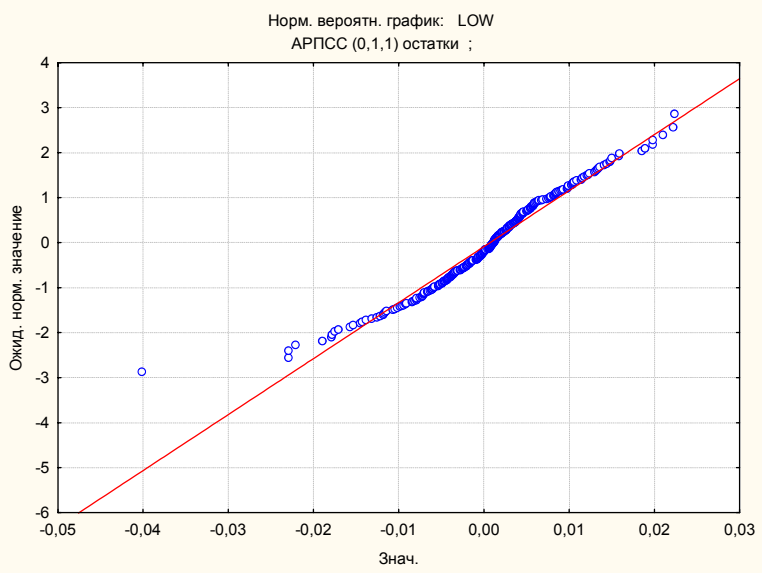


Рис 2.19

Нормальный вероятностный график остатков модели АРПСС(0,1,1)

Проверка остатков на независимость

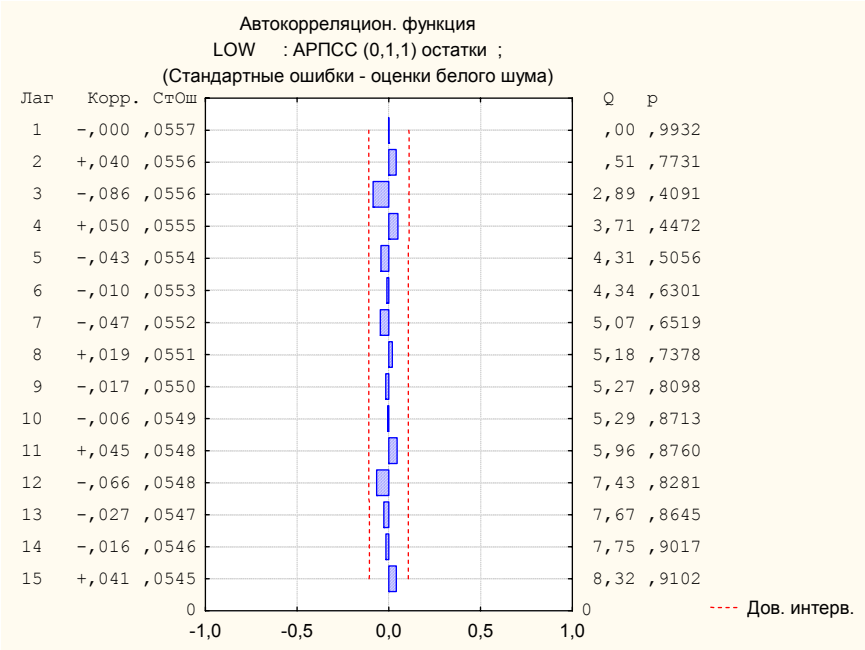


Рис 2.20

Автокоррелограмма остатков модели АРПСС(1,1,0)

Рис 2.21

Частная автокоррелограмма остатков модели АРПСС(0,1,1)

Как и в случае модели АРПСС(1,1,0) мы получили, что остатки с большой степенью точности удовлетворяют стандартным предположениям.

Построим прогноз.

	Прогноз	Нижний	Верхний
24 ямарта 2006 г.	549,1411	542,0562	556,3186
27 марта 2006 г.	549,1411	538,4981	559,9945
28 марта 2006 г.	549,1411	535,8812	562,7291
29 марта 2006 г.	549,1411	533,7151	565,0130
30 марта 2006 г.	549,1411	531,8278	567,0180
31 марта 2006 г.	549,1411	530,1351	568,8285
3 апреля 2006 г.	549,1411	528,5881	570,4932
4 апреля 2006 г.	549,1411	527,1556	572,0436
5 апреля 2006 г.	549,1411	525,8159	573,5010
6 апреля 2006 г.	549,1411	524,5537	574,8811
7 апреля 2006 г.	549,1411	523,3572	576,1953
10 апреля 2006 г.	549,1411	522,2176	577,4527
11 апреля 2006 г.	549,1411	521,1278	578,6603
12 апреля 2006 г.	549,1411	520,0819	579,8240
13 апреля 2006 г.	549,1411	519,0753	580,9484

Рис. 2.22

Прогноз модифицированного ряда при помощи модели АРПСС(1,1,0)

Рассмотренные модели не дают хороших прогнозов, потому как доверительный интервал слишком большой, а вид прогноза «подозрительно» простой. Во многом это объясняется тем, что мы принимали гипотезу об отсутствии сезонной компоненты.

Построенные вначале коррелограммы модифицированного ряда хоть и имели значения, большие доверительного уровня, но они были довольно малы и, выбрав уровень доверия чуть больше, можно было бы считать ряд случайным. Оценки значений параметров оказались весьма близкими к нулю (порядка 0.13).

Эти два аспекта говорят о том, что корреляция значений модифицированного ряда с предыдущими значениями была чрезвычайна, мала, что и даёт возможность принять гипотезу о случайности этих значений. Таким образом, остатки моделей слабо отличаются от самого ряда. При анализе остатков мы отметили, что их распределение очень близко к нормальному с нулевым математическим ожиданием. А поскольку нормальное распределение является симметричным, то отрицательные и положительные значения являются равновероятными. Отсюда и получаем, что прогноз является прямой горизонтальной линией.

Теперь рассмотрим

Случай 2. Сезонный лаг 40.

Снова рассмотрим обрезанный до интервенции ряд. Применим к нему процедуру автоматического поиска АРПСС модели. Результатом работы программы стали следующие результаты: (2,1,1)(0,0,1) и (1,1,2)(0,0,1)

Модель АРПСС(2,1,1)(0,0,1)

Применим модель к ряду. Найдем оценки параметров. Значения p-уровня в четвертом столбце таблицы (а именно то, что эти значения менее 0.05) сигнализируют о статистической значимости оцениваемых параметров.

Исход.:LOW before intervention: = v4 (GOLD_2005_2006_day.sta) Преобразования: ln(x),D(1) Модель(2,1,1)(0,0,1) Сезонный лаг: 40 MS Остаток= ,00006

	Парам.	Асимпт.	Асимпт.	p	Нижняя	Верхняя
p(1)	0,966880	0,362545	2,66692	0,008050	0,253564	1,680195
p(2)	-0,135909		-2,18229	0,029827	-0,258442	-0,013375
q(1)	0,828572	0,362065	2,28846	0,022773	0,116199	1,540944
Qs(1)	-0,206648	0,063811	-3,23841	0,001330	-0,332198	-0,081097

Проведем стандартную процедуру анализа остатков на предмет соответствия стандартным предположениям. Начнем с предположения о нормальности их распределения. Построим гистограмму частот и нормальный вероятностный график.

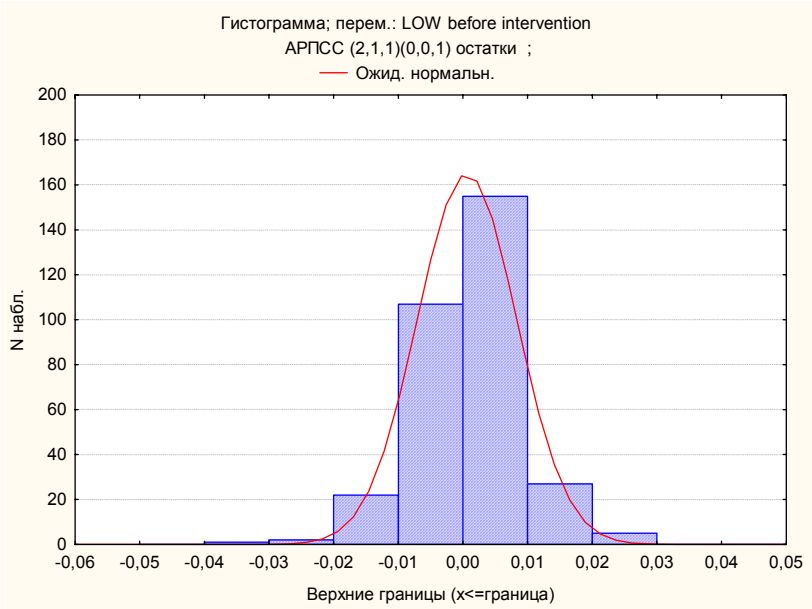


Рис. 2.23

Гистограмма распределения частот остатков в модели АРПСС(2,1,1)(0,0,1)

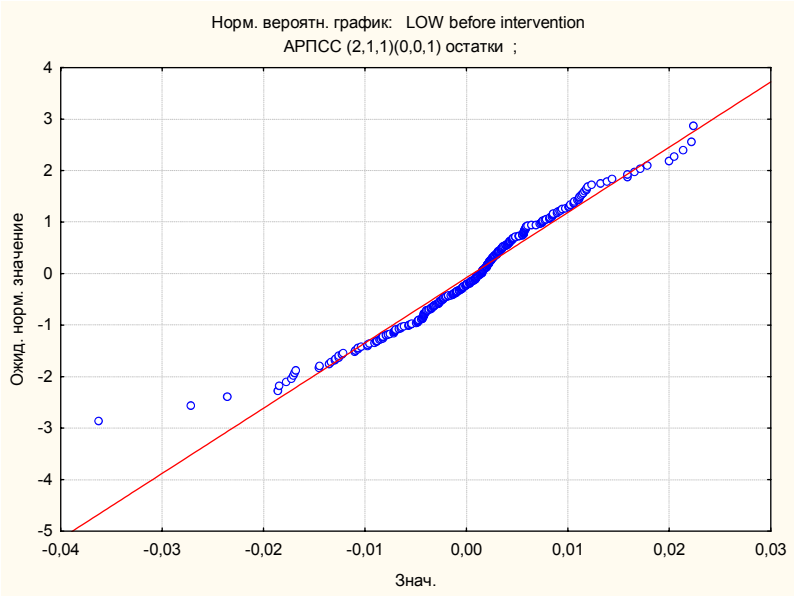


Рис 2.24

Нормальный вероятностный график остатков модели АРПСС(2,1,1)(0,0,1)

Исходя из построенных графиков можно сказать, что распределение остатков действительно является нормальным. На это указывает внешнее сходство гистограммы и графика плотности нормального распределения с соответствующими коэффициентами, а также тот факт, что нормальный вероятностный график лежит практически на биссектрисе координатного угла (за исключением нескольких выбросов).

Как и выше независимость остатков проверим, построив коррелограммы.

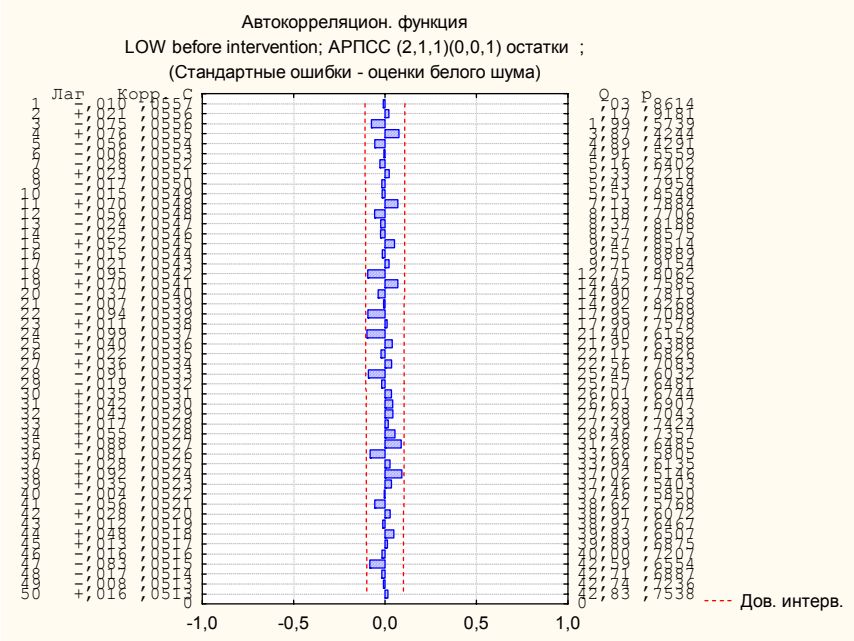


Рис 2.25

Автокоррелограмма остатков модели АРПСС(2,1,1)(0,0,1)

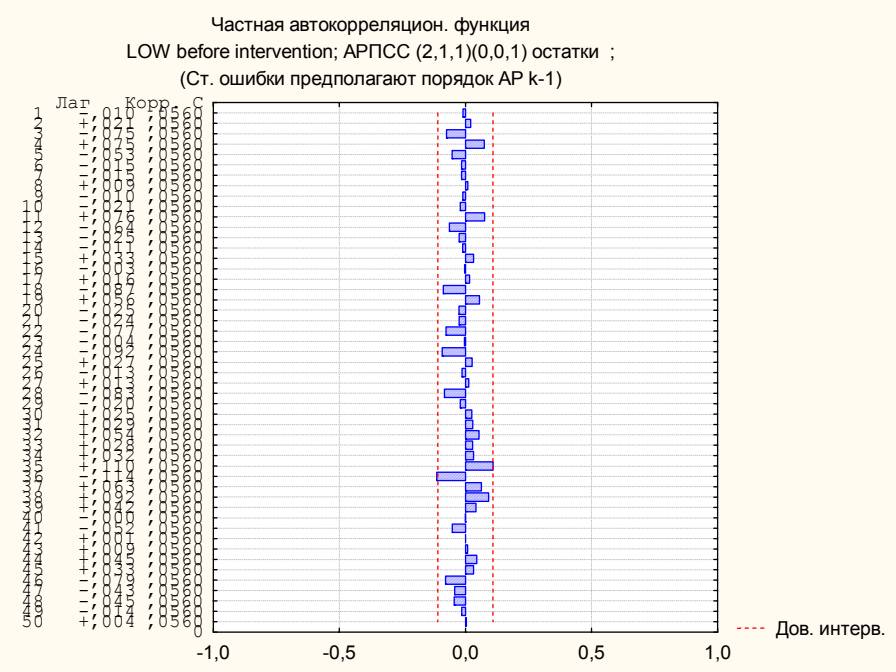


Рис 2.26

Частная автокоррелограмма остатков модели АРПСС(2,1,1)(0,0,1)

Построенные коррелограммы также указывают на то, что остатки удовлетворяют стандартным предположением (н.о.р.)

Модель АРПСС (1,1,2)(0,0,1)

Параметры этой модели также являются статистическими значимыми

Исход.:LOW before intervention: = v4 (GOLD_2005_2006_day.sta) Преобразования: ln(x),D(1) Модель(1,1,2)(0,0,1) Сезонный лаг: 40 MS Остаток= ,00006						
	Парам.	**Асимпт.**	**Асимпт.**	**p**	**Нижняя**	**Верхняя**
p(1)	-0,747197	0,151897	-4,91911	0,000001	-1,04606	-0,448337
q(1)	-0,904583	0,157131	-5,75688	0,000000	-1,21374	-0,595425
q(2)	-0,184220	0,060807	-3,02956	0,002652	-0,30386	-0,064580
Qs(1)	-0,200800	0,062842	-3,19530	0,001538	-0,32444	-0,077156

Проведем анализ остатков

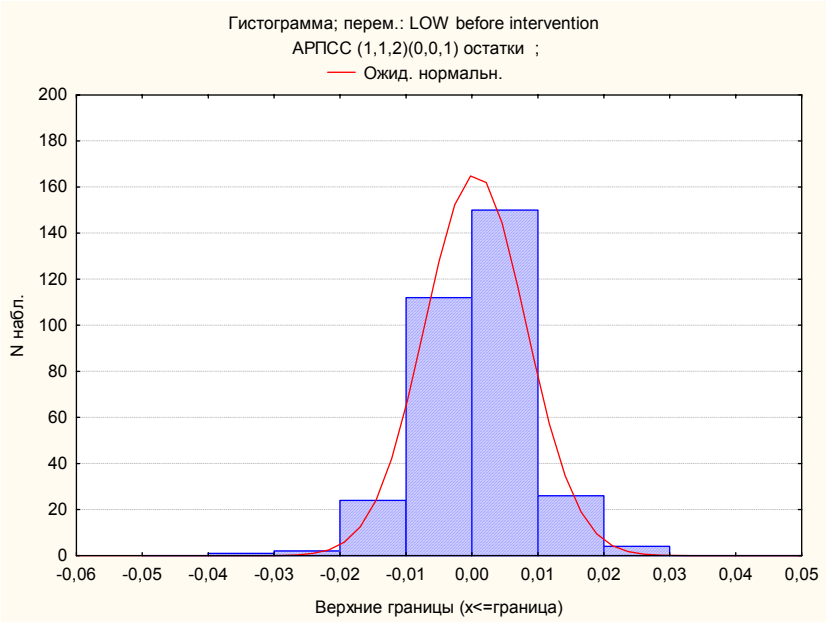


Рис. 2.27

Гистограмма распределения частот остатков в модели АРПСС(1,1,2)(0,0,1)

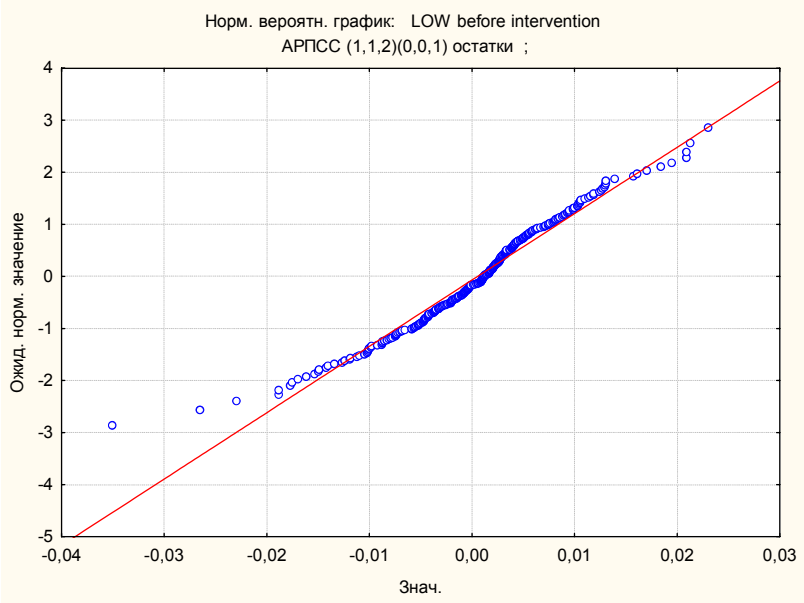


Рис 2.28

Нормальный вероятностный график остатков модели АРПСС(1,1,2)(0,0,1)

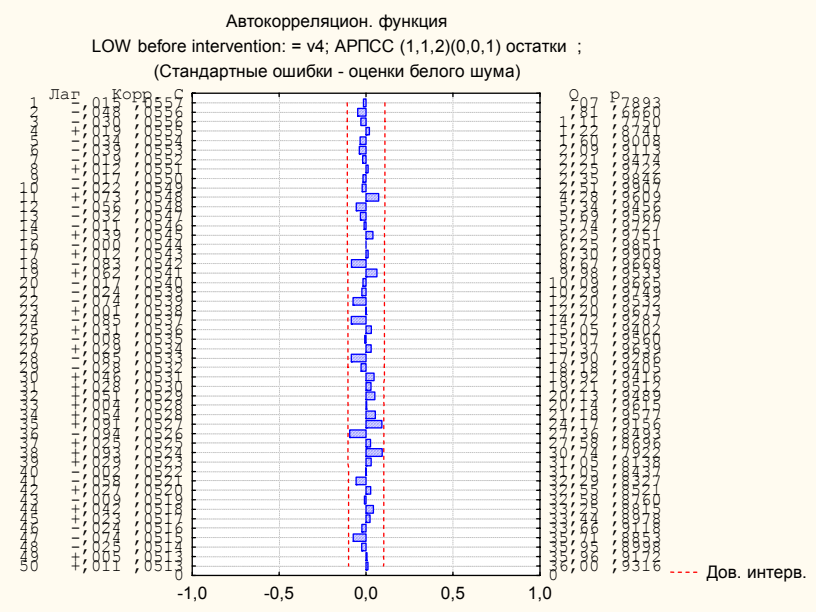


Рис 2.29

Автокоррелограмма остатков модели АРПСС(1,1,2)(0,0,1)

Опять же принимаем стандартные предположения об остатках.

АРПСС с интервенцией

Выше были сделаны некоторые предположения об интервенциях. Построенные модели для отрезка ряда попробуем применить к исходному ряду, дополняя их информацией об интервенциях. Более того, рассмотрим два случая, а именно случай с одной интервенцией и случай с тремя интервенциями.

Замечание: мы не будем рассматривать модели без сезонного лага, потому как даже на «хорошем» отрезке ряда они проявили себя гораздо хуже.

Модель АРПСС(2,1,1)(0,0,1)

Построив таблицу оценок параметров модели, заметим, что значимыми является лишь параметр Дельта(1), относящийся к первой интервенции. Именно этот факт и побуждает рассмотреть случай только с одной интервенцией.

Исход.:LOW (GOLD_2005_2006_day.sta) Преобразования: ln(x),D(1) (Прерванная АРПСС) Модель(2,1,1)(0,0,1) Сезонный лаг: 40 MS Остаток= ,00011

	Парам.	Асимпт.	Асимпт.	p	Нижняя	Верхняя	Интерв.	Интерв.
p(1)	0,229004	0,341953	0,669695	0,503355	-0,442802	0,900811		
p(2)	-0,062694	0,067761	-0,925220	0,355288	-0,195817	0,070430		
q(1)	0,068810	0,340334	0,202183	0,839854	-0,599816	0,737435		
Qs(1)	-0,037558	0,040655	-0,923817	0,356017	-0,117430	0,042314		
Омега(1)	0,013982	0,010349	1,351115	0,177256	-0,006349	0,034314	354	ск/врем
Дельта(1)	0,822909	0,279039	2,949085	0,003333	0,274705	1,371113	354	ск/врем
Омега(2)	0,009939	0,010603	0,937357	0,349018	-0,010892	0,030769	395	ск/уст
Омега(3)	-0,008904	0,010598	-0,840155	0,401215	-0,029725	0,011917	445	ск/уст

Оценим параметры для второго случая. Опять же статистически значимым является лишь параметр Дельта(1).

Исход.:LOW (GOLD_2005_2006_day.sta) Преобразования: ln(x),D(1) (Прерванная АРПСС) Модель(2,1,1)(0,0,1) Сезонный лаг: 40 MS Остаток= ,00011

	Парам.	Асимпт.	Асимпт.	p	Нижняя	Верхняя	Интерв.	Интерв.
p(1)	0,212194	0,345036	0,614990	0,538834	-0,465663	0,890051		
p(2)	-0,056139	0,067675	-0,829532	0,407189	-0,189094	0,076816		

q(1)	0,054810	0,343243	0,159682	0,873194	-0,619526	0,729145		
Qs(1)	-0,039590	0,040527	-0,976879	0,329090	-0,119208	0,040029		
Омега(1)	0,014312	0,010332	1,385255	0,166576	-0,005986	0,034610	354	ск/врем
Дельта(1)	0,817756	0,282051	2,899323	0,003900	0,263639	1,371873	354	ск/врем

Исходя из этого, делаем вывод о том, что данная модель с интервенцией (не зависимо от их количества) не адекватна.

Построив графики прогнозов, наглядно убедимся в этом.

Прогнозы; Модель:(2,1,1)(0,0,1) 3 интервенции (GOLD_2005_2006_day.sta) Исход.:LOW
Начало исходных: 1 Конец исходн.: 520

	Прогноз	**Нижний**	**Верхни**
1 января 2006 г.	632,6538	622,493	642,8141
2 января 2006 г.	632,4390	616,944	647,9339
3 января 2006 г.	632,1280	612,905	651,3507
4 января 2006 г.	632,1738	609,921	654,4258

5 января 2006 г.	632,4473	607,538	657,3564
8 января 2006 г.	632,3172	605,005	659,6290
9 января 2006 г.	632,2852	602,764	661,8059
10 января 2006 г.	632,1852	600,609	663,7606
11 января 2006 г.	632,2540	598,749	665,7583
12 января 2006 г.	632,1583	596,830	667,4863
15 января 2006 г.	632,3121	595,250	669,3743
16 января 2006 г.	632,3587	593,640	671,0774
17 января 2006 г.	632,4575	592,150	672,7647
18 января 2006 г.	632,5213	590,686	674,3567
19 января 2006 г.	632,5740	589,264	675,8836

Рис. 2.30

Прогноз исходного ряда при помощи модели АРПСС(2,1,1)(0,0,1) с 3 интервенциями.

Прогнозы; Модель:(2,1,1)(0,0,1) 1 интервенции (GOLD_2005_2006_day.sta) Исход.:LOW
Начало исходных: 1 Конец исходн.: 520

	Прогноз	Нижний	Верхний
1 января 2006 г.	632,6685	622,5107	642,8262
2 января 2006 г.	632,4759	617,0081	647,9437
3 января 2006 г.	632,1542	612,9525	651,3559
4 января 2006 г.	632,2070	609,9597	654,4544
5 января 2006 г.	632,5125	607,5939	657,4310
8 января 2006 г.	632,3673	605,0350	659,6996
9 января 2006 г.	632,3325	602,7822	661,8827
10 января 2006 г.	632,2219	600,6090	663,8349
11 января 2006 г.	632,2979	598,7489	665,8469
12 января 2006 г.	632,1913	596,8120	667,5706
15 января 2006 г.	632,3616	595,2421	669,4811
16 января 2006 г.	632,4128	593,6312	671,1944
17 января 2006 г.	632,5218	592,1464	672,8972
18 января 2006 г.	632,5914	590,6828	674,5001
19 января 2006 г.	632,6502	589,2625	676,0379

Рис. 2.31

Прогноз исходного ряда при помощи модели АРПСС(2,1,1)(0,0,1) с одной интервенцией.

Модель АРПСС(1,1,2)(0,0,1)

Для второй модели проводим те же рассуждения. Как и в предыдущем случае, получаем неутешительные результаты о статистической значимости параметров модели. Ниже приведена таблица оценок и график прогноза.

Исход.:LOW (GOLD_2005_2006_day.sta) Преобразования: ln(x),D(1) (Прерванная АРПСС) Модель(1,1,2)(0,0,1) Сезонный лаг: 40 MS Остаток= ,00011

	Парам.	Асимпт.	Асимпт.	p	Нижняя	Верхняя	Интерв.	Интерв.
p(1)	0,242467	0,807387	0,300311	0,764062	-1,34373	1,828660		
q(1)	0,088548	0,805772	0,109892	0,912538	-1,49447	1,671566		
q(2)	0,048303	0,130045	0,371431	0,710470	-0,20718	0,303789		
Qs(1)	-0,037933	0,040468	-0,937364	0,349012	-0,11744	0,041570		
Омега(1)	0,014812	0,010289	1,439593	0,150593	-0,00540	0,035027	354	ск/врем

Дельта1)	0,810609	0,288301	2,811681	0,005117	0,24421	1,377004	354	ск/врем

Прогнозы; Модель:(1,1,2)(0,0,1) 1 интервенции (GOLD_2005_2006_day.sta)
Исход.:LOW Начало исходных: 1 Конец исходн.: 520

	Прогноз	**Нижний**	**Верхний**
1 января 2006 г.	632,8675	621,8016	644,1303
2 января 2006 г.	632,8890	616,0434	650,1953
3 января 2006 г.	632,6079	611,5305	654,4117
4 января 2006 г.	632,7131	608,1188	658,3021
5 января 2006 г.	633,1242	605,4469	662,0668
8 января 2006 г.	632,9565	602,5335	664,9157
9 января 2006 г.	632,9313	599,9918	667,6792
10 января 2006 г.	632,8041	597,5392	670,1501
11 января 2006 г.	632,9118	595,4599	672,7194
12 января 2006 г.	632,7812	593,2827	674,9095
15 января 2006 г.	633,0072	591,5473	677,3728
16 января 2006 г.	633,0770	589,7587	679,5771
17 января 2006 г.	633,2189	588,1186	681,7778
18 января 2006 г.	633,3070	586,5006	683,8488
19 января 2006 г.	633,3863	584,9392	685,8460

Рис. 2.32

Прогноз исходного ряда при помощи модели АРПСС(2,1,1)(0,0,1) с одной интервенцией.

Вывод:

Проделанные попытки построить АРПСС модель, описывающую поведение данного временного ряда оказались безуспешными в том плане, что исходя из полученных результатов, нельзя сделать качественный вывод о поведении ряда с приемлемой степенью точности.

Причиной этому является, во-первых, то, что исходный ряд не обладает четкой сезонной составляющей. Это не позволяет найти повторяющиеся закономерности.

Во-вторых, ряд (а также все его приведенные преобразования) с очень большой степенью погрешности можно назвать стационарным. Этот факт обуславливает невозможность описать поведение ряда моделями авторегрессии и скользящего среднего, для которых это условие является необходимым.

В-третьих, очевидно, что структура предполагаемой взаимосвязи членов ряда с предыдущими очень сложная.

Следующим шагом на нашем пути прогнозирования будет применения модели Census I (применение сезонной декомпозиции).

3.Модель Census I(применение сезонной декомпозиции)

В диалоге «Временные ряды и прогнозирование» выберем метод **Census I(применение сезонной декомпозиции)**. После этого станет активным диалог.

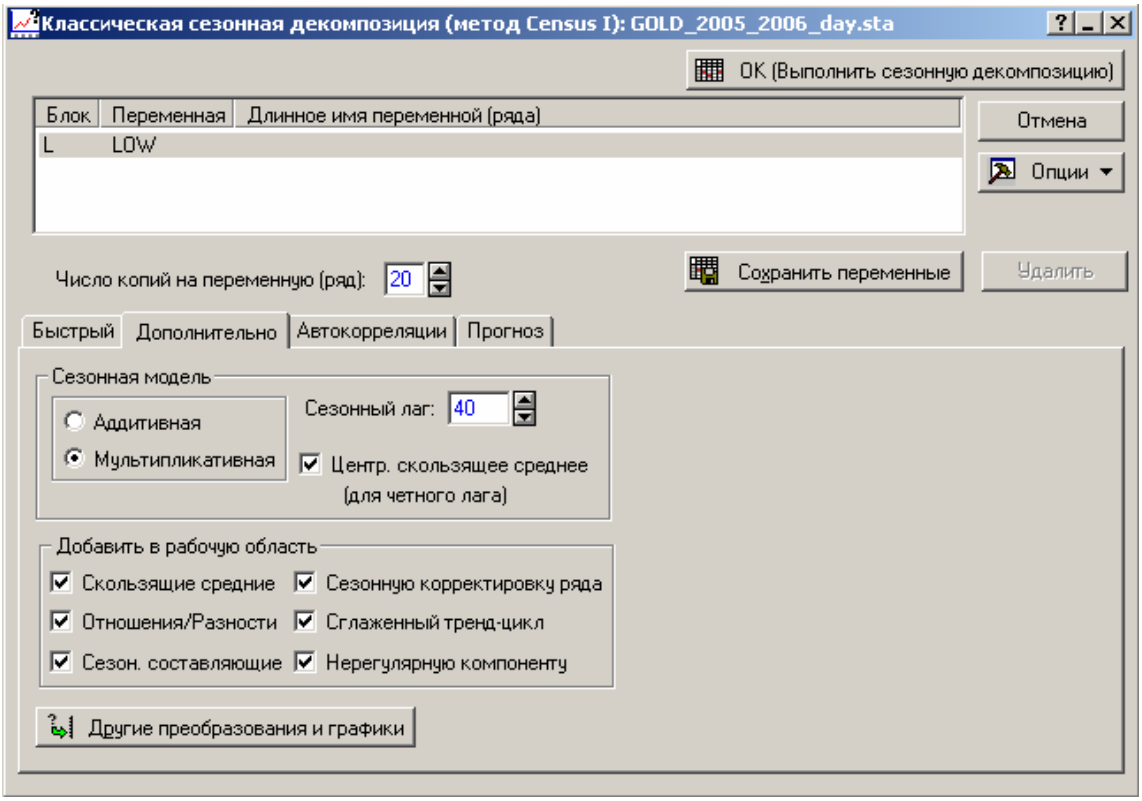


Рис. 3.1

Диалог «Классическая сезонная декомпозиция (метод Census I)»

На вкладке «Дополнительно» выберем значение *Мультипликативная сезонность*. В качестве значения величины сезонного цикла возьмем значение 40 (причина- анализ выше) Такой выбор можно аргументировать тем, что размах «колебаний» увеличивается с ростом номера наблюдений.

После нажатия кнопки «ОК» в рабочую область добавятся 6 рядов, которые мы потребовали добавить в настоящем диалоге.

Построим графики этих рядов.

1.Сглаживание скользящим средним

Сначала вычисляется скользящее среднее для временного ряда, при этом ширина окна берется равной периоду сезонности.

Рис 3.2

Сглаживание взвешенным 5-ти точечным средним

2.Отношение

После взятия скользящих средних вся сезонная (т.е. внутрисезонная) изменчивость будет исключена, поэтому отношение между наблюдаемым и сглаженным рядом будет выделять сезонную составляющую (плюс нерегулярную компоненту). Поскольку мы имеем дело с мультипликативной сезонностью, то сезонную составляющую нашего ряда будем выделять, деля значения ряда на значения скользящих средних

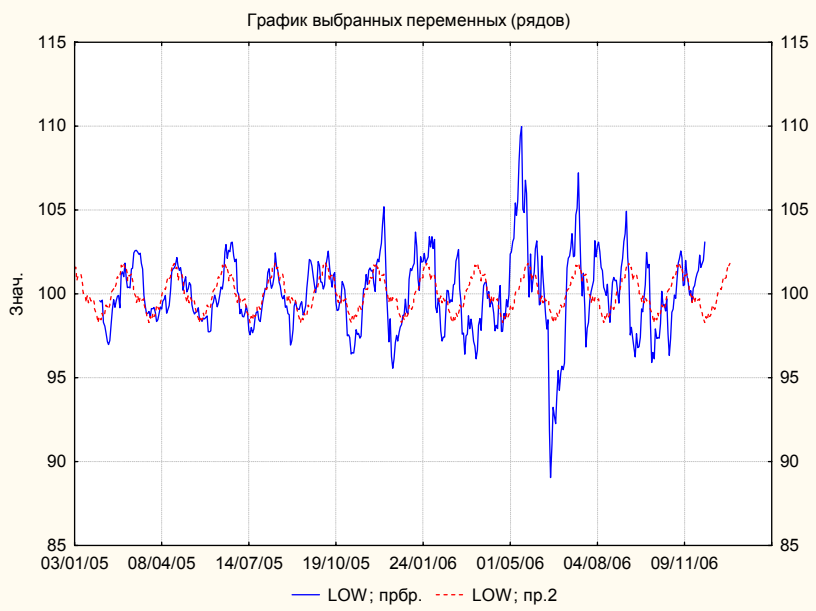


Рис 3.3

Отношение (пунктиром отмечена сезонная составляющая)

3.Сезонная составляющая

На следующем шаге вычисляется сезонная составляющая как урезанное среднее (поскольку наша модель мультипликативная по предположению) всех значений ряда, соответствующих данной точке сезонного интервала.

Урезанным средним некоторого множества чисел называется их среднее, взятое после удаления наибольшего и наименьшего значений. Полученный в результате этих действий ряд является (усредненной) сезонной составляющей исходного ряда.

Рис 3.4

Сезонная составляющая

4.Сглаженный тренд-цикл

Исходный ряд можно скорректировать, деля его значения на значения сезонной составляющей. Получающийся в результате ряд называется сезонной корректировкой ряда (т.е. из ряда убрана его сезонная составляющая).

Приближение для объединенной тренд-циклической компоненты можно получить, применяя к ряду с сезонной поправкой процедуру 5-точечного (центрированного) взвешенного скользящего среднего с весами 1, 2, 3, 2, 1.

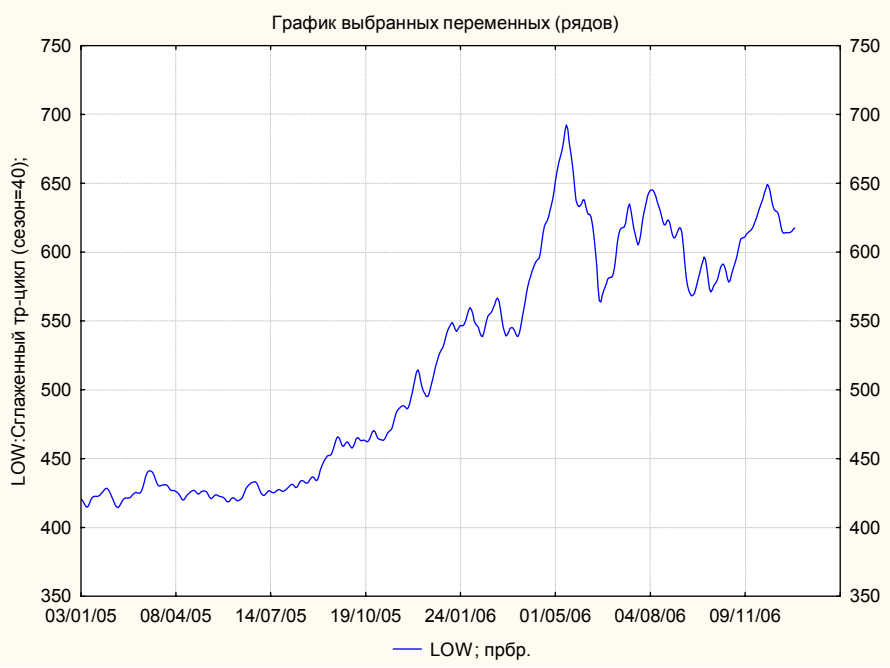


Рис 3.5

Сглаженный тренд-цикл

5.Нерегулярная компонента(ошибки)

На последнем шаге выделяется случайная или нерегулярная компонента (погрешность) путем делением этого ряда) на тренд-циклическую компоненту

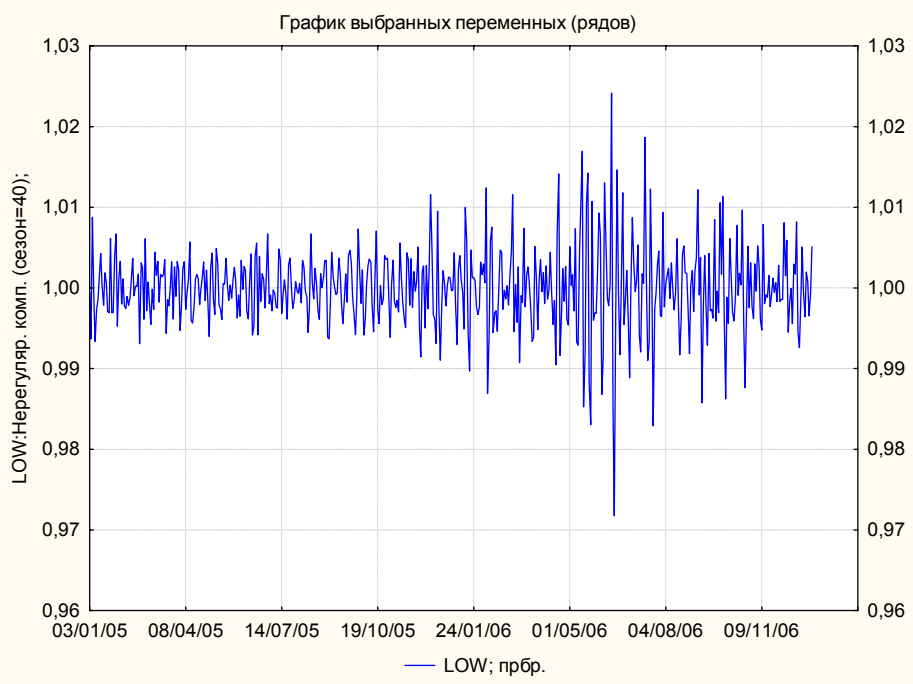


Рис 3.6

Нерегулярная компонента(ошибки)

6.Сглаженный тренд-цикл и ошибки

Для того, чтобы оценить качество сезонной декомпозиции целесообразно рассмотреть графики тренд-циклической компоненты и ошибок в одной системе координат (в разных масштабах). Как нетрудно заметить, например, наибольшие ошибки появляются в периоды резких скачков ряда (или интервенций, о которых говорилось выше)

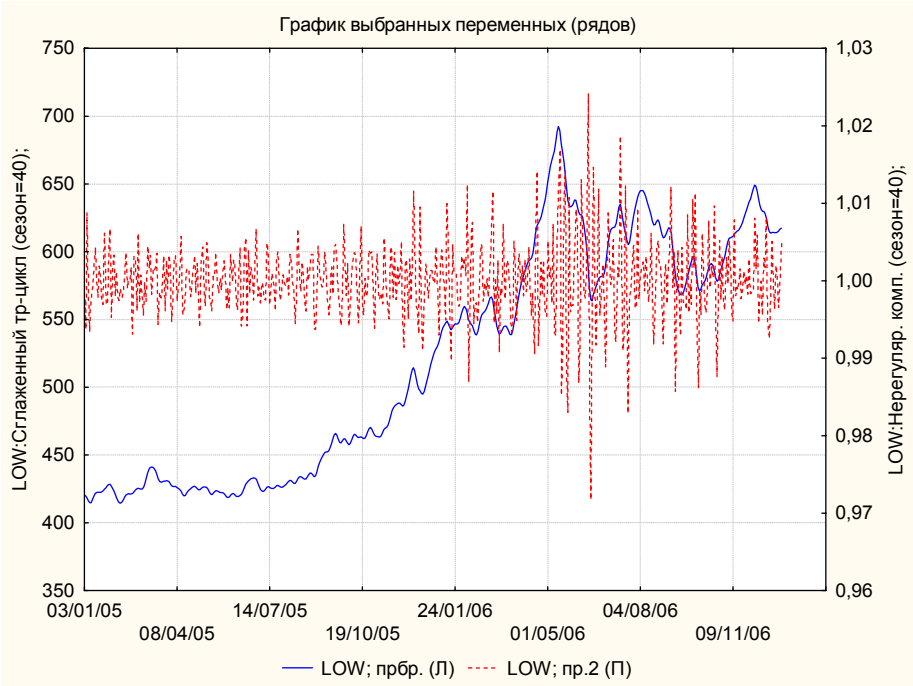


Рис 3.7

Сглаженный тренд-цикл и ошибки

Замечание:

Для того, чтобы построить какой-либо нелинейный график (например, гистограмму) одного из рядов (например, ошибок), находящегося в рабочей области, необходимо выделить этот ряд курсором и нажать кнопку построения необходимого графика.

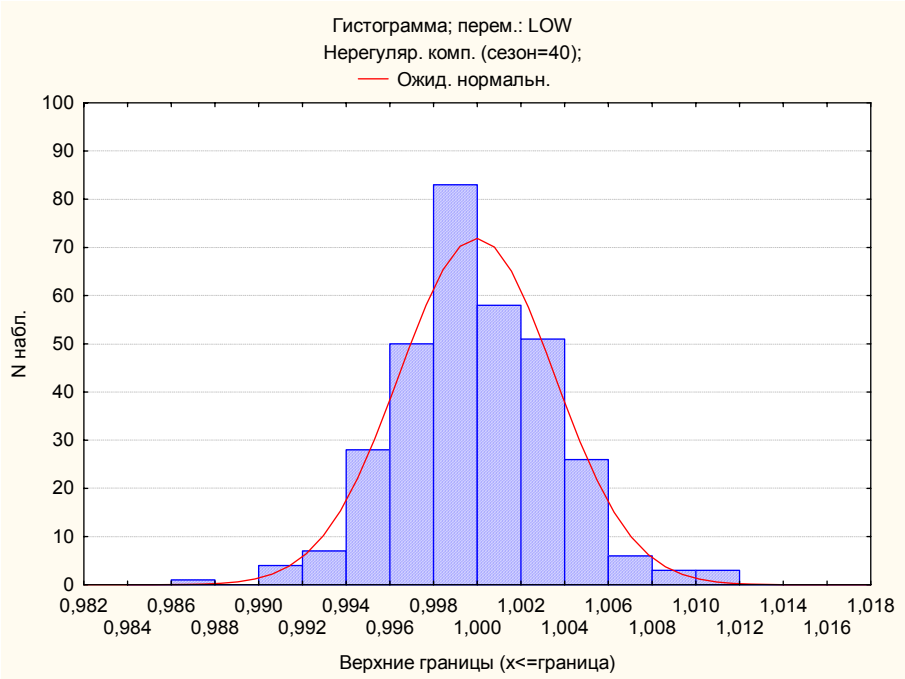


Рис 3.8

Гистограмма частот остатков при сезонной декомпозиции

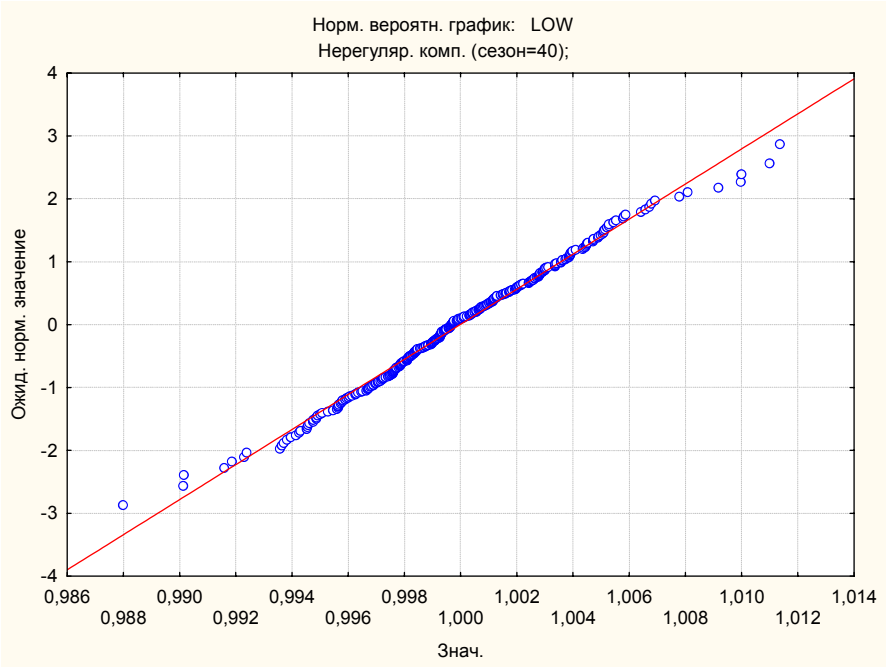


Рис 3.9

Нормальный вероятностный график остатков при сезонной декомпозиции.

Заметим, что сглаженный тренд-цикл не сильно отличается от исходного ряда. Причинами этого является то, что размах сезонной составляющей очень мал по сравнению со значениями ряда и его размахом (размах сезонной составляющей равен 4, размах ряда порядка 300!).

Для наглядности приведем таблицу корреляций всех компонент.

Корреляции (Сезонная декомпозиция: Мультип. сезон. (40) (GOLD_2005_2006_day.sta) в Workbook1) Отмеченные корреляции значимы на уровне p

	LOW	Скольз.	Отнош.	Сезонные	Скоррек.	Сглажен.	Нерег.
LOW	1,0000	,9861	,2300	,0946	,9977	,9970	,0733
	p= ---	p=0,00	p=,000	p=,041	p=0,00	p=0,00	p=,114
Скольз.	,9861	1,0000	,0666	,0212	,9889	,9908	,0044
	p=0,00	p= ---	p=,151	p=,648	p=0,00	p=0,00	p=,924
Отнош.	,2300	,0666	1,0000	,4819	,1977	,1829	,4146
	p=,000	p=,151	p= ---	p=0,00	p=,000	p=,000	p=0,00
Сезонные	,0946	,0212	,4819	1,0000	,0275	,0270	,0100
	p=,041	p=,648	p=0,00	p= ---	p=,553	p=,560	p=,829
Скоррек.	,9977	,9889	,1977	,0275	1,0000	,9993	,0724
	p=0,00	p=0,00	p=,000	p=,553	p= ---	p=0,00	p=,118
Сглажен.	,9970	,9908	,1829	,0270	,9993	1,0000	,0361
	p=0,00	p=0,00	p=,000	p=,560	p=0,00	p= ---	p=,437
Нерег.	,0733	,0044	,4146	,0100	,0724	,0361	1,0000
	p=,114	p=,924	p=0,00	p=,829	p=,118	p=,437	p= ---

В данной таблице коэффициент корреляции между переменной LOW и Сезонной составляющей хоть и является статистически значимым, но p-уровень практически на границе допустимого, а непосредственно сам коэффициент весьма мал (.0946). Также отметим, что коэффициент корреляции переменной LOW и Сезонной компоненты по уровню значимости не очень сильно отличается от соответствующего коэффициента между LOW и Нерегулярной компоненты, хотя корреляции между последними переменными быть не должно.

Корреляции (Сезонная декомпозиция: Мультип. сезон. (40) (GOLD_2005_2006_day.sta) в Workbook1) Отмеченные корреляции значимы на уровне p

	LOW	Сезонные	Нерег.
LOW	1,0000	,0489	,0708
	p= ---	p=,273	p=,112
Сезонные	,0489	1,0000	-,0087
	p=,273	p= ---	p=,845
Нерег.	,0708	-,0087	1,0000
	p=,112	p=,845	p= ---

В таблице выше представлены парные корреляции только между тремя переменными(LOW, Сезонная компонента и Нерегулярная компонента). Картина, наблюдаемая здесь, еще более наглядно показывает независимость LOW от Сезонной составляющей. Более того, судя по значениям p- уровня, Нерегулярная компонента оказывает большее влияние на исходный ряд, что невозможно по предположению.

Таким образом можно сделать вывод о том, что сезонную компоненту мы выделить не можем, а следовательно применение сезонной декомпозиции также не приводит к хорошим результатам

4.Экспоненциальное сглаживание

Для дальнейшего анализа воспользуемся методом экспоненциального сглаживания. В диалоге «Временные ряды и прогнозирование» выберем соответствующий метод. Диалог «Сезонное и несезонное экспоненциальное сглаживание» представлен ниже.

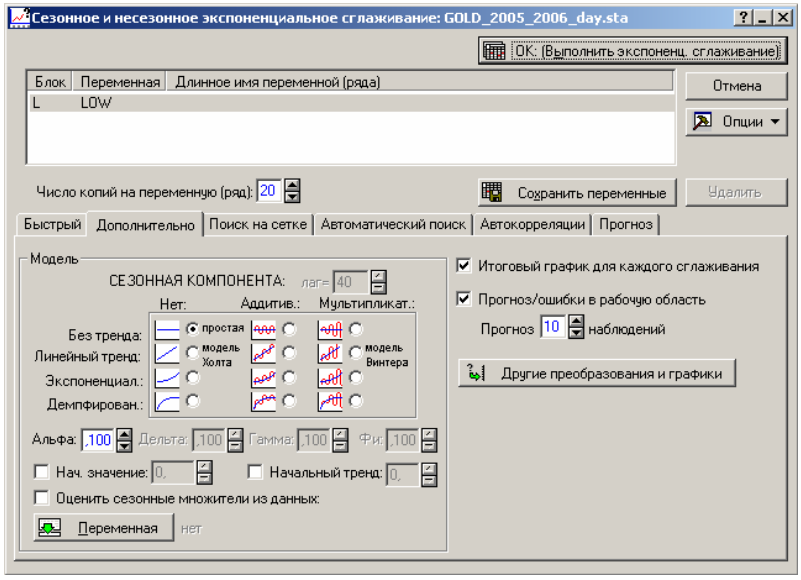


Рис.4.1

Диалог «Сезонное и несезонное экспоненциальное сглаживание»

На вкладке «Дополнительно» пользователю предлагается априорно оценить вид сезонной составляющей (нет, аддитивная сезонность, мультипликативная сезонность) и вид трендовой составляющей (без тренда, линейный тренд, экспоненциальный тренд, демпфированный тренд). Также мы можем выбрать параметры сглаживания и величину прогноза. Как было замечено выше сезонность мультипликативная. При визуальном анализе заметим, что тренд безусловно есть. Мы построим экспоненциальное сглаживание для линейного тренда и для экспоненциального тренда, потому как обе эти модели на первый взгляд возможны.

Количество задействованных параметров зависит от выбранной модели. Оптимальные для данной задачи значения этих параметров пользователь может найти двумя способами, а именно путем поиска на сетке или автоматически.

Для того чтобы выбрать первый вариант перейдите на вкладку «Поиск на сетке»

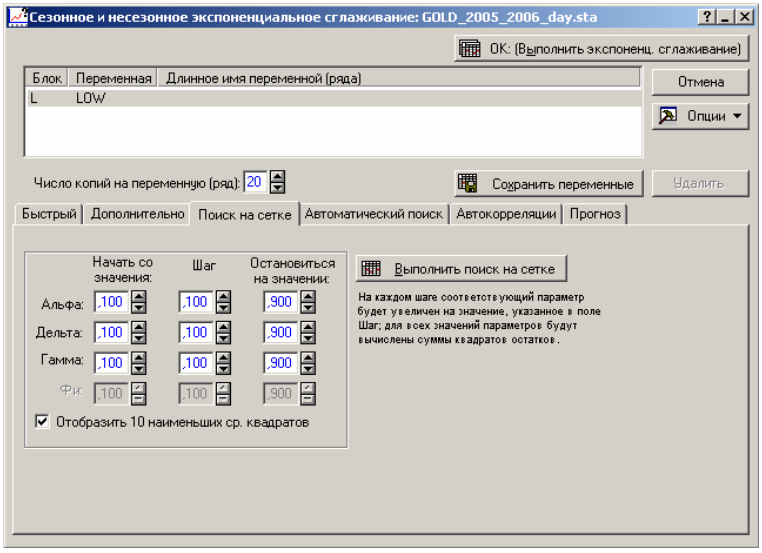


Рис. 4.2

Вкладка «Поиск на сетке» диалога «Сезонное и несезонное экспоненциальное сглаживание»

Алгоритм поиска на сетке сводится к перебору и сравнению по принципу наименьших квадратов ошибок моделей с параметрами, лежащими в узлах параметрической многомерной сетки с указанными границами и шагом ячейки.

Для осуществления автоматического поиска параметров необходимо выбрать вкладку «Автоматический поиск». Критерием качества модели с конкретными параметрами может является либо сумма квадратов ошибок, либо средняя абсолютная ошибка, либо средняя абсолютная относительная ошибка.

Для этой цели в модуле *Временные ряды и прогнозирование* реализован так называемый *квази-ньютоновский* метод (тот же метод используется при оценке параметров АРПСС)

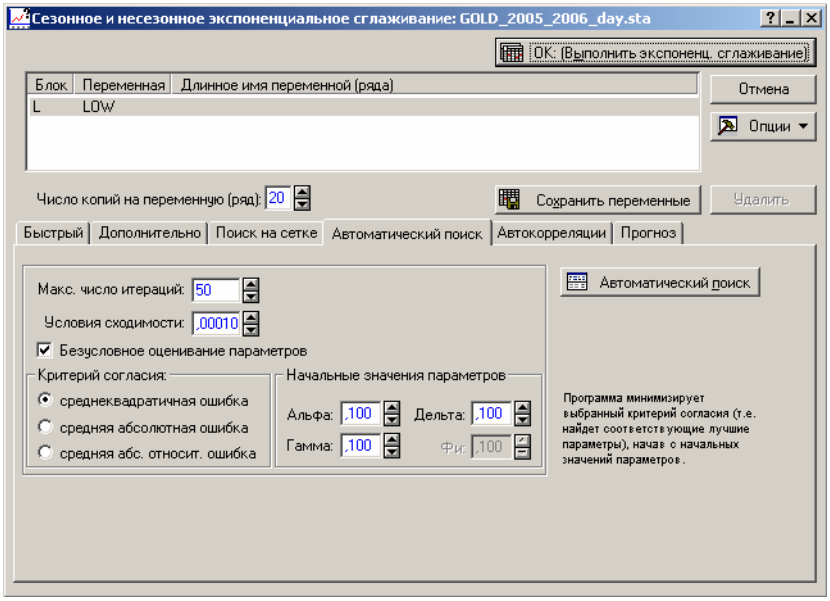


Рис.4.3

Вкладка «Автоматический поиск» диалога «Сезонное и несезонное экспоненциальное сглаживание»

Необходимо указать начальные значения параметров, а также выбрать критерий согласия.

Выберем сначала модель с линейным трендом.

Линейный тренд (Модель Винтера)

Будем искать параметры автоматически. Итак,

Эксп. сглажив.: Мульт. сезон. (40) S0=412,6 T0=,4230 (GOLD_2005_2006_day.sta) Лин.тренд,мульт.сезон.; Альфа= 1,00 Дельт=,131 Гамма=0,00 LOW

	Ошибка
Средн. ош.	-0,0240285198
Ср. абсол. ошибка	4,0093904646
Суммы квадратов	18160,5398800832
Ср. квадрат	34,9241151540
Средн. относ. ошибка	-0,0103516471
Ср. абс. отн. ошибка	0,7395283255

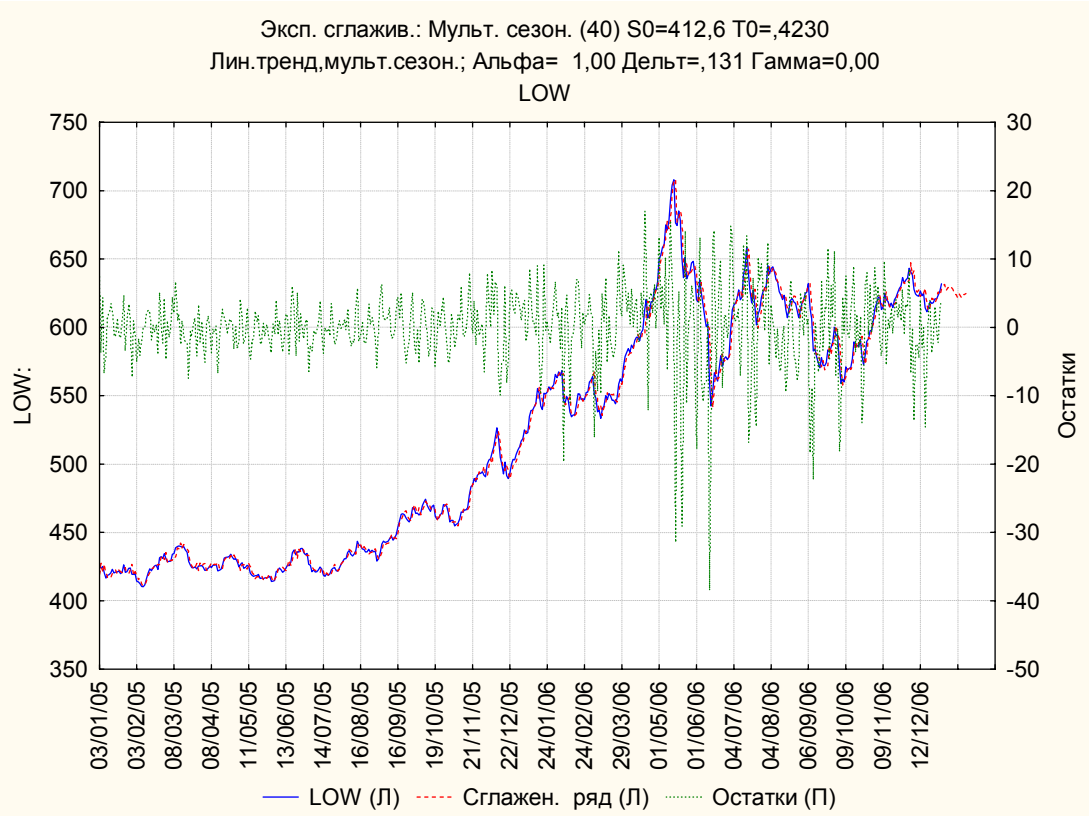
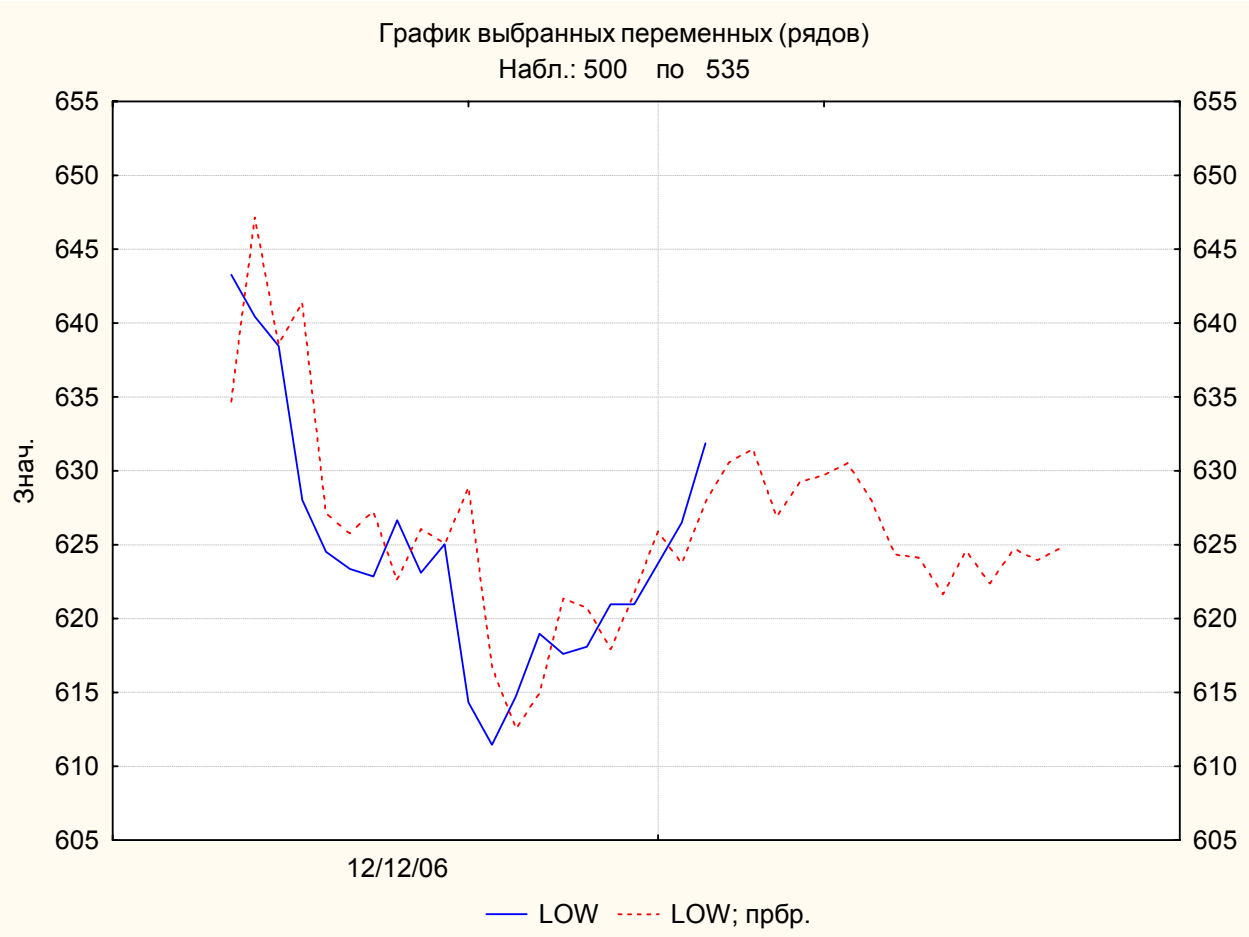


Рис.4.4

Экспоненциальное сглаживание

модель Винтера (мультипликативная сезонность, линейный тренд)

Для более детального рассмотрения отобразим только заключительный отрезок ряда (наблюдения 500 - 535) и прогноз экспоненциального сглаживания.



Эксп. сглажив.: Мульт. сезон. (40) S0=412,6
T0=,4230 (GOLD_2005_2006_day.sta)
Лин.тренд,мульт.сезон.; Альфа= 1,00
Дельт=,131 Гамма=0,00 LOW

	Сглажен.
1 января 2006 г.	630,5748
2 января 2006 г.	631,4620
3 января 2006 г.	626,8974
4 января 2006 г.	629,2557
5 января 2006 г.	629,6907
8 января 2006 г.	630,5116
9 января 2006 г.	628,0190
10 января 2006 г.	624,3222
11 января 2006 г.	624,1237

12 января 2006 г.	621,6381
15 января 2006 г.	624,5944
16 января 2006 г.	622,3807
17 января 2006 г.	624,7193
18 января 2006 г.	623,9565
19 января 2006 г.	624,8070

Рис. 4.5

Конечный отрезок (ноябрь 2006 г – январь 2007 г) исходного ряда и ряда, полученного в результате экспоненциального сглаживания в модели Винтера

Итак, видим довольно неплохую подгонку исходного ряда, но прогноз цены золота, построенный при помощи экспоненциального сглаживания говорит о её снижении, что идет вразрез с лучшими результатами, полученными при построении моделей АРПСС.

Также проведем анализ остатков.

Коррелограммы

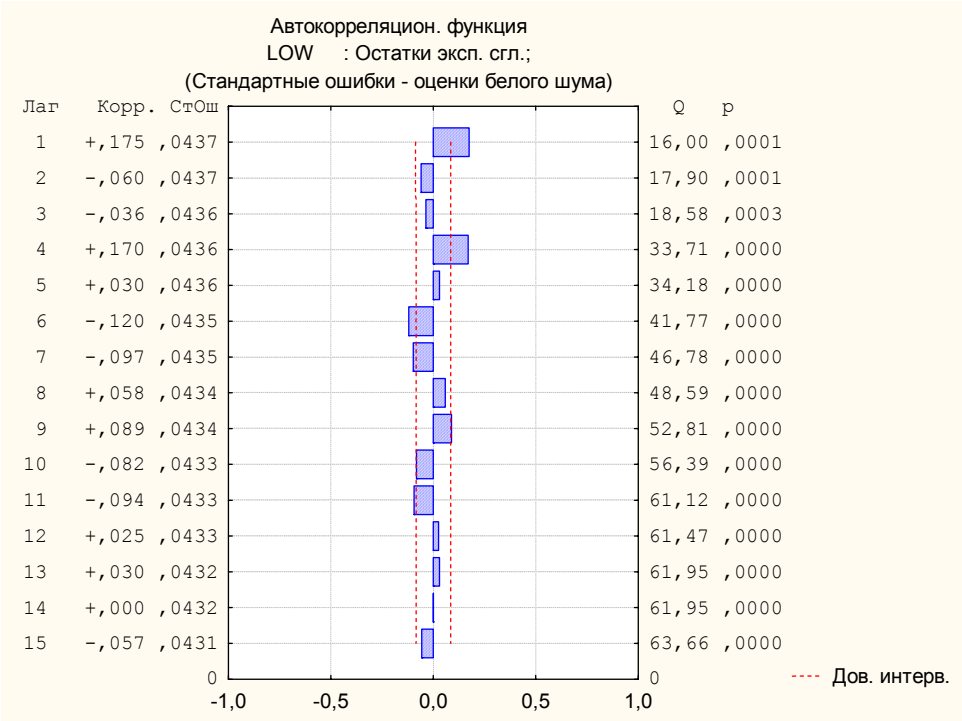


Рис. 4.6

Автокоррелограмма остатков экспоненциально сглаживания в модели Винтера.

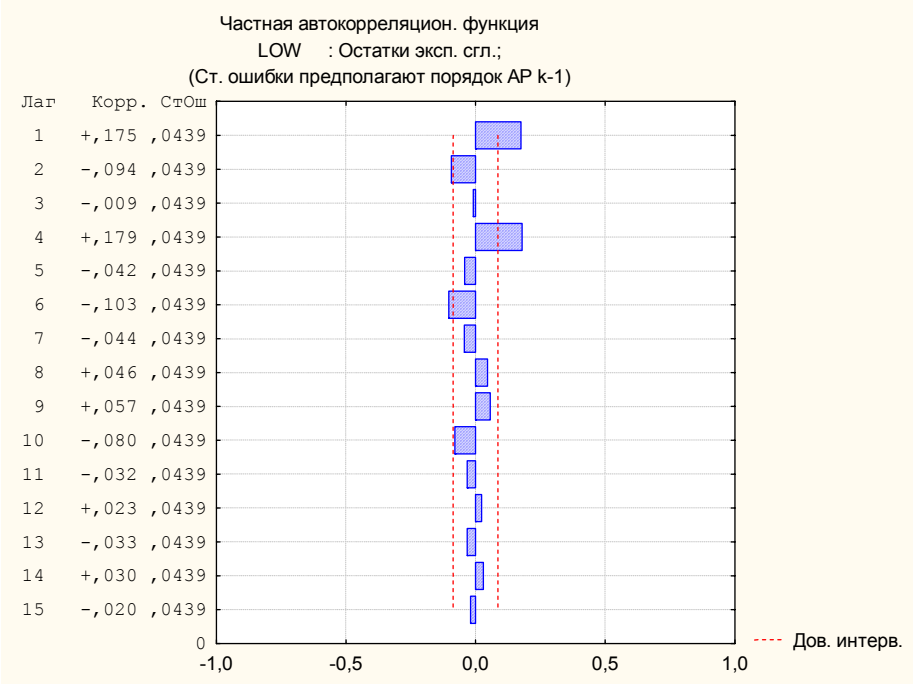


Рис. 4.7

Частная автокоррелограмма остатков экспоненциально сглаживания в модели Винтера.

Коррелограммы нельзя назвать «идеальными», но тем не менее на лицо убывание их значений на больших по величине лагах и сравнительно маленькие отклонения от допустимого уровня на начальных значениях лагов.

Анализ нормальности остатков

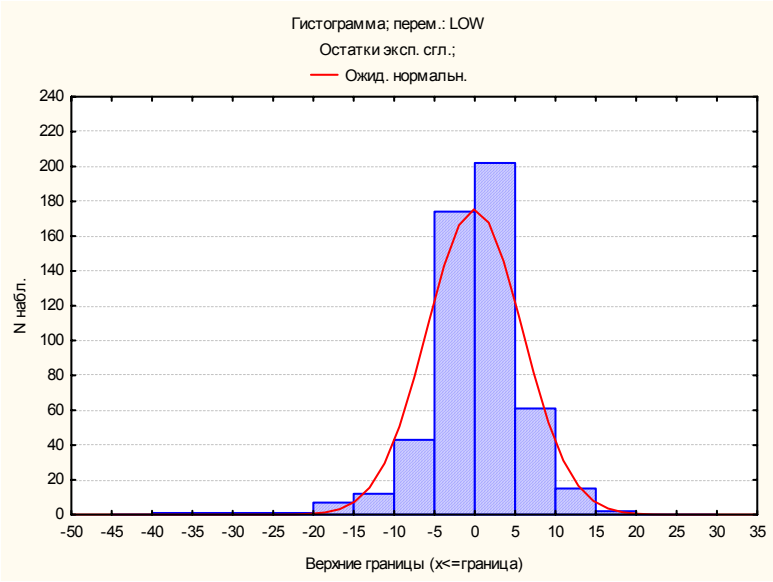


Рис. 4.8

Гистограмма частот остатков при экспоненциальном сглаживании в модели Винтера.

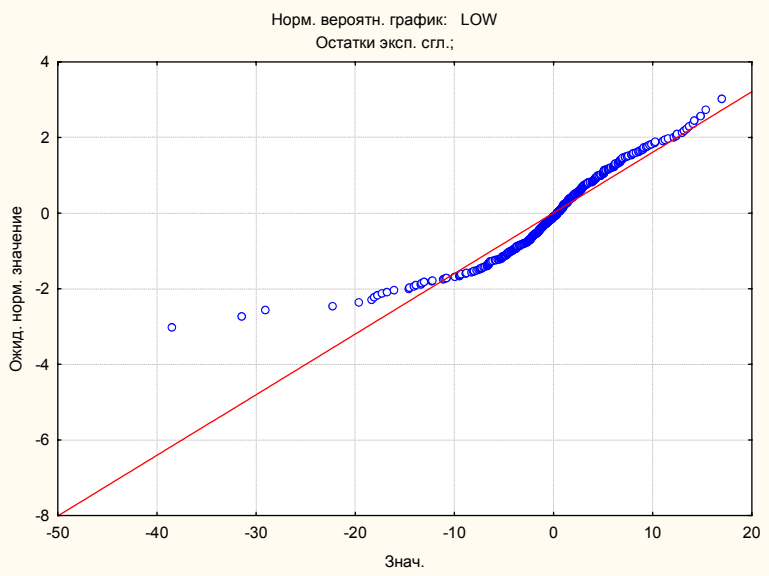


Рис 4.9

Нормальный вероятностный график остатков при экспоненциальном сглаживании в модели Винтера.

И гистограмма, и нормальный вероятностный график сообщают нам о том, что распределение остатков довольно близко к нормальному с нулевым математическим ожиданием. Тем не менее, погрешность принятия этой гипотезы все же существует и она предельна.

Экспоненциальный тренд

Опять ищем параметры автоматически

Эксп. сглажив.: Мульт. сезон. (40) S0=417,0 T0=1,000 (GOLD_2005_2006_day.sta) Эксп. тренд,мульт.сезон.; Альфа= 1,00 Дельт=,098 Гамма=,009 LOW	
	Ошибка
Средн. ош.	-0,0572667097
Ср. абсол. ошибка	3,9868016608
Суммы квадратов	18272,9852849727
Ср. квадрат	35,1403563173
Средн. относ. ошибка	-0,0094005631
Ср. абс. отн. ошибка	0,7353829667

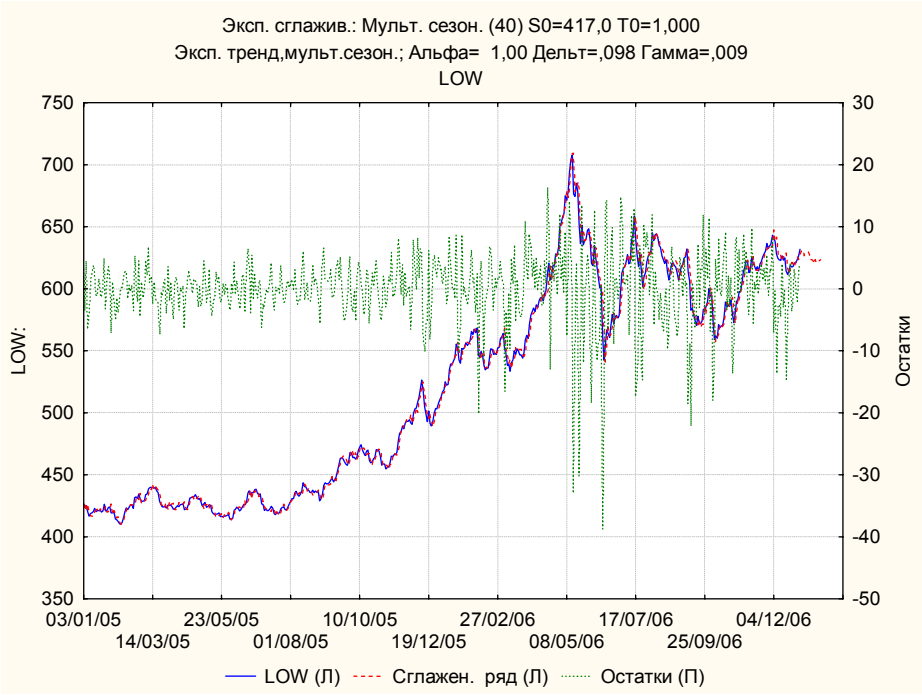
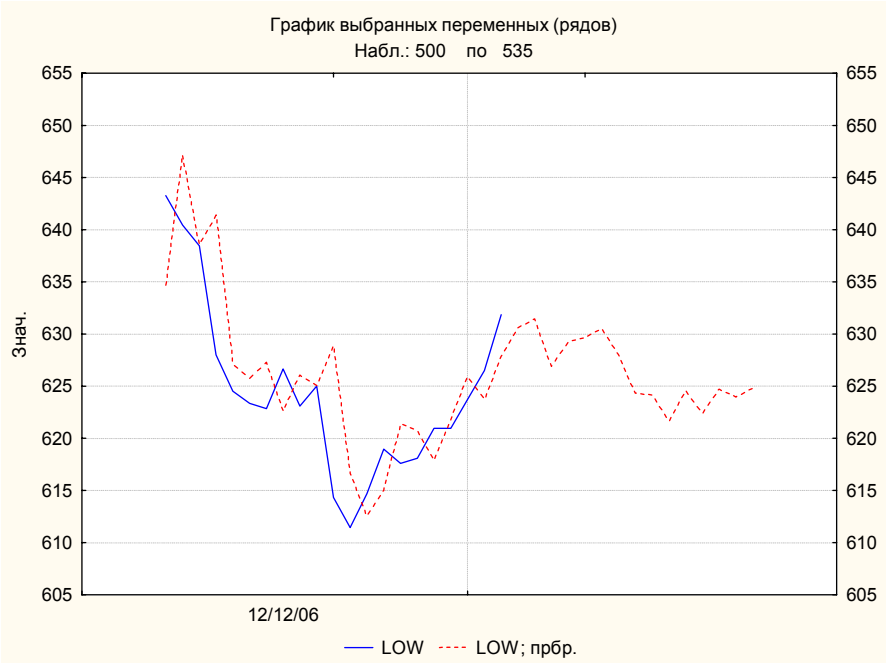


Рис.4.10

Экспоненциальное сглаживание

мультипликативная сезонность, экспоненциальный тренд.

Для более детального рассмотрения отобразим только заключительный отрезок ряда (наблюдения 500 - 535) и прогноз экспоненциального сглаживания.



Эксп. сглажив.: Мульт. сезон. (40) S0=417,0 T0=1,000 (GOLD_2005_2006_day.sta) Эксп. тренд,мульт.сезон.; Альфа= 1,00 Дельт=,098 Гамма=,009 LOW	
	Сглажен.
1 января 2006 г.	630,4833
2 января 2006 г.	631,2791
3 января 2006 г.	626,6255
4 января 2006 г.	628,8924
5 января 2006 г.	629,2370
8 января 2006 г.	629,9673
9 января 2006 г.	627,3875
10 января 2006 г.	623,6060
11 января 2006 г.	623,3196
12 января 2006 г.	620,7497
15 января 2006 г.	623,6142
16 января 2006 г.	621,3169
17 января 2006 г.	623,5646
18 января 2006 г.	622,7165
19 января 2006 г.	623,4788

Рис. 4.11

Конечный отрезок (ноябрь 2006 г– январь 2007 г) исходного ряда и ряда, полученного в результате экспоненциального сглаживания в модели с мультипликативной сезонностью и экспоненциальным трендом

Проведем анализ остатков на предмет справедливости стандартных предположений.

Построим коррелограммы, а также гистограммы частот остатков и нормальный вероятностный график этих остатков.

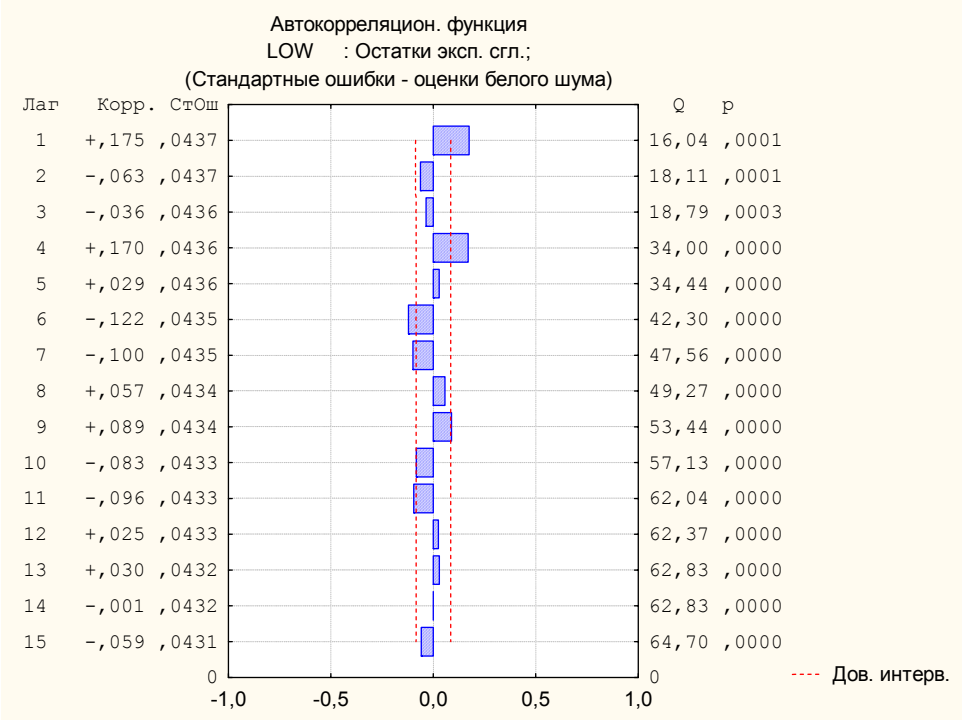


Рис. 4.12

Автокоррелограмма остатков экспоненциально сглаживания в модели с мультипликативной сезонностью и экспоненциальным трендом.

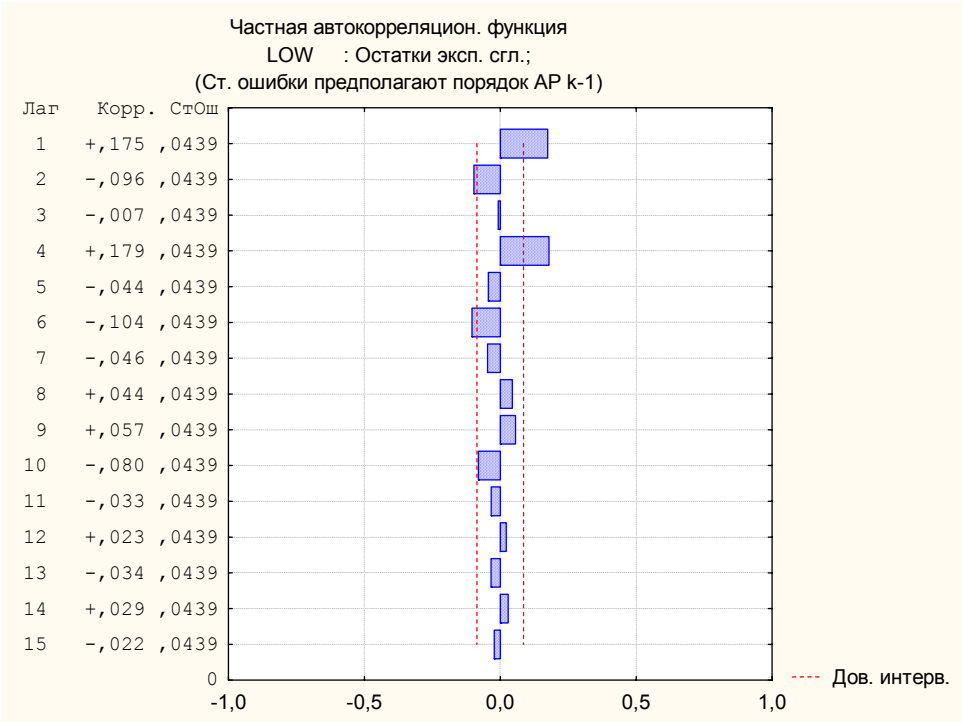


Рис. 4.13

Частная автокоррелограмма остатков экспоненциально сглаживания в модели с мультипликативной сезонностью и экспоненциальным трендом.

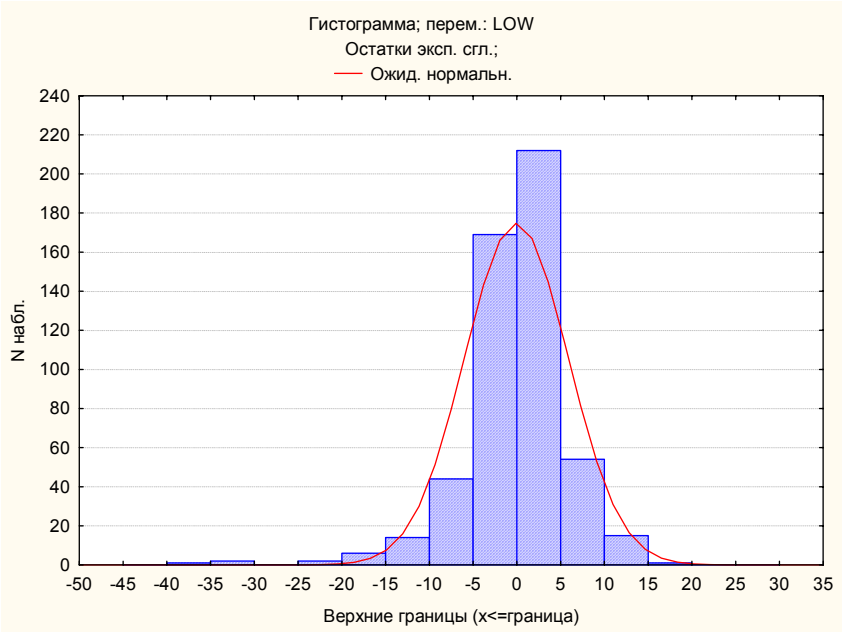


Рис. 4.14

Гистограмма частот остатков при экспоненциальном сглаживании в модели с мультипликативной сезонностью и экспоненциальным трендом.

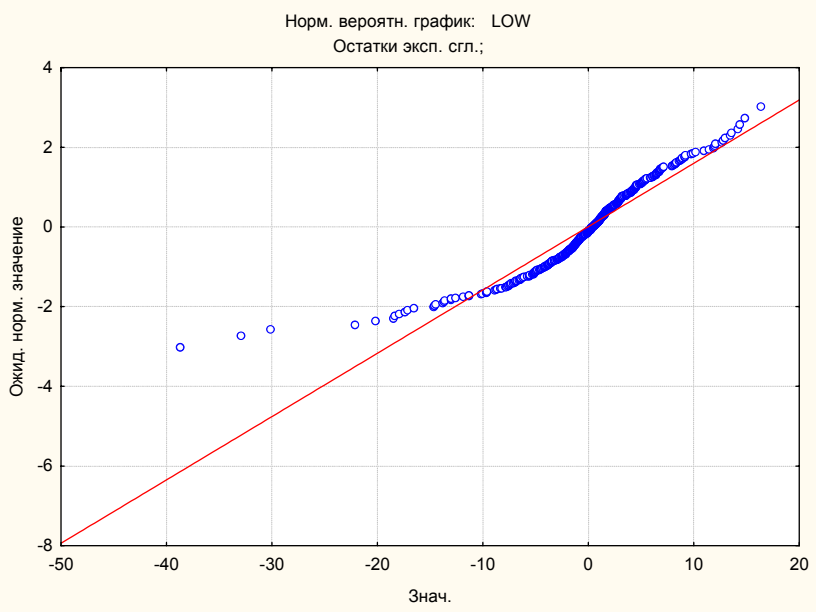


Рис 4.15

Нормальный вероятностный график остатков при экспоненциальном сглаживании в модели с мультипликативной сезонностью и экспоненциальным трендом.

Анализ остатков в этой модели дает приблизительно те же результаты, что и в модели Винтера, но судя по визуальному анализу погрешность несколько меньше. На это указывает то, что нормальный вероятностный график ближе прилегает к биссектрисе координатного угла, а отклонения значений коррелограммы от допустимого уровня меньше.

Таким образом, можно сделать вывод, что в данном ряде имеет место мультипликативная сезонность.

Следует также объяснить причину таких малых различий в принципиально разных по сути моделях. Как отмечалось выше при сезонной декомпозиции, корреляция ряда с выделенной сезонной составляющей предельно мала, что говорит о малой степени влияния последней на поведение ряда в целом. Было отмечено также, что размах сезонной компоненты несравним с размахом всего ряда, поэтому, как известно из математики, умножение на малоизменяющуюся величину с большой степенью точности можно аппроксимировать суммированием с другой малоизменяющейся величиной.

Вывод:

При рассмотрении экспоненциального сглаживания в моделях с обоими типами трендов мы получили очень похожие результатов прогнозов. Поведение остатков тоже очень похоже в двух случаях и по большому счету можно принять гипотезу о справедливости стандартных предположений.

Метод экспоненциального сглаживания дал нам более качественные результаты (в том смысле, что прогноз ряда не представляет собой прямую линию), но останавливаться на достигнутых результатах не стоит, потому как во-первых, ошибки «не идеальны» по структуре; во-вторых, структура самого ряда слишком сложная, чтобы доверять одному методу; в-третьих, ошибки прогнозирования довольно велики.

Применив классические методы и не получив полностью удовлетворяющих результатов, обратимся теперь к неклассическому методу в прогнозировании – применение нейронных сетей.

Для того чтобы запустить модуль «Нейронные сети», необходимо выбрать соответствующий тип анализа в меню «Анализ». Следует отметить, что применение нейронных сетей является независимым средством прогнозирования по отношению к разобранным выше методам.

5.Прогнозирование при помощи Нейронных Сетей

Концепция нейронный сетей позволяет осуществить качественно иной подход к прогнозированию временных сетей по отношению к классическим методам, рассмотренным выше. Выводы классических методов строились на поиске и анализе скрытых закономерностей: цикличностей, сезонных колебаний.

Отсутствие четко выраженных факторов, указанных выше, как мы убедились, делает применение классических методов практически бесполезным, а степень доверия полученным прогнозам весьма мала.

Нелинейные по своей сути нейронные сети, вообще говоря, позволяют с любой степенью точности аппроксимировать произвольную непрерывную функцию, не взирая на отсутствие или наличие какой-либо периодичности или цикличности.

Поскольку временной ряд представляет собой непрерывную функцию (на самом деле нам известно значение этой функции лишь в конечном числе точек, но её можно легко непрерывно продолжить на весь рассматриваемый отрезок), то применение нейронных сетей вполне оправдано и корректно.

Забегая вперед, отметим, что зачастую именно применение нейронных сетей оказывается ключевым моментом в процессе прогнозирования. Именно этот метод нередко позволяет добиться наиболее точных результатов.

Итак, запустим модуль «Нейронные сети» в меню «Анализ».

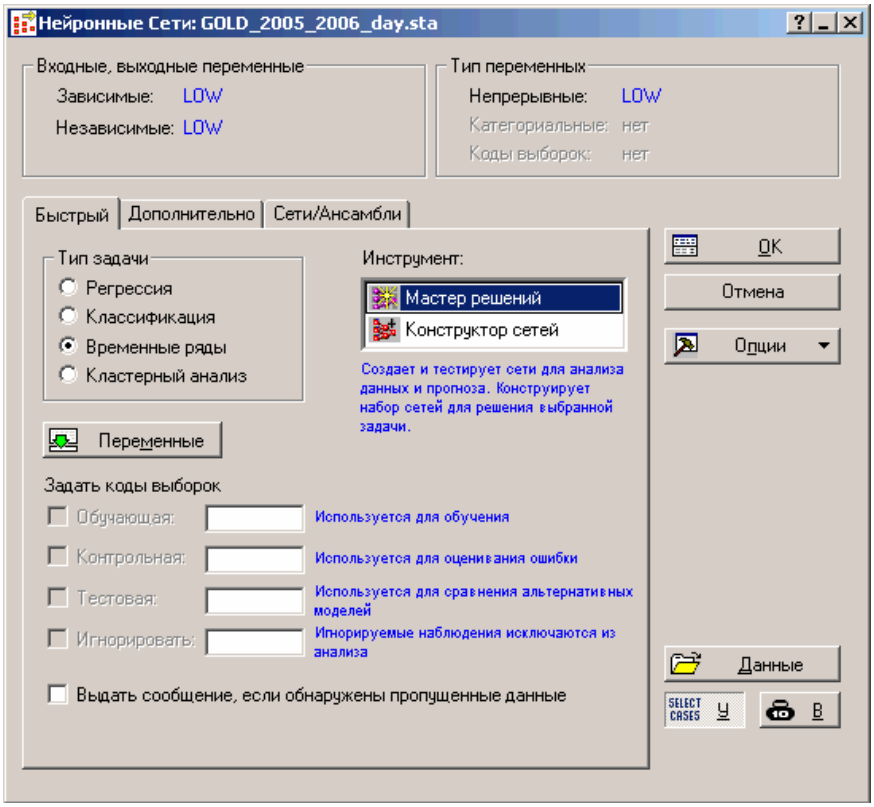


Рис 5.1

Диалог «Нейронные Сети»

В появившемся диалоге на вкладке «Быстрый» необходимо выбрать тип задачи «Временные ряды», а также задать входные и выходные переменные. Инструмент «Мастер решений» оставим по умолчанию. После этого нажмите «ОК» для введения параметров строящихся сети.

Отметим, что в задаче прогнозирования временных рядов одна и та же переменная может являться как входной, так и выходной одновременно. При конфигурировании сети для анализа временных рядов изменяется метод пре-процессирования данных (извлекаются не отдельные наблюдения, а их блоки), но обучение и работа сети происходят точно так же, как и в задачах других типов.

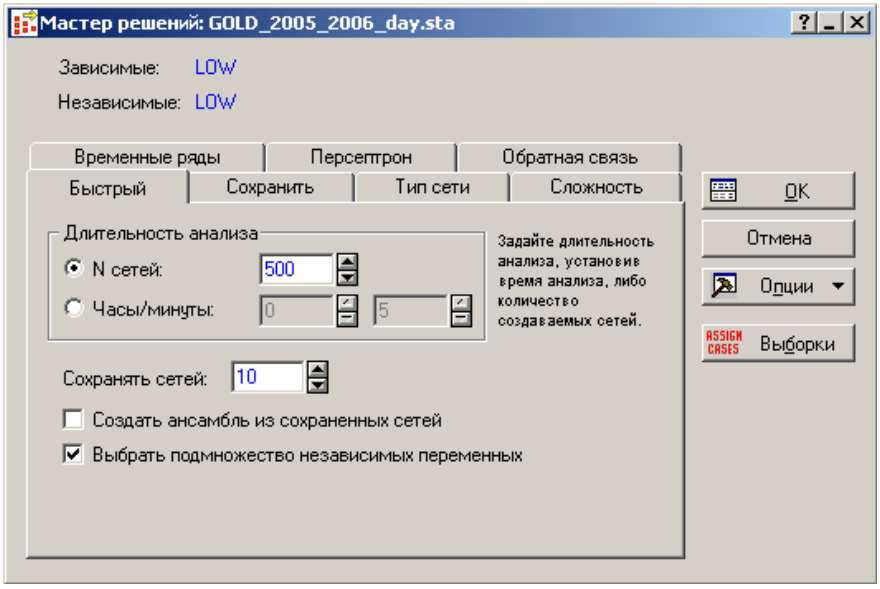


Рис 5.2

Диалог «Мастер решений»

Вкладка «Быстрый»

В появившемся диалоге есть много различных вкладок, предназначенных для уточнения параметров, как самих сетей, так и параметров вывода.

Зададим программе построить и обучить 500 сетей различного типа (о том, как этот тип выбрать см. ниже) и выбрать из них 10 наилучших в смысле критерия, который задается на вкладке «Сохранить»

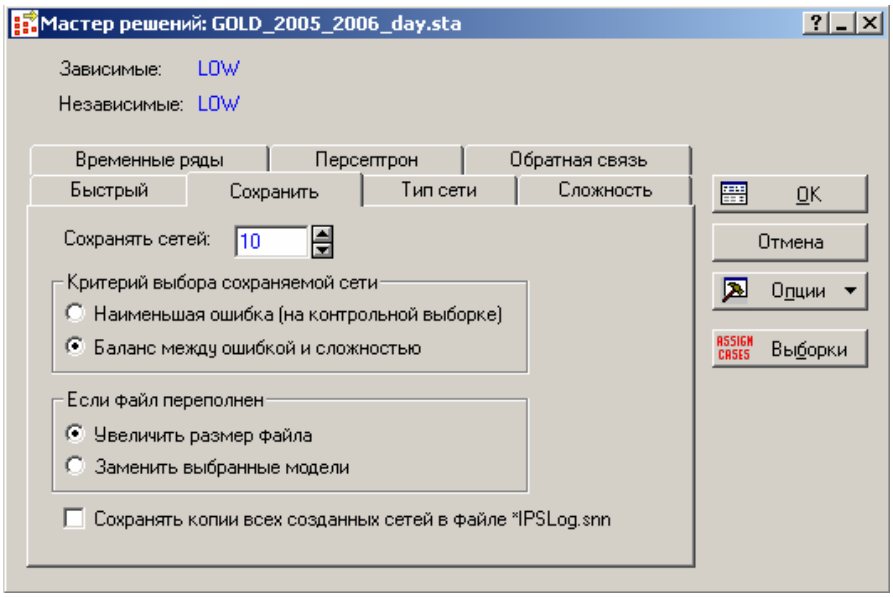


Рис 5.3

Диалог «Мастер решений»

Вкладка «Сохранить»

На вкладке «Временные ряды» необходимо задать размер окна анализа(задание диапазоном). Размер окна анализа и задает величину блока данных, которые используются в процедуре пре-процессирования.

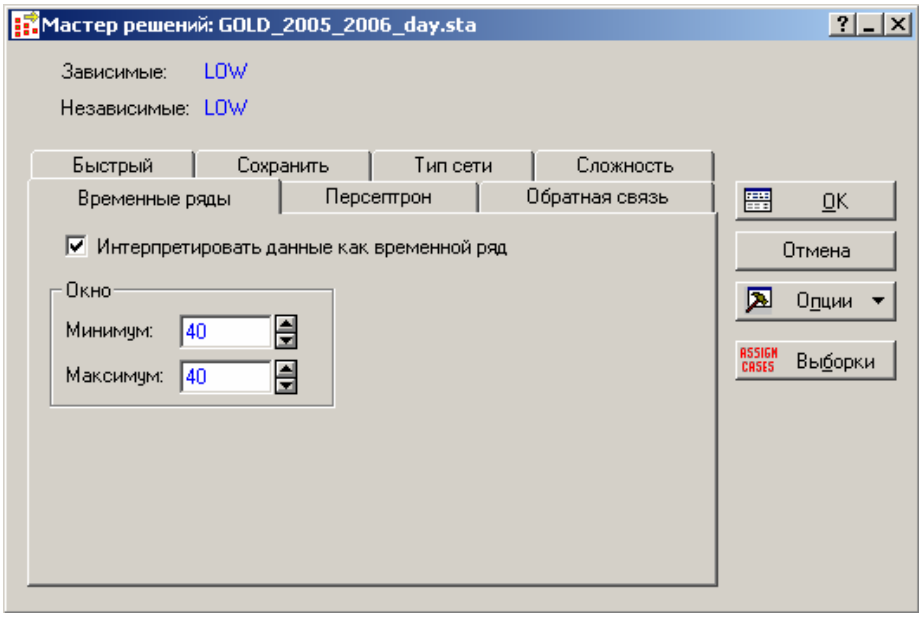


Рис 5.4

Диалог «Мастер решений»

Вкладка «Временные ряды»

Вкладка «Обратная связь» задает параметры выходной информации после процедуры выбора наилучших сетей.

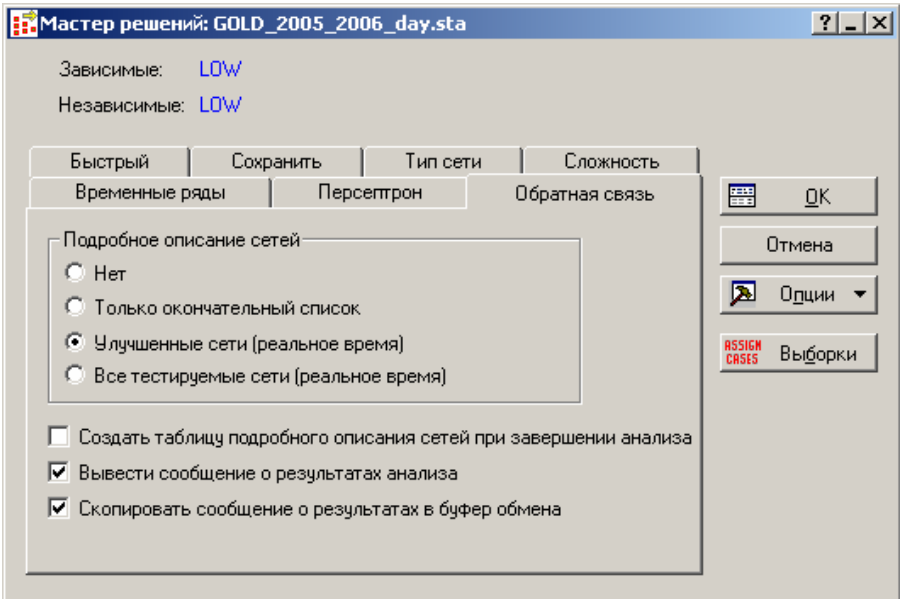


Рис 5.5

Диалог «Мастер решений»

Вкладка «Обратная связь»

Итак, после идентификации процесса построения сетей (который может длиться несколько минут) мы получим следующие результаты. Выбранные сети, как можно заметить, имеют различные конфигурации. Для отображения количественных характеристик выбранных сетей, необходимо нажать соответствующие кнопки на вкладке «Быстрый». Проанализируем полученные результаты.

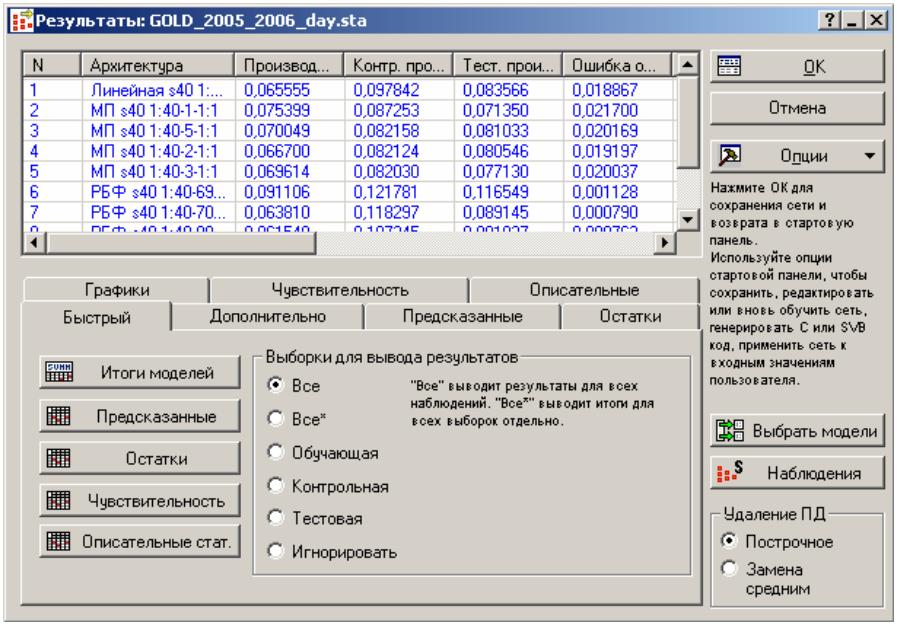


Рис 5.6

Диалог «Результаты»

Подробные результаты моделей (GOLD_2005_2006_day.sta)							
	Архитектура	**Произв. обуч.**	**Контр. Произв.**	**Тест. Произв.**	**Ошибка обуч.**	**Контрольная ошибка**	**Тестовая ошибка**
1	Линейная s40 1:40-1:1	0,065555	0,097842	0,083566	0,018867	0,028747	0,025563
2	МП s40 1:40-1-1:1	0,075399	0,087253	0,071350	0,021700	0,025572	0,021672
3	МП s40 1:40-5-1:1	0,070049	0,082158	0,081033	0,020169	0,024077	0,024578
4	МП s40 1:40-2-1:1	0,066700	0,082124	0,080546	0,019197	0,024064	0,024479
5	МП s40 1:40-3-1:1	0,069614	0,082030	0,077130	0,020037	0,024044	0,023435
6	РБФ s40 1:40-69-1:1	0,091106	0,121781	0,116549	0,001128	0,001535	0,001521
7	РБФ s40 1:40-70-1:1	0,063810	0,118297	0,089145	0,000790	0,001499	0,001177
8	РБФ s40 1:40-80-1:1	0,061549	0,107245	0,081937	0,000762	0,001354	0,001070
9	РБФ s40 1:40-97-1:1	0,067839	0,105739	0,091406	0,000840	0,001333	0,001192
10	РБФ s40 1:40-102-1:1	0,055761	0,098768	0,082171	0,000690	0,001245	0,001075

Рис. 5.7

Таблица результатов выбранных моделей

Нетрудно заметить, что производительность сетей с архитектурой Радиально Базисной Функции (РБФ) в среднем хуже производительности сетей с архитектурой Многослойного (в данном случае рассматриваются только 3-х слойные) Персептрона. Во многом это объясняется тем, что сети с архитектурой РБФ плохо экстраполируют данные (это связано с насыщением элементов скрытой структуры). Позже мы наглядно сможем в этом убедиться при визуальном анализе построенных всеми выбранными сетями прогнозов.

Также мы можем сравнить все сети по значениям различных ошибок.

	LOW.1	**LOW.2**	**LOW.3**	**LOW.4**	**LOW.5**	**LOW.6**	**LOW.7**	**LOW.10**
Среднее данных	528,0752	528,0752	528,0752	528,0752	528,0752	528,0752	528,0752	528,0752
Ст.откл. данных	82,6073	82,6073	82,6073	82,6073	82,6073	82,6073	82,6073	82,6073
Среднее ошибки	-0,3878	-0,0679	0,0421	-0,1267	-0,0381	-0,1193	-0,5540	-0,1064
Ст.откл. ошибки	6,5215	6,3485	6,2408	6,1089	6,1354	8,7150	7,1418	6,2009
Среднее абсолютной ошибки	4,5653	4,4018	4,3349	4,1999	4,2432	5,9004	4,9225	4,3125
Отношение ст.откл.	0,0789	0,0769	0,0755	0,0740	0,0743	0,1055	0,0865	0,0751
Корреляция	0,9969	0,9970	0,9971	0,9973	0,9972	0,9944	0,9963	0,9972

Рис. 5.8

Таблица ошибок регрессии исходного ряда и ряда, построенного выбранной сетью.

Значение ошибки «Отношение ст. отклонения» является наиболее пригодной для целей сравнения, потому как представляет собой число между 0 и 1, зависит лишь от величины каждой составляющей ошибки (не зависит от знака), а также довольно восприимчиво к выбросам.

Проанализировав значение ошибок, можно заметить, что для всех, выбранных программой сетей эти показатели довольно малы, что говорит об удачном подборе моделей.

На вкладке «Графики» диалога «Результаты» можно построить всевозможные диаграммы рассеяния для выбранных посредством опции «Выбрать модели» сетей.

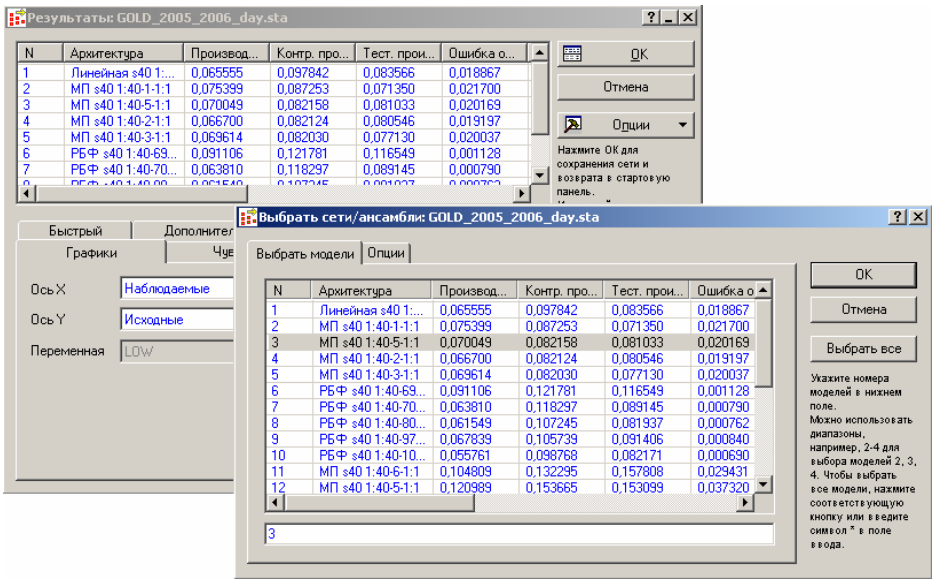


Рис. 5.9

Диалог «Результаты»

Вкладка «Графики»

Выбор моделей

Для того, чтобы оценить правдоподобность построенных моделей, построим график зависимости величины ошибки от номера наблюдения, а также график зависимости значений предсказанного ряда (ряда, построенного при помощи выбранной сети) от значений исходного.

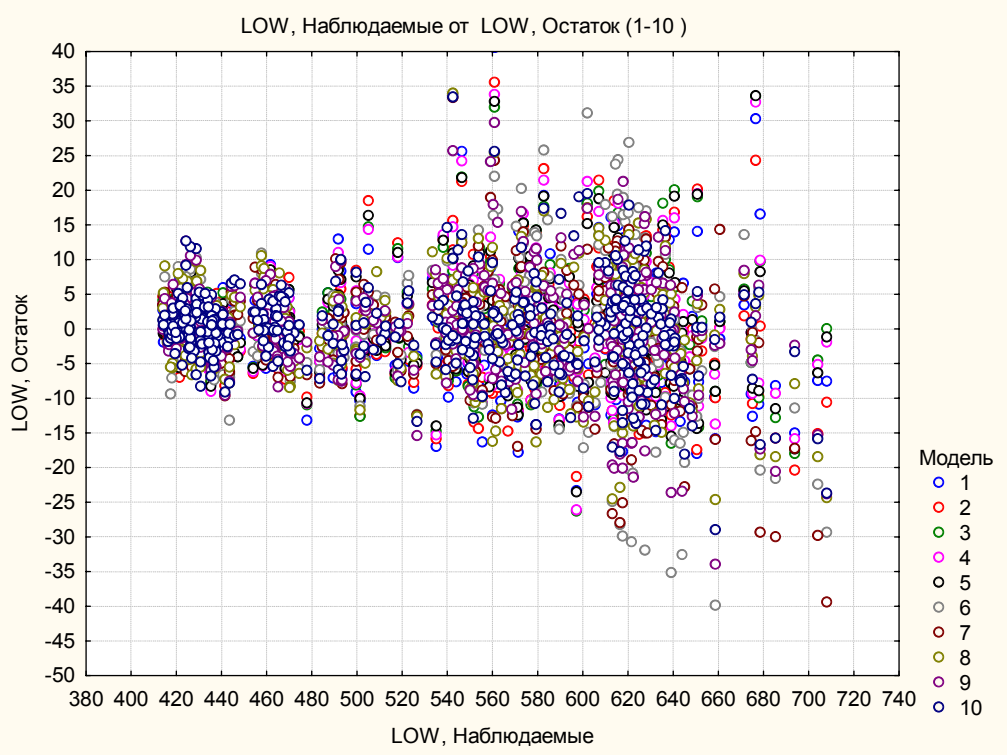


Рис. 5.10

Диаграмма рассеяния ошибок предсказания от номера наблюдения.

Увеличение размаха ошибок с номером наблюдения говорит о нарушениях предположения о том, что дисперсия ошибки является случайной величиной, но поскольку это увеличение незначительно (и происходит в основном за счет выбросов) считаем, что дисперсия постоянна.

Помимо диаграммы рассеяния для анализа остатков применим также метод построения гистограмм частот. Построим эти графики для первых пяти моделей.

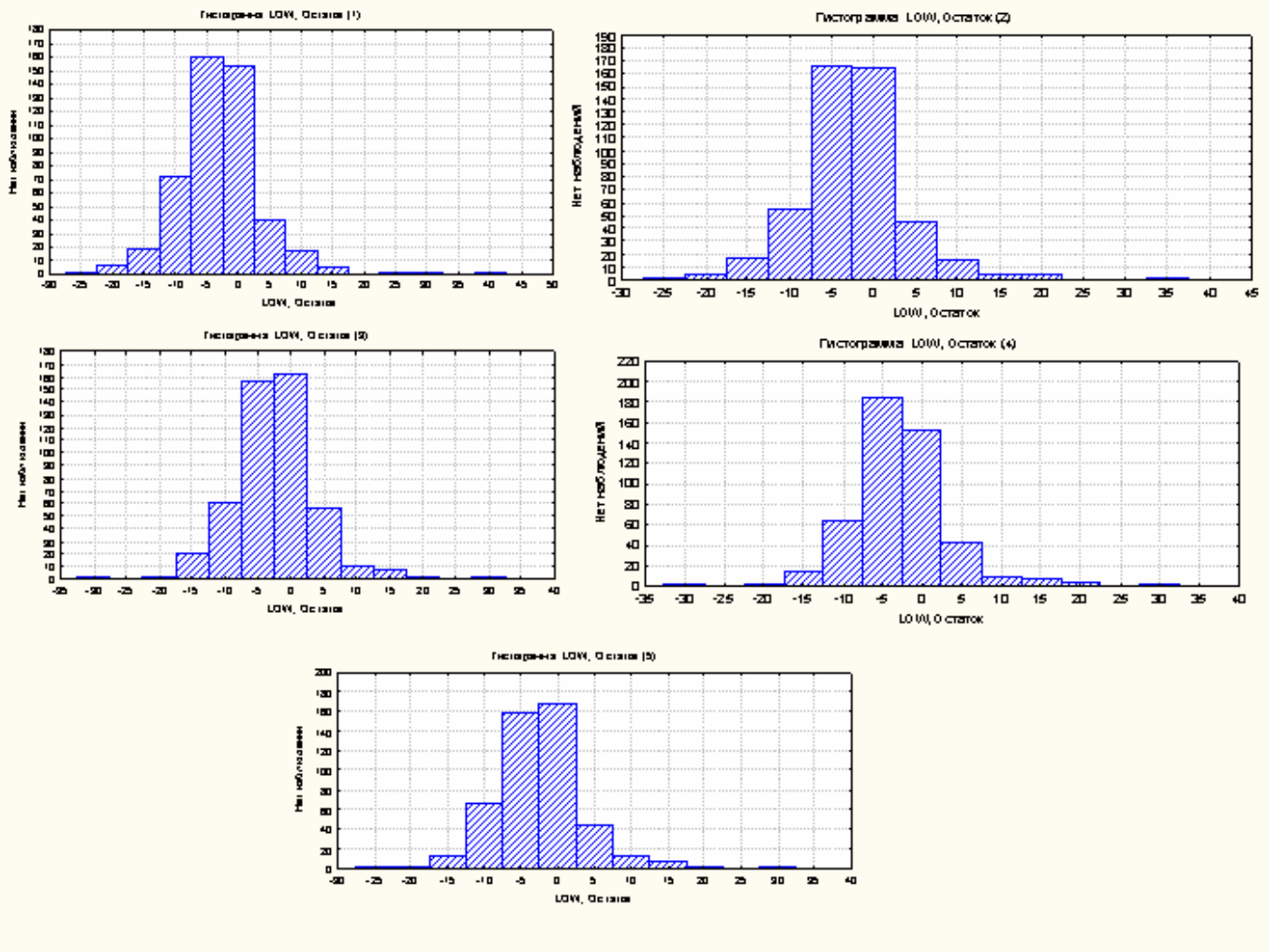


Рис. 5 .11

Гистограммы частот остатков первых пяти моделей нейронных сетей.

Нетрудно заметить, что 1,4 и 5 гистограммы являются весьма ассиметричными, что противоречит стандартным предположениям о нормальности остатков. Из всего выше сказанного об ошибках, можно заключить, что наиболее подходящая модель нейронной сети для данного ряда есть модель №3.

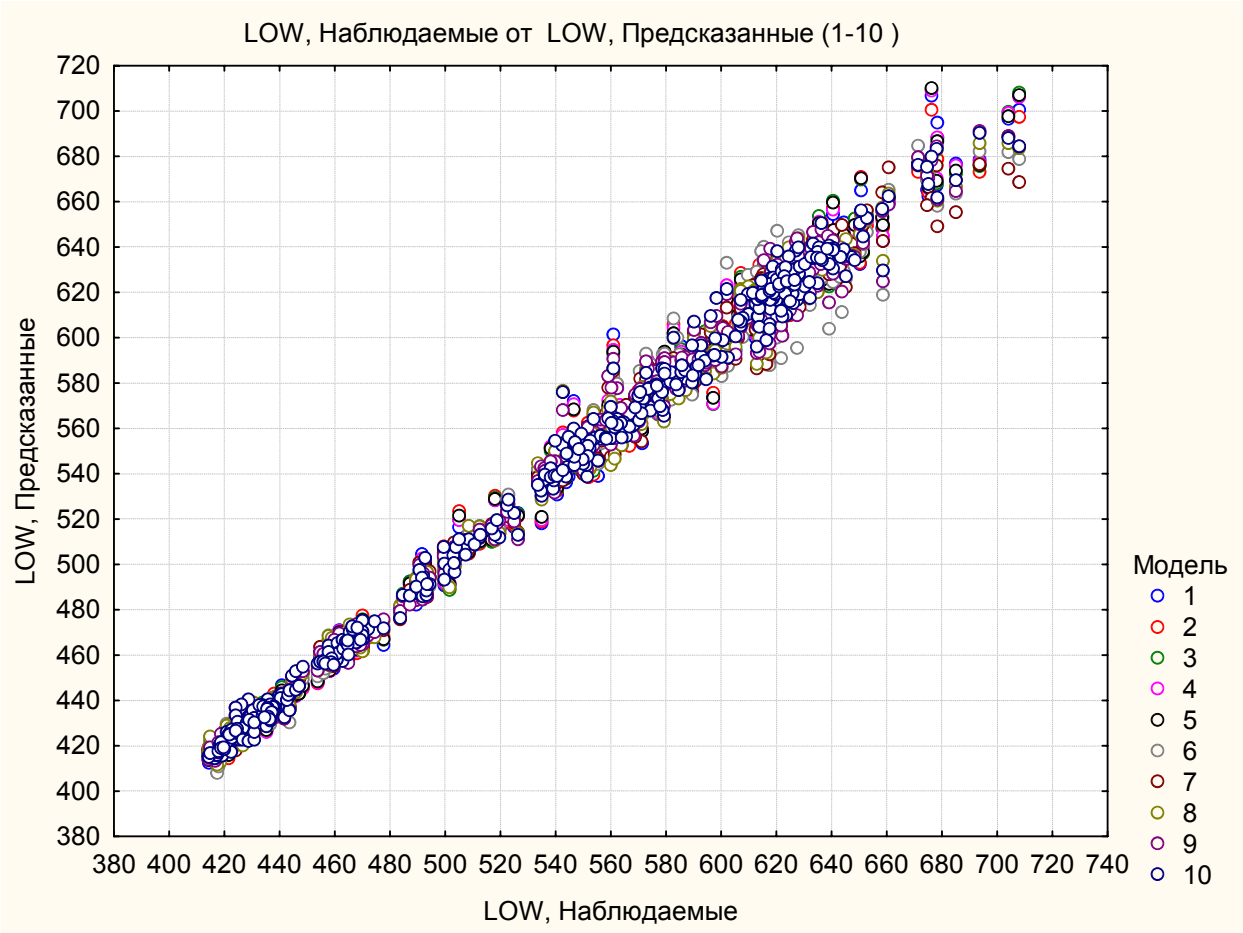


Рис. 5.11

Диаграмма рассеяния предсказанных значений в зависимости от значений исходного ряда.

Полученная диаграмма говорит об очень хорошем приближении исходного ряда одной из десяти рассмотренных сетей. Поскольку мы отметили сеть №3, как наиболее подходящую, целесообразным будет приведение её отдельной диаграммы рассеяния предсказанных значений.

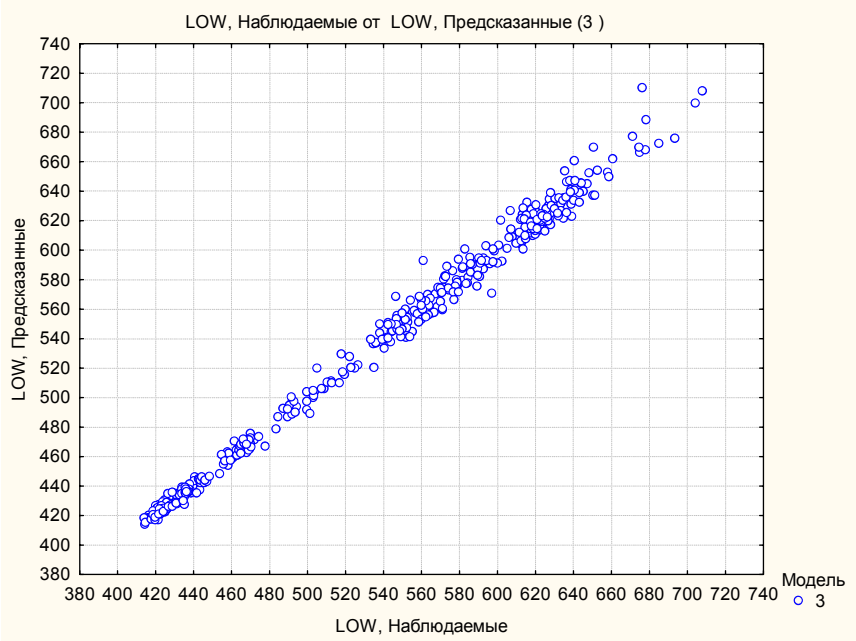
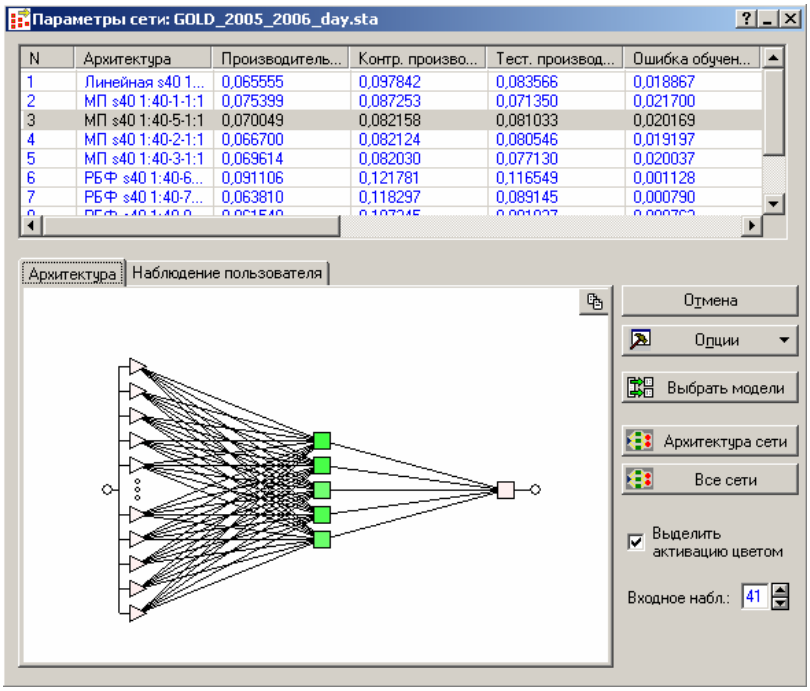


Рис. 5.12

Диаграмма рассеяния предсказанных значений моделью №3 в зависимости от значений исходного ряда.

Еще одной характеристикой качества построенной сети можно назвать следующий факт. На каждом скрытом слое, уровень активации в определенный момент времени должен быть одинаков. Знак активации роли не играет.

На вкладке «Дополнительно» нажмем кнопку «Архитектура сети» для отображения этой архитектуры.



N	Архитектура	Производитель...	Контр. произво...	Тест. производ...	Ошибка обучен...
1	Линейная s40 1...	0,065555	0,097842	0,083566	0,018867
2	МП s40 1:40-1-1:1	0,075399	0,087253	0,071350	0,021700
3	МП s40 1:40-5-1:1	0,070049	0,082158	0,081033	0,020169
4	МП s40 1:40-2-1:1	0,066700	0,082124	0,080546	0,019197
5	МП s40 1:40-3-1:1	0,069614	0,082030	0,077130	0,020037
6	РБФ s40 1:40-6...	0,091106	0,121781	0,116549	0,001128
7	РБФ s40 1:40-7...	0,063810	0,118297	0,089145	0,000790

Рис. 5.13

Вкладка «Архитектура сети»

Данная опция позволяет наглядно оценить уровни активации элементов сети в любой момент времени. Как видно, сеть №3 довольно хорошо отвечает этому требованию.

Построим 5 прогнозов на 15 наблюдений вперед. Следует напомнить, что прогнозирование производится по 40 последним наблюдением. Выбор этого значения был осуществлен ранее как величина окна прогноза.

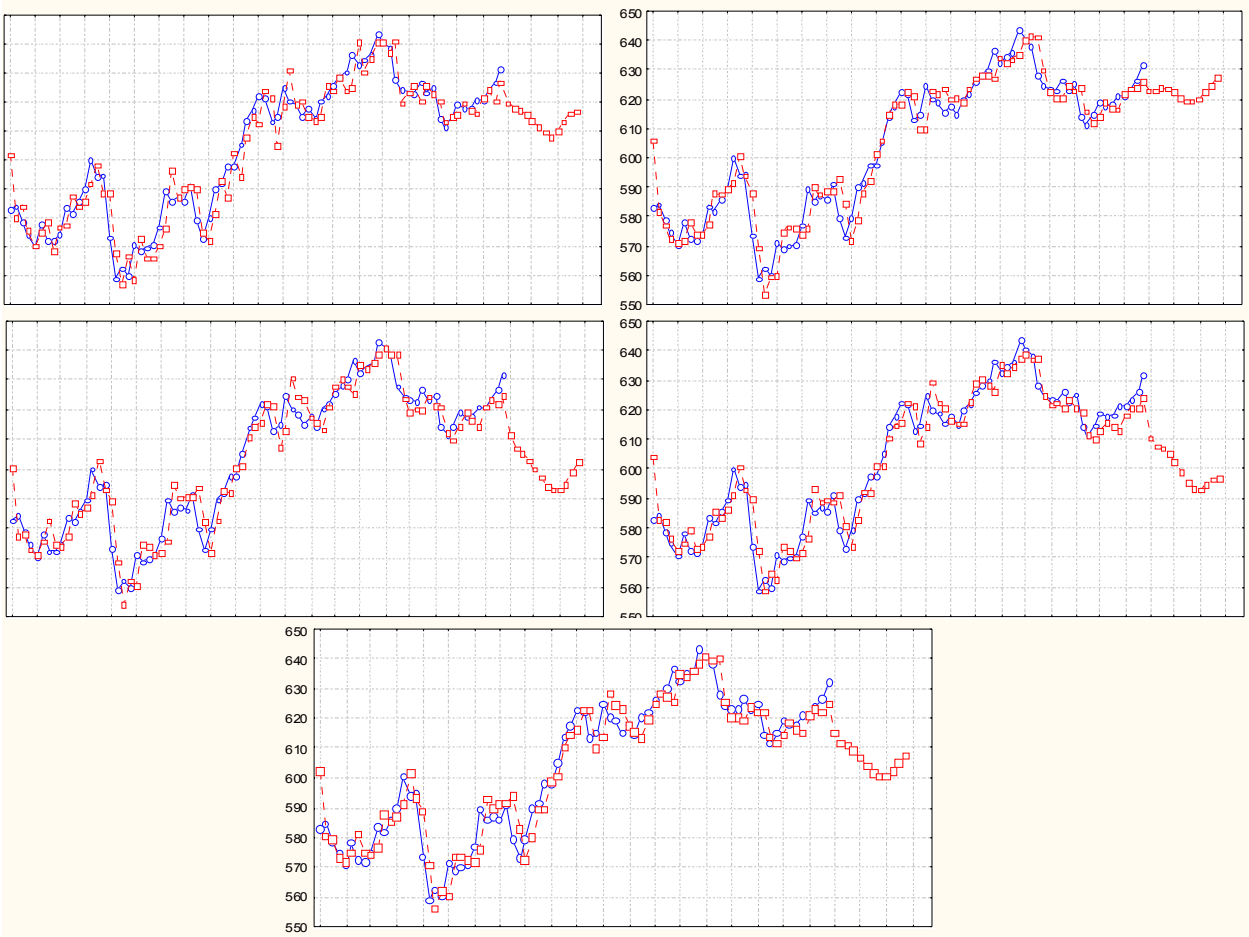


Рис. 5.14

Прогнозы моделей №№ 1-5

Синим – наблюдаемое значение

Красным – прогноз сети

Прогнозы в зависимости от моделей упорядочены слева - направо, сверху - вниз.

Поскольку мы остановились на модели №3, и будем считать результат построения прогноза при помощи этой модели за финальный прогноз. Действительно, визуально можно отметить, что полученный при помощи этой модели ряд довольно хорошо восстанавливает исходный (что впрочем, можно заключить и для результатов других моделей). Во всех моделях предсказывается качественно один и тот же результат - сначала цена падает, а затем столь же интенсивно растёт. Поскольку качественно такой же результат предсказывало нам и экспоненциальное сглаживание, то можно считать, что прогноз верен.

Ниже представлена проекция ряда на 15 наблюдений вперед

Проекция временного ряда, LOW (GOLD_2005_2006_day.sta)	
	LOW.3
1 января 2006 г.	625,069347
2 января 2006 г.	622,686824
3 января 2006 г.	617,987886
4 января 2006 г.	611,828524
5 января 2006 г.	607,272057
8 января 2006 г.	605,351248
9 января 2006 г.	602,996582
10 января 2006 г.	600,152152
11 января 2006 г.	597,07978
12 января 2006 г.	594,410874
15 января 2006 г.	592,876187
16 января 2006 г.	592,837292
17 января 2006 г.	595,069185
18 января 2006 г.	598,78757
19 января 2006 г.	602,166517

Рис. 5.15

Проекция ряда, построенного моделью №3, на 15 значений вперед

Как было показано выше, ошибки этой модели наиболее полно удовлетворяют стандартным предположениям (н.о.р.), а ошибка на тестовом множестве (которую можно считать критерием качества сети) порядка 0.025. Это очень хороший результат.

ВЫВОД:

Как и предполагалось, нейронные сети дали более надежный результат. Во многом это обусловлено сложностью и нелинейностью структуры данного ряда, потому как в отличие от рассматриваемых нейронных сетей, классические методы рассчитаны на применение к рядам с более заметными и очевидными структурными закономерностями. Но даже, несмотря на все видимые положительные качества нейронных сетей не стоит считать их некоей «панацеей», например, в задачах прогнозирования.

Во-первых, нейронные сети являются «черным ящиком», который не позволяет в явном виде определить вид зависимостей между членами ряда. Таким образом, конкретную нейронную сеть можно «научить» строить прогноз лишь на строго фиксированное количество шагов вперед (которое мы указываем в спецификации этой сети), следовательно, имеет место сильная зависимость от вида задачи.

Во-вторых, при наличии явной линейности, простоты структуры в задаче, способность нейронных сетей к обобщению оказывается более слабой по отношению к классическим методам. Объясняется это как раз нелинейностью сетей по своей сути (для того, чтобы убедиться в этом, необходимо ознакомиться с основными принципами работы), что так полезно в некоторых случаях. Таким образом, в некоторых случаях нейронные сети уступают по производительности классическим методам. Ситуация во многом напоминает «стрельбу из пушки по воробьям».

В общем случае для достижения наилучшего результата необходимо попробовать всевозможные методы, потому что на первый взгляд определить тип задачи и вид зависимости в ней порой довольно сложно.

6. Применение метода многомерного прогнозирования

Качественно отличающимся методом прогнозирования будет приём многомерного прогнозирования. Суть метода заключается в том, что вместо того, чтобы строить непосредственный прогноз самого ряда, мы будем строить прогноз факторов, оказывающих на значения рассматриваемого ряда наибольшее влияние. Для нахождения этих самых факторов используется следующий нехитрый алгоритм:

1) На основе каких-либо эвристических данных, либо научно обоснованных знаний полученных ранее, делаются предположения о множестве факторов, которые теоретически могут оказывать влияние на исследуемый ряд. Размер этого первоначального множества может быть весьма большим, но основное требование состоит в том, чтобы оно максимально полно охватывало круг гипотетически возможных факторов влияния.

2) Ввиду необходимости снижения размерности задачи и повышения обобщающих способностей (учет влияния очень малозначительных факторов является причиной сильной зависимости от случайной составляющей, что в свою очередь может сильно исказить картину ряда) применяется процедура отсеивания признаков. Существуют различные методы для достижения этой цели. В нижеследующем примере мы воспользуемся методом построением многомерных моделей регрессии и процедурой генетического алгоритма в построении нейронных сетей.

3) После того, как выбрано более узкое подмножество наиболее влиятельных факторов, для каждого из них строится собственный прогноз. При этом применимы все те методы прогнозирования, которые рассматривались выше. Используя необходимые значения спрогнозированных рядов факторов, строится соответствующее значение искомого ряда.

Такая процедура построения прогнозов наиболее точно описывает поведение ряда, потому как учитывает еще и внешние изменения среды, остающиеся за рамками стандартных методов прогноза. Именно такие внешние изменения зачастую бывают единственным разумным объяснением каких-либо вариаций исследуемого ряда.

Delta high (t+1) = HIGH (t) – (OPEN (t) + CLOSE (t))/2 –разница между наибольшей и средней ценой за прошедший день;

Delta low (t+1) = (OPEN (t) + CLOSE (t))/2 – LOW (t) –разница между наименьшей и средней ценой за прошедший день;

Delta (t+1) = HIGH (t) – LOW (t) –разница между наибольшей и наименьшей ценой за прошедший день;

Day delta (t+1) = ClOSE (t) – OPEN (t) – разница между ценой закрытия и открытия за прошедший день.

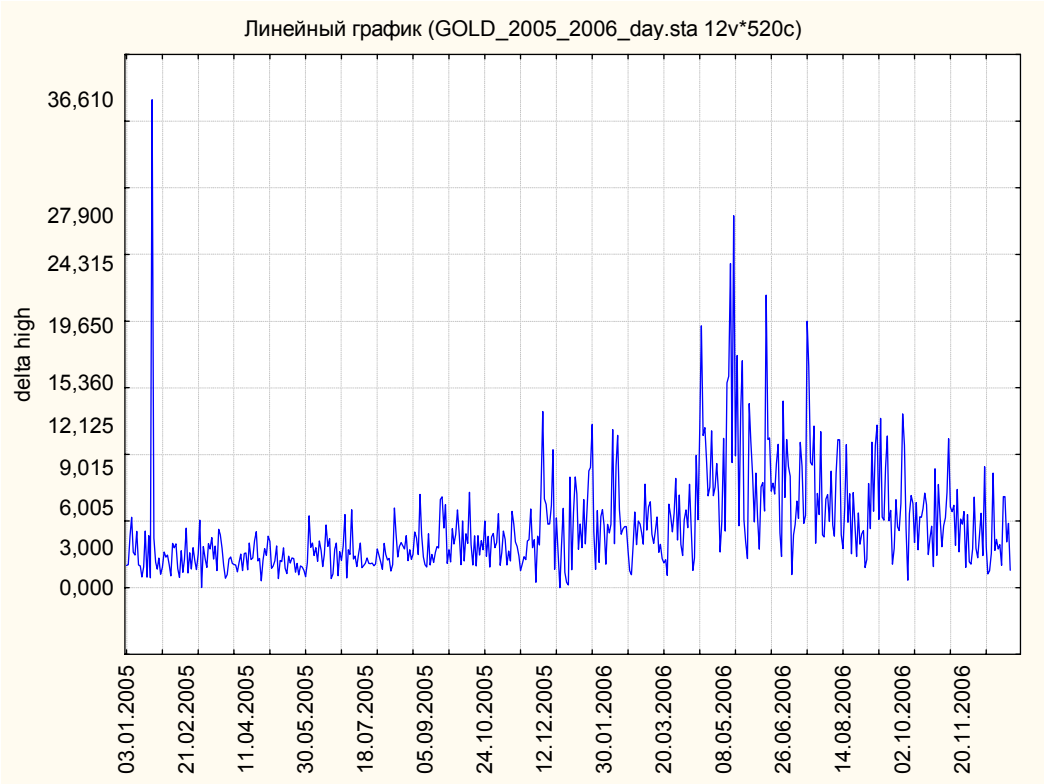


Рис.6.1

График переменной Delta high

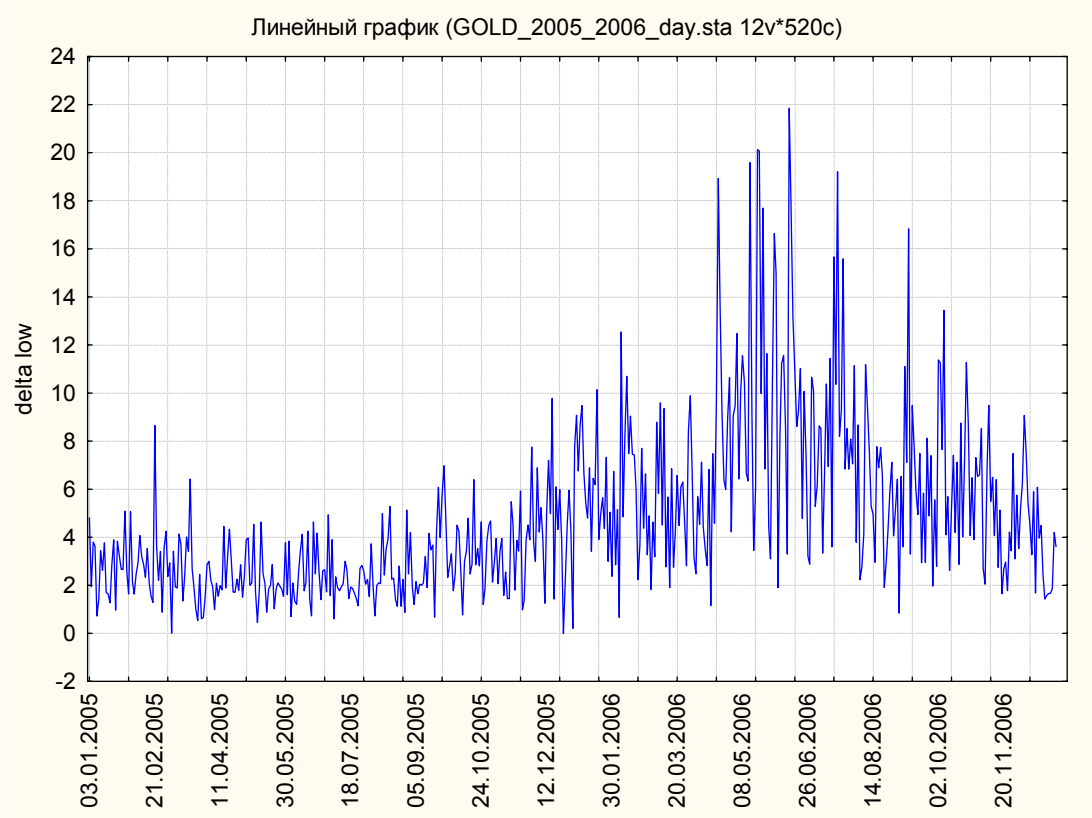


Рис.6.2

График переменной Delta low

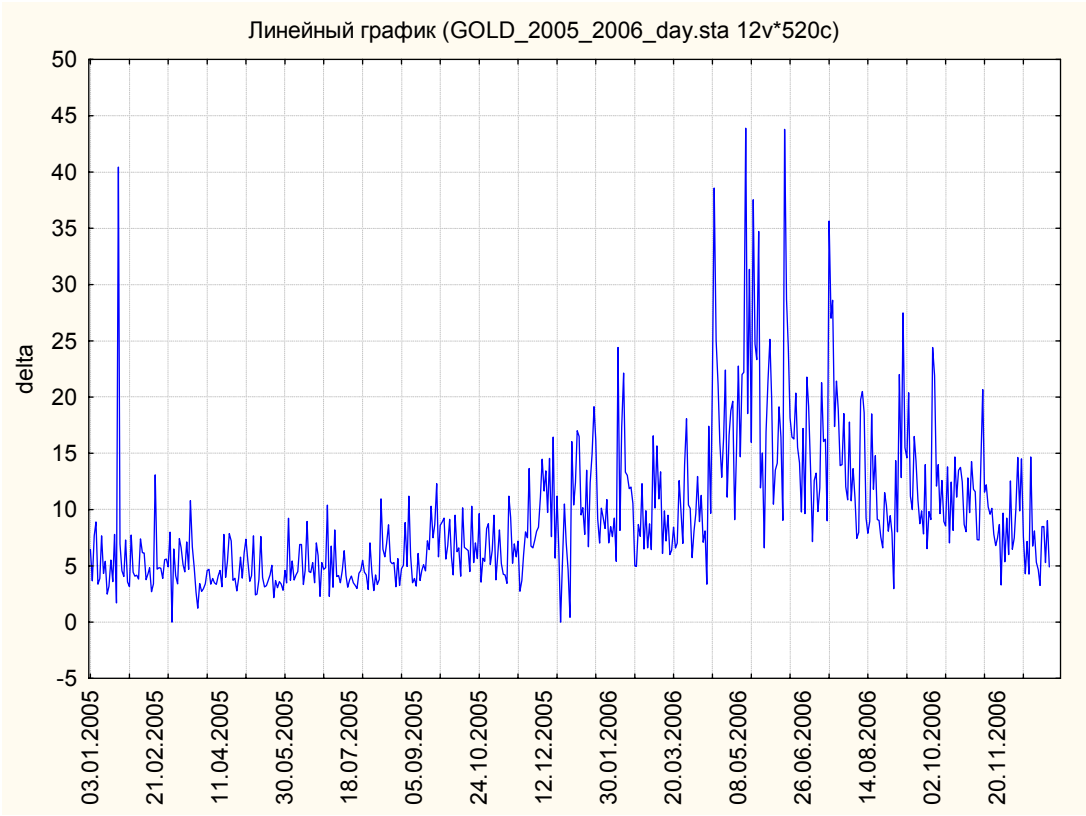


Рис.6.3

График переменной Delta

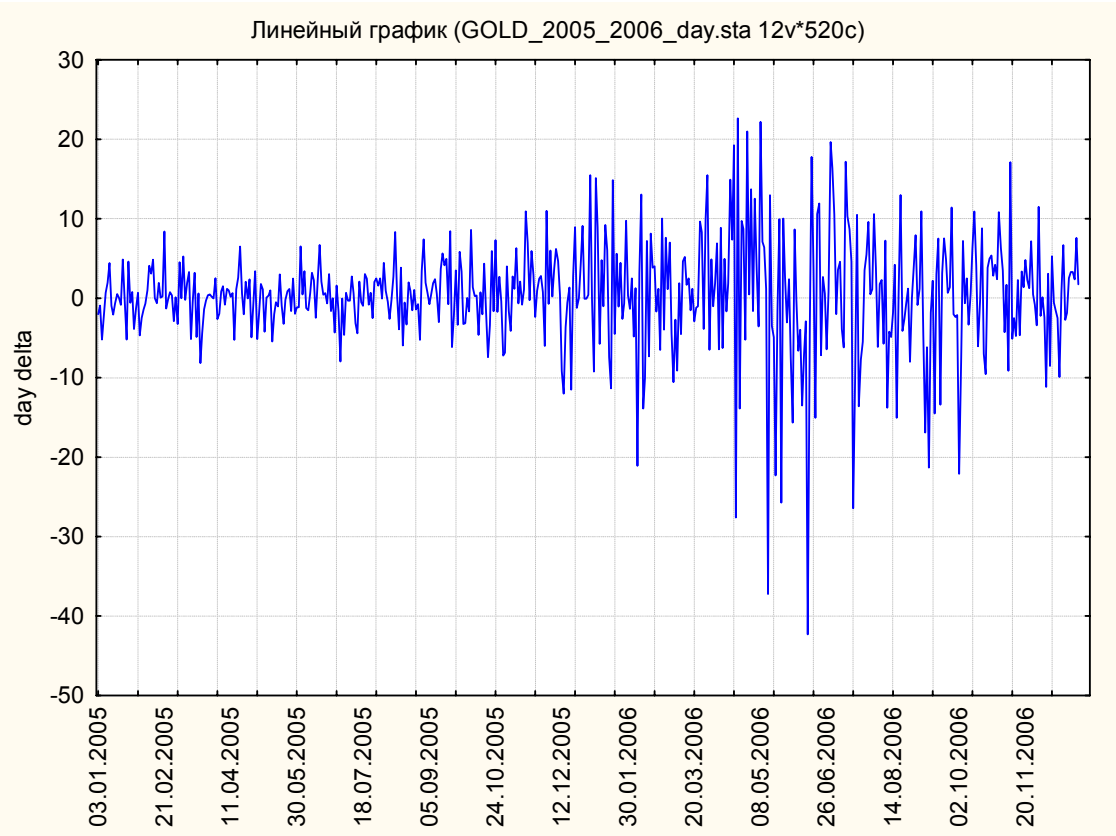


Рис.6.4

График переменной Day delta

Для того чтобы осуществить процедуру определения множества факторов влияния, воспользуемся методом множественной нелинейной регрессии. Для запуска соответствующего анализа в меню «Анализ» необходимо выбрать «Углубленные методы анализа», «Множественная нелинейная регрессия». После этого появится диалог:

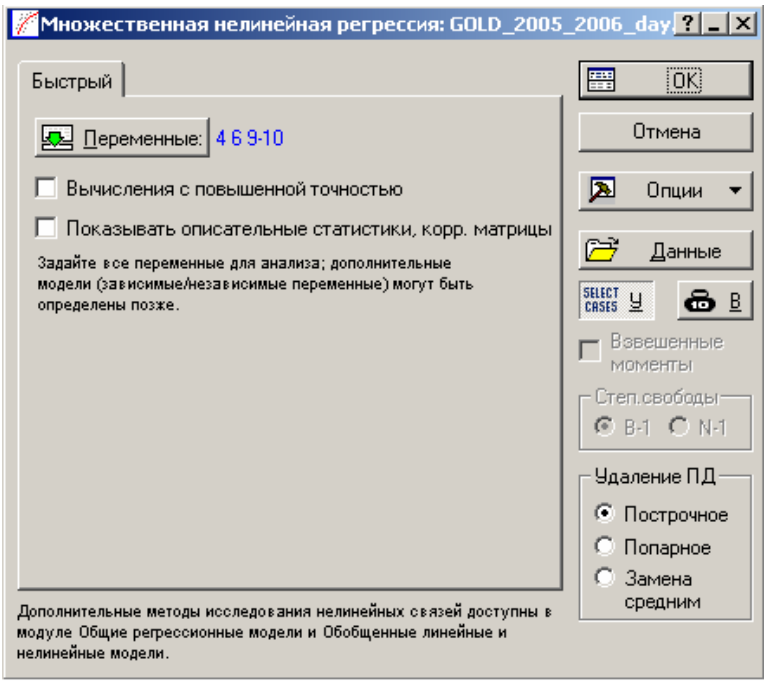


Рис. 6.5

Диалог «Множественная нелинейная регрессия»

На единственной вкладке настоящего диалога пользователю предлагается выбрать используемые в анализе переменные, а также определить некоторые опции. В нашем случае оставим все опции по умолчанию неактивным и нажмем «ОК».

Следует отметить, что суть метода нелинейной регрессии заключается в том, что помимо выбранных переменных в рассмотрение добавляются новые, полученные из исходных при помощи различного рода арифметических преобразований.

Полученное таким образом множество новых факторов используется уже в процедуре построения уравнения стандартной линейной регрессии.
Так, например, для выявления зависимости между переменной «А» и различными степенями переменной «Б» (это может быть задача нахождения коэффициентов аппроксимации какой-либо действительной функции полиномом) необходимо породить переменные «Б^2». «Б^3 », и т.д., а затем построить уравнение линейной регрессии, связывающее переменные «А», «Б^2», «Б^3», ...

Итак, в следующем диалоге предлагается выбрать те преобразования, которые будут применяться к исходным переменным для порождения новых.

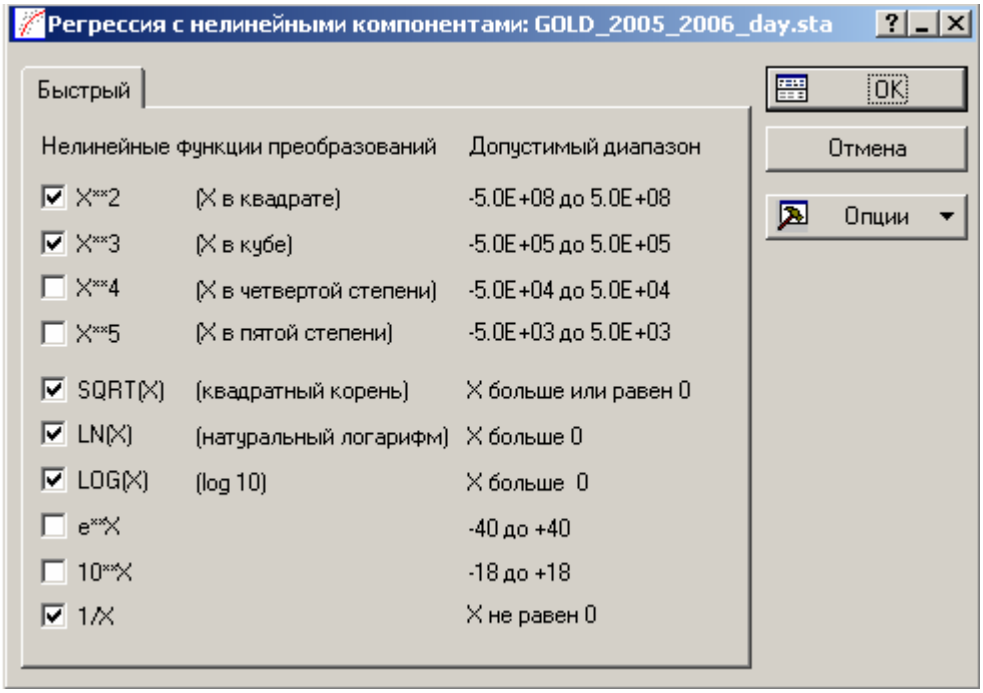


Рис. 6.6

Диалог «Регрессия с нелинейными компонентами»

Выбрав необходимые преобразования (в данном примере выбор видов преобразований проводился, следуя принципу максимизации мощности множества гипотетически возможных факторов влияния) и нажав кнопку «ОК» мы перейдем к следующему диалогу, в котором необходимо будет определить параметры строящихся моделей.

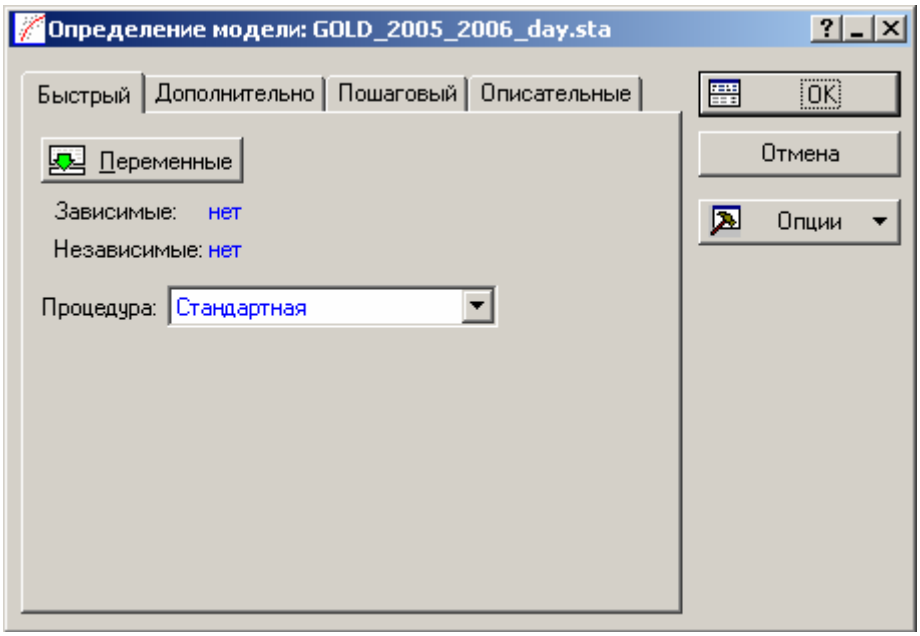


Рис. 6.7

Диалог «Определение модели»

В данном диалоге уже необходимо разделить переменные на Зависимые и Независимые, а также сделаем выбор процедуры построения уравнения регрессии в пользу Пошаговой процедуры с включением. Это значит, что на каждом шаге в модель будет включена еще одна переменная и оценена ошибка. Таким образом, будет выбрана модель с наименьшей ошибкой.

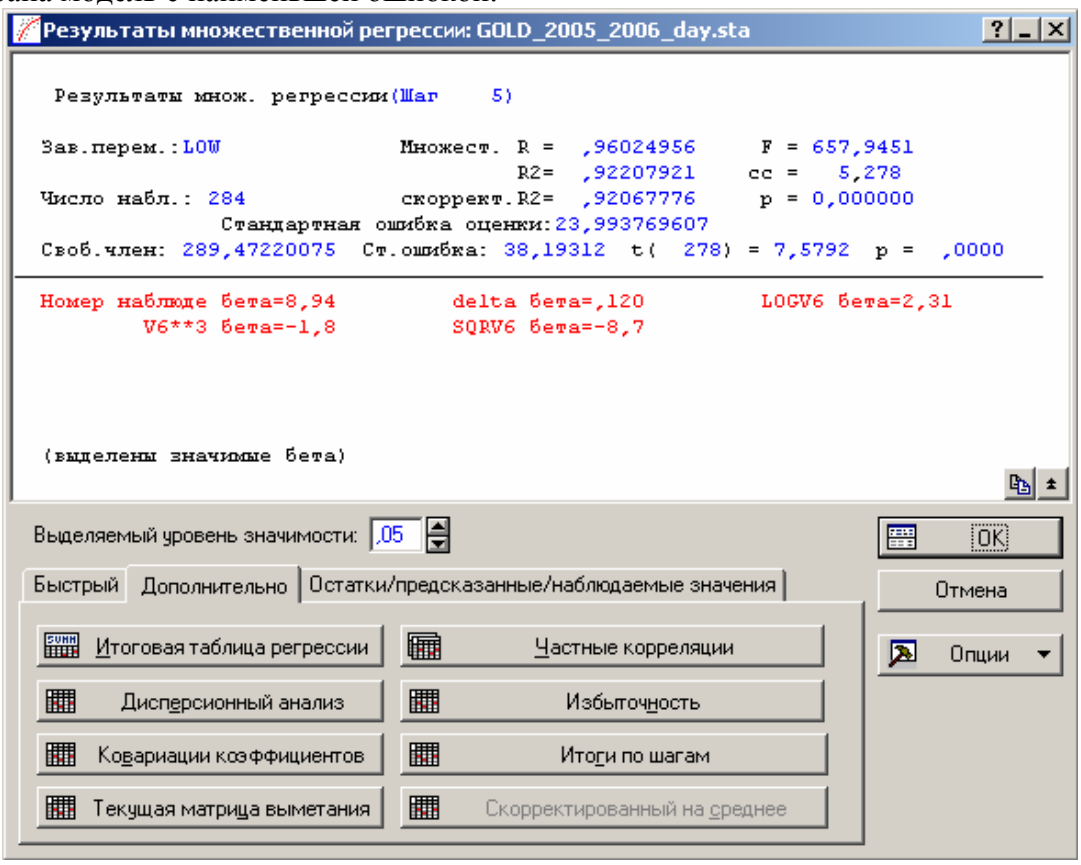


Рис. 6. 8

Диалог «Результаты множественной регрессии»

Итоги регрессии для зависимой переменной: LOW (GOLD_2005_2006_day.sta) R= ,96024956 R2= ,92207921 Скорректир. R2= ,92067776 F(5,278)=657,95 p

	БЕТА	Стд.Ош.	B	Стд.Ош.	t(278)	p-уров.
Св.член			289,472	38,19312	7,5792	0,000000
Номер наблюдения	8,94306	0,845723	5,102	0,48251	10,5745	0,000000
delta	0,11974	0,023753	1,957	0,38830	5,0409	0,000001
LOGV6	2,31396	0,345692	516,118	77,10519	6,6937	0,000000
V63**	-1,77204	0,161921	-0,000	0,00000	-10,9439	0,000000
SQRV6	-8,71900	1,025295	-142,095	16,70937	-8,5039	0,000000

Нажав «ОК» мы перейдем к диалогу, в котором дана возможность построить всевозможные графики, провести анализ остатков.

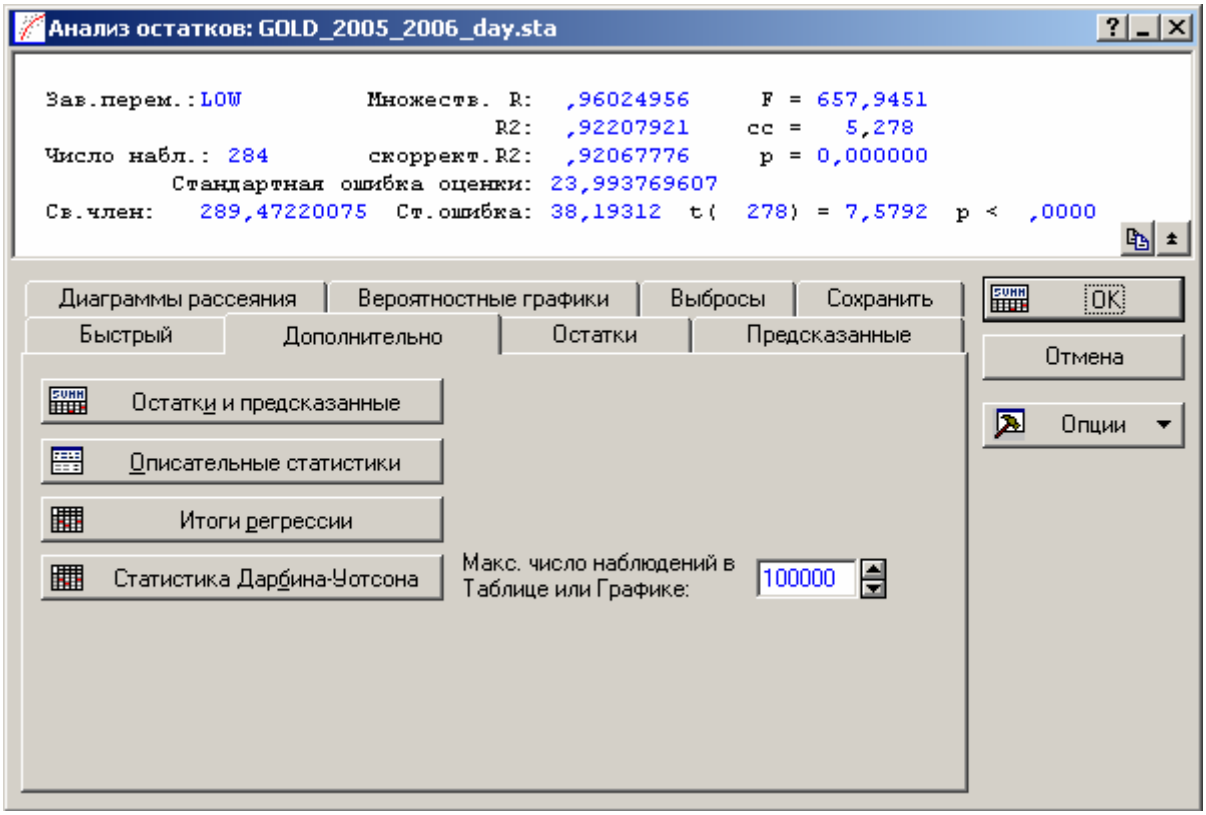


Рис. 6.9

Диалог «Анализ остатков»

Для того чтобы убедиться в том, что построенная нами модель аппроксимирует переменную, построим график зависимости Предсказанных значений от Наблюдаемых.

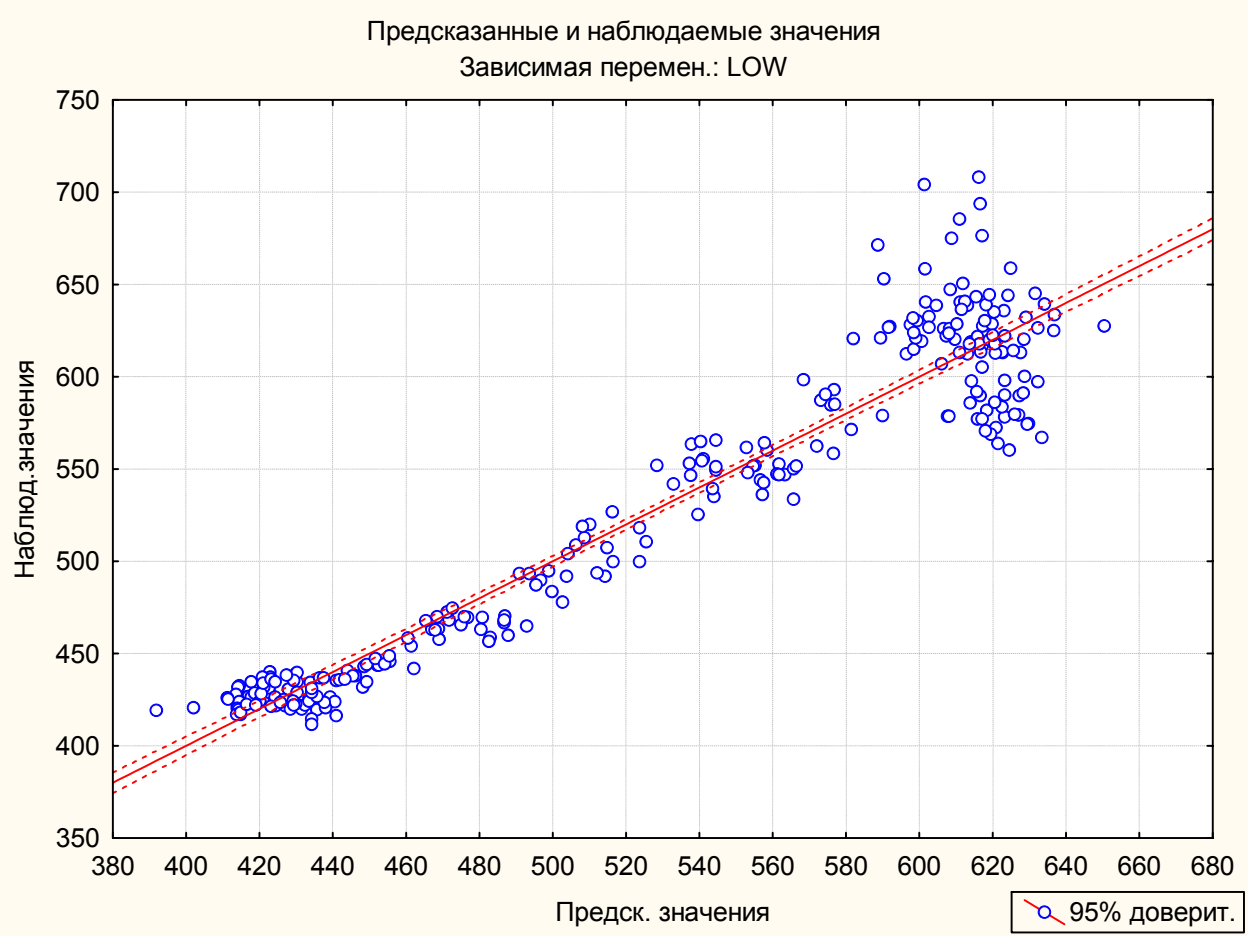


Рис.6.10

График зависимости Предсказанных значений от Наблюдаемых

Как показал результат множественной регрессии, значимыми являются лишь переменная «номер наблюдения» и «Delta», точнее либо они, либо их производные. Поэтому остается заняться прогнозом переменной «Delta». Будем строить прогноз данной переменной при помощи инструмента нейронных сетей.

Специфика процедуры построения и обучения нейронных сетей такая же такая, как и в примерах выше, поэтому не будем подробно останавливаться на этом. Отметим лишь, что будем искать наилучшую сеть из 500 различных сетей архитектур МП (Многослойный персептрон). РБФ (радиально-базисная функция) и Линейной архитектуры.

Ниже представлены архитектуры и параметры пяти наилучших сетей, а также таблица описательных статистик:

Подробные результаты моделей (GOLD_2005_2006_day.sta)

	Архитектура	Производ. обуч.	Контр. Производ.	Тест. Производ.	Ошибка обуч.	Контр. ошибка	Тест. ошибка
1	РБФ s3 1:3-16-1:1	0,696662	0,629469	0,860176	0,112073	0,108567	0,155799
2	Линейная s4 1:4-1:1	0,715741	0,692541	0,810769	0,100956	0,104398	0,128832
3	Линейная s3 1:3-1:1	0,734364	0,693726	0,821989	0,103583	0,104318	0,130472
4	МП s4 1:4-1-1:1	0,732245	0,681055	0,809760	0,103298	0,102147	0,128463

5	МП s4 1:4-8-1:1	0,714178	0,674200	0,799730	0,102228	0,100758	0,126878

	delta.1	delta.2	delta.3	delta.4	delta.5
Среднее данных	9,455891	9,455891	9,455891	9,455891	9,455891
Ст.откл. данных	6,483770	6,483770	6,483770	6,483770	6,483770
Среднее ошибки	-0,268375	-0,265732	-0,223451	-0,141418	0,486474
Ст.откл. ошибки	4,745718	4,789278	4,865555	4,814863	4,735207
Среднее абсолютной ошибки	3,155369	3,094925	3,193773	3,162946	3,261931
Отношение ст.откл.	0,731938	0,738656	0,750421	0,742602	0,730317
Корреляция	0,681999	0,674932	0,662140	0,669745	0,683374

Нетрудно заметить, что построенные модели не отличаются хорошими предсказательными качествами, а поэтому мы не будем строить прогнозы с их помощью.

Примечательным является тот факт, что наилучшими оказались сети с линейной архитектурой и с архитектурой МП с небольшим количеством нейронов на внутреннем слое. Это подталкивает нас к тому, чтобы попробовать метод построения АРПСС модели.

Для поиска параметров модели применим макрос «АвтоАРПСС» автоматического поиска параметров АРПСС моделей. Для определения сезонного лага воспользуемся методом анализа периодограмм.

Для переменной delta получаем следующую таблицу наибольших значений периодограмм.

Спектр. анализ: DELTA : = v3-v4 (GOLD_2005_2006_day.sta) Число набл.: 518 Наибольшие значения периодограммы

	Частота	Период	Косинус	Синус	Периодограмма	Плотн.
1	0,001931	518,0000	-1,04769	-2,34976	1714,333	1204,155
2	0,003861	259,0000	-1,51389	1,78506	1418,883	1289,982
3	0,005792	172,6667	0,58456	1,87029	994,483	885,342
6	0,011583	86,3333	0,37557	1,58679	688,673	413,810
5	0,009653	103,6000	-1,19445	0,34116	399,661	403,887
27	0,052124	19,1852	0,79407	-0,85457	352,455	219,159
24	0,046332	21,5833	-1,14952	0,10146	344,908	247,493

31	0,059846	16,7097	-0,05150	0,97763	248,228	181,045
32	0,061776	16,1875	0,94366	0,21739	242,876	175,888
9	0,017375	57,5556	0,53996	0,77795	232,262	150,878

Как видно, сильных всплесков на каких-либо периодах кроме длины всего ряда(или его делителей нет). Построим периодограмму для ряда с взятием одной разности с лагом 1.

Спектр. анализ: DELTA : = v3-v4; D(-1) (GOLD_2005_2006_day.sta) Число набл.: 518 Наибольшие значения периодограммы

	Частота	**Период**	**Косинус**	**Синус**	**Периодограмма**	**Плотн.**
220	0,424710	2,354545	-0,53157	1,434677	606,2836	342,3465
242	0,467181	2,140496	0,48716	1,371944	548,9633	401,1123
249	0,480695	2,080321	1,36722	0,177731	492,3276	348,8342
258	0,498069	2,007752	-1,16348	-0,441367	401,0615	232,9197
243	0,469112	2,131687	1,19153	0,345636	398,6571	339,7677
247	0,476834	2,097166	1,12276	0,108758	329,5548	227,8717
226	0,436293	2,292035	-1,08521	-0,259311	322,4344	213,4271
125	0,241313	4,144000	-1,03359	0,318169	302,9123	181,0887
231	0,445946	2,242424	0,62871	-0,879087	302,5297	191,6068
253	0,488417	2,047431	0,59217	0,903232	302,1205	185,2283

Наибольшее значение на 2.35. Заметим, что из вида коррелограммы тоже можно сделать заключение о том, что наблюдается сезонность со значением периода приблизительно равного трем.

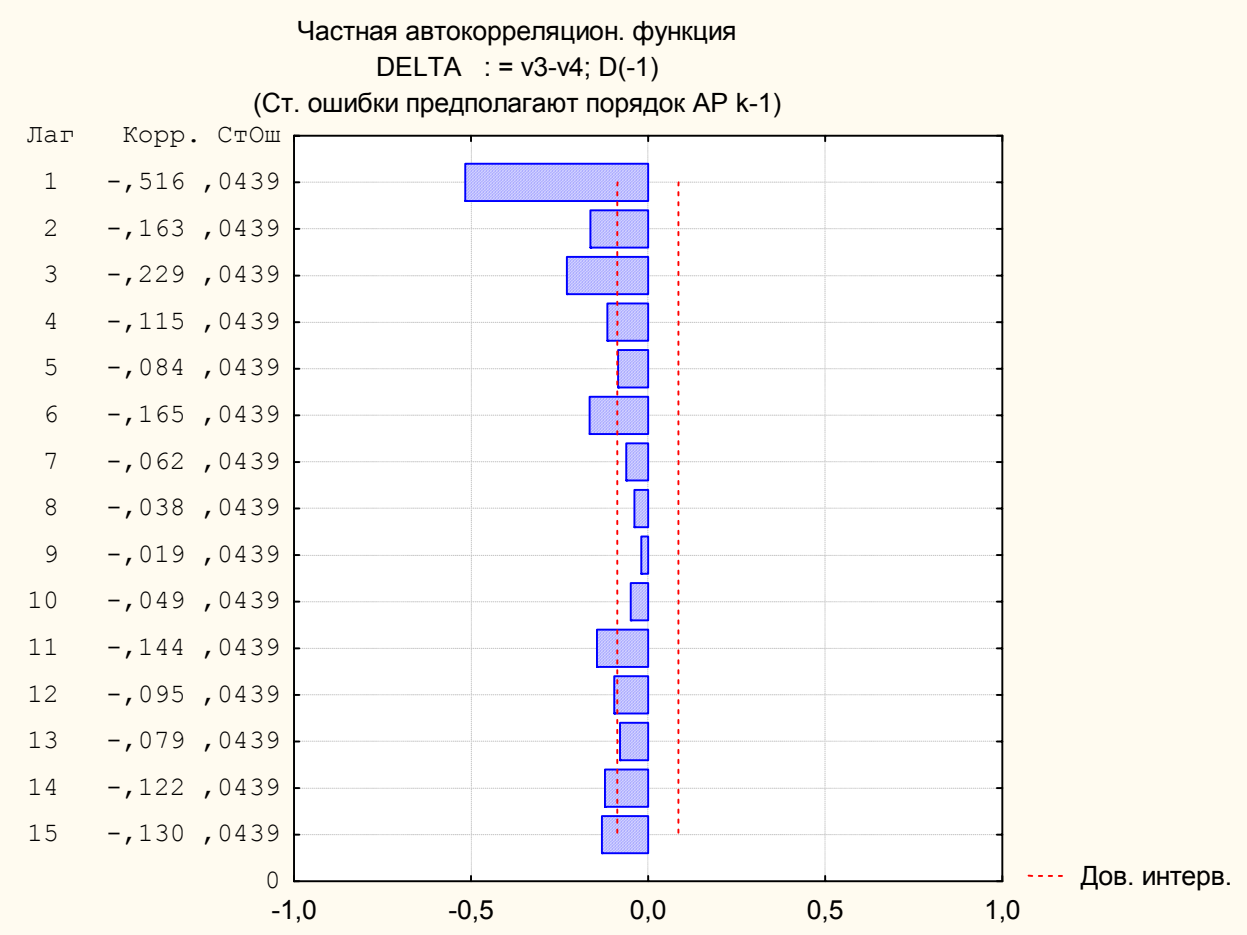


Рис. 6.11

Автокоррелограмма переменной Delta с одной разностью с лагом 1

Поскольку наибольший всплеск на нецелом значении периода, мы рассмотрим период сезонности равный 4 и 5 - т.е. длине рабочей недели. Такой выбор оправдывается еще и с точки зрения обычной логики.

Мы выберем следующие модели, потому что, во-первых, они различаются наличием параметров несезонного АРПСС, а во-вторых, они имеют наименьшую ошибку в этих классах моделей с лагом 5: АРПСС(0,1.0)(1,0,2), АРПСС(0,1,1)(2,2,1).

АРПСС(0,1,0)(1,0,2)

Оценки параметров оказались статистически незначимыми, на что указывает большое значение p-уровня.

Исход.:DELTA : = v3-v4 (GOLD_2005_2006_day.sta) Преобразования: D(1)
Модель(0,1,0)(1,0,2) Сезонный лаг: 12 MS Остаток= 35,336

	Парам.	Асимпт.	Асимпт.	p	Нижняя	Верхняя
Ps(1)	-0,070475	0,556344	-0,12667	0,899248	-1,16346	1,022509
Qs(1)	-0,110313	0,554655	-0,19889	0,842430	-1,19998	0,979351
Qs(2)	-0,055933	0,051987	-1,07592	0,282467	-0,15807	0,046199

АРПСС(0,1,1)(2,2,1)

В данном случае ситуация гораздо лучше.

Исход.:DELTA : = v3-v4 (GOLD_2005_2006_day.sta) Преобразования: D(1),2*D(5)
Модель(0,1,1)(2,2,1) Сезонный лаг: 5 MS Остаток= 43,171

	Парам.	Асимпт.	Асимпт.	p	Нижняя	Верхняя
q(1)	0,625693	0,036780	17,0118	0,000000	0,553433	0,697954
Ps(1)	-0,622758	0,044621	-13,9566	0,000000	-0,710425	-0,535092
Ps(2)	-0,340056	0,044442	-7,6517	0,000000	-0,427370	-0,252742
Qs(1)	0,852947	0,021351	39,9483	0,000000	0,810998	0,894895

Проведем анализ остатков.

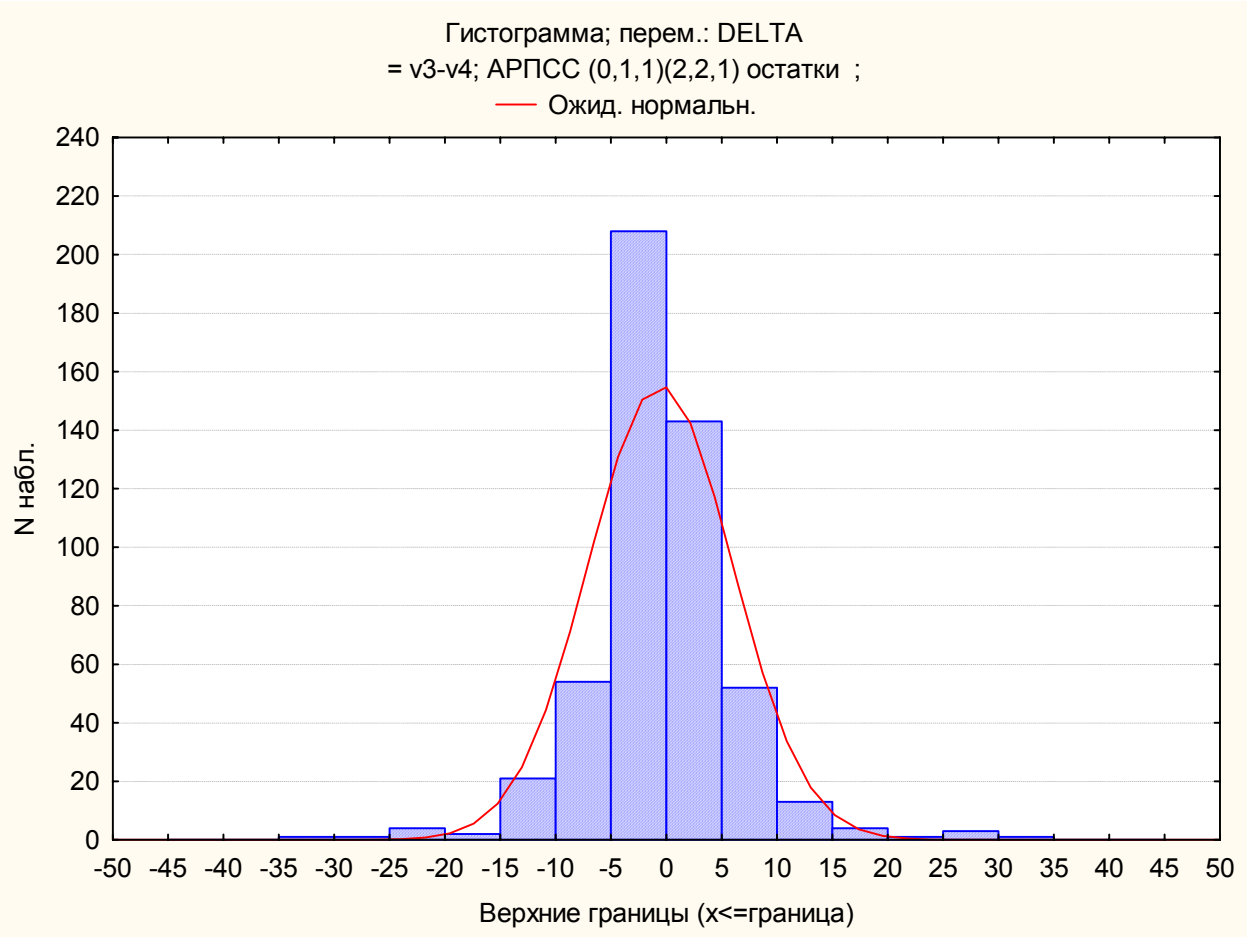


Рис. 6.12

Гистограмма частот ошибок модели АРПСС(0,1,1)(2,2,1)

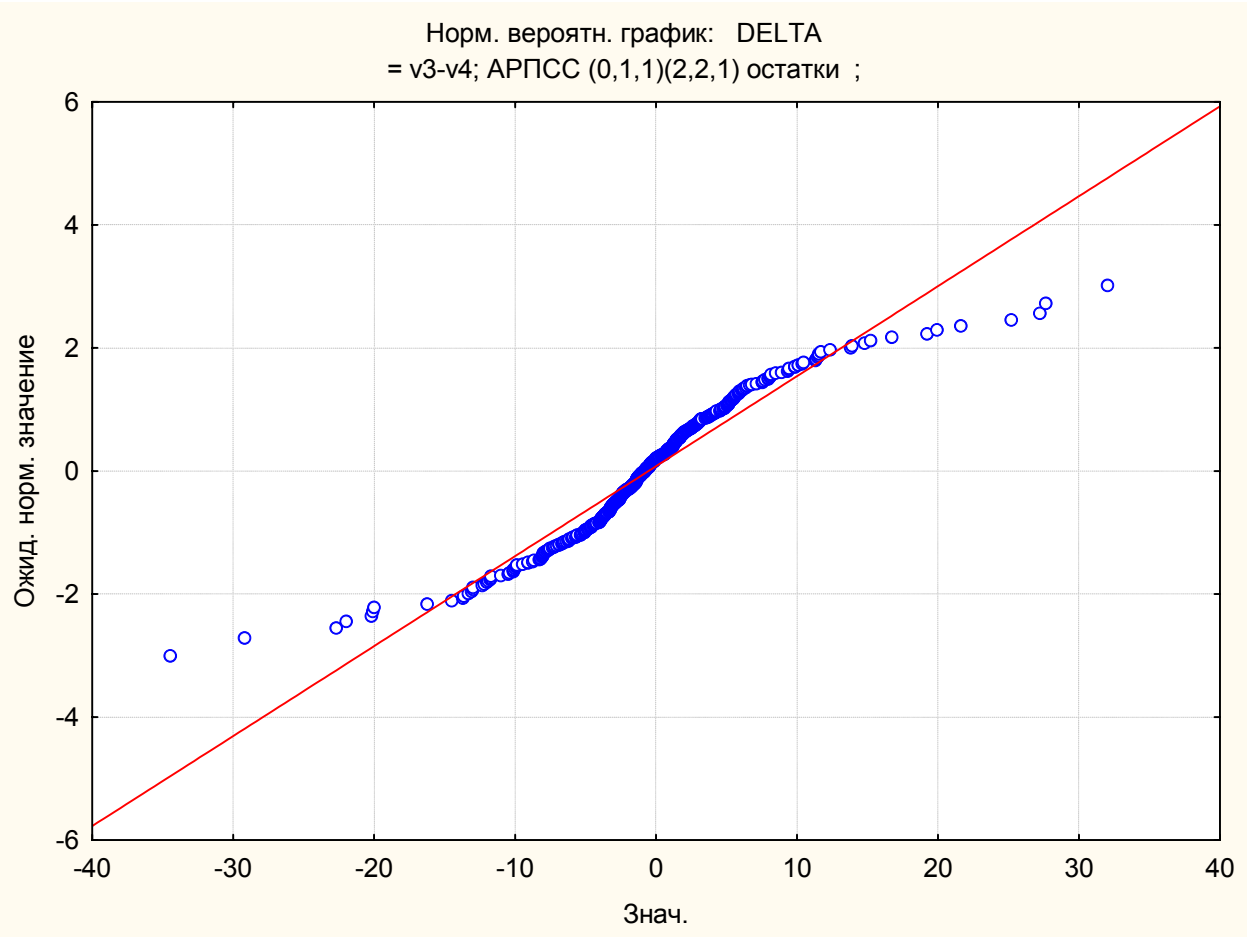
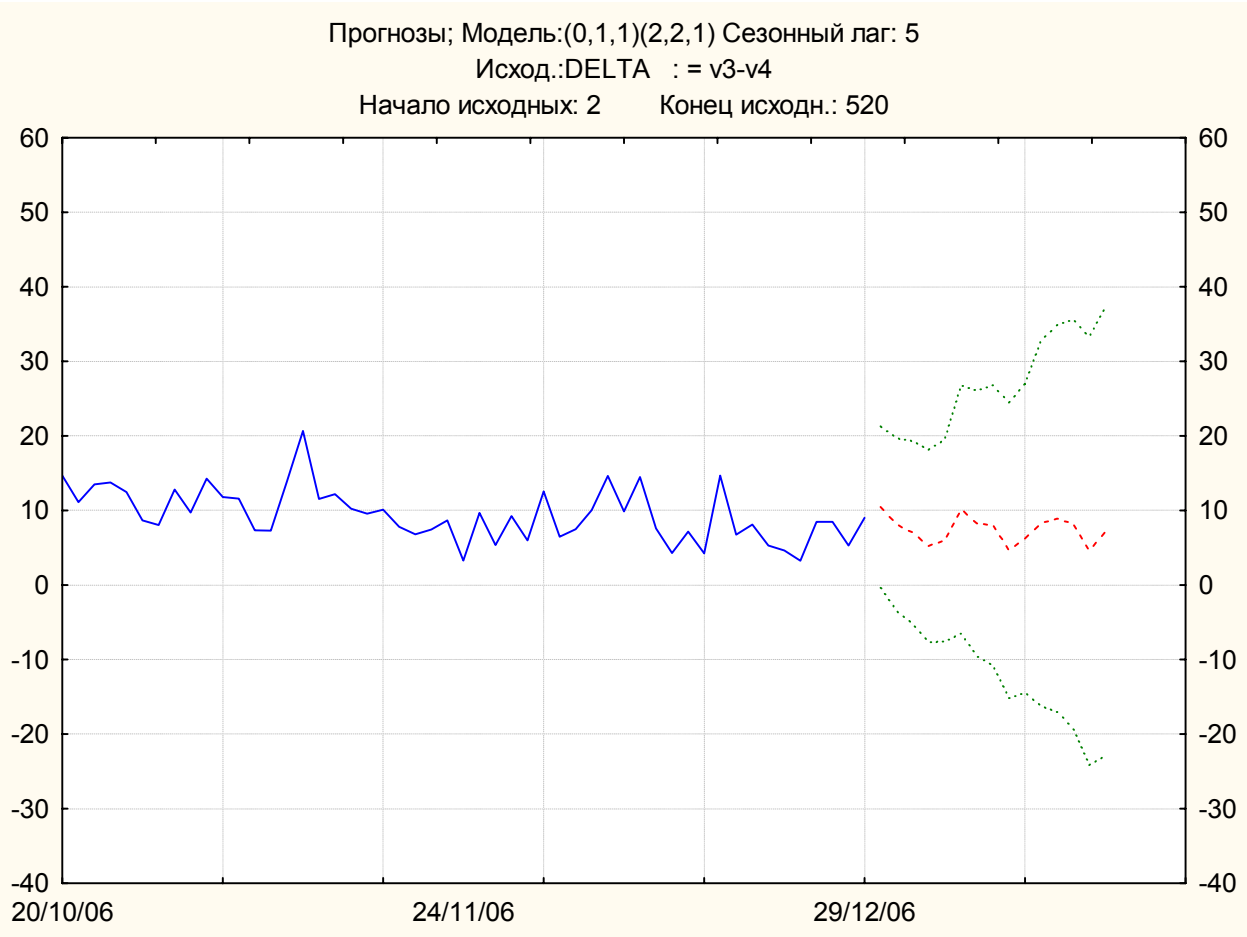


Рис. 6.13

Нормальный вероятностный график ошибок в модели АРПСС(0,!,1)(2,2,1)

Итак, приведенные графики говорят о «хорошем» поведении остатков. Прогноз следующий:



Прогнозы; Модель:(0,1,1)(2,2,1) Сезонный лаг: 5 (GOLD_2005_2006_day.sta)
Исход.:DELTA : = v3-v4 Начало исходных: 2 Конец исходн.: 520

	Прогноз	Нижний	Верхний	Ст.ошиб
1 января 2006 г.	10,46903	-0,3583	21,29637	6,57045
2 января 2006 г.	8,09792	-3,4631	19,65888	7,01565
3 января 2006 г.	7,03017	-5,2206	19,28091	7,43423
4 января 2006 г.	5,23850	-7,6652	18,14220	7,83047
5 января 2006 г.	5,94167	-7,5835	19,46684	8,20760
8 января 2006 г.	10,13766	-6,5234	26,79877	10,11062
9 января 2006 г.	8,26939	-9,5001	26,03889	10,78323
10 января 2006 г.	7,99846	-10,8142	26,81115	11,41628
11 января 2006 г.	4,69688	-15,1041	24,49789	12,01603
12 января 2006 г.	6,26381	-14,4785	27,00610	12,58724

15 января 2006 г.	8,28205	-16,1826	32,74671	14,84612
16 января 2006 г.	8,92352	-17,0490	34,89607	15,76116
17 января 2006 г.	8,19225	-19,2053	35,58981	16,62592
18 января 2006 г.	4,58319	-24,1688	33,33522	17,44786
19 января 2006 г.	7,11652	-22,9290	37,16202	18,23280

Рис. 6.14

График ряда DELTA и прогноз, построенный при помощи модели АРПСС(0,1,1)(2,2,1)

После построения прогноза, запустим макрос, который по полученному уравнению нелинейной регрессии вычислить прогноз переменной LOW на 15 наблюдений вперед.

Полученное нами уравнение следующее

LOW = 289,472 + 5,102 * (номер наблюдения) + 1,957*(delta) + 516,118* log (номер наблюдения) -0, 00000372260312778738*(номер наблюдения^3) - 142,094 * sqrt (номер наблюдения)

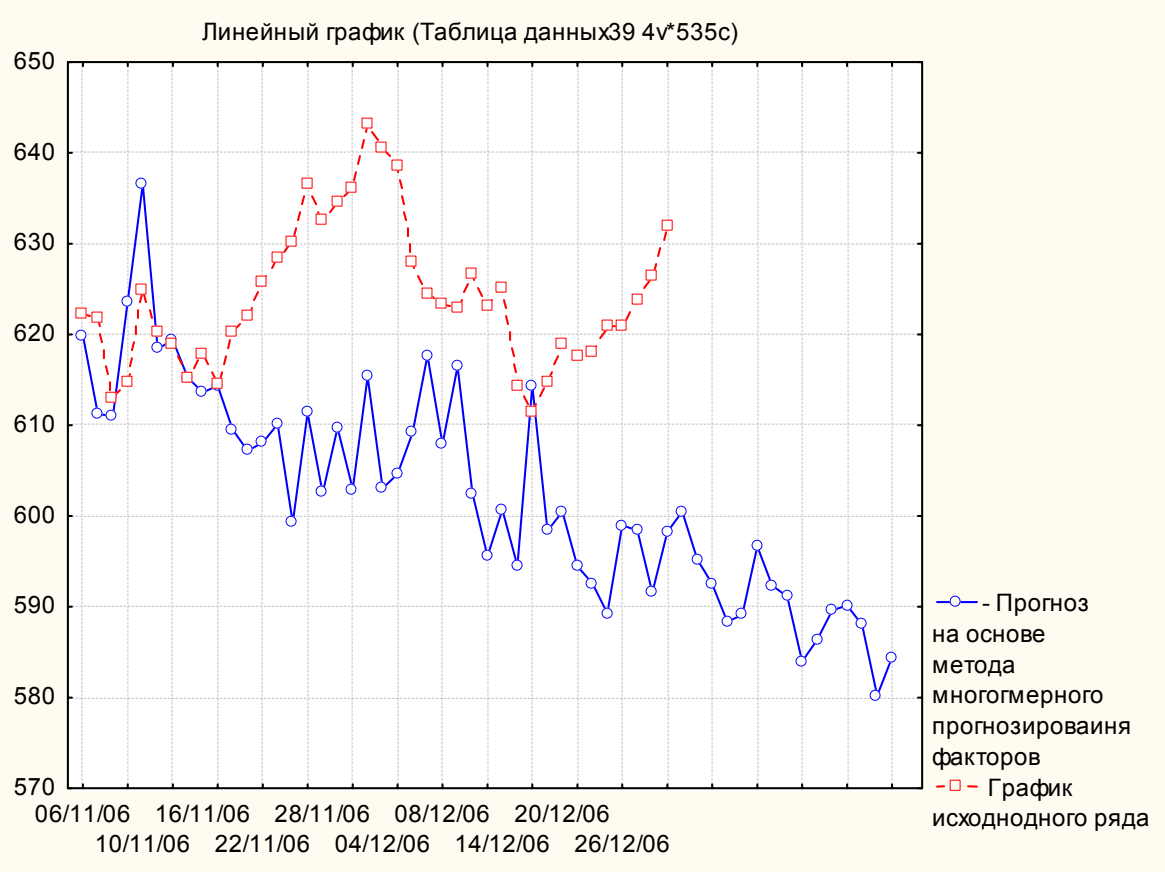


Рис. 6.15

Графики прогнозов ряда при использовании метода многомерного прогнозирования

Итак, мы получили следующие значения прогноза **(прогноз** 1). а также запишем в эту же таблицу значения прогноза, полученного при использовании нейронных сетей **(прогноз** 2).

	Прогноз 1	Прогноз 2(считаем этот прогноз наилучшим)
1 января 2006 г.	600,478698	625,069347
2 января 2006 г.	595,221855	622,686824
3 января 2006 г.	592,506202	617,987886
4 января 2006 г.	588,364316	611,828524
5 января 2006 г.	589,095252	607,272057
8 января 2006 г.	596,652068	605,351248
9 января 2006 г.	592,331508	602,996582
10 января 2006 г.	591,127282	600,152152
11 января 2006 г.	583,982425	597,07978
12 января 2006 г.	586,355546	594,410874
15 января 2006 г.	589,602138	592,876187
16 января 2006 г.	590,144667	592,837292
17 января 2006 г.	587,990935	595,069185

Также построим в одной координатной плоскости оба этих прогноза для сравнения.

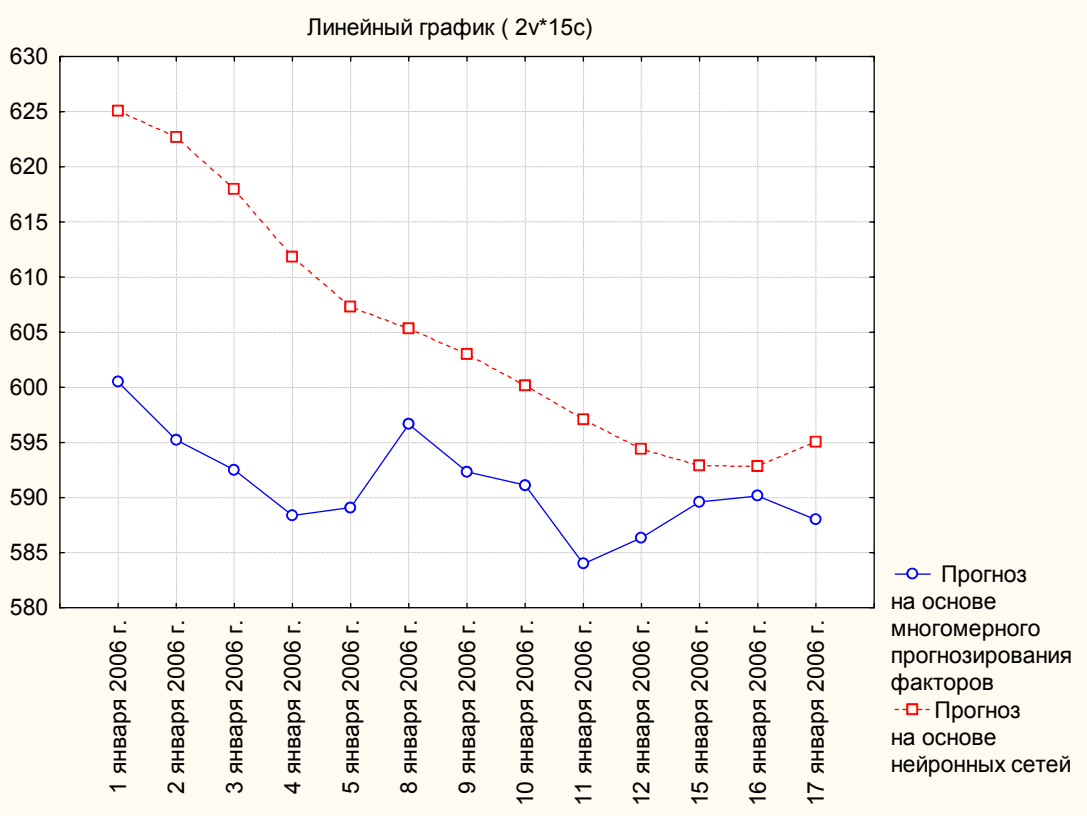


Рис. 6.16

Графики прогнозов ряда при использовании нейронных сетей и многомерного прогнозирования факторов

Вывод:

Проанализируем полученные результаты. Во-первых, в нашем распоряжении были не совсем независимые факторы. Они были получены нами искусственно из исходных рядов (HIGH, LOW, OPEN, CLOSE). Это стало причиной того, что мы буквально пытались «вытащить себя за волосы». В идеале же факторы не должны зависеть от анализируемой информации, а только лишь наоборот.

Во-вторых, построенная нами модель нелинейной регрессии показывала, что переменная LOW гораздо сильнее зависит от номера наблюдения, чем от других факторов. Это обстоятельство плохо согласуется со здравым смыслом. Оно утверждает, что цены на рынке больше всего зависят именно от времени, а не от других факторов, но в данной ситуации в нашем распоряжении было мало независимых (как от LOW, так и между собой) факторов, и возможно они на самом деле хуже объясняли поведении ряда, чем номер наблюдения.

Отметим, что уравнение регрессии можно достаточно точно аппроксимировать полиномом третьей степени от 1(номера наблюдения). Таким образом, получаем, что мы ищем полином третьей степени от t. который меньше всего отличается от рассматриваемого ряда на рассматриваемом интервале изменения значения t (наблюдение №1- наблюдение №520). В каком-то смысле это есть задача интерполяции. Но за пределами данного интервала построенный полином уже не обязан минимизировать отклонения от значений ряда (имеется виду его прогноз или реальные значения).

С ростом номера наблюдения функция регрессии будет очень быстро убывать (точнее как -const*t^3), а поэтому использовать её для прогноза нельзя. Ниже представлены графики ряда LOW и регрессионного уравнения.

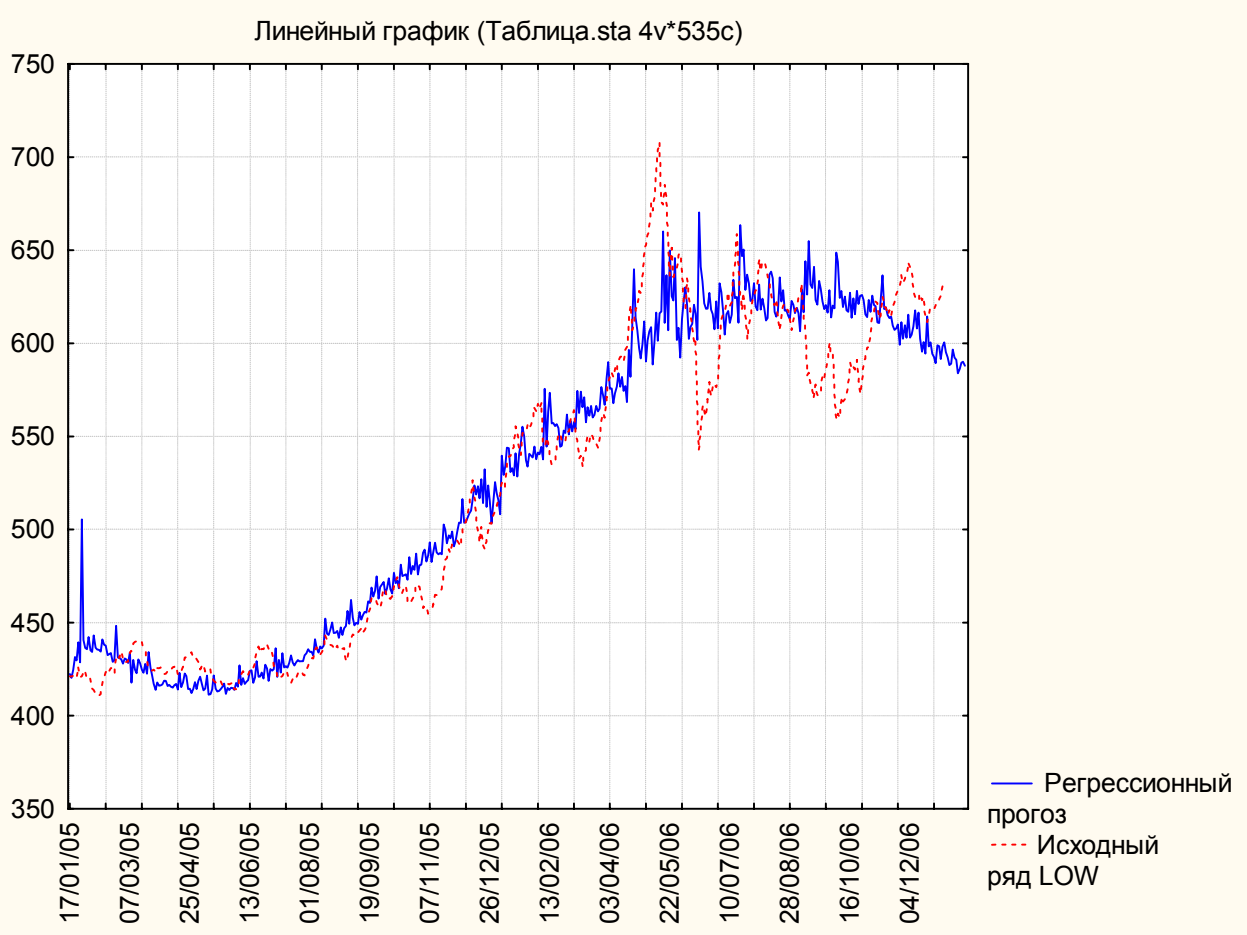


Рис. 6.17

Графики ряда LOW и прогноза, построенного при помощи метода многомерного прогнозирования факторов.

7.Заключение

Целью данной работы была демонстрация работы тех или иных методов прогнозирования временных рядов на реальных данных, а также интерпретация результатов. Были рассмотрены все доступные в системе *STATISTICA* методы, а именно модуль «Временные ряды и прогнозирования» и «Нейронные сети». Также был рассмотрен альтернативный подход прогнозирования факторов влияния.

Итогами работы стали следующие результаты:

1) Метод построения АРПСС моделей оказался не адекватным решаемой задаче, потому как в ряде имеет место очень слабая зависимость от предыдущих значений, а сезонность выражена слабо.

2) Сезонная декомпозиция также не дала хороших результатов по тем же причинам.

3) Метод экспоненциального сглаживания оказался наиболее продуктивным среди классических методов прогнозирования. Значения

спрогнозированного ряда сначала убывают, а затем колеблются около фиксированного значения.

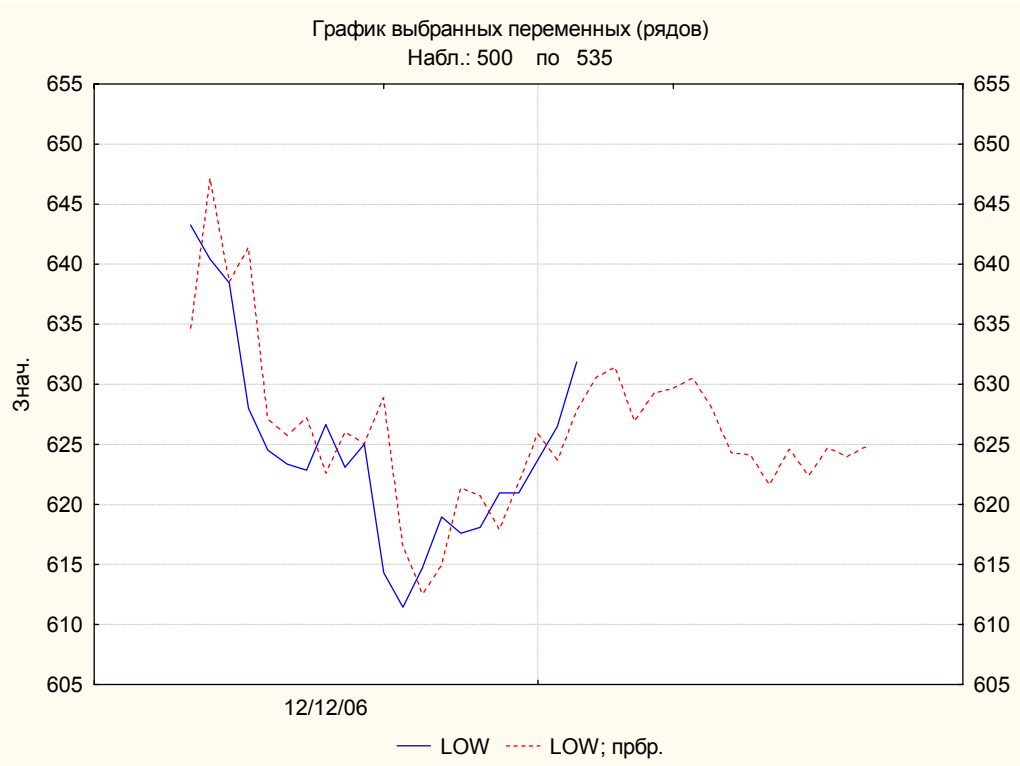


Рис. 7.1

Конечный отрезок (ноябрь 2006 г – январь 2007 г) исходного ряда и ряда, полученного в результате экспоненциального сглаживания в модели Винтера

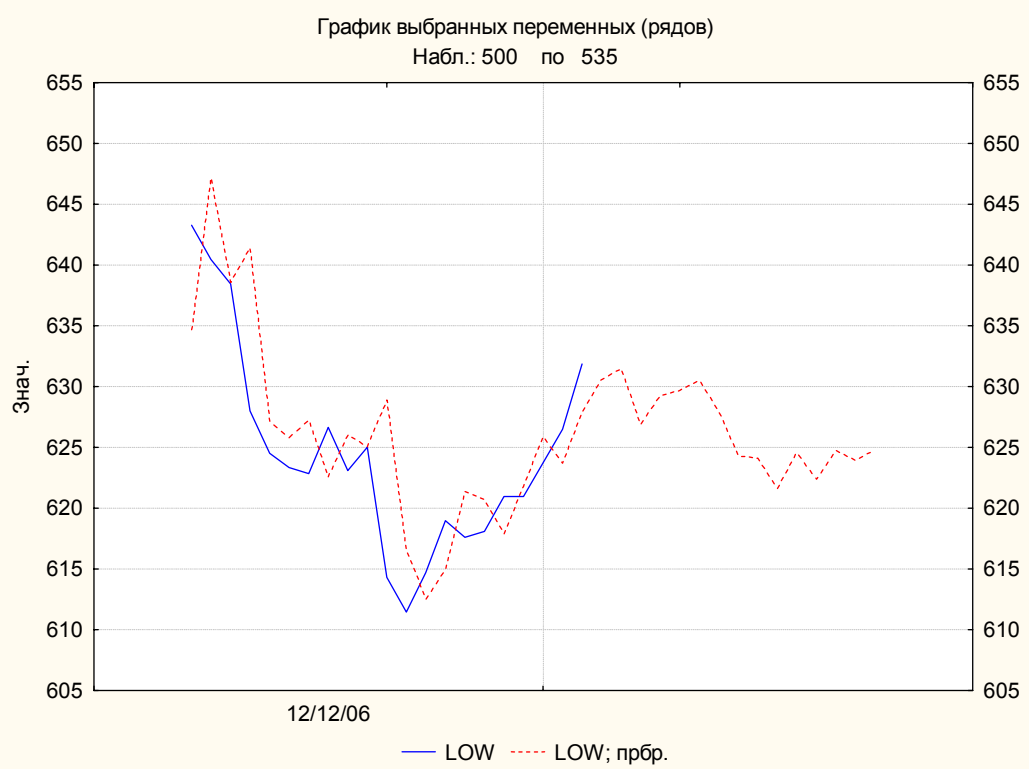


Рис. 7.2

Конечный отрезок (ноябрь 2006 г– январь 2007 г) исходного ряда и ряда, полученного в результате экспоненциального сглаживания в модели с мультипликативной сезонностью и экспоненциальным трендом.

Качественно прогноз этого метода одинаков для моделей с линейным и с экспоненциальным трендом.

4) Нейронные сети с большой степенью точности дали следующий результат.

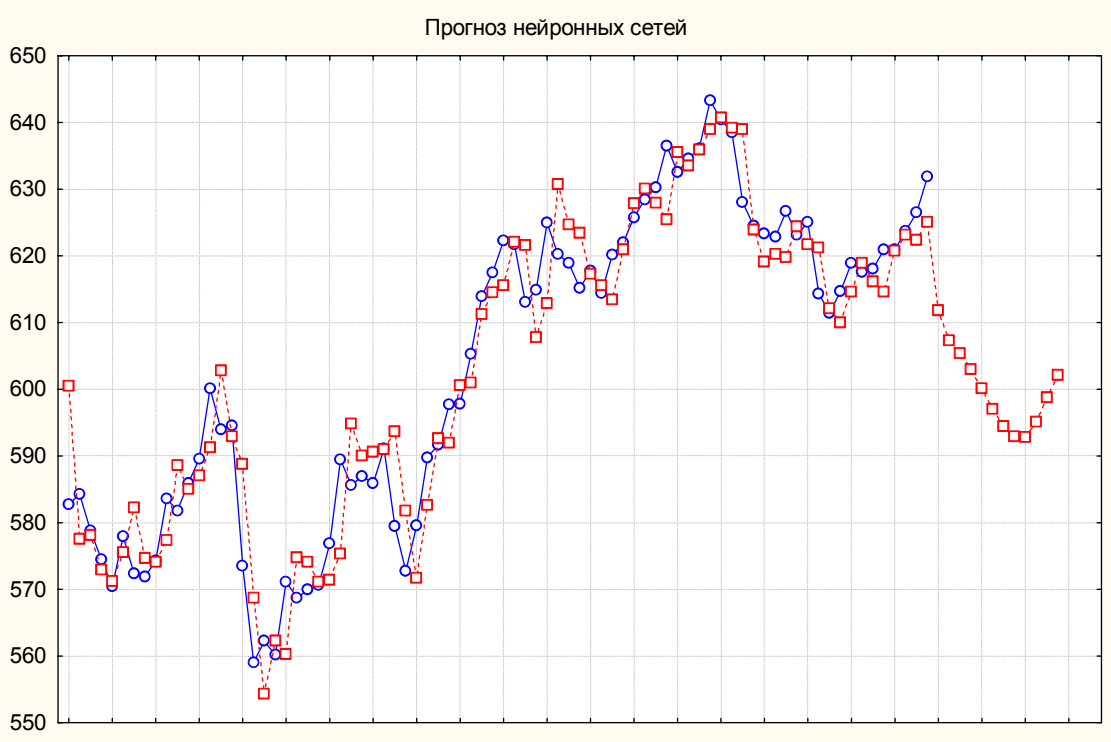


Рис.7.3

Прогноз, построенный при помощи нейронных сетей.

Вид прогноза схож с видом прогнозов экспоненциального сглаживания, что в свою очередь дает дополнительный повод убедиться в его правильности.

5) Ввиду отсутствия данных о независимых факторах, влияющих на поведение ряда, метод многомерного прогнозирования не смог описать этот ряд, хотя результаты оказались весьма интересными с точки зрения математики.

На основе всего вышеперечисленного можно сделать вывод о том, что прогнозирование с использованием нейронных сетей дало наиболее качественный результат. Именно его мы будем считать результатом всей работы.

	Итоговый прогноз
1 января 2006 г.	625,069347
2 января 2006 г.	622,686824

3 января 2006 г.	617,987886
4 января 2006 г.	611,828524
5 января 2006 г.	607,272057
8 января 2006 г.	605,351248
9 января 2006 г.	602,996582
10 января 2006 г.	600,152152
11 января 2006 г.	597,07978
12 января 2006 г.	594,410874
15 января 2006 г.	592,876187
16 января 2006 г.	592,837292
17 января 2006 г.	595,069185

Анализ и прогнозирование именно реального ряда дали возможность показать, насколько важно использовать в работе разнообразные приемы, а также их комбинации, учитывать специфику переменных для определения границ применения того или иного метода.

Также отмечена ценность, которую несут в себе неклассические методы.

8. Приложение

1.Построение линейных графиков.

Для того чтобы построить линейный график одной или нескольких переменных, необходимо в меню анализа «Графика» выбрать «2М графики» «Линейные графики (для переменных)».

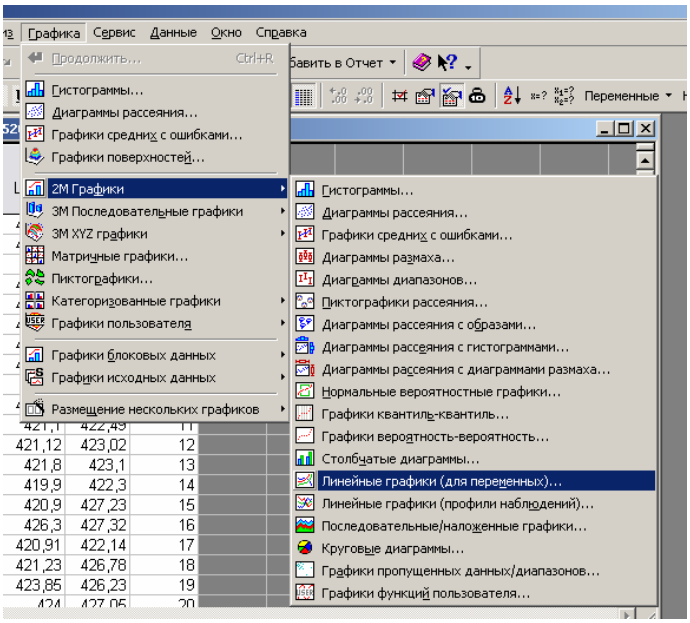


Рис.1.2

Меню анализа «Графика»

Затем, в появившемся диалоге на вкладке «Быстрый» выбрать соответствующие переменные и опцию «Составной» в окне тип графика. Остальные вкладки оставим без изменений.

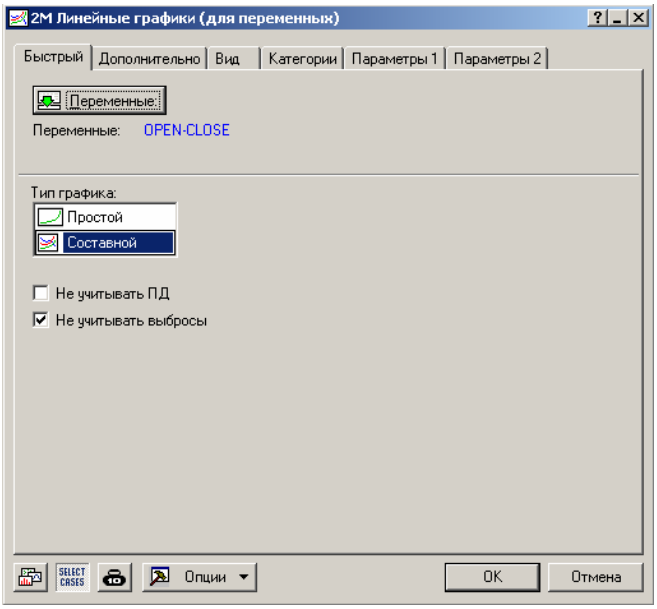


Рис.1.3

Диалог «Линейные графики»

После нажатия «ОК» отобразятся графики рядов.

2. Корреляции переменных

Корреляции (GOLD_2005_2006_day.sta) Отмеченные корреляции значимы на уровне p

	OPEN	HIGH	LOW	CLOSE
OPEN	1	0,998838811	0,998307315	0,996743185
HIGH	0,998838811	1	0,998233442	0,99811698
LOW	0,998307315	0,998233442	1	0,998901481
CLOSE	0,996743185	0,99811698	0,998901481	1

Рис. 8.3
Матрица корреляций переменных

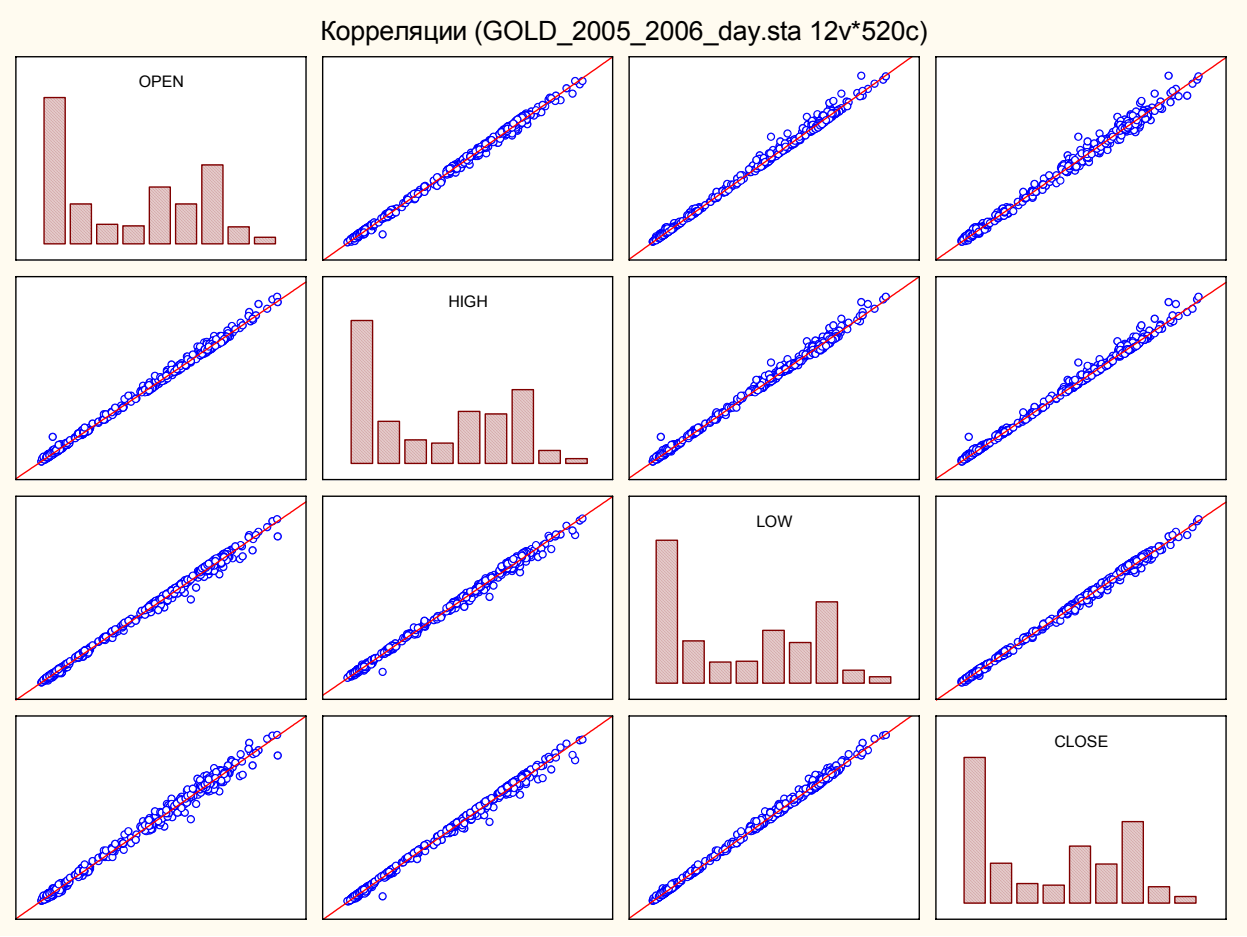


Рис. 8.4
Матричная диаграмма рассеяния

3. Построение графика временного ряда в модуле «Анализ временных рядов»

Для построения графика при анализе временных рядов необходимо проделать следующую процедуру:

1. Для любого метода анализа (т.е. для любой кнопки диалога «Анализ временных рядов») выбрать вкладку «Прогноз».
2. В окне диалога выбрать те ряды, графики которых необходимо построить.
3. При необходимости изменить опции самого графика:
   А) «Отображать только часть ряда», указав отображаемый отрезок
   Б) «Задать масштаб оси вручную», указав шаг масштабирования
   В) «Пометить точки над X», указав тип метки и в случае необходимости имя переменной, содержащей имена меток

4. В зависимости от числа выбранных рядов выбрать одну из опций «График выделенной переменной» или «График нескольких переменных».

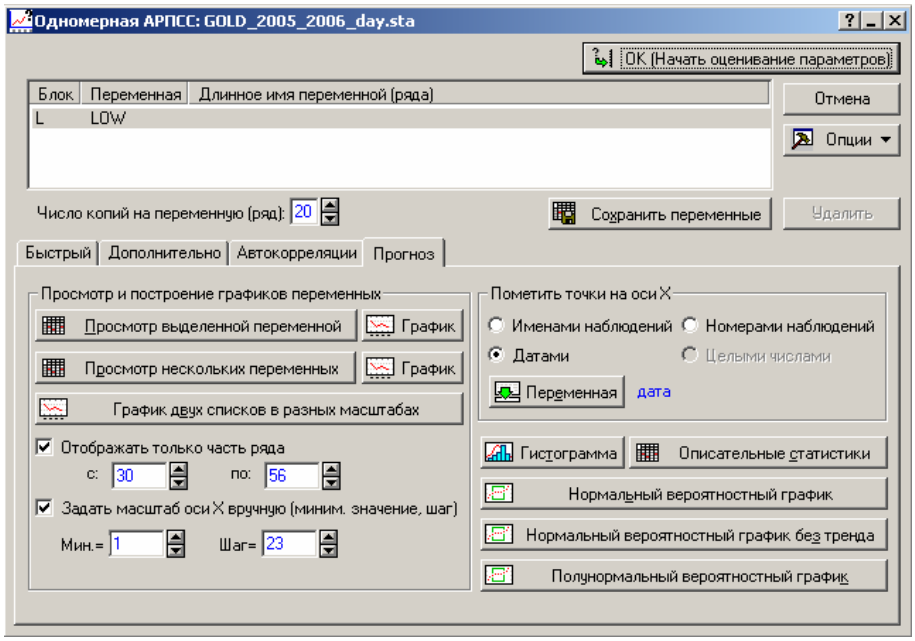


Рис 8.5

Построение графика ряда

4. Построение модели АРПСС с заданными параметрами в модуле «Анализ временных рядов» «АРПСС и автокорреляционные функции»

Для построения прогноза и анализа результатов временного ряда, используя модель АРПСС(p,d,q) необходимо проделать следующую процедуру:

1. В диалоге «Временные ряды и прогнозирование» выбрать метод «Одномерная АРПСС и корреляционные функции» (для метода «АРПСС с интервенцией» процедура построения прогноза та же, добавляются лишь данные об интервенциях - см. ниже)

2. На вкладке «Быстрый» или «Дополнительно» выбрать соответствующие значения параметров модели.

3. Предположим, что ранее было выполнено логарифмическое преобразование данных и взята одна разность с лагом 1. Все эти преобразования уже выполнены и результаты просмотрены. Преобразованный ряд можно теперь непосредственно использовать в АРПСС. Однако в большинстве случаев рекомендуется анализировать исходный ряд и задать необходимые преобразования внутри АРПСС (эти преобразования будут частью спецификации АРПСС). Если вы захотите построить прогноз (после оценки параметров АРПСС), то он будет вычислен из проинтегрированных рядов ("интегрирование", более точно суммирование, в данном случае означает просто операцию, обратную взятию разностей с соответствующими лагами). Таким образом, проводя обратные преобразования, вы возвращаетесь к исходному ряду, и прогноз соответствует исходным данным (что обеспечивает более легкую интерпретацию результатов).
   Заметим, внутри АРПСС доступны только преобразования логарифм, возведение в степень и взятие сезонных/несезонных разностей. В некоторых случаях определенные преобразования рекомендуется выполнять до работы в АРПСС. Речь идет о преобразованиях (например, сглаживание), не изменяющих диапазон данных и к которым не нужно применять обратные преобразования.

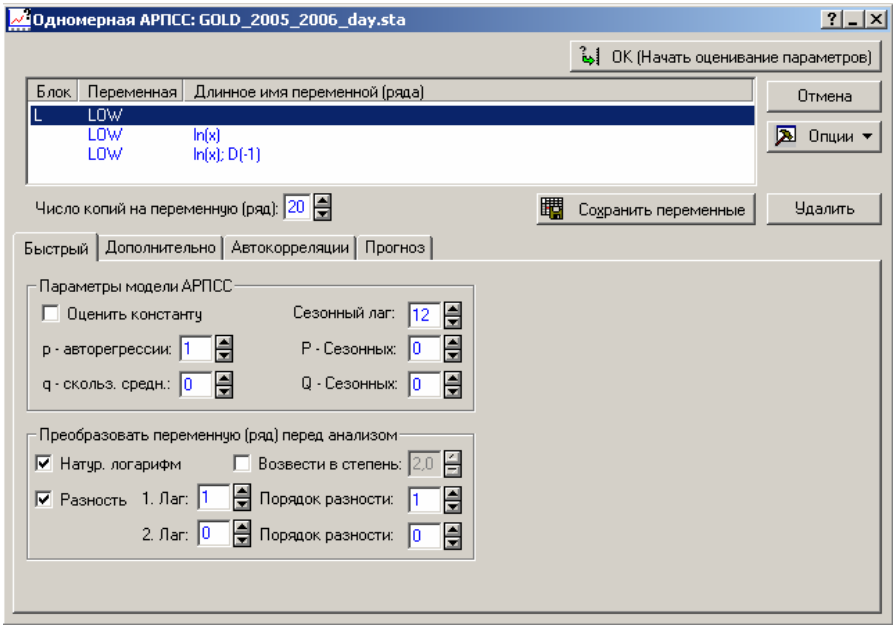


Рис. 8.6

Диалог «Одномерная АРПСС»

4. После окончания процесса оценивания параметров, отобразится диалог «Результаты АРПСС», в котором можно просмотреть таблицы параметров и их ошибок, построить таблицы ковариации и корреляции параметров, построить графики ряда и прогноза в одной системе координат. Помимо анализа параметров пользователю также доступен мощный аппарат анализа остатков, который позволяет, например, проверить соответствие стандартным предположениям. Просмотр характеристик остатков осуществляется соответствующими опциями на вкладках «Распределение остатков» и «Автокорреляции» настоящего диалога.

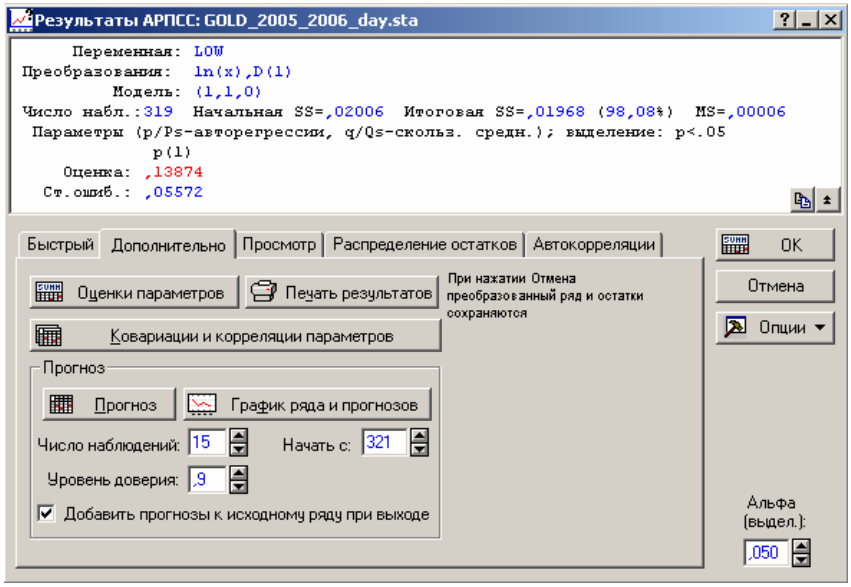


Рис. 8.7

Диалог «Результаты АРПСС»