



ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 36 за 2007 г.



ISSN 2071-2898 (Print)  
ISSN 2071-2901 (Online)

Ю. Н. Орлов, К.П. Осминин

## Анализ нестационарных временных рядов

Статья доступна по лицензии  
Creative Commons Attribution 4.0 International



**Рекомендуемая форма библиографической ссылки:** Орлов Ю. Н., Осминин К.П. Анализ нестационарных временных рядов // Препринты ИПМ им. М.В.Келдыша. 2007. № 36. 22 с.

<https://library.keldysh.ru/preprint.asp?id=2007-36>

## Введение

В данной работе предлагается новый подход к статистическому анализу временных рядов, основанный на использовании понятия эволюции выборочной функции распределения. Выборочными в общем случае могут быть как сама функция распределения (по набору данных за конкретный временной отрезок), так и ее эволюция, рассматриваемая либо как скользящее изменение функции распределения, либо как ее изменение при переходе от одного набора данных к другому, полученному на непересекающемся с первым временном отрезке. Эволюция выборочной функции распределения представляет собой прогноз временного ряда, поэтому предлагаемый метод может быть применен для построения прогнозных систем и нахождения различных индикаторов вероятности наступления того или иного события.

Работа методологически разделяется на две части. В первой части проводится стандартный корреляционный анализ временного ряда, во второй изучаются выборочные функции распределения и вводится понятие  $\varepsilon$ -стационарности выборки.

В качестве примера в работе рассматривается временной ряд, имеющий сильно выраженную циклическую составляющую: это ряд, образованный из индексов почасовых цен на электрическую энергию на оптовом рынке электроэнергии (мощности). Интерес к этому объекту вызван как его практической важностью, так и характерными свойствами указанного временного ряда, требующими разработки специфических методов анализа.

Практическая важность задачи определяется растущим уровнем либерализации оптового рынка электроэнергии и мощности в РФ, в связи с чем все более актуальной становится задача прогнозирования почасовых цен на электроэнергию в заданном регионе, по энергосистеме или в хабе. Хотя в настоящее время доля объемов, торгуемых по нерегулируемым ценам, составляет всего 8-10% от общего торгующегося объема, в будущем эта доля будет планомерно увеличиваться. В недалекой перспективе ожидается введение рынка производных финансовых инструментов, поэтому прогнозирование цен на электроэнергию на различные временные периоды (на сутки, неделю, месяц, квартал, либо год вперед) будет являться неотъемлемой частью аналитической работы по исследованию рынка.

Следует указать, что отличие рынка электроэнергии от рынка ценных бумаг (например, от биржевой торговли акциями) принципиально: спрос на электроэнергию имеет весьма высокую неслучайную составляющую, определяемую технологическими процессами выработки электроэнергии и суточной периодичностью спроса. Кроме того, спрос является сильно неэластичным (нелинейно зависит от уровня цен). На спрос оказывают влияние различные внешние факторы, такие, как суточная температура, сезонность, структура потребителей по секторам хозяйственной системы региона и др. На предложение влияют такие факторы, как гидрологическая ситуация, режимы работы ГЭС и ТЭЦ, графики ремонтных работ и др.

Таким образом, временной ряд, представляющий собой почасовые цены на электроэнергию в определенной области, например, в данной энергосистеме, наряду со случайной составляющей, определяется большим числом детерминированных процессов, которые могут быть описаны в рамках тех или иных физических моделей. Здесь следует сказать, что всякая физическая модель основана на некоторой идеализации реального процесса, из чего следует, что она изначально не точна. Поэтому следует определить предельно допустимый уровень точности детерминированного описания, выделив соответствующую составляющую из предоставляемых фактических данных и изучив чисто статистическое поведение «остатка» временного ряда. Желательно, чтобы точность описания неслучайной составляющей не превышала среднеквадратичного отклонения случайной составляющей, в силу чего методика определения последней представляет собой определенный практический интерес.

Подчеркнем также, что прогноз цен на электроэнергию наряду с чисто рыночным аспектом имеет и практическую важность для оптимизации стратегии управления и хеджирования рисков реально функционирующих производственных систем как производителями электроэнергии, так и ее потребителями.

## 1. Статистические свойства рассматриваемого временного ряда

В качестве исходного временного ряда рассматривается индекс системной равновесной цены на электроэнергию Европейской ценовой зоны в течение семи месяцев – с ноября 2006 года по май 2007 года включительно. Данный ряд находится в открытом доступе [1].

Из особенностей ряда следует отметить его естественную вложенную циклическую структуру суточной и недельной периодичности, а также зависимость от месяца года. Шагом ряда является 1 час. Для удобства анализа недельных циклов рассматриваются данные за  $w = 31$  полную неделю: с 30 октября 2006 г. (понедельник) по 03 июня 2007 г. (воскресенье). Всего, таким образом, анализируются  $N=5208$  значений; на каждый час одного из дней недели приходится  $d=217$  значений, что соответствует количеству дней в рассматриваемом периоде. Переход на летнее время отражается добавлением второго часа к 25 марта, при этом его значение берется из предыдущего часа.

Введем следующие обозначения.

1.  $p(i, j, k)$  есть значение ряда в час  $i$  – от 0 до 23 часов – в течение суток, нумеруемых индексом  $j$  – от 1 до 7, в течение  $k$  недель – от 1 до  $w$ .
2.  $\bar{p}_{k,l}(i, j)$  есть средний недельный график ряда, т.е. значения ряда  $p(i, j, n)$ , усредненные по периоду времени от  $k$ -ой недели до  $l$ -ой недели включительно;  $\bar{p}(i, j)$  есть средний недельный график за весь период наблюдений.
3.  $p_m(i)$  есть средний суточный график временного ряда, усредненный по данным в течение  $m$ -го месяца;

4.  $\bar{p}(i)$  есть средний суточный график за весь период наблюдений,  $i$  - номер часа (от 0 до 23).

5.  $c(t)$  - фактическое значение ряда в час  $t$  ( $t$  меняется от 1 до 5208).

6.  $x_m(t) = c(t) - p_m(t : \text{mod} 24)$  или  $x(t) = c(t) - \bar{p}(t)$  - остаток ряда в час  $t$ .

На Рис. 1. приведен для примера средний недельный график за четыре недели декабря 2006 г., т.е. в терминах  $\bar{p}_{k,l}(i, j)$  это график  $\bar{p}_{10,13}(i, j)$ .

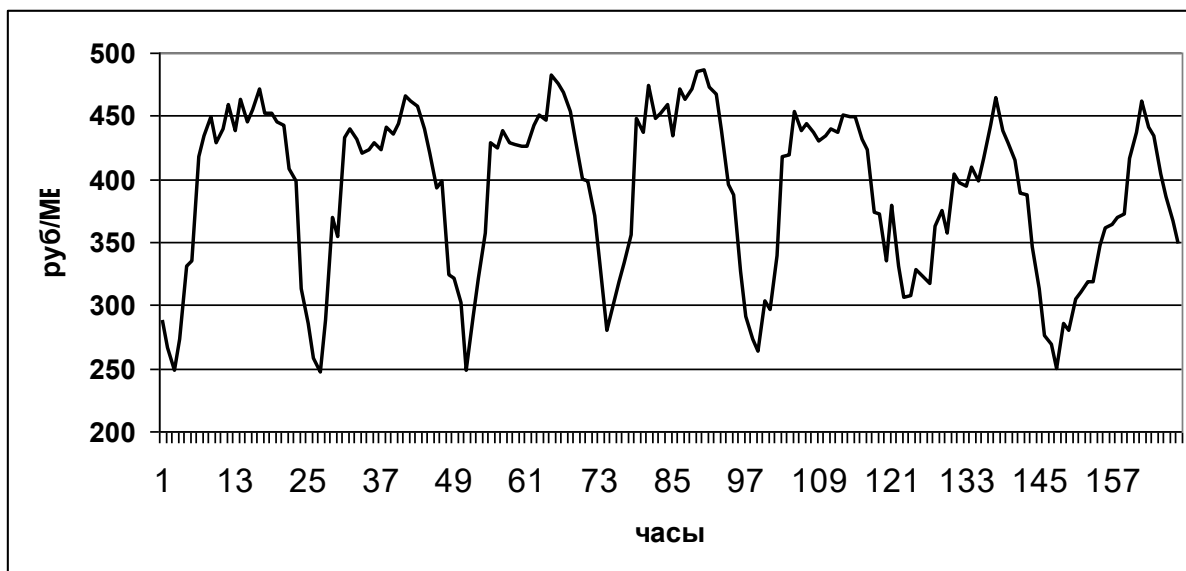


Рис. 1. График временного ряда за среднюю неделю декабря 2006 г.

Далее для краткости опускаем размерность цены (руб/МВт\*ч).

Из Рис. 1 следуют принципиальные выводы.

Во-первых, график имеет выраженную суточную периодичность.

Во-вторых, выходные и рабочие дни следует рассматривать отдельно, т.к. суточный профиль в эти дни существенно различается.

В-третьих, естественным суточным разбиением является не собственно период от 0 до 23 часов, а от 4 часов одних суток до 3 часов следующих. Указанный суточный период имеет три выраженных интервала для рабочего дня (от 4 до 9 часов, от 10 до 17 часов и от 18 до 3 часов следующих суток) и два для выходного (от 5 до 16 часов и от 17 до 4 часов следующих – выходных – суток, или до 3 часов, если следующий день – рабочий.)

Эти наблюдения весьма важны, поскольку они определяющим образом влияют на выбор модели, в рамках которой будет осуществляться прогнозирование.

Например, первое свойство (суточная периодичность) показывает, что удобно вычлест из  $c(t)$  средний за некоторый период наблюдений (неделю, месяц, год) суточный график  $\bar{p}(t)$  и рассматривать получившиеся остатки временного ряда как самостоятельный ряд. После этого составляется модель прогнозирования отклонений фактических значений ряда от их средних значений за выбранный период времени.

С другой стороны, разбиение временного ряда на подмножества, разделенные хронологически (например, только выходные дни), делает не

вполне корректным их изучение как статистического набора последовательных случайных событий. В этом случае методически более правильно прогнозировать не сам ряд, а скорость его изменения. Если же прогнозировать непосредственно сам временной ряд, то следует вводить отсутствующие в нем корректирующие факторы при переходе от одного подмножества ряда к другому (от выходных дней к будням).

Рассматриваемый ряд  $c(t)$  характеризуется следующими параметрами. Среднее значение за весь рассматриваемый период (31 неделя) равно

$$\bar{p} = \frac{1}{N} \sum_{t=1}^N c(t) = 472,29. \quad (1)$$

Относительное среднеквадратичное отклонение (ОСО или вариация):

$$\sigma = \frac{1}{\bar{p}} \sqrt{\frac{1}{N} \sum_{t=1}^N (c(t) - \bar{p})^2} \cdot 100\% = 33,3\%. \quad (2)$$

ОСО конкретного часа  $\sigma(i)$  не является постоянной величиной, а зависит от времени суток. Эта величина определяется как

$$\sigma(i) = \frac{1}{\bar{p}(i)} \sqrt{\frac{1}{d} \sum_{j=1}^7 \sum_{k=1}^w (p(i, j, k) - \bar{p}(i))^2}, i = 0...23. \quad (3)$$

Для формирования гипотезы относительно функции распределения случайной величины  $x(t) = c(t) - \bar{p}(t)$  полезно рассмотреть центральные моменты высших порядков:

$$\mu_k(i) = \frac{1}{d} \sum_{j=1}^7 \sum_{k=1}^w \left( \frac{p(i, j, k)}{\bar{p}(i)} - 1 \right)^k. \quad (4)$$

В этих терминах  $\sigma(i) = \sqrt{\mu_2(i)}$ . Если на основе расчета по формулам (4) величин  $(\mu_k(i))^{1/k}$  будет возможно принять некоторую гипотезу относительно характеристической функции, определяемую моментами  $\mu_k(i)$ , то это позволит дать оценку и самой функции распределения. Соответствующее утверждение носит название леммы Эссена (см. [2], стр. 33, а также [3], стр. 299). Согласно лемме, если  $F(x)$  и  $G(x)$  - две функции распределения, причем  $G(x)$  дифференцируема, и  $f(t)$  и  $g(t)$  - соответствующие характеристические функции, то

$$\sup_x |F(x) - G(x)| \leq \frac{1}{\pi} \int_{-T}^T \left| \frac{f(t) - g(t)}{t} \right| dt + \frac{24}{\pi T} \sup_x |G'(x)|. \quad (5)$$

Отметим, что вторые центральные моменты весьма заметно различаются для разных часов. Так, наименьшее ОСО за рассматриваемый период

соответствует 22 часам и составляет 12,5%, а наибольшее – для 3 и 4 часов, равное 45,3%. Такое большое различие между среднеквадратичными отклонениями (более чем трехкратное) показывает различие статистических свойств ряда даже в течение суточного интервала времени. Диаграмма ОСО в зависимости от часа показана на Рис. 2.



Рис. 2. Относительное среднеквадратичное отклонение суточных данных в зависимости от часа, %

Среднесуточные значения ряда  $c(t)$  в среднем за месяц и за весь период наблюдений показывают также наличие и недельной структуры (Рис. 3): слабо выраженный максимум в середине недели, явный минимум – в воскресенье.

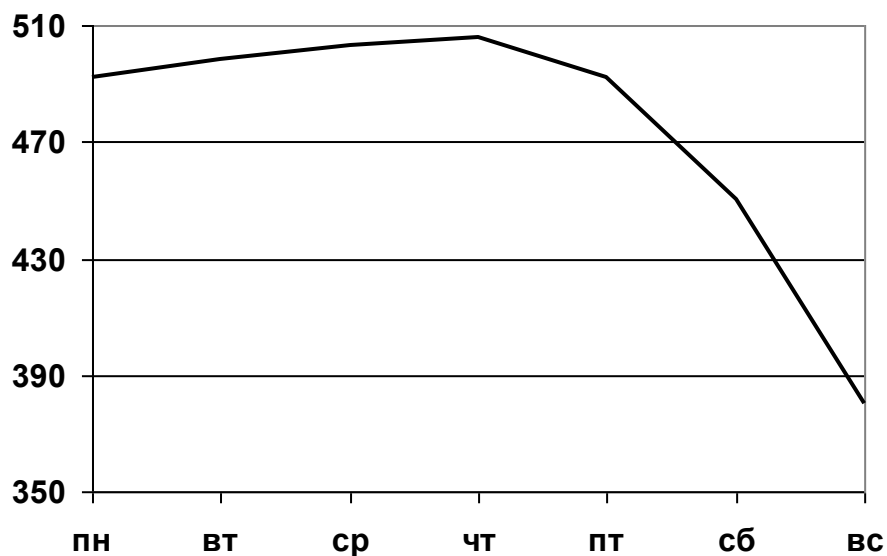


Рис. 3. Значения ряда за 7 рассматриваемых месяцев, усредненные по дням недели

Эти свойства ряда показывают, что его эволюцию правильно было бы описывать в терминах многомерных функций распределения, но для такого описания требуется достаточно репрезентативная выборка, чего на практике может не быть. Поэтому далее, в заключительном разделе данной работы, мы предложим упрощенный вариант анализа на основе одномерной выборочной функции распределения остатков, полученных после исключения суточной периодичности. Более тонкой, но практически важной задачей является исключение долговременных слабо выраженных колебаний без увеличения выборочной дисперсии вновь получаемых остатков. Эта задача будет рассмотрена в отдельной работе.

Для определения оптимальной модели аппроксимации данных рассмотрим выборочную автокорреляционную функцию  $F_N(n)$  временного ряда  $p(i)$ , определяемую как (см. [4, 5])

$$F_N(n) = \frac{(N-n) \sum_{k=1}^{N-n} p_k p_{k+n} - \sum_{k=1}^{N-n} p_k \sum_{j=1}^{N-n} p_{j+n}}{\sqrt{\left( (N-n) \sum_{k=1}^{N-n} (p_k)^2 - \left( \sum_{k=1}^{N-n} p_k \right)^2 \right) \left( (N-n) \sum_{k=1}^{N-n} (p_{k+n})^2 - \left( \sum_{k=1}^{N-n} p_{k+n} \right)^2 \right)}}. \quad (6)$$

Ее график (Рис. 4) имеет незатухающие периодические максимумы и минимумы амплитудой от  $-0,3$  до  $0,7$  с интервалом в 24 часа, что отражает суточную периодичность исходных величин. Кроме того, четко выражена недельная периодичность указанных экстремумов, максимальные значения которых приходятся на значения, кратные 168, что соответствует количеству часов в неделе. Менее четко, но все же заметна месячная периодичность как слабовыраженная длинноволновая модуляция амплитуды функции  $F_N(n)$ . Это означает, что естественным крупномасштабным промежутком усреднения является месяц, более мелким – неделя, и, наконец, сутки. В соответствии с этим делением ставится задача о прогнозе на указанных промежутках.

Автокорреляционная функция скорости, т.е. первых разностей данных  $v(n) = c(n+1) - c(n)$ , также имеет вид квазипериодической функции с явно выраженным суточным периодом и внутрисуточной структурой, отвечающей трем интервалам суточного разбиения (утро, день, вечер). Ее амплитуда не убывает с ростом шага корреляции и заключена в пределах от  $-0,2$  до  $0,5$ , как показано на Рис. 4.

Тот факт, что автокорреляция указанных величин не стремится к нулю с ростом шага, означает наличие неслучайной периодической составляющей, подлежащей исключению.

Рассмотрим статистику остатков временного ряда (отдельно для рабочих и выходных дней), полученных вычитанием из соответствующих величин  $p(i, j, k)$  среднего суточного графика  $\bar{p}(t \bmod 24)$ , определяемого усреднением по различным временным промежуткам. Важно отметить, что увеличение набора данных не обязательно приводит к уточнению статистики остатков.

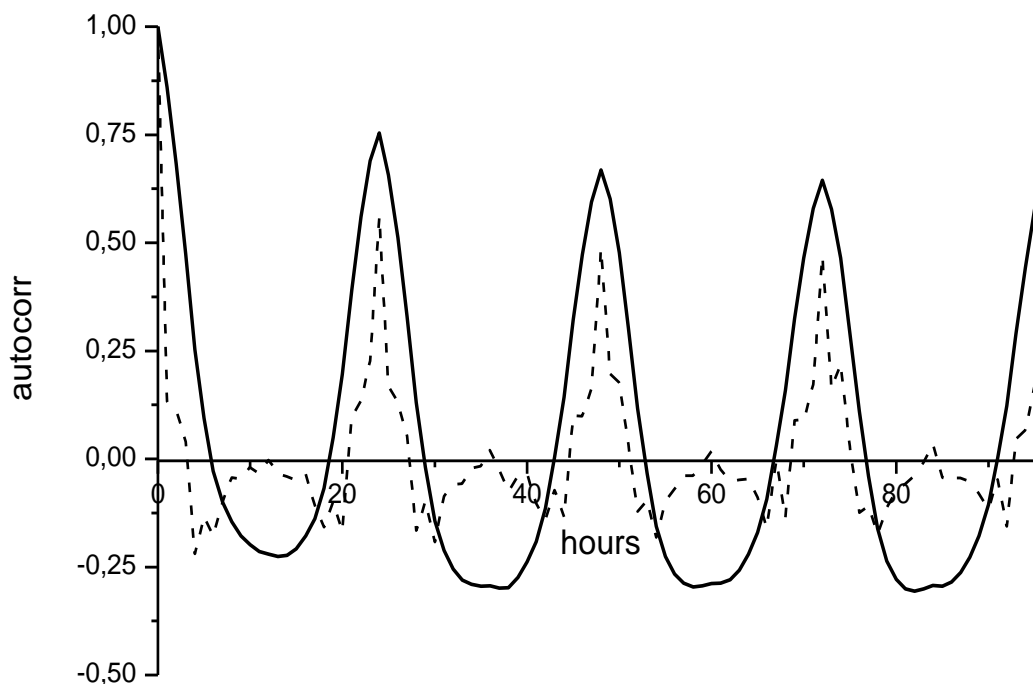


Рис. 4. Автокорреляционные функции временного ряда (сплошная кривая) и его первых разностей (пунктирная кривая) для рабочих дней декабря

На Рис. 5 показаны усредненные по дням недели значения относительных среднеквадратичных отклонений остатков от среднего суточного значения за весь период (7 месяцев) в целом и за один месяц, средний суточный график брался отдельно для рабочих и выходных дней.

Из графика Рис. 5 видно, что имеется статистически значимое повышение точности при усреднении данных в течение одного месяца по сравнению с усреднением по семи месяцам. В совокупности по рабочим дням (аналогично для выходных) ОСО для всего набора данных составило  $\sigma_w = 23,5\%$ , тогда как при усреднении в течение одного календарного месяца (соответственно, с ноября по май) получались следующие значения: 17,4%, 14,3%, 19,8%, 15%, 21,8%, 15,4% и 16,8%, так что среднеарифметическое среднемесячное ОСО составило 17,2%, а среднеквадратическое ОСО 17,4%.

Следует сказать, что среднеарифметическое ОСО для оценки точности помесечного усреднения более адекватно отражает свойства принятого усреднения, чем среднеквадратичное, поскольку в нашем случае одним статистическим экспериментом является определение ОСО за указанный период времени (1 месяц). Средним значениям по таким экспериментам будет именно среднее арифметическое найденных ОСО.

Еще меньшее значение ОСО получается при усреднении данных по рабочим дням в течение одной недели. Хотя в отдельные недели ОСО может превышать аналогичную среднемесячную величину, но в среднем за месяц (и за весь период в пять месяцев) получались меньшие значения. Так, за пять



месяцев среднеквадратичное ОСО на основе недельного усреднения данных по рабочим дням равно 13,0%, а его среднее арифметическое значение составило 12,1%.

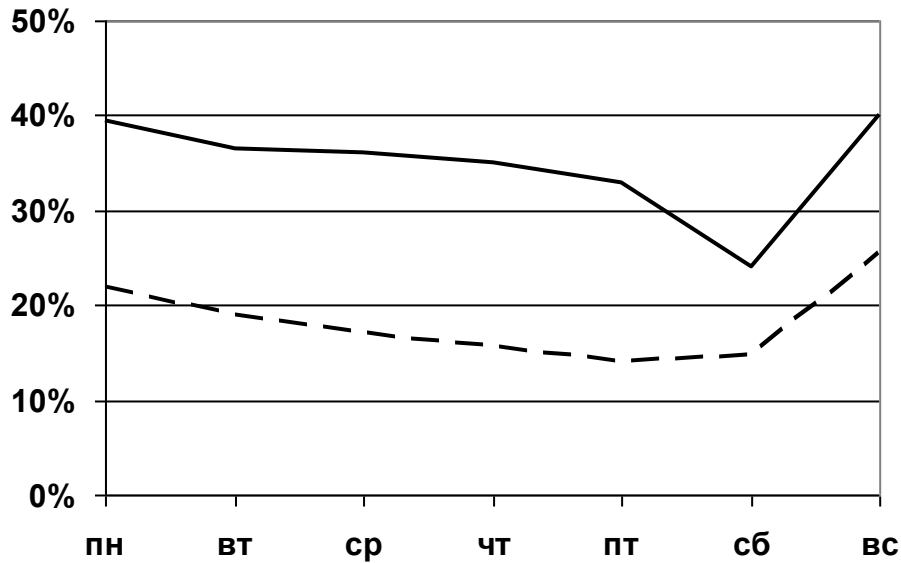


Рис. 5. ОСО по дням недели при вычитании среднего суточного графика за весь период (сплошная линия) и в среднем за месяц по месяцам за тот же период (пунктирная линия).

Из Рис. 5 также следует, что наименьший разброс данных (и наиболее точная аппроксимация) имеет место в пятницу и субботу, а наибольший разброс – в понедельник и воскресенье. Таким образом, если строить модели аппроксимации временного ряда по его моментам, то надо рассматривать ряды не только для фиксированного часа в течение суток, но и для каждого дня недели. Однако качественная зависимость ОСО от дней недели не является постоянной от месяца к месяцу, а также при переходе от одного объема выборки к другой: при других выборках наибольшее и наименьшее ОСО может приходиться на другие дни недели.

Автокорреляционная функция остатков (Рис. 6), в отличие от автокорреляции самих значений временного ряда или их первых разностей (Рис. 4), имеет уменьшающуюся амплитуду, поэтому остатки в первом приближении можно рассматривать как статистически независимые величины. Первые и вторые разности остатков в еще большей степени показывают свою статистическую независимость, но их автокорреляция хотя и имеет амплитуду меньше 0,2, обладает тем не менее явной суточной периодичностью. Корреляционный анализ остатков (или их первых и т.д. разностей) показывает, что дифференцирование ряда не может, в силу цикличности данных, полностью исключить зависимость моментов ряда от времени. Поэтому автокорреляционные модели могут применяться только на этапе качественного анализа, точность которого должна быть улучшена с использованием других подходов.

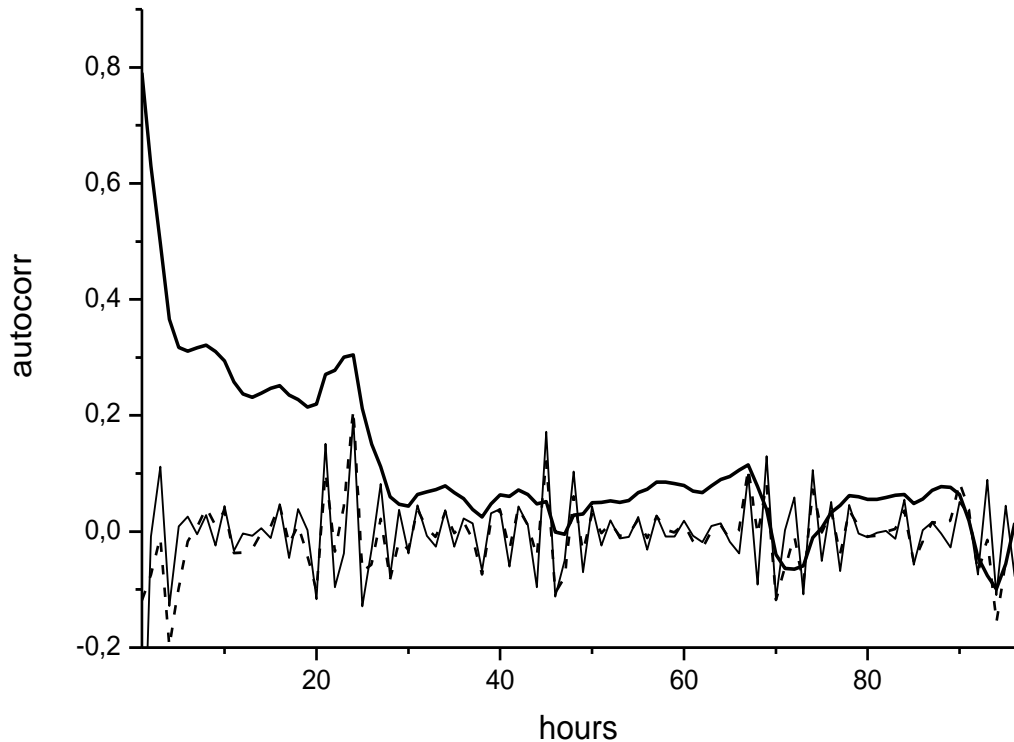


Рис. 6. Автокорреляционные функции остатков временного ряда (жирная кривая), их первых (пунктирная кривая) и вторых (тонкая кривая) разностей для рабочих дней

Наиболее детальная статистическая информация содержится в эмпирической (выборочной) функции распределения случайных величин. Рассмотрим выборочную функцию распределения (ВФР) относительных остатков, определяемых как отношение остатка к среднему значению в данный час:

$$\xi(i, j, k) = \frac{p(i, j, k)}{\bar{p}(i)} - 1. \quad (7)$$

В частности, по данным за весь период (7 месяцев) соответствующая гистограмма распределения ВФР  $f(\xi)$  представлена на Рис. 7. Максимум ВФР приходится на точку  $-0,13$ . Соответствующая вероятность равна  $0,024$ . В интервале 5%-го отклонения от суточного графика содержится 20% данных, от максимума – 21%. ВФР несимметрична, её центральная часть смещена в отрицательную область, в то время как хвосты смещены в положительную; коэффициент асимметрии равен  $0,06$ . Известно [4, 6], что дисперсии выборочных коэффициентов асимметрии и эксцесса распределены нормально, и при объеме данных, больших  $n=200$ , оцениваются соответственно как  $6/n$  и  $24/n$ . Если выборочные коэффициенты больше указанных величин, то гипотеза нормальности отклоняется, что и имеет место в нашем распределении.

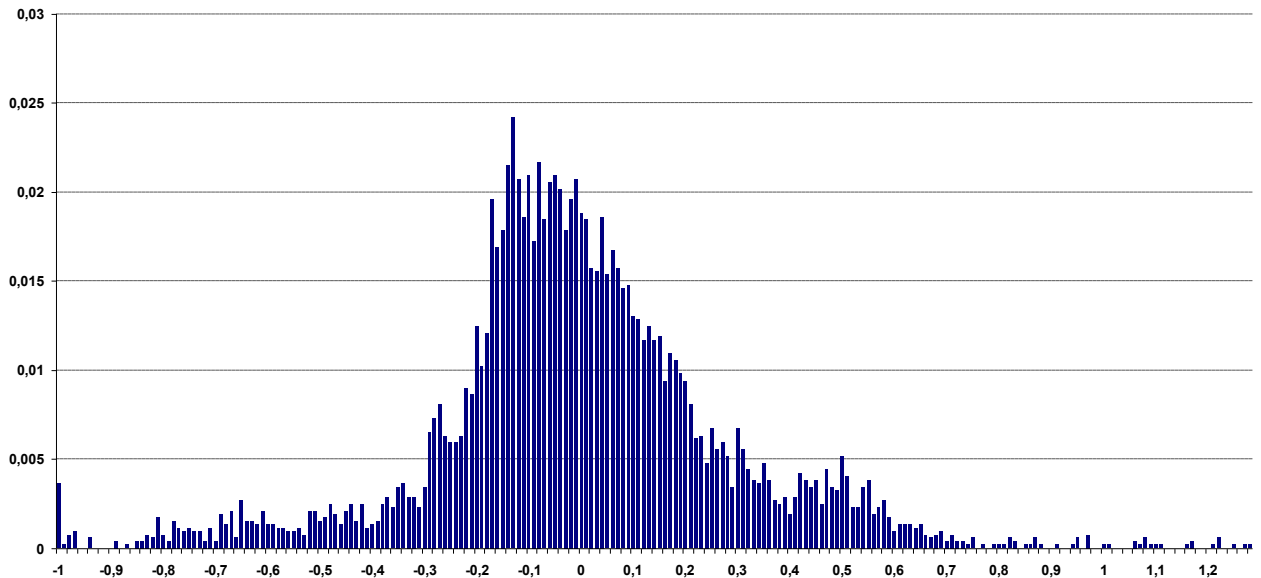


Рис. 7. ВФР, нормированная на полное количество часов (5208)

Коэффициент эксцесса, показывающий отличие данного распределения от нормального, велик: он равен 2 (для нормального распределения он равен нулю). Это означает, что ВФР убывает существенно медленнее, чем экспонента и имеет очень «толстые хвосты». Численный анализ данных [1] показал, что центральные моменты (4) этого распределения ведут себя таким образом, что  $\ln(\mu_n)^{1/n}$  асимптотически стремится к постоянной, приблизительно равной 0,53, а соответствующие величины  $(\mu_n)^{1/n}$  стремятся к 1,7.

## 2. Сравнительный анализ точности некоторых стандартных методов аппроксимации временного ряда

Существует несколько стандартных приемов исследования временных рядов и составления прогнозных моделей. Если ряд стационарный и представляет собой белый шум (БШ), то ошибка модели будет порядка ОСО, вычисленного для этого БШ. В этом случае вполне удовлетворительной на коротком промежутке времени является линейная регрессионная модель «ряд-время», которая будет иметь приблизительно ту же дисперсию, что и БШ. Напомним, что БШ – это такой стационарный процесс, Фурье-разложение которого является константой, а корреляционная функция –  $\delta$ -функция.

Если ряд нестационарен, то обычно используются модели со скользящими параметрами, которые определяются по периодически обновляемому набору вновь появляющихся значений временного ряда. Однако такие модели не всегда эффективны. Например, если временной ряд является квазистационарным на некотором промежутке времени (например, в течение одного месяца), а затем в силу внешних причин его параметры меняются, то использование данных за прошедший месяц при прогнозе на следующий будет приводить к систематической ошибке. Для прогноза тогда надо заранее знать,

как будут изменяться такие параметры, т.е. использовать дополнительные детерминистические модели.

Если есть основания считать (например, по виду автокорреляционной функции), что существует зависимость между членами временного ряда, то применяются автокорреляционные модели, в которых значение случайной величины в момент  $t+1$  определяется ее значениями в течение нескольких прошедших моментов времени. Но для слабо коррелированных случайных величин применение таких моделей дает обычно невысокую точность аппроксимации, поскольку тогда на чисто случайный процесс БШ оказывается наложенной некая не исключенная периодическая компонента, что приводит к увеличению суммарной дисперсии.

В конечном счете задачей статистического анализа данных является выявление корреляций между членами ряда, устранение из него трендов и сезонных колебаний и сведение ряда к виду БШ. Для этой цели можно применить различные итерационные методы уточнения модели, когда разность между временным рядом и его аппроксимацией, в свою очередь, рассматривается как временной ряд, к которому применяется та же (или, возможно, другая) аппроксимация и т.д.

Желание иметь дело со стационарным рядом вызвано возможностью обосновать аппроксимационные или прогнозные модели применением теоремы Вольда о разложении ([4], стр. 269), которая говорит о том, что всякий стационарный процесс может быть единственным образом представлен в виде суммы двух некоррелированных между собой процессов: детерминированного, прогноз которого на любое время вперед безошибочен, и чисто случайного БШ. Поэтому, хотя реальные процессы рассматриваемого нами происхождения, как правило, не являются стационарными, тем не менее, возникает желание в первом приближении считать их таковыми. Такой подход может дать удовлетворительный результат в задачах краткосрочного прогнозирования, горизонт которого, собственно, и определяется как период времени, в течение которого процесс можно считать стационарным с заданной точностью.

### **2.1. Регрессионная модель остатков.**

Построим прогноз на сутки для значений  $\xi(i)$ , где  $i$  меняется от 0 до 23. В качестве  $\bar{p}(i)$  в (7) можно взять среднемесячный или средненедельный суточные графики. Будем для определенности рассматривать рабочие дни. Прогнозные значения будем обозначать  $\tilde{\xi}(i)$ .

Напомним (см., напр., [4], стр. 573), что если требуется связать две величины  $Y$  и  $X$  линейной зависимостью вида  $Y = aX + b$  по имеющимся  $N$  парам значений  $(x_k, y_k)$  методом наименьших квадратов, то

$$Y - \bar{y} = a(X - \bar{x}), \quad a = \frac{\langle \Delta x \Delta y \rangle}{D^2(x)}, \quad \langle \Delta x \Delta y \rangle = \frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})(y_n - \bar{y}), \quad (8)$$

где  $D^2(x) = \langle (\Delta x)^2 \rangle$ ,  $\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n$ ,  $\bar{y} = \frac{1}{N} \sum_{n=1}^N y_n$ .

Эту зависимость, называемую регрессионной, можно записать в симметричном виде:

$$\frac{Y - \bar{y}}{D(y)} = Cor(x, y) \frac{X - \bar{x}}{D(x)}, \quad Cor(x, y) = \frac{\langle \Delta x \Delta y \rangle}{D(x)D(y)}.$$

Поскольку среднее значение рассматриваемых нами величин  $\xi(i)$  равно нулю, то из (8) следует, что линейная регрессионная модель (ЛРМ) как функция часа суток имеет вид

$$\tilde{\xi}(t) = a(t - 25/2), \quad \tilde{c}(t) = \bar{p}(t)(1 + \tilde{\xi}(t)). \quad (9)$$

Здесь  $a = cor(\xi, t)\sigma_\xi / \sigma_t$ , где  $\sigma^2$ -дисперсия соответствующих величин. Для модели аппроксимации корреляция определяется по данным за те же сутки. Ошибка аппроксимации по модели (9) равна

$$\sigma = \sqrt{\langle (\xi(t) - \tilde{\xi}(t))^2 \rangle} = \sigma_\xi \sqrt{1 - cor^2(\xi, t)}. \quad (10)$$

Если в (9) считать  $a = \bar{a} = const$  независимо от прогнозных суток, то тогда ЛРМ будет фактически давать прогноз на любые сутки. В то же время наибольший интерес представляет прогноз именно на ближайшие сутки. Следовательно, для такого прогноза необходимо моделировать сам коэффициент корреляции. При этом статистические данные для него могут браться как за скользящий фиксированный промежуток времени, так и за растущий от некоторого момента времени промежуток.

Пусть  $a_n$  есть значение параметра  $a$  при аппроксимации данных за день  $n$ . Например, для января 2007 г. эти значения представлены на Рис. 8.

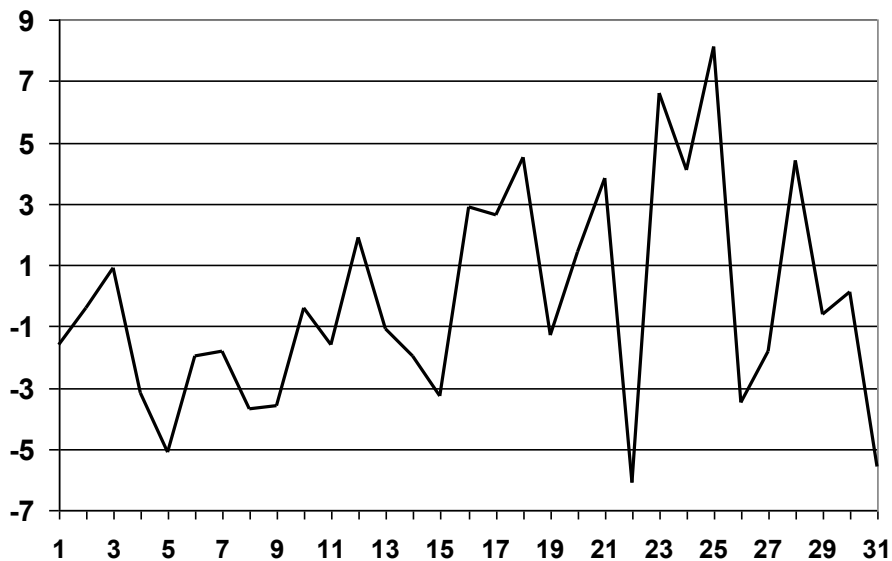


Рис. 8. Динамика параметра ЛРМ (7) по дням января 2007 г.

Видно, что именно разброс в этих параметрах и будет являться причиной увеличения ошибки аппроксимации суточных данных на месячном интервале времени. Если существует корреляция  $cor(a,b)$  между величинами  $a_n$  и, например, значениями некоторого параметра  $b_{n-1}$  на предыдущем шаге, то ошибка аппроксимации уменьшается путем оценки  $a_n$  в рамках той же регрессионной модели (7):

$$\tilde{a}_n = \bar{a} + cor(a,b) \frac{\sigma_a}{\sigma_b} (b_{n-1} - \bar{b}). \quad (11)$$

Чтобы получить прогноз величин  $\xi$ , необходимо вместо месячного промежутка усреднения использовать текущий. Это может быть интервал  $\Delta_n = (n - \tau, n - 1)$  - скользящий интервал, либо  $\Delta_n = (\tau, n - 1)$  - накапливающийся интервал, где  $\tau$  - некоторый заданный момент времени. Соответствующие средние по таким промежуткам будем отмечать  $\langle a \rangle_\Delta$ . Тогда прогнозная РЛМ имеет вид

$$\tilde{\xi}(t) = \left[ \langle a \rangle_\Delta + cor_\Delta(a,b) \frac{\sigma_{a,\Delta}}{\sigma_{b,\Delta}} (b_{t-1} - \langle b \rangle_\Delta) \right] (t - 25/2), \quad (12)$$

$$\tilde{c}(t) = \bar{p}(t) (1 + \tilde{\xi}(t)).$$

При этом средний суточный график может быть получен усреднением по промежутку, не обязательно равному  $\Delta_n$ , а, по предшествующему периоду времени, дающему наилучший прогноз в среднем на неделю или месяц. В частности, поскольку из предыдущего анализа следует, что усреднение за неделю дает наименьшую дисперсию при аппроксимации, то в качестве скользящего промежутка можно взять  $\Delta = (n - 7, n - 1)$ .

Важным частным случаем является вариант выбора в качестве параметра  $b_{n-1}$  величину  $a_{n-1}$ . В этом случае получается модель, основанная на автокорреляции между значениями параметра регрессии.

## **2.2. Автокорреляционные модели**

Для применения автокорреляционных моделей (АМ) желательно иметь временной ряд, автокорреляционная функция (6) которого имеет небольшое число максимумов и достаточно быстро спадает с ростом шага автокорреляции.

Из коррелограмм, представленных на Рис. 4, следует, что АМ небольшого порядка непосредственно для членов ряда или его остатков неприемлемы. Рассмотрим автокорреляционную функцию первых разностей отклонений почасовых значений ряда от среднемесячного суточного графика  $z(n) = c(n+1) - c(n) + \bar{p}(n) - \bar{p}(n+1)$ . Из графика Рис. 6 следует, что данные суточного графика можно приближенно считать независимыми или слабо связанными с «ближайшими соседями». В этом случае можно попытаться ограничиться АМ небольшого порядка, например, моделью Бокса-Дженкинса.



разностей отклонений, после чего для собственно прогноза получается рекуррентная формула

$$\tilde{c}(n) = \tilde{c}(n-1) + \bar{p}_1(n) - \bar{p}_1(n-1) + z(-1) * (F(1))^n, \quad n = 0, 1, \dots, 23, \quad (17)$$

$$\tilde{c}(-1) = c(23).$$

Для этой модели  $F(1) = -0,014$  для выходного дня и  $F(1) = -0,238$  для рабочего.

### 2.2.2. АМ второго порядка

$$x_n = \frac{[F(1)(1 - F(2))x_{n-1} + (F(2) - F^2(1))x_{n-2}]}{1 - F^2(1)} = U_1x_{n-1} + U_2x_{n-2}. \quad (18)$$

Коэффициенты модели (18) для остатков (аналог модели (14-15)) равны: для выходного дня  $U_1 = 0,949$ ;  $U_2 = -0,022$ , а для рабочего  $U_1 = 0,669$ ;  $U_2 = -0,14$ .

### 2.2.3. АМ третьего порядка

$$x_n = U_1x_{n-1} + U_2x_{n-2} + U_3x_{n-3}. \quad (19)$$

Для модели первых разностей остатков за январь 2007 коэффициенты равны: выходной день:  $U_1 = -0,0197$ ;  $U_2 = -0,0378$ ;  $U_3 = -0,15$ ;

рабочий день:  $U_1 = -0,255$ ;  $U_2 = -0,070$ ;  $U_3 = -0,067$ .

Ошибка аппроксимации данных на месяц по этой модели равна 5,6% для выходного дня и 7,4% для рабочего. Ошибка теста на феврале 2007 составила 9,5% для выходных дней и 13,9% для рабочих (по первым 10 рабочим дням ошибка равна 8,2%).

## **3. Оценка эволюции моментов временного ряда**

Преыдущий анализ показывает, что стандартные регрессионные или автокорреляционные модели имеют невысокую точность даже на уровне аппроксимации данных, поэтому ошибка их применения для прогнозирования будет еще больше. Для выбора адекватной модели, учитывающей периодическое изменение остатков ряда, следует проверить, насколько данный временной ряд можно считать стационарным. Подчеркнем, что на практике свести ряд к стационарному не всегда возможно. Существенным моментом является то, что, в отличие от идеальной теоретической ситуации, мы имеем дело с ограниченным набором данных. Поэтому, например, каждое разностное дифференцирование уменьшает размер выборки, и, следовательно, сколь угодно много раз эту операцию для исключения трендов высших порядков провести не удастся. В таком случае мы вынуждены работать с существенно нестационарными рядами.

Напомним, что случайный процесс называется стационарным в широком смысле ([4], стр. 240), если некоторое конечное число его моментов не зависит от времени. Если же в этом ключе рассматриваются сами распределения, то процесс называется стационарным в узком смысле. Нас будет интересовать в основном именно последний аспект, поскольку, например, одно и то же



среднее значение или дисперсия могут быть у существенно различных распределений. Поэтому прогноз, основанный на функции распределения, а не на первых ее моментах, объективно является более точным, хотя и требует применения более сложных процедур.

Если бы ряд из остатков или их первых (или более высоких) разностей оказался стационарным, то можно было бы, используя теорему Гливленко и критерий согласия Колмогорова [4], попытаться определить вид распределения, к которому относилась бы изучаемая выборка данных. Именно, если различные выборки с эмпирическим распределением  $F_n(x)$  принадлежат одной и той же генеральной совокупности с некоторым теоретическим распределением  $F(x)$ , параметры которого надо найти, то

$$\lim_{n \rightarrow \infty} D_n = 0, \quad D_n = \sup_x |F_n(x) - F(x)|, \quad (20)$$

причем функция распределения величины  $\sqrt{n}D_n$  стремится к  $K(z)$ :

$$\lim_{n \rightarrow \infty} P\{\sqrt{n}D_n < z\} = K(z), \quad K(z) = \begin{cases} 0, & z \leq 0 \\ \sum_{-\infty}^{+\infty} (-1)^k \exp(-2k^2 z^2), & z > 0 \end{cases}$$

Проведем анализ ряда (7) на наличие тренда, воспользовавшись критерием Фостера-Стюарта (см. [6], стр. 519, а также [7]). Строго говоря, этот критерий применяется в предположении, что случайные величины распределены нормально, но нам на данном этапе нужен лишь качественный ответ. Статистики этого критерия для обнаружения тренда в среднем значении и дисперсии используют производные случайные ряды из нулей и единиц, определяющие наличие тренда в максимумах или минимумах исходного ряда:

$$u_i = \begin{cases} 1, & x_i > x_{i-1}, x_{i-2}, \dots, x_1 \\ 0 & \text{в противном случае} \end{cases} \quad l_i = \begin{cases} 1, & x_i < x_{i-1}, x_{i-2}, \dots, x_1 \\ 0 & \text{в противном случае} \end{cases}.$$

Рассматриваются статистики

$$d = \sum_{i=2}^n d_i, \quad d_i = u_i - l_i; \quad S = \sum_{i=2}^n S_i, \quad S_i = u_i + l_i.$$

Известно [6], что при отсутствии тренда величины

$$t = \frac{d}{f}, \quad \tilde{t} = \frac{S - f^2}{l}, \quad \text{где } f = \sqrt{2 \ln n - 0,8456}, \quad l = \sqrt{2 \ln n - 3,4253}, \quad (21)$$

имеют распределение Стьюдента с  $n$  степенями свободы. Поэтому, если обозначить через  $t_\gamma$   $\gamma$ -квантиль распределения Стьюдента (это такое значение величины  $t$ , для которой вероятность того, что случайная величина  $x \leq t$ , равна  $\gamma$ ), то при условии, что  $|t|$  и  $\tilde{t}$  больше, чем  $t_{(1+\alpha)/2}$ , с доверительной вероятностью  $\alpha$  гипотеза отсутствия трендов отклоняется. Считая

доверительной вероятностью  $\alpha = 0,95$ , будем сравнивать статистики (21) с квантилями  $t_{0,975}(n)$ .

Рассмотрим ряд (7) для каждого часа суток, т.е.  $n = 217$  (это число данных для каждого часа суток, в соответствии с количеством дней). Тогда  $t_{0,975}(217) \approx 1,98$ . Из статистик (21) получаем, что тренд среднего значения с вероятностью 0,95 имеет место быть для следующих часов суток: 1, 8, 12, 23. Напротив, тренд маловероятен для 0, 5-7, 10-15 и 19 часов. В остальные часы можно считать, что среднее значение имеет тренд с вероятностью 0,8. Для дисперсии же гипотеза отсутствия тренда отклоняется с вероятностью 0,8 для 0, 1, 2, 6, 8, 11, 21 часов, а для остальных часов – с вероятностью 0,95. Таким образом, ряд не является стационарным в широком смысле, т.к. тренд в дисперсии имеет высокую степень достоверности.

#### 4. Выборочная функция распределения и критерий $\epsilon$ -стационарности

Поставим задачу прогнозирования временного ряда в иных терминах, используя понятие функции распределения. Новизна такой постановки заключается в том, что вместо скользящих средних (т.е. меняющихся моментов распределения) вводится оператор эволюции самой функции распределения. В данной работе мы опишем основные понятия такого подхода. В следующей статье будет дан пример его применения для временного ряда со свойствами, представленными в разделах 2-3.

Прогноз временного ряда в наперед заданный момент времени в общепринятом понимании представляет собой набор некоторых вполне конкретных чисел, которым, как предполагается, будут равны значения этого ряда. Однако вопрос «каким будет значение ряда в данный час  $t$ ?» является со статистической точки зрения не вполне корректным. Более правильной формой вопроса является следующая: «какова вероятность того, что значение ряда в данный час  $t$  заключено в пределах от  $p_1$  до  $p_2$ ?» Таким образом, для получения ответа на этот вопрос должна быть известна плотность распределения вероятности  $f(x, t)$  ряда  $x$  в момент времени  $t$ . Прогноз же случайной величины  $x$  в этих терминах определяется не однозначно, а по некоторому «соглашению». Например, прогнозным значением может быть аргумент наиболее вероятного значения плотности вероятности в данный момент времени. В качестве прогноза можно взять также среднее значение по данной функции распределения или иной функционал от  $f(x, t)$ . Очевидно, следует выбрать наилучший функционал, т.е. тот, который дает минимальную (в требуемом смысле) ошибку прогноза в течение некоторого периода времени. Возможно, что в разные периоды (например, месяцы или времена года) функционалы наилучшего приближения будут различны.

Для построения прогноза основной анализируемой величиной является отклонение значений ряда от среднего суточного графика в зависимости от ряда предыдущих значений. Соответствующая вероятность того, что отклонение  $c(t) - \bar{p}(t)$  лежит в пределах от  $x$  до  $x + dx$  при условии, что

предыдущие значения образуют определенный временной ряд (или, возможно, что некоторый функционал от прошлых ВФР равен определенному числу), называется условной выборочной функцией распределения (УВФР). Эта УВФР и является прогнозной моделью, совместно с выбором функционала наилучшего приближения.

Пусть первоначальная статистика для ВФР набирается по некоторому промежутку  $\Delta = (t; t + T)$ . Построенная по этим данным ВФР  $f_{\Delta}(x, t + T)$  будет давать распределение в момент времени  $t + T$ . Как и при анализе средних величин, в момент времени  $t + 2T$  у нас имеются две различных ВФР: одна, построенная скользящим образом по данным на промежутке  $[t + T; t + 2T]$ , т.е.  $f_{\Delta}(x, t + 2T)$ , и другая, «пополненная» ВФР  $f_{2\Delta}(x, t + 2T)$ .

Введем понятие  $\varepsilon$ -стационарности функции распределения. Будем говорить, что рассматриваемый процесс является  $\varepsilon$ -стационарным, если

$$\forall t, \tau \quad \int_{-\infty}^{\infty} |f_{\Delta}(x, t + \tau) - f_{\Delta}(x, t)| dx \leq \varepsilon. \quad (22)$$

Интегрирование по бесконечным пределам проводится здесь формально. В действительности ценовой ряд имеет ограниченное изменение, т.е. плотность распределения является финитной функцией аргумента  $x$ .

Введем функционал

$$V_{\Delta}(t, \tau) = \int_{-\infty}^{\infty} |f_{\Delta}(x, t + \tau) - f_{\Delta}(x, t)| dx. \quad (23)$$

Этот функционал аналогичен критерию Джини [4], вводимому для оценки расстояния между эмпирической и теоретической функциями распределения, но в нашем подходе он имеет смысл критерия стационарности. Если  $\tau = \Delta$ , то нижний индекс у функционала (23) будем иногда для краткости опускать.

Введем величины

$$U_{\Delta}(\tau) = \sup_t V_{\Delta}(t, \tau), \quad \tilde{U}_{\Delta}(t) = \sup_{\tau} V_{\Delta}(t, \tau). \quad (24)$$

Заметим, что в этих терминах любой процесс можно трактовать как  $\varepsilon$ -стационарный, где  $\varepsilon = \varepsilon(\Delta) = \sup_{\tau} U_{\Delta}(\tau) = \sup_t \tilde{U}_{\Delta}(t) \leq 2$ . Последнее неравенство

выполняется в силу того, что плотность распределения является неотрицательной функцией и нормирована на единицу. Разумеется, если  $\varepsilon \approx 1$  или больше 1, такое определение лишено практического смысла. Поэтому естественно считать, что процесс  $\varepsilon$ -стационарный, если только значение  $1 - \varepsilon$  не меньше точности, которая предъявляется априори к прогнозной системе.

Введенные величины  $U_{\Delta}(\tau)$ ,  $\tilde{U}_{\Delta}(t)$ , как и функционал  $V_{\Delta}(t, \tau)$ , имеют определенную аналогию, но уже с другим критерием согласия – Колмогорова-Смирнова [4]. Таким образом, в нашем подходе к определению  $\varepsilon$ -

стационарности временного ряда используются аналоги двух классических критериев согласия.

Пусть  $\varepsilon_0$  – допустимая ошибка в определении будущих значений временного ряда. Тогда квадрат ошибки прогнозирования равен дисперсии стационарного случайного процесса  $\sigma^2$  плюс квадрат ошибки  $D(t)$  за счет «ухода» средних значений:  $\varepsilon_0^2 = D + \sigma^2$ . Будем считать процесс квазистационарным, если он  $\varepsilon$ -стационарный и  $D \leq \sigma^2$ . В рамках этого определения получаем, что максимальное значение  $D$ , при котором процесс еще можно считать квазистационарным, равно  $D_{\max} = \varepsilon_0^2 / 2$ . Минимально допустимый объем выборки определяется тогда естественным условием, что одношаговое изменение ВФР (т.е. добавление одного значения во временной ряд) является квазистационарным процессом.

Величина  $\eta(\Delta) = \inf_{\tau \geq 1} U_{\Delta}(\tau)$  может быть принята за характеристику стационарности процесса с соответствующим объемом выборки  $\Delta$ , а величину  $\tau_0 : U_{\Delta}(\tau_0) = \min$  можно рассматривать как горизонт прогноза по стационарной модели. Отметим, что попытки анализировать нестационарный временной ряд на промежутках, где его можно считать приближенно стационарным, делались ранее в [8-10]. Однако в указанных работах интервал стационарности (т.е. горизонт прогноза) постулировался, а в нашем подходе он определяется совместно с необходимым для этого объемом выборки.

Аналогичные определения можно ввести и для пополненной ВФР, объем выборки для которой растет с течением времени.

Получим некоторые оценки для ряда (7) в вышеприведенных терминах. Сравнивая между собой данные [1], собранные в течение одного календарного месяца, на протяжении шести месяцев, находим, что значения функционала (23) оказываются следующими (первый аргумент – это порядковый номер месяца, второй – величина интервала набора статистики в месяцах):

$$V(1,1) = 0,208; V(1,1) = 0,346; V(2,1) = 0,413; V(3,1) = 0,382; V(4,1) = 0,313.$$

Видно, что процесс не является стационарным в узком смысле: функции распределения меняются от месяца к месяцу, и эти изменения составляют от 20 до 40%. С увеличением промежутка  $\Delta$  при фиксированном  $\tau$  величина функционала (23), очевидно, уменьшается. Важно отметить, что при этом уменьшается также и величина функционала  $V_{\Delta}(t, \Delta)$  по сравнению с аналогичными его значениями за меньшие промежутки:

$$V(1,2) = 0,192; V(3,2) = 0,297. \quad V(2,3) = 0,111.$$

Процесс будем называть асимптотически  $\varepsilon$ -стационарным, если

$$\exists \Delta_{\infty} : \forall \Delta > \Delta_{\infty} \quad V_{\Delta}(t, \Delta) \leq \varepsilon.$$

Если же существует предел  $\lim_{\Delta \rightarrow \infty} V_{\Delta}(t, \Delta) = 0$ , то процесс будем называть просто асимптотически стационарным.

Таким образом, практическая значимость критерия  $V(t, \tau) \leq \varepsilon$  состоит в том, что он позволяет определить величину промежутка  $\Delta$ , на котором данные временного ряда можно считать относящимися к одной генеральной совокупности локально или асимптотически.

Заметим, что при малых значениях  $\tau$ , например, при сдвиге на 1 шаг, т.е. при добавлении в ВФР одного нового значения, изменение ВФР будет пренебрежимо мало, если интервал набора данных достаточно велик. Именно, если ВФР набирается по статистике из  $N$  временных данных, то  $V_N(t, 1) \leq 2/N$ . Следовательно, если на основании выборочных данных можно будет выдвинуть гипотезу о том, что процесс асимптотически  $\varepsilon$ -стационарный, то на промежутке  $(1; \Delta_\infty)$  будет существовать такое значение  $\Delta_0$ , для которого функционал (23) при заданном значении времени  $t$  имеет максимум. Тогда ряд можно считать  $\Delta_0$ -стационарным на промежутке  $(1; \Delta_\infty)$ .

Обозначим через  $\tau_{\min}$  величину сдвига, при котором функционал (23) впервые достиг значения  $\varepsilon$  при некотором объеме выборки  $\Delta$ . Если  $\tau_{\min} \ll \Delta$ , то данные, сформировавшие первоначальную ВФР, нельзя считать принадлежащими одной генеральной совокупности. В этом случае возникает проблема оптимизации промежутка  $\Delta$  так, чтобы на нем ВФР была по возможности определена как  $\varepsilon$ -стационарная и разность  $\Delta - \tau_{\min}(\Delta)$  была минимальной. Если  $\Delta = \tau_{\min}(\Delta)$ , то такой объем выборки будем называть согласованным с  $\varepsilon$ -стационарной ВФР.

Пусть  $\sigma^2(\Delta)$  – дисперсия величины  $x$  по ВФР  $f_\Delta(x, t)$ . Будем уменьшать промежуток  $\Delta$ , следя за тем, чтобы, во-первых, разность  $\Delta - \tau_{\min}(\Delta)$  уменьшалась, и, во-вторых, чтобы при заданной погрешности  $\varepsilon_0$ , предъявляемой к аппроксимации данных,  $\sigma^2(\Delta)$  и ее вариация  $D(\sigma)/\bar{\sigma}$  по выборкам объема  $\Delta$  не превосходили бы  $\varepsilon_0^2$ . Минимальное значение объема выборки

$$\Delta_{\min} : \Delta - \tau_{\min}(\Delta) = \min \quad (25)$$

будем называть оптимальным объемом выборки.

Для прогнозирования на промежутке  $t \in (1; \tau_{\min})$  можно использовать методы анализа стационарных рядов. Этот промежуток времени определяет горизонт краткосрочного прогнозирования. На промежутке  $(\tau_{\min}, \Delta_0)$  временной ряд будет нестационарным, и требуется алгоритм, по которому следует переходить от ВФР, отнесенной к одному интервалу  $\varepsilon$ -стационарности, к ВФР в другом интервале. Такой алгоритм может быть построен на основе уравнения Лиувилля, выражающему закон сохранения полной вероятности со временем.

Пусть статистика для ВФР набирается по промежутку времени  $T$ , определяющему  $\Delta$  в соответствии с условием (25), так что к моменту  $t$  данные

были собраны на промежутке  $\Delta(t) = (t - T, t)$ . Пусть также на интервале  $\Delta_0$  укладывается  $p$  интервалов  $\tau_{\min}$ .

Обозначим через  $\hat{f}(x, t; n)$  функцию распределения, полученную по выборке из скользящего окна  $\Delta(t + n\tau_{\min}) = (t + n\tau_{\min} - T, t + n\tau_{\min})$ , т.е.  $\hat{f}(x, t; n) = f_{\Delta(t + n\tau_{\min})}(x, t + n\tau_{\min})$ . Определим скользящую скорость  $u(x, t; n)$  изменения ВФР  $\hat{f}(x, t; n)$ :

$$u(x, t; n) = \hat{f}(x, t; n+1) - \hat{f}(x, t; n), \quad \hat{f}(x, t; 0) = f_{\Delta(t)}(x, t). \quad (26)$$

Тогда средняя скорость изменения ВФР на промежутке  $\Delta_0$  равна

$$\bar{u}(x, t) = \frac{1}{p} \sum_{n=0}^{p-1} u(x, t; n) = \frac{\hat{f}(x, t; p) - \hat{f}(x, t; 0)}{p}. \quad (27)$$

Запишем формальное уравнение Лиувилля для  $\hat{f}(x, t; n)$ , изменение которой с течением текущего времени  $\tau$  определяется сдвигом вдоль координаты  $x$  с некоторой скоростью  $\hat{u}(x, t)$ :

$$\frac{\partial \hat{f}(x, t; \tau)}{\partial \tau} + \frac{\partial}{\partial x} \hat{u}(x, t) \hat{f}(x, t; \tau) = 0. \quad (28)$$

Мы записываем это уравнение в дифференциальной форме, хотя для нашего набора данных оно является разностным с шагом  $\tau_{\min}$ .

В качестве скорости  $\hat{u}(x, t)$  можно взять усредненное по промежуткам длительности  $\tau_{\min}$  эмпирическое выборочное выражение для скорости, фигурирующее в уравнении Лиувилля:

$$\hat{u}(x, t) = \frac{1}{n} \sum_{k=0}^{n-1} v(x, t; k), \quad v(x, t; k) = - \frac{\hat{f}(x, t; n+1) - \hat{f}(x, t; n)}{\hat{f}(x+1, t; n) - \hat{f}(x, t; n)}, \quad (29)$$

где величины  $x$  измеряются в единицах шага по координате, а величины скользящего времени – в единицах  $\tau^*$ .

В силу уравнения (28) легко получить уравнения эволюции моментов распределения  $\hat{f}$ . Обозначим угловыми скобками усреднение по переменной  $x$  с распределением  $\hat{f}$ :  $\langle g(x) \rangle = \int f(x) g(x) dx$ . Тогда

$$\frac{\partial \langle x \rangle}{\partial \tau} = \langle \bar{u} \rangle + \langle x \partial \bar{u} / \partial x \rangle, \quad (30)$$

$$\frac{\partial}{\partial \tau} \langle (x - \langle x \rangle)^2 \rangle = 2(\langle x \bar{u} \rangle - \langle x \rangle \langle \bar{u} \rangle) + \langle x^2 \partial \bar{u} / \partial x \rangle - 2\langle x \rangle \langle x \partial \bar{u} / \partial x \rangle.$$

Поскольку правые части уравнения (29) известны из предыдущих наблюдений, они представляют решение задачи о прогнозировании тренда в моментах ВФР.

## **Заключение**

Итак, в работе предложен метод анализа нестационарных временных рядов, который направлен на решение следующих задач:

- определение объема выборки для формирования стационарной ВФР с заданной точностью;
- минимизация совокупного функционала ошибки аппроксимации временного ряда на заданном интервале времени, а также определение интервала времени, на котором такая ошибка минимальна;
- прогнозирование временного ряда вне границ стационарности ВФР.

Применение описанного метода к прогнозированию временного ряда (7) будет проведено нами в следующей работе.

## **Цитированные источники**

1. НП «АТС». / <http://www.np-ats.ru/index.jsp?pid=36>
2. Каган А.М., Линник Ю.В., Рао С.Р. Характеризационные задачи математической статистики. – М.: Наука, 1972.
3. Лозв М. Теория вероятностей. – ИЛ, 1962.
4. Королук В.С., Портенко Н.И., Скороход А.В., Турбин А.Ф. Справочник по теории вероятностей и математической статистике. – М.: Наука, 1985.
5. Кремер Н.Ш., Путко Б.А. Эконометрика. – М.: ЮНИТИ-ДАНА, 2002.
6. Кобзарь А.И. Прикладная математическая статистика. – М.: Физматлит, 2006.
7. Foster F.G., Stuart A. A distribution-free test in time series dated on the breaking of records // JRSS, 1954. V. B16, 1, p. 1-22.
8. Кильдишев Г.С., Френкель А.А. Анализ временных рядов и прогнозирование. – М.: Статистика, 1973.
9. Четвериков Н.С. Статистические и стохастические исследования. – М.: Госстатиздат, 1963.
10. Леонтьев В. и др. Исследование структуры американской экономики (пер. с англ.). – М.: Статистика, 1968.