

УДК 519.246.8

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ СГЛАЖИВАНИЯ И МГУА ДЛЯ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ

Н.В. Кондрашова, Я.В. Павлов

*Международный научно-учебный центр информационных технологий
и систем НАН та МОН Украины,
nkondrashova@ukr.net, combatfree@i.net*

В данной работе содержатся анализ широко известных методов прогнозирования, реализованных в программном пакете *STATISTICA*, а также сравнение полученных результатов прогнозов некоторых гетероскедастических временных рядов с прогнозами моделей, построенных Методом Группового Учета Аргументов (*МГУА*).

Ключевые слова: прогнозирующая модель, *STATISTICA*, *МГУА*, гетероскедастический временной ряд, нефть марок *Brent* и *Urals*, адаптивный прогноз.

Analysis of the well-known forecasting methods which had been realized in the programmatic *STATISTICA* package and also the results comparison of their got forecasts of the some heteroskedastic time series and the results of *GMDH* forecasts models are contained in this work.

Key words: forecasting model, heteroskedastic time series, oil of sorts *Brent* and *Urals*, adaptive forecast

У даній роботі містяться аналіз широко відомих методів прогнозування, реалізованих в програмному пакеті *STATISTICA*, а також порівняння отриманих результатів прогнозів деяких гетероскедастических часових рядів з прогнозами моделей, побудованих методом групового урахування аргументів (*МГУА*).

Ключеві слова: прогнозуюча модель, *STATISTICA*, *МГУА*, гетероскедастичний часовий ряд, нафта морок *Brent* і *Urals*, адаптивний прогноз.

Введение. Рассматриваются следующие проблемы: во-первых, возможно ли с помощью известных методов, предназначенных описывать относительно устойчивые связи по однородным данным, краткосрочное прогнозирование *гетероскедастических* временных рядов? Во-вторых, если методы применимы для решения поставленной задачи, установит их наиболее перспективный класс, указать характерные особенности этих методов, уделив основным из них особое внимание. В-третьих, показать на примере практические результаты хотя бы некоторых попыток прогнозирования гетероскедастических рядов и провести сравнительный анализ с некоторым количеством прогнозов по *МГУА* [1].

Гетероскедастичность (неоднородный разброс) – свойство, общее у многих финансовых временных рядов, таких как: цены на акции, золото, нефть, курс валют и т.д.. Эти ряды являются результатом интегрированного воздействия огромного количества факторов. Мы рассматриваем краткосрочное адаптивное прогнозирование, как способ получения оценок ряда будущих значений переменной, причиной изменения которой является, главным образом, массовое поведение на товарном рынке мелких и крупных дилеров, совершающих основной объем торговых операций в предшествующие моменты времени. При таком упрощенном подходе динамику временного ряда можно прочитать как хронологическую запись о массовом поведении участников

товарного рынка. Это дает возможность при построении модели исходить из самого ряда, не привлекая дополнительной информации.

В качестве широко известных методов построения разностных моделей были выбраны 1) различные модификации метода экспоненциального сглаживания (*Exponential Smoothing*) [2]; 2) метод Бокса-Дженкинса, реализованные в пакете *STATISTICA* и 3) метод группового учета аргументов.

Детально уделяется внимание методу Бокса-Дженкинса [3] и следующим модификациям метода экспоненциального сглаживания: 1) методу Хольта (*Holt's method*) или сглаживания с линейным трендом [4]; 2) методу сглаживания с экспоненциальным трендом [5]; 3) Хольта-Винтера (*Holt-Winter's method*) или экспоненциального сглаживания с линейными трендом и сезонной компонентой [2]. Кроме того, кратко изложены особенности применения *МГУА* для прогнозирования временных рядов. Модели *МГУА* – метода построения линейных по параметрам моделей в классе заданных функций – были построены для тренда и остатков. Сконструированы модели на 1, 2 и 3 шага прогнозирования [6] и адаптивная модель прогноза [7], при помощи известных алгоритмов *COMBI* [6] и *МАКСО* [8], первый из которых является обновленной версией алгоритма *COMBI* пакета *ASTRID*. Второй – является усовершенствованной версией алгоритма с ортогонализацией переменных [9].

Подробнее остановимся на перечисленных выше методах.

1 Анализ методов сглаживания

К методам сглаживания мы относим не только те, которые в своем названии используют слово «сглаживание», но и другие методы, усредняющие данные, например, *метод скользящего среднего*.

1.1 Экспоненциальное сглаживание

Экспоненциальное сглаживание является простым вариантом самообучающейся модели. Вычисления выполняются итеративно. Они требуют меньше арифметических операций, чем для моделей со скользящим средним $S_t = \frac{1}{\tau} \sum_{i=\tau}^1 x_{t-i}$. Массив предыстории уменьшается до одного значения S_{t-1} .

Экспоненциальное сглаживание всегда использует предыдущее значение экспоненциальной средней:

$$S_t = \mu x_t + (1 - \mu) S_{t-1}, \quad (1)$$

где $0 < \mu < 1$ параметр сглаживания. Когда процесс только начинается, должна быть некоторая величина S_0 , например, среднее значение $S_0 = \frac{1}{\tau} \sum_{i=1}^{\tau} x_i$, которое может быть использовано в качестве значения, предшествующего S_1 .

Рассмотрим ряд, генерированный моделью:

$$x_t = a + \zeta_t$$

где $a = \text{const}$; ξ_t — случайные неавтокоррелированные отклонения, или шум со средним значением 0 и дисперсией σ^2 . Применим к нему процедуру экспоненциального сглаживания (1). Можно показать, что:

$$S_{t=\mu} = \sum_{i=0}^N \omega^i x_{t-i} + S_0 \omega^N,$$

является взвешенной суммой всех членов ряда. Так как $\omega < 1$ ($\omega = 1 - \mu$), то при $N \rightarrow \infty$ $\omega^N \rightarrow 0$, и $\mu \sum_{i=0}^{N-1} \omega^i \rightarrow 1$. $S_{t=\mu} = \sum_{i=0}^{\infty} \omega^i x_{t-i} = a + \mu \sum_{i=0}^{\infty} \omega^i \xi_{t-i}$. Следовательно, экспоненциальное сглаживание можно представить как фильтр, на вход которого в виде потока последовательно поступают члены исходного ряда, а на выходе формируются текущие значения экспоненциальной средней. Причем веса уменьшаются экспоненциально в зависимости от давности («возраста») наблюдения. Чем меньше μ , тем в большей степени фильтруются колебания исходного ряда.

Найдем математическое ожидание $M(S_t) = M(x_t) = a$ и дисперсию:

$$D(S_t) = M[(S_t - a)^2] = M \left[\left(\mu \sum_{i=0}^{\infty} \omega^i \xi_{t-i} \right)^2 \right] = \mu^2 \sum_{i=0}^{\infty} \omega^{2i} \sigma^2 = \sigma^2 \frac{\mu}{2 - \mu}.$$

Так как $0 < \mu < 1$, то $D(S_t) < D(x_t) = \sigma^2$. Как видно при большом значении μ дисперсия экспоненциальной средней незначительно отличается от дисперсии ряда x_t . Таким образом, экспоненциальная средняя S_t имеет то же математическое ожидание (МО), что и ряд x_t , но меньшую дисперсию. Чем меньше μ , тем в большей степени уменьшается дисперсия экспоненциальной средней. Экспоненциальное среднее используется для прогноза в текущий момент t на период упреждения l : $x_t(l) = S_t = S_{t-1} + \mu(x_t - S_{t-1})$. Новый прогноз получаем корректировкой предыдущего прогноза с учетом его ошибки. Поскольку любая дискретно-непрерывная функция может быть представлена в виде суммы ступенчатых функций вида: $a_t = a_t \cdot 1(t)$, то предполагается, что временной ряд генерируется моделью $x_t = a_t + \xi_t$, где a_t — в общем случае есть случайная величина. При краткосрочном прогнозировании желательно отобразить изменения x_t и в то же время наилучшим образом очистить ряд от случайных колебаний ξ_t . Для этого, с одной стороны, необходимо увеличить вес более свежих наблюдений, что можно достичь увеличением μ , а с другой стороны для сглаживания случайных отклонений величину μ нужно уменьшить. Как видно, эти два требования находятся в противоречии. Поиск компромиссного значения μ составляет задание оптимизации модели. В [10], [11] было показано для прогнозирования стационарного процесса с

автокорреляционной функцией вида $\rho_i = \rho_1^i$, что минимум среднего квадрата ошибки при прогнозировании такого ряда на 1 шаг вперед ($l=i=1$) будет при

$$\mu_{opt} = \begin{cases} (3\rho_1 - 1)/2\rho_1 & 1/3 < \rho_1 \leq 1 \\ 0 & -1 \leq \rho_1 \leq 1/3 \end{cases}$$

где ρ_1 - коэффициент автокорреляции при лаге 1. Если $l > 1$, то существенно повышается критическая величина $\rho_{1крит}$, ниже которой оптимальное значение μ_{opt} равно 0. Если данные сильно коррелированы $\rho_1 \geq 0.9$ и период упреждения l мал, то сглаживать не стоит. Целесообразно в качестве прогноза использовать наиболее позднее наблюдение. Если $\rho_1 \leq 1/3$, то наибольшее, что может дать простейшая модель, это оценка среднего уровня, вокруг которого варьирует процесс [4]. Для краткосрочных прогнозов значение μ , как правило, нужно брать большим, а для более долгосрочных – более малым. Выбор конкретных значений параметров μ во многом определяется опытом разработчика.

1.1.1 Метод Хольта

Метод Хольта или экспоненциального сглаживания с линейным трендом – родоначальник методов адаптивного прогнозирования:

$$\begin{cases} S_t = \mu(S_{t-1} + T_{t-1}) + (1 - \mu)x_t \\ T_t = \nu(S_t - S_{t-1}) + (1 - \nu)T_{t-1} \end{cases} \quad (2)$$

в котором кроме константы μ используются константа ν сглаживания тренда T_t . Для оценивания μ, ν пакет *STATISTICA* использует квази-ньютоновский метод, минимизирующий суммарную квадратичную ошибку.

1.1.2 Сглаживание с экспоненциальным трендом

В этой модели временного ряда прогноз простого экспоненциального сглаживания (1) "улучшается" с помощью экспоненциального тренда, сглаженного, как в (2), с параметром ν . Если при линейном тренде описываемый процесс изменяется во времени равномерно и имеет стабильный прирост линейного тренда, то в модели с экспоненциальным трендом стабильным является относительная скорость (темп) прироста:

$$\begin{cases} S_t = \mu(S_{t-1} + T_{t-1}) + (1 - \mu)x_t \\ T_t = \nu(\Delta S_t - \Delta S_{t-1}) + (1 - \nu)T_{t-1} \end{cases}$$

где $\Delta S_t = S_t - S_{t-1}$.

1.1.3 Метод Хольта-Винтера

Метод Хольта-Винтера или метод Винтера, в котором фигурирует дополнительная константа η сглаживания циклов с периодом p :

$$S_t = \mu(S_{t-1} + T_{t-1}) + (1 - \mu) \frac{x_t}{Z_{t-p}}$$

$$\begin{aligned} T_t &= \nu(S_t - S_{t-1}) + (1-\nu)T_{t-1} \\ Z_t &= \eta(x_t / S_{t-1}) + (1-\eta)Z_{t-p} \end{aligned}$$

Оптимальное значение параметра p есть то, которое минимизирует сумму квадратов остатков.

1.2 Метод Бокса-Дженкинса

В *методе Бокса-Дженкинса* построение модели авторегрессии проинтегрированного скользящего среднего *ARIMA – Auto Regression Integrated Moving Average* сводится к построению модели авторегрессии и скользящего среднего *ARMA - Autoregressive Moving Average*. Вкратце опишем этот процесс.

Пусть $\tilde{x}_t, \tilde{x}_{t-1}, \tilde{x}_{t-2} \dots$ будут отклонениями от математического ожидания \bar{x} , $\tilde{x}_t = x_t - \bar{x}$. Ряд можно трактовать как процесс авторегрессии плюс добавочный импульс a_t

$$\tilde{x}_t = \theta_1 \tilde{x}_{t-1} + \theta_2 \tilde{x}_{t-2} + \dots + a_t \quad (3)$$

где a_t - белый шум с $D(a_t) = \sigma_a^2$, $M(a_t) = 0$. Стохастический процесс (3) также можно рассматривать как выход линейного фильтра, на вход которого поступает белый шум a_t , т. е.

$$\tilde{x}_t = a_t + \varphi_1 a_{t-1} + \varphi_2 a_{t-2} + \dots$$

Но на практике для получения экономичной параметризации в модель авторегрессии включают скользящее среднее и называют это моделью смешанного процесса авторегрессии с остатками в виде скользящего среднего:

$$\tilde{x}_t = \theta_1 \tilde{x}_{t-1} + \dots + \theta_p \tilde{x}_{t-p} + a_t - \varphi_1 a_{t-1} - \dots - \varphi_q a_{t-q}.$$

В операторном виде это записывается как

$$(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^p) \tilde{x}_t = (1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_q B^q) a_t$$

где введен оператор $B^k \tilde{x}_t \stackrel{\Delta}{=} \tilde{x}_{t-k}$, $B^k a_t \stackrel{\Delta}{=} a_{t-k}$. Поэтому выражение

$$\theta(B) \tilde{x}_t = \varphi(B) a_t \quad (4)$$

определяет стационарный процесс при условии, что все корни (действительные и комплексные) характеристического уравнения $\theta(B) = 0$ ($|\lambda_i| \leq 1$) лежат внутри единичного круга. Аналогично, чтобы процесс был сходящимся, корни уравнения $\varphi(B) = 0$ тоже должны лежать внутри единичного круга ($|\phi_i| \leq 1$).

Если $\check{\theta}(B)$ - нестационарный оператор авторегрессии, такой, что d корней уравнения $\check{\theta}(B) = 0$ равны единице, а остальные B лежат вне единичного круга (так как уравнение $1 - \phi_i B = 0$ имеет корень $B = \phi_i^{-1}$ - условие устойчивости $|\phi_i| \leq 1$ эквивалентно $|B| \geq 1$), тогда модель (4) можно представить в виде:

$$\check{\theta}(B) \tilde{x}_t = \theta(B)(1 - B)^d \tilde{x}_t = \varphi(B) a_t, \quad (5)$$

в котором имеется $p+q+3$ неизвестных параметра: $d; \bar{x}, \theta_1, \dots, \theta_p, \varphi_1, \dots, \varphi_q; \sigma_a^2$, оцениваемых по наблюдениям.

Поскольку $\nabla^d \tilde{x}_t = \nabla^d x_t$, $\nabla = (1 - B)$ для $d \geq 1$, мы можем представить модель в виде :

$$\theta(B)\nabla^d x_t = \varphi(B)a_t,$$

Эквивалентное определение процесса можно задать двумя уравнениями:

$$\begin{cases} \theta(B)\varpi_t = \varphi(B)a_t \\ \varpi_t = \nabla^d x_t \end{cases}, \quad (6)$$

Модель *ARIMA* (5) отвечает предположению, что d -я разница ряда может быть представлена стационарным сходящимся процессом *ARMA*(p, q) (6).

Если в (5) оператор авторегрессии имеет порядок p , взята d -я разница, и оператор скользящего среднего имеет порядок q , то имеем модель *ARIMA*(p, d, q).

1.2.1 Мультипликативная модель сезонного ряда по методу Бокса-Дженкинса

Фундаментальным фактом, который относится к сезонным временным рядам, является сходство наблюдений, разделенных интервалом l (для квартальных данных $l=4$; для ежегодных $l=12$, для ежемесячных $l \in \{30, 31, 28\}$).

Следовательно, операция $B^l x_t = x_{t-l}$ будет играть особенно важную роль в анализе сезонных рядов дальше. Поскольку в ряду $x_t, x_{t-l}, x_{t-2l} \dots$ может быть нестационарность, то полезным является упрощающий оператор $\nabla_l x_t = (1 - B^l)x_t = x_t - x_{t-l}$. Устойчивый нестационарный оператор $(1 - B^l)$ имеет l нулей $e^{i(2\pi k/l)}$, $k = 0, l-1$, $i = \sqrt{-1}$, равномерно распределенных на единичном круге. Прогнозирующая на l шагов функция удовлетворяет уравнению $(1 - B^l)\hat{x}_t(l) = 0$ и потому может (но не обязана) быть представлена в виде набора синусоид и косинусоид:

$$\hat{x}_t(l) = b_0^t + \sum_{k=1}^{[l/2]} [b_{1k}^{(t)} \cos(2\pi kl/l) + b_{2k}^{(t)} \sin(2\pi kl/l)],$$

где b – подстраиваемые коэффициенты; $[l/2] = 0.5l$ для четных l , и $[l/2] = 0.5(l-1)$ для нечетных l .

Тогда, связь каждого значения временного ряда, взятого через сезонный интервал времени может быть описана моделью вида:

$$\Phi(B^l)\nabla_l^D x_t = \Theta(B^l)\alpha_t, \quad (7)$$

где t - значение времени в какой-то конкретной сезонной точке времени; $\nabla_t = 1 - B^t$ и $\Phi(B^t), \Theta(B^t)$ - полиномы B^t степеней P и Q соответственно, удовлетворяющие условиям стационарности и сходимости.

Для того, чтобы отобразить связь точек временного ряда, которые находятся внутри сезонного периода строится модель вида:

$$\theta(B)\nabla^d \alpha_t = \varphi(B)a_t, \quad (8)$$

$\theta(B), \varphi(B)$ - полиномы B степеней p и q соответственно, удовлетворяющие условиям стационарности и сходимости.

Подставив (8) в (7), получим общую мультипликативную модель:

$$\theta_p(B)\Phi_P(B^t)\nabla^d \nabla_t^D x_t = \Theta_Q(B^t)\varphi_q(B)a_t \quad (9)$$

Индексы p, P, q, Q введены для того, чтобы отличать порядки разных операторов. Говорят, что результирующий мультипликативный процесс имеет порядок $(p, d, q) \times (P, D, Q)_t$. Аналогичные рассуждения можно использовать для получения моделей с тремя и более периодическими компонентами, которые учитывают многообразие сезонных явлений. Отметим, что (9) есть нелинейная авторегрессионная модель.

1.2.2 Построение модели типа ARIMA(p,d,q) по реализации случайного процесса

Бокс и Дженкинс предложили несколько этапов:

I этап. а) Установить порядок интеграции d , то есть будем брать конечную разницу от x_t столько раз, сколько необходимо, чтобы обеспечить стационарность, в надежде свести процесс, который изучается, к смешанному процессу авторегрессии и скользящего среднего. Значительные изменения уровня (сильные скачки вверх или вниз) обычно требуют взятия несезонной разницы первого порядка (лаг=1). Сильные изменения наклона предполагает взятия разницы второго порядка. Для сезонной составляющей необходимо взятия соответствующих сезонных разностей. Если есть медленное спадание частных коэффициентов автокорреляции в зависимости от лага, обычно берут разницу первого порядка. Следует помнить, что для некоторых временных рядов нужно брать разницы небольшого порядка или совсем не брать их. Отметим, что избыточное количество взятых разностей приводит к менее стабильным оценкам коэффициентов.

б) После этого мы получаем временной ряд, которому нужно подобрать ARMA(p,q). Исходя из поведения автокорреляционной (ACF - autocorrelation function) и частной автокорреляционной функций (PACF - partial autocorrelation function), установить параметры p и q , соблюдая принцип экономии ($p+q \leq 2$).

Если исследуемый ряд не стационарен, то его автокорреляционная функция будет расти. Если ряд стационарен, то, начиная с какого-то номера, теоретические автокорреляции будут спадать. Поэтому можно рассчитать их оценки - выборочные автокорреляции, и посмотреть, спадают они или нет. Если

ряд окажется стационарным, перейти к определению параметров p и q . Иначе – построить ряд первых разностей и проверить его на стационарность.

II этап. Оценивание коэффициентов $\theta_1, \dots, \theta_p, \varphi_1, \dots, \varphi_q$; при условии, что мы уже знаем p и q . На этапе проводится уточнение оценок коэффициентов модели с использованием статистических методов. Для оцененных коэффициентов вычисляются приближенные стандартные ошибки, которые дают возможность, при дополнительных предположениях о распределениях случайных величин X_1, X_2, \dots , строить доверительные интервалы для этих коэффициентов и проверять гипотезы об их действительных значениях с целью уточнения спецификации модели. Одним из основных методов расчета оценок коэффициентов является *квази-ньютоновский метод*.

III этап. Стандартная для эконометрического подхода процедура. По остаткам осуществляется тестирование или диагностика построенной модели. Неадекватность, обнаруженная в процессе такой проверки, может указать на необходимую корректировку модели, после чего проводится новый цикл подбора, и так далее до тех пор, пока не будет получена удовлетворительная модель. Следует сказать, что если остатки систематически распределены в ряду (например, негативные в первой части ряда и приблизительно равные нулю – во второй) или включают некоторую периодическую компоненту, то это свидетельствует о неадекватности модели. Анализ остатков чрезвычайно важен и необходим при анализе временных рядов. Процедура оценивания допускает, что: остатки (наблюдаемые минус оцененные значения) нормально распределены; остатки независимы друг от друга, то есть между ними нет остаточной корреляции.

IV этап. Использование модели, в основном, для прогнозирования будущих значений временного ряда.

1.2.3 Идентификация стационарной модели ARMA

Основной отправной точкой для идентификации стационарной модели *ARMA* является отличие поведения *автокорреляционных (ACF)* и *частных автокорреляционных (PACF)* функций рядов, соответствующих разным моделям *ARMA*. По поведению только автокорреляционной функции $\rho(s)$ трудно идентифицировать даже порядок «чистого» (без МО составляющей) процесса авторегрессии. Решению этого вопроса помогает рассмотрение поведения *частной автокорреляционной функции (PACF)* стационарного процесса X_t . Ее значение $\rho_{part}(k)$ на лаге k определяется как значение коэффициента корреляции между случайными величинами X_t и X_{t+k} , очищенными от влияния случайных величин (СВ) X_{t+1}, X_{t+k-1} .

Это соответствует тому, что $\rho_{part}(k)$ является коэффициентом при X_{t-k} в линейной комбинации случайных величин X_{t-1}, X_{t-k} , что наилучшим образом приближает случайную величину X_t . В то же время, вместо неизвестных нам действительных последовательностей автокорреляций $\rho(k)$ и

частных автокорреляций $\rho_{part}(k)$ мы можем удовлетвориться только их оценками - выборочной ACF $r(k)$, и выборочной PACF $r_{part}(k)$:

$$r(k) = \frac{\frac{1}{T-k} \sum_{t=1}^{T-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sqrt{\frac{1}{T-k} \sum_{t=1}^T (x_t - \bar{x})^2} \sqrt{\frac{1}{T-k} \sum_{t=1}^T (x_{t+k} - \bar{x}_{+k})^2}} = \frac{\hat{R}(K)}{\sigma_t \sigma_{t+k}}, \quad k = \overline{1, T-1}, \quad \text{где}$$

$$\bar{x} = \frac{1}{T} \sum_{t=1}^T x_t \quad - \text{оценка МО} - M(X_t), \quad \hat{R}(k) = \frac{1}{T-k} \sum_{t=1}^{T-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \quad - \text{оценка}$$

ковариации $R(k)$ для ACF. Получить $r_{part}(k)$ можно, заменяя входные данные в выражении для функции автокорреляции $\rho(s)$, их оценками $r(s)$, что дает систему уравнений:

$$r(j) = r_{part}(1)r(j-1) + r_{part}(2)r(j-2) + \dots + r_{part}(k)r(j-k), \quad j = 1, \dots, k,$$

и решая получающиеся уравнения для $k=1, 2, \dots, p$.

Однако проще поступить иначе, исходя из того, что $\rho_{part}(k)$ является коэффициентом при X_{t-k} в линейной комбинации случайных величин $X_{t-1} \dots X_{t-k}$, наилучшим образом приближающих случайную величину X_t их можно просто оценить методом наименьших квадратов (МНК) в модели

$$x_t = \theta_1 x_{t-1} + \theta_2 x_{t-2} + \dots + \theta_k x_{t-k} + \varepsilon,$$

в которой составляющая ε определяется как невязка: $\varepsilon = x_t - \hat{x}_t$. Где \hat{x}_t - значения модели при оцененном векторе параметров $\hat{\theta}$. Полученная в результате МНК-оценка коэффициента $\hat{\theta}_k$ и есть $r_{part}(k)$, $k=1, \dots, p$.

Таким образом, в методе Бокса-Дженкинса находят начальные оценки несезонных параметров авторегрессии $\hat{\theta}$, начальные оценки несезонных параметров скользящего среднего $\hat{\phi}$, начальную оценку дисперсии белого шума $\hat{\sigma}_\varepsilon^2$.

Если СВ X_t является стационарным процессом типа $ARMA(p, q)$ и $E(X_t^4) < \infty$ то указанные оценки \bar{x} , $\hat{R}(k)$, $r(k)$ и $r_{part}(k)$ являются состоятельными оценками для $M(X_t)$, $R(k)$, $\rho(k)$ и $\rho_{part}(k)$ соответственно. Во многих случаях поведение теоретических ACF и PACF в какой-то мере отражается и на поведении их выборочных аналогов, вычисление которых может помочь в представлении поведения теоретических ACF, PACF и решении задачи идентификации соответствующих моделей в рамках общего класса моделей ARMA. По свойствам ACF и PACF судят о моделях для некоторых популярных стационарных временных рядов AR, MA, ARM, SAR, SMA (season moving average).

Для определения параметров p, d, q, i, P, D, Q , пакет *STATISTICA* предполагает многочисленные подгонки путем перебора. В результате подгонок оценивается стохастическая модель. Выходной информацией являются значения вышеперечисленных оценок на каждой итерации, а также следующая информация на последней итерации (т.е. когда процедура привела к сходящемуся результату или поиск оказался неудачным, или число выполненных итераций достигло заданного максимального числа, а сходимости – не достигнута): остаточные ошибки, соответствующие оценкам наименьших квадратов $\varepsilon_t, t=1, \dots, T$; оценка остаточной дисперсии (белого шума) $\hat{\sigma}_\varepsilon^2$; ковариационная матрица оценок; корреляционная матрица оценок; оценка общей константы φ_0 ; остаточные автокорреляции; χ^2 - статистика.

Построение сезонных моделей порядка $(p, d, q) \times (P, D, Q)_t$ не привело к удовлетворительным прогнозам. При получении прогнозов по моделям экспоненциального сглаживания мы увидели, что модели без сезонной составляющей имели более точные результаты.

При построении моделей *ARIMA*(p, d, q), во избежание большого перебора, для выбора параметра авторегрессии p мы воспользовались критериями: *AIC* (*Akaike Information Criterion*), предложенным *Акаике* и *BIC* (*Bayesian Information Criterion*), предложенным *Шварцом*. Оба критерия построены приблизительно одинаковым способом. Информационный критерий *AIC* для модели *ARMA*(p, q) выглядит таким образом:

$$AIC(p, q) = \ln \hat{\sigma}_\varepsilon^2 + 2 \frac{p+q}{N-n},$$

где $(N-n)$ - число наблюдений (точек), участвовавших в построении модели, n - число точек экзамена. Критерий Шварца имеет вид:

$$BIC(p, q) = \ln \hat{\sigma}_\varepsilon^2 + \frac{p+q}{N-n} \ln(N-n),$$

где значения $\hat{\sigma}_\varepsilon^2 = \frac{1}{N-n-p-q} \sum_{t=1}^{N-n} \varepsilon_t^2$ были рассчитаны для ограниченного числа (в наших примерах временных рядов - для девяти) моделей *ARMA*(p, q).

Известно, что оценка порядка модели по критерию *BIC* является состоятельной. Однако на практике чаще используют информационный критерий *AIC*, хотя он переусложняет модель.

Следует сказать, что метод *Бокса-Дженкинса* получил дальнейшее развитие, известное под названием *ARIMAX* на случай многофакторности модели [12]. Мы ограничились прогнозами временных рядов пакета *STATISTICA*, в котором *ARIMAX*-модели не реализованы.

1.2.4 Генерирование прогнозов для модели порядка $(p, d, q) \times (P, D, Q)_t$

Для трансформированного ряда x'_t прогнозы $x'_{N-b}(\ell)$ вычисляются по формуле:

$$x'_{N-b}(\ell) = \theta_0 + \sum_{t=1}^{p+tP+d+tD} \Phi_t^* [x'_{N-b-t+\ell}] + \sum_{t=1}^{q+tQ} \Theta_t^* [a_{N-b-t+\ell}],$$

$$\text{где } [x'_{N-b-t+\ell}] = \begin{cases} x'_{N-b}(\ell - t), & \ell > t \\ x'_{N-b-t+\ell}, & \ell \leq t, \end{cases}$$

$$[a_{N-b-t+\ell}] = \begin{cases} 0, & \ell > t \\ x'_{N-b-t+\ell} - x'_{N-b-t+\ell-1}, & \ell \leq t, \end{cases} \ell=1, \dots, n; n \leq K; K \geq p+q,$$

а значения преобразованного ряда для каждого набора параметров (d, D, λ, m) определяются:

$$x'_t = \begin{cases} (x_t + m)^\lambda, & \lambda \neq 0 \\ \ln((x_t + m)), & \lambda = 0 \end{cases}$$

где, если $\lambda \neq 0$, параметр m выбран так, что $x_t + m$ положительно для всех t , и, если $\lambda=1$, то m приравнивается нулю, так что x_t остаются неизменными.

Прогнозы находятся для каждого отодвигаемого назад момента времени $b=0, 1, \dots, n$. Прогнозы $\hat{x}_{N-b}(\ell)$ для исходного ряда находятся по формуле:

$$\hat{x}_{N-b}(\ell) = \begin{cases} x'_{N-b}(\ell), & \lambda = 1 \\ (x'_{N-b}(\ell))^{1/\lambda} - m, & \lambda \neq 0 \\ \exp\{x'_{N-b}(\ell)\}, & \lambda = 0. \end{cases}$$

2 Особенности применения МГУА для прогнозирования гетероскедастических рядов

В отличие от статистических методов, которым необходимы как можно большие и однородные выборки, в алгоритмах МГУА это не требуется, а разбиение на подвыборки используются для поиска модели со структурой и параметрами, оптимально подобранными из условия минимума дисперсии шума в эндогенной переменной. Начальная выборка W делится на три части: обучения A , проверки B и экзамена C . На подвыборке A оцениваются параметры альтернативных моделей, на заданном множестве функций, а на подвыборке B выбираются одна или некоторое множество лучших моделей. Подвыборка C не участвует при построении модели, а служит для проверки полученных прогнозов.

Особенностью применения МГУА является то, что модель предсказания представляется в виде суммы двух моделей: тренда (10) и остатка (11).

Вначале определяется тренд V_t как:

$$V_t = \boldsymbol{\theta}^T \mathbf{f}(t) \quad (10)$$

где $\boldsymbol{\theta}$ - вектор параметров; $\mathbf{f}(t)$ есть вектор из различного набора функций f_i :

для СОМВИ $f_i(t) \in \{1, t, t^2, t^3, \dots\}$;

для MACSO $f_i(t) \in \left\{ \frac{1}{t}, \sqrt[3]{t}, \frac{1}{\sqrt[3]{t}}, 1, t, t^2, t^3, \dots \right\}$. Затем строится линейная по параметрам модель для остатков $\Delta_t = x_t - V_t$ с учетом эффекта автокорреляции:

$$\Delta_{t+i} = \varphi(\boldsymbol{\eta}, \Delta_t, \Delta_{t-1}, \Delta_{t-2}, \Delta_{t-3}), \quad i = \overline{1, n}. \quad (11)$$

Векторы $\boldsymbol{\theta}$ и $\boldsymbol{\eta}$ определяются методом наименьших квадратов ошибок по обучающей подвыборке A , а структура выбирается критерием регулярности, вычисляемом на проверочной подвыборке B .

Рассмотрены два варианта: 1) алгоритмы *MGVA – COMBI* и *MACSO* конструируют модели прогнозирования *без адаптации* к будущим значениям и 2) *с адаптивным прогнозом*.

Обычно поступают так: строят для каждого i -го шага свои модели прогноза:

$$\begin{aligned} \hat{x}_{t+1} &= \psi_1(\boldsymbol{\theta}_1, x_t, x_{t-1} \dots), \\ \hat{x}_{t+2} &= \psi_2(\boldsymbol{\theta}_2, x_t, x_{t-1} \dots, x_{t-l}) \dots, \\ \hat{x}_{t+n} &= \psi_n(\boldsymbol{\theta}_n, x_t, x_{t-1} \dots, x_{t-r}) \end{aligned}$$

При *адаптивном прогнозе* каждое значение на один шаг вперед, принимает участие в дальнейшем предсказании и модели последующих прогнозов ($i = \overline{1, n}$) строятся заново с учетом этих предсказаний:

$$\begin{aligned} \hat{x}_{t+1} &= \phi_1(\boldsymbol{\theta}_1, x_t, x_{t-1} \dots), \\ \hat{x}_{t+2} &= \phi_2(\boldsymbol{\theta}_2, \hat{x}_{t+1}, x_t, x_{t-1} \dots, x_{t-l}) \dots, \\ \hat{x}_{t+n} &= \phi_n(\boldsymbol{\theta}_n, \hat{x}_{t+1}, \dots, \hat{x}_{t+n-1}, x_t, x_{t-1} \dots, x_{t-r}). \end{aligned}$$

Поэтому, фактически, на каждом шаге адаптивного прогноза заново строится модель предсказания только на один шаг вперед. Этот способ не является последовательной регрессией при адаптивной обработке сигналов, к нему нельзя применить рекуррентный МНК, т.к. при добавлении к исходным данным каждого из предсказанных значений \hat{x}_{t+i} $i = \overline{1, n-1}$ заново определяются не только параметры, но и структура модели.

3 Результаты идентификации

Общая формулировка задачи такая: по прошлым значениям наблюдаемой случайной величины x_{t-q} , $q = 1, 2, \dots, r$, $t = q+1, q+2, \dots$ оценить ненаблюдаемые будущие переменные x_{t+i}^0 , $i = 1, 2, \dots, n$, в общем случае, случайные. Последовательные среднемесячные значения цен нефти марок *Brent* и *Urals* за период январь 1999 – август 2008 гг. (113 точек) есть примеры гетероскедастических временных рядов. Они получены по данным x_1, x_2, \dots, x_N , размещенным на сайте [13]. График одного временного ряда нефти *Brent*

x_{1j} , $j = 1, 2, \dots, N$ представлен на рис.1. Необходимо построить прогнозы цены нефти на 1, 2 и 3 единицы времени (месяца), т.е. $n=3$ шага.

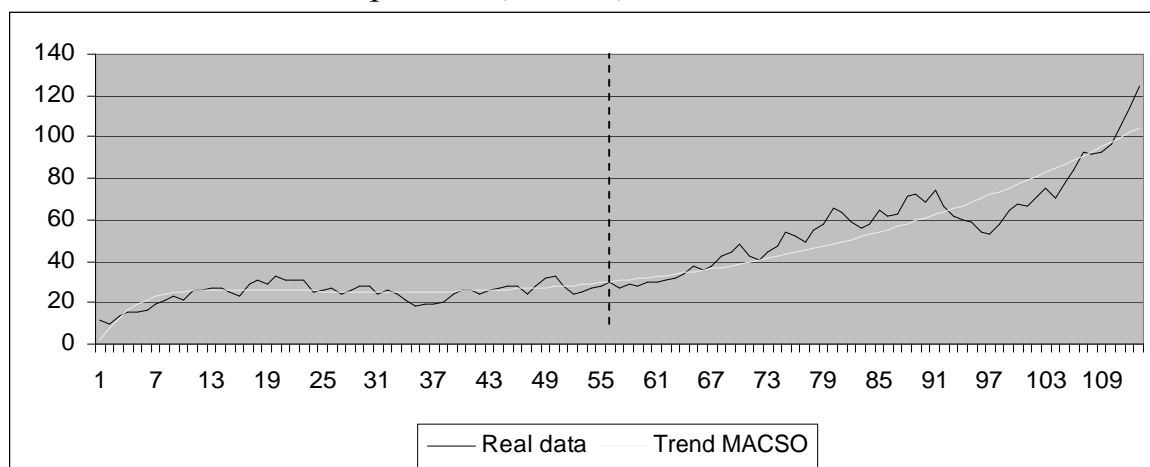


Рис. 1 Исходный временной ряд цен на нефть *Brent* и тренд по *MAKSO*

3.1 Метод Хольта

Для оценивания параметров по критерию минимальной квадратичной ошибки был использован квази-ньютоновский метод. Найденные значения для ряда нефти *Brent* $\mu=0,9$; $\nu = 0,11$. Прогнозы на 3 шага с марта по май 2008г. равны: 98,37; 100,67; 102,96; для *Urals* параметры – $\mu=0,86$; $\nu = 0,1$ прогнозы соответственно – 94,62; 96,77; 98,92.

3.2 Сглаживание с экспоненциальным трендом для ряда нефти *Brent*

Для оценивания параметров по критерию минимальной квадратичной ошибки был также использован квази-ньютоновский метод. Графики исходных данных и двух серийный прогноз по моделям изображены на рис. 2. Найденные значения равны $\mu=0,9$; $\nu = 0,12$. Прогноз *Brent* на 3 шага с марта по май 2008г: 99,22; 102,38; 105,63. Значения $\mu=1$; $\nu= 0,054$ и прогнозы для второй тройки на 1, 2, 3 шага с июня по август 2008г. равны: 128,87; 133,44; 138,18.

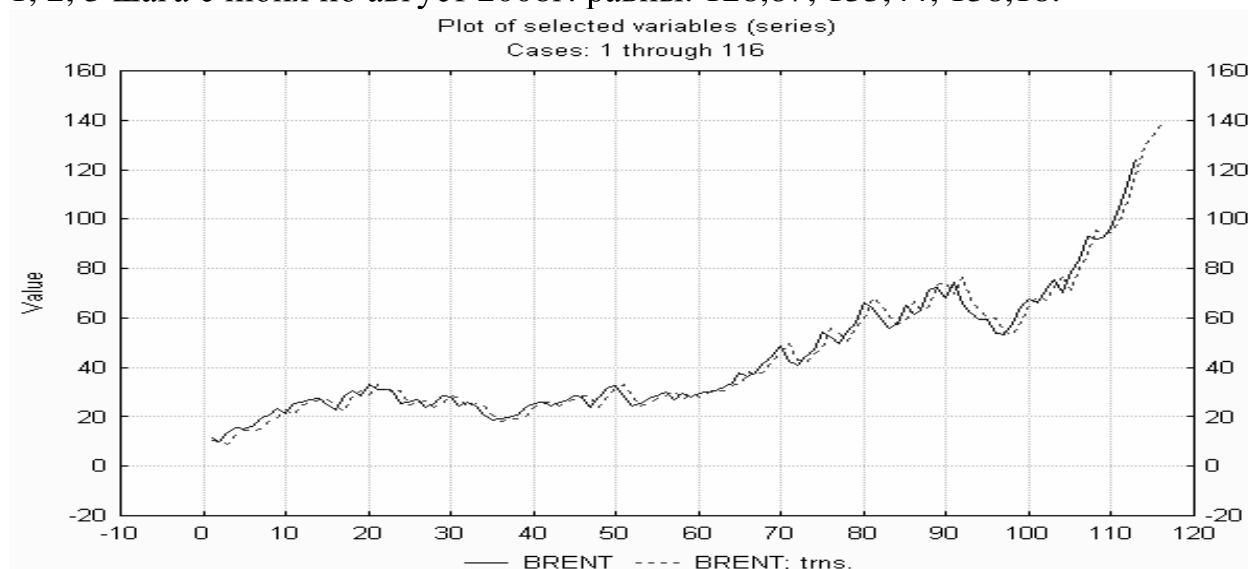


Рис. 2. Графики исходных данных цены на нефть *Brent* и два трех шаговых прогноза сглаживающей моделью с экспоненциальным трендом

3.3 Метод Хольта-Винтера

Для нахождения сезонного периода p были перебраны возможные значения $p = \overline{3,17}$ и для каждого значения методом перебора на сетке найдены значения параметров μ , ν , η , которые минимизируют сумму квадратов остатков. В методе поиска на сетке были использованы параметры: а) минимальное значение параметра: 0,1; б) максимальное значение параметра: 0,9; в) шаг: 0,05. Найденные значения для нефти *Brent* $\mu=0,9$; $\nu = 0,1$ $\eta = 0,1$; $p = 11$. Прогнозы на 3 шага с марта по май 2008г. равны: 98,7367; 96,6387; 100,6225.

3.4 Построение модели ARIMA для временного ряда цен марки Brent

На рис.1 видны четкие изменения наклона кривой, что требует взятие второй разности между элементами ряда для его стационарности.

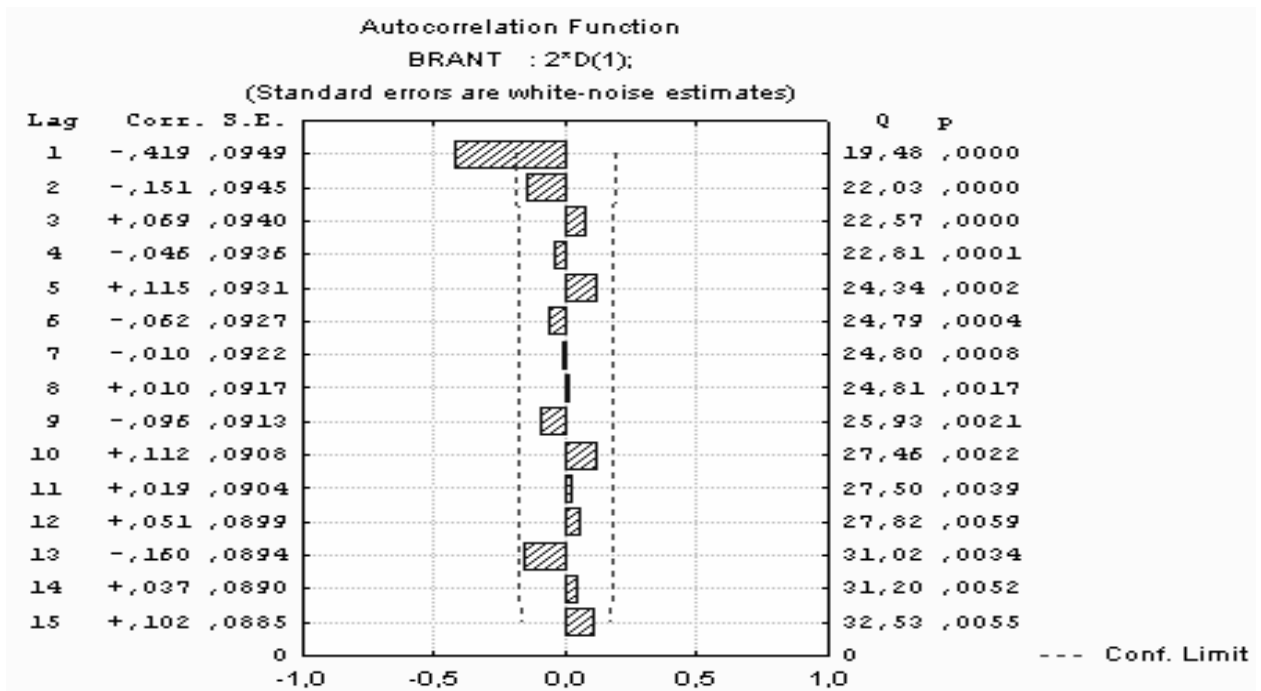


Рис.2 Автокорреляционная функция вторых разностей ряда нефти марки *Brent*

Как мы видим на рис. 2, значимой является только автокорреляция на лаге 1, что дает основания считать, что мы имеем модель скользящего среднего порядка 1.

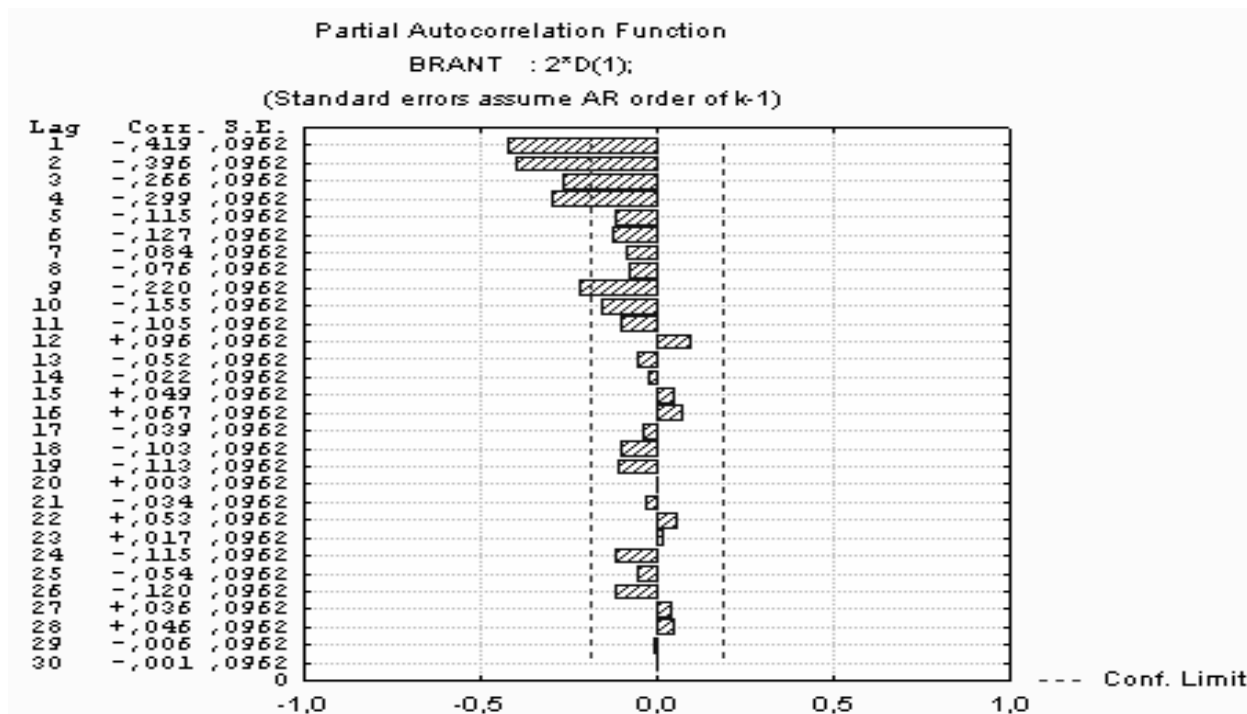


Рис. 3 – Частная автокорреляционная функция вторых разностей ряда нефти марки *Brent*

Из графика на рисунке 3 видно, что кандидатами на последующее рассмотрение являются модели с параметром p авторегрессии менее десяти. Построив возможные варианты моделей $ARIMA(1,2,1)$? $ARIMA(9,2,1)$ и вычислив значения критериев *Акаике* и *Шварца* для этих моделей, мы определили лучшего претендента.

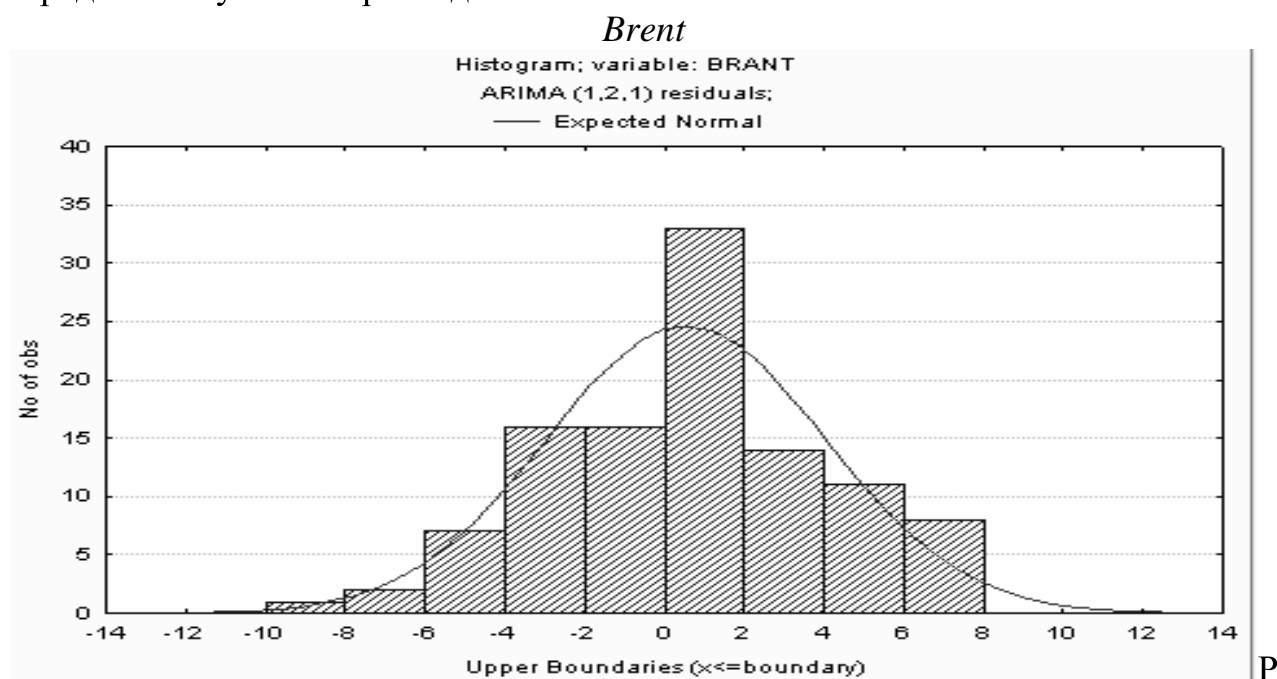


Рис. 4 – Распределение остатков модели $ARIMA(1,2,1)$

Минимум критериев AIC и BIC показал что, лучшей моделью является модель $ARIMA(1,2,1)$. Модель с оцененными коэффициентами имеет вид:

$$(1 - 0,03535B)(1 - B)^2 x_t = (1 - 0,95522B)a_t. \tag{12}$$

Анализ остатков, изображенных на рисунках 4 и 5 показывает, что модель является адекватной.

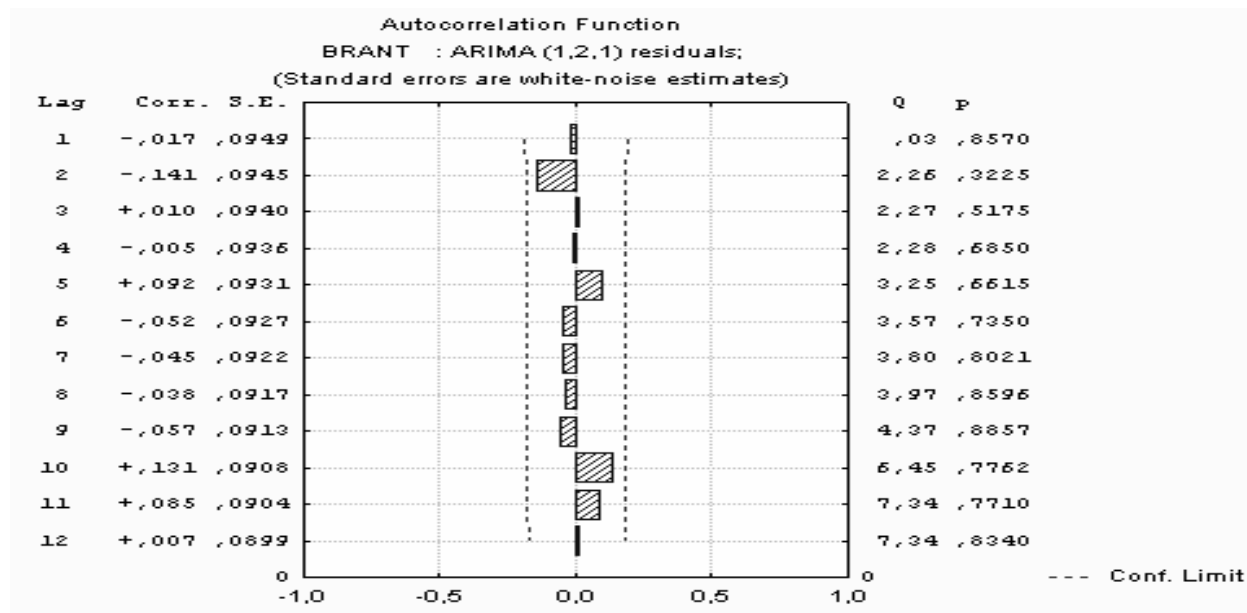


Рис. 5 – Автокорреляция остатков модели ARIMA(1,2,1)

Спрогнозированные значения ряда Brent по полученной модели (12) на $\ell=1, 2, 3$ шага вперед (март 08, апрель 08, май 08) равны соответственно: 97,84; 99,35; 100,85.

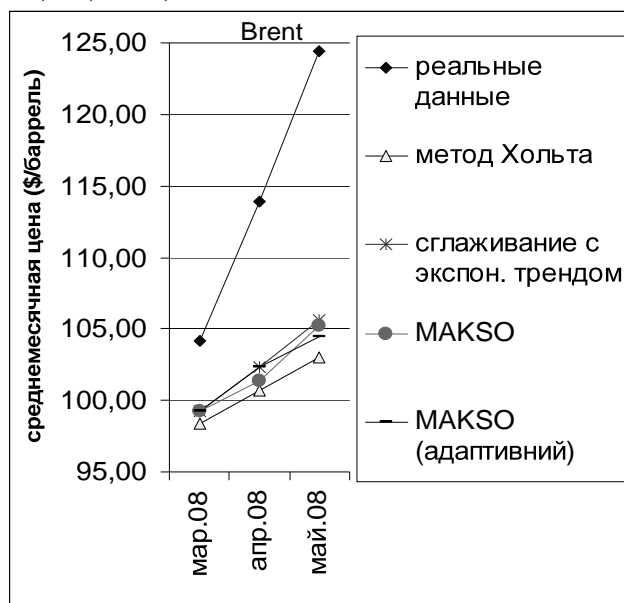


Рис. 6. Прогнозы цен на нефть Brent разными методами

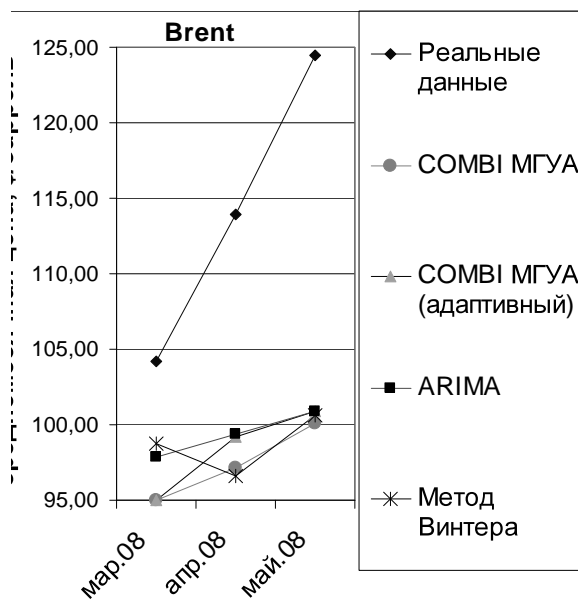


Рис. 7. Прогнозы цен того же периода на нефть Brent другими методами

Рисунки 6 и 7 представляют графики реальных данных среднемесячной цены на нефть марки Brent и этих же значений, предсказанных различными методами. Для Urals модель: $(1 + 0,0691B)(1 - B)^2 x_t = (1 - 0,97349B)\varepsilon$, прогнозы на март 08, апрель 08, май 08 соответственно равны: 93,43; 94,49; 95,55.

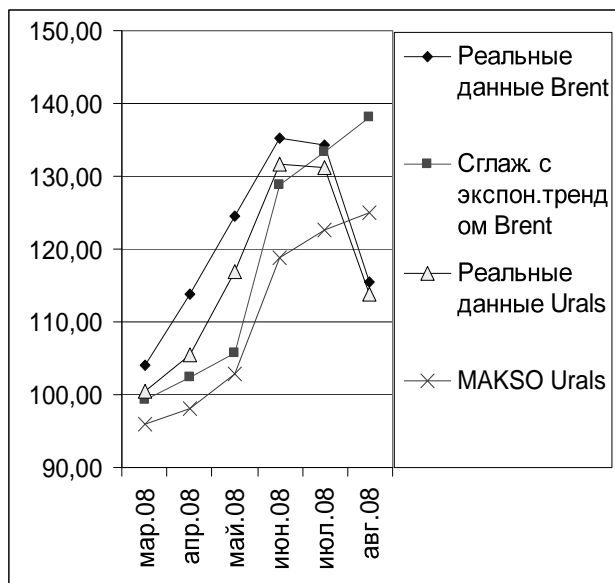


Рис. 8. Два трех шаговых прогноза цен нефти *Brent* – сглаживание с экспоненциальным трендом и цен нефти *Urals* – *MAKSO*

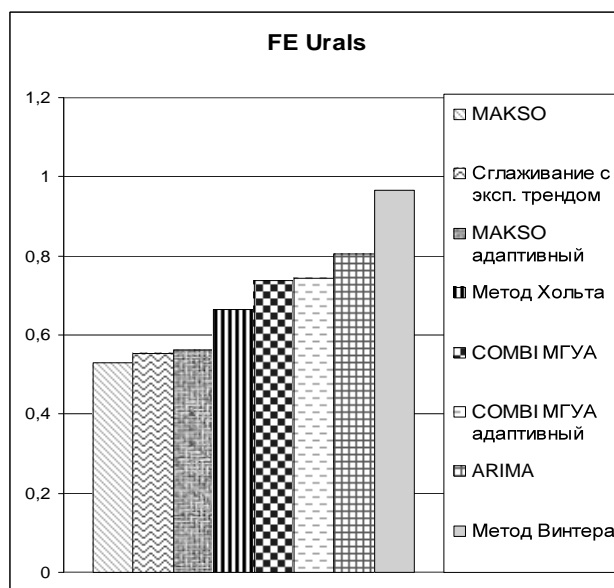


Рис. 9. Ошибка прогноза *FE* нефти *Urals* различными методами

Результатом алгоритма *MAKSO* явились достаточно сложные нелинейные модели для тренда и остатков, которые мы не приводим. Прогнозы для *Brent* на 3 шага с марта по май 2008г.: 99,23; 101,38; 105,20. Для *Urals* – с марта по май 2008г.: 96,01; 98,07; 102,77; с июня по август 2008г.: 118,79; 122,55; 125,07.

Для сравнения относительной ошибки использовались два критерия: отношение среднего модуля отклонения к диапазону изменения данных:

$$FE = \sum_{i=1}^n |x_{t+i} - \hat{x}_{t+i}^*| / n(x_{\max} - x_{\min}); \text{ критерий дисперсии } NMSE \text{ прогноза по}$$

$$\text{отношению к среднему значению } \bar{x}: NMSE = \sqrt{\frac{\sum_{i=1}^n (x_{t+i} - \hat{x}_{k+i}^*)^2}{\sum_{i=1}^n (x_{k+i} - \bar{x})^2}},$$

где x_{\max} , x_{\min} , \bar{x} определялись на n точках экзаменационной выборки. $NMSE^2$ – коэффициент остаточной дисперсии. На рисунке 8 представлены результаты двух наиболее точных методов прогнозирования для нефти *Brent* – сглаживание с экспоненциальным трендом, а для нефти *Urals* – *MAKSO MGYA* на первые – с марта по май 2008г. и последующие три шага с июня 2008 по август 2008гг..

Выводы. Из анализа методов следует, что каждый последующий описанный метод сглаживания был логическим обобщением, на основе преимущественности и усовершенствования предшествующего.

Анализ метода Бокса-Дженкинса показал, что если в пакете *STATISTICA* метод дополнить критерием штрафа за сложность моделей, таким, например, как критерий Акаике, критерий Шварца, критерий Малоуза или критерий регулярности МГУА, то можно избежать множества подгонок при окончательном выборе сложности модели. Тем более, что практическая

рекомендация авторов [3]: соблюдать принцип экономии ($p+q \leq 2$) – будет, по-видимому, при этом выполняться.

Модели всех методов за исключением сезонной модели метода Винтера правильно предсказывают тенденцию на восходящем участке прогнозируемого временного ряда. Ошибка прогноза нефти *Brent* различными методами по критерию *FE* изменяется в пределах от 0,53 до 0,83 и по критерию дисперсии *NMSE* – в пределах от 1,58 до 2,16, а у нефти *Urals* – *FE* варьирует в пределах 0,56?0,97 (см. рис. 9), *NMSE* – в пределах от 1,40 до 2,11 и является низкой. Наиболее точные прогнозы получены методами, строящими нелинейные модели трендов: сглаживание с экспоненциальным трендом и *МГУА МАКСО*.

Ни один метод не позволяет предсказывать *изменение знака приращения* временного ряда, т.е. «перелом» тенденции (см. рис. 8). Полученные модели не пригодны для прогнозирования *гетероскедастических* рядов.

Литература

1. Ивахненко А.Г. Долгосрочное прогнозирование и управление сложными системами.- Киев: Техніка, 1975.- 312 с.
2. Электронный учебник Stat Soft 1984-2001.
<http://www.statsoft.ru/home/textbook/modules/sttimser.html>.
3. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. (Вып. 1, 2.). – М.: Мир, 1972.
4. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов. –М.: Финансы и статистика, 2003. –416с.
5. Єріна А. М. Є 71 Статистичне моделювання та прогнозування: Навч. посібник. — К.: КНЕУ, 2001. –170 с.
6. Ивахненко А.Г., Степашко В.С. Помехоустойчивость моделирования. – К.: Наук. думка, 1985. – 214с.
7. Кондрашова Н.В. Влияние способа разбиения выборки в алгоритмах МГУА на точность прогнозирования. — УСиМ.–2003.–№2.– С.128-133.
8. Ванін В.В., Павлов О.В. Розробка та застосування алгоритмів самоорганізації для моделювання складних процесів та об'єктів, що відображаються точковим каркасом // Пр. Таврійської Держ. Агротехн. Акад., Мелітополь: ТДАТА, 2004, - 4. Т.24. - С.51-56.
9. Шелудько О.И. Алгоритм МГУА с ортогонализированным полным описанием для синтеза моделей по результатам планируемого эксперимента // Автоматика. - 1974. - №5. - С.32-41.
10. *Cox D.R.* Prediction by exponentially weighted moving averages and related methods //J. of the Royal Stat. Soc. - 1961.-Vol. 2 3 . -№ 2.
11. *Cohen G,D.* A note on exponential smoothing and autocorrelated inputs // Oper. Res. - 1963. - Vol. 11. - № 3.
12. Льюнг Л. Идентификация систем. Теория для пользователя. М.: Наука, 1991. – 432с.
13. <http://info.tatcenter.ru/informer>