УДК 004.0- 621.3

Компьютерное исследование и прогноз квазипериодических рядов

А.И. Андрюхин, В.С.Марченко
Донецкий национальный технический университет, г. Донецк,
alexandruckin@ramber.ru

Андрюхин А.И., В.А.Марченко. Компьютерное исследование и прогноз квазипериодических рядов. *В работе выполнен обзор методов и алгоритмов исследования, получения характеристик и прогноза квазипериодических временных последовательностей. Построена программная система исследования, определения свойств и прогноза квазипериодических рядов(КСИКР). Wolfram Mathematica является базой для построения КСИКР. Основное внимание уделено программной реализации таких моделей и методов исследования квазипериодических рядов, как скрытые цепи Маркова и вейвлетные преобразования. Выполнен анализ известного факта, что сумма периодических функций может не быть периодической функцией. Представлены примеры расчетов и основная структура КСИКР.*

Ключевые слова*: квазипериодические ряды, прогноз, скрытые цепи Маркова, вейвлет, Wolfram Mathematica.*

Введение

Квазипериодические временные ряд представляет собой совокупность наблюдений, сделанных последовательно во времени. Природа этих наблюдений может быть очень разнообразной. С другой стороны, времена, в которые были сделаны наблюдения, могут быть регулярно или нерегулярно разнесены и более того, время может быть непрерывным или дискретным.

В первую очередь в работе мы фокусируемся на описании методов обработки числовых квазипериодических временных рядов, наблюдаемых через равные промежутки времени.

В настоящее время существует множество программных систем, в которых реализованы современные методы исследования и прогноза квазипериодических временных рядов и последние представляют большой интерес для многих сфер современного общества, как бизнес, наука и т.п.

В статье описывается попытка обьединения современных возможностей доступа к данным различного типа с помощью Интернета и богатством уже существующих программных разработок в математических пакетах, в частности Wolfram Mathematica.Также в работе построены собственные модели прогноза квазипериодических последовательностей.

В работе рассматриваются возможности компьютерного исследования и прогноза квазипериодических временных рядов в следующих направлениях:

а) исследование и прогноз квазипериодических временных рядов на базе параметрических моделей;

б) использование цепей Маркова для исследования и прогноза квазипериодических временных рядов;

в) применение вейвлет-преобразований для исследования и прогноза квазипериодических временных рядов

Практической задачей в работе являлось построение программной оболочки компьютерной системы исследования квазипериодических рядов (КСИКР) на базе системы Wolfram Mathematica.

Эта оболочка должна обеспечить ввод исходных данных и программную реализацию первых трех пунктов для комфортного применения многочисленных наработок в этой области.

Актуальность.

Для подтверждения актуальности работы приведем несколько примеров временных данных из современных исследуемых областей, которые носят квазипериодический характер[1-12].

Финансовые данные. Финансы - это область, в которой временные ряды естественно возникают из эволюции индексов и цен. Мы приведем два основных примера: эволюцию известного фондового индекса Standard & Poor's и его объем биржевых операций.. На рисунке 1 показан логарифм дохода ежедневного фондового индекса S & P500 за период с января 1950 года по январь 2014 года. Обратите внимание, что этот показатель со временем растет, но есть некоторые периоды спада, которые обычно обозначаются как рынки с тенденцией с понижениv.

Для изучения этих индексов обычно приходится рассматривать логарифмический доход , который определяется как $r_t = \log P_t / P_{t-1} = \log P_t - \log P_{t-1}$, где P_t обозначает цену или значение индекса в момент времени t.

Этот показатель представлен на рис.1. Укажем на большое падение доходностей, произошедшем в октябре 1987 года, и резкие

изменения в течение 2009 года.

Экономические данные.На рис. 2 показана ежемесячная занятость в США в сфере искусства, развлечений и отдыха за период с января 1990 года по декабрь 2012 года, измеренная в тысячах человек.

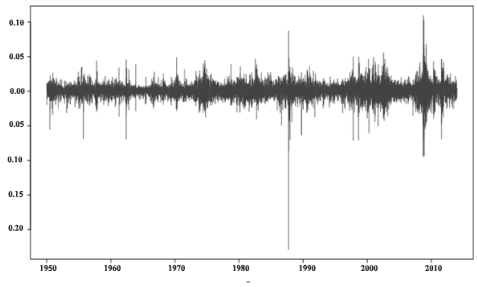


Рисунок 1. – Логарифмический доход фондового индекса Standard & Poor's с 1950 по 2014[1].

Согласно ему легко определяются сезонная картина и верхний тренд.

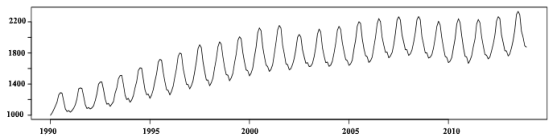


Рисунок 2. – Ежемесячная занятость (в тыс.) в США в сфере искусства, развлечений и отдыха за период с января 1990 года по декабрь 2012 года[7]

Гелио-физические данные. На рис.3 представлен квазипериодический ряд ежемесячного графика чисел Вольфа с 1906 по 2016 согласно [7].

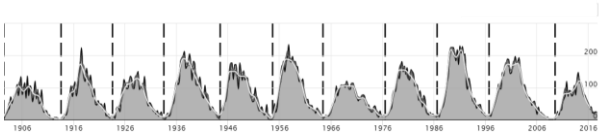


Рисунок 3. – Ежемесячный график чисел Вольфа с 1906 по 2016

Транспортные данные. На рисунке 4 показано количество ежемесячных авиапассажирских перевозок за период с января 2004 года по декабрь 2013 года Соединенных Штатах Америки.

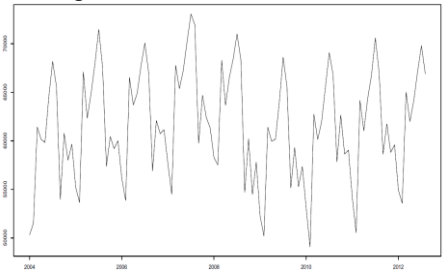


Рисунок 4. –Обьем ежемесячных авиапассажирских перевозок в США с января 2004 года по декабрь 2013[1]

Опять подчеркиваем сезонное поведение этой серии, полученное вследствие зимнего и летнего сезонов.

Также очевидно, что произошло падение в 2009 году, выявив вероятный эффект финансового кризиса этого года.

Социологические данные. На рис.5 представлены результаты серии ежемесячных обследований намерений в отношении голосования в Соединенном Королевстве за период с июня 1984 года по март 2012 года.

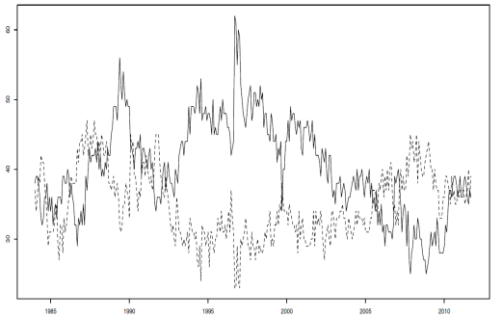


Рисунок 5. –Данные по намерениям голосования за лейбористов и консерваторов с июня 1984 года по март 2012 года[1].

Здесь показаны намерения голосовать за консервативную партию и лейбористскую партию. Толстая линия указывает на намерение голосовать за лейбористскую партии, а пунктирная линия соответствует намерениям голосования за консервативную партию.

Эти две политические партии исторически концентрируют большой процент голосов и мы наблюдаем зеркальный эффект в поведении этих двух линий на рисунке..

Кроме того, за период с 1993 по 2005 год существует большой разрыв между намерениями обеих сторон по голосованию. В течение этого периода лейбористская партия показывает более высокий уровень намерений по поддержке, нежели чем консервативная

Аналогичные квазиопериодические ряды можно привести для таких областей, как:

1)загрязнения окружающей среды;
2)гидрологические показатели;
3)биомедицинские данные;
4)климатологические характеристики;
5)данные в области энергетики и др[2-12].

Следовательно, исследование квазипериодических рядов теснейшим образом связано с наблюдаемыми в реальности процессами в различных проблемных областях.

К ним мы относим, как уже указано, социальные и политические процессы с их бурными преобразованиями, как кризисы и революции, экономические и финансовые процессы, природные явления и т.п.

Для формализации моделей этих процессов приведем некоторые системные и математические соображения.

Рассмотрим динамические системы, порожденные дифференциальными уравнениями

$$Dx/dt = F(x)$$

или соответствующее разностное уравнение

$$x(t+1) - x(t) = f(x(t)),\ x \in R^n.$$

Для этих уравнений для любой постоянной $s>0$ имеет место следующее очевидное свойство решений

$$X(t+s, s, x0) = x(t, 0, x0)$$

Это свойство справедливо и в случае более общих описаний динамических систем (с точки зрения их фазовых пространств или нелинейных операторов, действующих в этих пространствах).

Из последнего уравнения следует, что отрезок траектории $x(t, x0)$, выходящий из точки $x0$ в момент времени $t = 0$, совпадает с отрезком траектории, исходящим из точки $x0$ в момент времени $t = \tau$.

Из этого следует, что в тех же условиях физические экспериментальные данные повторяются, и поэтому мы можем теоретически прогнозировать некоторые процессы и управлять ими. Однако из-за возникновения неустойчивостей (которые в последние годы интенсивно рассматривались в [1]) упомянутое выше совпадение часто возможно в достаточно малых временных интервалах $(0, \tau)$ и $(s, s + \tau)$ только.

Мы описываем некоторые подходы к прогнозу и контролю, которые обусловлены общими законами неустойчивостей, возникающих в динамических системах.

Эти подходы, разработанные в рамках «экспериментальной математики», основаны на том, что мы не пытаемся строить, идентифицировать и анализировать приближенные модели довольно сложных реальных динамических объектов, а собирать определенный экспериментальный материал, связанный с реальными моделями, а затем использование его для прогнозирования и построения контроля.

Возникновение неустойчивостей подчиняется некоторым общим закономерностям, учет которых приводит к некоторым общим принципам теории прогнозирования и управления при исследовании квазипериодических временных последовательностей, которые являются объектом изучения в работе.

Сейчас ясно, почему нельзя составить прогноз погоды более чем на две недели по квазистационарному ряду наблюдений за ней ?

В 50-60-е годы прошлого века прогресс в области механики сплошных сред и вычислительной математики позволил генерировать более точные математические модели атмосферных изменений, строить более эффективные алгоритмы для решения дифференциальных уравнений этих моделей и реализовывать эти алгоритмы с помощью более быстродействующих компьютеров. Благодаря этому прорыву получило широкое распространение мнение, что, сделав некоторые дополнительные усилия в этих направлениях, мы можем сделать прогноз погоды на многие недели, месяцы и даже годы.

Однако впоследствии выяснилось, что длительный прогноз погоды в принципе невозможен. Этот факт был теоретически установлен в работах Э. Лоренца и его последователей, которые обнаружили неустойчивость математических моделей атмосферы. Эта неустойчивость следует из-за сильной чувствительности решений дифференциальных уравнений, описывающих атмосферные процессы в зависимости от исходных данных.

Понимание этого факта привело к экспериментальным наблюдениям в рамках нового направления, в так называемой «экспериментальной математике».

В Европе накоплен большой материал метеорологических наблюдений. Такие наблюдения проводились регулярно в течение многих десятилетий. Рассмотрим наблюдения, например, 31 мая 2014 года в определенном регионе Европы. Затем мы выбираем определенный год (например, 27: такой, что 31 мая 1927: в этом районе наблюдались примерно одинаковые метеорологические параметры (температура, давление, влажность воздуха, сила и направление ветра, облачность и т. д.).

Эти параметры используются в качестве начальных (и граничных) условий для решений дифференциальных уравнений атмосферной модели, которые описывают законы механики сплошных сред. Последние всегда справедливы для любого года. В то время как в разные годы решения этих уравнений однозначно определяются начальными данными, так как уравнения и исходные данные 30 июня 2014 г. и 30 июня 1927 совпадают, мы считаем, что решения, описывающие изменение метеорологических данных (такие как температура, давление, влажность воздуха и т. д.) должны быть одинаковы. Следовательно, для каждого из дней из выбранного интервала времени параметры, которые наблюдались в течение месяца к примеру должны совпадать с

достаточной точностью. Поэтому казалось бы, что погодные условия 1 июня 2014 г., например, и 1 июня 1927: должны быть очень близкими друг к другу. Однако эксперименты показывают, что такое совпадение возможно только на временных интервалах, не превышающих двух недель.

В то же время совпадение погодных условий на неделю может быть довольно близким и в метеорологии этот факт чаще всего используется для краткосрочного прогноза. Но результаты наблюдений во временных интервалах, превышающих две недели, сильно расходятся.

Это обусловлено тем, что небольшая дивергенция исходных данных в начальные моменты наблюдений приводит к большому расхождению наблюдаемых параметров уже через две недели. Таким образом, даже если математическая модель атмосферы достаточно правильна, и компьютерная техника обеспечивает высокую скорость расчетов, единственным результатом является то, что правильный прогноз погоды на две недели невозможен.

По этой причине в Японии отказались сделать прогноз погоды более чем на десять дней.

Рассмотрим прогноз поведения рынка как аналога прогноза погоды.

Можно утверждать, что в то время как физические законы и соответствующие уравнения конвекции справедливы для любого временного интервала, подобные законы рынка зависят от политики, финансовых обстоятельств и стремления участников рынка. Все они одинаковы только на коротком временном интервале [t0, T] (дни или часы). Ясно, что изменение переменных величин рынка x_k (t) подчиняется законам рынка.

Количество таких величин (используя сравнение с погодой) должно быть около десяти (k = 1, ..., 10). В случае рынка изменение начальных (и граничных) условий, подобных тем, которые используются для прогноза погоды, происходит в силу смены переменных на определенном «начальном» временном интервале [t0, t1] (t1 намного меньше ,чем T.) Это позволяет сделать некоторый аналог «начальной функции» для дифференциальных уравнений с запаздыванием.

Гипотезы. Существуют такие классы рынков, что «хорошее» совпадение всех наблюдаемых переменных характеристик рынка x_k (t) (k = 1, ..., N) на отрезках [t0, t1] и [t0 + τ, t1 + T] (t1 + τ <T) следует их «хорошее» совпадение на определенных отрезках [t1, t1 + dt] и [t1 + τ,t1 + τ + dt], в которых t1+ τ + dt <T. «Хорошее» означает некоторое предварительное «сглаживание» или усреднение величин x_k(t), что аналогично, например, тому, как мы учитываем некоторую «среднюю» скорость ветра, сглаживая его или ослабляя на малых временных интервалах.

Таким образом, мы можем, по-видимому, прогнозировать поведение определенных рынков на малых временных интервалах, или погоды, используя аналогичные параметры (характеристические переменные) предыдущих наблюдений. Разумеется, эта гипотеза нуждается в проверке конкретных многопараметрических рынков. Кроме того, в этом случае мы должны удачно выбрать (из экспериментов, как и раньше) временные масштабы (t, E, t1, T, τ, dt).

Основные направления исследования квазипериодических временных рядов

В работе основное внимание уделяется стохастическим квазиопериодическим процессам, которые наблюдаются в дискретные времена ..., t_0, t_1, t_2, ..., в отличие от непрерывного времени. Большинство моделей, разработанных в литературе временных рядов, связаны с равноотстоящими временами. В этом случае наблюдения могут быть записаны как $\{y_t: t \in Z\}$ для $t \in \{..., -2, -1,0,1,2, ...\}$. Существуют модели для обработки неравномерно распределенных интервалов времени, но их обычно сложнее определять и изучать. Можно также упомянуть проблему недостающих данных, где ряд не наблюдается при некоторых значениях t.

Основные определения и свойства.

Линейный процесс может быть записан как,

$$y_t = \beta(B)\varepsilon_t$$

где y_t - наблюдаемые значения, β(B) - линейный фильтр , B –оператор обратного сдвига, а ε_t - входной шум. Напомним, что B осуществляет сдвиг на временной такт назад при наблюдении, то есть $By_t = y_{t-1}$.

Тогда фильтр β(B) может быть записан как

$\beta(B) = \Sigma\beta_j B^j$, $j \in (-\infty,+\infty)$, где $\Sigma(\beta_j)^2 < +\infty$, $j \in (-\infty,+\infty)$.

Следовательно, мы можем записать наблюдаемый временной ряд в виде

$y_t = \Sigma\beta_j \varepsilon_{t-j}$ $j \in (-\infty,+\infty)$

Заметим, что этот фильтр называется линейным, так как он не содержит нелинейных слагаемых, как $\varepsilon_{t-i}\varepsilon_{t-j}$. Это определение говорит нам о том, что наблюдаемое значение в момент времени t, y_t, зависит от прошлых, настоящих и будущих значений входного шума, т.е.$\{..\varepsilon_{-3}, \varepsilon_{-2}, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3 \ldots\}$.

Стационарность. Это важнейшее

понятие при анализе временных рядов. Вообще говоря, мы можем различать два определения стационарности. Первое определение основное внимание обращает на совместном распределении процесса.

Строгая стационарность. Пусть $y_h(\xi) = \{y_{t1+h}(\xi),\ldots, y_{tn+h}(\xi)\}$ есть траектория процесса $\{y_t\}$ с временными моментами $\{t_1 + h,..,t_n + h\} \in Z$. Процесс называется строго стационарным тогда и только тогда, когда распределения y_h одинаковы и независимы от h.

Слабая стационарность. Процесс y_t называется слабо стационарным или стационарным второго порядка, если

а) он имеет постоянное среднее;

б) он имеет конечный и постоянный второй момент(дисперсия);

(с) существует функция γ, такая что $\gamma(k) = Cov(y_t, y_{t+|k|})$ для любых t и к.

Как было указано ранее, стационарность является важной концепцией анализа временных рядов, и это означает, что статистические свойства процесса остаются постоянными во времени. На практике это означает, что все значения процесса сопоставимы, независимо от того, в какое время они наблюдались. В свою очередь, сопоставимость наблюдения позволяют нам делать статистические выводы по всему процессу. Процесс может быть строго стационарным, но не обязательно слабо стационарным и наоборот.

Строгий процесс белого шума представляет собой последовательность независимых и идентично распределенных случайных величин, а слабый белый шум – последовательность некоррелированных случайных величин с нулевым средним и постоянной конечной дисперсией, т.е. с функцией автоковариации γ, удовлетворяющей условиям $\gamma(0) < \infty$ и $\gamma(h) = 0$ для $h \neq 0$.

Обратимость. Линейный процесс обратим, если существует такой фильтр f (B), что мы можем записать

$$f(B)y_t = \Sigma f_j y_{t-j} = \varepsilon_t,\ j \in (-\infty,+\infty),$$

Фильтр (B) можно рассматривать как обратный к фильтру β(B), если

$f(B)\beta(B) = 1$. Заметим, что обратимый временной ряд y_t можно выразить как

$$y_t = S^- + S^+ + \varepsilon_t,$$

где $S^- = \Sigma f_j y_{t-j}$, $j \in (-\infty,-1)$, а $S^+ = \Sigma f_j y_{t-j}$, $j \in (1, +\infty)$.

Причинность. Одним из способов описания случайного процесса с дискретным временем является его запись в виде фильтрации последовательности белого шума $\{\varepsilon_t\}$,

$$y_t = \beta(..\varepsilon_{-2}, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \varepsilon_2,..)$$

где β - измеримая функция, т. е. результирующая последовательность $\{y_t\}$ хорошо определенный случайный процесс.

Предположим теперь, что последовательность шума $\{\varepsilon_t\}$ генерируется одновременно как наблюдаемый процесс $\{y_t\}$, так что в любой момент времени t, генерируемая последовательность шума $(..\varepsilon_{-2}, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \varepsilon_2,.. \varepsilon_t)$ и наблюдаемый процесс предоставляется как $(..y_{-2}, y_{-1}, y_0, y_1, y_2,.. y_t)$

В этом контексте процесс $\{y_t\}$ является причинным, и его можно записать как

$$y_t = \beta(..y_{-2}, y_{-1}, y_0, y_1, y_2,.. y_t)$$

Таким образом, каузальный процесс зависит только от прошлого и настоящего шума и не зависит от будущих значений шума. Это важная особенность прооцесса $\{y_t\}$, что означает, что только прошлые или нынешние воздействия могут повлиять на него. Если процесс не является причинным, например, $y_t = \varepsilon_{t+2} + \varepsilon_{t+1} - \beta\varepsilon_t$., тогда будущие события $\varepsilon_{t+2}, \varepsilon_{t+1}$ могут оказывать влияние на его текущее значение. Хотя этот процесс не является причинным, мы все еще можем предсказывать его. Так известно, что лучший линейный прогноз y_t задается выражением

Последовательная зависимость. Рассмотрим стохастический процесс $\{y_t\}$ и предположим, что его среднее значение $\mu_t = M(y_t)$. Если этот процесс является гауссовским, то его можно разложить аддитивно, как $y_t = \mu_t + \eta_t$, где η_t - случайный процесс с нулевым средним значением. Чтобы указать процесс $\{y_t\}$, можно определить μ_t более конкретно. Например, для стационарного процесса среднее предполагается постоянным во времени, так что $\mu_t = \mu$ для всех t. В более общем смысле среднее может быть задано линейной моделью, которая зависит от времени $\mu_t = \beta_0 + \beta_1 t + \cdots + \beta_p t^p$ или зависит от других ковариантов $\mu_t = \beta_0 + \beta_1 x_1 t + \cdots + \beta_p x_p t$.

Стационарность означает, что статистические характеристики временных рядов сохраняются во времени. В частности, среднее и дисперсия рядов постоянны и относительная зависимость наблюдения от прошлых значений остается той же самой, независимо от момента, в который она оценивается.

То есть предположим, что существует такая функция γ, что

$$\gamma(h) = Cov(y_t, y_t + h).$$

Существование этой функции, обозначаемой как функция автоковариации, означает, что ковариация между наблюдениями y_t и $y_t + h$ не зависит от t.

Стационарность является ключевым допущением в анализе временных рядов для проведения статистических выводов и

прогнозирования.

Автокорреляционная функция (АКФ) определяется как:

$$\rho(h) = \gamma(h)/\gamma(0). \quad (1)$$

Эмпирические оценки АКФ даются так называемыми оценками моментов:

$$\rho_k = \gamma(k)/\gamma(0), \quad (2)$$

где $\gamma(k) = \Sigma(y_t - My)(y_{t+k} - My)/n$, t=1,n-k.

Нестационарность. Многие реальные временные ряды отображают нестационарные функции, такие как тренды или сезонность. Учитывая, что большинство методологий анализа временных рядов основано на предположении о стационарности, существует ряд методов, разработанных для преобразования нестационарных данных в стационарные. Среди этих подходов часто используются стабилизация дисперсии, оценка тренда с помощью линейной регрессии и дифференцирование ряда.

Многие статистические тесты и интервалы основаны на предположении о нормальности. Предположение о нормальности часто приводит к простым, математически корректным и мощным испытаниям по сравнению с тестами, которые не делают предположения о нормальности. К сожалению, многие реальные наборы данных на самом деле не являются приблизительно нормальными. Однако соответствующее преобразование набора данных может часто приводить к набору данных, который имеет статистические свойства примерно как нормальное распределением. Это повышает применимость и полезность статистических методов, основанных на предположении о нормальности.

Преобразование Box-Cox - это особенно полезное семейство преобразований [palma]. Оно определяется как:

$$T(Y) = (Y^{\lambda}-1)/\lambda$$

где Y - переменная отклика, а λ - параметр преобразования. При $\lambda = 0$ вместо используемой выше формулы берется натуральный логарифм .

Так стабилизация дисперсии обычно достигается путем преобразования данных Box-Cox.

Линейные модели - это инструменты для удаления детерминированного тренда из данных. Эта модель регрессии обычно включает в себя многочлен от t, гармонические компоненты или может содержать другие элементы. Таким образом, модель может быть записана как

$$y_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_p x_{pt} + \eta_t = X\beta + \eta,$$

где матрица $X = (1, x_{1t}, ..., x_{pt})$ являются ковариационной , а вектор $\eta = (\eta_1, \eta_2, ..., \eta_n)$ представляет собой несистематические ошибки. Коэффициенты $(\beta_0, \beta_1, ..., \beta_p)$ могут быть получены, например, оценками по методу наименьших квадратов (МНК). Согласно последнего мы определяем $\beta = (X'X)^{-1}X'y$. После оценки параметров регрессии исследуемый далее ряд e_t получается путем удаления регрессионной части из ряда $\{y_t\}$, $e_t = y_t - \beta_0 - \beta_1 x_{1t} - \cdots - \beta_p x_{pt}$. Впоследствии к этой результирующей последовательности можно применить методы временных рядов. Во многих приложениях регрессионные формы являются либо полиномами, либо гармоническими функциями, как, например, в случае сезонного поведения,

$$Y_t = \Sigma(\alpha_j \sin(\omega_j t) + \beta_j \cos(\omega_j t)) + \eta_t, \; j=1,m$$

где коэффициенты α_j и β_j неизвестны, но частоты ω_j обычно считаются известными или получаем их из спектрального анализа .

Дифференцирование. Другой подход для устранения тренда в данных - это дифференцирование. В этом случае, однако, основной тренд считается недетерминированным или стохастическим. В рамках этого подхода предполагается, что данные генерируются и интегрируются посредством некоторого стохастического процесса, например,

$$y_t = \Sigma\eta_k, \; k=1,t$$

где η_t - стационарный процесс с постоянной дисперсией с нулевым средним значением.

Таким образом, дифференцируя $\{y_t\}$, получаем ряд $\{z_t\}$

$$z_t = y_t - y_{t-1} = \eta_t,$$

Общей проблемой этого метода является решение, когда следует прекращать дифференцирование. Следует учесть два основных аспекта. Во-первых, продифференцированная последовательность должна выглядеть стационарной и во-вторых, ее дисперсия не должна быть больше дисперсии начальной последовательности. Несоразмерное увеличение дисперсии в результирующей серии может указывать на чрезмерную дифференциацию.

Другая дилемма заключается в выборе между построением регрессии или дифференциацией. Несмотря на отсутствие общих указаний по этому поводу, можно применить любой из этих методов и посмотреть, дают ли они адекватные результаты или нет.

Например, если процесс y_t удовлетворяет регрессионной модели $yt = \beta_0 + \beta_1 t + w_t$ с белым шумом w_t, то выполняя дифференцирование, получаем $z_t = y_t - y_{t-1} = \beta_1 + w_t - w_{t-1}$. Дисперсия ряда z_t в два раза превышает дисперсию исходных данных, что свидетельствует о неправильном применении процедуры дифференциации. Помимо подобных соображений, существует много других методов преобразования для

достижения стационарности. С другой стороны, существуют методологии, позволяющие непосредственно обрабатывать нестационарные данные без трансформации. Одним из примеров этих методов являются так называемые локально-стационарные модели.

Белый шум. Процесс белого шума представляет собой последовательность нулевых средних некоррелированных случайных величин. Если эта последовательность является гауссовой, то процесс также независим.

Фундаментальной процедурой анализа временных рядов является проверка, является ли последовательность белым шумом, или она имеет более сложное поведение. Учитывая последовательность $y_1, ..., y_n$, нулевая гипотеза равна H0: $\{y_t\}$ - белый шум по сравнению с H1: $\{y_t\}$ - не белый шум. Здесь необходимо обратить внимание, что проверка H0 может быть затруднена по многим причинам. Например, среднее значение процесса не является постоянным, его дисперсия не является постоянной и т.д,

Процедуры тестирования белого шума обычно не предусматривают проверку независимости, если только серию не считают гауссовой. Важно подчеркнуть, что определение белого шума относится только к некоррелированной последовательности.

В частности, это означает, что последовательность с коррелированными квадратами по-прежнему является белым шумом в соответствии с этим определением. Как правило, это относится к финансовым временным рядам: доходность часто не коррелирует, но волатильность или квадрат прибыли часто коррелируются.

Как правило, тест белого шума учитывает определяемые автокорреляции $r_1, ..., r_L$ с $r_k = \rho_k$, где ρ_k задается формулой (1) и (2).

Тест Box-Ljung, хорошо известная процедура проверки, является ли последовательность белым шумом или нет, может быть определена как

$$Q_L = n (n + 2)\Sigma r^2_k/(n-m), m=1,n-1$$

и можно показать, что статистика Q_L соответствует распределению χ^2 с L степенями свободы.

Основные модели временных квазипериодических рядов

Модель авторегрессии и скользящего среднего (autoregressive moving-average model, ARMA*).* Модели ARMA являются фундаментальными инструментами для анализа краткосрочных временных рядов. Можно показать, что этот класс моделей аппроксимирует любой линейный стационарный процесс с непрерывной спектральной плотностью. Кроме того, имеется большое число численных и вычислительных инструментов для диагностики и прогнозирования моделей ARMA. Они очень полезны для моделирования большого количества временных рядов, демонстрирующих слабую зависимость.

С другой стороны, авторегрессионные дробно интегрированные со скользящим средним процессы (ARFIMA) широко используются для данных временных рядов демонстрирующих долговременную зависимость. Процесс ARMA (p, q) {yt} может быть задан дискретно-временным уравнением,

$$\varphi (B)\, y_t = \theta (B)\, \varepsilon_t,$$

где, $\varphi(B) = 1-\varphi_1 B - \cdots - \varphi_p B^p$ - авторегрессионный многочлен от оператора обратного сдвига B, $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ - полином с значением подвижного среднего, корни которого отличаются от корней уравнения $\varphi(B)$. Ряд $\{\varepsilon_t\}$ - последовательность белого шума с нулевым средним и дисперсией σ^2. Подчеркнем, что процесс авторегрессии AR (p) соответствует модели ARMA (p, 0). С другой стороны, модели скользящего среднего MA (q) является частным случаем ARMA (0, q).

Мы фокусируем свое внимание на определенном классе линейных временных рядов, называемых длительной памятью или процессами, зависящими от дальнего расстояния. Существует несколько определений временных рядов этого типа в литературе. Один из основных аспектов связан с оценкой среднего значения процесса. Если функция автоковариации стационарного процесса суммируема, то выборочное среднее будет соответствовать корню из n, где n - размер выборки. Это имеет место, например, для последовательностей независимых и одинаково распределенных случайных величин или марковских процессов. Говорят, что эти процессы имеют короткую память. Наоборот, процесс имеет длительную память, если его автоковариации не являются абсолютно суммируемыми. В дальнейшем мы дадим краткий обзор этих классов моделей временных рядов.

Рассмотрим основные методологии для оценки моделей временных рядов.

Существует целый ряд хорошо известных методов, таких как метод максимального правдободобия и его различные вычислительные формулировки, такие как разложение Холецкого или уравнения системы в переменных состояния. С другой стороны, имеются аппроксимации метода максимального правдоподобия, включая, например, подход Уиттла, метод скользящего

среднего и авторегрессионное приближение.

Для выполнения первого этапа анализа временного ряда необходимо определить оценки среднего и АКФ ряда, которые необходимы как инструменты для определения модели.

Построение модели. Реальные временные ряды имеют ряд своих отличительных свойств. Среди них важно определить, будет ли ряд стационарным или нет.

В первом случае мы можем перейти к этапу моделирования, посмотрев на структуру автокорреляции, конкретный пример АКФ и ЧАКФ (частичная АКФ). Исходя из этих оценок момента, стационарные модели такие как, модели ARMA или ARFIMA могут быть предложены и реализованы.

С другой стороны, если ряд отображает ряд нестационарных характеристик, мы можем применить процедуры преобразований для получения стационарной серии. Среди этих процедур, которые мы предварительно пересмотрели путем декомпозиции данных с помощью средств методов регрессии и дифференциации данных временных рядов. Если регрессионная модель хороша, мы можем рассматривать остатки как ряды, которые должны быть проанализированы. Если происходит дифференциация, мы можем применить модель ARIMA.

Если ряд показывает сезонное поведение, тогда мы можем использовать гармоническую регрессию или модели SARIMA.

Спецификация модели обычно связана с выбором класса таких процессов, как ARIMA (p, d, q) или SARIMA (p, d, q)× (P, D, Q).

Эти модели могут быть выбраны согласно анализа автокорреляционной функции АКФ или частичных АКФ (PACF). Сейчас общепринято рассмотреть вложенное семейство моделей и затем оценить все модели класса. Например, ARMA (p, q) с порядком p,q = 0, 1, 2, 3.

Поскольку модели являются вложенными, мы можем использовать информационные критерии, такие как AIC или BIC, для выбора подходящих значений p, q.

Оптимальность моделей. Теоретически, если линейный процесс имеет непрерывный спектр, то мы всегда можем найти значения p и q, так что ARMA (p, q) достаточно хорошо его аппроксимирует. Следовательно, на практике мы всегда можем опираться на этот класс процессов для моделирования линейного временного ряда.

Однако этот общий математический результат не гарантируют, что значения p и q малы. Фактически, они могут быть довольно большими.

Наличие ARMA-модели с большими авторегрессионными и скользящими средними заказами может быть громоздким как с численной, так и с статистической точки зрения.

Предположим, что p = 30 и q = 28. Эта модель может хорошо подходить для набора данных, но она требует численного расчета 58 параметров и проверки, что она является стационарной и обратимой.

Обычно желательно, чтобы подходящая модель t была оптимальной, то есть значения p, q были относительно невелики. В этом смысле, есть выбор между качеством аппроксимации модели, которое обычно требует больших значений p и q, и простотой модели, которая требует небольшого числа параметров p и q.

Этот выбор обычно выполняется с помощью информативных критериев, которая штрафуют модель в соответствии с числом оцениваемых параметров. Они рассматриваются ниже.

Критерии информативности Акаике и Шварца. Информационный критерий Акаике (AIC) для краткости определяется как информативность Акаике:

$$AIC = -2 \log L(\beta) + 2r,$$

где, $\overline{\beta}$ - оценка максимального правдоподобия, r - число оценочных параметров модели. Например, для модели ARMA (p, q) имеем r = p + q + 1.

Информационный критерий Шварца или байесовский информационный критерий (БИК) определяется формулой

$$BIC = -2 \log L(\overline{\beta}) + r \log n.$$

Заметим, что для размера выборки n> 8 BIC гораздо более строго наказывает за увеличение числа параметров в модели, по сравнению с AIC.

Оценка среднего значения. Оценка среднего стационарного процесса является фундаментальной задачей анализа временных рядов. Несмотря на то, что имеется несколько оценок среднего значения, наиболее часто рассматриваются среднее значение выборки или наилучшая несмещенная линейная оценка (ННЛО).

Имея выборку $Y_n = (y_1, y_2, ..., y_n)'$ из стационарного процесса со средним μ и дисперсией D, среднее выборочное значение определяется как $\mu = 1/n \Sigma y_t, t= 1, n$, или $\mu = (I.Y_t)/n$, где $I = (1,1, ..., 1)'$, а для ННЛО $\mu = (I'D^{-1}I)^{-1}1'D^{-1}Y_n$. (′-знак транспонирования).

Разнообразное поведение этих двух хорошо известных оценок критически зависит от памяти процесса. Для процесса с малой памятью,

такого как ARMA-модель, среднее значение выборки и ННЛО сходятся к истинному среднему со скоростью O (n^{-1}). Аналогична оценка для их асимптотической дисперсии. В этом смысле среднее значение выборки является эффективной оценкой.

С другой стороны, для сильно зависимого процесса с большой памятью с параметром d скорость сходимости обеих оценок равна O (n^{2d-1}). Поскольку d>0, эта скорость сходимости медленнее, чем для случая короткой памяти. Кроме того, асимптотические дисперсии среднего значения выборки и ННЛО являются различными, подразумевая, что среднее значение выборки не является эффективной оценкой.

В частности, могут быть установлены следующие соотношения:

$$\sqrt{n}(y_n - \mu) \to N(0, v),$$

где, $v = 2\pi f(0) = \Sigma \gamma(h)$, $h \in (+\infty, -\infty)$.

Аналогично для ННЛО имеем

$$\sqrt{n}(\bar{\mu}_n - \mu) \to N(0, v).$$

Заметим, что для процесса ARMA (p, q) имеем $v = \sigma^2 |\theta(1)|^2 / |\varphi(1)|^2$.

Для процессов с большой памятью $n^{1/2-d}(\bar{y}_n - \mu) \to N(0, w)$:

и $w = \sigma^2 |\theta(1)|^2 / |\varphi(1)|^2 \Gamma(1-2d)/(d(1+2d)\Gamma(d)\Gamma(1-d))$,

а для ННЛО имеем $n^{1/2-d}(\bar{\mu}_n - \mu) \to N(0, w)$, где $w = \sigma^2 |\theta(1)|^2 / |\varphi(1)|^2 (\Gamma(1-2d)\Gamma(2-2d))/\Gamma(1-d)^2$.

Оценка автоковариаций. Для стационарного процесса автоковарианты обычно оцениваются с помощью оценки моментов. То есть, имея выборку $Y_n = (y_1, y_2, ..., y_n)'$ из стационарного процесса со средним μ и дисперсией D, обычная оценка автоковариации при запаздывании h, определяется формулой:

$$\gamma(h) = \Sigma(y_t - \bar{y})(y_{t+h} - \bar{y})/n, \quad t = 1, n-h.$$

Можно показать, что для фиксированного h эта оценка является асимптотически несмещенной $\lim \bar{\gamma}_n(h) \to \gamma(h)$ при $n \to \infty$.

Если автоковариации процесса абсолютно суммируемы

$$\sqrt{n}(\bar{\gamma}_n(h) - \gamma(h)) \to N(0, v),$$

где, $v = (n-3)\gamma((h)^2 + \Sigma(\gamma(j)^2 + \gamma(j-h)\gamma(j+h))$, $j \in (+\infty, -\infty)$.

Аналогичные выражения можно найти для асимптотического поведения АКФ. В этом случае мы имеем:

$$\sqrt{n}(\bar{\rho}_n(h) - \rho(h)) \to N(0, w),$$

где w определяется формулой Бартлета:

$w = \Sigma (1 + \rho((h)^2)) \rho(j)^2 + \rho(j)\rho(j+2h) + 4\rho(h)\gamma(j)\gamma(j+h))$, $j \in (+\infty, -\infty)$.

Оценка моментов. Обычно оценка моментов основан на сравнении выборки АКФ с ее теоретическим представлением. Имея выборку $Y_n = (y_1, y_2, ..., y_n)'$ из стационарного процесса со средним μ и дисперсией D и модель временного ряда с АКФ с $\gamma_p(h)$, запишем уравнения момента для различных значений h:

$\gamma_p(h) = \bar{\gamma}(h)$.

Решение этой системы уравнений $\bar{p}$ является истинной оценкой p.

Аналогично, эти уравнения можно записать в терминах типовых автокорреляций:

$$\rho_p(h) = \bar{\rho}(h).$$

В качестве примера рассмотрим авторегрессионую модель первого порядка AR (1), т.е. модель вида $y_t = \beta y_{t-1} + \varepsilon_t$. Тогда АКФ определяется формулой $\rho(h) = \beta^{|h|}$. Таким образом, мы можем провести оценку для β на основе уравнение $\rho_p(1) = \bar{\rho}(1)$, получая $\beta = \bar{\rho}(1)$. Для более сложных авторегрессионных моделей или моделей скользящего среднего MA(p) примеры оценок представлены в [1].

Оценка максимального правдоподобия. Предположим, что $\{y_t\}$ - стационарный гауссов процесс с нулевым средним. Функция логарифмического правдоподобия этого процесса определяется формулой:

$$L(\Theta) = -1/2 \operatorname{Log\,det} D_\Theta - 1/2 Y' D_\Theta^{-1} Y,$$

где, $Y = (y_1, y_2, ..., y_n)'$, $D_\Theta = \operatorname{Var}(Y)$ и Θ является вектором параметров.

Вид формулы определяет необходимость вычисления детерминанта и обратной матрицы дисперсии-ковариации D_Θ для определения максимума функции логарифмического правдоподобия. Эти расчеты могут быть проведены с помощью разложения Холецкого для матриц, алгоритма Дурбина-Левинсона и др.

Метод декомпозиции Холецкого. Учитывая, что матрица D_Θ является симметричной и положительно определенной матрицей, ее можно записать в виде:

$$D_\Theta = W'W,$$

где, W - верхняя треугольная матрица.

Согласно этому разложению Холецкого, определитель D_Θ можно вычислить по формуле:

$$\det D_\Theta = (\det W)^2 = \Pi(w_{jj})^2$$

где, w_{jj} обозначает j-й диагональный элемент матрицы W. Кроме того, обратная матрица D_Θ^{-1} может быть получена , как $D_\Theta^{-1}=W^{-1}(W^{-1})'$.

Алгоритм Дурбина-Левинсона. Разложение Холецкого может быть неэффективным для временных рядов и разработаны более быстрые методы вычисления логарифмической функции правдоподобия и одним из них является алгоритм Дурбина-Левинсона.

Алгоритм Дурбина-Левинсона использует структуру Теплица матрицы дисперсии-ковариации D_Θ.

Предположим, что $\bar{y}_1 = 0$ и $\bar{y}_{t+1} = \beta_{t1}y_t + .. + \beta_{tt}y_1$, t = 1,n-1, являются одношаговыми прогнозами процесса $\{y_t\}$ на основе конечного прошлого$(y_1, y_2, ..., y_{n-1})$, где коэффициенты регрессии β_{tj} задаются уравнениями:

$\beta_{tt} = [v_{t-1}]^{-1}(\gamma(t)+S)$, где $S=\Sigma\beta_{t-1,j}\gamma(t-j)$, j=1,t-1,

$\beta_{tj} = \beta_{t-1,j} - \beta_{tt}\beta_{t-1,t-j}$, j=1,t-1,

$v_0 = \gamma(0)$,

$v_t = v_{t-1}(1-(\beta_{tt})^2)$, j=1,t-1.

Кроме того, если $e_t = y_t - \bar{y}_t$ есть ошибка предсказания и $e = (e_1,..,e_n)'$, то E = LY, где L - нижняя треугольная матрица. Элементы матрицы L определяем следующим образом: $l_{ij}=-\beta_{i-1,n-j}$ при $i\geq j$ и $i>1$. Диагональные элементы $l_{ii}=1$ для i=1,n.

Отсюда D_Θ может быть представлена, как $D_\Theta = LI_VL'$, где I_V есть диагональная матрица с элементами $(v_0, v_t, .. v_{n-1})$.

Поэтому $\det D_\Theta = \Pi v_{jj}$, где j=0,n-1 и $Y' D_\Theta^{-1} Y = e'I_V^{-1}e$.

В результате логарифмическая функция правдоподобия может быть выражена как

$$L(\Theta) = -1/2\Sigma \operatorname{Log}(v_{t-1}) - 1/2\Sigma(e_t)^2/v_{t-1},$$

и границы суммирования t=1,n.

Временная сложность этого алгоритма для линейной стационарной последовательности равна $O(n^2)$. Для некоторых марковских процессов, таких как семейство ARMA, временная сложность алгоритм Дарбина-Левинсона может быть O (n). К сожалению, это сокращение числа операций невозможно для моделей ARFIMA, поскольку они не марковские.

Оценка долгосрочных процессов. Здесь мы обсудим некоторые специальные методы, разработанные для решения оценки долгосрочных зависимых временных рядов. Среди этих методов мы можем выделить процедуры максимального правдоподобия, основанные на приближениях авторегрессии (AR) и метода скользящего среднего (MA), регрессию логарифмической периодограммы, измененную масштабную статистику (R / S), анализ зависимости и анализ на основе вейвлет-метода.

Авторегрессионная аппроксимация. Так как вычисление точных оценок максимального правдободобия (MLE) требует больших вычислительных ресурсов, многие исследователи рассмотрели использование авторегрессионных приближений для ускорения расчета оценок параметров. Пусть $\{y_t: t\in Z\}$ процесс длительной памяти, определяемый авторегрессионным расширением:

$$y_t = \varepsilon_t + k_1(\beta)y_{t-1} + k_2(\beta)y_{t-2} + k_3(\beta)y_{t-3} + \dots ,$$

где, $k_j(\beta)$ – коэффициенты аппросимации. Поскольку на практике доступно лишь конечное количество наблюдений $(y_1, y_2, ..., y_n)$, рассматриваем усеченную модель для $m < t \leq n$:

$$y_t = \bar{\varepsilon}_t + k_1(\beta)y_{t-1} + k_2(\beta)y_{t-2} \dots + k_m(\beta)y_{t-m}.$$

Тогда приблизительная оценка максимального правдоподобия $\bar{\beta}_n$ получается путем минимизации функции:

$$L_1(\beta) = \Sigma(y_t - (k_1(\beta)y_{t-1} + k_2(\beta)y_{t-2} \dots + k_m(\beta)y_{t-m})^2$$

t = m+1,n.

Многие разработки могут быть сделаны на этой базовой структуре, чтобы получить лучшие оценки. Оценочный критерий, определяемый максимизацией аппроксимация функции гауссова правдоподобия называется оценкой квазимаксимального правдоподобия (QMLE).

Метод описания и оценки в переменных состояния. Эта общая методология может также использоваться для обработки авторегрессионного приближения. Например, рассматривая упрощенную авторегрессионную модель порядка m (AR (m)), получаем:

$$y_t = k_1y_{t-1} + k_2y_{t-2} + \dots + k_my_{t-m} + \varepsilon_t.$$

Отсюда, мы можем записать следующую систему в переменных состояния:

$$x_{t+1} = Fx_t + H\varepsilon_{t+1}, \quad y_t = Gx_t,$$

где, состояние задается вектором $x_t = [y_t, y_{t-1},\dots, y_{t-m+2}, y_{t-m+1}]'$.

Матрица перехода F имеет первую строку равную $(k_1,k_2,..,k_m)$, элементы $f_{i,i-1}=1$ для i=2,m. Все остальные элементы равны нулю.

Матрица наблюдений G = (1 0 0… 0). Матрица шума состояния дается формулой H = (1

0 0… 0) ′. Дисперсия шума наблюдения R = 0, ковариация между шумом состояния и шумом наблюдения S = 0, а ковариационная дисперсия матрица состояния Q имеет все нулевые элементы, кроме $q_{1,1}=\sigma^2$.

Скользящие средние аппроксимации. Альтернативной методологией приближений авторегрессии является усечение расширения Wold процесса с длинной памятью. Два преимущества этот подход представляет собой:

а) легкую реализацию рекурсивных фильтров Калмана;

б) простота анализа теоретических свойств ML по оценкам.

Кроме того, если временные ряды с большой памятью различаются, то результирующее скользящее среднее усечение имеет меньшую дисперсию ошибок, чем приближение авторегрессией.

Каузальное представление процесса ARFIMA (p; d; q) $\{y_t\}$ задается формулой $y_t = \Sigma \beta_j \varepsilon_{t-j}$, j=0,∞, но мы рассматриваем ее приближение $y_t = \Sigma \beta_j \varepsilon_{t-j}$, j=0,m. Таким образом мы вместо MA(∞) модели работаем с MA(m) моделью. Каноническое представление модели MA (m) моделью в пространстве состояний дается соотношениями:

$$x_{t+1} = Fx_t + H\,\varepsilon_t,\ y_t = Gx_t + \varepsilon_t,$$

где, $x_t = [y(t|t-1)\ y(t+1|t-1)\ \cdots\ y(t+m-1|t-1)]'$, а $y(t+j|t-1) = M[y_{t+j}|y_{t-1}, y_{t-2}, \cdots]$.

Матрицы модели системы в переменных состояния равны

$$G = (1\ 0\ \cdots\ 0),\ H = [\beta_1 \cdots \beta_m]',$$

а F имеет все элементы равны нулю, кроме $f_{i,i+1}=1$ для i=1,m-1.

Байесовская оценка. Рассмотрим некоторые применения байесовской методологии для анализа данных временных рядов. В нашем случае опишем байесовский подход для анализа процессов ARMA и ARFIMA с помощью марковской цепи (MCMC), важного вычислительного инструмента для получения образцов апостериорного распределения. В частности, мы описываем приложения алгоритма Метрополиса-Гастингса и сэмплера Гиббса для процессов с длинной памятью. Реализация этих вычислительных процедур проиллюстрирована на примере байесовской оценки стационарного гауссовского процесса. Важными при этом являются такие специфические вопросы, как выбор начальных значений и распределений.

Рассмотрим данные временного ряда $Y = (y_1, \ldots, y_n)'$ и статистическую модель, описываемую параметром θ. Пусть f (y|θ) - функция правдоподобия модели и π(θ) - предварительное распределение для параметра. Согласно теореме Байеса, эмпирическое распределение θ с учетом данных Y пропорционально:

$$\pi(\theta \mid Y) \propto f(Y \mid \theta)\,\pi(\theta).$$

Более конкретно, предположим, что временной ряд следует модели ARFIMA (p, d, q), описываемой $\varphi(B)(y_t - \mu) = \theta(B)(1-B)^{-d}\varepsilon_t$, где многочлены $\varphi(B) = 1 + \varphi_1 B + \cdots + \varphi_p B^p$ и $\theta(B) = 1 + \theta_1 B + \cdots + \theta_p B^q$ не имеют общих корней, а $\{\varepsilon_t\}$ - последовательность белого шума с нулевым средним значением и дисперсией σ^2. Обозначим $C_d = \{d: y_t$ стационарно и обратимо$\}$, $C_\varphi = \{\varphi_1, \ldots, \varphi_p: y_t$ стационарно$\}$, а $C_\theta = \{\theta_1, \ldots, \theta_q: y_t$ обратимо$\}$. Для этой модели вектор параметров может быть записан как:

$$\Theta = (d, \varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_q, \mu, \sigma^2),$$

и тогда пространство параметров может быть выражено как:

$$\Theta = C_d \times C_\varphi \times C_\theta \times IR \times (0, \infty).$$

Иногда, чтобы упростить спецификацию предварительного распределения по пространству параметров Θ, можно рассмотреть возможность назначения предыдущих распределений индивидуально подмножествам параметров. Например, мы можем предполагать предварительные равномерное распределения (U) для d, $\varphi_1, \ldots, \varphi_p$ и $\theta_1 \ldots, \theta_q$, т. е. $\Pi(d) = U(C_d)$, $\pi(\varphi_1, \ldots, \varphi_p) = U(C\varphi)$ и $\pi(\theta_1, \ldots, \theta_q) = U(C_\theta)$. Кроме того, мы можем предполагать неверным прежнее μ, $\pi(\mu) \propto 1$ и предшествующее $\pi(\sigma^2)$ для σ^2. С этим определением предварительное распределение θ является просто $\pi(\theta) \propto \pi(\sigma^2)$. Опытное распределение θ задается:

$$\pi(\theta \mid Y) \propto f(Y \mid \theta)\,\pi(\sigma^2).$$

Помимо расчета этого распределения, обычно важны определения оценок Байеса для θ. Например, можем рассмотреть возможность определения значения θ так, чтобы последующие потери были минимальными.

то есть, если L (θ, Y) - функция потерь, то:

$$\theta = \operatorname{argmin} \int L(\theta, Y)\,\pi(\theta \mid Y)\,dY.$$

В частном случае при квадратичных потерях $L(\theta, Y) = \|\theta - Y\|^2$ мы имеем, что оценка θ является опытным средним $\bar{\theta} = M[\theta|Y]$. Получение любой из этих величин требует интегрирования. Во многих практических ситуациях расчет этих интегралов может быть

чрезвычайно разным. Чтобы обойти эту проблему, в байесовской литературе были предложены несколько методологий, в том числе, численное интегрирование, моделирование методом Монте-Карло, аналитическое приближение Лапласа и методы Марковской цепи и Монте-Карло (MCMC). Мы сосредоточимся на методах MCMC.

Марковская цепь и метод Монте-Карло. Алгоритм MCMC создает образец представляющего интерес распределения методом, который объединяет методы Монте-Карло и цепи Маркова. Рассмотрим, например, что мы хотим получить образец заднего распределения π (θ|Y). Двумя хорошо известными процедурами для этой цели являются алгоритм Metropolis-Hastings и сэмплерГиббса.

Тренды. Наиболее распространенными характеристиками данных временных рядов являются наличие возрастающих или уменьшающихся трендов наряду с их разрывами. Основной вопрос, который возникает, состоит в том, являются ли эти тенденции результатом детерминированного базового шаблона или соответствует накоплению случайных шумов с течением времени. Конечно, наблюдаемые временные ряды могут быть результатом комбинаций этих двух или других более сложных механизмов генерации данных. Мы сейчас рассмотрим два хорошо известных подхода к пониманию и моделированию трендов: детерминированная и стохастическая методологии.

Детерминированные тренды. При детерминированном подходе наблюдаемый процесс является результатом обычно неизвестного основного шаблона f (t) и шума εt,

$y_t = f(t) + \varepsilon_t$.

Чтобы оценить тренд, функцию f(t) можно записать через некоторый вектор параметров β. Например, мы можем написать $f(t) = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_p x_{tp}$, где x_{t1}, x_{t2}, ..., x_{tp} - детерминированные ковариаты. В частности, устанавливая $x_{tj} = t^j$, мы можем генерировать полиномиальный тренд или полиномиальную регрессию. С другой стороны, если $x_{tj} = \exp(ij)$, то полученная модель соответствует гармонической регрессии.

Другой способ задания функции f (t) осуществляется через локальные полиномы, так что тренд достаточно гибкий, чтобы фиксировать краткосрочные изменения данных. В этом непараметрическом подходе тренд процесса y_t локально оценивается как $y_t = \Sigma w_j\, y_{t+j}$, $j \in (-m,+M)$ при t = m + 1, ..., N -m. Оптимальные веса w_j обычно получаются путем применения кубических многочленов для оценки зависимости в ряд y_t.

Стохастические тренды. Если основной тренд считается стохастическим, наблюдаемый процесс y_t обычно понимается как результат последовательности случайных ударов,

$y_t = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t$, где ε_t - белый шум или последовательность независимых случайных переменных. Заметим, что в этом случае заданная серия удовлетворяет $y_t - y_{t-1} = (1-B)\, y_t = \Delta\varepsilon_t$.

Таким образом, стохастический трендовый процесс в большей степени определяется:

$\Delta^d y_t = \varepsilon_t$,

где d - известный дифференциальный порядок, а ε_t - стационарный процесс.

Как правило, процесс, удовлетворяющий этому уравнению, называется интегрированным процессом порядка d, I (d). Предположим, что ε_t - последовательность из i.i.d. (т.е. случайные величины с нулевым средним и дисперсия σ^2).

Дисперсия процесса I(d) равна $\mathrm{Var}(y_t) = t\sigma^2$. Таким образом, интегрированный процесс обладает большой вариативностью, поскольку t стремится к бесконечности. Напротив, при тех же условиях дисперсия процесса с детерминированным трендом равна Var (yt) = σ. Естественно, в этом случае вариативность не является высокой. В этом смысле важно проверить, обладает ли процесс единичным корнем, то есть он соответствует интегрированному процессу.

Процессы АРИМА. Авторегрессионный интегрированный со скользящим средним процесс ARIMA (p, d, q) y_t определяется уравнением:

$\beta(B)\, \Delta^d y_t = \theta(B)\, \varepsilon_t$,

где, β (B) - авторегрессионный многочлен порядка p, а θ(B) представляет собой многочлен скользящего среднего порядка q, β (B) и θ (B) не имеют общих корней, а ε_t - последовательность белого шума. Заметим, что процесс $\Delta^d y_t$ удовлетворяет модели ARMA (p, q).

Модели скрытых цепей Маркова

Самый простой способ обработки последовательностных данных состоит в том, чтобы просто игнорировать последовательностные аспекты и рассматривать наблюдения как независимые одинаково распределенные случайные величины соответствующие рисунку 6. Однако при таком подходе невозможно рассматривать последовательностные явления в данных, такие как корреляции между наблюдениями, которые близки друг к другу во времени.

Рисунок 6 – Модель интерпретации наблюдения

Предположим, что мы наблюдаем двоичную переменную, обозначающую, был ли в определенный день дождем или нет. Учитывая временные ряды недавних наблюдений за этой переменной, мы хотим предсказать, будет ли дождь на следующий день. Если мы обрабатываем данные как независимые одинаково распределенные случайные величины, то единственной информацией, которую мы можем извлечь из данных, является относительная частота дождливых дней. Но на практике, мы знаем, что вероятность серии дождливых дней тенденция снижается в течение нескольких дней. Таким образом, наблюдение, идет ли сегодня дождь, - это значительная помощь в прогнозировании, будет ли завтра дождь. Чтобы выразить такие эффекты в вероятностной модели, одним из простейших способов нам сделать это является рассмотрение марковской модели.

Без ограничения общности мы можем использовать правило произведения для выражения совместного распределения для последовательности наблюдений в виде:

$$p(x_1, ..., x_n) = \Pi p(x \mid x_1, ..., x_{n-1}), n=1,N.$$

Если мы теперь предположим, что каждое из условных распределений в правой части не зависит от всех предыдущих наблюдений, кроме последнего, мы получаем цепь Маркова первого порядка, которая изображена на рисунке 7.

Рисунок 7 – Марковская цепочка наблюдений первого порядка

Совместное распределение для последовательности N наблюдений по этой модели дается формулой:

$$p(x1, ..., x_n) = p(x_1)\ \Pi p(x_n \mid x_{n-1}), n=2,N$$

Таким образом, если мы используем такую модель для предсказания следующего наблюдения, то распределение прогнозов будет зависеть от того, что будет иметь место непосредственно перед наблюдением и не будет зависеть от всех предыдущих наблюдений.

В большинстве применений таких моделей условные распределения $p(x_n|x_{n-1})$, определяющие модель, соответствует предположению о стационарном временном ряду. Модель тогда известна как однородная цепь Маркова. Например, если условные распределения зависят от настраиваемых параметров (значения которых могут быть выведены из набора данных обучения), то все условные распределения в цепочке будут иметь одни и те же значения этих параметров. Хотя это более общая, чем модель независимости, она по-прежнему очень ограничена.

Для многих последовательностных наблюдений мы ожидаем, что тенденции в данных по нескольким предыдущим наблюдениям предоставят важную информацию для прогнозирования следующего значения. Один из способов позволить более ранним наблюдениям иметь влияние - перейти к марковским цепям более высокого порядка. Если мы допустим, что предсказания будут зависеть и от предыдущего значения, мы получим цепь Маркова второго порядка, представленную на рисунке 8.

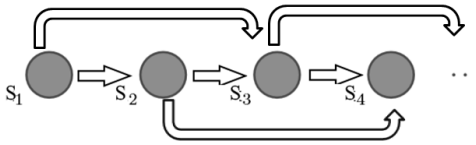


Рисунок 8 – Марковская цепочка наблюдений второго порядка

Совместное распределение для последовательности N наблюдений по этой модели дается формулой:

$$p(x_1, ..., x_n) = p(x_1)\ p(x_1|x_2)\ \Pi p(x_n|x_{n-1}, x_{n-1}), n=3,N.$$

Аналогичным образом мы можем рассмотреть расширения к цепи Маркова M-го порядка, в которой условное распределение для конкретной переменной зависит от предыдущих M переменных. Мы заплатили за эту общность, поскольку количество параметров в модели теперь намного больше. Предположим, что наблюдения являются дискретными переменными с K состояниями. Тогда условное распределение $p(x_n \mid x_{n-1})$ в марковской цепочке первого порядка будет задано набором из K-1 параметров для каждого из K-состояний x_{n-1}, дающим общее количество параметров K (K -1). Предположим теперь, что мы распространим модель на цепь Маркова M-го порядка, так что совместное распределение строится из условий $p(x_n \mid x_{n-M}, ..., x_{n-1})$. Если переменные дискретны, и если

условные распределения представлены общими условными таблицами вероятности, то число параметров в такой модели будет иметь параметры K^{M-1} (K -1). Поскольку это выражение растет экспоненциально с M, этот подход явно нецелесообразен для больших значений M.

Для непрерывных переменных мы можем использовать линейно-гауссовые условные распределения, в которых каждый узел имеет гауссовское распределение, среднее значение которого является линейной функцией его или родителей. Это известно как, авторегрессионная AR-модель.

Мы можем ввести дополнительные скрытые переменные, чтобы можно было построить богатый класс моделей из простых компонентов. Для каждого наблюдения x_n вводится соответствующая скрытая переменная z_n (которая может быть разного типа или иной размерности относительно наблюдаемой переменной).

Если считать, что эти скрытые переменные образуют цепь Маркова, то имеем известную модель пространства состояний, которая показана на рисунке 9.

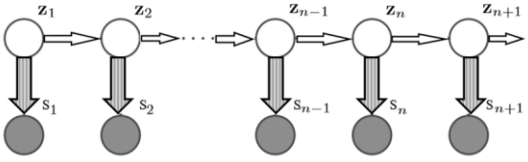


Рисунок 9 – Марковская цепочка для скрытых переменных

Совместное распределение для этой модели дается выражением:

$$p(x_1, ..., x_N, z_1, ..., z_N) = p(z_1)\,[\Pi p(z_n | z_{n-1}), n=2,N]\Pi p(x_k | z_k),\ k=1,N.$$

Мы видим, что всегда существует путь, соединяющий любые две наблюдаемые переменные x_n и x_m через скрытые переменные, и что этот путь никогда не блокируется.

Таким образом, предсказательное распределение $p(x_{n+1} | x_1, ..., x_n)$ для наблюдения x_{n+1} при всех предыдущих наблюдениях не имеет никаких условных свойств независимости, поэтому наши предсказания для x_{n+1} зависят от всех предыдущих наблюдений.

Однако наблюдаемые переменные не удовлетворяют марковскому свойству в любом порядке. Существуют две важные модели для последовательных данных, которые описываются рисунком 9.

Если скрытые переменные дискретны, то мы получаем скрытую марковскую модель или (HMM). Если и скрытые, и наблюдаемые переменные являются гауссовыми, то мы имеем линейную динамическую систему.

Как указано ранее, скрытую марковскую модель можно рассматривать, как конкретный экземпляр модели пространства состояний на рисунке 9, в которой скрытые переменные дискретны. Однако, если мы рассмотрим один временной срез модели, мы увидим, что оно соответствует вероятностному смешанному распределению смеси с плотностями компонент, заданными p(x|z). Поэтому его можно также интерпретировать как расширение смешанной модели, в которой выбор компонента смеси для каждого наблюдения не выбирается независимо, а зависит от выбора компонента для предыдущего наблюдения. HMM широко используется в распознавании речи, моделировании естественного языка, он-лайн распознавание рукописного ввода, анализу биологических последовательностей, таких как белки, ДНК и т.п.. Как и в случае стандартной модели смеси, скрытые переменные являются дискретными многочленными переменными z_n, описывающими, какая компонента смеси отвечает за генерирование соответствующего наблюдения.

Мы определяем распределение вероятности z_n в зависимости от состояния предыдущей скрытой переменной z_{n-1} через условное распределение $p(z_n | z_{n-1})$. Поскольку скрытые переменные являются K-мерными двоичными переменными, это условное распределение соответствует таблице чисел, которую мы обозначаем A, элементы которой известны как вероятности перехода. Они задаются $A_{jk} \equiv p(z_{nk} = 1 | z_{n-1,j} = 1)$, а так, как они являются вероятностями, они удовлетворяют условиям $0 \le A_{jk} \le 1$ и $\Sigma A_{jk} = 1, k=1,K$, так что матрица A имеет K (K-1) независимых параметра.

Матрицу перехода иногда иллюстрируют диаграммой, рисуя состояния как узлы на диаграмме перехода состояния, как показано на рисунке 10 для случая K = 3. Подчеркнем, что узлы не являются отдельными переменными, а состояниями одной переменной, поэтому мы показали состояния как прямоугольники, а не кружки.

Таким образом, мы имеем модель, чьи скрытые переменные имеют три возможных состояния. Иногда полезно использовать диаграмму перехода состояний, показанную на рисунке 10, и разворачивать ее со временем.

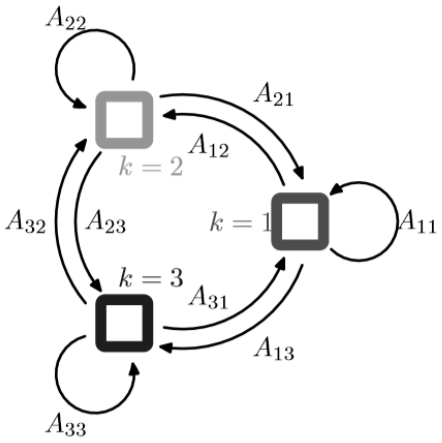


Рисунок 10 – Диаграмма перехода для модели с К=3.

Черные линии обозначают элементы переходной матрицы A_{jk}.

Если развернуть диаграмму перехода состояний на рис. 10 с течением времени, мы получим решетчатое представление скрытых состояний. Каждый столбец этой диаграммы соответствует одной из скрытых переменных z_n.

Это дает альтернативное представление переходов между скрытыми состояниями, известными как решетчатая диаграмма, и конкретная диаграмма показана для случая скрытой марковской модели на рисунке 11.

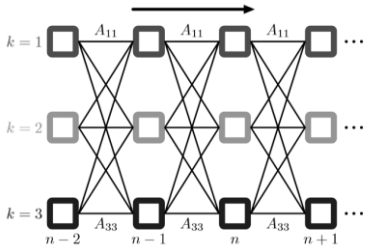


Рисунок 11 – . Развернутая диаграмма перехода для модели с К=3.

Реализованные алгоритмы определения параметров модели скрытых цепей Маркова

Одной из основных практических задач трейдинга на рынке акций является прогнозирование цен. Оно используется для принятия решения об открытии и закрытии позиции и для количественной оценки ценовых рисков.

Поэтому любая торговая система на фондовом рынке состоит из 3 основных функциональных блоков: блок прогнозирования цен и выработки сигналов к открытию / закрытию позиций; блок управления капиталом; блок управления рисками.

В качестве исследуемых данных использовались значения котировок индекса Доу-Джонса за 2016 год. Прогноз на следующий временной интервал (следующий день) осуществлялся на основе следующих двух алгоритма.

Алгоритм А.

а) Задаем файл исходных данных;

б) Определяем количество исходных точек (N) и число состояний (quantityInterv); Готовим массив последовательности (preData) с длиной, равной количеству исходных точек, con – матрица с quantityInterv строк и quantityInterv столбцов. Выполняем обнуление матриц.

в) Переменная quantityInterv определяет количество интервалов, на которые будет разделена исходная последовательность данных и рассчитывается как остаток от деления количества исходных точек N на число состояний k;

г) С помощью цикла определяем количество точек, попавших в каждый интервал, остаток разбрасывается в первые интервалов;

д) Далее устанавливаем границы интервалов и выполняем разброс выходных данных согласно установленным ранее границам;

е) Вычисляем вероятности перехода из одного состояния в другое;

ж) Выбирается кластер состояний максимального размера, то есть группа соседних состояний с максимальной вероятностью, на основе которого рассчитывается прогнозное значение на следующий временной интервал и вероятности изменения курса акций.

Алгоритм Б

Во втором варианте расчета весь диапазон изменения сигнала:

а) разбивается на интервалы определенной ширины (interv), начиная от минимального (MIN) уровня и завершая максимальным (MAX), каждый интервал считается состоянием Марковской цепи, и характеризуется численным значением;

б) сдвиг для разделения на интервалы определяем как разницу между минимальным и максимальным значениями разделенную на число состояний;

в) определяем начальное и конечное значение интервалов, что позволит включить первое и последнее значение исходной выборки при расчетах;

г) далее с помощью рассчитанного

ранее сдвига определяем границы интервалов, подсчитывается количество точек, попавших в каждый интервал, на основе этой информации определяются вероятности перехода из одного состояния в другое и строится матрица переходных вероятностей. Для прогнозирования следующего ценового уровня определяем, в каком состоянии попало это значение, берем вычисленные на предыдущем шаге вероятности переходов из этого состояния, взвешиваем по этим вероятностям значения приращений и получаем прогнозируемое значение.Примеры расчетов в системе КСИКР представлены ниже.

Модели параметрического класса

Важное различие между статистическими процедурами связано с параметрическими и непараметрическими моделями. Можно представить данные как исходящие из модели, обычно неизвестной, которая определяется рядом коэффициентов или параметров (парадигма Фишера). В этом контексте статистический анализ по существу угадывает, какие параметры модели создают наблюдаемые данные. Для достижения этой цели необходимо выбрать модель и оценить соответствующие параметры. Примерами этих процедур являются модели авторегрессионного скользящего среднего (ARMA). Для указания параметрической модели можно предоставить вектор параметров $\theta = (\theta_1, ..., \theta_p)$ и пространство параметров Θ, такое, что $\theta \in \Theta$. Заметим, что размерность параметра θ конечна и равна p. В качестве иллюстрации рассмотрим простую модель

$$y_t = \varepsilon_t + \beta\varepsilon_{t-1},$$

где $\{\varepsilon_t\}$ - последовательность белого шума, а β - одномерный параметр. Это пример так называемых моделей скользящего среднего. В этом случае модель может быть задана двумерным вектором $\theta = (\beta, \sigma)$, где σ - стандартное отклонение белого шума. Кроме того, пространство параметров задается $\Theta = (R, R^+)$, где R^+ обозначает положительные вещественные числа. Обобщение этой простой модели мы имеем рассматривая несколько параметров

$$y_t = \varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + \cdots + \beta_q\varepsilon_{t-q},$$

которая обозначается как модель скользящего среднего MA (q) .

В этом случае вектор параметра $\theta = (\beta_1, \beta_2, ..., \beta_q, \sigma)$.

Еще более общее расширение позволяет коэффициентам β_j зависеть от конкретного конечномерного параметра, $\beta_j(\theta)$ и можем записать

$$Yt = \Sigma\beta_j(\theta)\,\varepsilon_{t-j}\ ,\ j=0,+\infty.$$

В этом случае, хотя имеем бесконечное число коэффициентов $\beta_j(\theta)$, модель будет параметрической, так как они зависят от конечномерного параметра θ.

Накладывание субъективной параметрической модели на наблюдаемые данные чревато неправильными выводами и на это обращал внимание Ивахненко в связи с разработкой метода МГУА.

Далее мы приводим пример из работы [16], который показывает принципиальные трудности при использовании параметрических моделей и в частности общую ограниченность метода последовательного выделения периодических компонент при идентификации параметров квазипериодических последовательностей.

Так с помощью метода наименьших квадратов зачастую выделяют основную периодическую компоненту и далее последовательно определяют характеристики следующих по важности периодических составляющих, для которых исходными данными является временной ряд, получаемый из исходной последовательности путем вычитания найденных ранее компонент.

Однако простой пример показывает недостатки использования этого подхода. Так рассматривая сумму двух периодических последовательностей $\sin(x)+\sin(x/\sqrt{3})$ на рисунке 12, мы можем сказать, что она является квазипериодической, но не периодической [17, с. 219].

Последний математический факт указывает на невозможность нахождения точного определения параметров квазипериодической последовательности линейной комбинацией с периодическими компонентами.

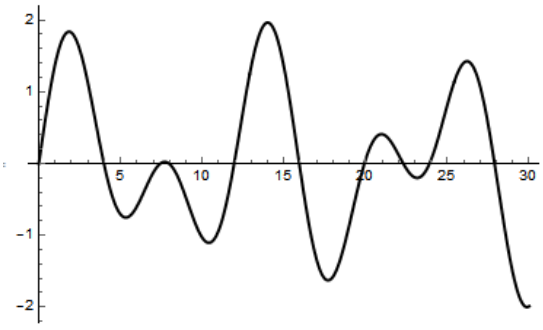


Рисунок 12 – $Y(x)= \sin(x)+\sin(x/\sqrt{3})$

Если мы попытаемся определить ряд значений функции $Y(x)=\sin(x)+\sin(x/\sqrt{3})$ при x=1,30 моделью $Y1(x)=A+B\cos(w\cdot x)+C\sin(w\cdot x)$, получим следующие результаты A = 0.0416901, B

= 1.0068, C = -0.37931, w=1.02416 и дисперсия равна 14.5888. График ошибки идентификации Y(x)-Y1(x) представлен на рисунке 13а.

Находим далее для ряда значений последовательности Y(x)-Y1(x) параметры ее модели Y2(x)=A2+B2*cos(w2·x)+C2*sin(w2·x). Имеем следующие результаты A2= -0.0368, B2= 0.00751, C2=0.9486, w2=0.576 и дисперсия равна 0.6668. График ошибки идентификации Y(x)-Y1(x)-Y2(x) представлен на рисунке 13б.

Заметим, что точные частоты исходной последовательности равны 1 и 0.57735, а следуя последовательной методике, мы получили 1.02416 и 0.576 с гораздо большей дисперсией . Основной вывод анализа этого примера, что следуя указанному выше подходу, мы не получим точных значений частот для подобного рода последовательностей (это возможно только для подобранных специальным образом данных).

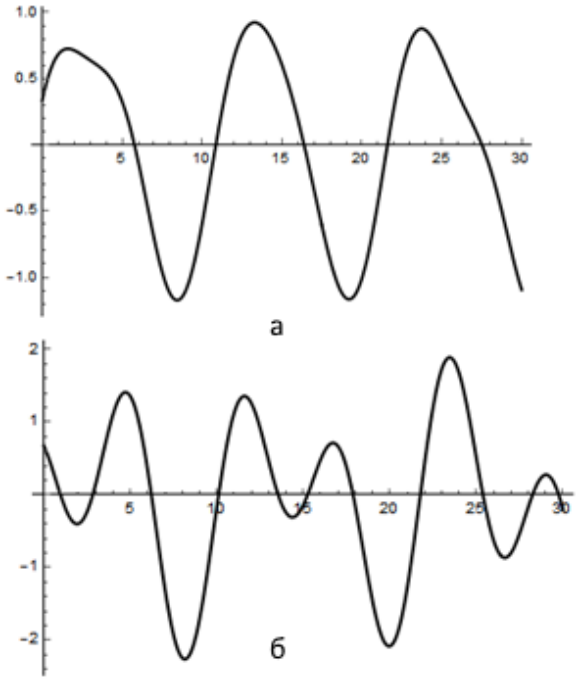


Рисунок 13 –Ошибка:а) Y(x)-Y1(x)
б) Y(x)-Y1(x)-Y2(x)

Если мы сразу находим параметры модели с дополнительным ограничением $w>w_2$, A+B*cos(wx)+C*sin(wx)+B2*cos(w_2x)+C2*sin(w_2x), то получаем почти идеальные значения A=-3.25763·10^{-11}, B=1.0, C=5.2498·10^{-10}, w=1.0, B2=1.0, C2=-5.31027·10^{-9}, w_2=0.57735 и дисперсия равна 1.74939·10^{-16}.

График ошибки идентификации представлен на рисунке 14 и обратим внимание на масштаб.

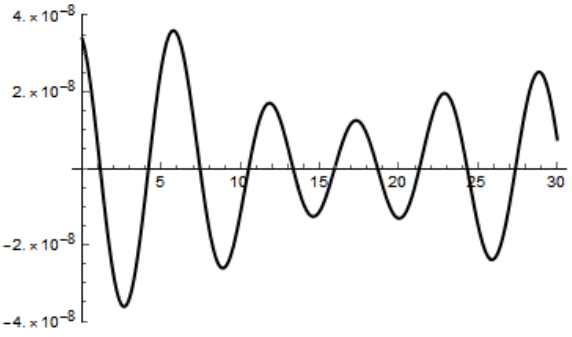


Рисунок 14 – Ошибка идентификации

В случае, когда для различения компонент с частотами w,w_2 используем ограничение по амплитудам $B^2+C^2>B1^2+C1^2$ для определения параметров модели A+B*cos(wx)+C*sin(wx)+B2*cos(w_2x)+C2*sin(w_2x), сразу приведем сравнение ошибок идентификации на рисунке 15(вторая линия-, график Y(x)-Y1(x)-Y2(x) на рис.13б), согласно которому различие результатов не так существенно.

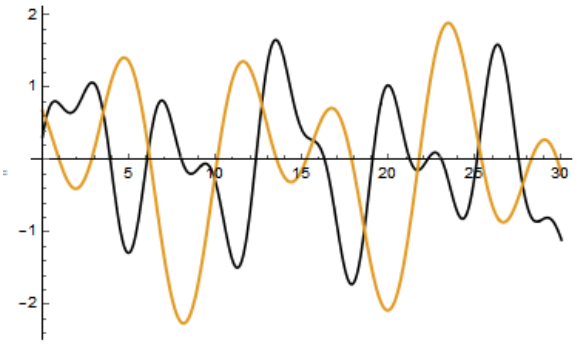


Рисунок 15 – Сравнение ошибок

Вышеизложенное указывает, что использование методики последовательного выделения параметрических периодических составляющих возможно при хорошем знании специфики проблемной области .

Выводы

В статье был выполнен обзор и сравнение методов исследования, получения характеристик и прогноза квазипериодических временных рядов.

На основе этих методов и алгоритмов была построена компьютерная система исследования и прогноза квазипериодических рядов на базе математического пакета Wolfram Mathematica. Система КСИКР объединяет в себе уже существующие программные разработок в математическом пакете Wolfram Mathematica(к

примеру, вейвлетные преобразования[13-15],18) и разработанные собственные модели прогноза квазипериодических временных рядов.

В системе КСИКР реализована возможность получения самых последних данных о более чем 4 тысячах финансовых показателей 88 тысяч различных компаний. Данные поступают через Интернет, данная возможность реализована благодаря конструкции сверхвысокого уровня – FinancialData. Данная конструкция позволяет получить данные по акциям, валютным курсам, индексам и другим финансовым инструментам.

Использование полученной программной системы позволяет разносторонне проанализировать и спрогнозировать самые актуальные данные различных мировых компаний, что позволит различным предприятиям и фирмам оптимизировать динамику и курс своего развития. Некоторые выходные результаты системы представлены на рис.16-

Система КСИКР может быть адаптирована для обработки метеорологических данных предоставляемых другой конструкцией сверхвысокого уровня – WeatherData. Приведённая выше конструкция предоставляет текущие и исторические данные о погоде со всех стандартных метеорологических станций по всему миру.

Система может быть расширена для обработки данных других сфер деятельности человека, как социальные, гидрологические и т.п.

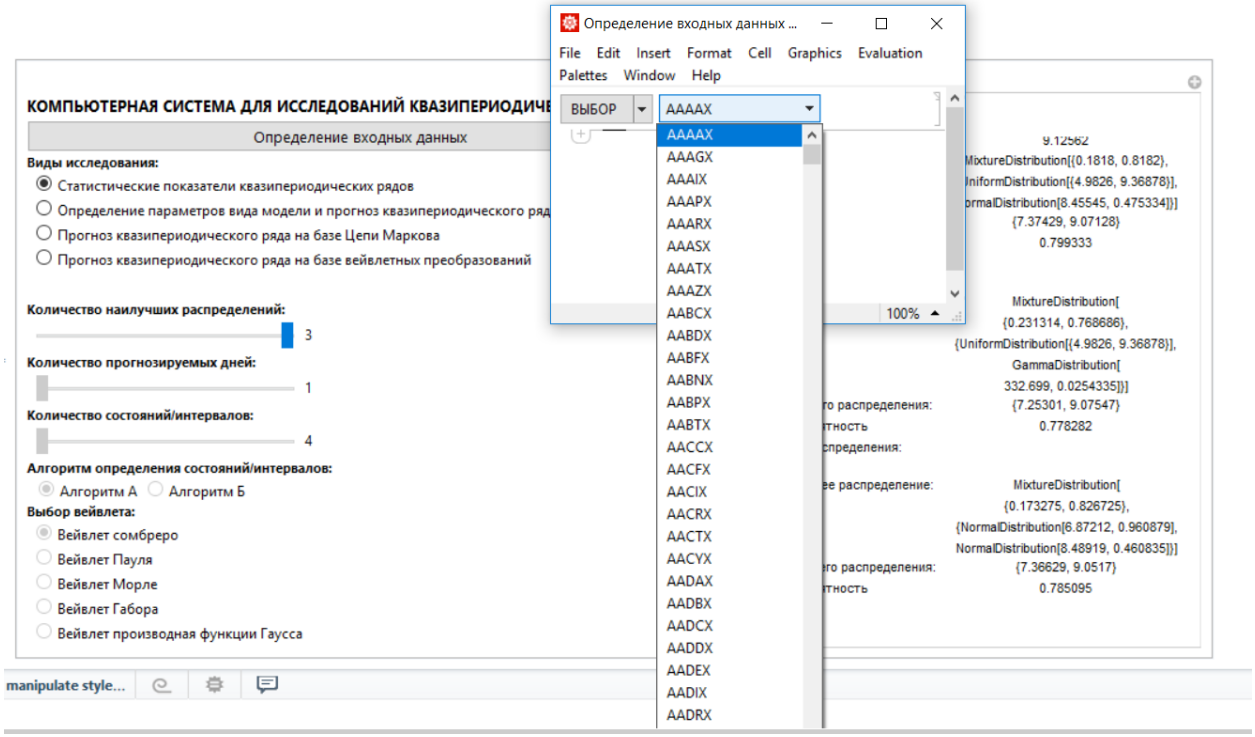
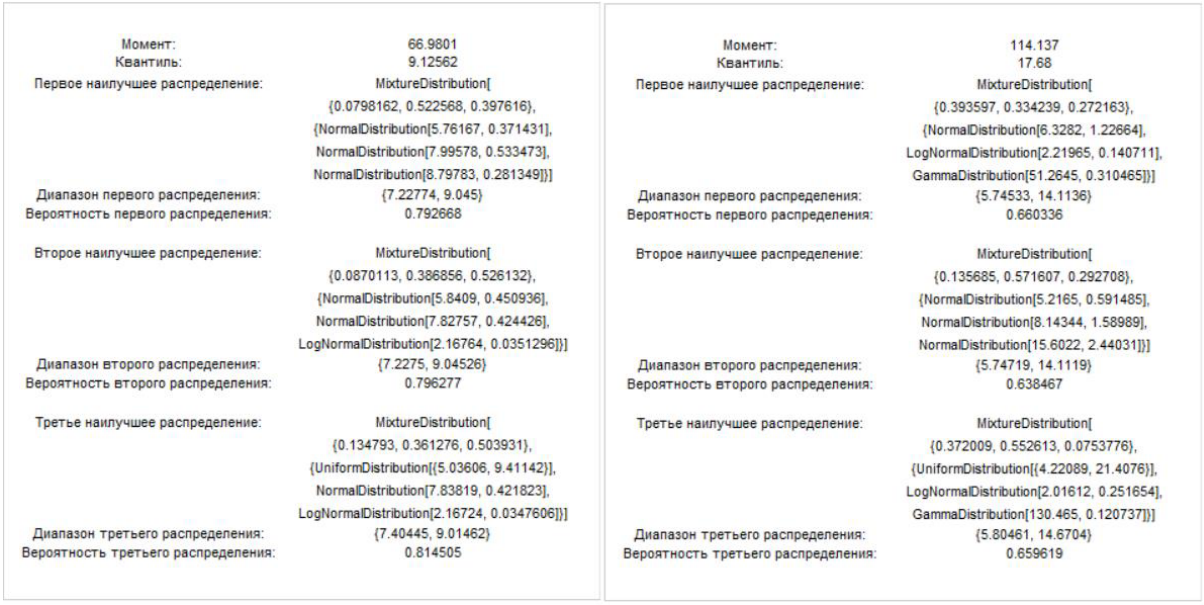


Рисунок 16 – Определение входных данных в FinancialData в системе КСИКР



Момент:	66.9801
Квантиль:	9.12562
Первое наилучшее распределение:	MixtureDistribution[{0.0798162, 0.522568, 0.397616}, {NormalDistribution[5.76167, 0.371431], NormalDistribution[7.99578, 0.533473], NormalDistribution[8.79783, 0.281349]}]
Диапазон первого распределения:	{7.22774, 9.045}
Вероятность первого распределения:	0.792668
Второе наилучшее распределение:	MixtureDistribution[{0.0870113, 0.386856, 0.526132}, {NormalDistribution[5.8409, 0.450936], NormalDistribution[7.82757, 0.424426], LogNormalDistribution[2.16764, 0.0351296]}]
Диапазон второго распределения:	{7.2275, 9.04526}
Вероятность второго распределения:	0.796277
Третье наилучшее распределение:	MixtureDistribution[{0.134793, 0.361276, 0.503931}, {UniformDistribution[{5.03606, 9.41142}], NormalDistribution[7.83819, 0.421823], LogNormalDistribution[2.16724, 0.0347606]}]
Диапазон третьего распределения:	{7.40445, 9.01462}
Вероятность третьего распределения:	0.814505

Момент:	114.137
Квантиль:	17.68
Первое наилучшее распределение:	MixtureDistribution[{0.393597, 0.334239, 0.272163}, {NormalDistribution[6.3282, 1.22664], LogNormalDistribution[2.21965, 0.140711], GammaDistribution[51.2645, 0.310465]}]
Диапазон первого распределения:	{5.74533, 14.1136}
Вероятность первого распределения:	0.660336
Второе наилучшее распределение:	MixtureDistribution[{0.135685, 0.571607, 0.292708}, {NormalDistribution[5.2165, 0.591485], NormalDistribution[8.14344, 1.58989], NormalDistribution[15.6022, 2.44031]}]
Диапазон второго распределения:	{5.74719, 14.1119}
Вероятность второго распределения:	0.638467
Третье наилучшее распределение:	MixtureDistribution[{0.372009, 0.552613, 0.0753776}, {UniformDistribution[{4.22089, 21.4076}], LogNormalDistribution[2.01612, 0.251654], GammaDistribution[130.465, 0.120737]}]
Диапазон третьего распределения:	{5.80461, 14.6704}
Вероятность третьего распределения:	0.659619

Рисунок 17 – Вывод статических показателей квазипериодических рядов в системе КСИКР для двух разных компаний

```
ftype == 2,
tsm = TimeSeriesModelFit[preData];
      |поиск модели временного ряда
forecast = TimeSeriesForecast[tsm, {switchsecond}];
           |прогнозировать значение временного ряда
normforecast = Normal[forecast];
               |нормальное выражение
rezult = {Grid[{{"Модель временного ряда: ", tsm}},
          |таблица
   ItemStyle → {Black, 10, Italic}],
    |стиль элемента |чёрный      |курсив
  ListPlot[{Flatten[preData], forecast},
  |диаграмма··· |уплостить
   Joined → True, Filling → Bottom]},
   |соедин··· |истина |заливка    |низ
```

Рисунок 18 – Реализация исследования определение параметров вида модели и прогноз квазипериодического ряда в системе КСИКР

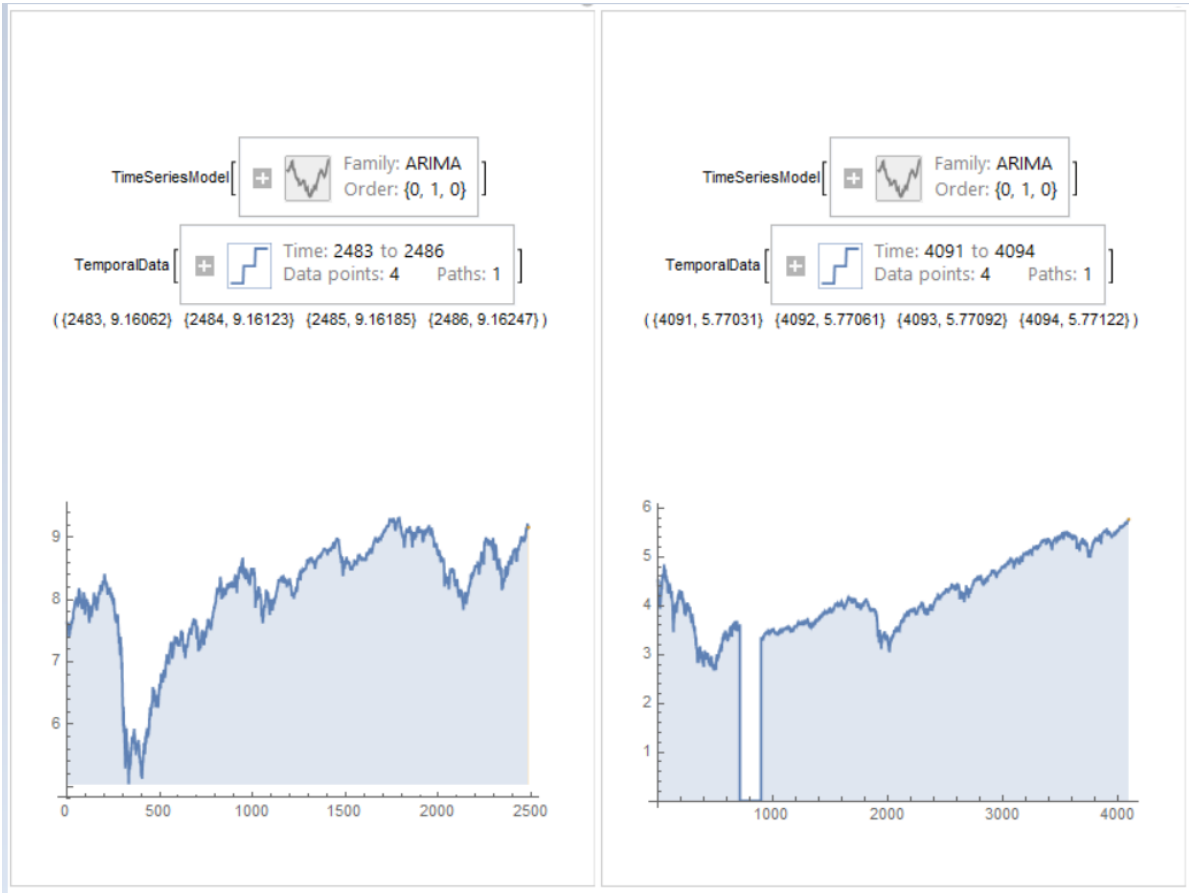


Рисунок 19 – Вывод определения параметров вида модели и прогноза квазипериодического ряда в системе КСИКР для двух разных компаний

Рисунок 20 – Результаты прогнозирования квазипериодического ряда на базе цепи Маркова по алгоритму А в системе КСИКР для двух разных компаний

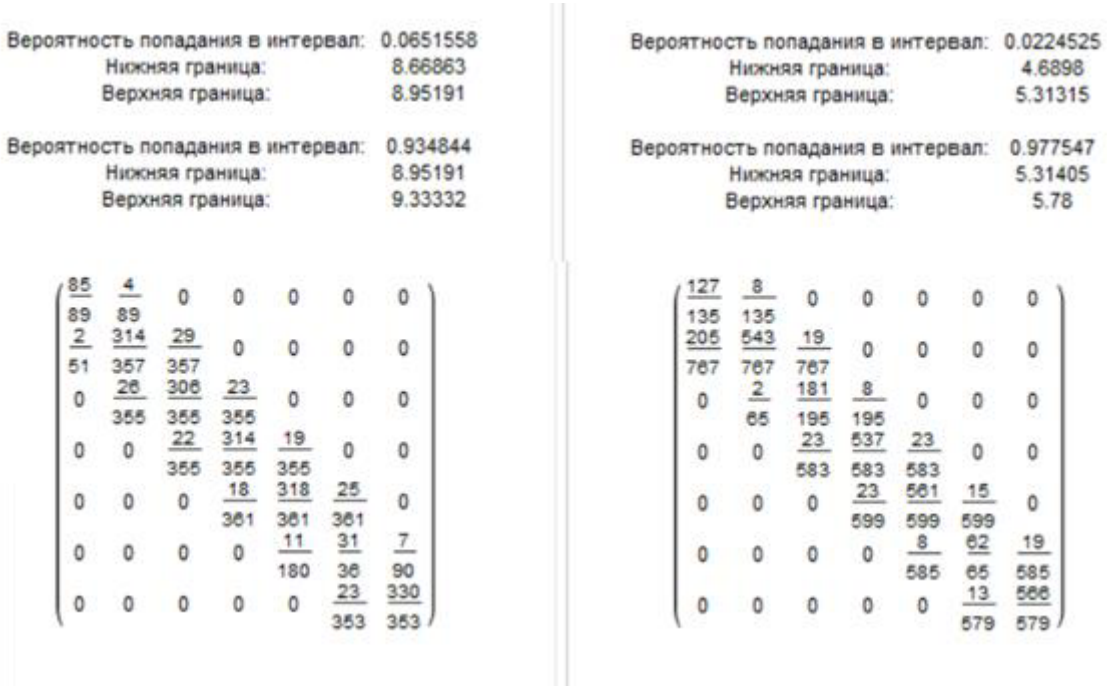


Рисунок 21 – Результаты прогнозирования квазипериодического ряда на базе цепи Маркова по алгоритму Б в системе КСИКР для двух разных компаний

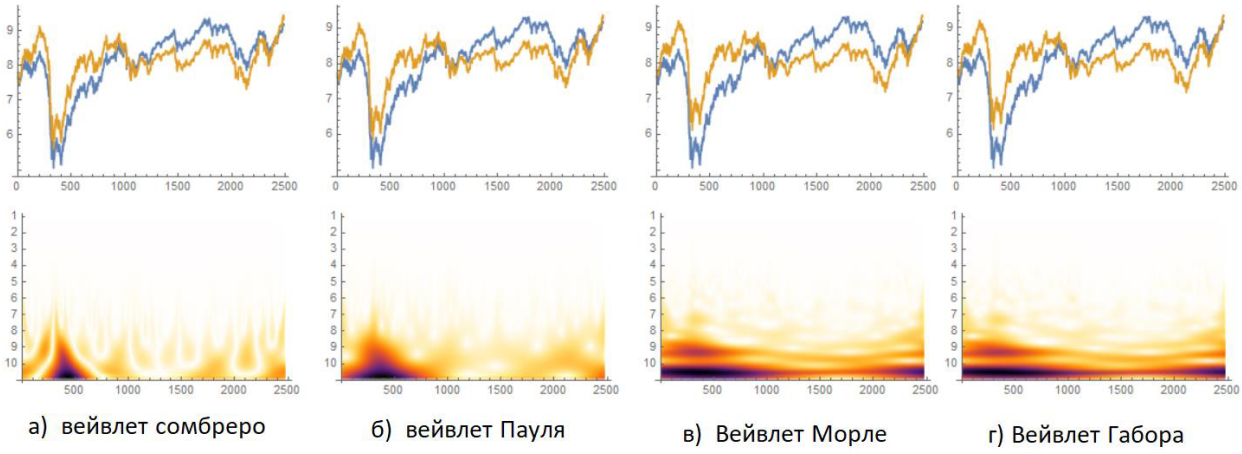


Рисунок 22 – Результаты прогнозирования квазипериодического ряда на базе вейвлетных преобразований в системе КСИКР для первой компании

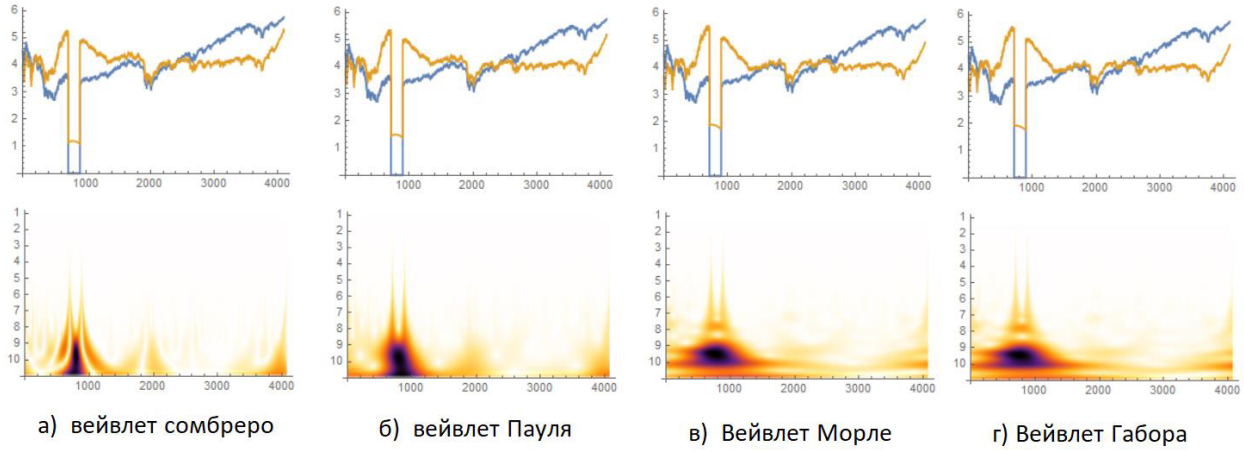


Рисунок 23 – Результаты прогнозирования квазипериодического ряда на базе вейвлетных преобразований в системе КСИКР для второй компании

Литература

1. Palma, W. Time series / W. Palam // Published by John Wiley & Sons, Inc., Hoboken, New Jersey, 2016

2. Канторович, Г. Г. Анализ временных рядов / Г. Г. Контарович // Экономический журнал ВШЭ, № 1, 2002. С.85-103

3. Montgomery, D. C. Introduction to Time Series Analysis and Forecasting./ D. C. Montgomery, C. L. Jennings, M. Kulanci.; Published by John Wiley & Sons. Inc .. Hoboken. New Jersey.

4. Медведев, Г. А. Практикум на ЭВМ по анализу временных рядов [Электронный ресурс] / Г. А. Медведев, В .А. Морозов; Учебное пособие. — Электрон. текст. дан. (1780 кб). — Мн.: “Электронная книга БГУ”, 2003. — Режим доступа: http://anubis.bsu.by/publications/elresources/Applied Mathematics/morozov.pdf

5. Brockwell, Peter J. Introduction to time series and forecasting / Peter J. Brockwell and Richard A. Davis.—2nd ed. p. cm. — (Springer texts in statistics), © 2002, 1996 Springer-Verlag New York, Inc.

6. Индекс Dow Jones: архив значений, экспорт в Exel, построение графиков [Электронный ресурс] – 2017 – Режим доступа: http://investfunds.ua/markets/indicators/indeks-dow-jones – Загл. с экрана|.

7. График среднемесячных чисел Вольфа [Электронный ресурс] – 2017 – Режим доступа: http://meteo-dv.ru/geospace/AverageMonthW – Загл. с экрана|.

8. Ежедневные числа Вольфа [Электронный ресурс] – 2017 – Режим доступа: http://www.gao.spb.ru/database/csa/daily_wolf_r.html – Загл. с экрана|.

9. Sunspot Nuber [Электронный ресурс] – 2017 – Режим доступа: http://www.sidc.be/silso/datafiles – Загл. с экрана|.

10. Мировой продукт как опережающий показатель индекса цен на нефть по циклам солнечной активности [Электронный ресурс] – Режим доступа: http://perfume007.livejournal.com/21002.html – Загл. с экрана|. – (08.01.17)

11. Time-series Forecasting Methods [Электронный ресурс] – 2017 – Режим доступа: http://www.ipredict.it/ForecastingMethods/ – Загл. с экрана|.

12. Bishop, C. M. Pattern Recognition and Machine Learning. / C. M. Bishop; © 2006 Springer Science+Business Media, LLC

13. Дремин, И. М. Вейвлеты и их использование / И. М. Дремин, О. В. Иванова, В. А. Нечитайло // Успехи физических наук, 2001. – Том 171, №5. – С. 465-501

14. Яковлев, А. Н. Введение в вейвлет-преобразования: учеб. Пособие / А. Н. Яковлев, – Новосибирск: Изд-во НГТУ, 2003. – 104 с.

15. Астафьева, Н. М. Вейвлет-анализ: основы теории и примеры применения / Н. М. Астафьева // Успехи физических наук, 1996. – Том 166, №11. – С. 1145-1170

16. Андрюхин, А. И. Компьютерный анализ

свойств решений ряда задач/ А. И. Андрюхин // Системный анализ в науках о природе и обществе, Донецк, Доннту, №1(4)-2(5)'2013. – С.39-45.

17. Гелбаум, Б. Контрпримеры в анализе. / Б. Гелбаум, Дж. Олмстед, – М.: Мир. – 1967.

18. Fractals, Wavelets, and their Applications. Contributions from the International Conference and Workshop on Fractals and Wavelets. / C. Bandt , M, Barnsley, R. Devaney, K. J. Falconer, V. Kannan , V. Kumar. – Editors. Springer International Publishing Switzerland 201

Andruckin A.I., V.S.Marchenko. Computer research and forecast of quasiperiodic series. *In this work, we review the methods and algorithms of the study, obtain characteristics and forecast quasiperiodic time sequences. A software system for studying, determining the properties and prognosis of quasiperiodic series (CSIQS) is constructed. Wolfram Mathematica is the basis for building CSIQS. The main attention is paid to the software implementation of such models and methods of investigation of quasiperiodic series, such as hidden Markov chains and wavelet transformations. The well-known fact that the sum of periodic functions may not be a periodic function is considered. Examples of calculations and the basic structure of CSIQS are presented.*

Keywords*: quasi-periodic series, forecast, hidden Markov chains, wavelet, Wolfram Mathematica.*

Андрюхин А.И., В.А.Марченко. Компьютерное исследование и прогноз квазипериодических рядов. В работе выполнен обзор методов и алгоритмов исследования, получения характеристик и прогноза квазипериодических временных последовательностей. Построена программная система исследования, определения свойств и прогноза квазипериодических рядов(КСИКР). Wolfram Mathematica является базой для построения КСИКР. Основное внимание уделено программной реализации таких моделей и методов исследования квазипериодических рядов, как скрытые цепи Маркова и вейвлетные преобразования. Выполнен анализ известного факта, что сумма периодических функций может не быть периодической функцией. Представлены примеры расчетов и основная структура КСИКР.

Ключевые слова*: квазипериодические ряды, прогноз, скрытые цепи Маркова, вейвлет, Wolfram Mathematica.*

Статья поступила в редакцию 21.4.2017
Рекомендована к публикации д-ром физ.-мат.. наук А.С. Миненко